Stochastic Processes

Daniel Adam Steck
Department of Physics, University of Oregon

Original revision posted ?? 2019.

This is a working draft, compiled 21 March 2023.

Cite this document as:
Daniel A. Steck, *Quantum and Atom Optics*, available online at `http://steck.us/teaching` (revision 0.0.0, 00 Schmanuary 2019).

Author contact information:
Daniel Steck
Department of Physics
1274 University of Oregon
Eugene, Oregon 97403-1274
dsteck@uoregon.edu

Contents

Part I

Fundamental Concepts

Chapter 1

Probability Theory

1.1 Basics of Probability

To define a probability distrbution axiomatically, a **probability space sample space** is simply a set Ω whose elements are called **outcomes**, which represent possible results of experiments with uncertain results. Outcomes are grouped into **events**, which are sets of outcomes (and subsets of the probability space). In the case of a finite probability space Ω, any subset of Ω is admissible as an event. However, in the case of infinite (in the sense of a continuum) spaces, we need to be more careful.

A ***σ*-algebra** on Ω is a collection Σ of subsets of Ω that satisfies the following axioms:

1. $\Omega \in \Sigma$.
2. if $A \in \Sigma$ then $\Omega - A \in \Sigma$. (And thus $\emptyset \in \Sigma$.)
3. the union of every countable subset of Σ is an element of Σ.

An example of a σ-algebra for $\Omega = \mathbb{R}$ is the **Borel algebra**, which consists of all open intervals of the form (a, b) with $a < b$, all possible countable unions and countable intersections of the open intervals, and all the corresponding complements. The idea of a σ-algebra is to exclude the possibility of pathological subsets of Ω—roughly speaking, those for which length or volume is ill-defined. The elements of Σ are the events that we mentioned before.

Now a **probability measure** P on a σ-algebra Σ on a probability space Ω is a function $P : \Sigma \longrightarrow \mathbb{R}$, such that

1. $P(E) \geq 0$ for every $E \in \Sigma$.
2. $P(\Omega) = 1$.
3. $P\Big(\bigcup_{\alpha \in C} A_\alpha\Big) = \sum_{\alpha \in C} P(A_\alpha)$ for every $C \subset \mathbb{Z}^+$ (i.e., for every countable indexing set C), provided all the $A_\alpha \in \Sigma$ and all the A_α are mutually disjoint.

That is, a probability measure assigns real numbers (probabilities) to events, such that the probabilities for nonoverlapping events simply add. These axioms—called the **Kolmogorov axioms**—also imply that probabilities are bounded (in fact, at most 1), and that $P(\emptyset) = 0$, or the probability of no event occurring is zero. The concept of the probability measure is obviously tied closely to the concepts of σ-algebra and probability space, and it is common to refer to these together as a **probability triple** (Ω, Σ, P).

1.2 Densities and Cumulative Probabilities

For a finite or countable probability space, it is convenient to label the outcomes by an integer index n, and then assign probabilities $P_n \geq 0$ to the outcomes (which as noted before are admissible as events), subject

to $\sum_n P_n = 1$. In the case where $\Omega = \mathbb{R}$, it is conventional to define a **probability density** $f(x)$, which is a function such that

$$f(x)\,dx := P\big[[x, x+dx)\big]. \tag{1.1}$$
(probability density)

This is a common definition in physics, which associates the density $f(x)$ with an arbitrarily narrow interval $[x, x+dx)$. A somewhat more careful definition could associate the probability measure

$$\int_a^b f(x)\,dx := P\big[(a,b)\big], \tag{1.2}$$
(probability density)

while requiring the axioms for probability measures to define the probabilities for other events (e.g., within the Borel measure). When referring to probabilities and probability densities, it is also common to refer to a **random variable** X which takes on an uncertain or random values. In this case we could write Eq. (1.2) as

$$\int_a^b f(x)\,dx := P\big[X \in (a,b)\big], \tag{1.3}$$
(probability density)

so that X is the random quantity, and the lower-case variable x refers to the outcomes of X, with (a,b) being the event (collection of outcomes) that is associated with a probability.

It is also common to use the **cumulative probability function** $F(x)$ of X, defined such that

$$F(x) := \int_{-\infty}^{x} f(x')\,dx' = P\big[X \in (-\infty, x)\big]. \tag{1.4}$$
(cumulative probability)

This function is well-defined in terms of the probability axioms. Because we can identify $dF(x) = f(x)\,dx$, the notation

$$\int_a^b dF(x) = P\big[X \in (a,b)\big] \tag{1.5}$$
(probability density)

is fairly popular in the mathemetical literature in place of the density integral (1.2).

1.2.1 Coordinate and Variable Transformations

Suppose that you want to make a coordinate transformation from old coordinate x to some new coordinate y. Then you can think of the new coordinate as a function of the old: $y = y(x)$. For this to be a "good" coordinate transformation, it should be smooth (continuous, and as many continuous derivatives as needed) and monotonic (i.e., dy/dx is continuous and everywhere nonvanishing). Then the coordinate transformation uniquely associates points in each coordinate system. That is, this function $y(x)$ is invertible, so you can also regard $x = x(y)$.

A simple function like the cumulative probability function $F(x)$ from Eq. (1.4) is something that associates a value to each coordinate x. Because the coordinate transformation $y(x)$ already uniquely associates x and y (by definition, they represent the same point in space, if you want to think of it that way), the same value applies to any coordinate system: $F(y) = F(x)$. A more formal way to think about this is that $F(y)$ is equivalent to $(F \circ y)(x) = F(y(x)) = F(x)$.

If $f(x)$ represents a probability density as in Eq. (1.1), then you have to be a little more careful, as $f(x)$ isn't meaningful on it's own. Rather, $f(x)\,dx$ is a meaningful quantity, as is the probability associated with the interval $[x, x+dx)$. To transform this, it has to transform to some equivalent probability

$$f_y(y)\,dy = f(y)\left|\frac{dx}{dy}\right|dy. \tag{1.6}$$

Here again $f(y)$ means $f(y(x))$, the pointwise association of the function. Thus, the "distribution in the y coordinate" is

$$f_y(y) = f(x)\left|\frac{dx}{dy}\right|. \tag{1.7}$$
(probability density under coordinate transformation.)

Only with this association will total probability be conserved under the coordinate transformation. Note the absolute value arises because both dy and dx are only sensibly positive in the expressions $f(x)\,dx$ and $f_y\,(y)$ (as they represent measures on the coordinate space). In this sense $|dy/dx|$ also represents the transformation of the integration measure for integrated probabilities of the form (1.2). It accounts for "squishing" of the coordinates in the transformation. In multiple dimensions, the $|dy/dx|$ generalizes to the Jacobian determinant, which describes the transformation of a volume element.

With probability densities it is also sensible to change variables where the change is noninvertible. For example, suppose that the probability density $f_x(x)$ is known for a random variable X. Then the question is, what is the density $f_y(y)$ for a transformed random variable $Y = Y(X)$? The idea is the same, except that it is important to sum over all the x's corresponding to a particular y:

$$f_y(y) = \sum_{x \in Y^{-1}(y)} f_x(x) \left| \frac{dx}{dy} \right|. \tag{1.8}$$
(noninvertible transformation)

Thie equation is also equivalent to the **Frobenius–Peron equation**

$$f_y(y) = \int dx'\, f_x(x')\, \delta[y - f(x')], \tag{1.9}$$
(Frobenius–Peron equation)

where the delta function contains the measure-transformation factor and also correctly associates the x's and y's.

1.3 Conditional Probability and Independence

Given two events $A, B \in \Sigma$, the **conditional probability** $P(A|B)$ ("the probability of A given that B occurred," whether the occurrence of B is by assumption or by observation) is defined in terms of the **joint probability** $P(A \cap B)$ such that

$$P(A \cap B) = P(A|B)\, P(B). \tag{1.10}$$
(conditional probability)

The two events A and B are said to be **independent** if

$$P(A \cap B) = P(A)\, P(B). \tag{1.11}$$
(independence of two events)

This condition is consistent with Eq. (1.10) provided

$$P(A|B) = P(A). \tag{1.12}$$

This makes intuitive sense: if the events are independent, then the observation that B occurred should not change anything about A. Also, putting $B = \Omega$ in Eq. (1.10),

$$P(A) = P(A \cap \Omega) = P(A|\Omega)\, P(\Omega) = P(A|\Omega), \tag{1.13}$$
(conditional probability)

demonstrates that conditioning on *any* outcome amounts to no conditioning at all.

Also the probability for a union of events is given by

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \tag{1.14}$$
(probability of union)

which follows from summing the probabilities of $A \cap B$ and the nonoverlapping parts of A and B. In the case of disjoint events, $A \cap B = \emptyset$, this becomes

$$P(A \cup B) = P(A) + P(B), \tag{1.15}$$

as consistent with the axioms for a probability density. In the case of independent events, we instead have

$$P(A \cup B) = P(A) + P(B) - P(A)\,P(B), \tag{1.16}$$

and the only way for this to reduce to the disjoint expression is for either $P(A) = 0$ or $P(B) = 0$.

In the analysis of stochastic processes, it is more useful to recast these ideas in terms of probability densities. In this case, for random variables X and Y, we can denote a joint probability density by $f(x, y)$. In the case that X and Y are independent random variables, the joint density factorizes as

$$f(x, y) = f(x)\,f(y). \tag{1.17}$$
(joint density for independent random variables)

We can also define a conditional density for X given an outcome $Y \in [y, y + dy)$ by $f(x|y)$, which must satisfy

$$f(x, y) = f(x|y)\,f(y). \tag{1.18}$$
(conditional-density relation)

In thinking of these quantities as representing infinitesimal probabilities, it is better to read this expression as $f(x, y)\,dx\,dy = [f(x|y)\,dx]\,f(y)\,dy$. Again, for independent random variables, $f(x|y) = f(x)$. The analogue here of conditioning on all possible events as in Eq. (1.13) is

$$\int_{-\infty}^{\infty} dy\, f(x, y) = \int_{-\infty}^{\infty} dy\, f(x|y)\,f(y) = f(x). \tag{1.19}$$
(integrating over one variable)

That is, we can ''integrate away'' one variable, or equivalently, by summing over all possible conditioning outcomes we end up with the unconditioned density.

1.4 Functionals of Probability Distributions

There are certain functions of probability distributions that are fundamental in probability theory. More specifically, these are **functionals**—functions that map functions to scalars. In particular, given a random variable X with probability density $f(x)$ **expectation value** of a function $g(X)$ is defined by

$$\langle g(X) \rangle := \int dx\, g(x)\, f(x). \tag{1.20}$$
(expectation value)

We will mostly deal with continuous random variables, but in the case of a discrete random variable N with probabilities P_n for the particular values n of N, the corresponding definition is

$$\langle g(X) \rangle := \sum_n g(n)\, P_n. \tag{1.21}$$
(expectation value)

One particularly important family of expectation values is the set of **moments**,[1] with the nth moment defined as the expectation value of X^n:

$$\langle X^n \rangle := \int dx\, x^n\, f(x). \tag{1.22}$$
(nth moment)

[1] The use of the word ''moment'' for these expectation values comes from analogy to classical mechanics, since an average of powers of the position $\mathbf{r}$ with respect to a mass distribution instead of a probability distribution yields total mass for the zeroth moment, the center of mass (times the total mass) for the first moment, and the moment of inertia for the second moment. In mechanics this term ''moment'' was in more common use in older textbooks; for example, see Sidney Wilcox McCuskey, *An Introduction to Advanced Dynamics* (Addison-Wesley, 1959), Section 1-4, p. 9 (archive: introductiontoad0000mccu), which introduces the ''moment of momentum'' (angular momentum) and ''moment of force'' (torque) in addition to moment of inertia. The origin of this terminology is an archaic use of the word ''moment'' to mean ''importance'' or ''consequence,'' in the case of mechanics referring to the importance of inertia, momentum, and force for the rotation of a rigid body; see Arthur Mason Worthington, *Dynamics of Rotation: An Elementary Introduction to Rigid Dynamics* (Longmans, Green, and Co., 1902), pp. 7-8 (archive: dynamicsrotatio00wortgoog). Incidentally, Newton used the word ''moment'' in a different archaic sense in the *Principia*, in the context of the ''moment of a genitum,'' which apparently refers to the derivative (moment) of a term (genitium) in a mathematical expression. See Isaac Newton, *Newton's Principia*, English translation by Andrew Motte (Daniel Adee, 1846) pp. 261-3 (archive: ost-physics-newtonspmathema00newtrich).

The first moment $\langle X \rangle$ is the **mean** of the random variable, and characterizes the ''middle'' value of the distribution. Higher moments characterize other aspects of the distribution. These are often given in terms of **centered moments**, which are expectation values of powers of $X - \langle X \rangle$:

$$\left\langle (X - \langle X \rangle)^n \right\rangle := \int dx\, (x - \langle X \rangle)^n\, f(x). \tag{1.23}$$
(nth centered moment)

The most famous example is the **variance**,

$$\mathrm{Var}[X] := \left\langle (X - \langle X \rangle)^2 \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2 \tag{1.24}$$
(variancedef)

(as a quick exercise, fill in the steps going between the second and third expressions here). The square root of the variance is the **standard deviation**

$$\sigma_x := \sqrt{\mathrm{Var}[X]} \tag{1.25}$$
(standard deviation)

is a common measure of the width of the probability distribution.

1.4.1 Characteristic Function

Given again a random variable X with probability density $f(x)$, another useful expectation value is the **characteristic function**, defined by

$$\tilde{f}(k) := \left\langle e^{ikX} \right\rangle = \int dx\, e^{ikx}\, f(x). \tag{1.26}$$
(characteristic function)

This, has the form of a Fourier transform of the density function $f(x)$, which is why we adopt the common notation $\tilde{f}(k)$ for the characteristic function. The characteristic function thus uniquely determines $f(x)$, because the Fourier transform uniquely identifies the original function; in particular, the inverse Fourier transform gives

$$f(x) = \frac{1}{2\pi} \int dk\, e^{-ikx}\, \tilde{f}(k).$$
(characteristic function determines the density function) (1.27)

Since the characteristic function (1.26) may be expanded in terms of moments,

$$\left\langle e^{ikX} \right\rangle = 1 + ik\langle X \rangle - \frac{k^2}{2} \langle X^2 \rangle + \cdots, \tag{1.28}$$

this implies that moments uniquely determine the probability density.

Unfortunately, it is not simple in practice to make use of the observation that the moments determine the distribution, with one notable exception. For a Gaussian distribution, as we will see in Section 2.1.2.1, the Fourier transform is also Gaussian; thus the characteristic function of a Gaussian random variable is Gaussian in k. In particular, then, all of the moments of a Gaussian can be characterized in terms of the mean and variance.

Another useful consequence of the definition of a characteristic function is that a shift in a probability density appears as a harmonic phase in the characteristic function. For example, suppose $\langle X \rangle = 0$ for some density function $f(x)$. This density can be shifted to have a nonzero mean $\langle X \rangle = \mu$ by letting $x \longrightarrow x - \mu$ in the integral in the definition (1.26):

$$\tilde{f}(k) = \int dx\, e^{ik(x-\mu)}\, f(x - \mu). \tag{1.29}$$

Then multiplying through by the new phase factor, we find

$$e^{ik\mu} \tilde{f}(k) = \int dx\, e^{ikx}\, f(x - \mu). \tag{1.30}$$

Thus, applying a factor $e^{ik\mu}$ to the characteristic function shifts the distribution mean by μ.

1.4.2 Moment-Generating Function

The **moment-generating function** is also an important characterization of the probability distribution, but it is essentially the same as the characteristic function (making the identification $z \longleftrightarrow ik$:

$$M(z) := \left\langle e^{tX} \right\rangle = \int dx\, e^{zx}\, f(x). \tag{1.31}$$
(moment-generating function)

The (very slight) advantage here is that $M(z)$ is real, but the more serious disadvantages are that the integral will have more problems with convergence than the characteristic function, and it is less obvious from this function that the moments determine the distribution. However, the function is named as it is because the expansion is slightly cleaner than the characteristic counterpart,

$$M(z) = 1 + z\langle X\rangle + \frac{z^2}{2}\left\langle X^2\right\rangle + \frac{z^3}{3}\left\langle X^3\right\rangle + \cdots, \tag{1.32}$$

and thus once we know $M(t)$, we can compute any moment via

$$\langle X^n\rangle = n!\, \left.\frac{\partial^n M}{\partial z^n}\right|_{z=0}. \tag{1.33}$$

Of course, we can have the analogous formula in terms of the characteristic function by letting $z \longrightarrow ik$ here. In either case, if the moments exist, we can consider the moment-generating function and the characteristic function to be well defined.

1.4.3 Cumulant-Generating Function

The **cumulant-generating function** is simply the logarithm of the moment-generating function:

$$K(z) := \log M(z) = \log\left\langle e^{zX}\right\rangle. \tag{1.34}$$
(cumulant-generating function)

If all the moments are well-defined, then we expect the cumulant-generating function to be well-defined and to have a series expansion

$$K(z) =: \kappa_1 z + \frac{1}{2}\kappa_2 z^2 + \frac{1}{3!}\kappa_3 z^3 + \cdots, \tag{1.35}$$
(cumulants)

where we used $K(0) = 0$ because $M(0) = 1$. The coefficients κ_n are the **cumulants** of X's probability distribution. The cumulants overall contain the same information as the moments (the nth cumulant is a function of the nth and lower moments), and they are sometimes preferred for characterizing probability distributions. Working out the first few terms of the series gives the first few cumulants:

$$\begin{aligned}
\kappa_1 &= \langle X\rangle \\
\kappa_2 &= \left\langle X^2\right\rangle - \langle X\rangle^2 \\
\kappa_3 &= \left\langle X^3\right\rangle - 3\langle X\rangle\left\langle X^2\right\rangle + 2\langle X\rangle^3 \\
\kappa_4 &= \left\langle X^4\right\rangle - 4\langle X\rangle\left\langle X^3\right\rangle - 3\left\langle X^2\right\rangle^2 + 12\langle X\rangle^2\left\langle X^2\right\rangle - 6\langle X\rangle^4.
\end{aligned} \tag{1.36}$$

Thus, the first cumulant is the mean (as is the first moment), and the second cumulant is the variance.

In the case of the Gaussian probability distribution, the characteristic and moment-generating functions are Gaussian. Thus, the cumulant-generating function is a quadratic function of z. That is, all cumulants beyond the second vanish for a Gaussian distribution; as we noted before, the mean and variance are sufficient to determine the distribution and thus the rest of the moments.

1.4.4 Probability-Generating Function

The concept of a generating function is quite general, and extends beyond moments and cumulants. Another important example is the **probabilty-generating function** for a discrete probability distribution P_n, which has the definition

$$P(z) := \sum_{n=0}^{\infty} P_n z^n. \tag{1.37}$$
(probability-generating function)

This is a compact way of representing the probability distribution, and of course individual probabilities can be recovered by appropriate differentiation of $P(z)$. However, this generating function is most useful in the case where $P(z)$ has an analytic form with a series expansion whose coefficients can be written down in closed form—in this case, the generating function is handy for analytic calculations. The proof of the Sparre Andersen theorem (Section 13.2.1)is a good example of the utility of the generating-function method.

The generating function has a few properties worth pointing out:

- $P(z = 1) = 1$, as a consequence of the normalization of the probability distribution. Although, of course, this notion of a generating functions generalizes to arbitrary sequences, in which case $P(1)$ is just the sum of the sequence.
- A linear combination of generating functions $aP(z)+bQ(z)$ corresponds to the same linear combinations of probabilities $aP_n + bQ_n$.
- For a product of generating functions $R(z) = P(z)Q(z)$, from matching powers of z it's easy to see that the coefficients satisfy
$$R_n = \sum_{j=0}^{n} P_k Q_{n-j}, \tag{1.38}$$
which has the form of a discrete convolution.
- Furthermore, if $P_M(z)$ and $Q_M(z)$ are the generating functions for (nonnegative) integer-valued random variables M and N, respectively, then the generating function for $M+N$ is the product $P_M(z)\,Q_M(z)$, as we can see from the previous convolution rule.

As a simple example of a generating function is for the probability distribution of an even coin toss to generate exactly n heads in a row before yields a tail on the next toss. In this case $P_n = (1/2)^{n+1}$ for $n \geq 0$, and so

$$P(z) = \sum_{n=0}^{\infty} \frac{z^n}{2^{n+1}} = \frac{1}{2-z}. \tag{1.39}$$

Another example is for the Catalan numbers C_n, which satisfy the recurrence

$$C_0 = 1, \qquad C_{n+1} = \sum_{j=0}^{n} C_j C_{n-j}. \tag{1.40}$$

Multiplying by z^{n+1},

$$C_{n+1} z^{n+1} = z \sum_{j=0}^{n} C_j z^j C_{n-j} z^{n-j}, \tag{1.41}$$

and then changing n to n' and summing from $n' = 0$ to n,

$$C(z) - C_0 = zC^2(z), \tag{1.42}$$

or

$$zC^2(z) - C(z) + 1 = 0. \tag{1.43}$$

Thus, the generating function must be of the form

$$C(z) = \frac{1 \pm \sqrt{1-4z}}{2z}, \tag{1.44}$$

but only the choice

$$C(z) = \frac{1 - \sqrt{1-4z}}{2z} \tag{1.45}$$

satisfies the correct boundary condition $C(0) = C_0 = 1$. The series expansion of the square root gives

$$C(z) = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} z^n, \tag{1.46}$$

so that we have

$$C_n = \frac{1}{n+1} \binom{2n}{n} \tag{1.47}$$

is an explicit expression for the nth Catalan number.

1.5 Likelihood and Statistical Inference

Let's return for a moment to the conditional probability density $f(x|y)$ defined in Eq. (1.18). This was the probability of a particular outcome x of a (continuous) random variable X *given* the occurrence of a particular outcome y of a random variable Y. With just a bit of reinterpretation, this conditional density acts as a powerful tool for inference of knowledge based on the outcome of a random event. (The same idea works for discrete probabilities, but we will stick to the notation for continuous random variables.)

Again, the notation $f(x|y)$ conveys the sense that the probability distribution for X is a function of the outcome of Y [which is assumed to be known, otherwise the expectation value of this function with respect to $f(y)$ gives the unconditioned density $f(x)$, as in Eq. (1.18)]. However, an alternate interpretation is that $f(x|y)$ says, given the particular outcome x of X, this gives information about the (uncertain) quantity Y. In this interpretation, $f(x|y)$ is said to be the **likelihood**

$$L(y|x) := f(x|y), \tag{1.48}$$
(likelihood)

which is regarded as a function of y, given a fixed x. Since $f(x|y)$ is a probability density for fixed y, $L(y|x)$ is not, in general, a probability density in y for fixed x (because it's not necessarily normalized), which is the reason for the different name. The likelihood is simply saying that since we know a particular outcome x occurred, this should give information about Y, because different values y are more or less consistent with the outcome x. Often Y is a parameter of a probability density that is unknown or partially known, in which case the outcome x gives information about the distribution, and thus about Y.

As a simple example, suppose we know that the probability density is a standard-normal distribution with some mean μ,

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}}\, e^{-(x-\mu)^2/2}, \tag{1.49}$$

but we don't know the value of μ. In fact, suppose that we know that either $\mu = 0$ or $\mu = 10$. If we have an outcome x of a random trial, and it turns out that $x = 0$ (this is our "datum"), then the likelihoods for the two cases are

$$L(\mu = 0|x) = f(x|\mu = 0) = \frac{1}{\sqrt{2\pi}}, \qquad L(\mu = 10|x) = f(x|\mu = 10) = \frac{1}{\sqrt{2\pi}}\, e^{-50}. \tag{1.50}$$

The likelihoods say that we should definitely favor $\mu = 0$ over $\mu = 10$, because the likelihood is a factor of 10^{50} in favor of $\mu = 0$—this value is *much* more consistent with the datum than is the large μ. The decision

to favor one value of the parameter μ here over the other is commonly discussed in terms of the **likelihood ratio**, here $L(\mu = 0|x)/L(\mu = 10|x)$. Exactly how large the likelihood ratio should be to favor one parameter value over another is subject to debate, but one suggestion[2] is that a likelihood ratio of at least 8 is ''fairly strong evidence'' for favoring one parameter value (''hypothesis'') over another, while a ratio of at least 32 constitutes ''strong evidence.''

1.5.1 Maximum Likelihood

The likelihood ratio gives a means for deciding between two alternatives based on data, but the likelihood function itself gives a means for deciding on the *best* parameter values, given a particular data set—that is, given a particular data set, the parameters that maximize the likelihood function are the most consistent with that data set. This principle of **maximum likelihood** is used to derive standard estimation statistics. For example, given the normal distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\mu)^2/2\sigma^2}, \tag{1.51}$$

and given a set of *independent* sample observations x_1, x_2, ... x_N, it turns out that the **sample mean**

$$\bar{x} := \frac{1}{N}\sum_{j=1}^{N} x_j \tag{1.52}$$

and the estimated standard deviation

$$\bar{\sigma} := \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(x_j - \bar{x}\right)^2} \tag{1.53}$$

are the maximum-likelihood values (''estimators'') for μ and σ, respectively. However, in the case of σ it turns out that $\bar{\sigma}$ is a *biased* estimator for σ, in the sense that for any finite N, $\langle\bar{\sigma}^2\rangle \neq \sigma^2$ (although this estimator is *asymptotically* correct, meaning that $\langle\bar{\sigma}^2\rangle \longrightarrow \sigma^2$ as $N \longrightarrow \infty$). Thus, the maximum-likelihood estimator $\bar{\sigma}$ is often replaced by the **sample standard deviation**

$$\bar{S} := \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(x_j - \bar{x}\right)^2}, \tag{1.54}$$

which has $\langle\bar{S}\rangle = \sigma$ for any $N \geq 2$.

1.5.2 Bayes' Theorem

The likelihood function shows up in another important interpretation of statistical inference. Recalling that the joint density $f(x, y)$ for X and Y to occur satisfies $f(x, y) = f(y, x)$, and that this may be written in terms of a conditional density as $f(x, y) = f(x|y)\, f(y)$, we can combine these last two equations to write

$$f(x|y)\, f(y) = f(y|x)\, f(x), \tag{1.55}$$

or

$$f(x|y) = \frac{f(y|x)\, f(x)}{f(y)} = \frac{L(x|y)\, f(x)}{f(y)}. \tag{1.56}$$

(Bayes' theorem)

[2]For a very readable introduction to inference via likelihood, see Greg Gandenberger, ''An Introduction to Likelihoodist, Bayesian, and Frequentist Methods,'' `http://gandenberger.org/2014/07/21/intro-to-statistical-methods` (2014). The likelihood ratio thresholds are discussed there and are from Richard Royall, ''On the Probability of Observing Misleading Statistical Evidence,'' *Journal of the American Statistical Association* **95**, 760 (2000) (doi: 10.2307/2669456).

This result, **Bayes' theorem**, is deceptively simple in its derivation, but is profound when interpreted correctly. Bayes' theorem gives the mathematical rule for incorporating statistical data into the probability distribution. More generally, it describes how new data influence our *state of knowledge*, in this case about the uncertain quantity X.[3] The original state of knowledge about X is given by the unconditioned density $f(x)$, which is called the **prior distribution**. The datum y is a particular outcome of another value Y that gives information about X; the new probability density $f(x|y)$ **conditioned** on the outcome y is given by multiplying the initial density by the likelihood function $L(x|y) = f(y|x)$. The resulting probability density is not in general normalized, but the normalization is restored simply by dividing by the unconditioned probability $f(y)$ for the datum y to occur. Often this normalization is established in terms of the likelihoods by

$$f(y) = \int dx\, f(y|x)\, f(x) = \int dx\, L(x|y)\, f(x), \qquad \begin{matrix}(1.57)\\ \text{(normalization factor)}\end{matrix}$$

since the likelihoods and unconditioned density must already be known for this Bayesian interpretation of statistical inference to make any sense.

In the case where the prior $f(x)$ is constant, perhaps representing a complete prior ignorance about X, the conditioned density $f(x|y)$ is proportional to the likelihood function $L(x|y)$, and in this case the Bayesian and likelihoodist approaches to statistical inferences should be equivalent. However, the Bayesian approach incorporates knowledge *prior* to the datum, which could reflect previous data or even subjective biases, via the prior distribution. Philosophically speaking, this could be good or bad. Scientists prefer to minimize subjective biases in analyzing experiments, which sounds like an argument against the Bayesian approach. However, such biases are arguably *always* present, so an argument in favor of the Bayesian approach is that it is best to be explicit about your assumptions.

[3] The definitive work on the Bayesian approach to inference is by E.T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge, 2003) (ISBN: 0521592712).

1.6 Exercises

Problem 1.1

Consider the Gaussian distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\mu)^2/2\sigma^2}, \tag{1.58}$$

given independent sample observations x_1, x_2, $\ldots$ x_N for X.

(a) Maximize the likelihood (actually, maximize the *logarithm* of the likelihood) with respect to μ and σ^2 to derive the maximum-likelihood estimators

$$\bar{x} := \frac{1}{N}\sum_{j=1}^{N} x_j \tag{1.59}$$

and

$$\bar{\sigma} := \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(x_j - \bar{x}\right)^2} \tag{1.60}$$

for μ and σ, respectively.

(b) Show that $\left\langle \bar{\sigma}^2 \right\rangle \neq \sigma^2$ for finite N, but that $\left\langle \bar{S} \right\rangle = \sigma$ for any $N \geq 2$, where

$$\bar{S} := \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(x_j - \bar{x}\right)^2}. \tag{1.61}$$

Chapter 2

Combinations of Random Variables and Random Walks

2.1 Finite Random Walks

The prototype problem for stochastic processes is the random walk. Suppose a random walker takes a random step of size X with probability density $f(x)$ between periodic intervals of duration Δt. Let's assume that all the steps are statistically independent, and the probability distribution is characterized by

$$\langle X \rangle = 0, \quad \mathrm{Var}[X] = \sigma^2. \tag{2.1}$$

After N steps (N large), where has the walker ended up? The **central limit theorem** says that the probability density of the accumulated displacement

$$S_N := \sum_{j=1}^{N} X_j \tag{2.2}$$

for N steps is Gaussian with zero mean and variance $N\sigma^2$. That is, the width (standard deviation, see Section 1.4) is $\sigma\sqrt{N}$. The probability distribution thus becomes asymptotically Gaussian with a time-dependent width of

$$\sigma(t) = \sigma\sqrt{\frac{t}{\Delta t}}. \tag{2.3}$$

This random-walk behavior is characteristic of a **diffusion process**, which is a transport process by which the distribution grows as $t^{1/2}$,

$$\Delta x \sim D\, t^{1/2}, \tag{2.4}$$

where for the random walker the **diffusion coefficient** is $D = \sigma/\sqrt{\Delta t}$. Note that within certain restrictions, the final distribution is asymptotically Gaussian, *independent of the one-step distribution.*

2.1.1 Two-Step Distribution

Before proving the full central limit theorem, we will examine the probability density after exactly two steps. The mathematical problem is as follows: let X_1 and X_2 be *independent* random variables with probability density functions $f_1(x)$ and $f_2(x)$, respectively. That is, the probability that $X_{1,2}$ is between x and $x + dx$ is $f_{1,2}(x)\, dx$. Then we can ask, what is the probability density of $X_1 + X_2$?

To answer this, we can note that $X_1 + X_2 = x$ for any pair of values of X_1 and X_2 that happen to add up to x. But then we must sum over all such pairs. The probability that both X_1 and X_2 will *both* have

particular probabilities is the product of the individual probabilities since the variables are independent. Thus, expressing what we said in equation form,

$$\text{Prob}(X_1 + X_2 \text{ between } x \text{ and } x + dx) = \sum_{x',x''} \text{Prob}(X_1 \text{ between } x' \text{ and } x' + dx') \times \text{Prob}(X_2 \text{ between } x'' \text{ and } x'' + dx'' \,|\, x = x' + x''). \tag{2.5}$$

We can translate this statement in terms of the probability densities and implement the constraint as a δ-function (with a factor of dx, so that the δ-function registers unity when the condition is met). Letting $f_+(x)$ denote the probability density of $X_1 + X_2$,

$$f_+(x)\,dx = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dx''\, f_1(x')\, f_2(x'')\, \delta(x' + x'' - x)\, dx. \tag{2.6}$$

Evaluating the x'' integral, we see that the probability density of the sum is the **convolution** of the individual densities,

$$f_+(x)\,dx = \int_{-\infty}^{\infty} dx'\, f_1(x')\, f_2(x - x')\, dx =: (f_1 * f_2)(x)\, dx, \tag{2.7}$$

where we use the $*$ symbol to denote convolution of two functions. Note that this result is general in that it doesn't assume any particular form for $f_1(x)$ or $f_2(x)$.

For the random walk, we assumed *identical*, indepedent steps, so that $f_1(x) = f_2(x) = f(x)$. Thus, the probability density for two steps is

$$f_{S_2}(x) = (f * f)(x), \tag{2.8}$$
(two-step probability density)

i.e., the convolution of the one-step distribution with itself. Recall that the convolution ''smears'' one function with another, and so as the effect of the second step is to smooth the one-step distribution. The idea behind the central limit theorem is that this smoothing continues until the distribution is Gaussian after many steps.

2.1.1.1 Example 1: Convolution with a Delta Function

As an example of the general idea of the convolution of two functions f and g,

$$(f * g)(x) = \int_{-\infty}^{\infty} dx'\, f(x')\, g(x - x'), \tag{2.9}$$

consider the convolution of $f(x)$ with the perfectly localized delta function $g(x) = \delta(x)$. The convolution is then

$$(f * \delta)(x) = \int_{-\infty}^{\infty} dx'\, f(x')\, \delta(x - x') = f(x). \tag{2.10}$$

The effect of convolution with a delta function is thus simply to do nothing: convolution with a delta function is just the identity operation.

In terms of the random walk, $\delta(x)$ as a one-step probability function simply corresponds to a step of zero length, or just taking no step at all. Thus, it makes intuitive sense that the distribution isn't changed by convolution with $\delta(x)$. In general, when $g(x)$ is some other function, the convolution ''smears'' $f(x)$ with the **convolution kernel** $g(x)$. Typically, we will use centered kernels; the effect of a displaced kernel is simply to displace the convolution by the same amount. For example, if

$$g(x) = \delta(x - x_0), \tag{2.11}$$

then

$$(f * g)(x) = \int_{-\infty}^{\infty} dx'\, f(x')\, \delta(x - x_0 - x') = f(x - x_0), \tag{2.12}$$

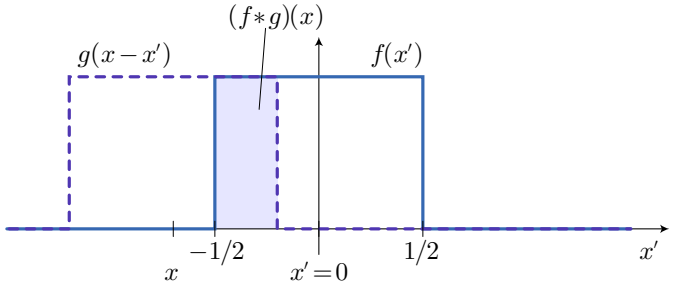
which is just the displaced version of the original.

2.1.1.2 Example 2: Convolution of Box Functions

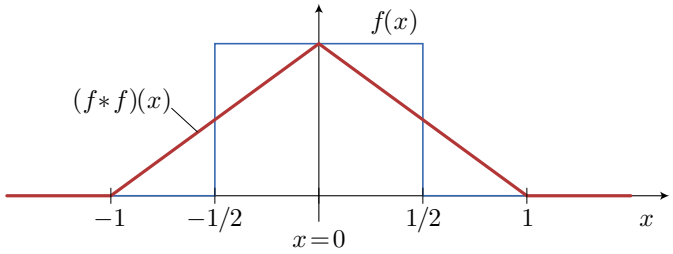
As a slightly more complicated example, consider the convolution of box functions, both given by

$$f(x) = g(x) = \begin{cases} 1, & |x| \le 1/2 \\ 0 & \text{elsewhere,} \end{cases} \tag{2.13}$$

which here are properly normalized to correspond to probability distributions. The convolution consists of displacing $g(x')$ by x, multiplying the functions together, and integrating. For this simple case (box functions of unit height), the convolution (product) just turns out to be the area where the two functions overlap.



When the displacement is large, $|x| > 1$, the boxes don't overlap at all, so the convolution is zero. Otherwise, the overlap area varies linearly with the displacement, so the convolution is a triangle function.



(Note that $f * f$ in the figure is the same as $f * g$ for this special case of $f = g$.) We see now the "smoothing" or "blurring" effect of the convolution. The original functions were discontinuous, but the convolution is continuous. The convolution is also wider than the original functions. As we will see, continued, successive convolutions will make the distribution look Gaussian.

2.1.2 Convolution Theorem

Now that we brought up the convolution, we may as well discuss how to compute it. The **convolution theorem** gives an easy way to evaluate the convolution integral in Eq. (2.7), both in an intuitive and a computational sense. The convolution theorem states that the Fourier transform of the convolution is the product of the Fourier transforms of the individual functions:

$$\mathscr{F}[f * g] = \mathscr{F}[f]\mathscr{F}[g]. \tag{2.14}$$
(convolution theorem)

To prove this, we'll just compute the explicit form of $\mathscr{F}[f * g]$. This will be very much a physicist's proof, not a mathematician's proof, in that we'll just assume the functions are nice enough that all the integrals simply exist.

First of all, in our notation here, the Fourier and inverse transforms have the form

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, \tilde{f}(k)\, e^{ikx}, \quad \tilde{f}(k) = \int_{-\infty}^{\infty} dx\, f(x)\, e^{-ikx}, \tag{2.15}$$

where $\tilde{f}(k) \equiv \mathscr{F}[f(x)]$. It's important to make this explicit, since the result depends on the normalization

convention we choose for the Fourier transform. Then computing the Fourier transform of $f * g$,

$$\begin{aligned}\mathscr{F}[f*g] &= \mathscr{F}\left[\int_{-\infty}^{\infty} dx'\, f(x')\, g(x-x')\right] \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, f(x')\, g(x-x')\, e^{-ikx} \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, f(x')\, e^{-ikx'} g(x-x')\, e^{-ik(x-x')}.\end{aligned} \tag{2.16}$$

Letting $x \longrightarrow x + x'$,

$$\begin{aligned}\mathscr{F}[f*g] &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx' f(x')\, e^{-ikx'} g(x)\, e^{-ikx} \\ &= \int_{-\infty}^{\infty} dx'\, f(x')\, e^{-ikx'} \int_{-\infty}^{\infty} dx\, g(x)\, e^{-ikx} \\ &= \mathscr{F}[f]\mathscr{F}[g].\end{aligned} \tag{2.17}$$

Thus, to convolve two functions, just follow this recipe: Fourier transform both functions, multiply them together, then compute the inverse Fourier transform. Mathematically, we can write

$$f * g = \mathscr{F}^{-1}\left\{\mathscr{F}[f]\mathscr{F}[g]\right\}. \tag{2.18}$$

Since Fourier transforms of common function are usually already known, the convolution theorem provides a shortcut for evaluating the full convolution integral.

2.1.2.1 Example: Convolution of Two Gaussians

Since it's easy to compute the Fourier transform of Gaussian distributions, let's use the convolution theorem to convolve two Gaussians. Let's write the two functions as

$$f(x) = Ae^{-x^2/\alpha^2}, \quad g(x) = A'e^{-x^2/\beta^2}. \tag{2.19}$$

The Fourier transform of a Gaussian is also a Gaussian, and in particular

$$\mathscr{F}[f](k) = \tilde{f}(k) = A\alpha\sqrt{\pi}e^{-\alpha^2k^2/4}, \quad \mathscr{F}[g](k) = \tilde{g}(k) = A'\beta\sqrt{\pi}e^{-\beta^2k^2/4}. \tag{2.20}$$

Then the product of the Fourier transforms is

$$\left(\mathscr{F}[f]\mathscr{F}[g]\right)(k) = AA'\alpha\beta\pi e^{-(\alpha^2+\beta^2)k^2/4}. \tag{2.21}$$

Finally, we invert the Fourier transform to obtain the convolution:

$$(f*g)(x) = \mathscr{F}^{-1}\left[AA'\alpha\beta\pi e^{-(\alpha^2+\beta^2)k^2/4}\right] = \frac{AA'\alpha\beta\sqrt{\pi}}{\sqrt{\alpha^2+\beta^2}} \exp\left(-\frac{x^2}{\alpha^2+\beta^2}\right). \tag{2.22}$$

Recall that the standard (normalized) form of the Gaussian is

$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{2.23}$$

where the μ is the mean and σ is the standard deviation (σ^2 is the variance). The standard deviation is a common measure of the width of a Gaussian function. Note that $f(x)$ has standard deviation $\alpha/\sqrt{2}$, $g(x)$ has standard deviation $\beta/\sqrt{2}$, and $(f*g)(x)$ has standard deviation $\sqrt{(\alpha^2+\beta^2)/2}$, so that the standard deviations add in quadrature as a result of the convolution. Thus, the convolution of Gaussians is still Gaussian, but the blurring effect of the convolution makes the convolved Gaussian wider than the original functions.

2.2 Central Limit Theorem

Now we extend the two-step analysis above analysis to N steps. Let $X_1, \ldots, X_N$ be independent, identically distributed random variables. Let $f(x)$ be the probability density function of each of the X_j. Defining the sum by

$$S_N := \sum_{j=1}^{N} X_j, \tag{2.24}$$

we will now ask, what is the probability density $f_{S_N}(x)$ of S_N? Evidently, we can iterate Eq. (2.8) to obtain

$$f_{S_N}(x) = (f * f * \cdots * f)(x), \tag{2.25}$$

where the result is the successive convolution of N copies of f (for $N-1$ total convolution operations). However, it turns out that this distribution becomes simple for large enough N.

The **central limit theorem** states that, provided that the mean and variance of the X_j exist, with the mean $\mu = \langle X_j \rangle$ and variance $\sigma^2 = \mathrm{Var}[X_j]$, the distribution $f_{S_N}(x)$ becomes asymptotically Gaussian for large N with

$$\langle S_N \rangle = N\mu, \quad \mathrm{Var}[S_N] = N\sigma^2. \tag{2.26}$$

(central limit theorem)

(The mean and variance are in fact exact, whereas the form of the distribution is valid for large N.) This is a rough statement, since ''becomes asymptotically Gaussian'' is an imprecise statement. So let's clean this up a bit.

The central limit theorem states that the probability density function $f_{Z_N}(x)$ of the centered, scaled statistic

$$Z_N := \frac{S_N - N\mu}{\sigma\sqrt{N}} \tag{2.27}$$

converges to the ''standard normal'' (Gaussian) distribution

$$f_{Z_N}(x) \longrightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{2.28}$$

which is the special Gaussian with mean 0 and unit variance.

Let's prove this now[1]. To evaluate the convolutions in Eq. (2.25), we need to employ the convolution theorem. Computing the characteristic function (1.26) of $f(x)$,

$$\begin{aligned} \tilde{f}(k) &= \int_{-\infty}^{\infty} dx\, f(x)\, e^{ikx} \\ &= \sum_{j=0}^{\infty} \int_{-\infty}^{\infty} dx\, f(x) \frac{(-ikx)^j}{j!} \\ &= 1 + ik\mu - \frac{k^2(\sigma^2 + \mu^2)}{2} + O(k^3). \end{aligned} \tag{2.29}$$

Here, we Taylor-expanded e^{-ikx} and then used the fact that the terms of the expansion were proportional to expectation values $\langle X^j \rangle$.

This is more cumbersome than necessary, so let's recompute the expansion in Eq. (2.29) for the centered, scaled variable

$$Z_j = \frac{X_j - \mu}{\sigma\sqrt{N}}, \tag{2.30}$$

with corresponding probability density $f_Z(x)$. The centering effectively zeroes the mean, and the rescaling changes the factor in front of the variance, with the result

$$\tilde{f}_Z(k) = 1 - \frac{k^2}{2N} + O\left[\left(\frac{k}{\sqrt{N}}\right)^3\right]. \tag{2.31}$$

[1] For a more rigorous version, see T. W. Körner, *Fourier Analysis* (Cambridge, 1988), starting on p. 349.

The convolution theorem says that to calculate the transform of the N-fold convolution, we just compute $\tilde{f}_Z(k)$ to the Nth power:

$$\tilde{f}_{Z_N}(k) = \left[\tilde{f}_Z(k)\right]^N = \left(1 - \frac{k^2}{2N} + O\left[\left(\frac{k}{\sqrt{N}}\right)^3\right]\right)^N. \tag{2.32}$$

As N becomes large, we can neglect the higher order terms beyond the first, and then use the formula

$$\lim_{n\longrightarrow\infty}\left(1+\frac{x}{n}\right)^n = e^x \tag{2.33}$$

to see that for large N, the transform becomes

$$\tilde{f}_{Z_N}(k) = \exp\left(-\frac{k^2}{2}\right). \tag{2.34}$$

But now the inverse Fourier transform of $\exp(-k^2/2)$ is $\exp(-x^2/2)/\sqrt{2\pi}$, so f_{Z_N} converges to a standard normal distribution as $N \longrightarrow \infty$.

2.2.1 Example: Square Distribution

As a simple example of the central limit theorem, let's try out the unit box function as the one-step distribution, as we tried out in Section 2.1.1.2:

$$f(x) = \begin{cases} 1, & |x| \le 1/2 \\ 0 & \text{elsewhere.} \end{cases} \tag{2.35}$$

First note that the this function is normalized, so it represents a proper probability distribution. Thus, so do all of its self-convolutions. Let $f^{*N}(x)$ denote the convolution of $f(x)$ with itself $N-1$ times. This is the same as $f_{S_N}(x)$ for the random-walk interpretation of this distribution. The central limit theorem says that asymptotically, the self-convolution becomes Gaussian,

$$f^{*N}(x) = \frac{1}{\sqrt{2\pi}\sigma_N} e^{-x^2/2\sigma_N^2}, \tag{2.36}$$

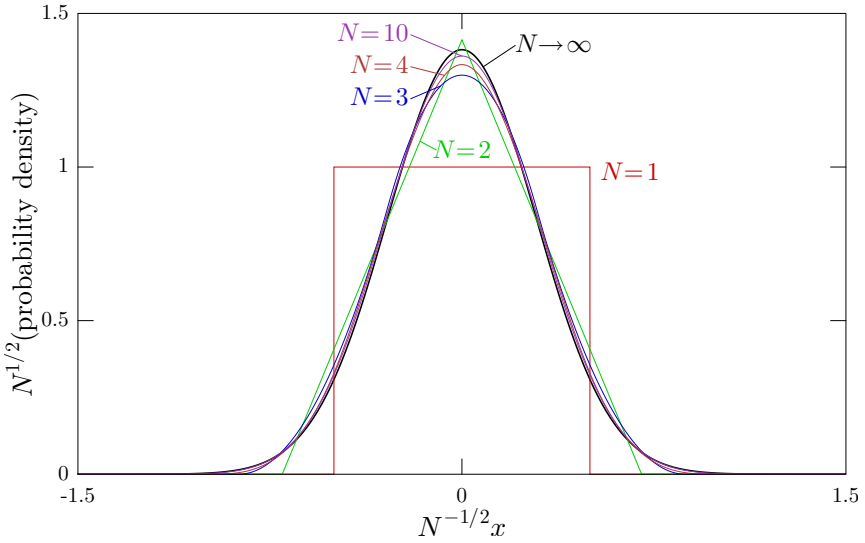
with zero mean, since $f(x)$ is centered. The variance of $f(x)$ is

$$\int_{-\infty}^{\infty} dx\, x^2\, f(x) = \int_{-1/2}^{1/2} x^2\, dx = \frac{1}{12}, \tag{2.37}$$

so that the width of the asymptotic Gaussian is

$$\sigma_N = \sqrt{\frac{N}{12}}. \tag{2.38}$$

Here, $f(x)$ is plotted with several self-convolutions $f^{*N}(x)$, along with the asymptotic form, the Gaussian of width $\sigma_N = \sqrt{N/12}$.



As N increases, the widths of the distributions increase and the peak values decrease, but we have rescaled the axes by appropriate factors of $\sqrt{N}$ to keep the distributions comparable at each step. The box function is very different from the asymptotic Gaussian. However, even the first self-convolution (a triangle function) is already pretty close to the Gaussian, and the successive self-convolutions converge fairly rapidly to the asymptotic form.

2.2.2 Standard Deviation of the Mean

Returning again to error analysis, suppose we make independent measurements $X_1, \ldots, X_N$ of some quantity in the laboratory. The **sample mean** is

$$\mu_N := \frac{1}{N}\sum_{j=1}^{N} X_j. \tag{2.39}$$

We can rewrite this as

$$\mu_N = \frac{S_N}{N} = \mu + \frac{\sigma Z_N}{\sqrt{N}}, \tag{2.40}$$

where the first term represents the *true mean*, and the second is the experimental error (statistical fluctuation in the sample mean). Applying the central limit theorem, Z_N is approximately standard normal for large N, so μ_N is Gaussian with mean μ and standard deviation $\sigma/\sqrt{N}$, where σ is the standard deviation of a single measurement. Thus, the **standard deviation of the mean** (also called the **standard error**) is $\sigma/\sqrt{N}$. This is why, by making many measurements, it is possible to increase the accuracy of a measured quantity.

2.3 Variances Add in Quadrature

In discussing random walks so far, we have been discussing the asymptotic, N-step probability distribution, and its scaling with time. However, we can make a simpler statement that does not explicitly refer to the distribution. Let $X_1, \ldots, X_N$ be independent random variables, but now we won't even require them to be identically distributed. For the moment, let's also assume $\langle X_n \rangle = 0$. Now consider the sum $X_1 + X_2$. Clearly

the mean vanishes, and thus the variance becomes

$$\begin{aligned}\mathrm{Var}[X_1 + X_2] &= \left\langle (X_1 + X_2)^2 \right\rangle \\ &= \left\langle X_1^2 \right\rangle + \left\langle X_2^2 \right\rangle + 2\langle X_1 X_2 \rangle \\ &= \left\langle X_1^2 \right\rangle + \left\langle X_2^2 \right\rangle + 2\langle X_1 \rangle \langle X_2 \rangle \\ &= \mathrm{Var}[X_1] + \mathrm{Var}[X_2],\end{aligned} \tag{2.41}$$

where we used the fact that X_1 and X_2 are independent, and thus their correlation function $\langle X_1 X_2 \rangle$ factorizes into $\langle X_1 \rangle \langle X_2 \rangle$ [the joint probability density $f(x_1, x_2)$ for independent processes must have the factored form $f_1(x_1) f_2(x_2)$]. Thus, the *variances* of independent random variables, and regarding the variance as the square of the ''width'' of the corresponding probability distributions, we see that the widths add *in quadrature* when we add together the random variables. By subtracting $\langle X_1 + X_2 \rangle^2 = \langle X_1 \rangle^2 + \langle X_2 \rangle^2 + 2\langle X_1 \rangle \langle X_2 \rangle$ from each intermediate expression, it isn't hard to see that the same result holds when $\langle X_n \rangle \neq 0$.

Iterating this process, we see that the variance of the sum defined as before,

$$S_N := \sum_{j=1}^{N} X_j, \tag{2.42}$$

is simply

$$\mathrm{Var}[S_N] = \sum_{j=1}^{N} \mathrm{Var}[X_j]. \tag{2.43}$$

Again, if we take each X_n to be identical as for the random walk, and we take the variance as the square of the width σ (i.e., $\mathrm{Var}[X_n] = \sigma^2$), then

$$\mathrm{Var}[S_N] = N\sigma^2, \tag{2.44}$$
(variance of the sum)

or

$$\sigma_N := \sqrt{\mathrm{Var}[S_N]} = \sqrt{N}\,\sigma. \tag{2.45}$$
(standard deviation of the sum)

This is the same as one of the results of the central limit theorem (2.26), but this is not an asymptotic statement, it is *exact*. Thus, we expect the width of the sum to be precisely $\sqrt{N}\,\sigma$. Nevertheless, we often expect this scaling of the distribution width to hold only asymptotically, since in general the ensemble of walkers will have an initial distribution that does not match the one-step distribution (or any N-step distribution), and thus we also need to include the convolution with this initial state.

2.3.1 Cauchy Walk

Consider independent, identically distributed random variables $X_1, \ldots, X_N$ with **Cauchy** (Lorentzian) probability density functions

$$f(x) = \frac{1}{\pi(1 + x^2)}. \tag{2.46}$$
(Cauchy distribution)

The Fourier transform is given by

$$\tilde{f}(k) = e^{-|k|}, \tag{2.47}$$

as we can see by computing the *inverse* Fourier transform of $\tilde{f}(k)$:

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-|k|} e^{ikx} \\ &= \frac{1}{2\pi}\left[\int_0^{\infty} dk\, e^{-k(1-ix)} + \int_0^{\infty} dk\, e^{-k(1+ix)}\right] \\ &= \frac{1}{2\pi}\int_0^{\infty} dk\, e^{-k(1-ix)} + \text{c.c.} \\ &= \frac{1}{2\pi(1-ix)} + \text{c.c.} \\ &= \frac{1+ix}{2\pi(1+x^2)} + \text{c.c.} \\ &= \frac{1}{\pi(1+x^2)}. \end{aligned} \tag{2.48}$$

We have now shown that $\mathscr{F}^{-1}[\tilde{f}(k)] = f(x)$. Both $f(x)$ and $\tilde{f}(k)$ are continuous and bounded, so the Fourier transform is invertible; thus, $\mathscr{F}[f(x)] = \tilde{f}(k)$.

Now let's compute the probability density of the mean

$$\mu_N := \frac{1}{N}\sum_{j=1}^{N} X_j. \tag{2.49}$$

The probability density function of the sum

$$S_N := \sum_{j=1}^{N} X_j \tag{2.50}$$

is $(f * f * \ldots * f)(x)$ (N copies of f or $N-1$ convolutions), so using the convolution theorem,

$$\tilde{f}_{S_N}(k) = \left[\tilde{f}(k)\right]^N = \left[e^{-|k|}\right]^N = e^{-|Nk|} = \tilde{f}(Nk). \tag{2.51}$$

In general, if $f(x)$ and $\tilde{f}(k)$ are a Fourier transform pair, then so are $\alpha f(\alpha x)$ and $\tilde{f}(k/\alpha)$. Thus, the inverse transform of $\tilde{f}(Nk)$ is $f(x/N)/N$. The variable μ_N is the same as S_N except for a scaling factor of $1/N$, so the probability density must be the same, but N times wider. So to get the probability density of μ_N, we make the replacement $x \longrightarrow Nx$ in the expression $f_{S_N}(x)\,dx = f(x/N)\,dx/N$ for the probability density of S_N, which gives $f(x)\,dx$. Thus, $f(x)$ is also the probability density of μ_N.

This is different than what we expect from the central limit theorem: there, we expect the mean to have a width that is smaller than that of the one-step distribution by a factor of $\sqrt{N}$. Stated otherwise, what we have shown is that the width of the sum distribution $f_{S_N}(x) = f(x/N)$ is N times larger than that of the one-step distribution, which says that the *widths add* for the Cauchy random walk. But the central limit theorem said that *variances add*, or the widths should add *in quadrature*. Is there a contradiction here?

Obviously there should be some simple resolution. The problem is that variance of X_j does not exist for a Cauchy distribution. This is because the Cauchy distribution only falls off as $1/x^2$ for large $|x|$, and so the variance integral

$$\int_{-\infty}^{\infty} dx\, f(x)\, x^2 \tag{2.52}$$

diverges. The central limit theorem implicitly assumes that the variance exists; thus, the central limit theorem does not apply to this case. This is one case of **anomalous diffusion**, where the diffusion coefficient diverges, because the width of the N-step distribution does not scale diffusively (i.e, it scales as t rather than $\sqrt{t}$). We will return to such **heavy-tailed distributions** and discuss them at length starting in Chapter 12.

2.4 Arbitrary Combinations of Random Variables

In considering random walks, we have been considering the probability distribution corresponding to the sum (2.2) of independent random variables, in terms of the probability distributions of the separate variables—it turned out to be just the convolution of the individual distributions. But how far can we push this? Here we will develop some concepts seemingly unrelated to probability, and use them to deduce the probability density for an arbitrary (scalar) function of a set of independent random variables.

2.4.1 Divergence Theorem

The divergence theorem is fundamental in the study of electrostatics, and the standard form states that for a vector field $\mathbf{A}$,

$$\int_V (\nabla \cdot \mathbf{A})\, d^3r = \oint_S \mathbf{A} \cdot \hat{n}\, dS, \tag{2.53}$$

where V is the volume of integration, S is the surface of the volume, $\hat{n}$ is the (outward-pointing) normal vector to the surface, and dS is the surface-area element for integration over the surface of S.

Let's briefly derive this in d dimensions. Consider a box of infinitesimal volume, given by

$$dV = dx_1 \cdots dx_d. \tag{2.54}$$

The flux of $\mathbf{A}(\mathbf{x})$ through the surface of this volume is given by summing over the fluxes of the two sides bounding each dimension; the "area" of the jth side is $dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_d = dV/dx_j$, and thus we have a total flux

$$\begin{aligned} \mathbf{A} \cdot \hat{n}\, dS &= \sum_j \left[\mathbf{A}(\mathbf{x} + dx_j \hat{x}_j) \cdot \hat{x}_j \frac{dV}{dx_j} - \mathbf{A}(\mathbf{x}) \cdot \hat{x}_j \frac{dV}{dx_j} \right] \\ &= \sum_j \frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_j} \cdot \hat{x}_j\, dV \\ &= (\nabla \cdot \mathbf{A})\, dV, \end{aligned} \tag{2.55}$$

where the divergence is interpreted in d dimensions. Now we integrate over the volume, adding up the volume and surface contributions due to all the infinitesimal elements in the integration volume. Whenever two elements contact, their fluxes on their common surface cancel, so the only contribution in the surface integral is due to the flux on the outer surface of the integration volume, and thus we have

$$\int_V d^d x\, (\nabla \cdot \mathbf{A}) = \oint_S \mathbf{A} \cdot \hat{n}\, dS \tag{2.56}$$

(divergence theorem, n dimensions)

as the generalized divergence theorem, converting a volume integral in n dimensions to an integral over the bounding hypersurface, a manifold in $n-1$ dimensions.

2.4.2 Transformation of Surface Delta Functions

This divergence theorem is useful in establishing a chain-rule formula for a delta function. Recall that given a function $f(x)$, the delta function obeys the chain rule

$$\delta[f(x-a)] = \sum_{x \in f^{-1}(a)} \frac{\delta(x-a)}{|f'(a)|}, \tag{2.57}$$

which follows from requiring that the delta function be normalized under integration over either $y = f(x)$ or x. For a coordinate change between coordinates $\mathbf{x}$ and $\mathbf{y}$ (both in d dimensions), this formula generalizes to

$$\delta^d(\mathbf{y}) = \frac{\delta^d(\mathbf{x})}{|\partial \mathbf{y}/\partial \mathbf{x}|}, \tag{2.58}$$

so that the scaling factor is now the Jacobian determinant of the coordinate transformation. But what if we have a *scalar* function $y = h(\mathbf{x})$ of a d-dimensional vector $\mathbf{x}$? The appropriate generalization of the determinant turns out to be the Euclidean norm of the vector of partial derivatives of h, as we will now show.

We will start by considering arbitrary scalar functions $h(\mathbf{x})$ and $\mathbf{A}(\mathbf{x})$, respectively, on $\mathbb{R}^d$. Now consider the step function $\Theta(h)$, which defines a volume (or possibly multiple, unconnected volumes). Then consider the integral

$$\int d^d x \, \nabla \cdot \{\Theta[h(\mathbf{x})]\mathbf{A}(\mathbf{x})\} = 0, \tag{2.59}$$

which follows from changing to a surface integral via the divergence theorem, and assuming that either the surface at infinity is outside the volume defined by h (any physically relevant volume should be finite), or that $\mathbf{A}$ vanishes when necessary at infinity. Then using

$$\nabla \cdot [\Theta(h)\mathbf{A}] = \mathbf{A} \cdot \nabla\Theta(h) + \Theta(h)\nabla \cdot \mathbf{A}, \tag{2.60}$$

we have

$$\begin{aligned} \int d^d x \, \Theta(h)\nabla \cdot \mathbf{A} &= -\int d^d x \, \mathbf{A} \cdot \nabla\Theta(h) \\ &= -\int d^d x \, \delta(h) \, \mathbf{A} \cdot \nabla h. \end{aligned} \tag{2.61}$$

Now put

$$f(\mathbf{x}) = \mathbf{A} \cdot \nabla h, \tag{2.62}$$

and consider

$$\begin{aligned} \int d^d x \, \delta[h(\mathbf{x})] \, f(\mathbf{x}) &= -\int d^d x \, \Theta[h(\mathbf{x})] \, \nabla \cdot \mathbf{A}(\mathbf{x}) \\ &= -\int_V d^d x \, (\nabla \cdot \mathbf{A}) \\ &= -\oint_S \mathbf{A} \cdot \hat{n} \, dS \\ &= \oint_{h^{-1}(0)} \mathbf{A} \cdot \frac{\nabla h}{|\nabla h|} \, dS \\ &= \oint_{h^{-1}(0)} \frac{f(\mathbf{x})}{|\nabla h|} \, dS, \end{aligned} \tag{2.63}$$

where we have identified V as the volume defined by $\Theta(h)$; S as the surface, defined by the locus of points where h vanishes; and the normal vector $\hat{n} = -\nabla h/|\nabla h|$, since the gradient is normal to the surface, but points towards the *interior* of the volume (where h is increasing away from the surface). To summarize, we have identified

$$\int d^d x \, \delta[h(\mathbf{x})] \, f(\mathbf{x}) = \oint_{h^{-1}(0)} \frac{f(\mathbf{x})}{|\nabla h|} \, dS, \tag{2.64}$$

(δ-function constraint in integration)

as the effect of a δ-function constraint in a volume integral, where $|\nabla h|$ is the Euclidean norm of the gradient vector:

$$|\nabla h| = \sqrt{\sum_j (\partial_{x_j} h)^2}. \tag{2.65}$$

Again, the constraint here changes the volume integral to an integral over the bounding hypersurface. In this derivation, we have assumed that any function $f(\mathbf{x})$ can be represented as $\mathbf{A} \cdot \nabla h$, but since $\mathbf{A}$ is arbitrary, this is always possible provided $\nabla h \neq 0$ anywhere (and we are free to assume this while still representing any surface manifold we like).

2.4.3 Direct Derivation

A more direct derivation of Eq. (2.64) proceeds as follows.[2] Recall that $h(\mathbf{x}) = 0$ defines the surface on which the delta function "fires," and we are considering vector coordinates $\mathbf{x} = (x_1, \ldots, x_d)$. Consider a neighborhood of a point $\mathbf{x}_0$ on the surface, in which one of the partial derivatives is nonvanishing, say $\partial h/\partial x_1 \neq 0$. Then we can consider the action of the δ function along this coordinate, as

$$\int dx_1\, \delta[h(\mathbf{x})]\, f(\mathbf{x}) = \int dx_1\, \frac{\delta(x_1 - x_{01})\, f(\mathbf{x})}{\left|\dfrac{\partial h}{\partial x_1}\right|_{\mathbf{x}=\mathbf{x}_0}} = \frac{f(x_{01}, x_2, \ldots, x_d)}{\left|\dfrac{\partial h}{\partial x_1}\right|_{\mathbf{x}=\mathbf{x}_0}}, \tag{2.66}$$

where $x_{10} = x_{10}(x_2, \ldots, x_d)$ is the x_1-component of $\mathbf{x}_0$, and we have used the ordinary δ-function chain rule (2.57). Then using

$$|\nabla h| = \sqrt{\sum_{j=1}^{d} \left(\frac{\partial h}{\partial x_j}\right)^2} = \left|\frac{\partial h}{\partial x_1}\right| \sqrt{1 + \sum_{j=2}^{d} \left(\frac{\partial x_1}{\partial x_j}\right)^2}, \tag{2.67}$$

and then identifying

$$dS = \sqrt{1 + \sum_{j=2}^{d} \left(\frac{\partial x_1}{\partial x_j}\right)^2}\, dx_2 \cdots dx_d = \sqrt{\sum_{j=1}^{d} \left(\frac{\partial x_1}{\partial x_j}\right)^2}\, dx_2 \cdots dx_d \tag{2.68}$$

as the local surface element (this bears more explanation), and finally integrating over the remaining coordinates [which now locally parameterize the surface $h(\mathbf{x}) = 0$], Eq. (2.64) then results.

Now to return to the business of identifying the surface element (2.68). We can define the surface element in terms of the volume element as

$$dV = dS\, d\ell, \tag{2.69}$$

where $d\ell$ is the line element normal to the surface. Then

$$dS = \frac{dV}{d\ell} = \frac{dx_1 \cdots dx_d}{d\ell}. \tag{2.70}$$

Now suppose that we express $d\ell$ in terms of x_1:

$$d\ell = (\hat{n} \cdot \hat{x}_1)\, dx_1, \tag{2.71}$$

where $\hat{n}$ is normal to the surface, so that $d\ell$ represents only the component of dx_1 normal to the surface. Then dividing the volume by $d\ell$, we only remove the dimension normal to the surface. Using $\hat{n} = \nabla h/|\nabla h|$, we have

$$d\ell = \frac{1}{|\nabla h|}\frac{\partial h}{\partial x_1}\, dx_1 = \frac{dx_1}{\sqrt{1 + \sum_{j=2}^{d} \left(\dfrac{\partial x_1}{\partial x_j}\right)^2}}, \tag{2.72}$$

after using Eq. (2.67), and then putting this expression into Eq. (2.70), we obtain the surface element (2.68).

2.4.4 Chain Rule for Coordinate Transformations

Using Eq. (2.64), we can also derive a chain rule for integrals of the form of the left-hand side, when we compare equivalent constraints specified by different functions. First, by rescaling the function $f(\mathbf{x})$, we can write

$$\int d^d x\, \delta[h(\mathbf{x})]\, |\nabla h|\, f(\mathbf{x}) = \oint_{h^{-1}(0)} f(\mathbf{x})\, dS(\mathbf{x}). \tag{2.73}$$

[2] This derivation and the connection to the divergence theorem in the previous section are adapted from (the much more terse treatment in) Lars Hörmander, *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis* (Springer-Verlag, 1983), p. 136 (ISBN: 9783642614972) (doi: 10.1007/978-3-642-61497-2).

Now consider the constraint $h(\mathbf{x})$ and the alternate but equivalent constraint $k(\mathbf{x})$:

$$h(\mathbf{x}) = 0 \iff k(\mathbf{x}) = 0. \tag{2.74}$$

Then the translation of Eq. (2.73) to the alternate constraint function is

$$\int d^d x\, \delta[k(\mathbf{x})]\, |\nabla k|\, f(\mathbf{x}) = \oint_{k^{-1}(0)} f(\mathbf{x})\, dS(\mathbf{x}). \tag{2.75}$$

The right-hand side here is equivalent to that of Eq. (2.73), so eliminating the surface integral, we have

$$\int d^d x\, \delta[h(\mathbf{x})]\, |\nabla h|\, f(\mathbf{x}) = \int d^d x\, \delta[k(\mathbf{x})]\, |\nabla k|\, f(\mathbf{x}) \tag{2.76}$$

This hold for any test function $f(\mathbf{x})$, so

$$\delta[h(\mathbf{x})] = \delta[k(\mathbf{x})]\, \frac{|\nabla k|}{|\nabla h|}.$$

(δ-function chain rule for scalar constraints) (2.77)

Therefore, the transformation just involves the ratio of Euclidean vector-gradient lengths for the two constraint functions, evaluated in each case at the boundaries. These should both exist for a sensible surface-constraint function, since these act to define the normal vector.

2.4.5 Probability Density for Combinations of Random Variables

Now we can return to the main point. Suppose we have random variables $X_1, \ldots, X_N$, with probability density function $f_{\mathbf{x}}(x_1, \ldots, x_N) = f_{\mathbf{x}}(\mathbf{x})$. Now suppose we define a combination of the random variables

$$Y = h(X_1, \ldots, X_N). \tag{2.78}$$

Then in the same way as in the result (2.6) for the two-step random walk, we can compute the probability density $f_y(y)$ for Y by integrating over all possible values of $\mathbf{x}$, using a delta function to enforce the relation between the variables:

$$f_y(y) = \int d^N x\, f(\mathbf{x})\, \delta[y - h(\mathbf{x})]. \tag{2.79}$$

Then using Eq. (2.64), we have

$$f_y(y) = \oint_{h(\mathbf{x})=y} \frac{f(\mathbf{x})}{|\nabla h|}\, dS = \oint_{h(\mathbf{x})=y} \frac{f(x_1, \ldots, x_N)}{\sqrt{\sum_{j=1}^{N} \left(\frac{\partial h(\mathbf{x})}{\partial x_j}\right)^2}}\, dS,$$

(transformation law, arbitrary function of random variables) (2.80)

where the integration is over all $\mathbf{x}$ such that $y = h(\mathbf{x})$ [i.e., all possible sets of $(X_1, \ldots, X_N)$ that give the desired value Y], a set that forms a hypersurface S of dimension $N - 1$.

2.4.5.1 Example: Convolution

For example, taking

$$Y = X_1 + X_2, \tag{2.81}$$

we have $|\nabla h| = \sqrt{2}$, taking $x_2 = y - x_1$, and parameterizing the "surface" integral with $s = x_1 + x_2$, so that

$$ds = \sqrt{dx_1^2 + dx_2^2} = \sqrt{2}\, dx_1, \tag{2.82}$$

Eq. (2.80) becomes

$$f_y(y) = \int ds\, \frac{f(x_1, y - x_1)}{\sqrt{2}} = \int dx_1\, f_1(x_1)\, f_2(y - x_1), \tag{2.83}$$

where in the last equality we have assumed independence of X_1 and X_2. This result recovers the convolution (2.7).

2.4.5.2 Example: Quotient of Normal Deviates

As a slightly more complicated example, consider two standard-normal deviates X_1 and X_2, with (separable) joint distribution

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}. \tag{2.84}$$

Now consider the *quotient* $Y = X_1/X_2$ of the two variables, such that the transformation function is

$$h(x_1, x_2) = \frac{x_1}{x_2}. \tag{2.85}$$

Then the gradient norm is

$$|\nabla h| = \sqrt{\frac{1}{x_2^2} + \frac{x_1^2}{x_2^4}} = \frac{\sqrt{x_1^2 + x_2^2}}{x_2^2} = \frac{\sqrt{1 + y^2}}{|x_2|}, \tag{2.86}$$

where we have set $x_1 = yx_2$, taking x_2 as the independent variable. The line element for the ''surface'' integration is

$$ds = \sqrt{dx_1^2 + dx_2^2} = \sqrt{1 + y^2}\, dx_2, \tag{2.87}$$

and thus

$$\begin{aligned} f_y(y) &= \int ds\, \frac{|x_2|\, f(yx_2, x_2)}{\sqrt{1 + y^2}} \\ &= \frac{1}{2\pi} \int dx_2\, |x_2|\, e^{-x_2^2(1+y^2)/2} \\ &= \frac{1}{2\pi} \left(\frac{2}{1 + y^2} \right). \end{aligned} \tag{2.88}$$

Then we see that the distribution function for the quotient is

$$f_y(y) = \frac{1}{\pi(1 + y^2)}, \tag{2.89}$$

which is a standard Cauchy distribution.

Part II

Gaussian Stochastic Processes

Chapter 3

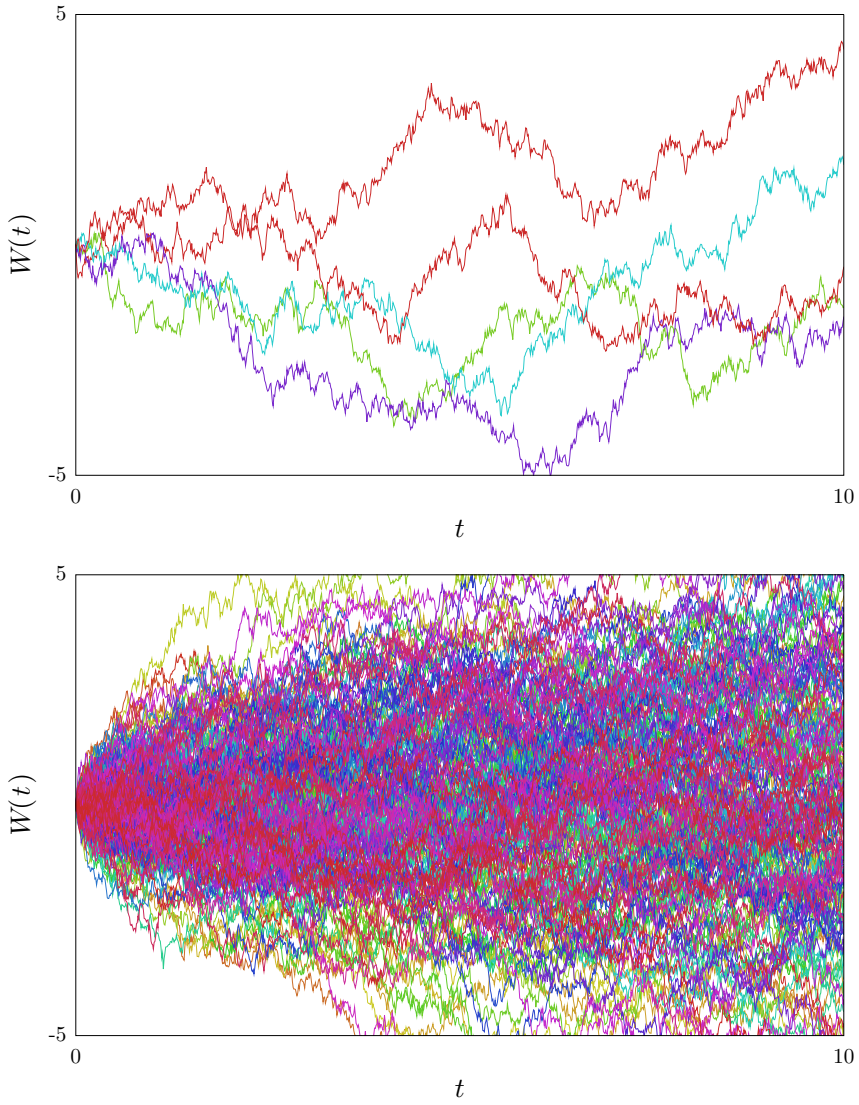
The Wiener Process

3.1 Continuous Random Walks: Wiener Process

Let's first define the **Wiener process** $W(t)$ as a sort of "ideal" random walk with arbitrarily small, independent steps taken arbitrarily often. That is, the usual random walk is usually taken to be a sequence of random steps of finite average (rms) size, taken after every finite time interval Δt. Recall from Section 2.1 that under fairly reasonable assumptions (such as the existence and finiteness of the one-step variance, and independence of the individual steps), that the central limit theorem guarantees that for long times, the probability density for the walker's location is Gaussian, *independent of the one-step distribution*, and the width (standard deviation) increases as $\sqrt{t}$. The Wiener process is essentially the idealized limit where $\Delta t \longrightarrow 0$, but where the size of each step *decreases* as appropriate to maintain the same asymptotic distribution. In this sense, the Wiener process is scale-free, since it has random steps on arbitrarily small time scales, and in fact is a fractal object: a Wiener process with appropriate but arbitrary rescaling (magnification) is still a Wiener process. We choose the Wiener process to correspond to a *symmetric* random walk, and so $W(t)$ is a normally distributed random variable with zero mean. To fix the scale of the random walk, we choose the variance of $W(t)$ to be simply t. That is, the (rms) width of the distribution is $\sqrt{t}$, as is characteristic of a diffusive process. In particular, $W(t)$ has the dimensions of $\sqrt{t}$. We can thus write the probability density for $W(t)$ as

$$P(W,t) = \frac{1}{\sqrt{2\pi t}}\, e^{-W^2/2t}. \tag{3.1}$$

Note that we have taken the convention that $W(0) = 0$, so that $P(W,0) = \delta(W)$. Again, it is important to emphasize that in view of the central-limit theorem, *any* simple random walk gives rise to a Wiener process in the continuous limit, independent of the one-step probability distribution (so long as the one-step variance is finite). To get an idea what these look like, 5 and 200 Wiener processes are respectively shown in the two plots below. (Actually, these are finite realizations of $\Delta W(t)$, with $\Delta t = 0.01$.)



Intuitively, $W(t)$ is a function that is continuous but everywhere nondifferentiable. (Of course, any such statement necessarily includes the proviso that the statement is true except for possibly a set of realizations of zero measure.) Naturally, the first thing we will want to do is to develop the analogue of the derivative for the Wiener process. We can start by defining the **Wiener increment**

$$\Delta W(t) := W(t + \Delta t) - W(t) \tag{3.2}$$

corresponding to a time interval Δt. Again, ΔW is a normally distributed random variable with zero mean and variance Δt. Note again that this implies that the rms amplitude of ΔW scales as $\sqrt{\Delta t}$. We can understand this intuitively since it is the *variances* that add for successive steps in a random walk, not the standard deviations. Mathematically, we can write the variance as

$$\left\langle\!\left\langle (\Delta W)^2 \right\rangle\!\right\rangle = \Delta t, \tag{3.3}$$

where the double angle brackets $\langle\!\langle\ \rangle\!\rangle$ denote an ensemble average over all possible realizations of the Wiener process. This relation suggests the notion that *second*-order terms in ΔW contribute at the same level as *first*-order terms in Δt, thinking about both of these variables in a small-time expansion of the evolution. In the infinitesimal limit of $\Delta t \longrightarrow 0$, we will write $\Delta t \longrightarrow dt$ and $\Delta W \longrightarrow dW$. Then $dW(t)$ is the **Wiener differential**, which is a fundamental object underlying stochastic calculus, and our analogue of the

derivative of the Wiener process. Thought of as a "signal," it is everywhere discontinuous. Notice that it is somewhat unusual: one sometimes writes a noise process as

$$\xi(t) := \frac{dW(t)}{dt}, \tag{3.4}$$

but this object is singular (i.e., has unbounded variance at any given time t), because as $\Delta t \longrightarrow 0$,

$$\frac{\Delta W}{\Delta t} \sim \frac{\sqrt{\Delta t}}{\Delta t} = \frac{1}{\sqrt{\Delta t}} \longrightarrow \infty. \tag{3.5}$$

It is possible to work with this singular fraction so long as you are careful with it, in the same sense that you can work with the singular delta function. We will tend to stick to the notation of differentials dt and $dW(t)$, but note that while dW is "zero," it is not "quite as small as" dt.

There is, in fact, a deeper connection of dW with the delta function. If we think of dW as a temporal "noisy" signal, the reason dW/dt is singular is that it contains contributions from *all frequencies with equal weights*—it's **white noise**—and that's the reason why the Wiener process contains random steps on all time scales. However, the total power for such a system, if the power in any band is finite, must be infinite. This is consistent with the fact that dW/dt diverges on average. On the other hand, if we do anything that *limits the bandwidth* of this signal, such as convolution with a finite function, or using a bandpass filter, it makes the resulting signal finite and well-behaved. Of course, any physical calculation or physical system involves just such a procedure, say, via dissipation through friction. This is exactly analogous to the delta function. The difference is that in the delta function, the frequency components have well-defined relative phases, while for the Wiener process they have effectively random phases.

3.2 Stochastic Process

Now that we have introduced a canonical example of a stochastic process—the Wiener process—it is worth stepping back for a minute to clarify what exactly we *mean*. Simply put, a **stochastic process** is a collection of random variables, indexed by some other variable. This could be written, for example, as $X(t)$ (or alternately and commonly as X_t), where at each value of time (where t is a real, continuous time variable), this variable has some random value, which may or may not be independent of its value at other times. The stochastic process could also be indexed by some *discrete* variable, and written for example as X_n, where n is an integer. Discrete processes could be thought of as samples of continuous processes [e.g., where $X_n = X(n\Delta t)$ for some time step Δt]. As we have already seen for the Wiener process, this connection is useful, and physicists often prefer to think of continuous processes as being *defined* in terms of discrete processes. This was our approach to the Wiener process above, and when confronted by a quagmire of infinitesimals, this point of view is often a handy way of making sense of things.

Stochastic processes can be classified in a number of different ways. One important distinction is termed **separability**, which comes up as something of a technical point with continuous-time processes. Basically, the idea is that if a joint probability density

$$f(x_1, t_1; x_2, t_2; \ldots) \tag{3.6}$$

[where x_j defines the distribution of $X(t_j)$] defines a stochastic process, then it is separable. Discrete processes are clearly separable. Continuous processes are where things get tricky; in this case a probability density of this form on a dense, countable subset of the set of all possible times should define the process. Since we will think of continuous processes as limits of discrete processes (in the sense of convergence of functionals of the processes), we won't have occasion to worry about separability—it's built into the construction of the continuous processes, as far as we're concerned.

An example of a simple type of stochastic process is a **completely independent** process. This means that the probability density factorizes,

$$f(x_1, t_1; x_2, t_2; \ldots) = f(x_1, t_1)\, f(x_2, t_2) \ldots, \tag{3.7}$$

which means that the values of $X(t)$ at different times are completely independent. This does *not* apply to $W(t)$, for example, although it does apply to $dW(t)$. A **Bernoulli trial** is a completely independent process where the individual density $f(x_j, t_j) = f(x_j)$, so the law is time independent—the process is just the same random experiment, repeated over and over. Again $dW(t)$ is an example of a Bernoulli trial. Bernoulli trials are an extreme example of a **stationary process**, which defines the notion of a process whose behavior does not drift in time. More generally, a process is stationary if a the form of any probability density $f(x_1, t_1; x_2, t_2; \ldots; x_n, t_n)$ describing the process is invariant under a shift of the time or the space variables. Both $W(t)$ and $dW(t)$ are stationary.

A stochastic process that, at any given time, has a zero expected change in the immediate future is called a **martingale**. For a discrete process, we can write this requirement mathematically as

$$\langle X_{n+1} | X_n = x_n \rangle = x_n, \tag{3.8}$$

where the conditional expectation value is defined in terms of the conditional probability

$$\langle X_{n+1} | X_n = x_n \rangle = \int dx_{n+1}\, x_{n+1}\, f(x_{n+1}, t_{n+1} | x_n, t_n). \tag{3.9}$$

More colloquially, we can write the martingale condition (3.8) as

$$\langle X_{n+1} \rangle = X_n, \tag{3.10}$$

where the expected value is understood to be conditioned on the particular value of X_n, once it is know. For a continuous-time process, we can then write the martingale condition as

$$\langle X(t + dt) \rangle = X(t), \tag{3.11}$$

in the spirit of representing continuous-time processes as effective discrete processes. Martingales are important in the theory of stochastic processes. The Wiener process $W(t)$ is a martingale, but $dW(t)$ is not [or, more properly, the discrete version $\Delta W(t)$ is not]. There are also less-restrictive forms of the martingale property, which apply for example to **sub-martingales** [where $\langle X(t + dt) \rangle \geq X(t)$] and **super-martingales** [where $\langle X(t + dt) \rangle \leq X(t)$].

3.2.1 Markov process

An important type of stochastic process is the **Markov process**. Colloquially, a Markov process is one in which the future behavior depends only on the *present* state of the process, not on the past. More precisely, we can express this in terms of a constraint on the conditional joint density

$$f(x_1, t_1; x_2, t_2; \ldots | x_0, t_0; x_{-1}, t_{-1}; \ldots) = f(x_1, t_1; x_2, t_2; \ldots | x_0, t_0), \tag{3.12}$$

where we assume the times to be ordered in the same way as the indices, $\cdots < t_{-1} < t_0 < t_1 < \cdots$, and x_0 here plays the role of the ''present state'' at the ''present time'' t_0. (A discrete-time Markov process is often called a **Markov chain**.) In fact, breaking the joint density via a conditional density [recall Eq. (1.18)] as

$$f(x_1, t_1; x_2, t_2; \ldots | x_0, t_0) = f(x_2, t_2; x_3, t_3; \ldots | x_1, t_1; x_0, t_0)\, f(x_1, t_1 | x_0, t_0), \tag{3.13}$$

the Markov condition (3.12) allows us to simplify this to

$$f(x_1, t_1; x_2, t_2; \ldots | x_0, t_0) = f(x_2, t_2; x_3, t_3; \ldots | x_1, t_1)\, f(x_1, t_1 | x_0, t_0). \tag{3.14}$$

Continuing in this way, we can factor the joint density into a product of simple conditional densities as

$$f(x_1, t_1; x_2, t_2; \ldots | x_0, t_0) = f(x_1, t_1 | x_0, t_0)\, f(x_2, t_2 | x_1, t_1)\, f(x_3, t_3 | x_2, t_2) \cdots . \tag{3.15}$$

Suppose, for simplicity, we throw out the dependence on times t_3 and later (e.g., by integrating over all possible x_3, x_4, …):

$$f(x_1, t_1; x_2, t_2 | x_0, t_0) = f(x_2, t_2 | x_1, t_1)\, f(x_1, t_1 | x_0, t_0). \tag{3.16}$$

Now integrating over all possible x_1 removes its dependence on the left-hand side:

$$f(x_2,t_2|x_0,t_0) = \int dx_1\, f(x_2,t_2|x_1,t_1)\, f(x_1,t_1|x_0,t_0).$$

(Chapman–Kolmogorov equation) (3.17)

This is an important result, called the **Chapman–Kolmogorov equation**; it is specific to Markov processes. For a non-Markovian process, we would instead need to have

$$f(x_2,t_2|x_0,t_0) = \int dx_1\, f(x_2,t_2|x_1,t_1;x_0,t_0)\, f(x_1,t_1|x_0,t_0) \tag{3.18}$$

in general for this conditional density; additionally, we could integrate over all possible values of x_0 to find

$$f(x_2,t_2) = \int dx_1\, f(x_2,t_2|x_1,t_1)\, f(x_1,t_1). \tag{3.19}$$

This equation is superficially similar to the Chapman–Kolmogorov equation, but again it is valid even for non-Markov processes—the conditional nature of the probabilities makes a huge difference.

3.3 Itō Calculus

Now that we have a white-noise process in hand, we will explore the formalism for handling this, particularly for handling the evolution of systems that are *driven* by white noise. It turns out that adding a white-noise stochastic process changes the basic structure of the calculus for treating the evolution equations. In particular, the usual Riemann integral is undefined for stochastic processes. There is more than one formulation to treat stochastic processes, but we will start out with **Itō calculus**,[1] which is the one most commonly used in treating quantum systems. We will start by showing how to *use* this calculus, since the rules are a bit different than what you're probably used to, and then we will justify the rules of usage.

3.3.1 Usage

First, let's review the usual calculus in a slightly different way. A differential equation

$$\frac{dy}{dt} = \alpha(y,t) \tag{3.20}$$

can be instead written in terms of differentials as

$$dy = \alpha\, dt. \tag{3.21}$$

The basic rule in the familiar *deterministic* calculus is that $(dt)^2 = 0$. To see what we mean by this, we can try calculating the differential dz for the variable $z = e^y$ in terms of the differential for dy as follows:

$$dz = e^{y+dy} - e^y = z\left(e^{\alpha\, dt} - 1\right). \tag{3.22}$$

Expanding the exponential and applying the rule $(dt)^2 = 0$, we find

$$dz = z\alpha\, dt. \tag{3.23}$$

This is, of course, the same result as that obtained by using the chain rule to calculate dz/dy and multiplying through by dy. The point here is that calculus breaks up functions and considers their values within short intervals Δt. In the infinitesimal limit, the quadratic and higher order terms in Δt end up being too small to contribute.

[1] After mathmatician Kiyoshi Itō (伊藤 清). Itō's name is also commonly transliterated as Ito or Itô, with Itō himself using "Itô."

In Itō calculus, we have an additional differential element dW representing white noise. The basic rule of Itō calculus is that $dW^2 = dt$, while $dt^2 = dt\,dW = 0$. We will justify this later, but to use this calculus, we simply note that we "count" the increment dW as if it were equivalent to $\sqrt{dt}$ in deciding what orders to keep in series expansions of functions of dt and dW. As an example, consider the stochastic differential equation (SDE)

$$dy = \alpha(y,t)\,dt + \beta(y,t)\,dW. \tag{3.24}$$

We obtain the corresponding differential equation for $z = e^y$ by expanding to *second* order in dy:

$$dz = e^y\left(e^{dy} - 1\right) = z\left(dy + \frac{(dy)^2}{2}\right). \tag{3.25}$$

Only the dW component contributes to the quadratic term; the result is

$$dz = z\left(\alpha + \frac{\beta^2}{2}\right)dt + z\beta\,dW. \tag{3.26}$$

The extra β^2 term is crucial in understanding many phenomena that arise in continuous-measurement processes.

3.3.2 Itō Rule: Justification

We now want to show that the Wiener differential dW satisfies the Itō rule $dW^2 = dt$. We already noted above that by definition, the *ensemble average* of $(\Delta W)^2$ is equal to Δt. However, in the infinitesimal limit, we will show that $dW^2 = dt$ holds *without* the ensemble average. This is surprising, since dW is a stochastic quantity, while dt obviously is not. To show this, consider the probability density function for $(\Delta W)^2$, which we can obtain by a transforming the Gaussian probability density for ΔW:

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\Delta t}}e^{-(\Delta W)^2/2\Delta t}. \tag{3.27}$$

We accomplish this for the coordinate transformation $y = f(x)$ by the transformation (see Section 1.2.1)

$$P_y(y)\,dy = \sum_{x\in f^{-1}(y)} P_x(x)\,dx. \tag{3.28}$$

Then we may write

$$P\left[(\Delta W)^2\right] = \frac{e^{-(\Delta W)^2/2\Delta t}}{\sqrt{2\pi\,\Delta t\,(\Delta W)^2}}, \tag{3.29}$$

so that the probability density for $(\Delta W)^2$ has the form of a gamma distribution (Section 10.8.3.4). In particular, the mean and variance of this distribution for $(\Delta W)^2$ are

$$\left\langle\!\left\langle (\Delta W)^2 \right\rangle\!\right\rangle = \Delta t \tag{3.30}$$

and

$$\mathrm{Var}\left[(\Delta W)^2\right] = 2(\Delta t)^2, \tag{3.31}$$

respectively. To examine the continuum limit, we will sum the Wiener increments over N intervals of duration $\Delta t_N = t/N$ between 0 and t. The corresponding Wiener increments are

$$\Delta W_n := W[(n+1)\Delta t_N] - W(n\Delta t_N). \tag{3.32}$$

Now consider the sum of the squared increments

$$\sum_{n=0}^{N-1}(\Delta W_n)^2, \tag{3.33}$$

which corresponds to a random walk of N steps, where a single step has average value t/N and variance $2t^2/N^2$. According to the central limit theorem, for large N the sum (3.33) is a Gaussian random variable with mean t and variance $2t^2/N$. In the limit $N \longrightarrow \infty$, the variance of the sum vanishes, and the sum becomes t with certainty. Symbolically, we can write

$$\int_0^t [dW(t')]^2 := \lim_{N\to\infty} \sum_{n=0}^{N-1} (\Delta W_n)^2 = t = \int_0^t dt'. \tag{3.34}$$

For this to hold over any interval $(0, t)$, we must make the formal identification

$$dt = dW^2. \tag{3.35}$$
(Itō rule)

This means that even though dW is a random variable, dW^2 is not, since it has no variance when integrated over any finite interval. Incidentally, we can also write down a similar expression for dW with itself, but at *different* times. The basis of this relation is the observation that $\langle\langle \Delta W(t)\, \Delta W(t') \rangle\rangle = 0$ for time increments $\Delta t < |t - t'|$, since the Wiener increments are independent. By a similar argument to the one above, the variance vanishes in the continuum limit—the variance of $\Delta W(t)\, \Delta W(t')$ is bounded above by the variance of $[\Delta W(t)]^2$—and thus it also follows that $dW(t)\, dW(t') = 0$ with certainty for $t \neq t'$, and thus we need not have an explicit ensemble average when replacing this product by zero.

3.3.3 Ensemble Averages

Finally, we need to justify a relation useful for averaging over noise realizations, namely that

$$\langle\langle y\, dW \rangle\rangle = 0 \tag{3.36}$$
(Itō ensemble average)

for a solution $y(t)$ of Eq. (3.24). This makes it particularly easy to compute averages of functions of $y(t)$ over all possible realizations of a Wiener process, since we can simply set $dW = 0$, even when it is multiplied by y. We can see this as follows. Clearly, $\langle\langle dW \rangle\rangle = 0$. Also, Eq. (3.24) is the continuum limit of the discrete relation

$$y(t + \Delta t) = y(t) + \alpha \Delta t + \beta \Delta W(t). \tag{3.37}$$

This discrete form here turns out to be the defining feature of Itō calculus, as we will see. Thus, $y(t)$ depends on $\Delta W(t - \Delta t)$, but is independent of $dW(t)$, which gives the desired result, Eq. (3.36). This gives the important feature of Itō calculus that makes it useful for computing ensemble averages: at a given time, the state of the noise process and the state of the system are independent. In particular, it is simple to write down an equation for the ensemble average of Eq. (3.24),

$$d\langle\langle y(t) \rangle\rangle = \langle\langle \alpha(y, t) \rangle\rangle\, dt, \tag{3.38}$$

which we obtain simply by setting $dW \longrightarrow 0$ in the SDE.

This leads us to some common terminology. Any process $y(t)$ satisfying an SDE of the form of Eq. (3.24) with no deterministic term ($\alpha = 0$) satisfies

$$d\langle\langle y(t) \rangle\rangle = \langle\langle y(t + dt) - y(t) \rangle\rangle = 0. \tag{3.39}$$

Any process satisfying this average condition is called a **martingale**, and is special in that each step in time is *unbiased* as a random walk.

3.3.4 Correlation Function

Now we are in a position to justify the ‘‘whiteness’’ of the noise. Recalling the singular noise signal

$$\xi(t) = \frac{dW(t)}{dt}, \tag{3.40}$$

let's compute the ensemble average

$$\langle\!\langle \xi(t)\,\xi(t')\rangle\!\rangle, \tag{3.41}$$

which is just the correlation function of the noise signal. Note that the *ensemble* average here can just as well be replaced by a *time* average, to get the correlation function in the time-averaged sense. If $t \neq t'$, we can simply write

$$\langle\!\langle \xi(t)\,\xi(t')\rangle\!\rangle = \left\langle\!\!\left\langle \frac{dW(t)\,dW(t')}{dt\,dt'} \right\rangle\!\!\right\rangle = \frac{\langle\!\langle dW(t)\rangle\!\rangle\langle\!\langle dW(t')\rangle\!\rangle}{dt\,dt'} = 0 \qquad (t \neq t'), \tag{3.42}$$

since in this case $dW(t)$ and $dW(t')$ are statistically independent. However, if $t = t'$, then

$$\langle\!\langle \xi(t)\,\xi(t)\rangle\!\rangle = \frac{\langle\!\langle (dW)^2\rangle\!\rangle}{(dt)^2} = \frac{1}{dt} \longrightarrow \infty. \tag{3.43}$$

Thus we see the divergent behavior. In fact, we can get the normalization from

$$\int dt\,\langle\!\langle \xi(t)\,\xi(t')\rangle\!\rangle = 1, \tag{3.44}$$

since there is a contribution of $dt \cdot (1/dt) = 1$ from the integration point $t = t'$, and no contributions from any other point in the integration range. Thus, we can infer that $\xi(t)$ is **delta-correlated**:

$$\langle\!\langle \xi(t)\,\xi(t')\rangle\!\rangle = \delta(t - t'). \tag{3.45}$$
(white-noise correlation function)

This justifies the notion of dW [equivalently, $\xi(t)$] as representing *white* noise, since the power spectrum is the Fourier transform of the correlation function according to the Wiener–Khinchin theorem, which in this case turns out to be a constant function over all frequencies. Note also the peculiarity that everything in this derivation carries through *without* the ensemble average, so that in Itō calculus, $\xi(t)\xi(t') = \delta(t - t')$.

3.3.5 Diffusion

The noise term in the Itō stochastic differential equation (SDE)

$$dy = \alpha(y,t)\,dt + \beta(y,t)\,dW \tag{3.46}$$

causes, as you might expect, *diffusion* of the trajectories $y(t)$. To see this, we need the evolution of the *width* of the ensemble. Using the Itō rule

$$d(y^2) = 2y\,dy + (dy)^2 = (2\alpha y + \beta^2)\,dt + 2\beta y\,dW, \tag{3.47}$$

we find the mean-square trajectory

$$d\langle\!\langle y^2\rangle\!\rangle = \langle\!\langle (2\alpha y + \beta^2)\,dt + 2\beta y\,dW\rangle\!\rangle = \langle\!\langle 2\alpha y + \beta^2\rangle\!\rangle dt. \tag{3.48}$$

Then defining the ensemble variance by

$$V_y := \langle\!\langle (y - \langle\!\langle y\rangle\!\rangle)^2\rangle\!\rangle = \langle\!\langle y^2\rangle\!\rangle - \langle\!\langle y\rangle\!\rangle^2, \tag{3.49}$$

we can use

$$\begin{aligned} d\langle\!\langle y\rangle\!\rangle &= \langle\!\langle \alpha\rangle\!\rangle dt \\ d\left[\langle\!\langle y\rangle\!\rangle^2\right] &= 2\langle\!\langle y\rangle\!\rangle d\langle\!\langle y\rangle\!\rangle = 2\langle\!\langle y\rangle\!\rangle\langle\!\langle \alpha\rangle\!\rangle dt, \end{aligned} \tag{3.50}$$

to write the variance evolution as

$$\begin{aligned} dV_y &= d\langle\!\langle y^2\rangle\!\rangle - d\left[\langle\!\langle y\rangle\!\rangle^2\right] \\ &= \Big[2\Big(\langle\!\langle \alpha y\rangle\!\rangle - \langle\!\langle \alpha\rangle\!\rangle\langle\!\langle y\rangle\!\rangle\Big) + \langle\!\langle \beta^2\rangle\!\rangle\Big]\,dt. \end{aligned} \tag{3.51}$$
(SDE variance evolution)

Thus, the variance is affected by gradients of α with y, or "spatial" dependence of the drift coefficient that can stretch or compact the distribution. This is the deterministic component of the variance. The noise part of the equation also contributes the β^2 term, so that the noise always tends to increase the ensemble variance, thus causing diffusion.

3.3.5.1 Fokker–Planck Equation

The evolution of the mean (3.50) and variance (3.51) are equivalent to the mean and variance according to the *deterministic* Fokker–Planck equation for the probability density $f(y,t)$ (Problem 3.3)

$$\partial_t f(y,t) = -\partial_y \alpha(y,t) f(y,t) + \frac{1}{2}\partial_y^2 \beta^2(y,t) f(y,t). \quad \text{(equivalent Fokker–Planck equation)} \tag{3.52}$$

In fact, this Fokker–Planck equation turns out to be the correct one to evolve the ensemble density. The standard form for the Fokker–Planck equation in one dimension is

$$\partial_t f(y,t) = -\partial_y A(y,t) f(y,t) + \frac{1}{2}\partial_y^2 D(y,t) f(y,t), \quad \text{(general Fokker–Planck equation)} \tag{3.53}$$

where $A(y,t)$ is the **drift coefficient**, and $D(y,t)$ is the **diffusion coefficient**. Thus, we identify the stochastic drift coefficient $\alpha(y,t)$ with the Fokker–Planck drift $A(y,t)$, while we identify the *squared* stochastic coefficient $\beta^2(y,t)$ with the diffusion coefficient $D(y,t)$. (For an alternate connection between stochastic trajectories and diffusion-type equations, see Section 7.1.)

To prove this, let's review a couple of concepts regarding probability theory. The **conditional probability density** $f(y,t|y_0,t_0)$, is a probability density in y, with $f(y,t|y_0,t_0)\,dy$ representing the probability density for finding the particle between y and $y+dy$ at time t, *given* the particle was at y_0 at time t_0. This is distinct from the **joint density** $f(y,t;y_0,t_0)$, which is a probability density in both y and y_0, where $f(y,t;y_0,t_0)\,dy\,dy_0$ is the probability for finding the particle between y and $y+dy$ at time t *and* between y_0 and y_0+dy_0 at time t_0. The individual probability densities are given by integrating out the other variable,

$$f(y,t) = \int dy_0\, f(y,t;y_0,t_0), \qquad f(y_0,t) = \int dy\, f(y,t;y_0,t_0). \tag{3.54}$$

The joint and conditional densities are related by the conditional probability relation, which states that the probability for A *and* B to occur is the product of the probability for A *given* that B occured and the probability for B to occur:

$$f(y,t;y_0,t_0) = f(y,t|y_0,t_0)\, f(y_0,t_0). \tag{3.55}$$

For Markovian evolution, that is, evolution where the entire state of the system for *all future* times is determined by the state of the system at the *present* time, this means that $f(y,t)$ is determined by $f(y_0,t_0)$. In this case, the conditional density satisfies the Chapman–Kolmogorov equation (3.17)

$$f(y,t|y_0,t_0) = \int dy'\, f(y,t|y',t')\, f(y',t'|y_0,t_0) \tag{3.56}$$

which certainly seems a reasonable property of the conditional density: two steps of the evolution of the density may be composed into a single step by integrating over all possible intermediate values.

Now to derive the Fokker–Planck equation.[2] To do this, consider the evolution of the ensemble average of an arbitrary function $g[y(t)]$, where $y(t)$ is a solution to the SDE (3.46):

$$\begin{aligned} d\langle\!\langle g(y)\rangle\!\rangle &= \left\langle\!\!\left\langle g'(y)\,dy + \frac{1}{2}g''(y)\,(dy)^2 \right\rangle\!\!\right\rangle \\ &= \left\langle\!\!\left\langle \alpha(y,t)g'(y)\,dt + \frac{\beta^2(y,t)}{2}\,g''(y)\,dt \right\rangle\!\!\right\rangle. \end{aligned} \tag{3.57}$$

[2] As in Crispin Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer, 2004) (ISBN: 9783540707127), Section 4.3.5, p, 93.

We can obviously rewrite this as

$$\partial_t \langle\!\langle g(y) \rangle\!\rangle = \left\langle\!\!\left\langle \alpha(y,t)\,\partial_y g(y) + \frac{\beta^2(y,t)}{2}\,\partial_y^2 g(y) \right\rangle\!\!\right\rangle. \tag{3.58}$$

The operator acting on $g(y)$ on the right-hand side,

$$\mathcal{G} := \alpha(y,t)\,\partial_y + \frac{\beta^2(y,t)}{2}\,\partial_y^2, \tag{3.59}$$

(generator of the SDE)

is often called the **generator** or **infinitesimal generator** corresponding to the SDE (3.52), because it "generates" the average change in a function of y over an infinitesimal time step dt.

Now let us write out the ensemble average explicitly, using the conditional density $f(y,t|y_0,t_0)$ for $y(t)$:

$$\int dy\, g(y)\,\partial_t f(y,t|y_0,t_0) = \int dy\, f(y,t|y_0,t_0)\left[\alpha(y,t)\,\partial_y g(y) + \frac{\beta^2(y,t)}{2}\,\partial_y^2 g(y)\right]. \tag{3.60}$$

Integrating by parts and discarding boundary terms,

$$\int dy\, g(y)\,\partial_t f(y,t|y_0,t_0) = \int dy\, g(y)\left[-\partial_y \alpha(y,t) f(y,t|y_0,t_0) + \frac{1}{2}\partial_y^2 \beta^2(y,t) f(y,t|y_0,t_0)\right]. \tag{3.61}$$

Since $g(y)$ is arbitrary, we may equate the integrands, and thus $f(y,t|y_0,t_0)$ satisfies an equation with the form of the Fokker–Planck equation:

$$\partial_t f(y,t|y_0,t_0) = -\partial_y \alpha(y,t) f(y,t|y_0,t_0) + \frac{1}{2}\partial_y^2 \beta^2(y,t) f(y,t|y_0,t_0).$$

(Kolmogorov forward equation) (3.62)

This equation is called the **Kolmogorov forward equation**, from which the Fokker–Planck equation (3.53) follows by multiplying through by $f(y_0,t_0)$ and integrating over y_0.

The above argument may also be adapted to give an evolution equation in terms of the *initial* time t_0.[3] Then writing

$$\begin{aligned}\partial_{t_0} f(y,t|y_0,t_0) &= \lim_{\delta t\to 0} \frac{f(y,t|y_0,t_0+\delta t) - f(y,t|y_0,t_0)}{\delta t} \\ &= \lim_{\delta t\to 0} \frac{1}{\delta t}\int dy'\, f(y',t_0+\delta t|y_0,t_0)\big[f(y,t|y_0,t_0+\delta t) - f(y,t|y',t_0+\delta t)\big],\end{aligned} \tag{3.63}$$

where the first term takes advantage of $f(y',t_0+\delta t|y_0,t_0)$ acting as a normalized distribution in y', and the second term uses the Chapman–Kolmogorov equation (3.56). Under the assumption that the distributions $f(y,t|y_0,t_0)$ are continuous functions, the two conditional densities in the difference above can be expanded to lowest order in δt, with first- and higher-order terms not contributing in the limit $\delta t \longrightarrow 0$:

$$\partial_{t_0} f(y,t|y_0,t_0) = \lim_{\delta t\to 0} \frac{1}{\delta t}\int dy'\, f(y',t_0+\delta t|y_0,t_0)\big[f(y,t|y_0,t_0) - f(y,t|y',t_0)\big]. \tag{3.64}$$

Now using the forward equation (3.62) to replace the first conditional distribution,

$$\begin{aligned}&\partial_{t_0} f(y,t|y_0,t_0) \\ &\quad = \int dy'\Big(-\partial_{y'}\alpha(y',t_0) f(y',t_0|y_0,t_0) + \frac{1}{2}\partial_{y'}^2 \beta^2(y',t_0) f(y',t_0|y_0,t_0)\Big)\big[f(y,t|y_0,t_0) - f(y,t|y',t_0)\big].\end{aligned} \tag{3.65}$$

Integrating by parts, we have (after discarding surface terms)

$$-\partial_{t_0} f(y,t|y_0,t_0) = \int dy'\, f(y',t_0|y_0,t_0)\Big(\alpha(y',t)\,\partial_{y'} f(y,t|y',t_0) + \frac{1}{2}\beta^2(y',t)\,\partial_{y'}^2 f(y,t|y',t_0)\Big). \tag{3.66}$$

[3] The beginning of this argument is as in C. Gardiner, *op. cit.*, Section 3.6, p. 55.

Finally using $f(y', t_0|y_0, t_0) = \delta(y' - y_0)$, we can carry out the remaining integral to obtain

$$-\partial_{t_0} f(y,t|y_0,t_0) = \alpha(y_0,t_0)\,\partial_{y_0} f(y,t|y_0,t_0) + \frac{\beta^2(y_0,t_0)}{2}\partial_{y_0}^2 f(y,t|y_0,t_0).$$

(Kolmogorov backward equation) (3.67)

This peculiar partial differential equation for the *initial* values y_0 and t_0 is called the **Kolmogorov backward equation**. It has a form similar to the Fokker–Planck equation, except for the order of the derivatives and the coefficients, and the minus sign on the time derivative.

3.3.5.2 Multidimensional Fokker–Planck Equation

In multiple dimensions, it is fairly straightforward to generalize the equivalent Fokker–Planck equation (3.52). We can start with a multidimensional generalization of the SDE (3.46),

$$dx_i = \alpha_i(\mathbf{x},t)\,dt + \beta_{ij}(\mathbf{x},t)\,dW_j, \tag{3.68}$$

where repeated indices are summed. In this case, the equivalent Fokker–Planck equation becomes (Problem 3.2)

$$\partial_t f(\mathbf{x},t) = -\partial_i \alpha_i(\mathbf{x},t) f(\mathbf{x},t) + \frac{1}{2}\partial_i\partial_j D_{ij}(\mathbf{x},t) f(\mathbf{x},t),$$

(equivalent Fokker–Planck equation) (3.69)

where the diffusion tensor is

$$D_{ij} := \beta_{ik}\beta_{jk} = \beta_{ik}\beta_{kj}^{\mathsf{T}} = (\boldsymbol{\beta}\boldsymbol{\beta}^{\mathsf{T}})_{ij}$$

(equivalent Fokker–Planck equation) (3.70)

in terms of the noise-coupling matrix $\boldsymbol{\beta}$. Note that the Fokker–Planck equation (3.69) has the form of a continuity equation,

$$\partial_t f(\mathbf{x},t) = -\nabla\cdot\mathbf{j}(\mathbf{x},t), \tag{3.71}$$

if we identify

$$j_i(\mathbf{x},t) := \alpha_i(\mathbf{x},t) f(\mathbf{x},t) + \frac{1}{2}\partial_j D_{ij}(\mathbf{x},t) f(\mathbf{x},t) \tag{3.72}$$

as the probability current density. This makes the probability-conserving property of the Fokker–Planck equation clearly evident: the current density merely shuffles around the probability density without creating or destroying it.

3.3.6 Ornstein–Uhlenbeck Process

As an example of using Itō calculus, we will consider the **Ornstein–Uhlenbeck process**, which we can define as the solution of the damped equation driven by a Wiener process:

$$dy = -\gamma y\,dt + dW.$$

(3.73)
(Ornstein–Uhlenbeck process)

This equation is the **Langevin equation**. To solve this, we write the equation in the form

$$d\left(ye^{\gamma t}\right) = e^{\gamma t} dW, \tag{3.74}$$

which we can integrate to obtain

$$y(t)e^{\gamma t} - y_0 = \int_0^t e^{\gamma t'}\,dW(t'), \tag{3.75}$$

or simply

$$y(t) = y_0\,e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} dW(t'). \tag{3.76}$$

The first term is clearly a decaying transient due to the initial condition, while the second is a convolution of the Wiener process with an exponential kernel, effectively smoothing the white noise. Note that since the Wiener differentials are Gaussian random variables, we see from this that the Ornstein–Uhlenbeck process is the sum over Gaussian random variables and is thus itself Gaussian. It is thus sufficient to completely characterize it by computing the mean and autocorrelation function. The mean is simply given by a decaying transient induced by the initial condition,

$$\langle\!\langle y(t) \rangle\!\rangle = y_0\, e^{-\gamma t}, \tag{3.77}$$

since in Itō calculus we compute ensemble averages by setting $dW = 0$. The correlation function is given by (taking $t' > t$)

$$\begin{aligned}
\langle\!\langle y(t)\, y(t') \rangle\!\rangle &= y_0^2 e^{-\gamma(t+t')} + \int_0^t ds \int_0^{t'} ds'\, e^{-\gamma(t-s)} e^{-\gamma(t'-s')} \left\langle\!\!\left\langle \frac{dW(s)}{ds} \frac{dW(s')}{ds'} \right\rangle\!\!\right\rangle \\
&= y_0^2 e^{-\gamma(t+t')} + \int_0^t ds \int_0^{t'} ds'\, e^{-\gamma(t-s)} e^{-\gamma(t'-s')} \delta(s-s') \\
&= y_0^2 e^{-\gamma(t+t')} + \int_0^t ds\, e^{-\gamma(t-s)} e^{-\gamma(t'-s)} \\
&= y_0^2 e^{-\gamma(t+t')} + e^{-\gamma(t+t')} \int_0^t ds\, e^{-2\gamma s} \\
&= \left(y_0^2 - \frac{1}{2\gamma} \right) e^{-\gamma(t+t')} + \frac{1}{2\gamma} e^{-\gamma(t'-t)}.
\end{aligned} \tag{3.78}$$

If we regard y_0 as an belonging to an ensemble of variance $1/2\gamma$, or if we consider the limit $t \longrightarrow \infty$ (with $t - t'$ fixed), we can ignore the transient part and take the correlation function to be

$$\langle\!\langle y(t)\, y(t') \rangle\!\rangle = \frac{1}{2\gamma} e^{-\gamma|t-t'|}, \tag{3.79}$$

(Ornstein–Uhlenbeck correlation)

where we have use the fact that for the real correlation function, $\langle\!\langle y(t)\, y(t') \rangle\!\rangle = \langle\!\langle y(t')\, y(t) \rangle\!\rangle$. The Ornstein–Uhlenbeck process, although corresponding to Gaussian noise, does thus not have independent increments at different times. We see that the damping introduces a ''memory'' in the dynamics. Further, since the correlation function is exponential, we immediately see that the power spectral density for the Ornstein–Uhlenbeck process is Lorentzian, and thus decays asymptotically as ω^{-2}, and thus corresponds to ''$1/f^2$'' noise. But also notice that the correlation function is finite for all times, and thus the damping, which has cut off the high frequencies, has made the noise bounded and well-behaved.

3.3.6.1 Brownian Motion

The Ornstein–Uhlenbeck process is a model for Brownian motion, if we use it to model the velocity of a particle subject to friction and to frequent ''kicks,'' as from collisions with many background-gas atoms:

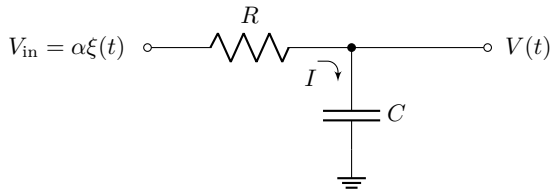
$$dv = -\gamma v\, dt + dW. \tag{3.80}$$

We are assuming that we are dealing with a scaled velocity such that the units come out right. Again, we can equivalently write this in the (possibly) more familiar form

$$\partial_t v = -\gamma v + \xi(t), \tag{3.81}$$

so that we have the usual equation of motion for the velocity, but driven by a white noise ''force'' $\xi(t)$.

The Ornstein–Uhlenbeck process also corresponds to a white-noise voltage signal $\alpha\xi(t)$ passing through a low-pass filter, with no connection at the output $V(t)$.



To see this, note that the current I flowing through the resistor is

$$I(t) = \frac{\alpha\xi(t) - V(t)}{R}, \tag{3.82}$$

and the voltage across the capacitor is related to the current by

$$\partial_t V = \frac{I}{C} = \frac{\alpha\xi(t) - V(t)}{RC}, \tag{3.83}$$

which we can write as

$$\partial_t V = -\gamma V + \alpha\gamma\xi(t), \tag{3.84}$$

where $\gamma = 1/RC$, as we expect for the low-pass filter. Thus, the output of the low-pass filter corresponds to a scaled Ornstein–Uhlenbeck process, and thus a physical signal despite the idealized input.

Note that from Eq. (3.79), an Ornstein–Uhlenbeck process of the form

$$dy = -\gamma y\, dt + dW \tag{3.85}$$

has $\langle\langle y^2 \rangle\rangle = 1/2\gamma$, and thus an rms fluctuation

$$y_{\text{rms}} = \frac{1}{\sqrt{2\gamma}}. \tag{3.86}$$

If we write Eq. (3.84) in the form

$$dV = -\gamma V\, dt + \alpha\gamma\, dW, \tag{3.87}$$

and then we let $t \longrightarrow t/(\alpha\gamma)^2$ and $W \longrightarrow W/\alpha\gamma$, we have

$$dV = -\frac{1}{\alpha^2\gamma} V\, dt + dW, \tag{3.88}$$

which is in the form of (3.85). Thus, the rms output voltage of the low-pass filter is

$$V_{\text{rms}} = \alpha\sqrt{\frac{\gamma}{2}}, \tag{3.89}$$

(amplitude of filtered white noise)

for an input signal of $\alpha\xi(t)$. Note that the rms output voltage *increases* as the square root of the filter bandwidth ($\gamma/2\pi = 1/2\pi RC$ is the corner frequency of the low-pass filter), and so this result shows that the transmitted *power* is proportional to the width of frequency band passed by the filter. Evidently, α has the dimensions of $\text{V}/\sqrt{\text{Hz}}$.

3.4 Stratonovich Calculus

The main alternative to Itō calculus is **Stratonovich calculus**, which we will introduce primarily to gain more insight into Itō calculus. Although Stratonovich calculus has some aesthetically nice features, it is often easier to perform calculations in Itō form, and we will mainly stick to Itō equations in our discussion of quantum measurement.

Consider the deterministic ODE

$$dy = \alpha(y(t), t)\, dt. \tag{3.90}$$

This is the continuous limit of the discrete relation

$$y(t + \Delta t) = y(t) + \alpha(y(\tau), \tau)\, \Delta t, \tag{3.91}$$

where τ is an *arbitrary* time in the range $[t, t + \Delta t]$. This is because the formal (implicit) solution of (3.90) is given by the Riemann integral

$$y(t) = y_0 + \int_0^t \alpha(y(t'), t')\, dt'. \tag{3.92}$$

The Riemann integral is approximated by successively finer refinements of discrete ''rectangle'' areas of width Δt, where the area of each rectangle is determined by the value of the integrand at *any* point within the interval Δt.

Returning to the SDE

$$dy = \alpha(y(t), t)\, dt + \beta(y(t), t)\, dW(t), \tag{3.93}$$

we must also interpret the solution of this equation in terms of the implicit integral

$$y(t) = y_0 + \int_0^t \alpha(y(t'), t')\, dt' + \int_0^t \beta(y(t'), t')\, dW(t'). \tag{3.94}$$

The first integral is an ordinary Riemann integral, but the second is of the form

$$\int_0^t \beta(y(t'), t')\, dW(t') = \int_0^t \beta(y(t'), t')\, \frac{dW}{dt'}\, dt'. \tag{3.95}$$

Due to the highly singular nature of $\xi(t) = dW/dt$, the Riemann integral does not in general exist. The way to save this is that in the successive finite approximations, if you consistently pick the *same* point within each interval of equal length Δt, the integral is defined. However, the *result* that you get by evaluating the integral will depend on your choice. Of course, if β is constant or even a smooth function of time—in the case of **additive noise**—then this won't be a problem, since the result amounts to integrating dW to get $W(t)$. The problem arises in the case of **multiplicative noise** when β is a function of y. Thus, when we regard the SDE (3.93) as the continuum limit of the finite-difference equation

$$y(t + \Delta t) = y(t) + \alpha(y(t), t)\, \Delta t + \beta(y(\tau), \tau)\, \Delta W(t), \tag{3.96}$$

where $\tau \in [t, t + \Delta t]$, we obtain a different limit depending on where in the interval we choose τ. In view of Eq. (3.37), Itō calculus takes the choice $\tau = t$, while Stratonovich calculus takes the choice $\tau = t + \Delta t/2$. Since we expect different results according to what calculus we intend to use to solve the equation, we must use a notation to distinguish the calculus that goes with the SDE. Thus, we will write as usual

$$dy = \alpha(y(t), t)\, dt + \beta(y(t), t)\, dW(t) \tag{3.97}$$
(notation: Itō SDE)

to denote an Itō SDE, while we will use the special notation

$$dy = \alpha(y(t), t)\, dt + \beta(y(t), t) \circ dW(t) \tag{3.98}$$
(notation: Stratonovich SDE)

to refer to a Stratonovich equation.

3.4.1 Example: Stochastic Integration

To illustrate the consequences of this choice, we will compute the sample Itō integral

$$I = \int_{t_0}^t W(t')\, dW(t'), \tag{3.99}$$

and compare it to the Stratonovich integral

$$J = \int_{t_0}^{t} W(t') \circ dW(t'), \tag{3.100}$$

of the same form, to see that they indeed give different results. Note that if these were ordinary Riemann integrals, we would simply have

$$\int_{t_0}^{t} f(t')\, df(t') = \frac{1}{2}\left[f^2(t) - f^2(t_0)\right] \tag{3.101}$$

for a sufficiently well-behaved function $f(t)$. The Itō integral follows from the continuum limit of the N-step approximation (with $t_j := t_0 + j(t-t_0)/N$)

$$\begin{aligned}
I &= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_j)\Delta W(t_j) \\
&= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_j)\left[W(t_{j+1}) - W(t_j)\right] \\
&= \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[W(t_j) + W(t_{j+1}) + W(t_j) - W(t_{j+1})\right]\left[W(t_{j+1}) - W(t_j)\right] \\
&= \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[W^2(t_{j+1}) - W^2(t_j)\right] - \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[W(t_{j+1}) - W(t_j)\right]^2 \\
&= \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[\Delta W(t_j)\right]^2 \\
&= \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \frac{1}{2}\int_{t_0}^{t} [dW(t')]^2 \\
&= \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \frac{1}{2}(t - t_0).
\end{aligned} \tag{3.102}$$

The shortcut for this calculation is to notice that by the Itō rule, $d(W^2) = 2W\, dW + (dW)^2 = 2W\, dW + dt$, so that

$$I = \int_{t_0}^{t} W(t')\, dW(t') = \frac{1}{2}\int_{t_0}^{t} \left[d(W^2) - dt\right] = \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \frac{1}{2}(t - t_0). \tag{3.103}$$

We thus see how the Itō rule enforces the choice of approximating integration intervals by the *beginning* point of the interval.

In the Stratonovich case, the integrand is evaluated at the intermediate time $t_{j+1/2} := (t_{j+1} + t_j)/2$. However, let us evaluate the most general case $t_{j+s} := (1-s)t_j + s t_{j+1}$, with $s \in [0,1]$, where Itō corresponds to $s = 0$ and Stratonovich to $s = 1/2$:

$$\begin{aligned}
J_s &:= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_{j+s})\Delta W(t_j) \\
&= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_{j+s})\left[W(t_{j+1}) - W(t_j)\right] \\
&= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_j)\left[W(t_{j+1}) - W(t_j)\right] + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]\left[W(t_{j+1}) - W(t_j)\right] \\
&= I + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]\left[W(t_{j+1}) - W(t_j)\right] \\
&= I + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]\left[W(t_{j+1}) - W(t_{j+s})\right] + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]^2
\end{aligned} \tag{3.104}$$

The second term corresponds to the continuum limit of a product of independent Wiener increments, which vanishes according to the last argument of Section (3.3.2). The last term is the sum of squared, independent Wiener increments corresponding to time intervals $s\Delta t$, and is thus given by $s(t-t_0)$. Thus,

$$J_s = I + s(t-t_0) = \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] + \left(s - \frac{1}{2}\right)(t-t_0). \tag{3.105}$$

In particular, the Stratonovich integral is

$$J = J_{1/2} = \frac{1}{2}\left[W^2(t) - W^2(t_0)\right]. \tag{3.106}$$

Note that this is exactly the same result as if we had just used *ordinary* calculus, so that in Stratonovich calculus it is appropriate to take $d(W^2) = 2W\,dW$. That is, the usual chain rule applies in Stratonovich calculus, so that $(dW)^2 = 0$. We will prove this after we see how to convert between Itō calculus and Stratonovich calculus.

3.4.2 Itō–Stratonovich Conversion

Itō and Stratonovich SDEs in general give different results for the *same* coefficients, but in what sense are they equivalent? That is, how do we convert between Itō and Stratonovich SDEs? Suppose we have the Itō SDE

$$dy(t) = \alpha(y,t)\,dt + \beta(y,t)\,dW(t), \tag{3.107}$$

and the Stratonovich SDE

$$dy(t) = \bar{\alpha}(y,t)\,dt + \beta(y,t) \circ dW(t). \tag{3.108}$$

Then what we will show is that these two SDEs are equivalent if and only if

$$\bar{\alpha} = \alpha - \frac{1}{2}\beta\partial_y\beta. \tag{3.109}$$

(Itō–Stratonovich conversion)

Clearly, this distinction only matters in the case of multiplicative noise. To show this, recall that the Itō SDE is the continuum limit of the discrete relation

$$y(t+\Delta t) = y(t) + \alpha(y(t),t)\Delta t + \beta(y(t),t)\Delta W(t), \tag{3.110}$$

while the Stratonovich SDE is the continuum limit of

$$y(t+\Delta t) = y(t) + \bar{\alpha}(y(t),t)\,\Delta t + \beta(y(t+\Delta/2), t+\Delta t/2)\,\Delta W(t). \tag{3.111}$$

Now noting that

$$\beta(y(t+\Delta t/2), t+\Delta t/2) = \beta(y(t),t) + \partial_y\beta(y(t),t)\left[\beta(y(t),t)\,\Delta W^{(1/2)}(t)\right] + O(\Delta t), \tag{3.112}$$

where the factor in square brackets is Δy, $\Delta W^{(1/2)}(t) := W(t+\Delta t/2) - W(t)$ is a Wiener "half increment," and we are dropping terms of order Δt and higher since this expression will be multiplied by ΔW. We can thus use this result to write

$$y(t+\Delta t) = y(t) + \bar{\alpha}(y(t),t)\,\Delta t + \beta(y(t),t)\,\Delta W(t) + \beta(y(t),t)\,\partial_y\beta(y(t),t)\,\Delta W(t)\,\Delta W^{(1/2)}(t). \tag{3.113}$$

In the continuous limit, we can write

$$\Delta W(t)\,\Delta W^{(1/2)}(t) \longrightarrow \frac{dt}{2}, \tag{3.114}$$

since $[\Delta W^{(1/2)}]^2 \longrightarrow dt/2$ with certainty, and the product of $\Delta W^{(1/2)}$ from two different time intervals will converge to zero with certainty. Thus, the continuum limit of (3.111) is the Itō-form SDE

$$dy = \left[\bar{\alpha} + \frac{1}{2}\beta\partial_y\beta\right]\,dt + \beta\,dW. \tag{3.115}$$

Since this is the continuum limit of the same equation as the Stratonovich form (3.108), we can thus conclude that the Itō and Stratonovich forms are equivalent in the case

$$\bar{\alpha} = \alpha - \frac{1}{2}\beta\partial_y\beta. \tag{3.116}$$

Thus, when writing down an SDE, we again see that it is crucial to specify which calculus it assumes, since the solution would otherwise be ambiguous.

3.4.3 Stratonovich Calculus and the Chain Rule

Recall that the Itō equation

$$dy = \alpha\, dt + \beta\, dW \tag{3.117}$$

transforms under the coordinate change $z = f(y)$ via an "extended chain rule" (often called **Itō's lemma**)

$$\begin{aligned} dz = df(y) &= f'(y)\, dy + \frac{1}{2}f''(y)(dy)^2 \\ &= \left[f'(y)\,\alpha + \frac{1}{2}f''(y)\,\beta^2\right] dt + f'(y)\,\beta\, dW. \end{aligned} \qquad \text{(Itō chain rule, } z = f(y)\text{)} \tag{3.118}$$

We will now show that in Stratonovich calculus, the "extra" β^2 term does not appear, and thus the usual chain rule applies (and thus $dW^2 = 0$).

Thus, consider the *Stratonovich* equation

$$dy = \alpha\, dt + \beta \circ dW \tag{3.119}$$

under the same transformation $z = f(y)$. The equivalent Itō equation is

$$dy = \left[\alpha + \frac{1}{2}\beta\partial_y\beta\right] dt + \beta\, dW, \tag{3.120}$$

and now we can use Itō rules to accomplish the transformation:

$$\begin{aligned} dz &= f'(y)\, dy + \frac{1}{2}f''(y)(dy)^2 \\ &= \left[f'(y)\left(\alpha + \frac{1}{2}\beta\partial_y\beta\right) + \frac{1}{2}f''(y)\beta^2\right] dt + f'(y)\beta\, dW. \end{aligned} \tag{3.121}$$

Now transform this back into Stratonovich form:

$$dz = \left[f'(y)\left(\alpha + \frac{1}{2}\beta\partial_y\beta\right) + \frac{1}{2}f''(y)\beta^2 - \frac{1}{2}f'(y)\beta\partial_z[f'(y)\beta]\right] dt + f'(y)\beta \circ dW. \tag{3.122}$$

Noting that

$$\partial_z = \frac{1}{f'(y)}\partial_y, \tag{3.123}$$

we can write the last dt term as

$$\begin{aligned} -\frac{1}{2}f'(y)\beta\partial_z[f'(y)\beta] &= -\frac{1}{2}\beta\partial_y[f'(y)\beta] \\ &= -\frac{1}{2}f''(y)\beta^2 - \frac{1}{2}f'(y)\beta\partial_y\beta. \end{aligned} \tag{3.124}$$

Thus, this term, which we obtained from switching from Itō to Stratonovich form, cancels the other two dt terms that involve β, leaving

$$dz = f'(y)\,\alpha\, dt + f'(y)\,\beta \circ dW. \qquad \text{(Stratonovich chain rule, } z = f(y)\text{)} \tag{3.125}$$

Thus, Stratonovich calculus obeys the usual chain rule, and we have no need for terms of order dW^2.

3.4.4 Comparison

We will now summarize the differences between Itō and Stratonovich SDEs, and then explain why we tend to favor Itō calculus. For Itō calculus:

- The rules of stochastic integration are slightly more complicated than for ordinary integration, since the ordinary chain rule does not apply (i.e., $dW^2 = dt$).
- The solution $y(t)$ of an Itō SDE and the driving Wiener process $dW(t)$ are statistically independent at equal times, so that ensemble averages are simply computed by setting $dW = 0$.
- An Itō SDE is ''natural'' as the continuum limit of an evolution constructed by a discrete-step process, since

  $$dy = \alpha(y)\,dt + \beta(y)\,dW \tag{3.126}$$

  is the continuous limit of

  $$y(t + \Delta t) = y(t) + \alpha(y(t))\,\Delta t + \beta(y(t))\,\Delta W(t). \tag{3.127}$$

On the other hand, for Stratonovich calculus:

- The rules of stochastic integration are those of ordinary Riemann integration, since the usual chain rule applies (i.e., $dW^2 = 0$).
- The solution $y(t)$ of a Stratonovich SDE and the driving Wiener process $dW(t)$ are *not* statistically independent at equal times. This is clear from the above Itō–Stratonovich conversion, since setting $dW = 0$ in a Stratonovich SDE does not give the same result as setting $dW = 0$ in the equivalent Itō SDE. In fact, the easiest rule for computing an ensemble average is to convert the SDE to Itō form and then set $dW = 0$.
- A Stratonovich SDE is ''natural'' as the idealization of a *physical* noise process in the following sense. If one models a stochastic system as being driven by a *physical* noise of finite bandwidth and bounded variance, then the normal rules of calculus apply. For example if dO represents an Ornstein–Uhlenbeck process, then we could model a system by the SDE

  $$dy = \alpha\,dt + \beta\,dO. \tag{3.128}$$

  If you take the white-noise limit for the driving process ($\gamma \longrightarrow 0$), then dO goes over to dW, but because the ODE always obeyed the rules of ordinary calculus, the white-noise limit

  $$dy = \alpha\,dt + \beta \circ dW. \tag{3.129}$$

  should be interpreted as a Stratonovich SDE.

Also note that most proofs, including the construction of Stratonovich calculus, are usually proved in Itō calculus, so its advantages tend to outweigh its peculiarities. For handling quantum measurements, we will often want to compute ensemble averages to obtain unconditioned master equations, and we will also in general construct continuous measurements as limits of discrete processes of the form (3.127). Thus, we will virtually always use Itō-form SDEs to handle continuous quantum measurements.

3.5 Particle Subject to a Stochastic Force

As an important example to tie together the stochastic-process concepts that we have developed so far, we will treat in some depth the dynamics of a particle subject to a stochastic force, together with damping and other external forces. In particular consider the equations of motion

$$\begin{aligned} dx &= \frac{p}{m}\,dt \\ dp &= \big[F(x,p;t) - \gamma p\big]\,dt + \sigma\,dW, \end{aligned}$$

(damped, stochastically forced particle) (3.130)

with F defining an external force, γ the damping coefficient, and σ defining the magnitude of the stochastic force. This is for a particle in one spatial dimension, but the generalization to multiple degrees of freedom is straightforward.

Also, note that γ could depend on x and p (and even t), as could σ. Note that if σ depends on p, then the particle is coupled to multiplicative noise (Section 3.4), and it is important to specify which calculus to use in treating the SDE system. As discussed in Section 3.4.4, if the noise arises as the continuum limit of a physical noise process of finite bandwidth, then Eqs. (3.130) should be interpreted in the Stratonovich sense. On the other hand, if the noise arises due to a process like spontaneous emission, where the emission probability depends on the present state (or arises in terms of a discrete Poisson process, which goes over to white noise in the limit of frequent emission events, as in Section 11.4), then the equations of motion should be interpreted in the Itō sense. For concreteness we will treat this system in the Itō sense, although it is useful to remember that as we make assumptions on σ in the discussion below, the importance of this distinction drops away.

From Eqs. (3.68) and (3.69), we can write out the equivalent Fokker–Planck equation for the probability density $f(x,p;t)$ of the particle state as

$$\partial_t f(x,p;t) = -\frac{p}{m}\partial_x f + \partial_p\big[F(x,p;t) - \gamma p\big] f + \frac{1}{2}\partial_p^2 \sigma^2 f. \tag{3.131}$$

(equivalent Fokker–Planck equation)

Note that any dependence of F, γ, and σ on p enforces the ordering of the derivative operators as shown; in the case where there is no such momentum dependence, the derivative operators ∂_p can commute with these functions and operate only on the probability density. But again notice that dependence on x causes no problem; for example, the diffusion term can have the form $\sigma^2(x)\,\partial_p^2 f$ in the case that σ depends only on position.

3.5.1 Free-Particle Limit

One of the simplest limits of the above equations of motion is to assume vanishingly small damping ($\gamma \longrightarrow 0$) and external force ($F \longrightarrow 0$). In this case we can integrate the momentum equation to obtain

$$p(t) = p(0) + \int_0^t dt'\, \sigma(x,p;t')\, dW(t'). \tag{3.132}$$

Then putting this into the position equation and integrating, we find

$$x(t) = x(0) + \frac{p(0)\,t}{m} + \frac{1}{m}\int_0^t dt' \int_0^{t'} dt''\, \sigma(x,p;t'')\, dW(t''). \tag{3.133}$$

It is difficult to carry this further in the general case. However, in the special case where σ is constant, we can characterize the transport of x quite precisely. In this case, we obtain

$$x(t) = \frac{\sigma}{m}\int_0^t dt'\, W(t'), \tag{3.134}$$

where we are also ignoring transients by setting $x(0) = p(0) = 0$. The integral factor here is important, and for obvious reasons it goes by the names **integrated Brownian motion** and **Brownian area** (this is also the iterated integral I_{10} that we discuss in Section 10.3), and here we will write it as

$$I(t) := \int_0^t dt'\, W(t'). \tag{3.135}$$

(integrated Brownian motion)

Since $I(t)$ is a linear combination of Gaussian increments, it is itself Gaussian, so we can characterize it fully by computing its mean and correlation. The mean is simple, since in Itō calculus we just set $dW = 0$:

$$\big\langle\!\big\langle I(t) \big\rangle\!\big\rangle = 0. \tag{3.136}$$

To compute the variance (see also Problem 10.1), we can write

$$\left\langle\!\left\langle I^2(t) \right\rangle\!\right\rangle = \left\langle\!\left\langle \left[\int_0^t dt' \int_0^{t'} dW(t'') \right]^2 \right\rangle\!\right\rangle. \tag{3.137}$$

Then using the integral identity $\int_0^t dt' \int_0^{t'} dt'' g(t', t'') = \int_0^t dt'' \int_{t''}^t dt' g(t', t'')$ to interchange the order of integration, we have

$$\begin{aligned} \left\langle\!\left\langle I^2(t) \right\rangle\!\right\rangle &= \left\langle\!\left\langle \left[\int_0^t dW(t'') \int_{t''}^t dt' \right]^2 \right\rangle\!\right\rangle \\ &= \left\langle\!\left\langle \left[\int_0^t (t - t'')\, dW(t'') \right]^2 \right\rangle\!\right\rangle. \end{aligned} \tag{3.138}$$

Then expanding out the product, using $dW(t)\, dW(t') = \delta(t - t')\, dt^2$,

$$\left\langle\!\left\langle I^2(t) \right\rangle\!\right\rangle = \left\langle\!\left\langle \int_0^t (t - t'')^2\, dt'' \right\rangle\!\right\rangle = \frac{t^3}{3}. \tag{3.139}$$

This implies a covariance

$$\left\langle\!\left\langle I(t) I(t') \right\rangle\!\right\rangle = \frac{1}{3} \big[\min\{t, t'\}\big]^3, \tag{3.140}$$

(covariance of integrated Brownian motion)

because for example if $t' > t$, then the Wiener increments associated with the time interval (t, t') are uncorrelated with any increments associated with the interval $(0, t)$, so the time interval (t, t') does not contribute to the correlation function.

Thus, the stochastically driven, undamped particle (3.134) tends to a Gaussian distribution in x, whose width grows as $(\sigma/m)t^{3/2}$, which is one power of t higher than the $t^{1/2}$ dependence we expect for diffusive growth in momentum.

3.5.2 Thermal Equilibrium

Now consider the particle equations (3.130) *with* damping at rate γ but no external force. This could correspond, for example, to a particle moving (in one dimension) in a static fluid at some temperature T. For simplicity we will assume constant temperature, noise coefficient σ, and damping rate γ.

First, note that the momentum equation in this case is basically a scaled Ornstein–Uhlenbeck process:

$$dp = -\gamma p\, dt + \sigma\, dW. \tag{3.141}$$

In fact, rescaling time via $\tilde{t} = \sigma^2 t$ and $d\tilde{W}(\tilde{t}) = \sigma\, dW(t)$, so that $d\tilde{W}^2 = [\sigma\, dW(t)]^2 = \sigma^2\, dt = d\tilde{t}$, gives

$$dp = -\frac{\gamma}{\sigma^2} p\, d\tilde{t} + d\tilde{W}, \tag{3.142}$$

which has the form of the Ornstein–Uhlenbeck process in Eq. (3.73). Then the results following from that analysis carry through here with the replacements $\gamma \longrightarrow \gamma/\sigma^2$ and $t \longrightarrow \sigma^2 t$. For example, the mean from Eq. (3.77) is only transient,

$$\left\langle\!\left\langle p(t) \right\rangle\!\right\rangle = p_0 e^{-\gamma t}, \tag{3.143}$$

(momentum mean)

and the correlation function (3.78) becomes

$$\left\langle\!\left\langle p(t)\, p(t') \right\rangle\!\right\rangle = \left(p_0^2 - \frac{\sigma^2}{2\gamma} \right) e^{-\gamma(t+t')} + \frac{\sigma^2}{2\gamma} e^{-\gamma|t'-t|}, \tag{3.144}$$

(momentum correlation)

where the first term is again transient and the second persists in steady state, and $t, t' \geq 0$.

In thermal equilibrium, we have

$$\frac{1}{2}m\langle\!\langle v^2\rangle\!\rangle = \frac{\langle\!\langle p^2\rangle\!\rangle}{2m} = \frac{1}{2}k_{\mathrm{B}}T. \tag{3.145}$$

We can obtain $\langle\!\langle p^2\rangle\!\rangle = \sigma^2/2\gamma$ in steady state from the last term in Eq. (3.144), which gives

$$\sigma = \sqrt{2m\gamma k_{\mathrm{B}}T}, \tag{3.146}$$
(temperature related to diffusion)

which fixes the noise coefficient in terms of the temperature and damping. Thus, for example, Eq. (3.144) becomes

$$\langle\!\langle p(t)\,p(t')\rangle\!\rangle = mk_{\mathrm{B}}T\,e^{-\gamma|t'-t|} \tag{3.147}$$

in terms of temperature, after ignoring any transients.

The position equation (3.130) then says that the position is the (scaled, shifted) integrated Ornstein–Uhlenbeck process:

$$x(t) = x(0) + \int_0^t dt'\,p(t'). \tag{3.148}$$

For the ''pure'' Ornstein–Uhlenbeck process (3.73),

$$dy = -\gamma y\,dt + dW, \tag{3.149}$$

we can define the integrated process by

$$G(t) := \int_0^t dt'\,y(t'). \tag{3.150}$$

Then clearly $G(t)$ is again a Gaussian process, with

$$\langle\!\langle G(t)\rangle\!\rangle = 0. \tag{3.151}$$

Computing the correlation function, we start with

$$\langle\!\langle G(t)\,G(t')\rangle\!\rangle = \int_0^t ds \int_0^{t'} ds'\,\langle\!\langle y(s)\,y(s')\rangle\!\rangle = \frac{1}{2\gamma}\int_0^t ds \int_0^{t'} ds'\left(e^{-\gamma|s-s'|} - e^{-\gamma(s+s')}\right). \tag{3.152}$$

where we used Eq. (3.78) for the Ornstein–Uhlenbeck correlation function, setting $y_0 = 0$ for simplicity but keeping the transient term, which is important when integrating the correlation function. Then if $t' \geq t$ (so that $t' \geq s$), we can split the second integral to simplify the absolute value:

$$\langle\!\langle G(t)\,G(t')\rangle\!\rangle = \frac{1}{2\gamma}\int_0^t ds\left[\int_0^s ds'\,e^{-\gamma(s-s')} + \int_s^{t'} ds'\,e^{\gamma(s-s')}\right] - \frac{1}{2\gamma}\int_0^t ds \int_0^{t'} ds'\,e^{-\gamma(s+s')}. \tag{3.153}$$

Carrying out the integration and simplifying gives

$$\langle\!\langle G(t)\,G(t')\rangle\!\rangle = \frac{1}{2\gamma^3}\left[2\left(e^{-\gamma t} + e^{-\gamma t'} - 1\right) - e^{-\gamma(t+t')} - e^{-\gamma(t'-t)}\right] + \frac{t}{\gamma^2}. \tag{3.154}$$

Again, using the requirement that the correlation function should be symmetric in t and t', we find

$$\langle\!\langle G(t)\,G(t')\rangle\!\rangle = \frac{1}{2\gamma^3}\left[2\left(e^{-\gamma t} + e^{-\gamma t'} - 1\right) - e^{-\gamma(t+t')} - e^{-\gamma|t'-t|}\right] + \frac{\min\{t,t'\}}{\gamma^2}.$$
(integrated-Ornstein–Uhlenbeck correlation) (3.155)

Note that $\langle\!\langle G(0)\,G(0)\rangle\!\rangle = 0$, as expected. Note also that $\langle\!\langle G(t)\,G(t')\rangle\!\rangle \geq 0$, because we derived it as the integral of $\langle\!\langle y(t)\,y(t')\rangle\!\rangle \geq 0$.

Returning to the transport of the particle in the fluid, we are interested in the asymptotic growth of the variance of $G(t)$. For large times, Eq. (3.155) gives an asymptotic growth

$$\left\langle\!\left\langle G^2(t) \right\rangle\!\right\rangle \sim \frac{t}{\gamma^2} \tag{3.156}$$

to leading order in t. The same procedure with $x(t) = \int_0^t dt'\, p(t')/m$ and Eq. (3.144) for the momentum correlation function gives

$$\left\langle\!\left\langle x^2(t) \right\rangle\!\right\rangle \sim \left(\frac{\sigma}{m\gamma}\right)^2 t = \frac{2k_{\mathrm{B}}T}{m\gamma} t. \tag{3.157}$$

(position diffusion)

That is, $x(t)$ grows diffusively with diffusion coefficient $(\sigma/m\gamma)^2$. Thus, the damping "transfers" the diffusive behavior from momentum (which is driven by the Wiener process) to position. We will see this more directly in the following section.

3.5.3 Strong-Damping Limit: Brownian Motion Revisited

Back in Section 3.3.6.1, we commented that the Ornstein–Uhlenbeck process is a model for Brownian motion—as we just showed, a damped and stochastically driven momentum leads to diffusion in position. At the same time, the Wiener process $W(t)$ itself is often referred to as "Brownian motion," although in the physical model it enters via the force, not directly in the particle's position. Here we will make the connection more directly of the stochastic force leading to diffusion in position, in the case of strong damping.

Starting again with the equations of motion (3.130),

$$\begin{aligned} dx &= \frac{p}{m}\, dt \\ dp &= \big[F(x,p;t) - \gamma p\big] dt + \sigma\, dW, \end{aligned} \tag{3.158}$$

with the same comments about the (x,p)-dependence of γ and σ, we will begin by assuming a large damping coefficient γ. In the limit of large damping, we will assume that the momentum is always approximately in equilibrium with respect to the position state, so that we may perform an adiabatic elimination of momentum. Setting $dp \approx 0$ then leads to the adiabatic approximation

$$p \approx \frac{1}{\gamma}\left[F + \sigma\, \frac{dW}{dt}\right]. \tag{3.159}$$

In assuming large damping, we are essentially assuming that the damping term "absorbs" the entire effect of the external and fluctuating forces, provided we coarse-grain on time scales much longer than γ^{-1}. The difficulty here is that dW/dt is divergent and fluctuates on all time scales. We will proceed for now, but we will soon see that we need to be somewhat more careful with this approximation.

Putting Eq. (3.159) into the position equation $\dot{x} = p/m$, we find

$$dx \approx \frac{F}{m\gamma}\, dt + \frac{\sigma}{m\gamma}\, dW. \tag{3.160}$$

(particle motion, adiabatic approximation)

In the absence of external forces ($F = 0$), then we see directly that $x(t) \propto W(t)$, which justifies the name Brownian motion for $W(t)$. And again appealing to Eqs. (3.68) and (3.69), the equivalent Fokker–Planck equation for this approximate SDE is

$$\partial_t f(x;t) = -\partial_x \frac{F}{m\gamma} f + \frac{1}{2}\partial_x^2 \frac{\sigma^2}{m^2\gamma^2} f. \tag{3.161}$$

(Smoluchowski equation)

Here, F, γ, and σ can still depend on position and momentum; however, the dependence on momentum should be in the form of $\langle\!\langle p(x) \rangle\!\rangle = F(x,p;t)/\gamma$ from the adiabatic relation (3.159), because the large damping

"concentrates" the momentum around the mean value. This diffusion equation is the counterpart of the original Fokker–Planck equation (3.131) after momentum is adiabatically eliminated. In the case of coupling to a uniform bath in thermal equilibrium, σ is constant and given by Eq. (3.146), such that

$$\partial_t f(x,t) = -\frac{1}{m\gamma}\partial_x F f + \frac{k_\mathrm{B}T}{m\gamma}\partial_x^2 f, \tag{3.162}$$
(Smoluchowski equation)

after also assuming constant damping γ. This is called the **Smoluchowski equation**.[4]

Note that in the form (3.162), at any given temperature, the steady-state solution is independent of m and γ. In particular, assuming that the force may be derived from a potential, $F(x) = -V'(x)$, we can write

$$\partial_x V' f + k_\mathrm{B}T\partial_x^2 f = 0 \tag{3.163}$$

in steady state. Removing one derivative gives

$$\partial_x f = -\frac{V'}{k_\mathrm{B}T} f, \tag{3.164}$$

where we can see that any constant of integration vanishes by comparing this equation with its $x \longrightarrow -x$ counterpart. Integrating this equation from 0 to x gives

$$f(x) = f(0)\, e^{-[V(x)-V(0)]/k_\mathrm{B}T}, \tag{3.165}$$
(Boltzmass distribution)

which is just the Boltzmann distribution.

3.5.3.1 Momentum Distribution

Given the reduced distribution $f(x,t)$ to the Smoluchowski equation (3.162), it is useful to see how to reconstruct the information about momentum that is implicit in the solution. First, as we already mentioned, the adiabatic relation (3.159) yields directly the mean momentum at each position:

$$\langle\!\langle p(x) \rangle\!\rangle \approx \frac{F(x, \langle\!\langle p(x) \rangle\!\rangle; t)}{\gamma}. \tag{3.166}$$

Here, to be precise, the ensemble average is over all dW, but at fixed position x. The mean position corresponding to the entire ensemble of particles must be averaged over the distribution $f(x,t)$:

$$\langle\!\langle p(t) \rangle\!\rangle_x \approx \int_{-\infty}^{\infty} dx\, f(x,t)\, \frac{F(x, \langle\!\langle p(x) \rangle\!\rangle; t)}{\gamma}. \tag{3.167}$$
(ensemble-mean momentum)

Here, the subscript on the ensemble average emphasizes the average over the spatial distribution.

However, the same approach applied to the second moment $\langle\!\langle p^2 \rangle\!\rangle$ is problematic, because the square of Eq. (3.159) contains a term with the divergent factor $dW^2/dt^2 = 1/dt$. This is a result of the infinite power associated with the white-noise force dW/dt. This is a sign that we need to treat the adiabatic approximation with more care. To do this we return to the momentum equation of motion in Eqs. (3.130),

$$dp = \big[F(x,p;t) - \gamma p\big]dt + \sigma\, dW, \tag{3.168}$$

and note that the same procedure that we used for the Ornstein–Uhlenbeck process in Section 3.3.6 applies here. Thus, we may rewrite the solution to Eq. (3.168) as

$$p(t) = p(0)\, e^{-\gamma t} + \int_0^t dt'\, e^{-\gamma(t-t')} \left[F + \sigma\, \frac{dW(t')}{dt'}\right]. \tag{3.169}$$

[4] M. San Miguel and J. M. Sancho, "A Colored-Noise Approach to Brownian Motion in Position Space. Corrections to the Smoluchowski Equation," *Journal of Statistical Physics* **22**, 605 (1980) (doi: 10.1007/BF01011341), Eq. (2.39).

The essential point here is that the exponential convolution kernel here smooths the white-noise force, endowing it with finite power and removing the divergence in the second momentum moment. What we did before was to assume that the bracketed factor in the integrand changed slowly over the time scale γ^{-1}, and that $t \gg \gamma^{-1}$, so that we could make the replacement

$$e^{-\gamma(t-t')}\Theta(t-t') \longrightarrow \frac{1}{\gamma}\,\delta(t-t'-0^+). \tag{3.170}$$

Making this replacement and carrying out the integral leads directly to the previous adiabatic relation (3.159). However, we are only justified in doing this for the external-force term; the white-noise term is *not* constant over *any* time scale. Thus, we may write Eq. (3.169) as

$$p(t) = p(0)\,e^{-\gamma t} + \frac{F}{\gamma} + \sigma \int_0^t e^{-\gamma(t-t')}\,dW(t'). \tag{3.171}$$

Under the assumption that F, γ, and σ vary slowly over the damping time γ^{-1}, we may regard these as constants. Coarse graining over the damping time scale, we can take the long-time limit of the above equation to remove transient behavior:

$$p(x) = \frac{F}{\gamma} + \sigma \int_0^t e^{-\gamma(t-t')}\,dW(t'). \tag{3.172}$$

Now we regard the momentum as a function of the particle's position, which is approximately constant. Under these conditions, the first term is the same mean position that we found before in Eq. (3.166), while the second represents (small) Gaussian fluctuations about the mean. Thus, the momentum distribution is a Gaussian tightly localized about the mean, and all that remains is to characterize the second moment. For this, note that $p(x) - F/\gamma = p(x) - \langle\!\langle p(x)\rangle\!\rangle$ is proportional to an Ornstein–Uhlenbeck process. Thus, we may adapt Eq. (3.79) to give the variance

$$V_p(x) = \left\langle\!\!\left\langle \big(p(x) - \langle\!\langle p(x)\rangle\!\rangle\big)^2 \right\rangle\!\!\right\rangle = \frac{\sigma^2}{2\gamma} = mk_{\text{B}}T, \tag{3.173}$$

where the latter expression follows from Eq. (3.146), and in the adiabatic regime can be regarded as correct even with spatial variation in temperature.

Now when measuring the momentum distribution of the entire ensemble, we should average the momentum variance over the spatial distribution. However, it is more appropriate to average the second *moment* of momentum rather than the variance, because local variations in the mean momentum are reflected as variance in the *global* distribution. Thus, we require

$$\langle\!\langle p^2(x)\rangle\!\rangle = V_p(x) + \langle\!\langle p(x)\rangle\!\rangle^2 = \frac{\sigma^2}{2\gamma} + \frac{F^2}{\gamma^2}. \tag{3.174}$$

Then averaging this over the spatial distribution from the Smoluchowski equation (3.162), we find

$$\langle\!\langle p^2(t)\rangle\!\rangle_x = \int_{-\infty}^{\infty} dx\, f(x,t)\left[\frac{\sigma^2}{2\gamma} + \frac{F^2}{\gamma^2}\right].$$

(distribution-average of second momentum moment) (3.175)

At this point one can subtract the square of the distribution-averaged mean (3.167) to obtain the momentum variance of the whole distribution. However, Eq. (3.174) is already sufficient to characterize the momentum distribution in important cases such as the steady-state distribution in a potential well, where the mean momentum vanishes.

3.6 Exercises

Problem 3.1

Geometric Brownian motion describes the time-dependence price S of a stock according to the SDE

$$dS = \mu S\, dt + \sigma S\, dW. \tag{3.176}$$

where μ represents the steady growth of the stock value, and σ represents the stock volatility (which is assumed to be constant within this model). Show that

$$S(t) = S_0\, e^{(\mu - \sigma^2/2)t + \sigma W(t)} \tag{3.177}$$

satisfies the above SDE.

Problem 3.2

Show that the vector SDE (3.68)

$$dx_i = \alpha_i(\mathbf{x}, t)\, dt + \beta_{ij}(\mathbf{x}, t)\, dW_j \tag{3.178}$$

is equivalent to the Fokker–Planck equation (3.69),

$$\partial_t f(\mathbf{x}, t) = -\partial_i \alpha_i(\mathbf{x}, t) f(\mathbf{x}, t) + \frac{1}{2}\partial_i \partial_j D_{ij}(\mathbf{x}, t) f(\mathbf{x}, t), \tag{3.179}$$

where

$$D_{ij} := \beta_{ik}\beta_{jk} = \beta_{ik}\beta^{\mathsf{T}}_{kj} = (\boldsymbol{\beta}\boldsymbol{\beta}^{\mathsf{T}})_{ij}. \tag{3.180}$$

Problem 3.3

Recall the one-dimensional Fokker–Planck equation:

$$\partial_t f(p, t) = -\partial_x \alpha(x) f(x, t) + \frac{1}{2}\partial_x^2 D(x) f(x, t). \tag{3.181}$$

Since the drift coefficient $\alpha(x)$ and diffusion coefficient $D(x)$ are state-dependent, we can't write down a general solution, and the interpretation of these coefficients is a bit tricky. However, computing the equations of motion for the first two moments helps with the interpretation.

(a) Show that

$$\partial_t \langle x \rangle = \langle \alpha(x) \rangle\,. \tag{3.182}$$

Here, the expectation value refers to an average with respect to the distribution $f(x, t)$. Assume that $f(x, t)$ falls off quickly enough with $|x|$ that any boundary terms are negligible. Thus, "drift" of the distribution mean is caused both by the drift coefficient, averaged over the distribution.

(b) Show that

$$\partial_t V_x = 2\langle x\alpha(x) \rangle - 2\langle x \rangle \langle \alpha(x) \rangle + \langle D(x) \rangle\,, \tag{3.183}$$

under the same assumptions as in part (a), where $V_x := \langle x^2 \rangle - \langle x \rangle^2$. Thus, the diffusion coefficient causes spreading of the variance, again weighted by the distribution $f(x, t)$. Also, variations in the drift coefficient can cause the distribution to spread or contract, since different accelerations for parts of the distribution at different momenta can tear them apart or squish them together, depending on the relative accelerations.

Chapter 4

Financial Applications

An important application of stochastic calculus and stochastic processes in general is to modeling the fluctuations in prices of financial assets (stocks, bonds, etc.). Here we will work through the most famous model of asset pricing. It makes use of the theory of Wiener-driven SDEs that we have developed in Chapter 3. This development has a flaw, which is that the underlying Gaussian noise fails to predict extreme events, such as market crashes. In this sense, heavy-tailed processes (Chapter 12) are more appropriate for constructing such models. However, Gaussian models can be handled analytically, and so much insight can be gained from them.

4.1 Geometric Brownian Motion

The basic model that we will use for the dynamics of financial assets is **geometric Brownian motion**, which is specified by the Itō SDE

$$dS = \mu S\,dt + \sigma S\,dW. \tag{4.1}$$
(SDE for geometric Brownian motion)

Here, $S(t)$ is the asset price. The parameter μ is the expected rate of return in the sense of the ensemble average

$$d\langle\!\langle S\rangle\!\rangle = \mu\langle\!\langle S\rangle\!\rangle\,dt, \tag{4.2}$$

which says that $\langle\!\langle S\rangle\!\rangle \sim e^{\mu t}$. The parameter μ thus represents the **interest rate** in the absence of random fluctuations ($\sigma = 0$); the increase as $e^{\mu t}$ corresponds to continuously compounded interest. The parameter σ is the **volatility**, and controls the magnitude of the random fluctuations.

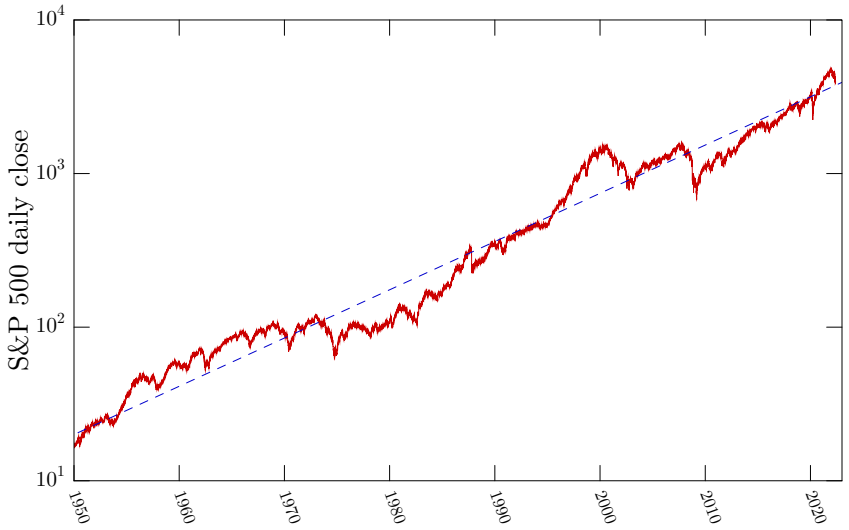
The solution to this Itō SDE is (Problem 3.1)

$$S(t) = S_0\,e^{(\mu-\sigma^2/2)t+\sigma W(t)}. \tag{4.3}$$
(geometric Brownian motion)

Because $\log S(t)$ has a Gaussian distribution, the distribution of $S(t)$ is said to be a **log-normal distribution**, and $S(t)$ is also called a **log-normal process**. Intuitively, this process models *fractional* changes in the asset price in each time interval, because dS/S changes as $\mu\,dt$ for deterministic interest and $\sigma\,dW$ for random fluctuations. These fractional changes mirrors the way financial changes occur in the real world.

Note that because of the interpretation of this SDE as an Itō equation, the volatility enters the exponentiated drift term. Also note that the noise here is multiplicative, but the functional form is simple enough to be amenable to analytic solution. This SDE may also be generalized to heavy-tailed noise, but the generalization is not trivial; we return to this in Section 12.7.2.

To further justify the plausibility of geometric Brownian motion for modeling financial data, the daily closing values of the S&P 500 stock-market index are plotted logarithmically below. Also shown is exponential-growth line at an annual rate of 7.5%. The index prices fluctuate around this exponential trend, although this noise is clearly not Gaussian. Thus the main value of studying geometric Brownian motion is in examining the structure of financial models; quantitatively speaking, this is all nonsense.



4.2 Financial Markets and Hedging

Given that $W(t)$ is equally likely to rise and fall, the price $S(t)$ of a financial asset from Eq. (4.3) can either rise or fall. Obviously, if you hold some of an asset—or **financial security** (tradable asset with some monetary value)—whose price falls, then you experience a financial loss. What may be somewhat more surprising is that it is possible to *gain* money when an asset loses value. The basic idea is to make a *bet* that the asset will go down in value, thereby profiting from its decline.

One means of implementing such a bet is an **option** or **option contract**. The purchaser of an option obtains the right to either buy or sell a security, at some time or in some time window, for some predetermined price. The price specified by the option is called the **strike price**. The buy vs. sell aspect breaks options down into two types:

- A **call option** gives the owner the right to *buy* a security at the strike price. The owner profits if the price of the security rises above the strike price by the time the option is exercised.
- A **put option** gives the owner the right to *sell* a security at the strike price. The owner profits if the price of the security falls below the strike price when the option is exercised.

In the case of a put option, the option holder profits from a downturn in the value of a security. The opposite obviously holds in the case of a call option. If the option is never exercised, the option simply expires, and the holder has lost the purchase price of the option. Because the value of an option is tied to the value of the underlying security (and it is *itself* a financial security), options are a type of **financial derivative**, the value of which is derived from the underlying security.

Besides being a means for financial traders to get rich as sin, why would anyone be interested in an option? A nice way to think about options is as an instrument for *mitigating risk*, or basically as an *insurance policy*. For example:

- An airline depends on the ability to buy large quantities of jet fuel at reasonable prices—fuel is the second-largest expense for airlines after labor. To guard against an unforeseen rise in fuel prices, an airline can buy call options on jet fuel. If fuel prices fall, the airline is out the (hopefully modest) price of the options, but the airline is safe if some world crisis causes a big price hike. But essentially, the airline is placing a bet that fuel prices go up; the airline "loses" the cost of the option if fuel prices don't exceed the strike price.
- A farmer depends on being able to sell a crop of corn for at least some minimum price in order to turn a product. In order to guard against a surplus of corn causing a crash in corn prices, the farmer could

buy put options. But viewed differently, the farmer places a bet that the corn price will crash, and will be out the price of the option if this doesn't happen.

This practice of mitigating risk is called **hedging**. In financial investing, holding options is a hedging method for a **portfolio**, or the total collection of securities that an investor holds. In particular, holding stocks along with put options reduces risk because the options rise in value when stocks fall, and vice versa.

There are also different classifications of options, depending on when they are exercised. The most important option types are:

- A **European option** can only be exercised at a particular, predefined point in time, the expiration date of the option or **at maturity**.
- An **American option** can be exercised at *any* time before the option matures (preferably at some time where the purchase or sale would be advantageous, because the price of the underlying security crossed through the strike price).

Being the most common possibilities, these are called **vanilla options**. Many other options are also possible. Because European options are the easiest to model mathematically, we will stick with them here.

A related derivative commonly used for hedging is the **futures contract** (or just **futures**). A futures contract is an agreement to buy or sell a security on a predefined future date. Actually, the agreement between two parties for one to sell and the other to buy is called a **forward contract**, and futures contracts are standardized forward contracts that are traded in financial markets. Unlike an option, the purchase or sale of the underlying security on the future date is *required*, whereas the option merely bestows the right to make the purchase or sale. Another related hedging maneuver is a **short sale**, where a trader borrows a security (like stock shares) and sells it at the current market price. The trader must return the security at a predefined time, but is betting that the security's price will fall by that time. If this happens, the trader profits, but if the security goes up in value the trader will be forced to rebuy at a higher price, potentially taking a large loss. Unlike options, the transaction in a futures contract or short sale is obligatory. In this sense, both futures and short sales are much riskier in terms of large potential losses than options, if the market doesn't go they way that the trader hopes. If a trader sells a security without actually owning it (i.e., sells short), this is also called a **short position** or **holding a security short**, as opposed to a **long position** or **holding a security long**, which is actually holding or owning the security. Importantly, a short position on a security is *debt*, and thus quantitatively amounts to a *negative* holding of that security. Also, note that while a short sale is analogous to but distinct from an option, because an option is tradable it can also be sold short.

Before proceeding, it is worth reiterating that hedging and risk mitigation are precisely the areas where Gaussian financial models can be problematic. They are often used to calculated how exposed an investor is, when implementing some investment strategy (especially when **leveraging**—investing/betting using borrowed money). Investment firms have gone broke by reassuring themselves based on Gaussian models that they were insulated from risk, after which a large financial crash (well outside the range of Gaussian prediction) came along and bankrupted them.

4.3 Black–Scholes Model

The most famous method for pricing options is based on geometric Brownian motion as in Eq. (4.1), and it is called the **Black–Scholes model** or **Black–Scholes–Merton model**.[1,2] We will assume that the option

[1] After Fischer Black and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* **81**, 637 (1973) (doi: 10.1086/260062); and Robert C. Merton "Theory of Rational Option Pricing," *The Bell Journal of Economics and Management Science* **4**, 141 (1973) (doi: 10.2307/3003143). This has been explained in many references, but a few good ones are Kurt Jacobs, *Stochastic Processes for Physicists: Understanding Noisy Systems* (Cambridge, 2010), Section 5.2 (ISBN: 9780511815980) (doi: 10.1017/CBO9780511815980); Rosario N. Mantegna and Eugene Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance*, (Cambridge, 1999) (ISBN: 9780511755767) (doi: 10.1017/CBO9780511755767); Crispin Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer, 2009) (ISBN: 9783540707127), Section 10.2.4.

[2] The work of Black, Scholes, and Merton was long preceded by the pioneering work of Bachelier, who considered Brownian models of option pricing, but through direct Brownian motion (rather than geometric motion). Bachelier, Louis, "The Theory

price O is a function of both the time t directly and of the underlying stock/security price $S(t)$. Applying the chain rule to compute the differential (keeping the stochastic differential to order dS^2 as appropriate for Itō calculus),

$$dO(S,t) = \frac{\partial O}{\partial t}\,dt + \frac{\partial O}{\partial S}\,dS + \frac{1}{2}\frac{\partial^2 O}{\partial S^2}\,dS^2, \tag{4.4}$$

and then using the SDE (4.1) and Itō rule $dW^2 = dt$,

$$dO = \left[\frac{\partial O}{\partial t} + \mu\frac{\partial O}{\partial S}S + \frac{\sigma^2}{2}\frac{\partial^2 O}{\partial S^2}S^2\right]dt + \sigma\frac{\partial O}{\partial S}S\,dW. \tag{4.5}$$

This gives an SDE for the evolution of the option price, driven by the stochastic evolution of the underlying stock.

4.3.1 Riskless Portfolio

To make use of the option SDE, consider a portfolio consisting of n shares of a stock or security of unit price $S(t)$ (where n need not be an integer), plus an option of value $O(t)$. This is an artificial portfolio, useful for computing the price of the option; it does not necessarily represent a good investment strategy (for an example, an investor might *want* risk in order to potentially increase profits). Then the portfolio value $\Pi(t)$ is given by

$$\Pi(t) = nS(t) + O(t). \tag{4.6}$$

The differential of this portfolio is

$$d\Pi = n\,dS + dO. \tag{4.7}$$

Then using (4.4) for dO, we find

$$d\Pi = \frac{\partial O}{\partial t}\,dt + \left(n + \frac{\partial O}{\partial S}\right)dS + \frac{1}{2}\frac{\partial^2 O}{\partial S^2}\,dS^2. \tag{4.8}$$

Note that the only stochastic component here is in the dS term; the dS^2 term is purely deterministic. The stochastic component of $d\Pi$ thus vanishes if we set

$$n = -\frac{\partial O}{\partial S}. \tag{4.9}$$
(risk-neutral condition)

Because the stochastic component vanishes under this condition, there are no fluctuations: the portfolio is now **risk-neutral**, at least over short time scales where $\partial O/\partial S$ is approximately constant. In this case the differential equation (4.8) for the portfolio value becomes

$$d\Pi = \frac{\partial O}{\partial t}\,dt + \frac{1}{2}\frac{\partial^2 O}{\partial S^2}\,dS^2, \tag{4.10}$$

which is now deterministic. Again using the SDE (4.1) and $dW^2 = dt$, the resulting differential equation is

$$\frac{d\Pi}{dt} = \frac{\partial O}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 O}{\partial S^2}S^2, \tag{4.11}$$

after dividing through by dt.

It is important to note the significance of the minus sign in the risk-neutral condition (4.9). Since we could well expect that $\partial O/\partial S > 0$ (i.e., the value of an option is proportionally larger for an increase in the security price, if say the strike price is proportionally larger), then the condition implies holding a *negative* quantity of the security. Remember that this is the same as a short position on the security. Alternatively, the portfolio could have a long position on the security, but then the option would have to be held short (in which case $\partial O/\partial S < 0$).

of Speculation'' (''Théorie de la Spéculation''), *Annales Scientifiques de l' École Normale Supérieure, Sér. 3* **17**, p. 21 (1900) (translated by D. May) `http://www.radio.goldseek.com/bachelier-thesis-theory-of-speculation-en.pdf`.

4.3.2 No Arbitrage

Another way to look at the risk-neutral portfolio (4.11) is that it should be equivalent to any *other* riskless security. An example is *risk-free bonds* (for example, United States Treasury bonds), which are essentially loans from a large, stable entity that pay back the loaned amount (the **principal**) at maturity, and pay back interest along the way. In practice, any *real* bond entails some risk (if the bond-issuer defaults on the interest or principal payments), but can be effectively risk-free. Such bonds will pay interest at the **riskless interest rate** r.

Now, the argument is that *any* risk-free security should accrue value at the same rate. To understand this requires the concept of **arbitrage**. If a particular security were priced in two different ways, say in different geographic locations (an inconsistency that is an example of a **price inefficiency** in the market), then a trader (**arbitrageur**) could make a risk-free profit by buying the security at the lower price and reselling at the higher price. This action drives up the lower price (by raising demand) and drives down the higher price (by raising supply) until an equilibrium at a single, global price for that security.

Now suppose that there are two different bonds that accrue interest at different rates; we will argue that there is an arbitrage opportunity here. Suppose that the two bonds have the same present value. Then one will be more valuable in the future. Then in a forward contract to exchange the two bonds, the trader buying the more valuable bond by trading for the less valuable one will obviously gain a risk-free profit at the expense of the seller. Because of this arbitrage opportunity, the forward contract itself have monetary value, which is why forward contracts are traded as futures contracts. In this case the value of the lower-interest bond plus the value of the futures contract should match the higher-interest bond, and effectively the interest rates are then the same.

Thus, an idealization is that arbitrageurs quickly remove any inefficiencies from the market, and so we assume that *all* riskless securities accrue interest at the same rate r. This is the **no-arbitrage assumption** behind the Black–Scholes model. This is often stated as the assumption that no risk-free profits are possible beyond the risk-free rate (any risk-entailing means of profit must yield expected profits at a higher rate, otherwise a trader would take the no-risk option).

Returning to the risk-free portfolio from Eq. (4.11), this means that the risk-neutral portfolio should accrue interest at the same rate:

$$\frac{d\Pi}{dt} = r\Pi. \tag{4.12}$$

Equating this expression with Eq. (4.11), rearranging, and using

$$\Pi = O + nS = O - \frac{\partial O}{\partial S}S \tag{4.13}$$

under the risk-neutral condition, we obtain the differential equation

$$\frac{\partial O}{\partial t} = rO - rS\frac{\partial O}{\partial S} - \frac{\sigma^2}{2}S^2\frac{\partial^2 O}{\partial S^2} \tag{4.14}$$

(Black–Scholes equation)

for the time-evolution of the option price in terms of the value of the underlying security. Note that the deterministic-growth parameter μ does not explicitly appear here, except of course in the time-dependence of $S(t)$.

4.3.3 Solution of the Option-Pricing Problem

To price the option at any time t, the idea is to regard the Black–Scholes equation (4.14) as a PDE, and to solve it for $O(S,t)$. The solution of course requires a boundary condition to be well-defined, which will depend on the option type. For a European call option, the option's value is well defined at the maturity time T, and the appropriate boundary condition is

$$O(S,T) = \max\{S(T) - K, 0\}, \tag{4.15}$$

(European-call-option value at maturity)

where K is the strike price. That is, if the security's value exceeds the strike price, then the holder of the option profits by buying the security at the strike price K and reselling it at the security price $S(T)$. Otherwise, the option is worthless. Similarly, for a European put option, the boundary condition is

$$O(S,T) = \max\{K - S(T), 0\}$$

(European-put-option value at maturity) (4.16)

in terms of the strike price. Here, if the security value $S(T)$ is below the strike price at maturity, the option holder profits by buying some of the security at market value and reselling at the strike price; otherwise the option is worthless.

Now beginning with the Black–Scholes equation (4.14), we can make the integrating-factor transformation

$$O = \tilde{O}\, e^{rt} \tag{4.17}$$

to obtain

$$\frac{\partial \tilde{O}}{\partial t} = -rS\frac{\partial \tilde{O}}{\partial S} - \frac{\sigma^2}{2}S^2\frac{\partial^2 \tilde{O}}{\partial S^2}. \tag{4.18}$$

Because the *final* condition $S(T)$ of this equation is defined, this should be interpreted as in terms of *backwards* evolution. Specifically, the PDE

$$\partial_t f(x,T|S,t) = -rS\partial_S f(x,T|S,t) - \frac{\sigma^2}{2}S^2\partial_S^2 f(x,T|S,t) \tag{4.19}$$

has the form of the *backwards* Kolmogorov equation (3.67), with boundary condition

$$f(x,T|S,T) = \delta(x - S). \tag{4.20}$$

Then the variable $\tilde{O}$ may be propagated backwards in time by applying the conditional density, because $f(x,T|S,t)$ in this case is a backwards Green function:

$$\tilde{O}(S,t) = \int dx\, \tilde{O}(x,T)\, f(x,T|S,t). \tag{4.21}$$

Transforming back to the original option price, this integral solution reads

$$O(S,t) = e^{r(t-T)} \int dx\, O(x,T)\, f(x,T|S,t). \tag{4.22}$$

The equivalent SDE to the backwards Kolmogorov equation is given by Eq. (3.46):

$$dx(t) = rx\, dt + \sigma x\, dW(t). \tag{4.23}$$

This SDE has multiplicative noise, but changing variables to

$$y = \log x \tag{4.24}$$

via

$$dy(t) = y(t+dt) - y(t) = \log[x(t+dt)] - \log[x(t)] = \frac{dx(t)}{x(t)} - \frac{dx^2(t)}{2x^2(t)} \tag{4.25}$$

eliminates the multiplicative nature of the noise:

$$dy(t) = \left(r - \frac{\sigma^2}{2}\right) dt + \sigma\, dW(t). \tag{4.26}$$

This is simple diffusion with diffusion coefficient σ^2 and drift rate $r - \sigma^2/2$; the density of solutions of this SDE thus evolves as a drifting, diffusing Gaussian:

$$f_y(y',t'|y,t) = \frac{1}{\sqrt{2\pi\sigma^2(t'-t)}} \exp\left[-\frac{y' - y - (r - \sigma^2/2)(t'-t)}{2\sigma^2(t'-t)}\right]. \tag{4.27}$$

Putting this solution into Eq. (4.22) yields the integral solution

$$O(S,t) = e^{r(t-T)} \int dy'\, O(e^{y'},T)\, f_y(y',T|\log S,t),$$

(integral solution of Black–Scholes equation) (4.28)

after implementing the proper variable changes. To proceed, we must implement a specific boundary condition. For example, putting in the boundary condition (4.15) for a European call option, we obtain the integral

$$O(S,t) = e^{r(t-T)} \int_{\log K}^{\infty} dy' \left(e^{y'} - K\right) f_y(y',T|\log S,t). \tag{4.29}$$

Because this is a linear combination of two Gaussian integrals, it may be carried out in closed form. The result is

$$\begin{aligned} O(S,t) &= S\Phi(z_1) - Ke^{r(t-T)}\Phi(z_2) \\ z_1 &:= \frac{\log(S/K) + (r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \\ z_2 &:= \frac{\log(S/K) + (r-\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \end{aligned}$$

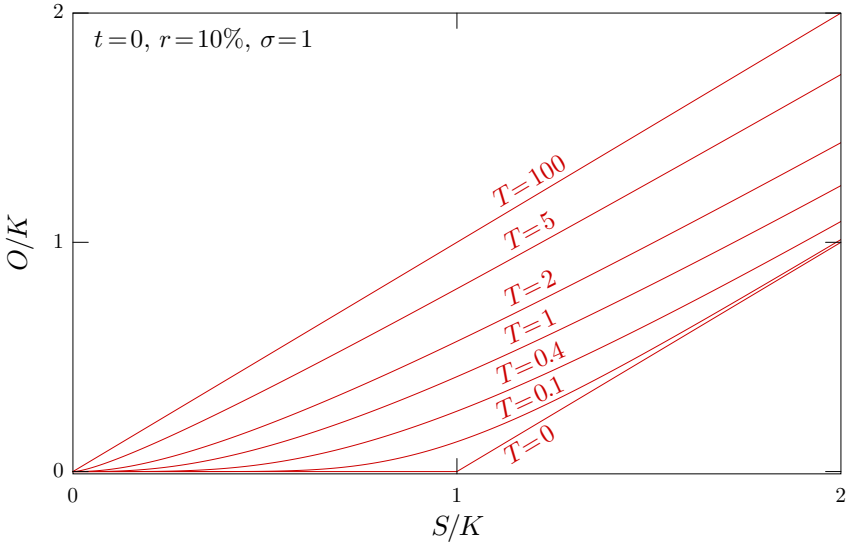
(European-call-option value at maturity) (4.30)

where

$$\Phi(z) := \int_{-\infty}^{z} dx\, \frac{e^{-x^2/2}}{\sqrt{2\pi}} = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right] \tag{4.31}$$

is the standard-normal cumulative distribution. Equation (4.30) is the **Black–Scholes option-pricing formula**.

To illustrate the behavior of the Black–Scholes formula, some European call option prices are plotted below, for an interest rate $r = 10\%$ per unit time, and a volatility $\sigma = 1$; different curves correspond to different times to maturity.



As the option matures ($T = 0$), the option value is just the payoff, as given in Eq. (4.15). For long times to maturity, the option value is the same as the value of the underlying security. The Black–Scholes solution interpolates smoothly between these extremes as time progresses.

4.3.4 Intuitive Derivation

An alternate derivation of the integral solution (4.28) may help to clarify its meaning, by bypassing the Black–Scholes PDE. The payoff of an option is given by the value at maturity, as in Eqs. (4.15). In general,

we can write this as $O(S_T, T)$, where $S_T = S(T)$ emphasizes that this is the underlying security price at maturity. Then the average payoff value is the average of $O(S_T, T)$ over all possible S_T, $\langle\langle O(S_T, T)\rangle\rangle$. This is a risk-neutral valuation of the option; at any prior time $t < T$, the value should then be discounted at the risk neutral rate r. Thus, the option should be valued on average at

$$O(S, t) = e^{-r(T-t)} \left\langle\left\langle O(S_T, T) \right\rangle\right\rangle_{S_T}. \tag{4.32}$$

Since in this equation the present value $S(t)$ is specified, we can write out the expectation value over the final value $S(T)$ by an average over the conditional density:

$$O(S, t) = e^{r(t-T)} \int dS_T \, O(S_T, T) \, f(S_T, T|S, t). \tag{4.33}$$

For example, for a European call option,

$$O(S, t) = e^{r(t-T)} \int_K^\infty dS_T \, (S_T - K) \, f(S_T, T|S, t). \tag{4.34}$$

These two integral solutions are equivalent to Eqs. (4.28) and (4.29), after a change of variables.

4.3.5 Construction of Options for Sale

Going back to the risk-neutral condition (4.13) and rearranging, we obtain

$$O = \Pi - nS. \tag{4.35}$$

Now since Π is a risk-neutral portfolio, suppose that we replace it by an equivalent amount $B(t)$ of bonds, and create a new portfolio Π' out of this combination of bonds and a quantity $-nS$ of stocks (remembering $n < 0$):

$$\Pi' = B - nS. \tag{4.36}$$

Now let's interpret this equation as follows. Constructing a portfolio of value $-nS$ in stocks, and an initial quantity $B = O + nS = O - S(\partial O/\partial S)$ bonds, this will have an equivalent value to a value O in options. Importantly, it will *track the value* of O, at least over short time scales. Over longer time scales, the portfolio Π' will need to be *rebalanced* by changing n to ensure the condition $B = O + nS = O - S(\partial O/\partial S)$ is satisfied.

This represents the procedure that allows banks to sell options without taking on the associated risks. Selling an amount O of options, a bank can create a portfolio Π' of equivalent value, with the appropriate mixture of bonds and the underlying security. With rebalancing, this portfolio tracks the value of the option, and the bank can use this to pay off the option at maturity. Because the initial setup and rebalancing require some effort, the bank sells the option at a small premium compared to the theoretical value O.

4.3.6 Greeks

The Black–Scholes model gives the value of an option as an *expectation*. It doesn't directly give an indication of risk; rather, it is necessary to look at the variation of the expected value with respect to the model parameters. These can of course be done in closed form for the Black–Scholes formula. Because the partial derivatives are conventionally labeled mostly by Greek letters, they are commonly called the **Greeks**.[3] Of course, many different partial derivatives are possible; we will discuss a few of them here.

- **Delta**: the most obvious Greek is the variation of the option value with respect to the underlying security price:

$$\Delta := \frac{\partial O}{\partial S} : \tag{4.37}$$

[3]For a good discussion of the Greeks, as well as some stories behind the publication of the Black–Scholes model, see D. M. Chance, "TN99-02: Derivation and Interpretation of the Black-Scholes Model," (version 15 October 2014), archived at `https://web.archive.org/web/20181007004632/http://www.bus.lsu.edu/academics/finance/faculty/dchance/Instructional/TN99-02.pdf`

Indeed, this is one of the derivatives that appears in the Black–Scholes equation (4.14), as well as in the risk-neutral condition (4.9), where we argued that it should be positive. From the Black–Scholes formula for a European call option, this turns out to have the value (Problem 4.1)

$$\Delta = \Phi(z_1), \tag{4.38}$$

where it has the form of a probability (and is thus positive as expected). This relation also predicts that a small fluctuation in a security price leads to a smaller, positively correlated fluctuation in the corresponding option price.

- **Gamma**: The next derivative with respect to the security price also appears in the Black–Scholes equation (4.14):

$$\Gamma := \frac{\partial^2 O}{\partial S^2} = \frac{\partial \Delta}{\partial S}. \tag{4.39}$$

From the Black–Scholes formula for a European call option, this derivative becomes

$$\Gamma = \frac{\Phi'(z_1)}{S\sigma\sqrt{T-t}}. \tag{4.40}$$

Note that this is also strictly positive, and indicates how Δ varies with S. This is interesting because of the risk-neutral condition (4.9): as the security price changes, the portfolio must be rebalanced to maintain its risk-free status (i.e., n must decrease as S increases), and Γ indicates the magnitude of the rebalance for a small change in S.

- **Theta**: The variation of the option price with time is the last derivative to appear in the Black–Scholes equation (4.14):

$$\Theta := \frac{\partial O}{\partial t}. \tag{4.41}$$

The specific expression for a European call option is

$$\Theta = -\frac{\sigma S}{2\sqrt{T-t}}\Phi'(z_1) - rKe^{r(t-T)}\Phi(z_2). \tag{4.42}$$

Note that this derivative is strictly negative: European options are more value the farther away they are from maturity.

- **Vega**: The sensitivity of the option price with respect to volatility is

$$\mathcal{V} := \frac{\partial O}{\partial \sigma}. \tag{4.43}$$

(Note that ‘‘Vega’’ isn’t a Greek letter, although it sounds like it should be; it is typeset here as a capital calligraphic V, which looks something like an enlarged ν.) For a European call option within the Black–Scholes universe, this has the value

$$\mathcal{V} = S\sqrt{T-t}\,\Phi'(z_1). \tag{4.44}$$

This is an important statistic, as the price of an option tends to depend sensitively on the volatility. (Remember that one purpose of an option is as a tool to *reduce* risk, and therefore the price should be positively and strongly correlated with increasing volatility, and thus risk.)

- **rho**: The last derivative that we will mention is the variation of option price with respect to the riskless interest rate,

$$\rho := \frac{\partial O}{\partial r}, \tag{4.45}$$

which takes the value

$$\rho = K(T-t)\,e^{-r(T-t)}\,\Phi(z_2) \tag{4.46}$$

for a European call option. Note that this is typically small: the cumulative density is of order 1, and the time-dependent factor is small for either small or large times from maturity. This means that option values are not very sensitive to interest rates, which also makes intuitive sense.

Also, what happened to α and β? The nomenclature of the above Greeks continued established notation in quantitative finance.[4] The α parameter, or **Jensen's alpha**, is a measure of the performance of a mutual fund beyond expectations, as set by a standard index (and thus a measure of the performance of the fund's managers). The expected performance is correlated with volatility, as we have seen, and β is the volatility parameter.

[4]Michael C. Jensen, ''The Performance of Mutual Funds in the Period 1945-1964,'' *Journal of Finance* **23**, 389 (1968) (doi: 10.2139/ssrn.244153).

4.4 Exercises

Problem 4.1

Derive the following expressions for the Greeks from the Black–Scholes formula for pricing European call options:

$$\begin{aligned}
\Delta &:= \frac{\partial O}{\partial S} = \Phi(z_1) \\
\Gamma &:= \frac{\partial^2 O}{\partial S^2} = \frac{\Phi'(z_1)}{S\sigma\sqrt{T-t}} \\
\Theta &:= \frac{\partial O}{\partial t} - \frac{\sigma S}{2\sqrt{T-t}}\Phi'(z_1) - rKe^{r(t-T)}\Phi(z_2) \\
\mathcal{V} &:= \frac{\partial O}{\partial \sigma} = S\sqrt{T-t}\,\Phi'(z_1) \\
\rho &:= \frac{\partial O}{\partial r} = K(T-t)\,e^{-r(T-t)}\,\Phi(z_2).
\end{aligned} \tag{4.47}$$

Chapter 5

Fixed-Endpoint Stochastic Processes

5.1 Stochastic Boundary-Value Problems: Brownian Bridges

We have already studied the formalism to handle the simple stochastic differential equation

$$dB = dW, \tag{5.1}$$

for which the solution is $B(t) = W(t)$, up to an arbitrary additive constant. (Recall that this is the same in either Itō or Stratonovich calculus, since the noise is additive.) However, suppose we add the additional constraints $B(0) = B(1) = 0$, and we want the solution $B(t)$ for $t \in [0, 1]$. The $B(0) = 0$ constraint is not a problem, as it defines the initial condition of the problem. But the *final* condition $B(1) = 0$ is considerably more difficult, as $W(t)$ in general tends to wander *away* from zero. However, a (vanishingly) small subset of solutions obey this final condition, so in principle we could simulate many possible realizations of $W(t)$, and discard them until we find one that returns sufficiently close to zero at $t = 1$. This kind of constrained random walk, or ''stochastic loop'', comes up, for example, in quantum field theory.[1] This problem is also a nice example, showing alternate approaches to solving stochastic equations, and providing more insight into regular diffusion $W(t)$.

One simple guess at a solution is simply to *force* a regular Wiener path $W(t)$ back to its initial point by subtracting off its final value $W(t = 1)$, where the subtraction is pro-rated over the path:

$$B(t) := W(t) - tW(1). \tag{5.2}$$
(Brownian bridge)

This is called a **Brownian bridge**, and somewhat surprisingly, this solution satisfies the conditions above for our constrained Wiener path (with some cautions). This is something like viewing a Wiener path $W(t)$ as composed of a linear drift to a final destinations plus fluctuations about zero, and then subtracting off the drift.

To see that this is the case, first we note that $B(t)$ is still a Gaussian random variable, since it is a linear combination of $W(t)$ and $W(1)$, both of which are random variables. Because $\langle W(t)\rangle = 0$, we must also have $\langle y(t)\rangle = 0$. Thus, we must only check the variance of the increments to establish that $B(t)$ is a proper Wiener process. Dividing the unit time interval into N increments, with time steps $\Delta t = 1/N$, with points $B_n := y(n\Delta t)$ and increments $\Delta B_n := B_{n+1} - V_n$, then we will work with

$$\Delta B_n = \Delta W_n - \Delta t\, W(1). \tag{5.3}$$

Using

$$\begin{aligned} \mathrm{Var}[X+Y] &= \left\langle (X+Y)^2 \right\rangle - \langle X+Y\rangle^2 \\ &= \left\langle X^2 \right\rangle + \left\langle X^2 \right\rangle + 2\langle XY\rangle - \langle X\rangle^2 + \langle Y\rangle^2 + 2\langle X\rangle\langle Y\rangle \\ &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y], \end{aligned} \tag{5.4}$$

[1] Holger Gies, Kurt Langfeld, and Laurent Moyaerts, ''Casimir effect on the worldline,'' *Journal of High Energy Physics* **06**, 018 (2003) (doi: 10.1088/1126-6708/2003/06/018).

we can compute the variance as

$$\begin{aligned}\mathrm{Var}[\Delta B_n] &= \mathrm{Var}[\Delta W_n] + (\Delta t)^2\,\mathrm{Var}[W(1)] - 2\Delta t\,\mathrm{Cov}[\Delta W_n, W(1)] \\ &= \Delta t + (\Delta t)^2 - 2(\Delta t)^2 \\ &= \Delta t - (\Delta t)^2,\end{aligned} \tag{5.5}$$

where $\mathrm{Var}[W(1)] = 1$ and $\mathrm{Var}[\Delta W_n] = \Delta t$. This means that $\mathrm{Var}[dB] = dt$, since $dt^2 = 0$, and so we have the statistics of Wiener noise. Notice that the subtraction of the drift *did* skew the statistics, but by a negligible amount in the continuum limit because the fluctuations become comparatively large. Therefore one should be careful with a naïve discretization of Eq. (5.2),

$$B_n = \sum_{j=1}^{n} \Delta W_j - \frac{n}{N}\sum_{j=1}^{N} \Delta W_j = \frac{1}{\sqrt{N}}\sum_{j=1}^{n} z_j - \frac{n}{N^{3/2}}\sum_{j=1}^{N} z_j, \tag{5.6}$$

where z_j are independent, unit-normal random deviates. This ''algorithm'' generates paths with increments that have variance smaller than $1/N$ by a factor $(N-1)/N$, so really the entire path should then be rescaled by $N/(N-1)$. As a coordinate transformation, this discretization is also somewhat pathological in mapping N independent coordinates z_j to $N-1$ independent coordinates y_n (see Problem 5.4). Below we will consider another algorithm that generates a Brownian bridge without these issues.

Note that by the above argument, we can add *any* finite, deterministic function to a Wiener process and still obtain a Wiener process. Thus, for example, we can define

$$B_{a\to b}(t) := a + t(b-a) + W(t) - tW(1) \tag{5.7}$$
(Brownian bridge from a to b)

to be a Brownian bridge[2] that connects a to b over the time interval from 0 to 1. (Other time intervals are possible by shifting and scaling the time variable.) A closely related property is that the Brownian bridge $W(t) - tW(1)$ is independent of $W(1)$, as we can see by computing the correlation function

$$\Big\langle\!\Big\langle \big[W(t) - tW(1)\big]\, W(1) \Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle W(t)W(1) - tW^2(1) \Big\rangle\!\Big\rangle = t - t = 0. \tag{5.8}$$

This is the continuous version of the coordinate-transform pathology in the discrete case that we noted above. This independence—along with our ability to stretch Wiener paths to create other Wiener paths—has an important meaning. A particular Wiener path $W(t)$ that wanders to $W(t=1)$ is possible with a Gaussian probability density in $W(1)$, so that in particular, large excursions are unlikely. However, once a *particular* value of $W(1)$ is stipulated, the possible paths are essentially Brownian bridges $B(t)$ that return to the initial point at $t = 1$, but with a uniform-velocity drift to $W(1)$. Even if a peculiarly large value of the overall drift $W(1)$ is stipulated, by far the most likely way to accomodate this is to distribute the drift uniformly over the whole time interval.

Finally, we can examine the fluctuations of the (loop-style) Brownian bridge,

$$\begin{aligned}\mathrm{Var}[B(t)] &= \mathrm{Var}[W(t) - tW(1)] \\ &= \mathrm{Var}[W(t)] + t^2\mathrm{Var}[W(1)] - 2t\mathrm{Cov}[W(t), W(1)] \\ &= t + t^2 - 2t^2,\end{aligned} \tag{5.9}$$

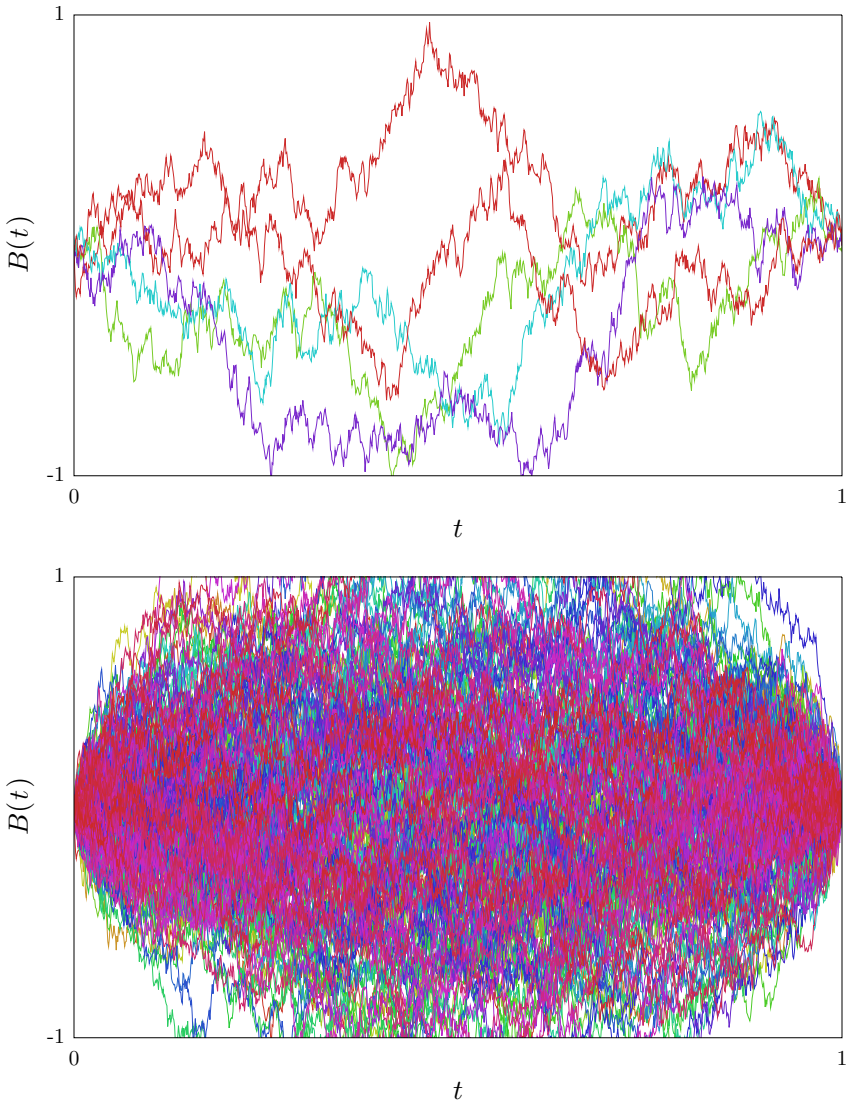
[2]While the name ''Brownian bridge'' may more sensibly apply to $B_{a\to b}(t)$, which ''bridges'' a to b, $B(t) = B_{0\to 0}(t)$ is a special case that bridges the origin to itself, and goes the name of a **standard Brownian bridge**. It is useful to think of $B(t)$ as a bridge in the sense of having a ''pinned'' solution at $t = 1$. Incidentally, other variations on conditioned Wiener paths are commonly defined, such as the **Brownian meander** $W^+(t)$ (a Wiener path, or Brownian motion $W(t)$, conditioned to be everywhere positive), and the **Brownian excursion**, which is a Brownian meander conditioned on $W^+(1) = 0$, or alternately a standard Brownian bridge $B(t)$ conditioned to be everywhere positive. See., e.g., Jim Pitman, ''Brownian Motion, Bridge, Excursion, and Meander Characterized by Sampling at Independent Uniform Times,'' *Electronic Journal of Probability* **4**, 11 (1999) (doi: 10.1214/EJP.v4-48).

so that we find

$$\mathrm{Var}[B(t)] = t(1-t). \tag{5.10}$$
(variance of Brownian bridge)

Thus, the bridge fluctuates most when it is farthest away from either fixed endpoint, which is sensible.

Again, to get a better idea of what these look like, 5 and 200 Brownian bridges are respectively shown in the two plots below. (Again, these are actually finite realizations of $B(t)$, with $\Delta t = 0.001$.)



Both the symmetry about $t = 0.5$ and the character of diffusing away and returning are readily apparent here.

5.1.1 Finite Bridge Generation: Homogeneous Case

Here we will consider generating a finite numerical approximation[3] to a closed Wiener path in a more direct way than before. If we again divide the path into N increments, then we have time steps

$$\Delta t = \frac{1}{N}, \tag{5.11}$$

[3] Gies *et al.*, *op. cit.*

and the points $B_n := B(n\Delta t)$, for consistency with Eq. (5.1), must be such that $B_{n+1} - B_n$ is a normally distributed random variable of zero mean and variance $\Delta t = 1/N$. Thus we have the multidimensional probability density

$$P(B_1, \ldots, B_{N-1}) \propto \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N}(B_j - B_{j-1})^2\right)\right], \tag{5.12}$$

where by construction $B_0 = 0$ and $B_N = 0$ are not dependent variables. We will proceed by a coordinate transformation, changing variables to obtain an standard normal Gaussian distribution in every dimension.

First, consider the sum in the exponent, which we may write as

$$\begin{aligned}\sum_{j=1}^{N}(B_j - B_{j-1})^2 &= \sum_{j=2}^{N-1}(B_j - B_{j-1})^2 + B_1^2 + B_{N-1}^2 \\ &= 2B_1^2 - 2B_1B_2 + 2B_2^2 - 2B_2B_3 + 2B_3^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2.\end{aligned} \tag{5.13}$$

Now separating out the B_1 dependence of the exponent,

$$\begin{aligned}\sum_{j=1}^{N}(B_j - B_{j-1})^2 &= 2\left(B_1 - \frac{1}{2}B_2\right)^2 + \frac{3}{2}B_2^2 - 2B_2B_3 + 2B_3^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2 \\ &= B_1'^2 + \frac{3}{2}B_2^2 - 2B_2B_3 + 2B_3^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2.\end{aligned} \tag{5.14}$$

where we completed the square of B_1, and we defined the transformed coordinate

$$B_1' := \sqrt{2}\left(B_1 - \frac{1}{2}B_2\right), \tag{5.15}$$

which encompasses all the dependence on B_1, and enters in the exponent to give a normally distributed random variable with zero mean and variance $1/N$. We now continue to complete squares and factor out the dependence on B_2, B_3, and so on. At the nth step, we have

$$\begin{aligned}\sum_{j=1}^{N}(B_j - B_{j-1})^2 &= B_1'^2 + \cdots + B_{n-1}'^2 + c_nB_n^2 - 2B_nB_{n+1} + B_{n+1}^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2 \\ &= B_1'^2 + \cdots + B_{n-1}'^2 \\ &\qquad + c_n\left(B_n - \frac{1}{c_n}B_{n+1}\right)^2 + \left(2 - \frac{1}{c_n}\right)B_{n+1}^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2 \\ &= B_1'^2 + \cdots + B_n'^2 + c_{n+1}B_{n+1}^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2\end{aligned} \tag{5.16}$$

where in the basis step above we began with $c_1 = 2$, and we have defined

$$c_{n+1} := \left(2 - \frac{1}{c_n}\right) \tag{5.17}$$

and

$$B_{n+1}' := \sqrt{c_n}\left(B_n - \frac{1}{c_n}B_{n+1}\right). \tag{5.18}$$

This has the same form as the $(n+1)$th step, which inductively completes all the squares. With these variables, the probability distribution is

$$P(B_1', \ldots, B_{N-1}') \propto \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N-1}B_j'^2\right)\right], \tag{5.19}$$

so that again the B'_j are independent, Gaussian random numbers of zero mean and variance $1/N$. These can be chosen independently, and Eq. (5.18) can be solved to give B_n in terms of B'_n and B_{n+1}:

$$B_n = \frac{B'_n}{\sqrt{c_n}} + \frac{B_{n+1}}{c_n}. \tag{5.20}$$

Thus, the bridge coordinates should be generated in a backwards recurrence, given the forward recurrence for the coefficients c_n. This is even more conveniently given in terms of standard-normal deviates z_j, which can replace the B'_j:

$$B_n = \frac{z_n}{\sqrt{Nc_n}} + \frac{B_{n+1}}{c_n}. \tag{5.21}$$

Note also that the recurrence (5.17) has the solution

$$c_n := \frac{n+1}{n}. \tag{5.22}$$

These relations give the bridge directly in terms of easily generated deviates z_n.

Note that in principle we must also involve the Jacobian determinant of the coordinate transformation from B_j to B'_j. Effectively, we have taken the exponent (5.13), which we can write as a quadratic form as

$$\sum_{j=1}^{N} (B_j - B_{j-1})^2 = B_a A_{ab} B_b, \tag{5.23}$$

where (A_{ab}) is a square, tridiagonal matrix of dimension $N-1$ with a twos along the diagonal and ones for every other nonzero element. Since the matrix is symmetric, it is diagonalized by an orthogonal transformation (P_{ab}), so that $B_a A_{ab} B_b = B'_a D_{ab} B'_b$, where (D_{ab}) is diagonal (and in fact has only one nonzero eigenvalue, whose value is $N-1$), and $B'_a := P_{ab} B_b$. We have effectively performed this diagonalization in constructing the above recurrence relations. Notice then that the Jacobian determinant is just the determinant of (P_{ab}), which is just a constant factor that only affects the normalization of the probability density. Thus, we have justified our coordinate transformation to a new Gaussian distribution.

To summarize, the algorithm to generate a Brownian bridge of N steps (i.e., $N+1$ points $B_0, \ldots B_N$, where $B_0 = B_N = 0$) is:

1. Generate standard normal random numbers (zero mean, unit variance) z_n from $n = 1, \ldots, N-1$.
2. Generate the positions B_n for $n = N, \ldots, 1$, according to the *backwards* recurrence

$$\begin{aligned} B_N &= 0 \\ B_n &= z_n \sqrt{\frac{n}{N(n+1)}} + \left(\frac{n}{n+1}\right) B_{n+1}, \qquad n = N-1, \ldots, 1 \\ B_0 &= 0. \end{aligned} \tag{5.24}$$

We can similarly write a *forward* recurrence

$$\begin{aligned} B_0 &= 0 \\ B_n &= z_n \sqrt{\frac{c_n}{N}} + c_n B_{n-1}, \qquad n = 1, \ldots, N-1 \\ B_N &= 0, \end{aligned} \tag{5.25}$$

where to simplify the notation, we have defined the recurrence coefficient

$$c_n := \frac{N-n}{N-n+1}. \tag{5.26}$$

This forward scheme gives finitely sampled Brownian bridges with the same statistics.

This algorithm gives a simulated (finite) realization of a closed, stochastic Wiener path in terms of easily generated standard-normal random deviates. Note that this algorithm is easily generalized to generate samples of a Brownian bridge $B_{a\to b}(t)$ running from a to b in unit time: the B_n in Eqs. (5.25) should be thought of as $B_n - b$ (i.e., measured in terms of their distance to the "target" b), so that

$$\begin{aligned} B_0 &= a \\ B_n &= z_n\sqrt{\frac{c_n}{N}} + c_n(B_{n-1} - b) + b, \qquad n = 1, \ldots, N-1 \\ B_N &= b, \end{aligned} \tag{5.27}$$

will act as a recurrence for this "open" bridge, with the c_n as before. In the case of a more general Brownian bridge $B_{t(a\to b)}(t')$ running from a to b in time t, this recurrence is further generalized by changing the first term on the right-hand side of the second equation of Eqs. (5.27) from $z_n\sqrt{c_n/N}$ to $z_n\sqrt{c_n\Delta t}$, where $\Delta t = t/N$ (the c_n coefficients again stay the same).

5.1.2 Finite Bridge Generation: Inhomogeneous Case

A slightly more complicated variation on the above recurrences arises when we allow for time-dependent drift and diffusion rates, according to

$$dy(t) = \alpha(t)\,dt + \sigma(t)\,dW(t), \tag{5.28}$$

and again impose the boundary condition $y(1) = y(0)$. We will treat this SDE as an Itō SDE for the purposes of finite differences, but strictly speaking this shouldn't matter since the noise is still additive. In finite form, we have the multivariate probability density

$$P(y_1, \ldots, y_{N-1}) \propto \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N}\frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2}\right)\right], \tag{5.29}$$

where again by construction $y_0 = 0$ and $y_N = 0$ are not dependent variables. We now thus have the exponent sum

$$\begin{aligned} \sum_{j=1}^{N}\frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2} &= \frac{(y_1 - \alpha_0/N)^2}{\sigma_0^2} + \frac{(y_2 - y_1 - \alpha_1/N)^2}{\sigma_1^2} + \sum_{j=3}^{N}\frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2} \\ &= \frac{\bar{y}_1^2}{\sigma_0^2} + \frac{(y_2 - \bar{y}_1 - \alpha_1/N - \alpha_0/N)^2}{\sigma_1^2} + \sum_{j=3}^{N}\frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2}, \end{aligned} \tag{5.30}$$

where we have defined

$$\bar{y}_1 := y_1 - \frac{\alpha_0}{N} \tag{5.31}$$

in order to begin eliminating the mean drifts. Continuing in this process, we have

$$\sum_{j=1}^{N}\frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2} = \frac{\bar{y}_1}{\sigma_0^2} + \frac{(\bar{y}_2 - \bar{y}_1)^2}{\sigma_1^2} + \cdots + \frac{(\bar{y}_N - \bar{y}_{N-1})^2}{\sigma_{N-1}^2}, \tag{5.32}$$

where

$$\bar{y}_n := y_n - \frac{1}{N}\sum_{j=0}^{n-1}\alpha_j, \qquad n \in 1, \ldots, N, \tag{5.33}$$

again remembering $y_N = 0$. Completing the first square as before,

$$\begin{aligned}\sum_{j=1}^{N}\frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2} &= \left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma_1^2}\right)\bar{y}_1^2 - \frac{2\bar{y}_1\bar{y}_2}{\sigma_1^2} + \frac{\bar{y}_2^2}{\sigma_1^2} + \frac{(\bar{y}_3-\bar{y}_2)^2}{\sigma_2^2} + \cdots + \frac{(\bar{y}_N - \bar{y}_{N-1})^2}{\sigma_{N-1}^2} \\ &= \left(\frac{\sigma_0^2+\sigma_1^2}{\sigma_0^2\sigma_1^2}\right)\left[\bar{y}_1 - \left(\frac{\sigma_0^2}{\sigma_0^2+\sigma_1^2}\right)\bar{y}_2\right]^2 + \left[\frac{1}{\sigma_1^2} - \frac{\sigma_0^2}{\sigma_1^2(\sigma_0^2+\sigma_1^2)}\right]\bar{y}_2^2 \\ &\qquad + \frac{(\bar{y}_3-\bar{y}_2)^2}{\sigma_2^2} + \cdots + \frac{(\bar{y}_N - \bar{y}_{N-1})^2}{\sigma_{N-1}^2} \\ &= y_1'^2 + \left[\frac{1}{\sigma_1^2} - \frac{\sigma_0^2}{\sigma_1^2(\sigma_0^2+\sigma_1^2)}\right]\bar{y}_2^2 + \frac{(\bar{y}_3-\bar{y}_2)^2}{\sigma_2^2} + \cdots + \frac{(\bar{y}_N - \bar{y}_{N-1})^2}{\sigma_{N-1}^2},\end{aligned} \tag{5.34}$$

where we have defined

$$y_1' := \sqrt{\left(\frac{\sigma_0^2+\sigma_1^2}{\sigma_0^2\sigma_1^2}\right)}\left[\bar{y}_1 - \left(\frac{\sigma_0^2}{\sigma_0^2+\sigma_1^2}\right)\bar{y}_2\right]. \tag{5.35}$$

At the nth stage of completing the square, we must handle terms of the form

$$\begin{aligned}c_n\bar{y}_n^2 - \frac{2\bar{y}_n\bar{y}_{n+1}}{\sigma_n^2} + \left(\frac{1}{\sigma_n^2}+\frac{1}{\sigma_{n+1}^2}\right)\bar{y}_{n+1}^2 &= c_n\left(\bar{y}_n - \frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}\right)^2 + \left(\frac{1}{\sigma_n^2}+\frac{1}{\sigma_{n+1}^2}-\frac{1}{c_n\sigma_n^4}\right)\bar{y}_{n+1}^2 \\ &= y_n'^2 + c_{n+1}\bar{y}_{n+1}^2,\end{aligned} \tag{5.36}$$

where we have defined the decoupled square

$$y_n' := \sqrt{c_n}\left(\bar{y}_n - \frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}\right) \tag{5.37}$$

and the recursion

$$c_{n+1} = \frac{1}{\sigma_n^2} + \frac{1}{\sigma_{n+1}^2} - \frac{1}{c_n\sigma_n^4}, \tag{5.38}$$

thus inductively completing all the squares. Again, the y_n' are Gaussian numbers, such that we may solve to find the shifted positions

$$\bar{y}_n = \frac{y_n'}{\sqrt{c_n}} + \frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}, \tag{5.39}$$

or in terms of standard-normal deviates,

$$\bar{y}_n = \frac{z_n}{\sqrt{Nc_n}} + \frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}. \tag{5.40}$$

Then solving Eq. (5.33),

$$y_n = \bar{y}_n + \frac{1}{N}\sum_{j=0}^{n-1}\alpha_j \tag{5.41}$$

we find the actual bridge positions.

To summarize, the algorithm is, to generate a stochastic, constrained path of N points $y_1, \ldots y_N$, where $y_0 = y_N = 0$:

1. Begin with means α_n and standard-deviations σ_n for $n \in 0, \ldots, N = 1$.
2. Generate the coefficients c_n for $n = 1, \ldots, N-1$, according to the recurrence

$$c_1 = \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2}, \qquad c_{n+1} = \frac{1}{\sigma_n^2} + \frac{1}{\sigma_{n+1}^2} - \frac{1}{c_n\sigma_n^4}. \tag{5.42}$$

If many paths are to be generated, these coefficients only need to be generated once.

3. Generate standard normal random numbers (zero mean, unit variance) z_n from $n = 1, \ldots, N-1$.

4. Generate the shifted positions $\bar{y}_n$ for $n = N, \ldots, 1$, according to the *backwards* recurrence

$$\begin{aligned} \bar{y}_N &= -\sum_{j=0}^{N-1} \alpha_j \\ \bar{y}_n &= \frac{z_n}{\sqrt{N c_n}} + \frac{\bar{y}_{n+1}}{c_n \sigma_n^2}. \end{aligned} \tag{5.43}$$

5. Generate the path positions y_n for $n = 1, \ldots, N$, using

$$y_n = \bar{y}_n + \frac{1}{N} \sum_{j=0}^{n-1} \alpha_j. \tag{5.44}$$

This gives a simulated realization of a closed, stochastic path with nonuniform drift and diffusion. Note that a path with *constant* $\alpha(t) \neq 0$ and $\sigma(t) = 1$ is equivalent to the homogeneous path generated by Eqs. (5.25) in the previous section, owing to the conditioning.

5.1.3 SDE and Integral Representations of the Brownian Bridge

The definition (5.2) of the Brownian bridge involves the pro-rated subtraction of the global drift of a Wiener path. However, given the completed-square construction of Eq. (5.1.1), we can also derive a *local* representation of the Brownian bridge as the solution of an SDE. Recall the *backward* recurrence, Eq. (5.24), which we may solve for B_{n+1}:

$$B_{n+1} = \left(\frac{n+1}{n} \right) B_n - z_n \sqrt{\frac{n+1}{nN}}. \tag{5.45}$$

We can further rewrite this as

$$B_{n+1} - B_n = \left(\frac{1}{n} \right) B_n - \Delta W_n \sqrt{\frac{n+1}{n}}, \tag{5.46}$$

where $\Delta W_n = z_n / \sqrt{N}$, and then

$$B_{n+1} - B_n = \left(\frac{N}{n} \right) B_n \Delta t - \Delta W_n \sqrt{\frac{n+1}{n}}, \tag{5.47}$$

where $\Delta t = 1/N$. Passing over to the continuous-time limit, we let $\Delta t \longrightarrow dt$, $n/N \longrightarrow t$, $\Delta W_n \longrightarrow dW(t)$, and we note that $(n+1)/n$ is only different from unity in a vanishingly small interval of small t, a correction we will ignore in the continuum limit:

$$dB = \frac{B}{t}\, dt - dW. \tag{5.48}$$

Note that due to the backward recurrence, this SDE has the "initial" condition $B(1) = 0$, and should be integrated *backwards* in time. We can fix this by letting $t \longrightarrow 1 - t$, so that $dt \longrightarrow -dt$ and $dW \longrightarrow -dW$, and thus

$$dB = -\left(\frac{B}{1-t} \right) dt + dW. \tag{5.49}$$
(SDE form of Brownian bridge)

Thus, we have a representation of a Brownian bridge as an SDE solution. Here, the SDE is similar to an Ornstein–Uhlenbeck process, where the damping rate increases with time, diverging at $t = 1$ to ensure the return of the bridge.

The solution to Eq. (5.49) is (see Problem 5.1)

$$B(t) = (1-t) \int_0^t \frac{dW(t')}{1-t'}, \tag{5.50}$$
(integral form of Brownian bridge)

which can be verified by differentiating this expression. According to this definition, the Brownian bridge is clearly a Gaussian process, and the correlation function computed from this definition

$$\langle\langle B(t)\, B(t')\rangle\rangle = \min(t,t') - tt'$$

(correlation function of Brownian bridge) (5.51)

matches the same correlation function as computed from the first definition (5.2) (see Problem 5.2). This is sufficient to characterize the Gaussian process, and thus the two definitions are equivalent (at least in the sense of ensemble averages, not in the sense of individual paths).

5.1.4 State-Dependent Diffusion in Brownian Bridges

Suppose we consider an inhomogeneous bridge problem that is slightly different from the inhomogeneous equation (5.28):

$$dy(t) = \alpha(y,t)\, dt + \sigma(y,t)\, dW(t).$$

(state-dependent boundary-value SDE) (5.52)

This is more complicated than our previous problem, since the drift and diffusion coefficients depend on the state of the system. While this equation is straightforward to integrate numerically, it is not at all straightforward to do this integration subject to a bridge condition. This problem could be handled iteratively. For example, start by generating a solution with zero drift and constant diffusion, use this fiducial trajectory to generate the drift and diffusion coefficients, which then generates a new path; continue this process until convergence is reached. That is, *if* convergence is reached. In principle, this is a many-dimensional root-finding procedure, but a more stable method such as Newton iteration can be numerically cumbersome.

We will, however, treat in more detail the somewhat simpler problem of state-dependent Stratonovich diffusion,

$$dy(t) = \sigma(y) \circ dW(t),$$

(bridge with state-dependent diffusion) (5.53)

subject to the closure condition $y(T) = y(0) = 0$. We will show that the solution to this SDE is equivalent to the initial-value SDE

$$dy(t) = \sigma(y) \circ dB(t),$$

(solution for state-dependent diffusion) (5.54)

where $B(t)$ is a standard Brownian bridge, *without* imposing the boundary condition at $t = T$. Since $dB(t)$ has the same *local* statistics (i.e., statistics of increments) as the Wiener process $dW(t)$, we need only verify the closure of the path.

Now we turn to the closure of the solution to the SDE (5.54). First, we rewrite this SDE as

$$\frac{1}{\sigma(y)} \circ dy(t) = dB(t), \tag{5.55}$$

still emphasizing the Stratonovich nature of this SDE. Now for some function $S(y)$, where $y(t)$ is defined by the SDE (5.54), the Stratonovich chain rule (3.125) gives

$$dS(y) = S'(y) \circ dy(t), \tag{5.56}$$

where we treat $dB(t)$ equivalently to $dW(t)$ as far as the chain rule is concerned [$dB(t)$ being a particular realization of $dW(t)$]. Suppose we take

$$S'(y) = \frac{1}{\sigma(y)}, \tag{5.57}$$

or that is, $S(y)$ is the antiderivative of $1/\sigma(y)$. Then Eq. (5.55) becomes

$$S'(y) \circ dy(t) = dB(t). \tag{5.58}$$

Integrating both sides from $t = 0$ to 1 gives

$$S[y(1)] - S[y(0)] = B(1) - B(0) = 0, \tag{5.59}$$

where we have again used the Stratonovich chain rule, so that the left-hand side was just the integral of $dS[y(t)]$. Thus,

$$S[y(1)] = S[y(0)], \tag{5.60}$$

which implies

$$y(1) = y(0) = 0 \tag{5.61}$$

as desired, so long as $S(y)$ is an invertible function. Since it is the antiderivative of $1/\sigma(y)$, this is guaranteed if $\sigma(y)$ is everywhere positive and finite, which is a reasonable restriction on the form of the SDE (5.53). However, the Stratonovich nature of this SDE is crucial: an Itō equation naïvely of the same form has a drift term when converted to a Stratonovich equation, and our solution here does not cover that case.

In fact, the integration procedure applied to arbitrary time t instead of $t = 1$ gives

$$S[y(t)] = B(t), \tag{5.62}$$

where we have chosen $S[y(0)] = 0$. Explicitly, this means that we have chosen

$$S(y) = \int_{y(0)}^{y} \frac{dy'}{\sigma(y')}. \tag{5.63}$$

Then the explicit solution in terms of a Brownian bridge is

$$y(t) = S^{-1}[B(t)], \tag{5.64}$$

(explicit solution for state-dependent diffusion bridge)

where as we have already stated, under reasonable assumptions S^{-1} exists. (For an example solution where $\sigma(y)$ is a step function, see Problem 7.3.)Note that the solutions here can be generalized to the case of a return at time $t = T$ instead of $t = 1$ by the replacements

$$B(t) \longrightarrow \sqrt{T}\, B(t/T), \qquad dB(t) \longrightarrow \sqrt{T}\, dB(t/T) \tag{5.65}$$

in the above solutions.

5.1.4.1 Drift

Returning to the more general SDE (5.52), we see that the presence of a drift term $\alpha(y, t)\, dt$ is more difficult because even when integrating with respect to a bridge, closure of the solution is no longer guaranteed: the drift will in general make the solution ''miss'' the initial point, because in the formal bridge solution (interpreting the SDE as a Stratonovich equation),

$$y(t) = \int_0^t \alpha(y, t')\, dt' + \int_0^t \sigma(y, t') \circ dB(t'), \tag{5.66}$$

the first term is not guaranteed to vanish, and in fact the *second* term is *also* no longer guaranteed to vanish due to the influence of the first term on $y(t)$. Additionally, any explicit time dependence in $\sigma(y, t)$ in general causes the solution (5.54) in the drift-free case to fail. As we already mentioned, an iterative procedure to generate a closed solution may work for this, but in practice iteration tends not to converge well if the SDE coefficients change rapidly with y.

An alternate approach is analogous to ''shooting'' methods for ODEs with boundary conditions. The idea here is to note that while the solution (5.66) may not close when integrated with respect to a a bridge $B(t)$, it will close for *some* bridge $B_{0\to c}(t)$ connecting 0 to c, as defined by Eq. (5.7). The extra drift compensates for any drift introduced by the SDE coefficients, and the existence of a value c that closes the path is guaranteed provided that the SDE is ''reasonable enough'' to guarantee continuous solutions (and solutions that vary continuously as the input noise varies). In general, this closing value c must be found numerically, via a root-finding algorithm.

A concern with this method is the *measure* with which we generate the solutions. When we perform the analogous procedure with *additive* noise by closing a Wiener path to form a bridge as in Eq. (5.2),

there is no problem: each Wiener path is associated with a unique bridge (with many paths associated with a single bridge), and choosing Wiener paths with the usual measure results in bridges being chosen with the correct measure (all Brownian bridges being equally likely). However, the nonlinear transformation in solving Eq. (5.52), as well as the *unique* association of a bridge $B_{0\to c}(t)$ with each solution $y(t)$, where c is different for each solution, makes things more complicated. In particular, when generating Wiener paths, the relative probability of generating a bridge from 0 to c is the usual probability density for a Wiener path to end up at c after time $t = 1$:

$$P[B_{0\to c}(t)] = \frac{1}{\sqrt{2\pi}}\, e^{-c^2/2}. \tag{5.67}$$

Therefore, if we require a bridge $B_{0\to c_\alpha}(t)$ to generate a particular closed solution $y_\alpha(t)$, the relative probability for this trajectory to occur is given by

$$w_\alpha = \frac{1}{\sqrt{2\pi}}\, e^{-c_\alpha^2/2}. \tag{5.68}$$

Thus, to generate an ensemble of solutions of Eq. (5.52), each generated solution should only be kept with a relative probability w_α (e.g., by the rejection method). Alternately, when computing an ensemble average with respect to the solutions $y_\alpha(t)$, the average should be computed as a *weighted* average, where the weight of each member of the ensemble is w_α. This procedure is valid for Itō or Stratonovich equations, provided that appropriate integration methods are used in each case. Finally, note that this measure business is not a concern for the solutions (5.54), since $c = 0$ for every solution, and thus they all occur with the same probability.

One final possibility for an algorithm in the case where the SDE coefficients are time-independent [i.e., $\alpha(y,t) = \alpha(y)$ and $\sigma(y,t) = \sigma(y)$], and a solution bridging $y(0) = a$ to $y(T) = b$ is as follows.[4] Generate numerical solutions $y_a(t)$ and $y_b(t)$, such that $y_a(0) = a$ and $y_b(0) = b$. If there is a crossing of the two paths at an appropriate time, i.e., $y_a(\tau) = y_b(T-\tau)$ for some t, to within the resolution of the time discretization, then the two paths can be spliced together to realize a bridge (i.e., the solution is $y_a(t)$ for $t < \tau$, and $y_b(t)$ thereafter). If there is no such crossing, the paths are rejected and the process is repeated until successful.

5.1.4.2 Lamperti Transform

The above considerations of rescaling stochastic processes can be elegantly viewed in the framework of the **Lamperti transform**,[5] which is a transformation that equalizes the variances within a stochastic process. This idea applies to SDEs as follows. Consider the Itō SDE

$$dy = \alpha(y)\, dt + \beta(y)\, dW, \tag{5.69}$$

and the transformation

$$z = S(y). \tag{5.70}$$

Then the Itō chain rule (3.118) gives

$$dz = \left[S'(y)\,\alpha + \frac{1}{2}S''(y)\,\beta^2\right] dt + S'(y)\,\beta\, dW, \tag{5.71}$$

and the Lamperti transformation obtains by choosing

$$S'(y) = \frac{1}{\beta(y)}, \tag{5.72}$$

or

$$z = S(y) = \int_{y_0}^{y} \frac{dy'}{\beta(y')}. \tag{5.73}$$

(Lamperti transform)

[4] Stefano M. Iacus, *Simulation and Inference for Stochastic Differential Equations* (Springer, 2008), Section 2.13.

[5] after J. W. Lamperti, "Semi-stable stochastic processes," *Transactions of the American Mathematical Society* **104**, 62 (1962) (doi: 10.1090/S0002-9947-1962-0138128-7). See Stefano M. Iacus, *op. cit.*, Section 1.11.4.

Also using $S''(y) = -\beta'/\beta^2$, Eq. (5.71) then becomes

$$dz = \left(\frac{\alpha}{\beta} - \frac{\beta'}{2}\right) dt + dW, \tag{5.74}$$
(Lamperti-transformed process)

which now is driven by additive noise. The complexity of the multiplicative noise in the dy SDE is thus moved from the stochastic to the deterministic term, and reconstructing the dynamics in z requires inverting $z = S(y)$ in Eq. (5.73), assuming that it is indeed invertible (which happens, for example, if β has the same sign over the domain of y).

Recall from the Itō–Stratonovich conversion (3.109) that a Stratonovich diffusion equation

$$dy = \beta(y) \circ dW \tag{5.75}$$

is equivalent to the Itō equation (5.69) provided $\alpha = \beta'\beta/2$. In Eq. (5.74) this leads to a vanishing drift term, and $dW = dz$, or $z(t) = W(t)$, so that this SDE can be solved in terms of the inversion of the function S—something not in general possible if the drift term remains in Eq. (5.69).

5.1.4.3 Temporal Rescaling

A similar transformation, closer to the original transformation introduced by Lamperti to study self-similar processes[6] is a trajectory-dependent temporal rescaling, with a transformation from time t to t', such that

$$dt = dt'\, \beta^2[y(t)]. \tag{5.76}$$
(temporal rescaling)

In the Itō SDE (5.69), this transformation leads to the equivalent SDE

$$dy(t') = \frac{\alpha(y)}{\beta^2(y)}\, dt' + dW(t'). \tag{5.77}$$
(rescaled Itō SDE)

Note that, as in Eq. (5.74), the variable diffusion rate disappears, and we are left with additive noise. However, the drift term is different, and in particular there is not a second-order Itō correction term.

On the other hand, the Stratonovich SDE (5.75) transforms simply to

$$dy(t') = dW(t'), \tag{5.78}$$
(rescaled Stratonovich SDE)

with no drift, just as in the case of the Lamperti transform. As in our discussion in Section 5.1.4, the interpretation here is simple if we consider Brownian bridges. In the Itō case, if the noise $dW(t)$ is a standard Brownian bridge, then the closure of the bridge is preserved under the temporal rescaling if and only if $\alpha = 0$. On the other hand, the Stratonovich SDE *always* preserves the closure of the path for any temporal rescaling. Note, however, that in the Itō case, even an Itō equation that is equivalent to the Stratonovich equation ($\alpha = \beta'\beta/2$) does not in general preserve path closure under temporal rescaling.

[6] J. W. Lamperti, *op. cit.*; see also Krzysztof Burnecki, Makoto Maejima, and Aleksander Weron "The Lamperti transformation for self-similar processes," *Yokohama Mathematical Journal* **44**, 25 (1997).

5.2 Exercises

Problem 5.1

Show that the integral expression (5.50)

$$B(t) = (1-t)\int_0^t \frac{dW(t')}{1-t'} \tag{5.79}$$

solves the SDE (5.49)

$$dB = -\left(\frac{B}{1-t}\right)dt + dW. \tag{5.80}$$

Problem 5.2

Derive the correlation function

$$\langle\langle B(t)\,B(t')\rangle\rangle = \min(t,t') - tt' \tag{5.81}$$

(a) using the definition (5.2)

$$B(t) := W(t) - tW(1). \tag{5.82}$$

(b) using the definition (5.50)

$$B(t) := (1-t)\int_0^t \frac{dW(t')}{1-t'}. \tag{5.83}$$

Problem 5.3

Show that if $B(t)$ is a standard Brownian bridge, the covariance of two Brownian-bridge increments $\mathrm{Cov}[dB(t), dB(t')] = 0$ provided $t \neq t'$.

Problem 5.4

Let $y(t)$ denote a Brownian bridge (i.e., $y(0) = y(1) = 0$), and let $y_0, y_1, \ldots, y_N$ denote samples of $y(t)$, taken at times $t_j = j\Delta t = j/N$. (Note that $y_0 = y_N = 0$.) Consider the Gaussian path integral

$$I := \int dy_1 \ldots dy_{N-1}\, \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N}(y_j - y_{j-1})^2\right)\right]. \tag{5.84}$$

(a) Evaluate this integral by using the recurrence (5.20)

$$y_n = \frac{y'_n}{\sqrt{c_n}} + \frac{y_{n+1}}{c_n}. \tag{5.85}$$

to decouple the integral into a product of Gaussian integrals.

(b) Use the $y(t) = W(t) - tW(1)$ construction of the Brownian bridge, along with the independence of $y(t)$ and $W(1)$, to evaluate this integral, to show the consistency of these approaches.

Problem 5.5

For the standard Brownian bridge $B(t)$:

(a) Use the statistics derived in Section 5.1 derived for the Brownian bridge to write down the probability density for $B(t)$ at any time $0 \le t \le 1$.

(b) Compute the *time-averaged* density for $B(t)$, where the average is taken over $0 \le t \le 1$.

Problem 5.6

(a) Using the recurrence for the finite Brownian bridge [Eq. (5.25)],

$$\begin{aligned} B_0 &= 0 \\ B_n &= z_n \sqrt{\frac{N-n}{N(N-n+1)}} + \left(\frac{N-n}{N-n+1}\right) B_{n-1}, \qquad n = 1, \ldots, N-1 \\ B_N &= 0, \end{aligned} \tag{5.86}$$

derive an explicit formula for the Brownian-bridge samples B_n in terms of (only) the standard-normal deviates z_n.

(b) Show that

$$\langle\langle B_n \, B_m \rangle\rangle = \frac{[N - \max(m,n)] \min(m,n)}{N^2}, \tag{5.87}$$

and show that this is consistent with what you know for the continuous-time limit.

(c) Show that

$$\langle\langle \Delta B_n \, \Delta B_m \rangle\rangle = \delta_{nm} \frac{(N-n-1)}{N(N-n)} + \frac{\min(n,m)}{N^2[N - \min(n,m)]} + \frac{(\delta_{nm} - 1)}{\sqrt{N}(N-n)(N-m)}. \tag{5.88}$$

and show that this is consistent with what you know for the continuous-time limit, if we take

$$\Delta B_n := B_{n+1} - B_n. \tag{5.89}$$

Chapter 6

Boundary-Crossing Problems

Now we treat the **boundary-crossing problem**: given a continuous random walk $W(t)$ with $W(0) = 0$, what is the probability that $W(t)$ will cross a ''boundary'' at d by time t? A variation on this ideas is the **escape problem**, which deals with the probability that $W(t)$ will leave some interval by some final time. These are useful statistics in a range of areas.[1] For example, in finance, a knock-in **barrier option** can be exercised only when the price of a security hits a ''barrier'' value, while the knock-out flavor option becomes invalid when the security price reaches the barrier. The same idea applies to **double barrier options**, which become exercisable (or invalid) when the value escapes some price interval.

Another famous problem where boundary crossings apply is the **gambler's ruin problem**. Suppose a gregarious gambler walks into a casino with a wad of money. A simple model of the gambling process is that the gambler wins or loses a small, fixed amount on each game with equal probability. In principle the gambler's cash balance is a discrete random walk, but we can idealize it as a continuous-time walk $W(t)$ in the limit of many games played for a small win/loss amount. The barrier in this case is where the gambler goes broke and must stop playing. We will show that if the gambling keeps going on indefinitely, the gambler will eventually go broke, with unit probability. Good news for the house, not so much for the gambler, and of course this is a for a fair game. The average time to go broke in the fair game is arbitrarily long; however, generally speaking, the house has the edge, so the gambler will typically go broke in finite time. The probability density for the time to go broke in the fair-game case is derived in Section 6.1.1

A variation on gambler's ruin is a sequence of games between two gamblers, where on each game one gambler wins a fixed amount of money at the expense of the other. The games end when one gambler goes broke. Again, this can be idealized as the escape of a Wiener process from an interval. The solution is that the probability of a particular gambler winning all the money is just the fraction of the total money initially possessed by that gambler. This result is derived in Sections 6.3.3.2 and 6.4.3.

Because the mathematical treatment is very similar, we will also consider crossing probabilities in the case of Brownian bridges. These statistics are useful in some quantum-mechanical path integrals, for example.

6.1 Boundary Crossings

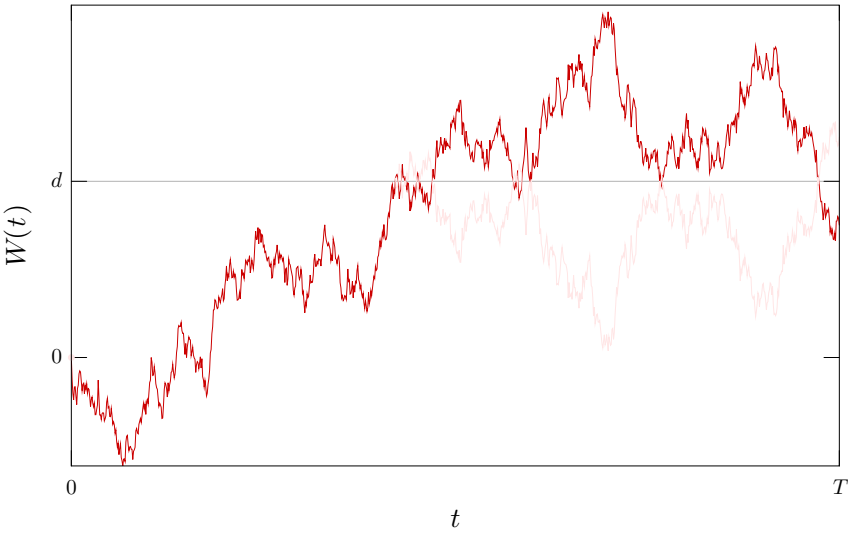
6.1.1 Wiener Process

Consider a Wiener path $W(t)$, where as usual $W(0) = 0$. What is the probability that $W(t)$ will cross a threshold $d > 0$ over this time interval from 0 to t? One approach to this problem is via the **Reflection Principle**.[2] The basic idea is illustrated in the plot below of a Wiener path. Once the path hits the threshold

[1] Sidney Redner, *A Guide to First Passage Processes* (Cambridge, 2001) (ISBN: 0521652480) (doi: 10.1017/CBO9780511606014); David Siegmund, ''Boundary crossing probabilities and statistical applications,'' *The Annals of Statistics* **14**, 361 (1986) (doi: 10.1214/aos/1176349928).

[2] A particularly readable reference for the Reflection Principle appears in Joseph T. Chang, *Stochastic Processes*, available at `http://www.stat.yale.edu/~jtc5/251/stochastic-processes.pdf`. We follow his arguments here for single-boundary crossing

at d, it continues to diffuse on. However, note that we can construct an *equally likely* path by mirroring the path after the crossing about d, as shown in the light, mirrored trace. This is true when the probability distribution for the increments is symmetric, which is of course the case for standard Brownian motion. Note that we are also implicitly assuming that there is a well-defined crossing, that is, given that $W(t) > d$, there is some time $t' < t$ such that $W(t') = d$, which is guaranteed by the continuity of the Wiener process.



Now let's apply this to the crossing probability, which we will write

$$P_{\text{cross}}(d,t) = P(\tau_d \le t), \tag{6.1}$$

where τ_d is the time of the first crossing of $W(t)$ through d:[3]

$$\tau_d := \inf\{t : W(t) \ge d\}. \tag{6.2}$$

Then we can partition the probability according to whether $W(t)$ ends up below or above the boundary:

$$P_{\text{cross}}(d,t) = P[\tau_d \le t \wedge W(t) < d] + P[\tau_d \le t \wedge W(t) \ge d]. \tag{6.3}$$

We can write out the first term via conditional probabilities according to $P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$, and in the second term, $\tau_d \le t$ is automatically satisfied if $W(t) \ge d$ by path continuity:

$$P_{\text{cross}}(d,t) = P[W(t) < d \,|\, \tau_d \le t]\, P[\tau_d \le t] + P[W(t) \ge d]. \tag{6.4}$$

According to the Reflection Principle,

$$P[W(t) < d \,|\, \tau_d \le t] = \frac{1}{2}. \tag{6.5}$$

By construction in the figure, for each path that has crossed through t and ends up with $W(t) < d$, there is an equally likely (mirrored) path with $W(t) > d$. So the probability of ending up above or below d is the same, given that the boundary is crossed. Then we have

$$P_{\text{cross}}(d,t) = \frac{1}{2} P_{\text{cross}}(d,t) + P[W(t) \ge d], \tag{6.6}$$

probabilities for Wiener paths and bridges. For the Reflection Principle for Wiener paths, see also Kurt Jacobs, *Stochastic Processes for Physicists: Understanding Noisy Systems* (Cambridge, 2010).

[3] For early work on first crossings or first passages in a general language for both Gaussian and general stochastic processes, along with a historical overview and wider motivation, see Arnold J. F. Siegert, "On the First Passage Time Probability Problem," *Physical Review* **81**, 617 (1951) (doi: 10.1103/PhysRev.81.617). Se also, for example, D. A. Darling and A. J. F. Siegert, "The First Passage Problem for a Continuous Markov Process," *The Annals of Mathematical Statistics* **24**, 624 (1953) (doi: 10.1214/aoms/1177728918).

and solving for $P_{\text{cross}}(d,t)$,

$$P_{\text{cross}}(d,t) = 2P[W(t) \geq d]. \tag{6.7}$$

Now $W(t)$ is normally distributed with variance t, so we have

$$P_{\text{cross}}(d,t) = 2\int_d^\infty dW\, \frac{1}{\sqrt{2\pi t}}\, e^{-W^2/2t}, \tag{6.8}$$

or finally[4]

$$P_{\text{cross}}(d,t) = \text{erfc}\left(\frac{d}{\sqrt{2t}}\right)$$
(crossing probability of $W(t)$ through d in time t) (6.9)

for the crossing probability past d in time t. Note that this probability converges to unity as $t \longrightarrow \infty$; even though sample paths can start off in the negative direction, given enough time, they will tend to return to the origin and go far enough into the positive direction to cross the boundary anyway. For small t, this probability reduces to

$$P_{\text{cross}}(d,t) \approx e^{-d^2/2t}\sqrt{\frac{2t}{\pi d^2}}, \tag{6.10}$$

so that the crossing probability is exponentially suppressed as $t \longrightarrow 0$.

The result (6.9) can also be interpreted as a cumulative probability distribution for the crossing to occur before time t. Then the probability density for the **first-passage time** τ_d is given by[5]

$$f_{\tau_d}(x) = \partial_t \text{erfc}\left(\frac{d}{\sqrt{2t}}\right)\bigg|_{t=x} = \frac{d}{\sqrt{2\pi x^3}}\, e^{-d^2/2x},$$
(probability density for first-passage time) (6.11)

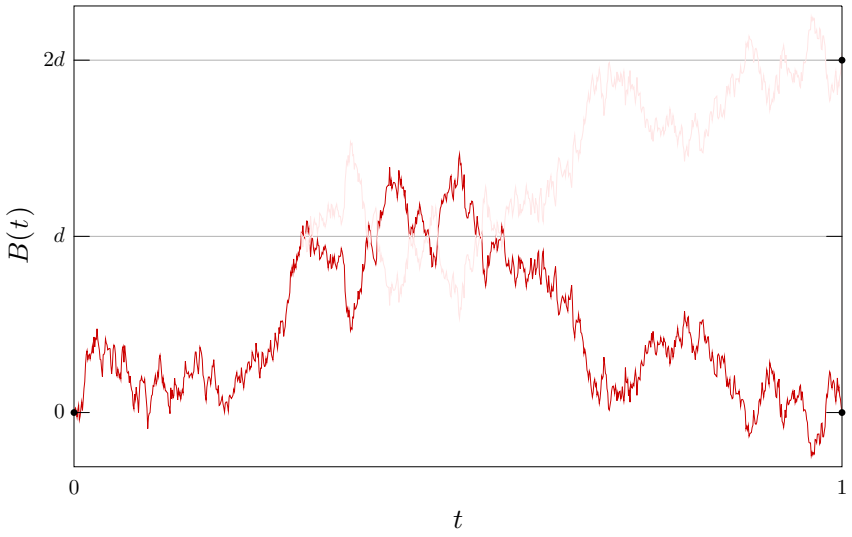
where we have simply differentiated the probability (6.9). Note that the probability density is suppressed exponentially at short times, as $e^{-d^2/2\tau_\text{d}}$, but decays only as $\tau_\text{d}^{-3/2}$ at large times. This distribution has no mean or variance, meaning that may well take a long time to cross a boundary, although it will almost surely happen. However, the *most likely* value of τ_d (i.e., that maximizes the probability density) is $d^2/3$.

6.1.2 Standard Brownian Bridge

A similar argument works to compute boundary-crossing probabilities for the standard Brownian bridge $B(t)$. Here, however, the Reflection Principle is slightly different. The path crossing the boundary at d must return to zero. The equivalent path that is mirrored after the crossing then must return to $2d$ instead of 0, as shown below.

[4] cf. Andrei N. Borodin and Paavo Salminen, *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. (Birkhäuser, 2002), p. 153, formula 1.1.4.

[5] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 198, formula 2.0.2.



Here we will represent the bridge $B(t)$ as an ordinary Wiener path $W(t)$, but subject to the condition $W(1) = 0$. Thus, our crossing probability is the conditional probability

$$P_{\text{cross}}(d) = P[\tau_d \le 1 \,|\, W(1) = 0]. \tag{6.12}$$

Using again the conditional probability relation in the form to $P(A \wedge B)/P(B) = P(A|B)$, we have

$$P_{\text{cross}}(d) = \frac{P[\tau_d \le 1 \wedge W(1) = 0]}{P[W(1) = 0]}. \tag{6.13}$$

To compute the numerator,

$$P[\tau_d \le 1 \wedge W(1) = 0] = P[W(1) = 0 \,|\, \tau_d \le 1]\, P[\tau_d \le 1], \tag{6.14}$$

and using the Reflection Principle,

$$\begin{aligned} P[\tau_d \le t \wedge W(1) = 0] &= P[W(1) = 2d \,|\, \tau_d \le 1]\, P[\tau_d \le 1] \\ &= P[W(1) = 2d \wedge \tau_d \le 1] \\ &= P[W(1) = 2d]. \end{aligned} \tag{6.15}$$

Thus, the crossing probability becomes

$$P_{\text{cross}}(d) = \frac{P[W(1) = 2d]}{P[W(1) = 0]}. \tag{6.16}$$

To interpret these probabilities carefully, the probability $P[W(1) = x]$ refers to the probability of $W(1)$ being between x and $x + dx$; that is, we are referring to probability *densities* here, so that we are taking the ratios of two zero probabilities. Since $P[W(1) = x] = \exp(-x^2/2)\, dx/\sqrt{2\pi}$, then we have[6]

$$P_{\text{cross}}(d) = e^{-2d^2} \tag{6.17}$$

(crossing probability of $B(t)$ through d)

for the crossing probability of the bridge.

Note that the crossing probability is the same as the probability for the peak of the bridge to cross d:

$$P_{\text{cross}}(d) = P\{\sup[B(t)] \ge d\} = 1 - P\{\sup[B(t)] < d\}. \tag{6.18}$$

[6] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 154, formula 1.2.8.

Thus, the probability density function for the maximum of $B(t)$ is

$$f_{\sup\{B(t)\}}(x) = \partial_x P\{\sup[B(t)] < x\} = -\partial_x P_{\rm cross}(x), \tag{6.19}$$

or[7]

$$f_{\sup\{B(t)\}}(x) = 4xe^{-2x^2} \qquad (x \geq 0)$$

(probability density for Brownian-bridge maximum) (6.20)

We can also compute the moments of this distribution, via

$$\left\langle\!\!\left\langle [\sup\{B(t)\}]^n \right\rangle\!\!\right\rangle = \int_0^\infty x^n f_{\sup\{B(t)\}}(x) = \int_0^\infty 4x^{n+1} e^{-2x^2}, \tag{6.21}$$

with the result

$$\left\langle\!\!\left\langle [\sup\{B(t)\}]^n \right\rangle\!\!\right\rangle = 2^{-n/2}\Gamma\left(1+\frac{n}{2}\right),$$

(moments for Brownian-bridge maximum) (6.22)

which gives $(1/4)\sqrt{\pi/2}$, 1/2, $(3/8)\sqrt{\pi/2}$, and 1/2, for $n = 1$, 2, 3, and 4, respectively. Note that the results here are easy to generalize for a Brownian bridge that is pinned such that $B(T) = 0$ instead of $B(1) = 0$, since the pinning essentially rescales the size of the bridge by $\sqrt{T}$, which is equivalent to scaling any distances by $1/\sqrt{T}$. Thus, for example, the crossing probability (6.17) is given by letting $d^2 \longrightarrow d^2/T$, and the peak density (6.20) is given by letting $x^2 \longrightarrow x^2/T$ and $x\,dx \longrightarrow x\,dx/T$ in $f_{\sup\{B(t)\}}(x)\,dx$.

6.1.3 Brownian Bridge

The treatment above for the standard Brownian bridge is easy to generalize to a bridge $B_{0\to c}(t)$ that connects $B_{0\to c}(0) = 0$ to final point $B_{0\to c}(1) = c$, where we obtain

$$P_{\rm cross}(d) = \frac{P[W(1) = 2d - c]}{P[W(1) = c]} = e^{-(2d-c)^2/2 + c^2/2} \tag{6.23}$$

such that Eq. (6.17) generalizes to

$$P_{\rm cross}(d, c) = e^{-2d(d-c)} \qquad (d \geq 0, c).$$

(crossing probability of $B_{0\to c}(t)$ through d) (6.24)

Then Eq. (6.20) similarly becomes

$$f_{\sup\{B_{0\to c}(t)\}}(x) = 2(2x - c)e^{-2x(x-c)} \qquad (x \geq 0, c).$$

(probability density for Brownian-bridge maximum) (6.25)

These results can be generalized to a bridge pinned to c at time t by scaling $d \longrightarrow d/\sqrt{t}$ and $c \longrightarrow c/\sqrt{t}$.

6.1.4 First-Passage Time of the Brownian Bridge

In analyzing the crossing probability for the Wiener path, we were able to simply derive the result (6.11) for the density of the first-passage time τ_d. It is only slightly more complicated to do this for a Brownian bridge, so we will carry out the derivation here. For the Brownian bridge, we can define the first-passage time as

$$\tau_d := \inf\{\tau \leq t : B_{0\to c}(\tau) \geq d\}, \tag{6.26}$$

where $d \geq 0$ and the running time of the bridge is t, such that $B_{0\to c}(t) = c$. Then the crossing probability up to time τ (*before* the final time t) is

$$P_{\rm cross}(d, \tau \leq t) = P[\tau_d \leq \tau \leq t | W(t) = c], \tag{6.27}$$

[7] cf. Eq. (2) in Jim Pitman and Marc Yor, "On the distribution of ranked heights of excursions of a Brownian bridge," *Annals of Probability* **29**, 361 (2001) (doi: 10.1214/aop/1008956334).

when written in terms of a Wiener path. We don't know this crossing probability, but from Eq. (6.24) we *do* know the crossing time of the bridge up to the final running time t. To rewrite Eq. (6.27) in terms of this crossing probability, we can proceed as follows:

$$\begin{aligned} P_{\text{cross}}(d,\tau\le t) &= \sqrt{2\pi t}\,e^{c^2/2t}\,P[\tau_d\le\tau\le t\wedge W(t)=c] \\ &= \sqrt{2\pi t}\,e^{c^2/2t}\int_{-\infty}^{\infty}dz\,P[\tau_d\le\tau\le t\wedge W(\tau)=z]\,P[W(t)=c|W(\tau)=z]. \end{aligned} \tag{6.28}$$

The second factor in the integrand is just a Gaussian probability density,

$$P[W(t)=c|W(\tau)=z] = \frac{1}{\sqrt{2\pi(t-\tau)}}\,e^{-(c-z)^2/2(t-\tau)}, \tag{6.29}$$

and the first factor is given by Eq. (6.24) as

$$\begin{aligned} P[\tau_d\le\tau\le t\wedge W(\tau)=z] &= \frac{1}{\sqrt{2\pi\tau}}\,e^{-z^2/2\tau}P[\tau_d\le\tau\le t|W(\tau)=z] \\ &= \frac{1}{\sqrt{2\pi\tau}}\,e^{-z^2/2\tau}e^{-2d(d-z)/\tau} \\ &= \frac{1}{\sqrt{2\pi\tau}}\,e^{-(2d-z)^2/2\tau} \qquad (z\le d) \end{aligned} \tag{6.30}$$

for the case where the intermediate point z is *not* past the boundary d, and

$$\begin{aligned} P[\tau_d\le\tau\le t\wedge W(\tau)=z] &= \frac{1}{\sqrt{2\pi\tau}}\,e^{-z^2/2\tau}P[\tau_d\le\tau\le t|W(\tau)=z] \\ &= \frac{1}{\sqrt{2\pi\tau}}\,e^{-z^2/2\tau} \qquad (z\ge d) \end{aligned} \tag{6.31}$$

for the case where the path has already crossed the boundary by time τ. Putting these expressions into Eq. (6.28) and carrying out the resulting two integrals for the cases $z\le d$ and $z\ge d$, we obtain the crossing probability

$$P_{\text{cross}}(d,\tau\le t) = \frac{1}{2}\operatorname{erfc}\left(\frac{\tau(d-c)+d(t-\tau)}{\sqrt{2t\tau(t-\tau)}}\right) + \frac{1}{2}\,e^{-2d(d-c)/t}\operatorname{erfc}\left(\frac{\tau(c-d)+d(t-\tau)}{\sqrt{2t\tau(t-\tau)}}\right).$$

(first-passage cumulative probability, Brownian bridge) (6.32)

To check the normalization of this probability note that $\operatorname{erfc}(x)\longrightarrow 0$ as $x\longrightarrow\infty$ and $\operatorname{erfc}(x)\longrightarrow 2$ as $x\longrightarrow-\infty$. Then as $\tau\longrightarrow t$, Eq. (6.32) becomes

$$P_{\text{cross}}(d,t) = \Theta(c-d) + e^{-2d(d-c)/t}\,\Theta(d-c). \tag{6.33}$$

which gives the correct crossing probability from Eq. (6.24). The probability density for the first-passage time is then simply given by differentiating with respect to τ:[8]

$$f_{\tau_d}(x) = \frac{\sqrt{t}}{\sqrt{2\pi x^3(t-x)}}\,d\,e^{-(d\,t-cx)^2/2tx(t-x)}.$$

(probability density for Brownian-bridge first-passage time) (6.34)

Note that in the limit of large t, this reduces to the Wiener-process result (6.11), as it should. This expression is also unnormalized if $c<d$: in this case the normalization is given by Eq. (6.33), since the first-passage time is undefined in the event that a path does not cross the boundary.

[8] L. Beghin and E. Orsingher, "On the maximum of the generalized Brownian bridge," *Lithuanian Mathematical Journal* **39**, 157 (1999) (doi: 10.1007/BF02469280), Eq. (2.15).

The moments of the first-passage time can be written in terms of the density (6.34) as

$$\left\langle\!\left\langle \tau_d^{\,n} \right\rangle\!\right\rangle = \sqrt{\frac{t}{2\pi}}\, d \int_0^t dx\, \frac{x^{n-3/2}}{\sqrt{t-x}}\, e^{-(d\,t-cx)^2/2tx(t-x)}. \tag{6.35}$$

Changing variables to $y = t/x - 1$ gives the alternate expression

$$\left\langle\!\left\langle \tau_d^{\,n} \right\rangle\!\right\rangle = \frac{t^{n-1/2} d}{\sqrt{2\pi}} \int_0^\infty dy\, \frac{(1+y)^{-n}}{\sqrt{y}}\, e^{-[(1+y)d-c]^2/2ty}. \tag{6.36}$$

For $n = 0$, this expression gives the correct normalization

$$\left\langle\!\left\langle \tau_d^{\,0} \right\rangle\!\right\rangle = e^{-d(d-c+|d-c|)/t}. \tag{6.37}$$

The integral is easier to evaluate for $n < 0$ than $n > 0$. For example

$$\left\langle\!\left\langle \tau_d^{-1} \right\rangle\!\right\rangle = \frac{d|d-c| + t + d^2}{td^2}\, e^{-d(d-c+|d-c|)/t} \tag{6.38}$$

gives the first inverse moment—the well-defined value here and of the other inverse moments is an indication of how heavily the value $\tau_d = 0$ is suppressed.

The quantity $(t - \tau_d)$ is the time *after* the first passage time. In the case $c = 0$, for a closed bridge, this is also the statistic for the *last* passage time, because of the time-reversal symmetry of the bridge. Moments of the last passage time can be calculated in the same way as the regular moments. Adapting the same variable change leading to Eq. (6.36), the post-first-passage-time moments are

$$\left\langle\!\left\langle (t-\tau_d)^n \right\rangle\!\right\rangle = \frac{t^{n-1/2} d}{\sqrt{2\pi}} \int_0^\infty dy\, y^{n-1/2} (1+y)^{-n}\, e^{-[(1+y)d-c]^2/2ty}. \tag{6.39}$$

Then, for example, the first inverse moment is given by

$$\left\langle\!\left\langle (t-\tau_d)^{-1} \right\rangle\!\right\rangle = \frac{d + |d-c|}{|d-c|t}\, e^{-d(d-c+|d-c|)/t}. \tag{6.40}$$

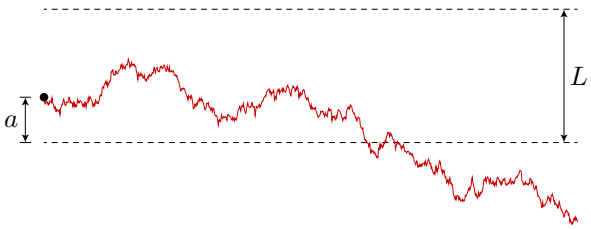
Note that, in the case of $c = 0$, that the mean of the inverse *fractional* last-passage time $t/(t - \tau_d)$ happens to have the same form as the boundary-crossing probability $e^{-2d^2/t}$ from Eq. (6.17), but with an extra factor of 2 (recall that this applies to the case where paths that do *not* cross the boundary count as zero in this ensemble average).

6.2 Escape Probability

Similar to the boundary-crossing problem is the *escape* problem, which is concerned whether a stochastic process leaves an *interval*. There are implicitly two boundaries involved, and the idea is to see whether the process touches *either* boundary. [This is closely related to whether the process touches *both* boundaries, and we can also calculate this via $P(A \wedge B) = P(A) + P(B) - P(A \vee B)$, where the individual probabilities $P(A)$ and $P(B)$ are given by the appropriate single-boundary-crossing probabilities.]

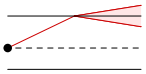
6.2.1 Wiener Process

We will set up the problem as follows: a Wiener process $W(t)$ begins between two barriers separated by distance L, and the starting point is a distance a from one of the barriers. For concreteness, we take the Wiener process' starting point to be distance a from the *lower* barrier, and with $W(0) = 0$, the barriers define the interval $[-a, L - a]$.



Of course, we would obtain the same answers by instead using the interval $[a - L, a]$. The question is, what is the probability to touch either boundary, and thus to escape the interval, in time t?

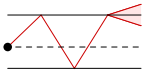
We will approach this problem in the same way as the boundary-crossing, making use of the boundary-crossing probability (6.9) in the process.[9] Actually we will first consider a slightly different problem, which is, what is the probability that $W(t)$ touches the *upper* boundary *first*? That is, we only count the escapes where the *first* escape is through the upper boundary. To compute this, we will define some events (sets of outcomes). First, we define event U_1 to be set of all outcomes where the process touches the upper boundary:

U_1 : 

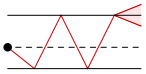
We are illustrating the trajectory schematically here; the trajectory may be much more complicated and touch either boundary many more times than we have indicated. The conical "spray" of trajectories to the right indicates that we don't particularly care what happens to the trajectory afterwards. Now we are only interested in the cases where the process touches the upper boundary first, but we have included cases where the process touches the lower boundary and *then* the upper boundary, since U_1 includes *any* outcome that touches the upper boundary. We will denote this set L_1:

L_1 : 

To properly count the events we want, we should delete events in L_1. But not *all* of them! In L_1 we included paths that touch the upper boundary before touching the lower boundary, and we want to count these. We will denote this set U_2:

U_2 : 

But again, in this set, we are counting paths that touch the lower boundary before the indicated touchings, and we don't want to count these. We will denote this set L_2.

L_2 : 

Continuing in this way, we should define the set U_j to be the set of all paths that touch the upper boundary j times, with $j - 1$ touchings of the lower boundary "interleaved," and L_j to be the set of all paths that touch the lower boundary and then alternating between the upper boundary and lower boundary, with j touchings of each boundary. (Once a boundary touches a boundary, it is okay for it to touch it again before touching the other boundary in these definitions.)

Thus, we have argued that the set of all paths that touch the upper boundary *first* is

$$A_{\text{upper first}} = U_1 - L_1 + U_2 - L_2 + \ldots \tag{6.41}$$

The probabilities of the events on the right-hand side are easy to compute, using the Reflection Principle. (Indeed, notice how the "reflections" pop up here, in a way analogous to the infinity of reflections in a Fabry–Perot cavity.) The idea as before is to "unwrap" trajectories via reflections, and the resulting probability

[9] The method here was applied to the Wiener process by J. L. Doob, "Heuristic Approach to the Kolmogorov–Smirnov Theorems," *The Annals of Mathematical Statistics* **20**, 393 (1949) (doi: 10.1214/aoms/1177729991); and T. W. Anderson, "A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size," *The Annals of Mathematical Statistics* **31**, 165 (1960) (doi: 10.1214/aoms/1177705996).

is just the single-boundary crossing probability (6.9), where the distance d is the total vertical distance traversed in each diagram (noting that the distances to the boundaries are $L-a$ and a above and below the dashed line, respectively. Thus, for example,

$$\begin{aligned} P(U_1) &= \operatorname{erfc}\left(\frac{L-a}{\sqrt{2t}}\right) \\ P(L_1) &= \operatorname{erfc}\left(\frac{L+a}{\sqrt{2t}}\right) \\ P(U_2) &= \operatorname{erfc}\left(\frac{3L-a}{\sqrt{2t}}\right) \\ P(L_2) &= \operatorname{erfc}\left(\frac{3L+a}{\sqrt{2t}}\right), \end{aligned} \tag{6.42}$$

and so on. The probability to touch the upper boundary before the lower boundary is then

$$\begin{aligned} P_{\text{upper first}}(t) &= \operatorname{erfc}\left(\frac{L-a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{L+a}{\sqrt{2t}}\right) + \operatorname{erfc}\left(\frac{3L-a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{3L+a}{\sqrt{2t}}\right) + \dots \\ &= \sum_{j=1}^{\infty}\left[\operatorname{erfc}\left(\frac{(2j-1)L-a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{(2j-1)L+a}{\sqrt{2t}}\right)\right]. \end{aligned} \tag{6.43}$$

The probability to touch the *lower* boundary *before* the upper boundary is simply given by the replacement $a \longrightarrow L-a$ in the above expression:

$$\begin{aligned} P_{\text{lower first}}(t) &= \sum_{j=1}^{\infty}\left[\operatorname{erfc}\left(\frac{(2j-2)L+a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{(2j)L-a}{\sqrt{2t}}\right)\right] \\ &= \operatorname{erfc}\left(\frac{a}{\sqrt{2t}}\right) + \sum_{j=1}^{\infty}\left[\operatorname{erfc}\left(\frac{2jL+a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{2jL-a}{\sqrt{2t}}\right)\right]. \end{aligned} \tag{6.44}$$

The escape probability is the sum of the above two probabilities,

$$P_{\text{escape}}(t) = P_{\text{upper first}}(t) + P_{\text{lower first}}(t), \tag{6.45}$$

since they represent two disjoint sets of outcomes:[10]

$$\begin{aligned} P_{\text{escape}}(t) &= \operatorname{erfc}\left(\frac{a}{\sqrt{2t}}\right) + \sum_{j=1}^{\infty}(-1)^j\left[\operatorname{erfc}\left(\frac{jL+a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{jL-a}{\sqrt{2t}}\right)\right] \\ &= \sum_{j=-\infty}^{\infty}(-1)^j \operatorname{sgn}(j+0^+)\operatorname{erfc}\left(\frac{|a+jL|}{\sqrt{2t}}\right). \end{aligned}$$

(escape probability for Wiener process) (6.46)

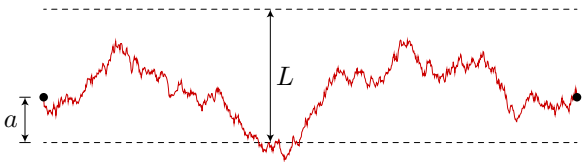
Note that the 0^+ is included in the sgn function so that the $j=0$ term is positive.

6.2.2 Standard Brownian Bridge

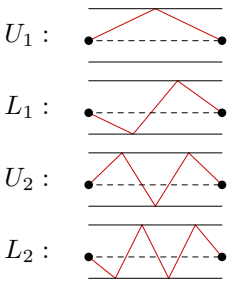
The calculation for the Brownian bridge goes in essentially the same was as for the Wiener path.[11] We set up the problem with the same dimensions as before, the only difference being the constraint $B(1)=0$ on the Wiener path $B(t)$.

[10] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 167, formula 1.7.4(2), and see p. 641 for the function definition.

[11] Bruno Casella and Gareth O. Roberts, "Exact Monte Carlo Simulation of Killed Diffusions," *Advances in Applied Probability* **40**, 273 (2008) (doi: 10.1239/aap/1208358896). (Note the typo in the expression for q_j.) Also Alexandros Beskos, Stefano Peluchetti, and Gareth Roberts, "ϵ-Strong Simulation of the Brownian Path," arXiv.org preprint (arXiv: 1110.0110v1).



Again, we consider the altered problem of: what is the probability that $W(t)$ touches the upper boundary first? Again, we define events U_j and L_j as in the Wiener-path case, except that now the final points of the paths are pinned down:



and so on. Continuing in this way, we again define the set U_j to be the set of all paths that touch the upper boundary j times, with $j-1$ touchings of the lower boundary "interleaved," and L_j to be the set of all paths that touch the lower boundary and then alternating between the upper boundary and lower boundary, with j touchings of each boundary.

Thus, the set of all paths that touch the upper boundary first is again

$$A_{\text{upper first}} = U_1 - L_1 + U_2 - L_2 + \dots \tag{6.47}$$

The probabilities of the events on the right-hand side are easy to compute, using the Reflection Principle. The probability in each diagram is just the single-boundary (bridge) crossing probability (6.17), where the distance d is *half* the total vertical distance traversed in each diagram. Thus, for example,

$$\begin{aligned} P(U_1) &= e^{-2(L-a)^2} \\ P(L_1) &= e^{-2L^2} \\ P(U_2) &= e^{-2(2L-a)^2} \\ P(L_2) &= e^{-2(2L)^2}, \end{aligned} \tag{6.48}$$

and so on. The probability to touch the upper boundary before the lower boundary is then

$$\begin{aligned} P_{\text{upper first}} &= e^{-2(L-a)^2} - e^{-2L^2} + e^{-2(2L-a)^2} - e^{-2L^2} + \dots \\ &= \sum_{j=1}^{\infty} \left[e^{-2(jL-a)^2} - e^{-2(jL)^2} \right]. \end{aligned} \tag{6.49}$$

The probability to touch the lower boundary before the upper boundary is again simply given by the replacement $a \longrightarrow L - a$ in the above expression:

$$\begin{aligned} P_{\text{lower first}} &= \sum_{j=1}^{\infty} \left[e^{-2[(j-1)L+a)^2} - e^{-2(jL)^2} \right] \\ &= e^{-2a^2} + \sum_{j=1}^{\infty} \left[e^{-2(jL+a)^2} - e^{-2(jL)^2} \right] \end{aligned} \tag{6.50}$$

The escape probability is once again the sum of the above two probabilities,

$$P_{\text{escape}} = P_{\text{upper first}} + P_{\text{lower first}}. \tag{6.51}$$

The resulting expression is

$$P_{\text{escape}} = e^{-2a^2} + \sum_{j=1}^{\infty} \left[e^{-2(jL-a)^2} + e^{-2(jL+a)^2} - 2e^{-2(jL)^2} \right] \qquad (0 < a < L)$$
$$= 1 + \sum_{j=-\infty}^{\infty} \left[e^{-2(a+jL)^2} - e^{-2(jL)^2} \right].$$

(escape probability, standard Brownian bridge) (6.52)

Notice that besides the obvious difference in the functions appearing here compared to the Wiener case, the *structure* is different: an a-independent term (the last term here) does not appear in the Wiener case.

6.2.3 Brownian Bridge

It is not difficult to generalize the above escape probability to a more general Brownian bridge $B_{0\to c}(t)$, pinned to c at $t = 1$. In the derivation, we summed over terms of the form $\exp(-2d^2)$ from Eq. (6.17) for various distances d. According to (6.24) we just need to change these to terms of the form $\exp[-2d(d-c)]$, with

$$P_{\text{escape}} = e^{-2a(a-c)} + \sum_{j=1}^{\infty} \left[e^{-2(jL-a)(jL-a-c)} + e^{-2(jL+a)(jL+a-c)} - 2e^{-2(jL)(jL-c)} \right]$$
$$(0 < a < L;\ -a < c < L-a)$$

(escape probability, Brownian bridge) (6.53)

as the result.

6.3 Diffusion-Equation Solution

The predictions for crossing probabilities and first-passage times derived through the probabilistic arguments above can be reproduced via solutions of the **diffusion equation** [i.e., the Fokker–Planck equation (3.52) with $\alpha = 0$ and $\beta = 1$]

$$\partial_t f(x,t) = \frac{1}{2}\partial_x^2 f(x,t). \qquad (6.54)$$

(diffusion-equation propagator)

The solution to the diffusion equation, given the initial condition $f(x,t_0) = \delta(x - x_0)$, is is the Gaussian function

$$f(x,t|x_0,t_0) = \frac{1}{\sqrt{2\pi(t-t_0)}}\, e^{-(x-x_0)^2/2(t-t_0)}$$

(diffusion-equation Green function) (6.55)

with time-dependent uncertainty $\sigma(t) = \sqrt{t-t_0}$. Note that this solution only makes sense for $t \geq t_0$. Since it corresponds to the completely localized initial condition, this solution is special, and is often referred to as the **propagator** or **Green function**. The reason is that via the usual conditional Chapman–Kolmogorov equation (3.17), a general solution can be written as

$$f(x,t) = \int dx_0\, f(x,t|x_0,t_0)\, f_0(x_0,t_0), \qquad (6.56)$$

which is a combination of propagators weighted by the initial condition $f_0(x_0,t_0)$.

6.3.1 Image-Method Picture of Boundary Crossing

Now let's return to the boundary-crossing problem from Section 6.1: what is the probability that a Wiener path will cross a boundary at $d > 0$, and how long does it take? In terms of the diffusion equation it will be slightly simpler to set up the initial position of the Wiener paths to be d, and to ask when do they cross the origin? A good way to set up this boundary-crossing problem is to make the region $x < 0$

completely *absorbing*—that is, any particles from the ensemble that cross the boundary are removed from the ensemble, and we can track the crossing probability by computing the (decaying) norm of the probability distribution. In particular, paths are excluded from existing within the absorbing retion beyond $d > 0$. In this sense, the absorbing boundary imposes a **Dirichlet boundary condition** on the probability distribution, $f(x = 0, t) = 0$. It's easiest to satisfy this boundary condition by the **method of images**, which states that if $f(x, t)$ solves the diffusion equation (6.54), then

$$\check{f}(x,t) := f(x,t) - f(-x,t) \tag{6.57}$$

solves the diffusion equation with a Dirichlet boundary at $x = 0$. This is the probability-density counterpart of the reflection principle that we used before. With the solution (6.55), starting at $x_0 = d$ at $t_0 = 0$, we have

$$\check{f}(x,t) = \frac{1}{\sqrt{2\pi t}}\left(e^{-(x-d)^2/2t} - e^{-(x+d)^2/2t}\right) \tag{6.58}$$

as the probability density for Brownian motion starting at $x = d$ with a perfect absorber at $x = 0$.

The survival probability—the probability to *not* cross the origin— is given by integrating over the probability density that has not been absorbed which amounts to integrating the density (6.58) over $x \in [0, \infty)$:

$$P_{\text{surv}}(d,t) = \int_0^\infty dx\, \check{f}(x,t) = \frac{1}{\sqrt{2\pi t}}\left[\sqrt{\frac{\pi t}{2}}\left[1 + \text{erf}\left(\frac{d}{\sqrt{2t}}\right)\right] - \sqrt{\frac{\pi t}{2}}\left[1 - \text{erf}\left(\frac{d}{\sqrt{2t}}\right)\right]\right], \tag{6.59}$$

or simplifying,

$$P_{\text{surv}}(d,t) = \text{erf}\left(\frac{d}{\sqrt{2t}}\right). \tag{6.60}$$

Then the crossing probability is

$$P_{\text{cross}}(d,t) = 1 - P_{\text{surv}}(d,t) = 1 - \text{erf}\left(\frac{d}{\sqrt{2t}}\right) = \text{erfc}\left(\frac{d}{\sqrt{2t}}\right). \tag{6.61}$$

This is the same as the result (6.9) that we obtained before via probability arguments. The first-passage time is obtained from this crossing probability in the same way as before.

6.3.2 Image-Method Picture of Escape

The method of images can of course also be adapted to solving the escape problem of Section 6.2, where a Wiener path escapes from the spatial interval $[0, L]$. This problem requires absorbing boundary conditions at both $x = 0$ and $x = L$. The method of images is more complicated in this case, giving the solution

$$\check{f}(x,t) := \sum_{j=-\infty}^{\infty}\Big[f(x + 2jL, t) - f(2jL - x, t)\Big] \tag{6.62}$$

that vanishes at the boundaries, in terms of the Green function $f(x, t) = f(x, t|x_0, t_0)$. Intuitively, because there are two Dirichlet boundaries, there is an infinity of images (much like the infinity of optical images if you look into one mirror with a parallel mirror behind you). It is not too difficult to verify that this solution satisfies the boundary conditions, and that it is invariant under x translations by $2L$ (see Problem 6.3).

Starting with the initial condition $f(x, 0) = \delta(x - x_0)$, the image-method solution gives

$$\check{f}(x,t) := \frac{1}{\sqrt{2\pi t}}\sum_{j=-\infty}^{\infty}\left[e^{-(x-x_0+2jL)^2/2t} - e^{-(x+x_0+2jL)^2/2t}\right] \tag{6.63}$$

in terms of the explicit Green function. This calculation will be a little easier if we let $j \longrightarrow -j$ in the last term:

$$\check{f}(x,t) = \frac{1}{\sqrt{2\pi t}}\sum_{j=-\infty}^{\infty}\left[e^{-(x-x_0+2jL)^2/2t} - e^{-(x+x_0-2jL)^2/2t}\right]. \tag{6.64}$$

The survival probability is the spatial integral of the distribution,

$$\begin{aligned} P_{\rm surv}(t) &= \int_0^L dx\, \check{f}(x,t) \\ &= \frac{1}{2}\sum_{j=-\infty}^{\infty}\left[\operatorname{erf}\!\left(\frac{L-x_0+2jL}{\sqrt{2t}}\right)+2\operatorname{erf}\!\left(\frac{x_0+2jL}{\sqrt{2t}}\right)-\operatorname{erf}\!\left(\frac{L+x_0-2jL}{\sqrt{2t}}\right)\right] \\ &= \sum_{j=-\infty}^{\infty}\left[\operatorname{erf}\!\left(\frac{x_0+2jL}{\sqrt{2t}}\right)-\frac{1}{2}\operatorname{erf}\!\left(\frac{x_0-2(j+1/2)L}{\sqrt{2t}}\right)-\frac{1}{2}\operatorname{erf}\!\left(\frac{x_0-2(j-1/2)L}{\sqrt{2t}}\right)\right], \end{aligned} \tag{6.65}$$

after using the antisymmetry of $\operatorname{erf}(x)$. Note that it is important to be careful in interpreting this series, because the error functions all attain unit magnitude at fixed t and as $|j| \longrightarrow \infty$. Additionally, the terms for fixed j, all have the same sign, so some cancellation must happen between the $\pm j$ terms. With this in mind, this expression for the survival probability may be rewritten in the form (Problem 6.4)

$$P_{\rm surv}(t) = 1 - \sum_{j=-\infty}^{\infty}(-1)^j \operatorname{sgn}(j+0^+)\operatorname{erfc}\!\left(\frac{|x_0+jL|}{\sqrt{2t}}\right), \tag{6.66}$$

which gives the escape probability

$$P_{\rm escape}(t) = \sum_{j=-\infty}^{\infty}(-1)^j \operatorname{sgn}(j+0^+)\operatorname{erfc}\!\left(\frac{|x_0+jL|}{\sqrt{2t}}\right). \tag{6.67}$$

This is equivalent to the expression (6.46) derived from probability arguments (note that the sum over images parallels the sum over more complicated crossing scenarios in the previous calculation).

6.3.2.1 Short-Time Asymptotics

The escape-probability expression (6.67) is most suited to analyzing the behavior at short times. For the midpoint starting position $x_0 = L/2$, the probability becomes

$$P_{\rm escape}(t) = \sum_{j=-\infty}^{\infty}(-1)^j \operatorname{sgn}(j+0^+)\operatorname{erfc}\!\left(\frac{|(j+1/2)L|}{\sqrt{2t}}\right). \tag{6.68}$$

Differentiating to find the first-escape distribution, using $\partial_t \operatorname{erfc}(c/\sqrt{t}) = ce^{-c^2/t}/\sqrt{\pi t^3}$,

$$f_{\rm escape}(\tau) = \frac{L}{\sqrt{2\pi\tau^3}}\sum_{j=-\infty}^{\infty}(-1)^j \operatorname{sgn}(j+0^+)\,|j+1/2|\,e^{-(j+1/2)^2L^2/2\tau}. \tag{6.69}$$

The short-time behavior is dominated by the $j = 0$ and $j = -1$ terms:

$$f_{\rm escape}(\tau) \sim \frac{L}{\sqrt{2\pi\tau^3}}\,e^{-L^2/8\tau}. \tag{6.70}$$

Note that this short-time expression is double the first-passage density (6.11) for a Wiener process past a barrier at $d = L/2$. This is sensible, since in this regime, there are two options for crossing the boundary, and the presence of a given boundary hasn't had time to "interfere" with the other.

6.3.3 Eigenfunction Approach

The diffusion equation has eigenfunctions in the sense

$$-\lambda f(x,t) = \partial_t f(x,t) = \frac{1}{2}\partial_x^2 f(x,t), \tag{6.71}$$

which thus have a decaying exponential time dependence

$$f(x,t) = f(x)\, e^{-\lambda t} \tag{6.72}$$

in terms of the eigenvalue (decay rate) λ. The eigenfunctions may be written

$$f_n(x) = \sqrt{\frac{2}{L}}\, \sin \frac{n\pi x}{L}, \tag{6.73}$$

where the corresponding eigenvalues

$$\lambda_n = \frac{n^2\pi^2}{2L^2}. \tag{6.74}$$

The eigenfunctions are normalized such that the square integrates to unity over $[0, L]$, so that they are orthonormal in the sense

$$\int_0^L dx\, f_m(x)\, f_n(x) = \delta_{mn}. \tag{6.75}$$

Starting with the initial condition

$$f(x, t_0 = 0) = \delta(x - x_0), \tag{6.76}$$

we can write the general solution as the linear combination

$$f(x,t) = \sum_{n=1}^{\infty} c_n f_n(x)\, e^{-n^2\pi^2 t/2L^2}, \tag{6.77}$$

where the coefficients c_n are determined by the overlap integral at $t = 0$:

$$c_n = \int_0^L dx\, f_n(x)\, f(x,0) = \sqrt{\frac{2}{L}} \int_0^L dx\, \sin\Big(\frac{n\pi x}{L}\Big)\, \delta(x - x_0) = \sqrt{\frac{2}{L}} \sin\Big(\frac{n\pi x_0}{L}\Big). \tag{6.78}$$

Thus, the general solution is

$$f(x,t) = \frac{2}{L} \sum_{n=1}^{\infty} \sin\Big(\frac{n\pi x}{L}\Big)\, \sin\Big(\frac{n\pi x_0}{L}\Big)\, e^{-n^2\pi^2 t/2L^2} \tag{6.79}$$

for the decaying (unescaped) density $\psi(x,t)$.

The survival probability is just the spatial integral of this function, which involves

$$\int_0^L dx\, \sin\Big(\frac{n\pi x}{L}\Big) = \frac{L}{n\pi}(1 - \cos n\pi) = \frac{L}{n\pi}\big[1 - (-1)^n\big]. \tag{6.80}$$

Computing $P_{\text{escape}} = 1 - P_{\text{surv}}$ gives the the escape-probability expression

$$P_{\text{escape}}(L,t) = 1 - \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{[1 - (-1)^n]}{n} \sin\Big(\frac{n\pi x_0}{L}\Big)\, e^{-n^2\pi^2 t/2L^2} \tag{6.81}$$

as an alternative to Eq. (6.67).

6.3.3.1 Escape-Time Distribution and Long-Time Asymptotics

To carry on with Eq. (6.81) a bit farther, we can assume the midpoint condition, $x_0 = L/2$, so that

$$P_{\text{escape}}(t) = 1 + \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1}\, e^{-(2n-1)^2\pi^2 t/2L^2}. \tag{6.82}$$

The derivative of this function gives the distribution of first-escape times:

$$f_{\text{escape}}(\tau) = \frac{d}{d\tau} P_{\text{escape}}(\tau) = \frac{2\pi}{L^2} \sum_{n=1}^{\infty} (-1)^{(n-1)} (2n-1)^2 \, e^{-(2n-1)^2 \pi^2 \tau / 2L^2}. \tag{6.83}$$

In the long-time limit, the escape is dominated by the term involving the smallest eigenvalue ($n = 1$):

$$f_{\text{escape}}(\tau) \sim \frac{2\pi}{L^2} \, e^{-\pi^2 \tau / 2L^2}. \tag{6.84}$$

The long-time decay is thus exponential in this regime, at rate $\pi^2/2L^2$.

6.3.3.2 Probability Currents and the First-Exit Problem

In the probability-based treatment of the escape problem, we segregated terms corresponding to the two ends of the interval where the sample path escaped. This is the **first-exit problem**—given that the random walker will exit a region (and we know that this will eventually happen with probability 1), where on the boundary will the walker exit? We can also get the same information based on further analysis of the unescaped probability density. In terms of the unescaped density, the survival probability at any time is simply its integral:

$$P_{\text{surv}}(t) = \int_0^L dx \, f(x,t). \tag{6.85}$$

Differentiating and using the diffusion equation (6.54),

$$\partial_t P_{\text{surv}} = \int_0^L dx \, \partial_t f(x,t) = \frac{1}{2} \int_0^L dx \, \partial_x^2 f(x,t) = \frac{1}{2} \big[f'(L,t) - f'(0,t) \big]. \tag{6.86}$$

Now we are left with two surface terms, which represent the flux of probability through the two boundaries. In particular, we can define the probability current

$$j(x,t) := -\frac{1}{2} \partial_x f(x,t), \tag{6.87}$$

which is consistent with the definition of probability current density (3.72) for the Fokker–Planck equation. In terms of the current density, the rate of change of the survival probability is given by

$$\partial_t P_{\text{surv}} = j(0,t) - j(L,t), \tag{6.88}$$

which says that the survival probability decreases according to a probability current at $x = L$, and according to a *negative* (i.e., leftward) probability current at $x = 0$. We can thus separate these two processes; integrating the current at $x = 0$ over all time gives the total probability of exiting to the left as

$$P_{\text{left}} = -\int_0^\infty dt \, j(0,t) = \frac{1}{2} \int_0^\infty dt \, f'(0,t). \tag{6.89}$$

Then using the unescaped density (6.79), this probability becomes

$$P_{\text{left}} = \int_0^\infty dt \sum_{n=1}^{\infty} \frac{n\pi}{L^2} \sin\Big(\frac{n\pi x_0}{L}\Big) \, e^{-n^2 \pi^2 t / 2L^2}, \tag{6.90}$$

and then carrying out the integral yields

$$P_{\text{left}} = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin\Big(\frac{n\pi x_0}{L}\Big). \tag{6.91}$$

This is the Fourier series for a sawtooth wave of amplitude 2 and period $2L$; over the relevant domain we may thus write this solution as

$$P_{\text{left}} = 1 - \frac{x_0}{L} \tag{6.92}$$
(left-exit probability)

for the total probability of exiting to the left. The right-exit probability is similarly

$$P_{\text{right}} = \int_0^\infty dt\, j(L,t) = -\frac{1}{2}\int_0^\infty dt\, f'(L,t), \tag{6.93}$$

leading to

$$P_{\text{right}} = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\cos(n\pi)\sin\Big(\frac{n\pi x_0}{L}\Big)\,. \tag{6.94}$$

This of course is also the series for a sawtooth wave, but phase-shifted, and the right-exit probability is then

$$P_{\text{right}} = \frac{x_0}{L}. \tag{6.95}$$
(right-exit probability)

The limits $x_0 = 0, L/2, L$ all make sense from symmetry or the trajectory immediately encountering the boundary. That the behavior in between is so simple (i.e., linear) may be surprising. However, we will see in the following more general analysis of the Dirichlet problem that these which-exit probabilities must be harmonic functions, and thus the linear functions are in some sense expected. (We will rederive these solutions in Section 6.4.3.)

6.4 Dirichlet Problem and Connection to Electrostatics

6.4.1 Laplace Equation

Another interesting context in which boundary crossing and escape arises is in the Laplace equation, subject to Dirichlet boundary conditions. That is, suppose we want the solution to the Laplace equation

$$\nabla^2\phi(\mathbf{r}) = 0, \qquad \phi(\mathbf{r}) = \phi_\partial(\mathbf{r}) \quad \forall_{\mathbf{r}\in\partial D}. \tag{6.96}$$
(Dirichlet problem)

on some bounded domain D, where $\phi_\partial(\mathbf{r})$ fixes the solution on the domain boundary ∂D. (That is, find the electrostatic potential in a charge-free region, given the potential/voltage on a bounding surface.)

We will now show that the solution can be written as the path average

$$\phi(\mathbf{r}) = \Big\langle\!\Big\langle \phi[\mathbf{r} + \mathbf{W}(\tau_\partial)] \Big\rangle\!\Big\rangle_{\mathbf{W}(\tau_\partial)}, \tag{6.97}$$
(stochastic solution)

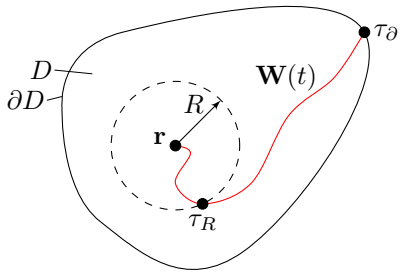
where $\mathbf{W}(t)$ is a vector Wiener process, and τ_∂ is the first passage time of $\mathbf{W}(t)$ through ∂D. That is, we start a bunch of Wiener paths from $\mathbf{r}$, let them go until they hit the boundary, and then take the average of the boundary values where the paths hit. Actually, this solution requires that the boundary be sufficiently nice, which is certainly true for physical boundaries in electrostatic problems. (More technically, the surface should satisfy the "Poincaré cone condition," which basically says that at each point on the surface, you can attach a code of finite angle and length that doesn't intersect the boundary except at the attachment point. This rules out, for example, interior boundaries of arbitrarily small area, or a sufficiently severe "kink" in the surface.[12])

That the stochastic solution (6.97) has the correct boundary values is reasonably obvious, because as the point $\mathbf{r}$ moves close to a point on the boundary, the paths starting from that point will hit the nearest surface with probability approaching unity. This can be seen, for example, from the first-passage-time density

[12] Kiyosi Itô and Henry P. McKean, Jr., *Diffusion Processes and their Sample Paths* (Springer, 1974), pp. 257, 261-4 (doi: 10.1007/978-3-642-62025-6).

for the Wiener path in Eq. (6.11), where the peak of the density shifts to zero as $d \longrightarrow 0$, or that the crossing probability (6.9) in any finite time converges to unity as $d \longrightarrow 0$. However, the boundary function ϕ_∂ should be continuous, as the paths from $\mathbf{r}$ will always average over a small region of the boundary, even as the source point approaches the boundary.

Now to show that the expression (6.97) satisfies the Laplace equation. Suppose we draw a sphere of radius R, centered at $\mathbf{r}$, such that the sphere lies entirely within D, as shown below.



The idea is that the Wiener path must cross the sphere before it crosses the boundary. Then suppose we rewrite Eq. (6.97) as

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \left\langle\!\!\left\langle \phi[\mathbf{r} + \mathbf{W}(\tau_R) + \Delta\mathbf{W}(\tau_\partial)] \right\rangle\!\!\right\rangle_{\Delta\mathbf{W}(\tau_\partial)} \right\rangle\!\!\right\rangle_{\mathbf{W}(\tau_R)}, \tag{6.98}$$

where τ_R is the first crossing time of the sphere, and

$$\Delta\mathbf{W}(\tau_\partial) := \mathbf{W}(\tau_\partial) - \mathbf{W}(\tau_R). \tag{6.99}$$

We haven't really done much here, except to split (pathwise) the time interval into pre- and post-τ_R, and we are explicitly taking the ensemble average after τ_R separately from the ensemble average over all possible first crossings of the sphere represented by τ_R. However, since the path $\Delta\mathbf{W}(\tau_\partial)$ after touching the sphere acts itself like a Wiener path, we can use the solution (6.97) to replace the inner ensemble average:

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \phi[\mathbf{r} + \mathbf{W}(\tau_R)] \right\rangle\!\!\right\rangle_{\mathbf{W}(\tau_R)}. \tag{6.100}$$

Now since the Wiener path is equally likely to have its first touching point $\mathbf{W}(\tau_R)$ at any point on the sphere, we can simplify the notation a bit to write

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \phi(\mathbf{r} + \mathbf{R}) \right\rangle\!\!\right\rangle_{|\mathbf{R}|=R}, \tag{6.101}$$

where $\mathbf{R}$ determines some point on the sphere, and the ensemble average is simply a uniform average over the surface of the sphere. That $\phi(\mathbf{r})$ is the average value of ϕ on a sphere centered on $\mathbf{r}$ is a necessary and sufficient condition for ϕ to be a **harmonic function** (i.e., a solution to the Laplace equation).[13]

6.4.2 Laplacian as Spherical Average

We can make the above argument about the averaging property of harmonic functions more precise by working it out somewhat more explicitly, and in the process arrive at a useful representation of the Laplacian. First, consider the Taylor expansion

$$\phi(\mathbf{r} + \mathbf{R}) = \phi(\mathbf{r}) + \big[\partial_\alpha \phi(\mathbf{r})\big] R_\alpha + \frac{1}{2}\big[\partial_\alpha \partial_\beta \phi(\mathbf{r})\big] R_\alpha R_\beta + O(R^3), \tag{6.102}$$

[13] See David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice Hall, 2013), Section 3.1.4, p. 117. The idea is to consider a point charge outside the sphere, and show that the averaging statement is true for this case. Then the same statement must be true for any collection of charges, and conversely for any solution of the Laplace equation, since any physical solution may be regarded as being produced by some charge distribution.

where repeated indices are summed. Averaging over all orientations of $\mathbf{R}$ at fixed distance R gives

$$\left\langle\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\right\rangle_{|\mathbf{R}|=R} = \phi(\mathbf{r}) + \frac{1}{2}\left\langle\!\left\langle \left[\partial_\alpha^2 \phi(\mathbf{r})\right] R_\alpha^2 \right\rangle\!\right\rangle_{|\mathbf{R}|=R} + O(R^3), \tag{6.103}$$

where all first derivatives have vanished under the symmetric average. Now we can rewrite the last term using

$$\begin{aligned} \left\langle\!\left\langle \sum_{\alpha=1}^{d} \left[\partial_\alpha^2 \phi(\mathbf{r})\right] R_\alpha^2 \right\rangle\!\right\rangle_{|\mathbf{R}|=R} &= \sum_{\alpha=1}^{d} \left[\partial_\alpha^2 \phi(\mathbf{r})\right] \left\langle\!\left\langle R_\alpha^2 \right\rangle\!\right\rangle_{|\mathbf{R}|=R} \\ &= \sum_{\alpha=1}^{d} \left[\partial_\alpha^2 \phi(\mathbf{r})\right] \frac{R^2}{d} \\ &= \frac{R^2}{d} \nabla^2 \phi(\mathbf{r}), \end{aligned} \tag{6.104}$$

where we are writing out the sum explicitly now over d dimensions, and we used the fact that R_α^2 is independent of α once averaged over all orientations. Then putting this result into Eq. (6.103) and taking the limit $R \longrightarrow 0$

$$\nabla^2 \phi(\mathbf{r}) = \lim_{R\to 0^+} \frac{2d}{R^2} \left[\left\langle\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\right\rangle_{|\mathbf{R}|=R} - \phi(\mathbf{r}) \right], \tag{6.105}$$

(Laplacian as spherical average)

which is a representation of the Laplacian operator in terms of an average over a small sphere around $\mathbf{r}$. This equation immediately implies that Eq. (6.101) is equivalent to the Laplace equation (6.96), completing our proof of the Wiener-path solution.

6.4.3 First-Exit Problem, Revisited

Returning to the first-exit problem of Section 6.3.3.2, the solution (6.97) in one dimensions gives an easy way to calculate the probability of exiting one particular side of an interval. Namely, consider a Wiener path starting at a position x_0 somewhere in the interval $[0, L]$. We want to compute the probability that the particle first exits at L. So let's assign a score of 1 to paths that first exit at L, and a score of 0 to paths that first exit at 0. The solution (6.97) to the Dirichlet problem says that the average score over all paths starting at x_0 is a harmonic function ($\partial_{x_0}^2 \phi(x_0) = 0$) of x_0, with boundary conditions $\phi(0) = 0$ and $\phi(L) = 1$. The solution is clearly $\phi(x_0) = x_0/L$, which is the first-exit probability to the right from Eq. (6.95).

6.4.4 Poisson Equation

The same basic approach[14] to solving the Laplace equation via an average over stopped paths also works for the Poisson equation

$$\nabla^2 \phi(\mathbf{r}) = \rho(\mathbf{r}), \qquad \phi(\mathbf{r}) = \phi_\partial(\mathbf{r}) \quad \forall_{\mathbf{r}\in\partial D}. \tag{6.106}$$

(Dirichlet Poisson problem)

This is the same as the original problem (6.96), but with the addition of a source $\rho(\mathbf{r})$ (a factor of $1/\epsilon_0$ is absorbed into the source for simplicity). The solution is the same as before, but with the addition of a source-averaging term:

$$\phi(\mathbf{r}) = \left\langle\!\left\langle \phi[\mathbf{r}+\mathbf{W}(\tau_\partial)] - \frac{1}{2}\int_0^{\tau_\partial} dt\, \rho[\mathbf{r}+\mathbf{W}(t)] \right\rangle\!\right\rangle_{\mathbf{W}(\tau_\partial)}. \tag{6.107}$$

(stochastic Poisson solution)

[14] For a rigorous version of the argument here, see Sidney C. Port and Charles J. Stone, *Brownian Motion and Classical Potential Theory* (Academic Press, 1978) (ISBN: 0124335942), Proposition 5.2 on p. 14 and Section 4.6 on p. 114.

The same argument leading to Eq. (6.98) applies, but the second term splits into two parts, before and after τ_R. The result is then

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\!\right\rangle_{|\mathbf{R}|=R} - \left\langle\!\!\left\langle \frac{1}{2}\int_0^{\tau_R} dt\, \rho[\mathbf{r}+\mathbf{W}(t)] \right\rangle\!\!\right\rangle_{\mathbf{W}(\tau_R)}. \tag{6.108}$$

The portion of the integral from τ_R to τ_∂ was already absorbed into the solution in the first term here. Rearranging and multiplying by $2d/R^2$,

$$\frac{2d}{R^2}\left[\left\langle\!\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\!\right\rangle_{|\mathbf{R}|=R} - \phi(\mathbf{r})\right] = \frac{d}{R^2}\left\langle\!\!\left\langle \int_0^{\tau_R} dt\, \rho[\mathbf{r}+\mathbf{W}(t)] \right\rangle\!\!\right\rangle_{\mathbf{W}(\tau_R)}. \tag{6.109}$$

Taking the limit $R \longrightarrow 0$, we can use Eq. (6.105) on the left-hand side, and on the right-hand side, we can treat ρ as being a constant with respect to the integral. Thus

$$\nabla^2\phi(\mathbf{r}) = \rho(\mathbf{r})\frac{d}{R^2}\langle\!\langle \tau_R \rangle\!\rangle_{\mathbf{W}(\tau_R)}. \tag{6.110}$$

At this point, we have dropped the limit $R \longrightarrow 0$ on the right-hand side, as it is not necessary. The statistic involved here is the mean of the first-passage time through a sphere of radius R in d dimensions.

Now to calculate the remaining expectation value. First, recall that for a vector Wiener process in d dimensions, the Euclidean norm is given on average by

$$\left\langle\!\!\left\langle \|\mathbf{W}(t)\|^2 \right\rangle\!\!\right\rangle = td, \tag{6.111}$$

since there are d independent directions of displacement, each contributing t to the variance. Now what we would like to show is that this is still true if t is replaced by the stopping time τ_R, in the sense that

$$\left\langle\!\!\left\langle \|\mathbf{W}(\tau_R)\|^2 \right\rangle\!\!\right\rangle = \langle\!\langle \tau_R \rangle\!\rangle d, \tag{6.112}$$

or indeed any other stopping time. First, let's take t to be some very large time, such that almost certainly $\tau_R < t$ (i.e., $0 < \tau_R < t$). For some paths this will not be true, such that the argument below will miss them; in these cases we can take τ_R to be equal to t, which will produce an error that vanishes in the limit $t \longrightarrow \infty$. Then starting with Eq. (6.111),

$$\begin{aligned} td &= \left\langle\!\!\left\langle \|\mathbf{W}(t)\|^2 \right\rangle\!\!\right\rangle \\ &= \left\langle\!\!\left\langle \|\mathbf{W}(t) - \mathbf{W}(\tau_R) + \mathbf{W}(\tau_R)\|^2 \right\rangle\!\!\right\rangle \\ &= \left\langle\!\!\left\langle \|\mathbf{W}(t) - \mathbf{W}(\tau_R)\|^2 \right\rangle\!\!\right\rangle + \left\langle\!\!\left\langle \|\mathbf{W}(\tau_R)\|^2 \right\rangle\!\!\right\rangle - \left\langle\!\!\left\langle [\mathbf{W}(t) - \mathbf{W}(\tau_R)] \cdot \mathbf{W}(\tau_R) \right\rangle\!\!\right\rangle, \end{aligned} \tag{6.113}$$

where the last term is zero, because the parts of the path $\mathbf{W}(t)$ before and after τ_R are independent, and at least the first factor has zero mean. Then taking on the first term on the right-hand side, we can think of all paths that start at a particular stopping point $\mathbf{W}(\tau_R)$, such that the ensemble average of $\|\mathbf{W}(t) - \mathbf{W}(\tau_R)\|$ is just $(t - \tau_R)d$. Then continue the ensemble average over all stopping times τ_R, so that

$$\left\langle\!\!\left\langle \|\mathbf{W}(t) - \mathbf{W}(\tau_R)\|^2 \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle (t-\tau_R)d \right\rangle\!\!\right\rangle = td - \langle\!\langle \tau_R \rangle\!\rangle d. \tag{6.114}$$

Putting this together with Eq. (6.113), we find the desired result

$$\left\langle\!\!\left\langle \|\mathbf{W}(\tau_R)\|^2 \right\rangle\!\!\right\rangle = \langle\!\langle \tau_R \rangle\!\rangle d, \tag{6.115}$$

of course taking the limit $t \longrightarrow \infty$ (which only appears in the sense that we forced $\tau_R \leq t$) so that the ensemble averages involving τ_R are correct. By definition of τ_R, the left-hand side is R^2, so

$$\langle\!\langle \tau_R \rangle\!\rangle = \frac{R^2}{d}. \tag{6.116}$$

Putting this into Eq. (6.109), we see that it reduces to the Poisson equation (6.106) as desired.

6.5 Exercises

Problem 6.1

Recall that for a generalized Brownian bridge $B_{0\to c}(t)$ that connects $B_{0\to c}(0) = 0$ to final point $B_{0\to c}(1) = c$, the probability of crossing a boundary at d is [Eq. (6.24)]

$$P_{\text{cross}}[B_{0\to c}(t); d, c] = e^{-2d(d-c)} \qquad (d \geq 0, c). \tag{6.117}$$

Using *only* this expression (along with basic probability theory and your own cunning) to obtain the analogous expression [Eq. (6.9)]

$$P_{\text{cross}}[W(t); d, t] = \text{erfc}\left(\frac{d}{\sqrt{2t}}\right) \tag{6.118}$$

for a standard Wiener path $W(t)$ to cross the boundary at $d \geq 0$ in the time interval $[0, t]$.

Hint: rescale the bridge probability to account for the alternate time interval, and interpret this as a conditional probability on a Wiener path $W(t)$, given that $W(t) = c$.

Problem 6.2

Lord Kelvin used a setup similar to the image-method solution to the boundary-crossing problem in Section 6.3.1 to estimate the age of the Earth. First, the **heat-conduction equation** is the same as the diffusion equation (6.54), but with the replacement $t \longrightarrow 2\alpha t$,

$$\partial_t \psi = \alpha \partial_x^2 \psi, \tag{6.119}$$

where α is the **thermal diffusivity**. Here, $\psi(x, t)$ represents the space- and time-dependent temperature profile of a thermally conductive medium. Kelvin's model of the Earth is that it started as a ball of uniform temperature, which then cooled with time due to radiation at the surface. To model this, assume that we are looking at a relatively shallow depth near Earth's surface, so we can treat Earth's surface as flat, and let $\psi(x, t)$ denote the temperature profile at depth $x > 0$. We will also assume constant diffusivity α with depth.

The initial condition is

$$\psi(x, 0) = \psi_0, \tag{6.120}$$

where ψ_0 is a ''core temperature,'' as compared to the surface temperature (we'll reinterpret ψ_0 shortly). The temperature profile is subject to the Dirichlet boundary condition

$$\psi(0, t) = 0, \tag{6.121}$$

representing cooling at the surface (i.e., thermal contact with outer space). Now show that the solution

$$\psi(x, t) = \frac{\psi_0}{\sqrt{4\pi\alpha t}} \int_0^\infty dx' \left(e^{-(x-x')^2/4\alpha t} - e^{-(x+x')^2/4\alpha t}\right) \tag{6.122}$$

satisfies the initial condition and the boundary condition. Carry out the integration, and compute the temperature gradient at the surface

$$G := \left.\frac{\partial \psi}{\partial x}\right|_{x=0}. \tag{6.123}$$

Solving for t, you should find

$$t = \frac{\psi_0^2}{\pi \alpha G^2}. \tag{6.124}$$

The Earth's temperature gradient and thermal diffusivity can both be measured. For ψ_0, this can be taken to be the melting point of the Earth's crust (note that, handily, no *distance* appears, so the

temperature at the edge of the region where α is constant will work nicely here). Lord Kelvin, using the measurements available at the time, gave an estimate[15] of 10^8 years for the age of the Earth, which is not too bad compared to the current estimate of 4.5×10^9 years.

Problem 6.3

Let $f(x,t)$ be a solution to the diffusion equation in one dimension:

$$\partial_t f(x,t) = \frac{1}{2}\partial_x^2 f(x,t). \tag{6.125}$$

Show that

$$\check{f}(x,t) := \sum_{j=-\infty}^{\infty} \Big[\psi(x+2jL,t) - \psi(2jL-x,t)\Big] \tag{6.126}$$

is a (possibly unnormalized) solution to the diffusion equation with absorbing (i.e., Dirichlet) boundaries at $x=0$ and $x=L$. Also verify that this solution is periodic in x.

Problem 6.4

Starting with the expression (6.65)

$$P_{\text{surv}}(t) = \sum_{j=-\infty}^{\infty} \left[\operatorname{erf}\left(\frac{x_0+2jL}{\sqrt{2t}}\right) - \frac{1}{2}\operatorname{erf}\left(\frac{x_0-2(j+1/2)L}{\sqrt{2t}}\right) - \frac{1}{2}\operatorname{erf}\left(\frac{x_0-2(j-1/2)L}{\sqrt{2t}}\right) \right] \tag{6.127}$$

for the survival probability of a Wiener process escaping from $[0,L]$ and starting at $x_0 \in (0,L)$, show that this expression can be written

$$P_{\text{surv}}(t) = 1 - \sum_{j=-\infty}^{\infty} (-1)^j \operatorname{sgn}(j+0^+) \operatorname{erfc}\left(\frac{|x_0+jL|}{\sqrt{2t}}\right). \tag{6.128}$$

[15] T. W. Körner, *Fourier Analysis* (Cambridge, 1988) Chapters 56-8 (ISBN: 0521389917).

Chapter 7

Feynman-Kac Formula

Here we will introduce a powerful method of computing expectation values for SDEs, based on connecting the expectation to a (deterministic) PDE. This should intuitively work, as the diffusion equation solves the same problem as an ensemble of random walkers, but at a general level the exact connection is not especially obvious.

7.1 The Formula

To jump right in, the **Feynman–Kac formula**,[1] solves diffusion problems in terms of integrals over solutions to SDEs. There are numerous forms of this formula, and we will develop here a relatively simple version[2] that considers a forced diffusion equation for the distribution $f(x,t)$,

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x,t)f + g(x,t), \tag{7.1}$$
(PDE for Feynman–Kac formula)

subject to the initial condition

$$f_0(x) = f(x,0). \tag{7.2}$$
(initial condition for Feynman–Kac formula)

The Feynman–Kac formula gives the solution of (7.1) for $t > 0$ as

$$f(x,t) = \left\langle\!\!\left\langle f_0[x+W(t)] \exp\left(-\int_0^t dt'\, V[x+W(t'),t-t']\right) + \int_0^t dt'\, g[x+W(t'),t-t'] \exp\left(-\int_0^{t'} dt''\, V[x+W(t''),t-t'']\right) \right\rangle\!\!\right\rangle, \tag{7.3}$$
(Feynman–Kac formula)

where the ensemble average is over all realizations of $W(t)$. Before continuing to prove this formula, notice that it reduces correctly to $f_0(x)$ as $t \longrightarrow 0$. Also, for a simple, undriven diffusion,

$$\partial_t f = \frac{1}{2}\partial_x^2 f, \tag{7.4}$$

[1] R. P. Feynman, ''Space-Time Approach to Non-Relativistic Quantum Mechanics,'' *Reviews of Modern Physics* **20**, 367 (1948) (doi: 10.1103/RevModPhys.20.367); M. Kac, ''On Distributions of Certain Wiener Functionals,'' *Transactions of the American Mathematical Society* **65**, 1 (1949) (doi: 10.1090/S0002-9947-1949-0027960-X); M. Kac, ''On Some Connections between Probability Theory and Differential and Integral Equations,'' in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, 1951), p. 189 (`http://projecteuclid.org/euclid.bsmsp/1200500229`).

[2] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 103.

the formula reduces to

$$\begin{aligned} f(x,t) &= \Big\langle\!\Big\langle f_0[x+W(t)] \Big\rangle\!\Big\rangle \\ &= \int_{-\infty}^{\infty} dW\, f_0(W)\, \frac{1}{\sqrt{2\pi t}} e^{-(W-x)^2/2t}, \end{aligned} \tag{7.5}$$

which is the convolution of $f_0(x)$ with a Gaussian of variance t, as we expect. The extra terms V and g introduce damping (or "killing" of diffusing particles) and particle sources, respectively, that clearly make the solution more complicated.

7.1.1 Proof: Simple Diffusion

But now to prove the Feynman–Kac formula (7.3), which we will do in increasingly complex stages. First, take again the simple case where $V = g = 0$, so that we have the simple diffusion equation,

$$\partial_t f = \frac{1}{2}\partial_x^2 f, \tag{7.6}$$

with solution

$$f(x,t) = \Big\langle\!\Big\langle f_0[x+W(t)] \Big\rangle\!\Big\rangle. \tag{7.7}$$

We will show explicitly that this is the solution by simply differentiating it with respect to time. Technically, we will just compute the differential $df(t)$, regarding x as a fixed parameter,

$$\partial_t f\, dt = \Big\langle\!\Big\langle \partial_x f_0[x+W(t)]\, dW \Big\rangle\!\Big\rangle + \frac{1}{2}\Big\langle\!\Big\langle \partial_x^2 f_0[x+W(t)]\, dt \Big\rangle\!\Big\rangle, \tag{7.8}$$

where we have differentiated according to the Itō rule. The dW term vanishes in the ensemble average,

$$\partial_t f\, dt = \frac{1}{2}\Big\langle\!\Big\langle f_0''[x+W(t)] \Big\rangle\!\Big\rangle\, dt, \tag{7.9}$$

and differentiating (7.7) with respect to x to evaluate the averages on the right-hand side, we see that the first derivatives cancel, and the second-derivative term gives the remaining term we need in Eq. (7.6), after canceling factors of dt.

7.1.2 Proof: Diffusion with Damping

Now consider the case where we have the damping term $V(x,t)$, but $g = 0$. Then the PDE we wish to generate is

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x,t) f, \tag{7.10}$$

and we want to show that this PDE is solved by

$$f(x,t) = \left\langle\!\!\left\langle f_0[x+W(t)]\, \exp\!\left(-\int_0^t dt'\, V[x+W(t'), t-t']\right) \right\rangle\!\!\right\rangle. \tag{7.11}$$

The procedure here is somewhat more complicated than for the simple diffusion in Eqs. (7.6) and (7.7). To keep the calculation organized, consider the quantity[3]

$$M(t) = f[x+W(t), t''-t]\, \exp\left(-\int_0^t dt'\, V[x+W(t'), t''-t']\right), \tag{7.12}$$

[3]The proof here follows the basic idea of Richard Durrett, *Stochastic Calculus: A Practical Introduction* (CRC Press, 1996), pp. 137-41.

where we assume $f(x,t)$ to satisfy the PDE (7.10), with x and t'' effectively fixed parameters. This is something like the solution (7.11) with the initial condition replaced by the time-dependent solution. We will now compute the differential of M:

$$\begin{aligned} dM(t) = \Big(&- \partial_t f[x+W(t), t''-t]\,dt + \partial_x f[x+W(t), t''-t]\,dW \\ &+ \frac{1}{2}\partial_x^2 f[x+W(t), t''-t]\,dt \Big)\, e^{-\int_0^t dt'\, V[x+W(t'), t''-t]} \\ &+ f[x+W(t), t''-t]\, V[x+W(t), t''-t]\, e^{-\int_0^t dt'\, V[x+W(t'), t''-t]}\, dt. \end{aligned} \tag{7.13}$$

The first, third, and fourth terms here vanish together since we assumed $f(x,t)$ to satisfy (7.10). Thus,

$$dM(t) = \partial_x f[x+W(t), t''-t]\, \exp\left(-\int_0^t dt'\, V[x+W(t'), t''-t'] \right)\, dW. \tag{7.14}$$

In particular, $\langle\langle dM(t) \rangle\rangle = 0$, so $M(t)$ is a martingale, which says that $M(t)$ tracks (at least locally) the average behavior of our solution. We will make use of this as follows. Note that evaluating $\langle\langle M(t) \rangle\rangle$ at $t = t''$, we have

$$\Big\langle\Big\langle M(t'') \Big\rangle\Big\rangle = \left\langle\left\langle f[x+W(t''), 0]\, \exp\left(-\int_0^{t''} dt'\, V[x+W(t'), t''-t'] \right) \right\rangle\right\rangle, \tag{7.15}$$

which is just the desired solution (7.11) with the replacement $t \longrightarrow t''$, while evaluating $\langle\langle M(t) \rangle\rangle$ at $t = 0$ gives the general solution

$$\Big\langle\Big\langle M(0) \Big\rangle\Big\rangle = f(x, t''). \tag{7.16}$$

It follows from integrating $\langle\langle dM(t) \rangle\rangle = 0$ from $t = 0$ to t'' that we have a constant of the motion,

$$\Big\langle\Big\langle M(0) \Big\rangle\Big\rangle = \Big\langle\Big\langle M(t'') \Big\rangle\Big\rangle, \tag{7.17}$$

which directly leads to Eq. (7.11).

7.1.3 Proof: Diffusion with Source

Next, we proceed to the case where we have the source term $g(x,t)$, but $V = 0$. Then the PDE we wish to generate is

$$\partial_t f = \frac{1}{2}\partial_x^2 f + g(x,t), \tag{7.18}$$

and we want to show that this PDE is solved by

$$f(x,t) = \left\langle\left\langle f_0[x+W(t)] + \int_0^t dt'\, g[x+W(t'), t-t'] \right\rangle\right\rangle. \tag{7.19}$$

Again, consider the quantity[4]

$$M(t) = f[x+W(t), t''-t] + \int_0^t dt'\, g[x+W(t'), t''-t'], \tag{7.20}$$

with $f(x,t)$ satisfying Eq. (7.18). The differential is

$$\begin{aligned} dM(t) &= -\partial_t f[x+W(t), t''-t]\,dt + \partial_x f[x+W(t), t''-t]\,dW + \frac{1}{2}\partial_x^2 f[x+W(t), t''-t]\,dt \\ &\quad + g[x+W(t), t''-t] \\ &= \partial_x f[x+W(t), t''-t]\,dW. \end{aligned} \tag{7.21}$$

[4]Richard Durrett, *op. cit.*, pp. 130-6.

Again, we have that $M(t)$ is a martingale,

$$d\Big\langle\!\Big\langle M(t) \Big\rangle\!\Big\rangle = 0, \tag{7.22}$$

which upon integration gives

$$\Big\langle\!\Big\langle M(0) \Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle M(t'') \Big\rangle\!\Big\rangle, \tag{7.23}$$

where

$$\Big\langle\!\Big\langle M(0) \Big\rangle\!\Big\rangle = f(x, t''), \tag{7.24}$$

and

$$\Big\langle\!\Big\langle M(t'') \Big\rangle\!\Big\rangle = \left\langle\!\!\left\langle f[x + W(t), 0] + \int_0^{t''} dt'\, g[x + W(t'), t'' - t'] \right\rangle\!\!\right\rangle, \tag{7.25}$$

thus establishing the desired solution (7.19).

7.1.4 Proof: Diffusion with Damping and Source

Now we return to the general case, Eqs. (7.1) and (7.3). In analogy with the simpler cases, we consider the candidate quantity

$$\begin{aligned} M(t) = {} & f[x + W(t), t'' - t] \, \exp\left(- \int_0^t dt'\, V[x + W(t'), t'' - t'] \right) \\ & + \int_0^t dt'\, g[x + W(t'), t'' - t'] \, \exp\left(- \int_0^{t'} dt'''\, V[x + W(t'''), t'' - t'''] \right). \end{aligned} \tag{7.26}$$

The differential is

$$\begin{aligned} dM(t) = {} & \left(-\partial_t f[x + W(t), t'' - t]\, dt + \partial_x f[x + W(t), t'' - t]\, dW \right. \\ & \left. + \frac{1}{2} \partial_x^2 f[x + W(t), t'' - t]\, dt \right) \exp\left(- \int_0^t dt'\, V[x + W(t'), t'' - t'] \right) \\ & - f[x + W(t), t'' - t]\, V[x + W(t), t'' - t] \, \exp\left(- \int_0^t dt'\, V[x + W(t'), t'' - t'] \right) dt \\ & + g[x + W(t), t'' - t] \, \exp\left(- \int_0^t dt'''\, V[x + W(t'''), t'' - t'''] \right) dt \\ = {} & \partial_x f[x + W(t), t'' - t]\, dW, \end{aligned} \tag{7.27}$$

after using the PDE (7.1) as usual to eliminate terms. Once again, $M(t)$ is a martingale,

$$d\Big\langle\!\Big\langle M(t) \Big\rangle\!\Big\rangle = 0, \tag{7.28}$$

so that

$$\Big\langle\!\Big\langle M(0) \Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle M(t'') \Big\rangle\!\Big\rangle, \tag{7.29}$$

where

$$\Big\langle\!\Big\langle M(0) \Big\rangle\!\Big\rangle = f(x, t''), \tag{7.30}$$

and

$$\begin{aligned} \Big\langle\!\Big\langle M(t'') \Big\rangle\!\Big\rangle = {} & \left\langle\!\!\left\langle f[x + W(t), 0] \, \exp\left(- \int_0^{t''} dt'\, V[x + W(t'), t'' - t'] \right) \right.\right. \\ & \left.\left. + \int_0^{t''} dt'\, g[x + W(t'), t'' - t'] \, \exp\left(- \int_0^{t'} dt'''\, V[x + W(t'''), t'' - t'''] \right) \right\rangle\!\!\right\rangle, \end{aligned} \tag{7.31}$$

which establishes the solution (7.3).

7.1.5 Other Forms

Other forms of the Feynman–Kac theorem are common, for example, that specify *final* conditions for the diffusion equation, or that employ more complicated diffusions. As a simple example of the latter, consider

$$\partial_t f = \alpha(x)\,\partial_x f + \frac{1}{2}\beta^2(x)\,\partial_x^2 f, \tag{7.32}$$

(generalized diffusion equation)

which is solved by

$$f(x,t) = \left\langle\!\!\left\langle f_0[y(t)] \right\rangle\!\!\right\rangle_{y(0)=x},$$

(Feynman–Kac path average) (7.33)

where again $f_0(x) = f(x,0)$, and we assume state-dependent drift and diffusion,

$$dy = \alpha(y)\,dt + \beta(y)\,dW, \qquad y(0) = x,$$

(trajectories for Feynman–Kac equation) (7.34)

with no explicit time dependence in the SDE. Defining

$$M(t) = f[y(t), t' - t], \tag{7.35}$$

the differential is

$$\begin{aligned} dM(t) &= -\partial_t f[y(t),t'-t]\,dt + \partial_x f[y(t),t'-t]\,dy + \frac{1}{2}\partial_x^2 f[y(t),t'-t]\,(dy)^2 \\ &= -\partial_t f[y(t),t'-t]\,dt + \alpha(y,t'-t)\,\partial_x f[y(t),t'-t]\,dt + \beta(y,t'-t)\,\partial_x f[y(t),t'-t]\,dW \\ &\qquad + \frac{1}{2}\beta^2(y,t'-t)\,\partial_x^2 f[y(t),t'-t]\,dt \\ &= \beta(y,t'-t)\,\partial_x f[y(t),t'-t]\,dW. \end{aligned} \tag{7.36}$$

As usual, $M(t)$ is a martingale, meaning $d\langle\!\langle M(t)\rangle\!\rangle = 0$, so that $\langle\!\langle M(0)\rangle\!\rangle = \langle\!\langle M(t')\rangle\!\rangle$, where

$$\left\langle\!\!\left\langle M(0) \right\rangle\!\!\right\rangle = f[y(0),t'] = f(x,t'), \tag{7.37}$$

and

$$\left\langle\!\!\left\langle M(t') \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle f[y(t'),0] \right\rangle\!\!\right\rangle, \tag{7.38}$$

establishing (7.33).

7.1.5.1 General form for State-Dependent Diffusion

For the same state-dependent drift and diffusion represented by the SDE (7.33), but adding decay and source terms to the SDE as in

$$\partial_t f = \alpha(x)\,\partial_x f + \frac{1}{2}\beta^2(x)\,\partial_x^2 f - V(x,t)f + g(x,t),$$

(PDE for generalized Feynman–Kac formula) (7.39)

the corresponding Feynman–Kac formula becomes

$$\begin{aligned} f(x,t) = \Bigg\langle\!\!\Bigg\langle & f_0[y(t)]\,\exp\!\left(-\int_0^t dt'\,V[y(t'),t-t']\right) \\ & + \int_0^t dt'\,g[y(t'),t-t']\,\exp\!\left(-\int_0^{t'} dt''\,V[y(t''),t-t'']\right)\Bigg\rangle\!\!\Bigg\rangle_{y(0)=x}. \end{aligned}$$

(generalized Feynman–Kac formula) (7.40)

The proof for this goes in the same way by considering the martingale function

$$M(t) = f[y(t), t'' - t] \exp\left(-\int_0^t dt'\, V[y(t'), t'' - t']\right) + \int_0^t dt'\, g[y(t'), t'' - t'] \exp\left(-\int_0^{t'} dt'''\, V[y(t'''), t'' - t''']\right), \tag{7.41}$$

where the differential is

$$\begin{aligned} dM(t) &= \left(-\partial_t f[y(t), t'' - t]\, dt + \partial_x f[y(t), t'' - t]\, dy + \frac{1}{2}\partial_x^2 f[y(t), t'' - t]\, dy^2\right) \\ &\qquad \times \exp\left(-\int_0^t dt'\, V[y(t'), t'' - t']\right) \\ &\qquad - f[y(t), t'' - t]\, V[y(t), t'' - t] \exp\left(-\int_0^t dt'\, V[y(t'), t'' - t']\right) dt \\ &\qquad + g[y(t), t'' - t] \exp\left(-\int_0^t dt'''\, V[y(t'''), t'' - t''']\right) dt \\ &= \left(-\partial_t f[y(t), t'' - t]\, dt + \Big[\alpha[y(t)]\, dt + \beta[y(t)]\, dW\Big]\partial_x f[y(t), t'' - t]\right. \\ &\qquad \left. + \frac{1}{2}\beta^2(y,t)\, \partial_x^2 f[y(t), t'' - t]\, dt\right) \exp\left(-\int_0^t dt'\, V[y(t'), t'' - t']\right) \\ &\qquad - f[y(t), t'' - t]\, V[y(t), t'' - t] \exp\left(-\int_0^t dt'\, V[y(t'), t'' - t']\right) dt \\ &\qquad + g[y(t), t'' - t] \exp\left(-\int_0^t dt'''\, V[y(t'''), t'' - t''']\right) dt \\ &= \beta[y(t)]\, \partial_x f[y(t), t'' - t]\, dW. \end{aligned} \tag{7.42}$$

We have already implemented the diffusion equation (7.39) as usual to show that this is a martingale, $\langle\langle dM(t) \rangle\rangle = 0$. It is important to notice here that this argument does not carry through if α and β depend explicitly on time, because then the time dependences do not match in the correct way to permit the diffusion equation to cancel all the deterministic terms of $dM(t)$.

Then setting

$$\left\langle\!\left\langle M(0) \right\rangle\!\right\rangle = \left\langle\!\left\langle M(t'') \right\rangle\!\right\rangle, \tag{7.43}$$

where

$$\left\langle\!\left\langle M(0) \right\rangle\!\right\rangle = f[y(0), t''], \tag{7.44}$$

and

$$\left\langle\!\left\langle M(t'') \right\rangle\!\right\rangle = \left\langle\!\left\langle f[y(t''), 0] \exp\left(-\int_0^{t''} dt'\, V[y(t'), t'' - t']\right) + \int_0^{t''} dt'\, g[y(t'), t'' - t'] \exp\left(-\int_0^{t'} dt'''\, V[y(t'''), t'' - t''']\right) \right\rangle\!\right\rangle, \tag{7.45}$$

and dropping the primes on t'', we arrive at the result (7.41).

7.1.5.2 Time-Dependent Drift and Diffusion: First Form

Again, for

$$dy = \alpha(y,t)\, dt + \beta(y,t)\, dW,$$

(trajectories for Feynman–Kac equation) (7.46)

the approach from the previous section does not carry through because the explicit time dependence of α and β runs forward in time, but the time dependence of f runs backwards in time. However, both should be the same, as they appear in the PDE. Thus consider the alternate martingale function

$$\begin{aligned} M(t) = f[y(t''-t), t''-t] \, \exp\!\left(-\int_0^t dt' \, V[y(t''-t'), t''-t']\right) \\ + \int_0^t dt' \, g[y(t''-t'), t''-t'] \, \exp\!\left(-\int_0^{t'} dt''' \, V[y(t''-t'''), t''-t''']\right), \end{aligned} \tag{7.47}$$

with differential

$$\begin{aligned} dM(t) &= \left(-\partial_t f[y(t''-t), t''-t] \, dt + \partial_x f[y(t''-t), t''-t] \, dy + \frac{1}{2}\partial_x^2 f[y(t''-t), t''-t] \, dy^2\right) \\ &\qquad \times \exp\!\left(-\int_0^t dt' \, V[y(t''-t'), t''-t']\right) \\ &\qquad - f[y(t''-t), t''-t] \, V[y(t''-t), t''-t] \, \exp\!\left(-\int_0^t dt' \, V[y(t''-t'), t''-t']\right) dt \\ &\qquad + g[y(t''-t), t''-t] \, \exp\!\left(-\int_0^t dt''' \, V[y(t''-t'''), t''-t''']\right) dt \\ &= \left(-\partial_t f[y(t''-t), t''-t] \, dt - \Big[\alpha(y, t''-t) \, dt + \beta[y(t), t''-t] \, dW(t''-t)\Big] \partial_x f[y(t), t''-t]\right. \\ &\qquad \left. + \frac{1}{2}\beta^2(y, t''-t) \, \partial_x^2 f[y(t), t''-t] \, dt\right) \exp\!\left(-\int_0^t dt' \, V[y(t''-t'), t''-t']\right) \\ &\qquad - f[y(t''-t), t''-t] \, V[y(t''-t), t''-t] \, \exp\!\left(-\int_0^t dt' \, V[y(t''-t'), t''-t']\right) dt \\ &\qquad + g[y(t''-t), t''-t] \, \exp\!\left(-\int_0^t dt''' \, V[y(t''-t'''), t''-t''']\right) dt \\ &= -\beta(y, t''-t) \, \partial_x f[y(t''-t), t''-t] \, dW(t''-t). \end{aligned} \tag{7.48}$$

Note that all the evolution now is explicitly backwards, including the time dependence of the trajectories $y(t''-t)$. In the last step, we used the PDE

$$\partial_t f = -\alpha(x,t) \, \partial_x f + \frac{1}{2}\beta^2(x,t) \, \partial_x^2 f - V(x,t) f + g(x,t), \tag{7.49}$$

(PDE for generalized Feynman–Kac formula)

which gives the relevant diffusion equation. As usual, setting

$$\Big\langle\!\Big\langle M(0) \Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle M(t'') \Big\rangle\!\Big\rangle, \tag{7.50}$$

where

$$\Big\langle\!\Big\langle M(0) \Big\rangle\!\Big\rangle = f[y(t''), t''], \tag{7.51}$$

and

$$\begin{aligned} \Big\langle\!\Big\langle M(t'') \Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle f[y(0), 0] \, \exp\!\left(-\int_0^{t''} dt' \, V[y(t''-t'), t''-t']\right) \\ + \int_0^{t''} dt' \, g[y(t''-t'), t''-t'] \, \exp\!\left(-\int_0^{t'} dt''' \, V[y(t''-t'''), t''-t''']\right) \Big\rangle\!\Big\rangle, \end{aligned} \tag{7.52}$$

gives the generalized Feynman–Kac formula

$$f(x,t) = \left\langle\!\!\left\langle f_0[y(0)] \exp\!\left(-\int_0^t dt'\, V[y(t-t'),t-t']\right) + \int_0^t dt'\, g[y(t-t'),t-t'] \exp\!\left(-\int_0^{t'} dt''\, V[y(t-t''),t-t'']\right) \right\rangle\!\!\right\rangle_{y(t)=x} .$$

(generalized Feynman–Kac formula) (7.53)

The time dependence can be simplified somewhat by writing

$$f(x,t) = \left\langle\!\!\left\langle f_0[y(0)] \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) + \int_0^t dt'\, g[y(t'),t'] \exp\!\left(-\int_{t-t'}^t dt''\; V[y(t''),t'']\right) \right\rangle\!\!\right\rangle_{y(t)=x} .$$

(generalized Feynman–Kac formula) (7.54)

However, the form (7.53) is somewhat better for interpreting the formula, because it explicitly traces each trajectory $y(t)$, ''beginning'' at $y(t) = x$, *backwards* in time to the initial condition $y(0)$. The formula then weights the average in the first term by the initial distribution f_0 at the corresponding state $y(0)$. The second term corresponds to the creating of a trajectory by the source $g(x,t)$ at time t' (summed over all possible t'), propagating backwards to $t = 0$. In both terms, the trajectory amplitudes were damped by $V(x,t)$, integrated over the temporal extent of each path. The backwards-propagating nature of the solutions is reflected in the above PDE, where the drift coefficient has a minus sign compared to the earlier, ''forward-propagating'' PDE (7.39). (The sign change does not apply to the diffusion term, since the backward propagation is conditioned on a *final* condition, not an *initial* condition.) Note also that while more general than Eq. (7.40) in accounting for explicit time dependence in the SDE coefficients, Eq. (7.40) is much more convenient, from the point of view of simulation—the trajectories here have to be propagated backwards from a particular final point (in which case the SDE acts as an anticipating SDE), and weighted according to the corresponding zero-time point.

7.1.5.3 Time-Dependent Drift and Diffusion: Second Form

An alternative, ''forward-time'' version of the generalized Feynman–Kac formula arises by considering the forward-propagating martingale function

$$M(t) = f[y(t),t] \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) + \int_0^t dt'\, g[y(t'),t'] \exp\!\left(-\int_0^{t'} dt''\, V[y(t''),t'']\right) . \tag{7.55}$$

The differential is

$$\begin{aligned}
dM(t) &= \Big(\partial_t f[y(t),t]\,dt + \partial_x f[y(t),t]\,dy + \frac{1}{2}\partial_x^2 f[y(t),t]\,dy^2\Big)\exp\Big(-\int_0^t dt'\,V[y(t'),t']\Big) \\
&\qquad - f[y(t),t]\,V[y(t),t]\,\exp\Big(-\int_0^t dt'\,V[y(t'),t']\Big)\,dt \\
&\qquad + g[y(t),t]\,\exp\Big(-\int_0^t dt''\,V[y(t''),t'']\Big)\,dt \\
&= \Big(\partial_t f[y(t),t]\,dt + \Big[\alpha(y,t)\,dt + \beta(y,t)\,dW\Big]\partial_x f[y(t),t] \\
&\qquad + \frac{1}{2}\beta^2(y,t)\,\partial_x^2 f[y(t),t]\,dt\Big)\exp\Big(-\int_0^t dt'\,V[y(t'),t']\Big) \\
&\qquad - f[y(t),t]\,V[y(t),t]\,\exp\Big(-\int_0^t dt'\,V[y(t'),t']\Big)\,dt \\
&\qquad + g[y(t),t]\,\exp\Big(-\int_0^t dt''\,V[y(t''),t'']\Big)\,dt \\
&= \beta(y,t)\,\partial_x f[y(t),t]\,dW(t),
\end{aligned} \tag{7.56}$$

where in the last step we used the PDE

$$-\partial_t f = \alpha(x,t)\,\partial_x f + \frac{1}{2}\beta^2(x,t)\,\partial_x^2 f - V(x,t)f + g(x,t).$$

(PDE for generalized Feynman–Kac formula) (7.57)

Then as we have done so many times, we set

$$\Big\langle\!\Big\langle M(0)\Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle M(t)\Big\rangle\!\Big\rangle, \tag{7.58}$$

where

$$\Big\langle\!\Big\langle M(0)\Big\rangle\!\Big\rangle = f[y(0),0], \tag{7.59}$$

and

$$\begin{aligned}
\Big\langle\!\Big\langle M(t)\Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle & f[y(t),t]\,\exp\Big(-\int_0^t dt'\,V[y(t'),t']\Big) \\
& + \int_0^t dt'\,g[y(t'),t']\,\exp\Big(-\int_0^{t'} dt''\,V[y(t''),t'']\Big)\Big\rangle\!\Big\rangle,
\end{aligned} \tag{7.60}$$

leading to the alternate generalized formula

$$f_0(x) = \Big\langle\!\Big\langle f[y(t),t]\,\exp\Big(-\int_0^t dt'\,V[y(t'),t']\Big) + \int_0^t dt'\,g[y(t'),t']\,\exp\Big(-\int_0^{t'} dt''\,V[y(t''),t'']\Big)\Big\rangle\!\Big\rangle_{y(0)=x}.$$

(generalized Feynman–Kac formula) (7.61)

Like Eq. (7.54), this form is in a sense less convenient for simulation than Eq. (7.40), because it obtains the initial distribution from the final distribution. The reverse-time nature of this formula is reflected in the negative time derivative in the PDE (7.57).

Note also that the diffusion equation (7.57) has the form of the Kolmogorov backward equation (3.67). Thus, the derivation of the above is effectively an alternate derivation of the Kolmogorov backward equation, for the evolution of $P(x,t|x_0,t_0)$ with t_0, given a final distribution at t. This is implicit in the form of the solution (7.61), which gives the initial distribution at t_0, given the final distribution at t, such that the time derivative in the PDE (7.57) can be regarded here as a derivative with respect to t_0, and the distribution $f(x,t)$ should be regarded as equivalent to $P(x,t|x_0,t_0)$.

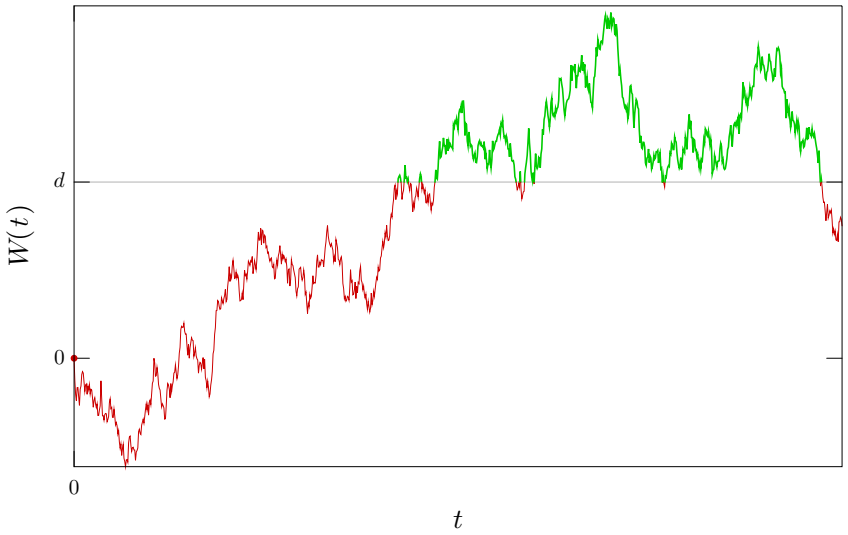
7.2 Sojourn Times

7.2.1 Wiener Process

As an example application of the Feynman–Kac formula, we wish to consider the **sojourn time** of a Wiener path $W(t)$ past a boundary at d, which is defined as the functional

$$T_{\rm s}[W(t);d] := \int_0^t dt'\,\Theta[W(t')-d] \qquad (0 \le T_{\rm s}[W(t);d] \le t), \tag{7.62}$$
(sojourn time)

or the total time that $W(t)$ spends across the boundary. In the plot below, this counts the portion of time that the path is highlighted in green.



We will only consider the case $d \ge 0$, since we can get the $d < 0$ probabilities as complements to the probabilities we calculate.

Let $f_{T_{\rm s}}(x)$ for $0 \le x \le t$ denote the probability density for the sojourn time $T_{\rm s}[W(t);d]$ of $W(t)$ above d, with cumulative probability $P(T_{\rm s} \le x)$, satisfying

$$f_{T_{\rm s}}(x) = \partial_x P(T_{\rm s} \le x). \tag{7.63}$$

Then the Laplace transform of $f_{T_{\rm s}}(x)$ is

$$\int_0^t dx\, e^{-sx}\, f_{T_{\rm s}}(x) = \left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[W(t);d]\right\} \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \exp\left[-s\int_0^t dt'\,\Theta[W(t')-d]\right] \right\rangle\!\!\right\rangle. \tag{7.64}$$

Our goal will be to compute the expectation value on the right via the Feynman–Kac formula, and then to obtain $f_{T_{\rm s}}(x)$ by inverting the Laplace transform.[5]

Consider now the driven diffusion equation

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x)f - \lambda f + g(x), \tag{7.65}$$

where $V(x)$ is the occupation function we seek here (i.e., the step function)—we will take it to be the constant s past the barrier, and 0 before it:

$$V(x) = s\Theta(x-d). \tag{7.66}$$

[5]here we are adapting the method of Gerard Hooghiemstra, "On explicit occupation time distributions for Brownian processes," *Statistics & Probability Letters* **56**, 405 (2002) (doi: 10.1016/S0167-7152(02)00037-8).

The Feynman–Kac formula (7.3) gives the solution of this equation as

$$f(x,t) = \left\langle\!\!\left\langle f_0[x+W(t)] \exp\left(-\lambda t - \int_0^t dt'\, V[x+W(t')]\right) \right.\right. \\ \left.\left. + \int_0^t dt'\, g[x+W(t')] \exp\left(-\lambda t' - \int_0^{t'} dt''\, V[x+W(t'')]\right) \right\rangle\!\!\right\rangle. \tag{7.67}$$

We will seek the steady-state solution $f(x) := f(x, t \longrightarrow \infty)$. In particular, $f(x)$ must be independent of $f_0(x)$, so we will take this initial condition to be zero. Thus, we have the steady state

$$\begin{aligned} f(x) &= \left\langle\!\!\left\langle \int_0^\infty dt \exp\left(-\lambda t - \int_0^t dt'\, V[x+W(t')]\right) \right\rangle\!\!\right\rangle \\ &= \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\, \Theta[x+W(t')-d]\right) \right\rangle\!\!\right\rangle, \end{aligned} \tag{7.68}$$

after dropping some primes and setting $g(x) = 1$. This contains our desired expectation value from Eq. (7.64) upon setting $x = 0$, except that we still here have *another* Laplace transform of the desired expectation value. This is then the solution of the steady-state version of Eq. (7.65):

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\Theta(x-d)f(x) + 1. \tag{7.69}$$

For $x > d$, the ODE is

$$f'' = 2(\lambda+s)f - 2, \tag{7.70}$$

with the case of $x < d$ given by setting $s = 0$. Then setting $h = f - 1/(\lambda+s)$, we have $h'' = 2(\lambda+s)h$, so that for $x > d$,

$$h(x) = Ae^{-\sqrt{2(\lambda+s)}\,x}, \tag{7.71}$$

or

$$f(x) = \begin{cases} Ae^{-\sqrt{2(\lambda+s)}\,x} + \dfrac{1}{\lambda+s} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{1}{\lambda} & (x < d) \end{cases} \tag{7.72}$$

for some undetermined constants A and B, where we have chosen the bounded solution on either side of d. Demanding continuity of $f(x)$ at $x = d$ gives

$$Ae^{-\sqrt{2(\lambda+s)}\,d} + \frac{1}{\lambda+s} = Be^{\sqrt{2\lambda}\,d} + \frac{1}{\lambda}, \tag{7.73}$$

and continuity of $f'(x)$ at $x = d$ gives

$$-\sqrt{2(\lambda+s)}Ae^{-\sqrt{2(\lambda+s)}\,d} = \sqrt{2\lambda}Be^{\sqrt{2\lambda}\,d}. \tag{7.74}$$

The solution of these two equations fixes the coefficients as

$$A = \frac{se^{\sqrt{2(\lambda+s)}\,d}}{(\lambda+s)[\lambda+\sqrt{\lambda(\lambda+s)}]}, \qquad B = -\frac{se^{-\sqrt{2\lambda}\,d}}{\lambda[\lambda+s+\sqrt{\lambda(\lambda+s)}]}. \tag{7.75}$$

Now, we can equate Eqs. (7.68) and (7.72) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\, \Theta[W(t')-d]\right) \right\rangle\!\!\right\rangle = B + \frac{1}{\lambda} = \frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda[\lambda+s+\sqrt{\lambda(\lambda+s)}]}. \tag{7.76}$$

Then using Eq. (7.64) on the left-hand side,

$$\int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\{-sT_{\rm s}[W(t)]\} \right\rangle\!\!\right\rangle = \frac{1}{\lambda} - \frac{s e^{-\sqrt{2\lambda}\, d}}{\lambda[\lambda + s + \sqrt{\lambda(\lambda+s)}]}. \tag{7.77}$$

Now we can use the formula for the inverse Laplace transform

$$y(t) = \frac{1}{2\pi i} \int_{0^+ - i\infty}^{0^+ + i\infty} ds\, e^{s\tau}\, \mathscr{L}[y](s), \tag{7.78}$$

where $\mathscr{L}[y](s)$ is the Laplace transform of $y(t)$:

$$\mathscr{L}[y](s) := \int_0^\infty dt\, e^{-st}\, y(t). \tag{7.79}$$

Then we have

$$\left\langle\!\!\left\langle \exp\{-sT_{\rm s}[W(t)]\} \right\rangle\!\!\right\rangle = \frac{1}{2\pi i} \int_{0^+ - i\infty}^{0^+ + i\infty} d\lambda\, e^{\lambda t} \left[\frac{1}{\lambda} - \frac{s e^{-\sqrt{2\lambda}\, d}}{\lambda[\lambda + s + \sqrt{\lambda(\lambda+s)}]} \right]. \tag{7.80}$$

Evaluating the first term is simple by rotating the integration direction and then completing the contour integral around the great half-plane:

$$\frac{1}{2\pi i} \int_{0^+ - i\infty}^{0^+ + i\infty} d\lambda\, \frac{e^{\lambda t}}{\lambda} = \frac{1}{2\pi i} \int_{-\infty - i0^+}^{\infty - i0^+} d\lambda\, \frac{e^{i\lambda t}}{\lambda} = 1, \qquad (t > 0). \tag{7.81}$$

The second term is more involved, and to evaluate it we will stick to ''recipes'' for known Laplace transforms. We will need[6]

$$\mathscr{L}\left[\frac{e^{-st/2}}{\sqrt{s}} I_{1/2}(st/2) \right](\lambda) = \frac{1}{\sqrt{\lambda}\left(\lambda + s + \sqrt{\lambda(\lambda+s)}\right)}, \tag{7.82}$$

where $I_\nu(x)$ is a modified Bessel function, and[7]

$$\mathscr{L}\left[\frac{1}{\sqrt{\pi t}} e^{-k^2/4t} \right](\lambda) = \frac{1}{\sqrt{\lambda}} e^{-k\sqrt{\lambda}}, \tag{7.83}$$

along with the convolution theorem for Laplace transforms,[8]

$$\mathscr{L}\left[\int_0^t d\tau\, f(t-\tau) g(\tau) \right] = \mathscr{L}[f]\, \mathscr{L}[g], \tag{7.84}$$

which combine to give the Laplace transform appropriate to the second term in Eq. (7.77),

$$\mathscr{L}\left[\int_0^t d\tau\, \frac{e^{-k^2/4(t-\tau)}}{\sqrt{\pi(t-\tau)}} \frac{e^{-s\tau/2}}{\sqrt{s}} I_{1/2}(s\tau/2) \right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\lambda\left(\lambda + s + \sqrt{\lambda(\lambda+s)}\right)}, \tag{7.85}$$

provided we take $k = -\sqrt{2}\, d$ and insert an overall factor of $-s$. Thus, Eq. (7.77) becomes

$$\int_0^t dx\, e^{-sx}\, f_{T_{\rm s}}(x) = 1 - \sqrt{\frac{s}{\pi}} \int_0^t d\tau\, \frac{e^{-d^2/2(t-\tau) - s\tau/2}}{\sqrt{t-\tau}} I_{1/2}(s\tau/2) \tag{7.86}$$

[6] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), p. 1024, Eq. (29.3.54).
[7] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.84).
[8] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1020, Eq. (29.2.8).

after using Eq. (7.64) to replace the ensemble average on the left-hand side. Now we will use an integral representation of the Bessel function,[9]

$$I_{1/2}(x) = \sqrt{\frac{x}{2\pi}} \int_{-1}^{1} du\, e^{-ux} = \sqrt{\frac{2x}{\pi}} e^{x} \int_{0}^{1} dv\, e^{-2vx} = \sqrt{\frac{2x}{\pi}} \frac{e^{x}}{t} \int_{0}^{t} dv\, e^{-2vx/t}, \tag{7.87}$$

where we have changed variables by setting $1 + u = 2v$ and then letting $v \longrightarrow v/t$. Then Eq. (7.86) becomes

$$\begin{aligned} \int_0^t dx\, e^{-sx}\, f_{T_s}(x) &= 1 - \frac{s}{\pi t} \int_0^t d\tau \int_0^t dv\, e^{-vs\tau/t} e^{-d^2/2(t-\tau)} \sqrt{\frac{\tau}{t-\tau}} \\ &= 1 + \frac{1}{\pi} \int_0^t d\tau \left(e^{-s\tau} - 1\right) \frac{e^{-d^2/2(t-\tau)}}{\sqrt{\tau(t-\tau)}} \\ &= 1 - I(d,t) + \frac{1}{\pi} \int_0^t d\tau\, e^{-s\tau} \frac{e^{-d^2/2(t-\tau)}}{\sqrt{\tau(t-\tau)}} \\ &= \int_0^t dx\, e^{-sx}\, \delta(x - 0^+) \left[1 - I(d,t)\right] + \frac{1}{\pi} \int_0^t dx\, e^{-sx} \frac{e^{-d^2/2(t-x)}}{\sqrt{x(t-x)}}, \end{aligned} \tag{7.88}$$

where in the last step we set $\tau = x$, and we have also named the integral

$$I(d,t) := \frac{1}{\pi} \int_0^t dx\, \frac{e^{-d^2/2(t-x)}}{\sqrt{x(t-x)}}, \tag{7.89}$$

which we will evaluate shortly. Eq. (7.88) now equates (finite-time) Laplace transforms. Differentiating with respect to time undoes them, which then yields

$$f_{T_s}(x) = \left[1 - I(d,t)\right] \delta(x - 0^+) + \frac{1}{\pi} \frac{e^{-d^2/2(t-x)}}{\sqrt{x(t-x)}}. \tag{7.90}$$

Clearly, this result is normalized, since integration over x produces canceling terms of $I(d,\tau)$ from the δ function and the integral. The δ function indicates that there is a (possibly) finite probability for having zero sojourn time, whereas positive sojourn times have infinitesimal probability, as we expect for a distribution. In particular, the probability for having zero sojourn time is just

$$\begin{aligned} P(T_s = 0) &= \lim_{\epsilon \to 0} \int_0^{\epsilon} dx\, f_{T_s}(x) \\ &= 1 - I(d,t). \end{aligned} \tag{7.91}$$

However, we have already calculated this: the probability to *not* sojourn across the boundary at d is equivalent to the non-crossing probability for a boundary at d. This is the complement of the crossing probability from Eq. (6.9), and thus

$$P(T_s = 0) = 1 - P_{\text{cross}}(d,t) = 1 - \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) = \operatorname{erf}\left(\frac{d}{\sqrt{2t}}\right). \tag{7.92}$$

Thus, in computing the boundary-crossing probability before, we have essentially used a path-integral method to evaluate the integral (7.89), with the result

$$I(d,t) = \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right). \tag{7.93}$$

Putting this result into Eq. (7.88), we finally have the probability density for the sojourn time of $W(t)$

[9] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 376, Eq. (9.6.18).

beyond d:[10]

$$f_{T_s}(x) = \operatorname{erf}\left(\frac{d}{\sqrt{2t}}\right)\delta(x-0^+) + \frac{e^{-d^2/2(t-x)}}{\pi\sqrt{x(t-x)}} \qquad (0 \le x \le t, d \ge 0).$$

(probability density for sojourn time of $W(t)$ past d) (7.94)

The probability density for the time the particle *stays under* the barrier at d is given by the replacement $x \longrightarrow t - x$. The cumulative probability is given by simply integrating this from 0 to x:

$$P(T_s \le x) = \operatorname{erf}\left(\frac{d}{\sqrt{2t}}\right) + \frac{1}{\pi}\int_0^x dx' \, \frac{e^{-d^2/2(t-x')}}{\sqrt{x'(t-x')}} \qquad (0 \le x \le t, d \ge 0).$$

(cumulative probability for sojourn time of $W(t)$ past d) (7.95)

For the case $d \le 0$, we simply have $P(T_s \le x, -|d|) = 1 - P(T_s \le x, |d|)$, since there are only two, nonintersecting regions in which to sojourn.

The moments are given simply by integrating the density (7.94) as

$$\begin{aligned}\left\langle\!\left\langle {T_s}^n \right\rangle\!\right\rangle &= \int_0^t dx\, x^n f_{T_s}(x) \\ &= \int_0^t dx\, x^n \frac{e^{-d^2/2(t-x)}}{\pi\sqrt{x(t-x)}} \qquad (d \ge 0).\end{aligned} \tag{7.96}$$

A somewhat nicer integral expression comes from changing variables via $x = t(1-\sigma^2)$, such that the moments becomes

$$\left\langle\!\left\langle {T_s}^n \right\rangle\!\right\rangle = \frac{2t^n}{\pi}\int_0^1 d\sigma \left(1-\sigma^2\right)^{n-1/2} e^{-d^2/2t\sigma^2} \qquad (d \ge 0). \tag{7.97}$$

Thus, for example, the mean sojourn time is

$$\left\langle\!\left\langle T_s \right\rangle\!\right\rangle = \frac{1}{2}(d^2+t)\operatorname{erfc}\left[\frac{d}{\sqrt{2t}}\right] - \sqrt{\frac{td^2}{2\pi}}\, e^{-d^2/2t} \qquad (d \ge 0). \tag{7.98}$$

Again, for $d < 0$, this expression for $d \ge 0$ can be adapted by subtracting it from t.

7.2.1.1 Arcsine Laws

One well-known special case of the probability distribution (7.95) is the case of $d = 0$

$$P(T_s \le x) = \frac{1}{\pi}\int_0^x dx' \, \frac{1}{\sqrt{x'(t-x')}}. \tag{7.99}$$

Evaluating this integral gives

$$P(T_s \le x) = \frac{2}{\pi}\sin^{-1}\sqrt{\frac{x}{t}} \qquad (0 \le x \le t, d = 0),$$

(Levy's arcsine law) (7.100)

which is known as **Lévy's arcsine law**.[11] This gives the probability distribution of the time a Wiener path spends on a particular side of its starting point. The density following from Eq. (7.99) is

$$f(x) = \frac{1}{\pi\sqrt{x(t-x)}}. \tag{7.101}$$

(density for Levy arcsine law)

[10] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 156, formula 1.4.4, though note the absence of the first (δ-function) term; Paul Lévy, "Sur certains processus stochastiques homogènes," *Compositio Mathematica* **7**, 283 (1940), p. 327, Eq. (58) (`http://www.numdam.org/item?id=CM_1940__7__283_0`); Marc Yor, "The Distribution of Brownian Quantiles," *Journal of Applied Probability* **32**, 405 (1995), Eq. (3.a) (doi: 10.2307/3215296); Lajos Takács, "Sojourn Times," *Journal of Applied Mathematics and Stochastic Analysis* **9**, 415 (1996), Eq. (3) (doi: 10.1155/S1048953396000366).

[11] Paul Lévy, "Sur un problème de M. Marcinkiewicz," *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **208**, 318 (1939) (`http://gallica.bnf.fr/ark:/12148/bpt6k3160g/f300.image`); Paul Lévy, "Sur certains processus stochastiques homogènes," *op. cit.*

Since this density diverges at $x = 0$ and $x = t$, it says that the path is most likely to spend essentially all its time on one side or the other of the origin.

7.2.1.2 Alternate Arcsine Law: Last Crossing time

The same arcsine law also applies to other statistics related to Wiener processes. To explore the first statistic, suppose we fix a time t, and ask what was the most recent time before t when $W(t)$ crossed the origin?[12] That is, consider the statistic

$$\tau_{\mathrm{L}0}(t) := \sup\{t' : t' \in [0,t] \wedge W(t') = 0\}. \tag{7.102}$$

Then let's compute the cumulative probability $P(\tau_{\mathrm{L}0}(t) \leq \tau)$. For $\tau_{\mathrm{L}0}(t) \leq \tau$ to occur, the Wiener path must not touch $W = 0$ for all $t' > \tau$ (up to t). We could have $W(\tau) = 0$ or not. Then let's integrate over all possible values x of $W(\tau)$, using the Gaussian probability density for $W(\tau)$:

$$P\big[\tau_{\mathrm{L}0}(t) \leq \tau\big] = \int_{-\infty}^{\infty} dx\, f_{W(\tau)}(x)\, P[W(\tau < t' \leq t) \neq 0 | W(\tau) = x], \tag{7.103}$$

where of course the condition "$W(\tau) = x$" means $W(\tau)$ is between x and $x + dx$. Since the Gaussian probability density is even, we can change to evaluating only half of the integral $x \geq 0$:

$$P\big[\tau_{\mathrm{L}0}(t) \leq \tau\big] = 2\int_{0}^{\infty} dx\, f_{W(\tau)}(x)\, P\big[W(t') < x\ \forall_{t' \in [0,t-\tau)}\big]. \tag{7.104}$$

Note we have also changed the second probability statement: it was that a Wiener process, starting at x, should not touch the origin for a time interval of length $t - \tau$. This is the same as the probability that a Wiener process, starting at 0, should not touch x over an interval of the same duration. The latter probability is just the complement of a crossing probability, which we can obtain from the complement of Eq. (6.9). Putting in this probability as an error function, and also putting in the Gaussian density, we have

$$P\big[\tau_{\mathrm{L}0}(t) \leq \tau\big] = 2\int_{0}^{\infty} dx\, \frac{1}{\sqrt{2\pi\tau}}\, e^{-x^2/2\tau}\, \mathrm{erf}\left(\frac{x}{\sqrt{2(t-\tau)}}\right). \tag{7.105}$$

Evaluating this integral, we have

$$P\big[\tau_{\mathrm{L}0}(t) \leq \tau\big] = \frac{2}{\pi}\sin^{-1}\sqrt{\frac{\tau}{t}}.$$

(second arcsine law for last crossing time) (7.106)

Note that this implies the density (7.101) for $\tau_{\mathrm{L}0}(t)$, which is someone odd in that it is a symmetric function over $[0,t]$. This means that the most recent crossing most likely occurred close to $t' = t$ or close to $t' = 0$, and the symmetry is peculiar, given that $W(t')$ itself does not share the symmetry. Note that while the last crossing is well-defined, asking something like when was the *next-to-last* crossing occur is much more complicated: the crossings of $W(t')$ form a set of measure zero, but are nevertheless uncountable, much like the Cantor set. Thus, the last crossing is not isolated, but rather has other, arbitrarily close crossings.

7.2.1.3 Alternate Arcsine Law: Maximum-Value Time

The remaining statistic governed by the arcsine law is the time $\tau_{\max}$ at which a Wiener path attains its maximum value over the time interval $[0,t]$. To see this, it is first necessary to make some comments about the maxima of a Wiener process.[13] First, note that over any time interval $[a,b]$, the maximum of $W(t)$ over the interval does not occur at either endpoint (with unit probability). This follows at the lower endpoint a from noting the equivalence (via some shifting) of $W(t)$ over this interval to a Wiener path starting at

[12] The proof here is from Jim Pitman's course notes, `http://www.stat.berkeley.edu/~pitman/s205s03/lecture18.pdf`, Theorem 18.4.

[13] Peter Mörters and Yuval Peres, *Brownian Motion* (Cambridge, 2010) (ISBN: 0521760186).

$t = 0$, and the crossing probability (6.9): over an interval of any finite duration, the probability of crossing a barrier just above the starting point is arbitrarily close to 1. The maximum does not occur at the upper endpoint b for the same reason (i.e., the same argument under a time-reversal). By subdividing the $[a, b]$ into smaller and smaller intervals, each of which has its own local maximum, we can see that the local maxima are dense (but countable). Also by subdividing $[a, b]$ into two intervals (which do not overlap except at the boundary point), each subinterval will have a local maximum. But because the Wiener process progresses independently in each interval, the local maximum can take on any value, and in particular the probability for attaining the *same* local maximum in each interval is zero. The point is that the maximum over $[a, b]$ is unique, so that $\tau_{\max}$ is well-defined.

Now consider a Wiener process $W(t)$, with the **maximum-value process**

$$M(t) := \max_{0 \le t' \le t} W(t'), \tag{7.107}$$

representing the running maximum value attained by $W(t)$ up to time t. Defining the difference process

$$R(t) := M(t) - W(t), \tag{7.108}$$

we will aim to show that $R(t)$ is statistically equivalent[14] to the **absolute-value process** or **reflection Brownian motion** $|W(t)|$ (about which we will have more to explore in Section 7.3.4.1). To do this, consider a fixed time $t_0 > 0$, and an evolution over a time interval Δt. The Wiener process changes in this interval by

$$\Delta W_0(\Delta t) := W(t_0 + \Delta t) - W(t_0), \tag{7.109}$$

and the maximum changes by

$$\Delta M_0(\Delta t) := \max_{0 \le \delta t \le \Delta t} \Delta W_0(\delta t). \tag{7.110}$$

The idea will be to show that $R(t_0 + \Delta t)$ is statistically equivalent to $|R(t_0) + \Delta W_0(\Delta t)|$. First, note that the maximum at $t_0 + \Delta t$ is

$$M(t_0 + \Delta t) = \max\Big\{M(t_0), W(t_0) + \Delta M_0(\Delta t)\Big\}. \tag{7.111}$$

That is, the final maximum is either the initial maximum, or the initial value of $W(t)$ plus the maximum added over the interval Δt, whichever is larger. Now writing

$$\begin{aligned} R(t_0 + \Delta t) &= M(t_0 + \Delta t) - W(t_0 + \Delta t) \\ &= \max\Big\{M(t_0), W(t_0) + \Delta M_0(\Delta t)\Big\} - \Big[W(t_0) + \Delta W_0(\Delta t)\Big] \\ &= \max\Big\{R(t_0), \Delta M_0(\Delta t)\Big\} - \Delta W_0(\Delta t). \end{aligned} \tag{7.112}$$

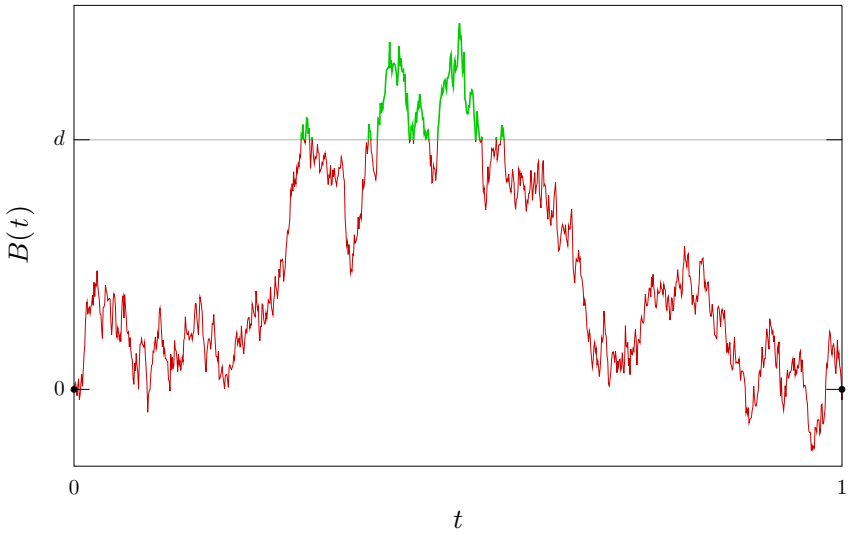
Now to analyze the two cases here. If $R(t_0)$ is the larger value in the maximum, then this reduces to $R(t_0 + \Delta t) = R(t_0) - \Delta W_0(\Delta t)$. Since $\Delta W_0(\Delta t)$ and $-\Delta W_0(\Delta t)$ are statistically equivalent, $R(t_0 + \Delta t)$ is equivalent to $R(t_0) + \Delta W_0(\Delta t)$ in this case, and furthermore the resulting $R(t_0 + \Delta t) \ge 0$ because the maximum value that we are subtracting from $R(t_0)$ is less than $R(t_0)$, by the definition of the maximum and $\Delta M_0(\Delta t)$. In the other case, $\Delta M_0(\Delta t) > R(t_0)$, which means that $R(t_0) - \Delta W_0(\Delta t)$ crosses through zero, and so this case requires more careful treatment. In fact it is useful to identify the time $\delta t_{\max}$ at which the maximum value of $\Delta W_0(\Delta t)$ occurs, so that $\Delta W_0(\delta t_{\max}) = \Delta M_0(\Delta t)$. This is the time when R reaches zero, $R(t_0 + \delta t_{\max}) = 0$, according to Eqs. (7.112). This is intuitive, as $R(t)$ should hit zero anytime $W(t)$ attains a local maximum, according to its definition. Then continuing the evolution, $R(t_0 + \Delta t)$ is the magnitude of the difference between $W(t_0 + \delta t_{\max})$ and $W(t_0 + \Delta t)$. So $R(t_0 + \Delta t)$ is still positive as required, and if we apply the reflection principle to $\Delta W_0(\Delta t)$ at this time $t_0 + \delta t_{\max}$, the reflected version of $R(t_0 + \Delta t)$ is exactly equivalent to a Wiener process that happened to cross through zero at time $t_0 + \delta t_{\max}$.

[14] This argument follows that of Peter Mörters and Yuval Peres, *op. cit.*, at least until near the end.

Returning to the main point, since $M(t) - W(t)$ behaves like $|W(t)|$, the *last* zero of which correspond to the time $\tau_{\max}$ when $W(t)$ achieves its maximum value. Since $|W(t)|$ has the same set of zeros as $W(t)$, and the last zero has the arcsine distribution, the time $\tau_{\max}$ of the maximum likewise has the arcsine distribution (7.106). Somewhat counterintuitively, then, the Wiener process tends to achieve its maximum value near the beginning or near the end of the time interval in question.

7.2.2 Standard Brownian Bridge

Now we will repeat the above analysis to compute the probability distributions for the sojourn time of a Brownian bridge above a boundary at d.[15] As before, in the plot here, this counts the portion of time that the path is highlighted in green; the only difference is the pinning of the path at $t = 0$ to $B(t) = 0$.



We will again take $d \geq 0$ in the analysis here. As before, we define the sojourn time for a bridge $B(t)$ as the functional

$$T_{\rm s}[B(t); d] := \int_0^1 dt\, \Theta[B(t) - d],$$

(sojourn time for standard Brownian bridge) (7.113)

or the total time that $B(t)$ spends across the boundary. The calculation here is similar to the calculation for the Wiener path, and we will refer to that calculation frequently to save effort. The main difference is that we will take $g(x) = e^{ikx}$ in the diffusion equation (7.65) instead of $g(x) = 1$, and then we will integrate with respect to k to introduce a δ function that will pin the endpoint of the Wiener path to make a bridge. This will lead to somewhat more complicated algebra, but nothing conceptually different.

The procedure is the same as for the Wiener path up to Eq. (7.68), which we will leave as

$$f(x) = \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle g[x + W(t)] \, \exp\left(-s \int_0^t dt'\, \Theta[x + W(t') - d]\right) \right\rangle\!\!\right\rangle. \tag{7.114}$$

Now for the steady-state PDE (7.69); this time, taking $g(x) = e^{ikx}$ leads to

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\Theta(x - d) f(x) + e^{ikx}, \tag{7.115}$$

which for $x > d$ is

$$f'' = 2(\lambda + s) f - 2e^{ikx}, \tag{7.116}$$

[15] this is the first calculation in Gerard Hooghiemstra, *op. cit.*

again with the case of $x < d$ given by setting $s = 0$. Then setting $h = f - 2e^{ikx}/[2(\lambda + s) + k^2]$, we have

$$\begin{aligned} h'' &= f'' + \frac{2k^2 e^{ikx}}{2(\lambda+s)+k^2} \\ &= 2(\lambda+s)f - 2e^{ikx} + \frac{2k^2 e^{ikx}}{2(\lambda+s)+k^2} \\ &= 2(\lambda+s)h + \frac{4(\lambda+s)e^{ikx}}{2(\lambda+s)+k^2} - 2e^{ikx} + \frac{2k^2 e^{ikx}}{2(\lambda+s)+k^2} \\ &= 2(\lambda+s)h. \end{aligned} \tag{7.117}$$

so that for $x > d$,

$$h(x) = Ae^{-\sqrt{2(\lambda+s)}\,x}, \tag{7.118}$$

or

$$f(x) = \begin{cases} Ae^{-\sqrt{2(\lambda+s)}\,x} + \dfrac{2e^{ikx}}{2(\lambda+s)+k^2} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{2e^{ikx}}{2\lambda+k^2} & (x < d), \end{cases} \tag{7.119}$$

where A and B must be determined by requiring continuity of f and f' at $x = d$. Solving the two resulting equations gives

$$A = \frac{2\sqrt{2}\,s\left(\sqrt{2\lambda} - ik\right)e^{\sqrt{2(\lambda+s)}\,d+ikd}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}, \qquad B = -\frac{2\sqrt{2}\,s\left(\sqrt{2(\lambda+s)} + ik\right)e^{-\sqrt{2\lambda}\,d+ikd}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}. \tag{7.120}$$

Now, we can equate Eqs. (7.114) and (7.119) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle e^{ikW(t)}\,\exp\left(-s\int_0^t dt'\,\Theta[W(t') - d]\right)\right\rangle\!\!\right\rangle = B + \frac{2}{2\lambda+k^2}. \tag{7.121}$$

Integrating with respect to k and then dividing through by 2π,

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \delta[W(t)]\,\exp\left(-s\int_0^t dt'\,\Theta[W(t') - d]\right)\right\rangle\!\!\right\rangle = \frac{1}{2\pi}\int_{-\infty}^\infty dk\left(B + \frac{2}{2\lambda+k^2}\right). \tag{7.122}$$

The δ function here picks out only the Wiener paths that return to $W(t) = 0$ (i.e., standard Brownian bridges).

We will treat this carefully to obtain the correct normalization for the ensemble average. In the following, we will use t' as a dummy variable and t as the temporal endpoint, so that $0 \le t' \le t$. We have an ensemble average of a functional $F[W(t')]$ (note that $\delta[W(t)]$ only operates on the final value $W(t)$, and so is not itself a functional),

$$\begin{aligned} \left\langle\!\!\left\langle \delta[W(t)]\,F[W(t')]\right\rangle\!\!\right\rangle &= \sum_{W(t')} P[W(t')]\,\delta[W(t)]\,F[W(t')] \\ &= \sum_{\{W(t')|W(t)=0\}} \frac{P[W(t')\wedge W(t)=0]}{dW}\,F[W(t')], \end{aligned} \tag{7.123}$$

where we have expanded into a sum over all possible paths, weighted by the probability $P[W(t')]$ of $W(t')$ to occur, and then implemented the δ function in selecting the probabilities for Wiener paths satisfying $0 \le W(t) < dW$, which we write as $W(t) = 0$ for short. The factor of $1/dW$ comes from thinking of

$\delta(W=0) = 1/dW$, with $\delta(W \neq 0) = 0$ for proper normalization of the δ function. Then we change to a conditional probability via $P(A \wedge B) = P(A|B)P(B)$:

$$\begin{aligned}
\left\langle\!\!\left\langle \delta[W(t)]\, F[W(t')] \right\rangle\!\!\right\rangle &= \sum_{\{W(t')\}} \frac{P[W(t')|W(t)=0]\, P[W(t)=0]}{dW}\, F[W(t')] \\
&= \sum_{\{W(t')|W(t)=0\}} \frac{P[W(t')|W(t)=0]\, dW/\sqrt{2\pi t}}{dW}\, F[W(t')] \\
&= \sum_{\{W(t')|W(t)=0\}} \frac{P[W(t')|W(t)=0]}{\sqrt{2\pi t}}\, F[W(t')] \\
&= \sum_{\{B_t(t')\}} \frac{P[B_t(t')]}{\sqrt{2\pi t}}\, F[B_t(t')] \\
&= \frac{1}{\sqrt{2\pi t}} \left\langle\!\!\left\langle F[B_t(t')] \right\rangle\!\!\right\rangle,
\end{aligned} \tag{7.124}$$

where we have switched to Brownian bridges, which are equivalent to pinned Wiener paths, and we are using the notation $B_T(t) = \sqrt{T}\, B(t/T)$ for a bridge that is pinned at time $t = T$.

Changing the ensemble average in Eq. (7.122) to Brownian bridges, we have

$$\int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s \int_0^t dt'\, \Theta[B_t(t') - d]\right) \right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty dk \left(B + \frac{2}{2\lambda + k^2}\right), \tag{7.125}$$

where we are using the notation $B_T(t) = \sqrt{T}\, B(t/T)$ for a bridge that closes at time T. Then using Eq. (7.64) on the left-hand side, adapted for the bridge, and (7.120) on the right-hand side,

$$\begin{aligned}
\int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\{-sT_{\rm s}[B_t]\} \right\rangle\!\!\right\rangle &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty dk \left[\frac{2}{2\lambda + k^2} - \frac{2\sqrt{2}\, s \left(\sqrt{2(\lambda+s)} + ik\right) e^{-\sqrt{2\lambda}\, d + ikd}}{(2\lambda + k^2)[2(\lambda+s) + k^2] \left(\sqrt{\lambda} + \sqrt{\lambda+s}\right)}\right] \\
&= \sqrt{\frac{\pi}{\lambda}} - \sqrt{\pi}\, \frac{e^{-2\sqrt{2\lambda}\, d}}{\sqrt{\lambda}} \left(\frac{\sqrt{\lambda+s} - \sqrt{\lambda}}{\sqrt{\lambda+s} + \sqrt{\lambda}}\right).
\end{aligned} \tag{7.126}$$

To invert the Laplace transform here, we will use the formula[16]

$$\mathscr{L}\left[\frac{1}{t} e^{-st/2} I_1(st/2)\right](\lambda) = \frac{\sqrt{\lambda+s} - \sqrt{\lambda}}{\sqrt{\lambda+s} + \sqrt{\lambda}} \tag{7.127}$$

along with formula (7.83) (with $k = 2\sqrt{2}\, d$) and the Laplace convolution rule (7.84) to write

$$\mathscr{L}\left[\int_0^t d\tau\, \frac{1}{\sqrt{\pi\tau(t-\tau)}} e^{-2d^2/(t-\tau)} e^{-s\tau/2} I_1(s\tau/2)\right] = \frac{e^{-2\sqrt{2\lambda}\, d}}{\sqrt{\lambda}} \left(\frac{\sqrt{\lambda+s} - \sqrt{\lambda}}{\sqrt{\lambda+s} + \sqrt{\lambda}}\right), \tag{7.128}$$

which takes care of the second term. The first term on the right-hand side of (7.126) is also covered by the formula (7.83) with $k = 0$. Thus, undoing the Laplace transform in Eq. (7.126) gives

$$\left\langle\!\!\left\langle \exp\{-sT_{\rm s}[B_t]\} \right\rangle\!\!\right\rangle = 1 - \int_0^t d\tau\, \frac{1}{\tau} \sqrt{\frac{t}{t-\tau}} e^{-2d^2/(t-\tau)} e^{-s\tau/2} I_1(s\tau/2). \tag{7.129}$$

We can simplify things by considering only a standard Brownian bridge $B(t)$ by setting $t = 1$. Doing this and using Eq. (7.64) on the left-hand side,

$$\int_0^1 dx\, e^{-sx}\, f_{T_{\rm s}}(x) = 1 - \int_0^1 d\tau\, \frac{1}{\tau\sqrt{1-\tau}}\, e^{-2d^2/(1-\tau)} e^{-s\tau/2} I_1(s\tau/2). \tag{7.130}$$

[16] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1024, Eq. (29.3.52).

We will proceed as before by using the integral representation of the Bessel function,[17]

$$I_1(x) = \frac{x}{\pi}\int_{-1}^{1} du\,\sqrt{1-u^2}\,e^{-ux} = \frac{4x}{\pi}e^{x}\int_0^1 dv\,\sqrt{v(1-v)}\,e^{-2vx}, \tag{7.131}$$

where we have changed variables by setting $1+u=2v$. Using this in Eq. (7.130), we find

$$\int_0^1 dx\,e^{-sx}\,f_{T_s}(x) = 1 - \frac{2s}{\pi}\int_0^1 d\tau\int_0^1 dv\,\sqrt{\frac{v(1-v)}{1-\tau}}e^{-2d^2/(1-\tau)}\,e^{-s\tau v}, \tag{7.132}$$

and letting $v = x/\tau$,

$$\begin{aligned}\int_0^1 dx\,e^{-sx}\,f_{T_s}(x) &= 1 - \frac{2s}{\pi}\int_0^1 d\tau\int_0^\tau dx\,\frac{1}{\tau^2}\sqrt{\frac{x(\tau-x)}{1-\tau}}\,e^{-2d^2/(1-\tau)}\,e^{-sx}\\ &= \int_0^1 dx\,e^{-sx}\,\delta(x-0^+) - \frac{2s}{\pi}\int_0^1 dx\,e^{-sx}\int_x^1 d\tau\,\frac{1}{\tau^2}\sqrt{\frac{x(\tau-x)}{1-\tau}}\,e^{-2d^2/(1-\tau)}.\end{aligned} \tag{7.133}$$

Note how we changed the integration limits when interchanging the order of integration in the last step, in order to integrate over the same triangular area in the (x,τ)-plane. We can regard $f_{T_s}(x)$ here as vanishing for $x>1$, since it is nonsensical to consider the standard bridge past $t=1$. If we regard the last term in the same way, we can extend the upper integration limits to ∞ and write out the Laplace transforms as

$$\mathscr{L}\left[f_{T_s}(x)\right] = \mathscr{L}\left[\delta(x-0^+)\right] - s\mathscr{L}\left[\frac{2}{\pi}\int_x^1 d\tau\,\frac{1}{\tau^2}\sqrt{\frac{x(\tau-x)}{1-\tau}}\,e^{-2d^2/(1-\tau)}\right]. \tag{7.134}$$

The Laplace transform of a derivative satisfies

$$s\mathscr{L}[f(x)](s) = \mathscr{L}[f'(x)](s) + f(0), \tag{7.135}$$

so that

$$\begin{aligned}\mathscr{L}\left[f_{T_s}(x)\right] &= \mathscr{L}\left[\delta(x-0^+)\right] - \mathscr{L}\left[\partial_x\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right] - I(d)\\ &= \mathscr{L}\left\{\left[1-I(d)\right]\delta(x-0^+)\right\} - \mathscr{L}\left[\partial_x\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right],\end{aligned} \tag{7.136}$$

where

$$I(d) := \lim_{x\to 0}\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}. \tag{7.137}$$

Thus, we can write the sojourn-time density as

$$f_{T_s}(x) = \left[1-I(d)\right]\delta(x-0^+) - \partial_x\left[\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right]. \tag{7.138}$$

Again, the $[1-I(d)]$ coefficient of the δ function is associated with the probability of having a zero sojourn time. This is the probability of a bridge to *not* touch the boundary. We have already calculated the touching probability in Eq. (6.17), so

$$I(d) = e^{-2d^2}, \tag{7.139}$$

Thus, our first complete expression for the sojourn-time density is

$$f_{T_s}(x) = \left[1-e^{-2d^2}\right]\delta(x-0^+) - \partial_x\left[\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right] \qquad (0\le x\le 1;\ d>0),$$

(probability density for bridge sojourn time) (7.140)

[17] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 376, Eq. (9.6.18).

with cumulative probability[18]

$$P(T_{\rm s} \le x) = 1 - \frac{2\sqrt{x}}{\pi} \int_x^1 d\tau \sqrt{\frac{\tau - x}{1-\tau}} \, \frac{e^{-2d^2/(1-\tau)}}{\tau^2} \qquad (0 \le x \le 1;\ d > 0),$$

(cumulative probability distribution for bridge sojourn time) (7.141)

though we will continue a bit more in massaging the expression here into nicer forms.

Next, we change variables via

$$\tau = \frac{x(1+\sigma^2)}{1 + x\sigma^2}, \tag{7.142}$$

with the result

$$f_{T_{\rm s}}(x) = \left[1 - e^{-2d^2}\right] \delta(x - 0^+) - \partial_x \left[\frac{4(1-x)}{\pi} \, e^{-2d^2/(1-x)} \int_0^\infty d\sigma \, \frac{\sigma^2}{(1+\sigma^2)^2} \, e^{-2\sigma^2 d^2 x/(1-x)} \right].$$

(probability density for bridge sojourn time) (7.143)

Note that here we can identify

$$\begin{aligned} I(d) &= -\lim_{x\to 0} \frac{4(1-x)}{\pi} \, e^{-2d^2/(1-x)} \int_0^\infty d\sigma \, \frac{\sigma^2}{(1+\sigma^2)^2} \, e^{-2(1+s^2 x)d^2} \\ &= \frac{4}{\pi} \, e^{-2d^2} \int_0^\infty d\sigma \, \frac{\sigma^2}{(1+\sigma^2)^2} \\ &= \frac{4}{\pi} \, e^{-2d^2} \int_0^\infty d\sigma \, \frac{\sigma^2}{(1+\sigma^2)^2} \\ &= e^{-2d^2}, \end{aligned} \tag{7.144}$$

which serves as a separate verification of the boundary-crossing result (6.17). Additionally, we can now integrate Eq. (7.143) from 0 to x to obtain the cumulative distribution for the sojourn time. In doing so, the integral of the last term, evaluated at the lower integration limit, will cancel the e^{-2d^2} from the first term, and the result is[19]

$$P(T_{\rm s} \le x) = 1 - \frac{4(1-x)}{\pi} \, e^{-2d^2/(1-x)} \int_0^\infty d\sigma \, \frac{\sigma^2}{(1+\sigma^2)^2} \, e^{-2\sigma^2 d^2 x/(1-x)} \qquad (0 \le x \le 1;\ d \ge 0).$$

(cumulative probability distribution for bridge sojourn time) (7.145)

Finally, we can evaluate the σ integral, with the result[20]

$$P(T_{\rm s} \le x) = 1 + 4d\sqrt{\frac{x(1-x)}{2\pi}} \, e^{-2d^2/(1-x)} - (1 - x + 4d^2 x) \, e^{-2d^2} \, \mathrm{erfc}\left(\sqrt{\frac{2d^2 x}{1-x}} \right) \qquad (0 \le x \le 1;\ d \ge 0).$$

(cumulative probability distribution for bridge sojourn time) (7.146)

The same integration applied to (7.143) gives the explicit probability density[21]

$$f_{T_{\rm s}}(x) = \left[1 - e^{-2d^2}\right] \delta(x - 0^+) + \sqrt{\frac{8d^2(1-x)}{\pi x}} \, e^{-2d^2/(1-x)} + (1 - 4d^2) \, e^{-2d^2} \, \mathrm{erfc}\left(\sqrt{\frac{2d^2 x}{1-x}} \right).$$

(probability density for bridge sojourn time) (7.147)

As usual, these formulae can be extended to a bridge $B_T(t) = \sqrt{T} B(t/T)$ pinned to zero at $t = T$ by the replacements $d \longrightarrow d/\sqrt{T}$ and $x \longrightarrow x/T$. Also, recall that we explicitly assumed $d \ge 0$; for $d < 0$, this formula for $f_{T_{\rm s}}(x)$ can be regarded as the density for $1 - T_{\rm s}$.

[18] Gerard Hooghiemstra, *op. cit.*, Eq. (6).

[19] Gerard Hooghiemstra, *op. cit.*, before Eq. (10).

[20] Lajos Takács, "The Distribution of the Sojourn Time for the Brownian Excursion," *Methodology and Computing in Applied Probability* **1**, 7 (1999), Eq. (6) (doi: 10.1023/A:1010060107265).

[21] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.8.

We can then compute the moments for the sojourn time as

$$\begin{aligned}\left\langle\!\left\langle (T_{\rm s})^n \right\rangle\!\right\rangle &= \int_0^1 dx\, x^n f_{T_{\rm s}}(x) \\ &= 1 - n\int_0^1 dx\, x^{n-1} P(T_{\rm s} \le x) \\ &= n\int_0^1 dx\, x^{n-1}[1 - P(T_{\rm s} \le x)] \\ &= \frac{2n}{\pi}\int_0^1 dx\, x^{n-1/2}\int_x^1 d\tau\, \sqrt{\frac{\tau - x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2} \\ &= \frac{2n}{\pi}\int_0^1 d\tau\, \frac{e^{-2d^2/(1-\tau)}}{\tau^2\sqrt{1-\tau}}\int_0^\tau dx\, x^{n-1/2}\,\sqrt{\tau - x},\end{aligned} \tag{7.148}$$

where we integrated by parts to integrate with respect to the cumulative distribution, for which we used the form (7.141). After performing the x integration, the result is[22]

$$\left\langle\!\left\langle (T_{\rm s})^n \right\rangle\!\right\rangle = \frac{n\Gamma[n+1/2]}{\sqrt{\pi}(n+1)!}\int_0^1 d\tau\, \frac{\tau^{n-1}e^{-2d^2/(1-\tau)}}{\sqrt{1-\tau}}. \tag{7.149}$$

Notice that this moment formula also arises from the moment-generating function in the form (7.129), where the power-series expansion of $e^{-s\tau/2}I_1(s\tau/2)$ in s yields the individual moments here. A somewhat easier form to handle arises by changing variables via $\tau = 1 - \sigma^2$, with the result

$$\left\langle\!\left\langle (T_{\rm s})^n \right\rangle\!\right\rangle = \frac{2n\Gamma[n+1/2]}{\sqrt{\pi}(n+1)!}\int_0^1 d\sigma\,(1-\sigma^2)^{n-1}e^{-2d^2/\sigma^2}. \tag{7.150}$$

(sojourn-time moments)

The integral here does not have a simple general form, but it can be readily evaluated for particular n. For example, we have

$$\begin{aligned}\left\langle\!\left\langle T_{\rm s} \right\rangle\!\right\rangle &= \frac{e^{-2d^2}}{2} - \sqrt{\frac{\pi}{2}}\,d\,\mathrm{erfc}\left[\sqrt{2}\,d\right] \\ \left\langle\!\left\langle (T_{\rm s})^2 \right\rangle\!\right\rangle &= \frac{(1+2d^2)\,e^{-2d^2}}{3} - \sqrt{2\pi}d\left(1 + \frac{4d^2}{3}\right)\mathrm{erfc}\left[\sqrt{2}\,d\right] \\ \left\langle\!\left\langle (T_{\rm s})^4 \right\rangle\!\right\rangle &= \frac{(12 + 87d^2 + 80d^4 + 16d^6)\,e^{-2d^2}}{60} - \frac{\sqrt{2\pi}d\left(105 + 420d^2 + 336d^4 + 64d^6\right)}{240}\,\mathrm{erfc}\left[\sqrt{2}\,d\right]\end{aligned} \tag{7.151}$$

for the first, second, and fourth moments of the sojourn time, remembering that $d \ge 0$ in these expressions.

7.2.3 Brownian Bridge

We can now generalize the above treatment to a more general Brownian bridge $B_{0\to c}(t)$, pinned to c at $t = 1$, so that in this section we will consider the sojourn-time functional

$$T_{\rm s}[B_{0\to c}(t); d] := \int_0^1 dt\,\Theta[B_{0\to c}(t) - d]. \tag{7.152}$$

(sojourn time for Brownian bridge)

To modify the above derivation, after Eq. (7.114), we will choose $g(x) = e^{ik(x-c)}$. The solution of the PDE is then essentially the same, with the result with $\exp(ikx) \longrightarrow \exp[ik(x-c)]$ in (7.119), and $\exp(ikd) \longrightarrow \exp[ik(d-c)]$ in the expression for B in Eqs. (7.120). Then after setting $x = 0$ as before, we have

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\left\langle e^{ik[W(t)-c]}\,\exp\left(-s\int_0^t dt'\,\Theta[W(t') - d]\right)\right\rangle\!\right\rangle = B + \frac{2e^{-ikc}}{2\lambda + k^2}. \tag{7.153}$$

[22] Gerard Hooghiemstra, *op. cit.*, Eq. (10).

in place of (7.153), where we should keep in mind that the form of B is modified here. Then integrating with respect to k and dividing through by 2π gives

$$\int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle \delta[W(t)-c]\, \exp\left(-s\int_0^t dt'\, \Theta[W(t')-d]\right) \right\rangle\!\!\right\rangle = \frac{1}{2\pi}\int_{-\infty}^\infty dk \left(B + \frac{2e^{-ikc}}{2\lambda+k^2}\right). \tag{7.154}$$

in place of Eq. (7.122). Thus we have introduced the correct modification to force the δ function to pin the Wiener paths that return to $W(t)=c$.

Next, we should generalize the result (7.124) to the case of pinning $W(t)$ to c:

$$\begin{aligned}
\left\langle\!\!\left\langle \delta[W(t)-c]\, F[W(t')] \right\rangle\!\!\right\rangle &= \sum_{W(t')} P[W(t')]\, \delta[W(t)-c]\, F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} \frac{P[W(t')\wedge W(t)=c]}{dW}\, F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} \frac{P[W(t')|W(t)=c]\, P[W(t)=c]}{dW}\, F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} \frac{P[W(t')|W(t)=c]\, e^{-c^2/2t}\, dW/\sqrt{2\pi t}}{dW}\, F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} P[W(t')|W(t)=c]\, \frac{e^{-c^2/2t}}{\sqrt{2\pi t}}\, F[W(t')] \\
&= \sum_{\{B_t(t')\}} P[B_{t(0\to c)}(t')]\, \frac{e^{-c^2/2t}}{\sqrt{2\pi t}}\, F[B_{t(0\to c)}(t')] \\
&= \frac{e^{-c^2/2t}}{\sqrt{2\pi t}} \left\langle\!\!\left\langle F[B_{t(0\to c)}(t')] \right\rangle\!\!\right\rangle.
\end{aligned} \tag{7.155}$$

The change here is straightforward, and involves evaluating the Gaussian probability density for $W(t)$ at c instead of 0.

Then changing the average in Eq. (7.154) to encompass the appropriate Brownian bridges, we have

$$\int_0^\infty dt\, \frac{e^{-c^2/2t}}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\, \Theta[B_{t(0\to c)}(t')-d]\right) \right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk \left(B + \frac{2e^{-ikc}}{2\lambda+k^2}\right), \tag{7.156}$$

in place of Eq. (7.125). Carrying out the following integration gives

$$\begin{aligned}
&\int_0^\infty dt\, \frac{e^{-c^2/2t}}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\} \right\rangle\!\!\right\rangle \\
&\qquad = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk \left[\frac{2e^{-ikc}}{2\lambda+k^2} - \frac{2\sqrt{2}\, s\left(\sqrt{2(\lambda+s)}+ik\right) e^{-\sqrt{2\lambda}\, d + ik(d-c)}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}\right] \\
&\qquad = \left[\sqrt{\frac{\pi}{\lambda}}\, e^{-\sqrt{2\lambda}\,|c|} - \sqrt{\pi}\, \frac{e^{-\sqrt{2\lambda}\,(2d-c)}}{\sqrt{\lambda}} \left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right)\right] \qquad (c\le d,\ d\ge 0)
\end{aligned} \tag{7.157}$$

in place of Eqs. (7.126) in the case $c \le d$. If $c \ge d$ this is instead

$$\int_0^\infty dt\, \frac{e^{-c^2/2t}}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\} \right\rangle\!\!\right\rangle = \sqrt{4\pi}\, \frac{e^{-\sqrt{2\lambda}\, d - \sqrt{2(\lambda+s)}\,(c-d)}}{\sqrt{\lambda+s}+\sqrt{\lambda}} \qquad (c\ge d\ge 0). \tag{7.158}$$

The first expression (7.157) is the same result as in Eq. (7.126), except for a factor $\exp(-c^2/2t)$ on the left-hand side, the factor of $\exp(-\sqrt{2\lambda}\, c)$ in the first term on the right-hand side, and the replacement

$2d \longrightarrow 2d - c$ in the second term on the right-hand side. To invert the Laplace transform, the second term on the right-hand side inverts in the same way with the replacement $2d \longrightarrow 2d - c$, while Eq. (7.83) applies to the first term with $k = \sqrt{2}\,|c|$, such that Eq. (7.129) becomes

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = 1 - e^{c^2/2t}\int_0^t d\tau\,\frac{1}{\tau}\sqrt{\frac{t}{t-\tau}}e^{-(2d-c)^2/2(t-\tau)}e^{-s\tau/2}I_1(s\tau/2) \qquad (c \le d, d \ge 0), \tag{7.159}$$

or that is, the last term on the right-hand side is multiplied by $\exp(c^2/2t)$, and subject to the replacement $2d \longrightarrow 2d - c$. The same modifications (with $t = 1$ for the Brownian bridge we want here) carry through the rest of the treatment in the previous section. In particular, Eq. (7.137) becomes

$$\begin{aligned} I(d) &:= \lim_{x\to 0}\frac{2\sqrt{x}}{\pi}\,e^{c^2/2}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-(2d-c)^2/2(1-\tau)}}{\tau^2} \\ &= e^{c^2/2}e^{-(2d-c)^2/2} \\ &= e^{-2d(d-c)}, \end{aligned} \tag{7.160}$$

which is the correct boundary-crossing probability (6.24) for the same bridge pinned to c. Then we can adapt the probability-density expressions with these modifications, with the result[23]

$$\begin{aligned} f_{T_{\rm s}}(x) &= \left[1 - e^{-2d(d-c)}\right]\delta(x - 0^+) - \partial_x\left[\frac{2\sqrt{x}}{\pi}\,e^{c^2/2}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-(2d-c)^2/2(1-\tau)}}{\tau^2}\right] \\ &= \left[1 - e^{-2d(d-c)}\right]\delta(x - 0^+) \\ &\qquad - \partial_x\left[\frac{4(1-x)}{\pi}\,e^{c^2/2-(2d-c)^2/2(1-x)}\int_0^\infty d\sigma\,\frac{\sigma^2}{(1+\sigma^2)}\,e^{-\sigma^2(2d-c)^2x/2(1-x)}\right] \\ &= \left[1 - e^{-2d(d-c)}\right]\delta(x - 0^+) + (2d-c)\sqrt{\frac{2(1-x)}{\pi x}}\,e^{c^2/2-(2d-c)^2/2(1-x)} \\ &\qquad + \left[1 - (2d-c)^2\right]\,e^{-2d(d-c)}\,\mathrm{erfc}\left(\sqrt{\frac{(2d-c)^2x}{2(1-x)}}\right) \qquad (0 \le x \le 1;\ c \le d;\ d \ge 0). \end{aligned}$$

(probability density for bridge sojourn time) (7.161)

Similarly, the cumulative-probability expressions become

$$\begin{aligned} P(T_{\rm s} \le x) &= 1 - \frac{2\sqrt{x}}{\pi}\,e^{c^2/2}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-(2d-c)^2/2(1-\tau)}}{\tau^2} \qquad (0 \le x \le 1;\ c \le d;\ d \ge 0) \\ &= 1 - \frac{4(1-x)}{\pi}\,e^{c^2/2-(2d-c)^2/2(1-x)}\int_0^\infty d\sigma\,\frac{\sigma^2}{(1+\sigma^2)^2}\,e^{-\sigma^2(2d-c)^2x/2(1-x)} \\ &= 1 + (2d-c)\sqrt{\frac{2x(1-x)}{\pi}}\,e^{c^2/2-(2d-c)^2/2(1-x)} \\ &\qquad - \left[1 - x + (2d-c)^2x\right]\,e^{-2d(d-c)}\,\mathrm{erfc}\left(\sqrt{\frac{(2d-c)^2x}{2(1-x)}}\right), \end{aligned}$$

(cumulative probability distribution for bridge sojourn time) (7.162)

and the moment formula (7.150) becomes

$$\left\langle\!\!\left\langle (T_{\rm s})^n \right\rangle\!\!\right\rangle = \frac{2n\Gamma[n+1/2]}{\sqrt{\pi}(n+1)!}\int_0^1 d\sigma\,(1-\sigma^2)^{n-1}e^{c^2/2-(2d-c)^2/2\sigma^2} \qquad (c \le d;\ d \ge 0)$$

(sojourn-time moments) (7.163)

[23] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.8.

under the same replacements. For example, the explicit expression for the mean is

$$\left\langle\!\left\langle T_{\rm s} \right\rangle\!\right\rangle = \frac{e^{-2d(d-c)}}{2} - \sqrt{\frac{\pi}{8}}\,(2d-c)\,e^{c^2/2}\,\mathrm{erfc}\left[\frac{2d-c}{\sqrt{2}}\right] \qquad (c \le d;\ d \ge 0), \tag{7.164}$$

which reduces to the standard-bridge mean in Eqs. (7.151).

In the case $c \ge d$, the inversion of the Laplace transform (7.158) is somewhat different. First using[24]

$$\mathscr{L}\left[\frac{e^{-st/2}}{2\sqrt{s}\,t} I_{1/2}(st/2)\right](\lambda) = \mathscr{L}\left[\frac{1-e^{-st}}{2s\sqrt{\pi t^3}}\right](\lambda) = \frac{1}{\sqrt{\lambda+s}+\sqrt{\lambda}}, \tag{7.165}$$

where $I_\nu(x)$ is a modified Bessel function, along with[25]

$$\mathscr{L}\left[\frac{k}{2\sqrt{\pi t^3}}\,e^{-k^2/4t}\right](\lambda) = e^{-k\sqrt{\lambda}}, \tag{7.166}$$

these transforms can be combined via the convolution formula (7.84) to give

$$\mathscr{L}\left[\int_0^t d\tau\,\frac{k}{4\pi s\sqrt{\tau^3(t-\tau)^3}}\,e^{-k^2/4(t-\tau)}\left(1-e^{-s\tau}\right)\right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda+s}+\sqrt{\lambda}}. \tag{7.167}$$

Then using the shifted-argument version of Eq. (7.166),

$$\mathscr{L}\left[\frac{k'}{2\sqrt{\pi t^3}}\,e^{-st}\,e^{-k'^2/4t}\right](\lambda) = e^{-k'\sqrt{\lambda+s}}, \tag{7.168}$$

we can again employ convolution to combine this with Eq. (7.167) to give

$$\mathscr{L}\left[\int_0^t d\sigma\int_0^{t-\sigma} d\tau\,\frac{kk'}{8\pi^{3/2}s\sqrt{\tau^3(t-\sigma-\tau)^3\sigma^3}}\,e^{-k^2/4(t-\sigma-\tau)}\,e^{-k'^2/4\sigma}\left(1-e^{-s\tau}\right)e^{-s\sigma}\right](\lambda) = \frac{e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}}}{\sqrt{\lambda+s}+\sqrt{\lambda}}. \tag{7.169}$$

Thus, setting $k = \sqrt{2}\,d$ and $k' = \sqrt{2}\,(c-d)$, we may invert the Laplace transform in Eq. (7.158) to give

$$\left\langle\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\}\right\rangle\!\right\rangle = e^{c^2/2t}\,\frac{d(c-d)\sqrt{t}}{2\pi s}\int_0^t d\sigma\int_0^{t-\sigma} d\tau\,\frac{e^{-d^2/2(t-\sigma-\tau)}\,e^{-(c-d)^2/2\sigma}}{\sqrt{\tau^3(t-\sigma-\tau)^3\sigma^3}}\left(1-e^{-s\tau}\right)e^{-s\sigma} \qquad (c \ge d \ge 0). \tag{7.170}$$

We can simplify this to $t = 1$ for a Brownian bridge that runs over a unit time interval:

$$\left\langle\!\left\langle \exp\left\{-sT_{\rm s}[B_{(0\to c)}]\right\}\right\rangle\!\right\rangle = e^{c^2/2}\,\frac{d(c-d)}{2\pi s}\int_0^1 d\sigma\int_0^{1-\sigma} d\tau\,\frac{e^{-d^2/2(1-\sigma-\tau)}\,e^{-(c-d)^2/2\sigma}}{\sqrt{\tau^3(1-\sigma-\tau)^3\sigma^3}}\left(1-e^{-s\tau}\right)e^{-s\sigma} \qquad (c \ge d \ge 0). \tag{7.171}$$

Now using

$$\frac{e^{-s\sigma}-e^{-s(\sigma+\tau)}}{s} = \int_\sigma^{\sigma+\tau} dx\,e^{-sx} \tag{7.172}$$

to remove the difference in the final factor, while introducing a new integral, we can shift the order of integration according to

$$\int_0^1 d\sigma\int_0^{1-\sigma} d\tau\int_\sigma^{\sigma+\tau} dx = \int_0^1 d\sigma\int_\sigma^1 dx\int_{x-\sigma}^{1-\sigma} d\tau = \int_0^1 dx\int_0^x d\sigma\int_{x-\sigma}^{1-\sigma} d\tau, \tag{7.173}$$

[24] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1024, Eq. (29.3.53).
[25] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.82).

and shifting $\tau \longrightarrow \tau - \sigma$, Eq. (7.171) becomes

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{(0\to c)}]\right\}\right\rangle\!\!\right\rangle = e^{c^2/2}\,\frac{d(c-d)}{2\pi}\int_0^1 dx\,e^{-sx}\int_0^x d\sigma\int_x^1 d\tau\,\frac{e^{-d^2/2(1-\tau)-(c-d)^2/2\sigma}}{\sqrt{(\tau-\sigma)^3(1-\tau)^3\sigma^3}} \qquad (c\geq d\geq 0). \tag{7.174}$$

Now the Laplace transform in s may be inverted to give the probability density:

$$f_{T_{\rm s}}(x) = e^{c^2/2}\,\frac{d(c-d)}{2\pi}\int_0^x d\sigma\int_x^1 d\tau\,\frac{e^{-d^2/2(1-\tau)-(c-d)^2/2\sigma}}{\sqrt{(\tau-\sigma)^3(1-\tau)^3\sigma^3}} \qquad (0\leq x\leq 1; c\geq d\geq 0). \tag{7.175}$$

Changing variables via $\tau \longrightarrow 1-\tau$ gives

$$f_{T_{\rm s}}(x) = e^{c^2/2}\,\frac{d(c-d)}{2\pi}\int_0^x d\sigma\int_0^{1-x} d\tau\,\frac{e^{-d^2/2\tau-(c-d)^2/2\sigma}}{\sqrt{(1-\sigma-\tau)^3\sigma^3\tau^3}} \qquad (0\leq x\leq 1; c\geq d\geq 0), \tag{7.176}$$

and then letting $u = x/\sigma$ and $v = (1-x)/\tau$ leads to the alternate integral expression

$$f_{T_{\rm s}}(x) = e^{c^2/2}\,\frac{d(c-d)}{2\pi\sqrt{x(1-x)}}\int_1^\infty du\int_1^\infty dv\,\frac{uv\,e^{-d^2v/2(1-x)-(c-d)^2u/2x}}{[uv-(1-x)u-xv]^{3/2}} \qquad (0\leq x\leq 1; c\geq d\geq 0), \tag{7.177}$$

from where it is difficult to proceed with the integration.

There is a second approach to inverting the Laplace transforms here that will lead to expression with only a single integral, in both the density and the Laplace transform of the density.[26] As an alternate form of the right-hand side in Eq. (7.169), we can consider

$$\begin{aligned}\frac{e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}}}{\sqrt{\lambda+s}+\sqrt{\lambda}} &= \frac{1}{s}\left(\sqrt{\lambda+s}-\sqrt{\lambda}\right)e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}} \\ &= \frac{1}{s}\,e^{-k\sqrt{\lambda}}\left(\sqrt{\lambda+s}\,e^{-k'\sqrt{\lambda+s}}\right) - \frac{1}{s}\,e^{-k'\sqrt{\lambda+s}}\left(\sqrt{\lambda}\,e^{-k\sqrt{\lambda}}\right).\end{aligned} \tag{7.178}$$

Using again the Laplace transform (7.166) and its shifted version (7.168), along with[27]

$$\mathscr{L}\left[\frac{k^2-2t}{4\sqrt{\pi t^5}}\,e^{-k^2/4t}\right](\lambda) = \sqrt{\lambda}\,e^{-k\sqrt{\lambda}}, \tag{7.179}$$

and the shifted version

$$\mathscr{L}\left[\frac{k'^2-2t}{4\sqrt{\pi t^5}}\,e^{-st}\,e^{-k'^2/4t}\right](\lambda) = \sqrt{\lambda+s}\,e^{-k'\sqrt{\lambda+s}}, \tag{7.180}$$

the convolution theorem allows us to combine these transforms to form the right-hand side of Eq. (7.178), with the result

$$\mathscr{L}\left[\int_0^t d\tau\,\frac{k(t-\tau)(k'^2-2\tau)-k'\tau[k^2-2(t-\tau)]}{8\pi s\sqrt{\tau^5(t-\tau)^5}}\,e^{-k^2/4(t-\tau)}\,e^{-k'^2/4\tau}\,e^{-s\tau}\right](\lambda) = \frac{e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}}}{\sqrt{\lambda+s}+\sqrt{\lambda}}. \tag{7.181}$$

Then again setting $k = \sqrt{2}\,d$ and $k' = \sqrt{2}\,(c-d)$, we may invert the Laplace transform in Eq. (7.158) to give

$$\begin{aligned}\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = e^{c^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,&\frac{d(t-\tau)[(c-d)^2-\tau]-(c-d)\tau[d^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}} \\ &\times\; e^{-d^2/2(t-\tau)}\,e^{-(c-d)^2/2\tau}\,e^{-s\tau} \qquad (c\geq d\geq 0),\end{aligned} \tag{7.182}$$

[26] Vadim Linetsky, "Step Options," *Mathematical Finance* **9**, 55 (2001) (doi: 10.1111/1467-9965.00063). See in particular Eq. (C.9) and the discussion just before Eq. (A.9).

[27] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.87) for $n = 2$, which $H_2(x) = 2(2x^2-1)$.

or for $t = 1$,

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = \frac{e^{c^2/2}}{\sqrt{2\pi}s}\int_0^1 d\tau\,\frac{d(1-\tau)[(c-d)^2-\tau]-(c-d)\tau[d^2-(1-\tau)]}{\sqrt{\tau^5(1-\tau)^5}} \times e^{-d^2/2(1-\tau)}\,e^{-(c-d)^2/2\tau}\,e^{-s\tau} \qquad (c\ge d\ge 0). \tag{7.183}$$

Then writing the s-dependence as

$$\frac{e^{-s\tau}}{s} = \int_\tau^\infty dx\,e^{-sx} = \int_0^\infty dx\,e^{-sx} - \int_0^\tau dx\,e^{-sx}, \tag{7.184}$$

and then massaging the two resulting integrals (with the rest of the integrand suppressed for brevity) via

$$\int_0^1 d\tau\int_0^\infty dx\,e^{-sx} - \int_0^1 d\tau\int_0^\tau dx\,e^{-sx} = \int_0^\infty dx\,e^{-sx}\int_0^1 d\tau - \int_0^1 dx\,e^{-sx}\int_x^1 d\tau = \int_0^1 dx\,e^{-sx}\int_0^x d\tau, \tag{7.185}$$

where we have made use of the fact that the sojourn-time density has its support on $[0,1]$. Thus we can now invert the Laplace transform (7.183) of the sojourn-time density to obtain

$$f_{T_{\rm s}}(x) = \frac{e^{c^2/2}}{\sqrt{2\pi}}\int_0^x d\tau\,\frac{d(1-\tau)[(c-d)^2-\tau]-(c-d)\tau[d^2-(1-\tau)]}{\sqrt{\tau^5(1-\tau)^5}}\,e^{-d^2/2(1-\tau)-(c-d)^2/2\tau} \qquad (0\le x\le 1;\ c\ge d\ge 0). \tag{7.186}$$

This integral can be performed analytically, with result[28]

$$f_{T_{\rm s}}(x) = \sqrt{\frac{2}{\pi}}\,\frac{(c-d)x+d(1-x)}{\sqrt{x(1-x)}}\,\exp\left(\frac{c^2}{2}-\frac{(c-d)^2}{2x}-\frac{d^2}{2(1-x)}\right) \qquad (0\le x\le 1;\ c\ge d\ge 0)$$
$$+\left[1-(2d-c)^2\right]e^{-2d(d-c)}\,\mathrm{erfc}\left(\frac{(c-d)(1-x)+d\,x}{\sqrt{2x(1-x)}}\right), \tag{7.187}$$

(probability density for bridge sojourn time)

as can be verified by differentiating this expression to obtain the integrand of Eq. (7.186). This then gives the density for the boundary-crossing case of the bridge sojourn time.

The cases with $d < 0$ can be generated from the cases with $d \ge 0$ by replacing $x \longrightarrow 1-x$. And again, all of these expressions can be generalized to a bridge $B_{t(0\to c)}(t)$ that is pinned to c at time T via the replacements $c \longrightarrow c/\sqrt{t}$, $d \longrightarrow d/\sqrt{t}$, and $x \longrightarrow x/t$. Overall factors of t must be restored to make the density come out with "units" of $1/t$ [including, e.g., an overall factor of $\sqrt{t}$ that came from the path-pinning factor $e^{c^2/2}$ in Eq. (7.156)]. Also, a bridge starting at a different location a than 0 can be obtained by shifting c and d by $-a$. Writing these out explicitly from Eqs. (7.161) and (7.187),

$$f_{T_{\rm s}}(x) = \left[1-e^{-2(d-a)(d-c)/t}\right]\delta(x-0^+) + (2d-a-c)\sqrt{\frac{2(t-x)}{\pi t^3 x}}\,e^{(c-a)^2/2t-(2d-a-c)^2/2(t-x)}$$
$$+\frac{1}{t}\left[1-\frac{(2d-a-c)^2}{t}\right]e^{-2(d-a)(d-c)/t}\,\mathrm{erfc}\left(\sqrt{\frac{(2d-a-c)^2x}{2t(t-x)}}\right) \qquad (0\le x\le t;\ a\le d;\ c\le d)$$
$$f_{T_{\rm s}}(x) = \sqrt{\frac{2}{\pi}}\,\frac{(c-d)x+(d-a)(t-x)}{\sqrt{t^3x(t-x)}}\,e^{(c-a)^2/2t-(c-d)^2/2x-(d-a)^2/2(t-x)} \qquad (0\le x\le t;\ a\le d\le c)$$
$$+\frac{1}{t}\left[1-\frac{(2d-a-c)^2}{t}\right]e^{-2(d-a)(d-c)/t}\,\mathrm{erfc}\left(\frac{(c-d)(t-x)+(d-a)\,x}{\sqrt{2tx(t-x)}}\right), \tag{7.188}$$

(probability density for bridge sojourn time)

[28] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.8.

while to obtain the other cases, we can change the signs of a, c, and d, while replacing x with $t-x$,

$$
\begin{aligned}
f_{T_{\rm s}}(x) &= \left[1-e^{-2(a-d)(c-d)/t}\right]\delta(t-x-0^+) + (a+c-2d)\sqrt{\frac{2x}{\pi t^3(t-x)}}\,e^{(a-c)^2/2t-(a+c-2d)^2/2x} \\
&\quad + \frac{1}{t}\left[1-\frac{(a+c-2d)^2}{t}\right]e^{-2(a-d)(c-d)/t}\,\mathrm{erfc}\left(\sqrt{\frac{(a+c-2d)^2(t-x)}{2tx}}\right) \quad (0\le x\le t;\ d\le c;\ d\le a) \\
f_{T_{\rm s}}(x) &= \sqrt{\frac{2}{\pi}}\,\frac{(a-d)x+(d-c)(t-x)}{\sqrt{t^3x(t-x)}}\,e^{(a-c)^2/2t-(a-d)^2/2x-(d-c)^2/2(t-x)} \qquad (0\le x\le t;\ c\le d\le a) \\
&\quad + \frac{1}{t}\left[1-\frac{(a+c-2d)^2}{t}\right]e^{-2(a-d)(c-d)/t}\,\mathrm{erfc}\left(\frac{(a-d)(t-x)+(d-c)\,x}{\sqrt{2tx(t-x)}}\right),
\end{aligned}
$$

(probability density for bridge sojourn time, reflected cases) (7.189)

which takes advantage of the fact that the problem is reflection symmetric if the occupation time is taken to be the non-occupation time.

It is also useful to write out the moment-generating functions corresponding to these probability densities. We have the explicit expression for ($c\ge d\ge 0$) in Eq. (7.182), but we bypassed the solution for ($c\le d, d\ge 0$) by generalizing the treatment from the standard Brownian bridge. In fact, it is best to replace that treatment with one that avoids obtaining a Bessel function, as we will do here. First, to set up the Laplace transforms, we will need to arrive at something of the form

$$
\begin{aligned}
\frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right) &= \frac{e^{-k\sqrt{\lambda}}}{s\sqrt{\lambda}}\left(2\lambda+s-2\sqrt{\lambda(\lambda+s)}\right) \\
&= \frac{2}{s}\left(\sqrt{\lambda}-\sqrt{\lambda+s}\right)e^{-k\sqrt{\lambda}} + \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}}.
\end{aligned}
\tag{7.190}
$$

Then using the Laplace-transform formula[29]

$$
\mathscr{L}\left[\frac{1}{2\sqrt{\pi t^3}}\left(e^{-st}-1\right)\right](\lambda) = \sqrt{\lambda}-\sqrt{\lambda+s}, \tag{7.191}
$$

we can combine this with the transform formula (7.166) using the convolution theorem to obtain the formula

$$
\mathscr{L}\left[\int_0^t d\tau\,\frac{k}{4\pi\sqrt{\tau^3(t-\tau)^3}}\,e^{-k^2/4\tau}\left(e^{-s(t-\tau)}-1\right)\right](\lambda) = \left(\sqrt{\lambda}-\sqrt{\lambda+s}\right)e^{-k\sqrt{\lambda}}. \tag{7.192}
$$

Then adding this to the transform formula (7.83) to hit the last term on the right-hand side of Eq. (7.190),

$$
\mathscr{L}\left[\frac{1}{\sqrt{\pi t}}\,e^{-k^2/4t} + \int_0^t d\tau\,\frac{k}{2\pi s\sqrt{\tau^3(t-\tau)^3}}\,e^{-k^2/4\tau}\left(e^{-s(t-\tau)}-1\right)\right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right). \tag{7.193}
$$

Changing k to k' and combining this again with Eq. (7.83) gives

$$
\begin{aligned}
&\mathscr{L}\left[\frac{1}{\sqrt{\pi t}}\left(e^{-k^2/4t}-e^{-k'^2/4t}\right) - \int_0^t d\tau\,\frac{k'}{2\pi s\sqrt{\tau^3(t-\tau)^3}}\,e^{-k'^2/4\tau}\left(e^{-s(t-\tau)}-1\right)\right](\lambda) \\
&\qquad = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}} - \frac{e^{-k'\sqrt{\lambda}}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right).
\end{aligned}
\tag{7.194}
$$

[29] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1023, Eq. (29.3.36) for $n=2$, which $H_2(x) = 2(2x^2-1)$.

Then using this to invert Eq. (7.157) with $k = \sqrt{2}\,|c|$ and $k' = \sqrt{2}(2d-c)$, we find[30]

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = 1 - e^{-2d(d-c)/t} \qquad (c \le d,\ d \ge 0)$$
$$+ e^{c^2/2t}\frac{\sqrt{t}(2d-c)}{\sqrt{2\pi}\,s}\int_0^t d\tau\,\frac{1}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(2d-c)^2/2\tau}\left(1-e^{-s(t-\tau)}\right), \tag{7.195}$$

The non-integral part of the expression here is related to the boundary-touching probability of the bridge. Note that the normalization can be verified by taking the limit $s \longrightarrow 0$ and carrying out the resulting integral, recovering the unit-normalization result. Now putting this together with Eq. (7.182) and putting in a starting point of a,

$$\left\langle\!\!\left\langle e^{-sT_{\rm s}}\right\rangle\!\!\right\rangle = 1 - e^{-2(d-a)(d-c)/t} \qquad (a \le d;\ c \le d)$$
$$+ e^{(c-a)^2/2t}\frac{\sqrt{t}(2d-a-c)}{\sqrt{2\pi}\,s}\int_0^t d\tau\,\frac{1}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(2d-a-c)^2/2\tau}\left(1-e^{-s(t-\tau)}\right)$$
$$\left\langle\!\!\left\langle e^{-sT_{\rm s}}\right\rangle\!\!\right\rangle = e^{(c-a)^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,\frac{(d-a)(t-\tau)[(c-d)^2-\tau]-(c-d)\tau[(d-a)^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}}$$
$$\times\ e^{-(d-a)^2/2(t-\tau)-(c-d)^2/2\tau-s\tau} \qquad (a \le d \le c)$$
$$\left\langle\!\!\left\langle e^{-sT_{\rm s}}\right\rangle\!\!\right\rangle = \left[1-e^{-2(a-d)(c-d)/t}\right]e^{-st} \qquad (d \le c;\ d \le a)$$
$$+ e^{(a-c)^2/2t}\frac{\sqrt{t}(a+c-2d)}{\sqrt{2\pi}\,s}\int_0^t d\tau\,\frac{1}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(a+c-2d)^2/2\tau}\left(e^{-s\tau}-e^{-st}\right)$$
$$\left\langle\!\!\left\langle e^{-sT_{\rm s}}\right\rangle\!\!\right\rangle = e^{(a-c)^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,\frac{(d-c)(t-\tau)[(a-d)^2-\tau]-(a-d)\tau[(d-c)^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}}$$
$$\times\ e^{-(d-c)^2/2(t-\tau)-(a-d)^2/2\tau-s\tau} \qquad (c \le d \le a).$$

(sojourn-time moment generating function) (7.196)

The last two cases here follow from the first two via the symmetry of the problem. Specifically, ''flipping'' the geometry by reversing the signs of a, c, and d allows us to calculate the ''complimentary'' generating function $\langle\!\langle e^{-s(t-T_{\rm s})}\rangle\!\rangle$. Thus in addition to reversing these signs, we also need to reverse the sign of s in the first two expressions and then multiply through by e^{-st} to obtain the desired generating function. In the fourth expression, we also changed integration variables, letting $\tau \longrightarrow t-\tau$, to bring it into a form more similar to the second expression. In this case it turns out to have the same form, but with a and c interchanged.

The singularity of the integrand in Eqs. (7.196) can be problematic, and merits some further discussion. For example, in the case of the second expression with $d = a$, the integral naïvely becomes

$$\left\langle\!\!\left\langle e^{-sT_{\rm s}}\right\rangle\!\!\right\rangle = e^{(c-a)^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,\frac{(c-a)}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(c-a)^2/2\tau-s\tau} \qquad (a \le c;\ d = a). \tag{7.197}$$

However, this integral is ill-defined because the integrand diverges at $t = \tau$ (note that it is okay at $\tau = 0$ provided $c \neq a$). Similarly, the expression is problematic if instead $d = c$, because the integrand diverges at $\tau = 0$. Thus, some cases above may require some caution in their evaluation. One solution to regularize this example[31] is to multiply through by s in Eq. (7.182), and then take the limit $s \longrightarrow 0$ to obtain

$$e^{c^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}}\int_0^t d\tau\,\frac{d(t-\tau)[(c-d)^2-\tau]-(c-d)\tau[d^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}}\,e^{-d^2/2(t-\tau)}\,e^{-(c-d)^2/2\tau} = 0 \qquad (c \ge d \ge 0). \tag{7.198}$$

[30] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.7.

[31] Vadim Linetsky, *op. cit.*, in the discussion after Eq. (C.9).

[That the result vanishes may be somewhat more obvious from performing the same maneuver in Eq. (7.181).] After restoring the initial point a, this is a useful counterpart to the second and fourth expressions in Eqs. (7.196). For example, multiplying Eq. (7.198) by e^{-st}/s and subtracting from the second equation in Eqs. (7.196) leads to a similar expression, but with the replacement $e^{-s\tau} \longrightarrow e^{-s\tau} - e^{-st}$. This cures the problematic case (7.197), which becomes

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle = \frac{\sqrt{t}}{\sqrt{2\pi}s} \int_0^t d\tau \, \frac{(c-a)}{\sqrt{\tau^3(t-\tau)^3}} \, e^{-(c-a)^2(\tau^{-1}-t^{-1})/2} \left[e^{-s\tau} - e^{-st} \right] \qquad (a \le c;\ d = a) \tag{7.199}$$

after this regularization; this cures the divergence at $t = \tau$, which is now cut off by the difference in exponentials. Note that this is the same expression that follows from the third expression in Eqs. (7.196). To cure the divergence in the case $d = c$, it is better to multiply Eq. (7.198) by $1/s$ and subtract from the second equation in Eqs. (7.196); this again leads to a similar expression, but with the replacement $e^{-s\tau} \longrightarrow e^{-s\tau} - 1$, cutting off the divergence at $\tau = 0$. However, the integrals are still pathological when $a = d = c$, in which case it is easiest to setting $d = 0$ in Eq. (7.129), in which case the result is

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle = \frac{1 - e^{-st}}{st} \qquad (a = c = d). \tag{7.200}$$

To see how this limit follows directly from Eq. (7.199), note that as $c \longrightarrow a$, the factor of $(c-a)$ makes the integrand small everywhere except for the range of small τ—the divergent factor $\tau^{-3/2}$ is important here, and although it is cut off by the Gaussian factor $e^{-(c-a)^2/2\tau}$, it is still important for $\tau \sim \sqrt{c-a}$. Thus, since the significant part of the integrand moves towards vanishingly small τ, we can expand the integrand in τ to obtain

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle = \frac{1}{\sqrt{2\pi}st} \int_0^t d\tau \left[\frac{(c-a)}{\tau^{3/2}} \, e^{-(c-a)^2(\tau^{-1}-t^{-1})/2} \big(1 - e^{-st}\big) + O\big(\tau^{-1/2}\big) \right] \qquad (a \le c;\ d = a). \tag{7.201}$$

This integral may then be carried out, with result

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle \approx e^{(c-a)^2/2t} \left(\frac{1 - e^{-st}}{st} \right) \operatorname{erfc}\left[\frac{c-a}{\sqrt{2t}} \right] \qquad (a \le c;\ d = a). \tag{7.202}$$

At this point, the limit $c \longrightarrow a$ yields Eq. (7.200).

In general, the moments must then be computed by integrating the probability density or differentiating the moment-generating functions listed above. Although we wrote out an integral for $c \le d$ and $d \ge 0$ in Eq. (7.163), the case for $0 \le d \le c$ is not as simple. For the mean sojourn time, however, it is possible to work out a relatively nice expression by integrating the local time. The result is given in Eq. (7.287) as

$$\begin{aligned} \left\langle\!\left\langle T_s[B_{a\to c}(t); d] \right\rangle\!\right\rangle = \frac{t}{2} + \operatorname{sgn}(2d - a - c) \, \frac{t}{2} \left[e^{-2[(d-a)(d-c)\,\Theta(d-a)\,\Theta(d-c) + (a-d)(c-d)\,\Theta(a-d)\,\Theta(c-d)]/t} - 1 \right] \\ - \sqrt{\frac{\pi t}{8}} \, (2d - a - c) \, e^{(c-a)^2/2t} \operatorname{erfc}\left(\frac{|d-a| + |d-c|}{\sqrt{2t}} \right), \end{aligned} \tag{7.203}$$

(mean sojourn time)

after restoring the initial point a of the Brownian bridge. While it may not be obvious from this expression, when d is in the interval (a, c), the second term does not contribute, and the last term leads to a decreasing, straight-line dependence on d.

7.2.4 Path-Pinning Normalization as a Constrained Integration

In deriving sojourn-time statistics, we made use of the relation (7.155) to remove the delta function that pinned the endpoint of a Wiener path such that $W(T) = c$:

$$\left\langle\!\left\langle \delta[W(T) - c] \, F[W(t)] \right\rangle\!\right\rangle = \frac{e^{-c^2/2T}}{\sqrt{2\pi T}} \left\langle\!\left\langle F\big[B_{T(0\to c)}(t)\big] \right\rangle\!\right\rangle. \tag{7.204}$$

We also used this in restricted form in Eqs. (7.124). In both cases we used probabilistic arguments to justify this relation. However, it is also useful to see how this arises as a constrained integration problem, by considering the explicit probability measure of the path. To begin, suppose that we consider N discrete steps $(\Delta W_0, \ldots, \Delta W_{N-1})$ of the Wiener path, each of time step $\Delta T = T/N$, and write out the explicit integration over the multidimensional Gaussian probability density:

$$\begin{aligned}\left\langle\!\!\left\langle \delta[W(T)-c]\, F[W(t)] \right\rangle\!\!\right\rangle \\ = \int d\Delta W_0 \ldots d\Delta W_{N-1}\, \frac{1}{(2\pi\Delta T)^{N/2}}\, \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-1}\left(\Delta W_j\right)^2\right]\, \delta\left[\sum_{j=0}^{N-1}\Delta W_j - c\right]\, F[W(t)].\end{aligned} \tag{7.205}$$

Note that the path functional $F[W(t)]$ is still written in continuous notation, since its details are not important to this calculation. Now we will define shifted increments that move linearly towards the ''target'' c,

$$\Delta B_j := \Delta W_j + \frac{c}{N}, \tag{7.206}$$

which will become the increments of the Brownian bridge. Changing variables, the integral becomes

$$\begin{aligned}\left\langle\!\!\left\langle \delta[W(T)-c]\, F[W(t)] \right\rangle\!\!\right\rangle \\ = \int d\Delta B_0 \ldots d\Delta B_{N-1}\, \frac{1}{(2\pi\Delta T)^{N/2}}\, \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-1}\left(\Delta B_j - \frac{c}{N}\right)^2\right]\, \delta\left[\sum_{j=0}^{N-1}\Delta B_j\right]\, F\big[B_{T(0\to c)}(t)\big] \\ = \int d\Delta B_0 \ldots d\Delta B_{N-1}\, \frac{e^{-c^2/2T}}{(2\pi\Delta T)^{N/2}}\, \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-1}\left(\Delta B_j\right)^2 + \frac{c}{T}\sum_{j=0}^{N-1}\Delta B_j\right] \\ \times \delta\left[\sum_{j=0}^{N-1}\Delta B_j\right]\, F\big[B_{T(0\to c)}(t)\big].\end{aligned} \tag{7.207}$$

Now carrying out the integration over ΔB_{N-1} will remove the delta function and enforce the replacement

$$\Delta B_{N-1} = -\sum_{j=0}^{N-2}\Delta B_j. \tag{7.208}$$

However, some care must be taken in the resulting integral. The delta-function integration formula (2.64) reads

$$\int d^d q\, \delta[h(\mathbf{q})]\, f(\mathbf{q}) = \oint_{h^{-1}(0)} \frac{f(\mathbf{q})}{|\nabla h|}\, dS, \tag{7.209}$$

where the reduced integral involves the Euclidean norm $|\nabla h|$ of the gradient of the constraint function $h(\mathbf{q})$. In Eq. (7.207), the constraint function is a simple sum over all the ΔB_j, and the derivatives are taken with respect to each ΔB_j (i.e., each of the N derivatives in the gradient has unit magnitude). The corresponding norm $|\nabla h|$ is then simply $\sqrt{N}$. Thus, the result of carrying out the ΔB_{N-1} integral in Eq. (7.207) gives the

desired result:

$$\begin{aligned}
&\left\langle\!\!\left\langle \delta[W(T)-c]\,F[W(t)] \right\rangle\!\!\right\rangle \\
&\quad= \int d\Delta B_0 \dots d\Delta B_{N-2} \frac{e^{-c^2/2T}}{\sqrt{N}(2\pi\Delta T)^{N/2}} \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-2}\left(\Delta B_j\right)^2 - \frac{1}{2\Delta T}\left(\sum_{j=0}^{N-2}\Delta B_j\right)^2\right] \\
&\qquad\qquad \times F\big[B_{T(0\to c)}(t)\big] \\
&\quad= \frac{e^{-c^2/2T}}{\sqrt{2\pi T}} \int d\Delta B_0 \dots d\Delta B_{N-2} \frac{1}{(2\pi\Delta T)^{(N-1)/2}} \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-2}\left(\Delta B_j\right)^2 - \frac{1}{2\Delta T}\left(B_{N-1}\right)^2\right] \\
&\qquad\qquad \times F\big[B_{T(0\to c)}(t)\big]. \\
&\quad= \frac{e^{-c^2/2T}}{\sqrt{2\pi T}} \left\langle\!\!\left\langle F\big[B_{T(0\to c)}(t)\big] \right\rangle\!\!\right\rangle.
\end{aligned} \tag{7.210}$$

In the second-to-last expression here, there are $N-1$ steps of variance ΔT, in addition to a constraint that the variance of B_{N-1} is ΔT (B_{N-1} is close to zero rather than c because we defined the B_j to be ''drifting'' towards c; thus, $B_N = 0$ in these coordinates). Again, the coefficient on the final path average is the probability density for $W(T)$ evaluated at $W(T) = c$, but in this calculation part of this factor came from the path measure, and part came from the integration over a delta function.

7.3 Local Time

We will define the **local time** of a stochastic process $y(t)$ at displacement d as

$$\ell[y(t);d] := \int_0^t dt'\,\delta[y(t')-d]. \tag{7.211}$$
(local time)

The integrand here only ''activates'' when $y(t)$ passes through d, and the local time is a measure of how much time $y(t)$ spends at the displacement d, but normalized so that the answer is not merely zero. Recalling that we defined the sojourn time for $y(t)$ as

$$T_{\mathrm{s}}[y(t);d] := \int_0^t dt'\,\Theta[y(t')-d], \tag{7.212}$$

we can immediately deduce that

$$\ell[y(t);d] = -\partial_d T_{\mathrm{s}}[y(t);d], \tag{7.213}$$
(local time)

so that we may simply adapt our sojourn-time results to obtain local-time statistics.

One useful aspect of the local time arises in calculating functionals of stochastic processes of the form

$$\int_0^t dt'\,F[y(t')] = \int_0^t dt' \int da\,F(a)\,\delta[y(t')-a] = \int da\,F(a)\int_0^t dt'\,\delta[y(t')-a]. \tag{7.214}$$

The definition of the local time then implies

$$\int_0^t dt'\,F[y(t')] = \int_{-\infty}^{\infty} da\,F(a)\,\ell[y(t);a], \tag{7.215}$$
(local-time density formula)

so that the local time acts as an occupation density for $y(t)$. The local time is also commonly thought of as a time-dependent process, here through the time dependence of the process itself. Intuitively, the local time ''accumulates'' as the stochastic process continues, so $\ell[y(t);a]$ is a nondecreasing function of time.

As an alternate representation of the local time, recall the property of the δ function

$$\delta[f(x)] = \sum_{x_0 \in f^{-1}(0)} \frac{\delta(x - x_0)}{|f'(x_0)|}, \tag{7.216}$$

where the sum is over all real roots x_0 of $f(x)$. Then the definition (7.211) becomes

$$\ell[y(t); d] = \int_0^t dt' \sum_{\{t_d : y(t_d) = d\}} \frac{\delta(t' - t_d)}{|y'(t_d)|}, \tag{7.217}$$

or carrying out the integral,

$$\ell[y(t); d] = \sum_{\{t_d : y(t_d) = d\}} \frac{1}{|y'(t_d)|}. \tag{7.218}$$

(intersection representation of local time)

Thus, the local time is given by summing over the intersections of the process $y(t)$ with the boundary at d, where at each intersection the contribution is the reciprocal of the ''speed'' $|y'|$ during the intersection. Intuitively, this makes sense, since the greater the speed, the less the time spent at the level d during the intersection. To get a better feeling for this, consider the case of a Wiener process, $y(t) = W(t)$, in discrete form:

$$\ell[W(t); d] = \lim_{\Delta t \to 0} \sum_{\{j : (W_{j+1} - d)(W_j - d) < 0\}} \frac{\Delta t}{|\Delta W_j|}. \tag{7.219}$$

Note that as $\Delta t \longrightarrow 0$, the contribution from each intersection in the sum decreases as $\Delta t^{1/2}$. As we will show, the local time can converge to a nonzero value; evidently, this means that the smaller step size ''reveals'' extra intersections in the neighborhood of each intersection to compensate for this decrease.

It is also interesting to consider possible generalizations of the local time. For example, consider the functional

$$\ell'[y(t); d] := \int_0^t dt'\, \delta'[y(t') - d], \tag{7.220}$$

which we can see is related to the local time via a $-\partial_d$ derivative, as the local time is related to the sojourn time. Using the composition rule (Problem 7.1)

$$\delta'[f(x)] = \sum_{x_0 \in f^{-1}(0)} \left[\frac{\delta'(x - x_0)}{f'(x_0)|f'(x_0)|} + \frac{f''(x_0)\, \delta(x - x_0)}{|f'(x_0)|^3} \right], \tag{7.221}$$

the local-time derivative becomes

$$\ell'[y(t); d] = \int_0^t dt' \sum_{\{t_d : y(t_d) = d\}} \left[\frac{\delta'(t' - t_d)}{y'(t_d)|y'(t_d)|} + \frac{y''(t_d)\, \delta(t' - t_d)}{|y'(t_d)|^3} \right]. \tag{7.222}$$

The first term under the integral vanishes so long as the velocity in the denominator does not vanish (something that occurs with zero probability, and which we also ignored in the local-time analysis), and we have

$$\ell'[y(t); d] = \sum_{\{t_d : y(t_d) = d\}} \frac{y''(t_d)}{|y'(t_d)|^3}. \tag{7.223}$$

(local-time derivative)

For a Wiener process, $y(t) = W(t)$, in discrete time increments Δt we can count $y'(t)$ as $O(\Delta t^{-1/2})$ and $y''(t)$ as $O(\Delta t^{-3/2})$, so the summand here is of order unity. However, recall from our discussion of local-time intersections in Eq. (7.218) that the number of intersection times t_d grows as $\Delta t^{-1/2}$. Thus, while we can assign an ensemble average to this statistic,

$$\left\langle\!\!\left\langle \ell'[y(t); d] \right\rangle\!\!\right\rangle = -\partial_d \left\langle\!\!\left\langle \ell[y(t); d] \right\rangle\!\!\right\rangle = \partial_d^2 \left\langle\!\!\left\langle T_{\mathrm{s}}[y(t); d] \right\rangle\!\!\right\rangle, \tag{7.224}$$

evidently the *variance* of the derivative statistic is arbitrarily large. (See Problem 7.5.)

7.3.1 Wiener Process

To compute the probability distribution for the local time of the Wiener process, we will follow closely the procedure of Section 7.2.1 for the sojourn time of the Wiener process. Let $f_\ell(x)$ denote the probability density for the local time $\ell[W(t);d]$ of $W(t)$ at d, with cumulative probability $P(\ell \le x)$, satisfying

$$f_\ell(x) = \partial_x P(\ell \le x). \tag{7.225}$$

Then the Laplace transform of $f_\ell(x)$ is

$$\int_0^\infty dx\, e^{-sx} f_\ell(x) = \left\langle\!\!\left\langle \exp\{-s\ell[W(t);d]\} \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \exp\left[-s\int_0^t dt'\,\delta[W(t')-d]\right] \right\rangle\!\!\right\rangle. \tag{7.226}$$

Note that $f_\ell(x)$ is not limited in domain to $x < t$ as was the case for the sojourn time, but the domain *is* limited to $x > 0$. Consider then the driven diffusion equation

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x)f - \lambda f + g(x), \tag{7.227}$$

where $V(x)$ is the occupation function, which here is a delta function:

$$V(x) = s\delta(x-d). \tag{7.228}$$

We will also take $g(x) = 1$. The steady-state solution is given by the Feynman–Kac formula (7.3) as

$$\begin{aligned} f(x) &= \left\langle\!\!\left\langle \int_0^\infty dt\, \exp\left(-\lambda t - \int_0^t dt'\, V[x+W(t')]\right) \right\rangle\!\!\right\rangle \\ &= \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\delta[x+W(t')-d]\right) \right\rangle\!\!\right\rangle. \end{aligned} \tag{7.229}$$

This is then the solution of the steady-state version of Eq. (7.227):

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\delta(x-d)f(x) + 1. \tag{7.230}$$

For $x \neq d$, the ODE is

$$f'' = 2\lambda f - 2. \tag{7.231}$$

Setting $h = f - 1/\lambda$, we have $h'' = 2\lambda h$, so that for $x \neq d$,

$$h(x) \propto e^{\pm\sqrt{2\lambda}\,x}, \tag{7.232}$$

or choosing the bounded solutions,

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} + \dfrac{1}{\lambda} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{1}{\lambda} & (x < d) \end{cases} \tag{7.233}$$

for some undetermined constants A and B. Demanding continuity of $f(x)$ at $x = d$ gives

$$B = Ae^{-2\sqrt{2\lambda}\,d}. \tag{7.234}$$

The δ function in the ODE says that the derivative $f'(x)$ should jump by $2sf(d)$ at $x = d$, so that

$$-\sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d} - 2s\left(Ae^{-\sqrt{2\lambda}\,d} + \frac{1}{\lambda}\right) = \sqrt{2\lambda}Be^{\sqrt{2\lambda}\,d} = \sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d}, \tag{7.235}$$

or that is, $f'(d-0^+) = f'(d+0^+) - 2sf(d+0^+)$. The solution of these two equations fixes the coefficients as

$$A = -\frac{se^{\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}, \qquad B = -\frac{se^{-\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}. \tag{7.236}$$

Now, we can equate Eqs. (7.229) and (7.233) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\delta[W(t')-d]\right)\right\rangle\!\!\right\rangle = B + \frac{1}{\lambda} = \frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}, \tag{7.237}$$

where we have assumed $d > 0$. Then using Eq. (7.226) on the left-hand side,

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\left\{-s\ell[W(t);d]\right\}\right\rangle\!\!\right\rangle = \frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}. \tag{7.238}$$

Now using the Laplace-transform formulae

$$\mathscr{L}\left[1\right](\lambda) = \frac{1}{\lambda} \tag{7.239}$$

and[32]

$$\mathscr{L}\left[-e^{ak}e^{a^2t}\operatorname{erfc}\left(a\sqrt{t}+\frac{k}{2\sqrt{t}}\right)+\operatorname{erfc}\left(\frac{k}{2\sqrt{t}}\right)\right](\lambda) = \frac{ae^{-k\sqrt{\lambda}}}{\lambda\left(\sqrt{\lambda}+a\right)}, \tag{7.240}$$

which becomes with $k = d\sqrt{2}$, $a = s/\sqrt{2}$,

$$\mathscr{L}\left[-e^{sd}e^{s^2t/2}\operatorname{erfc}\left(s\sqrt{t/2}+\frac{d}{\sqrt{2t}}\right)+\operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right)\right](\lambda) = \frac{se^{-d\sqrt{2\lambda}}}{\lambda\left(\sqrt{2\lambda}+s\right)}. \tag{7.241}$$

Thus, Eq. (7.238) becomes

$$\int_0^\infty dx\, e^{-sx}\, f_\ell(x) = 1 - \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) + e^{sd}e^{s^2t/2}\operatorname{erfc}\left(s\sqrt{t/2}+\frac{d}{\sqrt{2t}}\right) \tag{7.242}$$

after using Eq. (7.226) to replace the ensemble average on the left-hand side. Now using

$$\int_0^\infty dx\, e^{-sx}\frac{1}{\sqrt{2\pi t}}e^{-(x+d)^2/2t} = \frac{1}{2}\,e^{sd}\,e^{s^2t/2}\operatorname{erfc}\left[\sqrt{\frac{t}{2}}\left(s+\frac{d}{t}\right)\right], \tag{7.243}$$

we can invert the Laplace transforms on both sides, with the result[33]

$$f_\ell(x) = \operatorname{erf}\left(\frac{|d|}{\sqrt{2t}}\right)\delta(x-0^+) + \sqrt{\frac{2}{\pi t}}\,e^{-(x+|d|)^2/2t}, \tag{7.244}$$

(local-time probability density)

which is normalized in view of

$$\sqrt{\frac{2}{\pi t}}\,e^{-(x+d)^2/2t} = \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) = 1 - \operatorname{erf}\left(\frac{d}{\sqrt{2t}}\right). \tag{7.245}$$

[32] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1027, Eq. (29.3.89).

[33] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 155, formula 1.3.4; A. N. Borodin, "Brownian local time," *Russian Mathematical Surveys* **44**, 1 (1989), p. 5 (doi: 10.1070/RM1989v044n02ABEH002050); Jim Pitman, "The distribution of local times of a Brownian bridge," in *Séminaire de Probabilités XXXIII*, Jacques Azéma, Michel Émery, Michel Ledoux, and Marc Yor, Eds. (Springer, 1999), p. 388, Eq. (1) (doi: 10.1007/BFb0096528).

Again, the δ-function term here gives the boundary-noncrossing probability, with the crossing probability (6.9) appearing, which is the same as the probability to have zero local time at d:

$$P(\ell = 0) = 1 - P_{\text{cross}}(d,t) = 1 - \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) = \operatorname{erf}\left(\frac{d}{\sqrt{2t}}\right). \tag{7.246}$$

We have also inserted absolute-value symbols for d: while we assumed $d > 0$ for this derivation, the result should be exactly symmetric in d owing to the same symmetry in $W(t)$. The cumulative probability is again given by integrating 0 to x:[34]

$$P(\ell \le x) = \int_0^x dx'\, f_\ell(x') = \operatorname{erf}\left(\frac{|d|}{\sqrt{2t}}\right) + \left[\operatorname{erf}\left(\frac{x+|d|}{\sqrt{2t}}\right) - \operatorname{erf}\left(\frac{|d|}{\sqrt{2t}}\right)\right], \tag{7.247}$$

or

$$P(\ell \le x) = \operatorname{erf}\left(\frac{x+|d|}{\sqrt{2t}}\right) \tag{7.248}$$

(local-time cumulative density)

after cancelling terms.

7.3.2 Standard Brownian Bridge

The local-time distributions for the standard Brownian bridge $B(t)$ are special cases of the results in the following section, so we just quote them here.

$$f_\ell(x) = \left[1 - e^{-2|d|^2}\right]\delta(x - 0^+) + \big(x + 2|d|\big)\, e^{-(x+2|d|)^2/2} \tag{7.249}$$

(local-time probability density for $B(t)$)

is the probability density for the standard bridge, while

$$P(\ell \le x) = 1 - e^{-(x+2|d|)^2/2} \tag{7.250}$$

(local-time cumulative density for $B(t)$)

is the cumulative density.

Note that as in Eq. (7.148), it is not difficult to compute moments of the local time using

$$\begin{aligned} \left\langle\!\left\langle \ell^n \right\rangle\!\right\rangle &= n \int_0^\infty dx\, x^{n-1}[1 - P(\ell \le x)] \\ &= n \int_0^\infty dx\, x^{n-1} e^{-(x+2|d|)^2/2}. \end{aligned} \tag{7.251}$$

Thus, for example,

$$\left\langle\!\left\langle \ell \right\rangle\!\right\rangle = \sqrt{\frac{\pi}{2}}\, \operatorname{erfc}\left(\sqrt{2}\,|d|\right) \tag{7.252}$$

for the mean local time of a standard Brownian bridge.

7.3.3 Brownian Bridge

For the local time of the Brownian bridge $B_{0\to c}(t)$ from 0 to c as $t = 0$ to 1, we will merge the procedures of Sections 7.2.2, 7.2.3, and 7.3.1. Here $f_\ell(x)$ for $0 \le x \le \infty$ will denote the probability density for the local time $\ell[B_{0\to c}(t); d]$ of $B_{0\to c}(t)$ at d, with cumulative probability $P(\ell \le x)$. The procedure is the same as in Section 7.3.1 up to Eq. (7.229), where we will leave $g(x)$ in the Feynman–Kac formula:

$$f(x) = \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\left\langle g[x + W(t)] \exp\left(-s \int_0^t dt'\, \delta[x + W(t') - d]\right) \right\rangle\!\right\rangle. \tag{7.253}$$

[34] cf. Lajos Takács, "On the Local Time of the Brownian Motion," *The Annals of Applied Probability* **5**, 741 (1995), Eq. (3) (doi: 10.1214/aoap/1177004703).

In this case, we will take $g(x) = e^{ik(x-c)}$, as we did in Section 7.2.3. This is then the solution of the steady-state diffusion equation

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\delta(x-d)f(x) + e^{ik(x-c)}. \tag{7.254}$$

For $x \neq d$, the ODE is

$$f'' = 2\lambda f - 2e^{ik(x-c)}. \tag{7.255}$$

Then setting $h = f - 2e^{ik(x-c)}/(2\lambda + k^2)$, we have

$$\begin{aligned} h'' &= f'' + \frac{2k^2 e^{ik(x-c)}}{2\lambda + k^2} \\ &= 2\lambda f - 2e^{ik(x-c)} + \frac{2k^2 e^{ik(x-c)}}{2\lambda + k^2} \\ &= 2\lambda h + \frac{4\lambda e^{ik(x-c)}}{2\lambda + k^2} - 2e^{ik(x-c)} + \frac{2k^2 e^{ik(x-c)}}{2\lambda + k^2} \\ &= 2\lambda h, \end{aligned} \tag{7.256}$$

so that for $x \neq d$,

$$h(x) \propto e^{\pm\sqrt{2\lambda}\,x}, \tag{7.257}$$

or picking the bounded solutions in each domain,

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} + \dfrac{e^{ik(x-c)}}{\lambda + k^2/2} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{e^{ik(x-c)}}{\lambda + k^2/2} & (x < d), \end{cases} \tag{7.258}$$

for undetermined constants A and B. Demanding continuity of $f(x)$ at $x = d$ again gives

$$B = Ae^{-2\sqrt{2\lambda}\,d}. \tag{7.259}$$

The δ function in the ODE causes the derivative $f'(x)$ to jump by $2sf(d)$ at $x = d$, so that

$$-\sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d} - 2s\left(Ae^{-\sqrt{2\lambda}\,d} + \frac{e^{ik(d-c)}}{\lambda + k^2/2}\right) = \sqrt{2\lambda}Be^{\sqrt{2\lambda}\,d} = \sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d}, \tag{7.260}$$

or that is, $f'(d-0^+) = f'(d+0^+) - 2sf(d)$. The solution of these two equations fixes the coefficients as

$$A = -\frac{se^{\sqrt{2\lambda}\,d}\,e^{ik(d-c)}}{(\lambda + k^2/2)\left(\sqrt{2\lambda} + s\right)}, \qquad B = -\frac{se^{-\sqrt{2\lambda}\,d}\,e^{ik(d-c)}}{(\lambda + k^2/2)\left(\sqrt{2\lambda} + s\right)}. \tag{7.261}$$

Now we can equate Eqs. (7.229) and (7.258) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle e^{ik[W(t)-c]} \exp\left(-s\int_0^t dt'\,\delta[W(t') - d]\right)\right\rangle\!\!\right\rangle = B + \frac{e^{-ikc}}{\lambda + k^2/2}, \tag{7.262}$$

where we have again assumed $d > 0$ [but note that if we assume $d < 0$, the following treatment is the same, but with $d \longrightarrow -d$, so we will simply replace d by $|d|$ in the factor $\exp(-\sqrt{2\lambda}\,d)$]. Integrating with respect to k and dividing through by 2π introduces the δ function that pins the Wiener path to c at time t:

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \delta[W(t) - c] \exp\left(-s\int_0^t dt'\,\delta[W(t') - d]\right)\right\rangle\!\!\right\rangle = \frac{1}{2\pi}\int_{-\infty}^\infty dk\,\left(B + \frac{e^{-ikc}}{\lambda + k^2/2}\right). \tag{7.263}$$

Using Eq. (7.155) to change the Wiener-path average into a bridge average,

$$\int_0^\infty dt\,\frac{e^{-c^2/2t}}{\sqrt{t}}\,e^{-\lambda t}\left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\delta[B_{t(0\to c)}(t')-d]\right)\right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\,\left(B+\frac{e^{-ikc}}{\lambda+k^2/2}\right). \quad (7.264)$$

Then using Eq. (7.226) on the left-hand side and carrying out the k integration,

$$\begin{aligned}\int_0^\infty dt\,\frac{e^{-c^2/2t}}{\sqrt{t}}\,e^{-\lambda t}\left\langle\!\!\left\langle \exp\left\{-s\ell[B_{t(0\to c)};d]\right\}\right\rangle\!\!\right\rangle &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\left(\frac{e^{-ikc}}{\lambda+k^2/2}-\frac{s\,e^{-\sqrt{2\lambda}|d|}\,e^{ik(d-c)}}{(\lambda+k^2/2)\left(\sqrt{2\lambda}+s\right)}\right)\\ &= \sqrt{\frac{\pi}{\lambda}}\,e^{-\sqrt{2\lambda}|c|}-\frac{\sqrt{\pi}\,s\,e^{-\sqrt{2\lambda}\,(|d|+|c-d|)}}{\sqrt{\lambda}\left(\sqrt{2\lambda}+s\right)}.\end{aligned} \quad (7.265)$$

Now using the Laplace-transform formulae[35]

$$\mathscr{L}\left[\frac{1}{\sqrt{\pi t}}e^{-k^2/4t}\right](\lambda) = \frac{1}{\sqrt{\lambda}}e^{-k\sqrt{\lambda}}, \quad (7.266)$$

and[36]

$$\mathscr{L}\left[e^{ak}e^{a^2t}\,\mathrm{erfc}\left(a\sqrt{t}+\frac{k}{2\sqrt{t}}\right)\right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}\left(\sqrt{\lambda}+a\right)}, \quad (7.267)$$

which becomes with $k=(|d|+|c-d|)\sqrt{2}$, $a=s/\sqrt{2}$,

$$\mathscr{L}\left[e^{s(d+|c-d|)}e^{s^2t/2}\,\mathrm{erfc}\left(s\sqrt{t/2}+\frac{(d+|c-d|)}{\sqrt{2t}}\right)\right](\lambda) = \frac{e^{-(d+|c-d|)\sqrt{2\lambda}}}{\sqrt{\lambda}\left(\sqrt{\lambda}+s/\sqrt{2}\right)} = \frac{\sqrt{2}\,e^{-(d+|c-d|)\sqrt{2\lambda}}}{\sqrt{\lambda}\left(\sqrt{2\lambda}+s\right)}. \quad (7.268)$$

Thus, Eq. (7.264) becomes

$$\int_0^\infty dx\,e^{-sx}\,f_\ell(x) = 1-\sqrt{\frac{\pi t}{2}}\,s\,e^{c^2/2t}e^{s(|d|+|c-d|)}e^{s^2t/2}\,\mathrm{erfc}\left(s\sqrt{\frac{t}{2}}+\frac{(|d|+|c-d|)}{\sqrt{2t}}\right) \quad (7.269)$$

after using Eq. (7.226) to replace the ensemble average on the left-hand side. Now again using

$$\int_0^\infty dx\,e^{-sx}\frac{1}{\sqrt{2\pi t}}e^{-(x+d)^2/2t} = \frac{1}{2}\,e^{sd}\,e^{s^2t/2}\,\mathrm{erfc}\left[\sqrt{\frac{t}{2}}\left(s+\frac{d}{t}\right)\right], \quad (7.270)$$

with the derivative rule

$$s\mathscr{L}[f(x)](s) = \mathscr{L}[f'(x)](s)+f(0), \quad (7.271)$$

so that

$$-\int_0^\infty dx\,e^{-sx}\frac{(x+d)}{\sqrt{2\pi t^3}}e^{-(x+d)^2/2t}+\frac{1}{\sqrt{2\pi t}}e^{-d^2/2t} = \frac{s}{2}\,e^{sd}\,e^{s^2t/2}\,\mathrm{erfc}\left[\sqrt{\frac{t}{2}}\left(s+\frac{d}{t}\right)\right], \quad (7.272)$$

we can invert the Laplace transforms on both sides, with the result[37]

$$f_\ell(x) = \left[1-e^{[c^2-(|d|+|c-d|)^2]/2t}\right]\delta(x-0^+)+\frac{1}{t}\left(x+|d|+|c-d|\right)e^{[c^2-(x+|d|+|c-d|)^2]/2t},$$

(local-time probability density for $B_{t(0\to c)}(t')$) (7.273)

[35] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.84).
[36] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1027, Eq. (29.3.90).
[37] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 155, formula 1.3.8; A. N. Borodin, "Brownian local time," *Russian Mathematical Surveys* **44**, 1 (1989), p. 6 (doi: 10.1070/RM1989v044n02ABEH002050).

The δ-function term here gives the boundary-noncrossing probability in the case of $c < d$, where the exponential part reduces to $\exp[-2d(d-c)]$, in agreement with Eq. (6.24).

The corresponding cumulative probability is

$$P(\ell \le x) = \int_0^x dx'\, f_\ell(x') = 1 - e^{[c^2-(|d|+|c-d|)^2]/2t} - \Big[e^{[c^2-(x+|d|+|c-d|)^2]/2t}\Big]_0^x, \tag{7.274}$$

or

$$P(\ell \le x) = 1 - e^{[c^2-(x+|d|+|c-d|)^2]/2t}$$

(local-time cumulative density for $B_{t(0\to c)}$) (7.275)

after cancelling terms.

7.3.3.1 Moments

Note that as in Eqs. (7.148) and (7.251), we can compute moments of the local time via the cumulative as probability function

$$\begin{aligned} \Big\langle\!\Big\langle \ell^n \Big\rangle\!\Big\rangle &= n \int_0^\infty dx\, x^{n-1}[1 - P(\ell \le x)] \\ &= n \int_0^\infty dx\, x^{n-1}\, e^{[c^2-(x+|d|+|c-d|)^2]/2t} \end{aligned} \tag{7.276}$$

for $n > 0$. For example, the mean is

$$\big\langle\!\big\langle \ell \big\rangle\!\big\rangle = \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t}\, \operatorname{erfc}\!\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \tag{7.277}$$

Note that in the case where the endpoints of the bridge straddle the interval (i.e., $0 \le d \le c$ or $c \le d \le 0$) the argument $|d| + |c-d|$ reduces to $|c|$, in which case the mean local time is independent of the boundary location, which seems peculiar.[38] Since the probability (7.275) and density (7.273) depend on c and d in exactly the same way, these probabilities are also invariant to shifts of the interface, provided the shift keeps the interface d between 0 and c.[39]

7.3.3.2 Moment-Generating Function

Since the moments are fairly straightforward to calculate, it shouldn't come as much surprise that the moment-generating function is easy to obtain. Writing this out,

$$\begin{aligned} \Big\langle\!\Big\langle e^{-s\ell} \Big\rangle\!\Big\rangle &= \int_0^\infty dx\, e^{-sx}\, f_\ell(x) \\ &= s \int_0^\infty dx\, e^{-sx} P(\ell \le x) \\ &= 1 - s \int_0^\infty dx\, e^{-sx} \big[1 - P(\ell \le x)\big] \\ &= 1 - s \int_0^\infty dx\, e^{-sx} e^{[c^2-(x+|d|+|c-d|)^2]/2t}, \end{aligned} \tag{7.278}$$

where we used Eq. (7.275) for the cumulative density. Note that in the first step, we discarded the boundary terms from integrating by parts; the term at $x = 0$ vanishes if we regard the delta function in Eq. (7.275) to

[38] Zhiyi Chi, Vladimir Pozdnyakov, and Jun Yan, "On expected occupation time of Brownian bridge," *Statistics and Probability Letters* **97**, 83 (2015) (doi: 10.1016/j.spl.2014.11.009).

[39] Peter Howard and Kevin Zumbrun, "Shift invariance of the occupation time of the Brownian bridge process," *Statistics and Probability Letters* **45**, 379 (1999) (doi: 10.1016/S0167-7152(99)00080-2).

be displaced *above* $x = 0$ until the end of the calculation (which is a shortcut for treating the delta-function component separately). Carrying out the final integral, we find[40]

$$\left\langle\!\left\langle e^{-s\ell} \right\rangle\!\right\rangle = 1 - s\sqrt{\frac{\pi t}{2}}\, e^{c^2/2t + s^2 t/2 + s(|d| + |c-d|)} \operatorname{erfc}\!\left(\frac{|d| + |c-d| + st}{\sqrt{2t}}\right)$$

(local-time moment-generating function for $B_{t(0\to c)}$) (7.279)

for the moment-generating function.

7.3.3.3 Application to the Sojourn Time

The local-time mean can also be used to obtain an expression for the mean of the sojourn time for a general Brownian bridge. By taking the expectation value of Eq. (7.213), we can simply write the mean sojourn time as the integral

$$\left\langle\!\left\langle T_\mathrm{s}[y(t); d] \right\rangle\!\right\rangle = \int_d^\infty dx \left\langle\!\left\langle \ell[y(t); x] \right\rangle\!\right\rangle. \tag{7.280}$$

To handle this integral, we can use the integral formula

$$\int dx \operatorname{erfc}(ax - b) = \frac{b}{a} - \frac{1}{a\sqrt{\pi}} + \left(x - \frac{b}{a}\right) \operatorname{erfc}(ax - b), \tag{7.281}$$

and handle the integral in several cases. First, if $d \geq 0$ and $d \geq c$, then we have from Eq. (7.277)

$$\begin{aligned} \left\langle\!\left\langle T_\mathrm{s}[y(t); d] \right\rangle\!\right\rangle &= \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t} \int_d^\infty dx \operatorname{erfc}\!\left(\frac{2x - c}{\sqrt{2t}}\right) \\ &= \frac{t}{2} e^{-2d(d-c)/t} - \sqrt{\frac{\pi t}{8}}\, e^{c^2/2t} (2d - c) \operatorname{erfc}\!\left(\frac{2d - c}{\sqrt{2t}}\right). \end{aligned} \tag{7.282}$$

In the case $d \geq 0$ and $c \geq d$, we can divide the integrand into the parts with $x \leq c$ and $x \geq c$, the latter of which we just did:

$$\begin{aligned} \left\langle\!\left\langle T_\mathrm{s}[y(t); d] \right\rangle\!\right\rangle &= \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t} \int_d^c dx \operatorname{erfc}\!\left(\frac{|x|}{\sqrt{2t}}\right) + \left\langle\!\left\langle T_\mathrm{s}[y(t); c] \right\rangle\!\right\rangle \\ &= \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t} (c - d) \operatorname{erfc}\!\left(\frac{c}{\sqrt{2t}}\right) + \frac{t}{2} - \sqrt{\frac{\pi t}{8}}\, e^{c^2/2t} c \operatorname{erfc}\!\left(\frac{c}{\sqrt{2t}}\right) \\ &= \frac{t}{2} - \sqrt{\frac{\pi t}{8}}\, e^{c^2/2t} (2d - c) \operatorname{erfc}\!\left(\frac{c}{\sqrt{2t}}\right). \end{aligned} \tag{7.283}$$

Combining Eqs. (7.282) and (7.283) into a single case valid for $d \geq 0$,

$$\left\langle\!\left\langle T_\mathrm{s}[y(t); d] \right\rangle\!\right\rangle = \frac{t}{2}\, e^{-2[d(d-c)\Theta(d)\Theta(d-c)]/t} - \sqrt{\frac{\pi t}{8}}\, (2d - c)\, e^{c^2/2t} \operatorname{erfc}\!\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \tag{7.284}$$

For the case $d \leq 0$, we can change the signs of both c and d to obtain the mirror image, and replace the sojourn time by the sojourn time subtracted from t. The net result for $d \leq 0$ is

$$\left\langle\!\left\langle T_\mathrm{s}[y(t); -d] \right\rangle\!\right\rangle = t - \frac{t}{2}\, e^{-2[d(d-c)\Theta(-d)\Theta(c-d)]/t} - \sqrt{\frac{\pi t}{8}}\, (2d - c)\, e^{c^2/2t} \operatorname{erfc}\!\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \tag{7.285}$$

Equations (7.284) and (7.285) can then be combined into a single case as

$$\begin{aligned} \left\langle\!\left\langle T_\mathrm{s}[y(t); d] \right\rangle\!\right\rangle = \frac{t}{2} + \operatorname{sgn}(d) \frac{t}{2} \Big[e^{-2[d(d-c)\Theta(d)\Theta(d-c) + d(d-c)\Theta(-d)\Theta(c-d)]/t} - 1 \Big] \\ - \sqrt{\frac{\pi t}{8}}\, (2d - c)\, e^{c^2/2t} \operatorname{erfc}\!\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \end{aligned} \tag{7.286}$$

[40] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 155, formula 1.3.7.

Actually, since the second term vanishes as long as $0 \le d \le c$, the $\mathrm{sgn}(d)$ can be shifted, with the result

$$\begin{aligned}\left\langle\!\left\langle T_{\mathrm{s}}[y(t);d]\right\rangle\!\right\rangle = \frac{t}{2} + \mathrm{sgn}(2d-c)\,\frac{t}{2}\left[e^{-2[d(d-c)\Theta(d)\Theta(d-c)+d(d-c)\Theta(-d)\Theta(c-d)]/t}-1\right] \\ - \sqrt{\frac{\pi t}{8}}\,(2d-c)\,e^{c^2/2t}\,\mathrm{erfc}\left(\frac{|d|+|c-d|}{\sqrt{2t}}\right),\end{aligned} \tag{7.287}$$

which makes the expression symmetric about $c/2$.

7.3.4 Local Time and Discontinuities in Stochastic Processes

Because of the relation of the local time to the delta function, the local time tends to show up in the theory of stochastic processes especially when discontinuities are involved. We will consider two examples of this here: the absolute-value process, and a diffusion process with a discontinuity in the diffusion rate.

7.3.4.1 Reflected Brownian Motion

Consider the absolute-value diffusion process $|W(t)| = \mathrm{sgn}[W(t)]\,W(t)$. This is an example of **reflected Brownian motion**, in the sense that this is Brownian motion on $(0,\infty)$, but when the process encounters the origin and attempts to cross it, it is ''reflected'' back into the positive real axis.

The Itō differential for this process is, expanding to second order using the Itō chain rule (3.118),

$$\begin{aligned} d|W(t)| &= \frac{d|W|}{dW}dW + \frac{1}{2}\frac{d^2|W|}{dW^2}dW^2 \\ &= \mathrm{sgn}(W)\,dW + \delta(W)\,dt,\end{aligned} \tag{7.288}$$

where we have used

$$\frac{d|x|}{dx} = \mathrm{sgn}(x), \qquad \frac{d|x|^2}{dx^2} = 2\delta(x), \tag{7.289}$$

thinking of the signum function as twice the Heaviside function plus a constant offset. Then integrating this transformed SDE from 0 to t,

$$|W(t)| = \int_0^t \mathrm{sgn}[W(t')]\,dW(t') + \ell[W(t);0], \tag{7.290}$$
(Tanaka formula)

where we used the local-time definition (7.211). This is the **Tanaka formula**.[41] This can also be written as

$$|W(t)-d| - |d| = \int_0^t \mathrm{sgn}[W(t')-d]\,dW(t') + \ell[W(t);d], \tag{7.291}$$
(Tanaka formula)

if $W(t)$ is shifted by d at the start of the derivation. This says that in Itō calculus, the reflecting effect of the absolute value is reflected in a deterministic ''pushing'' term that activates whenever the process hits the origin.

7.3.4.2 Discontinuous Diffusion Rate

Consider the Itō-form SDE

$$dy = \beta(y)\,dW, \tag{7.292}$$

representing state-dependent diffusion without an explicit drift term. We want to consider the case where $\beta(y)$ has a discontinuity (or a number of isolated discontinuities). As a particular model, consider

$$\beta(y) = \begin{cases} \beta_>, & y > 0 \\ \beta_<, & y < 0, \end{cases} \tag{7.293}$$

[41] Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 43.

with the point $\beta(0)$ being, say, the average of the left- and right-hand values. As we argued before in Section 3.3.5.1, the probability density of particles governed by this SDE evolves by the Fokker–Planck equation

$$\partial_t P(y,t) = \frac{1}{2}\partial_y^2 \beta^2(y) P(y,t), \tag{7.294}$$

so that the diffusion coefficient $D(y) = \beta^2(y)$ has a discontinuity at the origin. Making the Lamperti transform (Section 5.1.4.2), we construct the function

$$z = S(y) = \int_0^y \frac{dy'}{\beta(y')} = \begin{cases} y/\beta_>, & y > 0 \\ -y/\beta_<, & y < 0 \end{cases} \tag{7.295}$$

(which is invertible provided $\beta_{\gtrless}$ have the same sign), we can transform our SDE to the form of Eq. (5.74),

$$dz = -\frac{1}{2}\beta'(y)\,dt + dW. \tag{7.296}$$

Evaluating the derivative of β as the derivative of a step function,

$$dz = -\frac{1}{2}(\beta_> - \beta_<)\,\delta(y)\,dt + dW. \tag{7.297}$$

Now we have

$$\delta(z) = \delta[S(y)] = \frac{\delta(y)}{|S'(y)|} = \frac{2\delta(y)}{\beta_> + \beta_<}, \tag{7.298}$$

where we think of the delta function as a limit of centered distributions, so that the derivative of $S(y)$ is the average of the derivatives on either side of the discontinuity. Thus, we have the fully transformed SDE

$$dz = \left(\frac{\beta_< - \beta_>}{\beta_< + \beta_>}\right)\delta(z)\,dt + dW. \tag{7.299}$$

In this form, where we have scaled away the variance, the discontinuity arises as a deterministic ''kick'' term that acts only at the discontinuity's location. Using the local-time definition (7.211), we can integrate this to find

$$z(t) - z_0 = W(t) + \left(\frac{\beta_< - \beta_>}{\beta_< + \beta_>}\right)\ell[W(t);0], \tag{7.300}$$

(skew Brownian motion)

so that $z(t)$ has the form of a regular Wiener path, with a shift that depends on the time the particle spent at the discontinuity (and thus being kicked by the local gradient). The motion of $z(t)$ is called **skew Brownian motion**.

7.3.4.3 Skew Brownian Motion

Skew Brownian motion[42] is equivalent to ordinary Brownian motion, except that the probability is skewed towards one direction at the origin. Suppose we take a random walk of steps taken every Δt, and probability density

$$f(x) = p\,\delta\left(x - \sqrt{\Delta t}\right) + (1-p)\,\delta\left(x + \sqrt{\Delta t}\right), \tag{7.301}$$

so that the probability of a step of size $x = +\sqrt{\Delta t}$ is p, and the probability is $1-p$ of stepping in the opposite direction. The second moment of each step is

$$\int_{-\infty}^{\infty} dx\, x^2 f(x) = \Delta t, \tag{7.302}$$

[42] Skew Brownian motion was introduced by Kiyosi Itô and Henry P. McKean, Jr., *Diffusion Processes and their Sample Paths* (Springer-Verlag, 1974). See also J. M. Harrison and L. A. Shepp, ''On Skew Brownian Motion,'' *The Annals of Probability* **9**, 309 (1981) (doi: 10.1214/aop/1176994472); and Antoine Lejay, ''On the constructions of the skew Brownian motion,'' *Probability Surveys* **3** 413 (2006) (doi: 10.1214/154957807000000013).

so that the unbiased case $p = 1/2$ leads to the usual Wiener path $W(t)$ as $\Delta t \longrightarrow 0$. The more general case gives a bias of

$$\int_{-\infty}^{\infty} dx\, x f(x) = p\sqrt{\Delta t} - (1-p)\sqrt{\Delta t} = (2p-1)\sqrt{\Delta t} \tag{7.303}$$

in each step (and thus a step variance of $\Delta t - (2p-1)^2 \Delta t = 4p(1-p)\Delta t \leq \Delta t$ for $p \in [0,1]$). If we then consider an inhomogeneous walk where we take a biased stepping probability p in the region $x \in [-\sqrt{\Delta t}/2, \sqrt{\Delta t}/2)$, and an unbiased probability $p = 1/2$ outside this region, then we have the discrete-step equation

$$\Delta x = \Delta W + \mathrm{rect}\left(\frac{x}{\sqrt{\Delta t}}\right)(2p-1)\sqrt{\Delta t}, \tag{7.304}$$

where $\mathrm{rect}(t)$ is a square pulse of unit height and unit width, centered at $t = 0$, and we have replaced the stochastic component of the random walk by an equivalent Gaussian process in view of the limit $\Delta t \longrightarrow 0$. Note that this process has the correct mean

$$\langle\!\langle \Delta x \rangle\!\rangle = \mathbb{1}_{[-\sqrt{\Delta t}/2, \sqrt{\Delta t}/2)}(x), \tag{7.305}$$

where $\mathbb{1}_A(x)$ is the indicator function for $x \in A$, and the process has the correct second moment,

$$\langle\!\langle \Delta x^2 \rangle\!\rangle = \langle\!\langle \Delta W^2 \rangle\!\rangle = \Delta t, \tag{7.306}$$

provided we ignore terms of order $\Delta W\, \Delta t$ and Δt^2. Then as $\Delta t \longrightarrow 0$, we can also note that

$$\frac{1}{\sqrt{\Delta t}}\,\mathrm{rect}\left(\frac{x}{\sqrt{\Delta t}}\right) \longrightarrow \delta(x), \tag{7.307}$$

so that in this limit we obtain the SDE

$$dx = dW + (2p-1)\,\delta(x)\,dt, \tag{7.308}$$
(skew Brownian motion)

with solution

$$x(t) - x_0 = W(t) + (2p-1)\,\ell[W(t); 0]. \tag{7.309}$$
(skew Brownian motion)

These equations are equivalent to Eqs. (7.299) and (7.300) if we identify

$$p = \frac{\beta_<}{\beta_< + \beta_>}, \qquad 1 - p = \frac{\beta_>}{\beta_< + \beta_>}. \tag{7.310}$$

As an example at the extreme end of this range of probabilities, note that the reflected Brownian motion in Eq. (7.288) has a similar form to skew Brownian motion with $p = 1$ (if we modify the process by absorbing the sgn into dW, which is possible due to the symmetry of dW). In this limit, the process is always reflected upward at the origin. In terms of the inhomogeneous diffusion, this corresponds to the limit where $\beta_< \gg \beta_>$, though the diffusion picture does not necessarily show the same repulsion from the $y < 0$ region; this is more an artifact of the transformation (7.295), which "smooshes" the $y < 0$ region up against the origin if $\beta <$ is large.

7.3.4.4 Reflections, Images, and Transition Densities for Skew Brownian Motion

The original construction of skew Brownian motion[43] began with reflected Brownian motion, and then introduced probabilities p and $1-p$ of reflecting $|W(t)|$ upwards or downwards, respectively, on each encounter with the origin.

In this sense,

$$r = 2p - 1 = \frac{\beta_< - \beta_>}{\beta_< + \beta_>} \tag{7.311}$$
("reflection coefficient" for diffusion)

[43] Kiyosi Itô and Henry P. McKean, Jr., *op. cit.*

acts as a ''reflection coefficient'' from above the interface for the Wiener path. The reflection occurs always up or down for $p = 1$ or 0, respectively (corresponding to $r = +1$ or -1), and has no reflection (bias) for $p = 1/2$ ($r = 0$). Indeed, identifying the $\beta_{\lessgtr}$ with the inverse ''refractive indices'' $1/n_{\lessgtr}$ on either side of a refractive interface, the reflection coefficient r has the form of the Fresnel coefficient for the electric field at normal incidence from above a dielectric interface [and $t = 1 + r = 2\beta_</(\beta_> + \beta_<)$ has the form of the Fresnel transmission coefficient from above].

Indeed, the reflection principle can give the transition density for the trajectories,[44] making the connection to a reflection coefficient more explicit. To compute the transition densities for skew Brownian motion, we will use a procedure similar to the calculation for boundary crossings of Brownian bridges in Section 6.1.2. To begin, let's compute the probability density for a skew-Brownian-motion process $z(t)$ starting at $x_0 > 0$ and ending at $x > 0$. Let τ_0 denote the first-crossing time through the boundary $x = 0$. Then we start with

$$P[z(t) = x] = P[z(t) = x \wedge \tau_0 \le t] + P[z(t) = x \wedge \tau_0 > t]. \tag{7.312}$$

That is, the path either hits the origin or it doesn't before time t. We can compute the first term using the Reflection Principle, similar to the calculation in Section 6.1:

$$\begin{aligned} P[z(t) = x \wedge \tau_0 \le t] &= P[z(t) = x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= P[z(t) = x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= 2pP[W(t) + x_0 = x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= 2pP[W(t) + x_0 = -x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= 2pP[W(t) + x_0 = -x \wedge \tau_0 \le t] \\ &= 2pP[W(t) + x_0 = -x] \\ &= 2p\phi(x + x_0)\, dx, \end{aligned} \tag{7.313}$$

where $\phi(x)$ is the transition density for $W(t)$ from 0,

$$\phi(x) = \frac{1}{\sqrt{2\pi t}}\, e^{-x^2/2t}. \tag{7.314}$$

In changing from skew Brownian to regular Brownian motion, we used the ratio of probabilities for taking an upward step at the point of crossing $z(\tau) = 0$:

$$\frac{P[z(0) = 0 \wedge z(t) = x]}{P[W(t) = x]} = \frac{p}{1/2} = 2p. \tag{7.315}$$

(This ratio is unity for the unbiased case $p = 1/2$.) We then used the Reflection Principle for $W(t)$, given that it crossed $W(\tau) + z_0 = 0$.

We can evaluate second term in Eq. (7.312) using the **method of images**, which works in the same way as the method of images for potentials due to charges in the presence of boundary conditions. Here, we have $P[z(t) = x \wedge \tau_0 > t]$, which is the same as $P[z(t) = x]$, but subject to the boundary condition $P[z(t) = x = 0] = 0$, since no crossings happen here. We can achieve this by considering the density

[44] J. B. Walsh, ''A diffusion with a discontinuous local time,'' *Société Mathématique de France Astérisque* **52-53**, 37 (1978); J. F. Le Gall, ''One-dimensional stochastic differential equations involving the local times of the unknown process,'' in *Stochastic Analysis and Applications: Proceedings of the International Conference held in Swansea, April 11-15, 1983*, A. Truman and D. Williams, Eds. (Springer-Verlag, 1984), p. 51; G. J. M. Uffink, ''A random walk method for the simulation of macrodispersion in a stratified aquifer,'' in *Relation of Groundwater Quantity and Quality (Proceedings of the Hamburg Symposium, August 1983)* (IAHS Publ. no. 146, 1985), p. 103; Antoine Lejay, *op. cit.*, around Eq. (42); Thilanka Appuhamillage, Vrushali Bokil, Enrique Thomann, Edward Waymire, and Brian Wood, ''Occupation and local times for skew Brownian motion with applications to dispersion across an interface,'' *Annals of Applied Probability* **21**, 183 (2011), Eq. (2.2) (doi: 10.1214/10-AAP691); Eric M. LaBolle and Graham E. Fogg, ''Reply [to ''Comment on 'Diffusion theory for transport in porous media: Transition-probability densities of diffusion processes corresponding to advection-dispersion equations' by Eric M. LaBolle et al.''],'' *Water Resources Research* **36**, 823 (2012) (doi: 10.1029/1999WR900325).

diffusing from a distribution initially localized at x_0, balanced by a distribution of *negative* density at $-x_0$, such that perfect cancellation occurs at $x = 0$:

$$P[z(t) = x \wedge \tau_0 > t] = \phi(x - x_0)\,dx - \phi(x + x_0)\,dx. \tag{7.316}$$

Putting this all together in Eq. (7.312), we then have

$$P[z(t) = x] = \phi(x - x_0) + (2p - 1)\,\phi(x + x_0)\,dx, \tag{7.317}$$

again for $x > 0$ and $x_0 > 0$.

For $x_0 > 0$ and $x < 0$, the second term in Eq. (7.312) corresponds to an impossible event, and by a similar Reflection-Principle argument, the second term is $2(1 - p)\,\phi(x + x_0)$:

$$P[z(t) = x] = 2(1 - p)\,\phi(x + x_0)\,dx. \tag{7.318}$$

Collecting both cases together, we have for $z(0) = z_0$, the probability density

$$P[z(t)] = \begin{cases} \phi(z - z_0)\,dz + r\phi(z + z_0)\,dz, & z > 0, z_0 > 0 \\ (1 - r)\,\phi(z + z_0)\,dz, & z < 0, z_0 > 0, \end{cases}$$

(transition density for skew Brownian motion) (7.319)

with $\phi(x)$ defined in Eq. (7.314), and the reflection coefficient r given by Eq. (7.311). The same formula applies for $z_0 < 0$ with minor changes,

$$P[z(t)] = \begin{cases} \phi(z - z_0)\,dz - r\phi(z + z_0)\,dz, & z < 0, z_0 < 0 \\ (1 + r)\,\phi(z + z_0)\,dz, & z > 0, z_0 < 0, \end{cases}$$

(transition density for skew Brownian motion) (7.320)

where both r and z change sign.

If we translate these results back into the language of the discontinuous-diffusion SDE (7.292), we can invert the Lamperti transform (7.295) to obtain

$$y = S^{-1}(z) = \begin{cases} z\beta_>, & z > 0 \\ -z\beta_<, & z < 0. \end{cases} \tag{7.321}$$

Then transforming the probability density, Eq. (7.319) becomes

$$P[y(t)] = \begin{cases} \phi\left(\dfrac{y - y_0}{\beta_>}\right)\dfrac{dy}{\beta_>} + r\phi(x + x_0)\,\dfrac{dy}{\beta_>}, & y > 0, y_0 > 0 \\ (1 - r)\,\phi\left(\dfrac{y - (\beta_</\beta_>)y_0}{\beta_<}\right)\dfrac{dy}{\beta_<}, & y < 0, y_0 > 0, \end{cases}$$

(transition density for discontinuous diffusion) (7.322)

where note that in the $y < 0$ case, the y_0 was determined by x_0 in the Reflection-Principle argument for $x > 0$, and should correspond to the same distance from the origin as $y_0/\beta_>$ to properly match the boundary condition at $y = 0$ (that is, ''equivalent'' trajectories from the effective sources on either side will hit $y = 0$ at the same time only with this rescaling). Again, the $y_0 < 0$ case can be obtained by exchanging $\beta_{\gtrless}$ everywhere, including in the reflection coefficient r.

7.3.4.5 Stratonovich Discontinuous Diffusion

We began the discussion of a discontinuous diffusion rate with the Itō-form diffusion (7.292), and the choice of Itō calculus was important in generating the ''pumping'' effect at the discontinuity. If we instead consider a Stratonovich-form diffusion SDE

$$dy = \beta(y) \circ dW, \tag{7.323}$$

corresponding to the Itō SDE

$$dy = \frac{1}{2}\beta'(y)\beta(y)\,dt + \beta(y)\,dW, \tag{7.324}$$

then things are different; in fact, the Lamperti-transformed version of this SDE is simply

$$dz = dW, \tag{7.325}$$

as we discussed in Section 5.1.4.2. The equivalent Fokker–Planck equation from Eq. (3.53) is

$$\partial_t P(y,t) = -\frac{1}{2}\partial_y \beta'(y)\beta(y)P(y,t) + \frac{1}{2}\partial_y^2 \beta^2(y)P(y,t), \tag{7.326}$$

or commuting a derivative,

$$\partial_t P(y,t) = \frac{1}{2}\partial_y \beta(y)\partial_y \beta(y)P(y,t).$$

(equivalent Fokker–Planck equation) (7.327)

Note that by changing variables to z according to

$$\frac{dz}{dy} = \frac{1}{\beta}, \tag{7.328}$$

which is the same as the transformation (7.295), we can rewrite the Fokker–Planck equation as a standard diffusion equation,

$$\partial_t \rho = \frac{1}{2}\partial_z^2 \rho, \tag{7.329}$$

(equivalent Fokker–Planck equation)

where

$$\rho(y,t) = \beta(y)P(y,t). \tag{7.330}$$

This simplification is the same as the result $dz = dW$ from the Lamperti transformation of the Stratonovich-diffusion SDE.

7.4 Exercises

Problem 7.1

(a) Show that

$$\delta[f(x)] = \sum_{x_0 \in f^{-1}(0)} \frac{\delta(x - x_0)}{|f'(x_0)|}, \tag{7.331}$$

where the sum is over the (isolated) zeros of $f(x)$. Good ways to proceed are: to write the delta function as a derivative of the step function $\Theta[f(x)]$, or to treat the delta function as a limit of a more "reasonable" function, such as a box function of width a and height $1/a$.

(b) Show that

$$\delta'[f(x)] = \sum_{x_0 \in f^{-1}(0)} \left[\frac{\delta'(x - x_0)}{f'(x_0)|f'(x_0)|} + \frac{f''(x_0)\,\delta(x - x_0)}{|f'(x_0)|^3} \right]. \tag{7.332}$$

The same suggested approaches apply here.

Problem 7.2

Rederive the probability density 7.273

$$f_\ell(x) = \left[1 - e^{[c^2 - (|d| + |c-d|)^2]/2t}\right] \delta(x - 0^+) + \frac{1}{t}\left(x + |d| + |c-d|\right) e^{[c^2 - (x + |d| + |c-d|)^2]/2t}, \tag{7.333}$$

for the local time of a Brownian bridge $B_{t(0\to c)}(t)$ pinned to c at time t. Use an alternate derivation: begin with the same Feynman–Kac-based approach as in Section 7.3.3, but take $g(x) = \Theta(c - x)$ instead of $g(x) = e^{ik(x-c)}$, and then after working out the solution of the appropriate ODE, differentiate with respect to c to obtain the pinning delta function.

Note that the Wiener path can be pinned even more directly by taking $g(x) = \delta(x - c)$. The advantage of the method using $g(x) = e^{ik(x-c)}$ is that it replaces the book-keeping of several different cases with a single extra integral over k. Also, note that by avoiding the c-derivative and completing the calculation, it is possible to obtain the joint probability of $W(t)$ occupying $[0, \infty)$ and the local time $\ell[W(t); d]$ taking on a particular value.

Problem 7.3

Given a state-dependent diffusion function of the form

$$\sigma(y) = \sigma_0 + (\sigma_1 - \sigma_0)\,\Theta(y - d), \tag{7.334}$$

where the diffusion rate is σ_0 for $y < d$ and σ_1 for $y > d$, show that an explicit solution (5.64) for the SDE (5.53)

$$dy(t) = \sigma(y) \circ dW(t) \tag{7.335}$$

with the conditions $y(1) = y(0) = 0$, is given by

$$y(t) = \sigma_0 B(t) + \left[(\sigma_1 - \sigma_0)B(t) + d\left(1 - \frac{\sigma_1}{\sigma_0}\right)\right] \Theta[B(t) - d], \tag{7.336}$$

where $\Theta(x)$ is the Heaviside function and $B(t)$ is a standard Brownian bridge.

Problem 7.4

(a) Using the calculation of the probability density for the local time $\ell[W(t); d]$ as a template, derive

the following formula for the expectation value[45]

$$\lambda\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\Big(-s\ell[W(t);d]-s'\ell[W(t);d']\Big)\right\rangle\!\!\right\rangle_{W(t)} = 1-\frac{s\left[\sqrt{2\lambda}+s'\left(1-e^{-\sqrt{2\lambda}|d-d'|}\right)\right]e^{-\sqrt{2\lambda}|d|}+s'\left[\sqrt{2\lambda}+s\left(1-e^{-\sqrt{2\lambda}|d-d'|}\right)\right]e^{-\sqrt{2\lambda}|d'|}}{\left(\sqrt{2\lambda}+s\right)\left(\sqrt{2\lambda}+s'\right)-ss'e^{-2\sqrt{2\lambda}|d-d'|}}, \tag{7.337}$$

giving the dual moment-generating function for the two local times $\ell[W(t);d]$ and $\ell[W(t);d']$, given that the process $W(t)$ is stopped exponentially at rate λ.

(b) Repeat for a Brownian bridge $B_{t(0\to c)}(t)$, pinned at the exponentially distributed stopping time t to $B(t)=c$, to find the analogous result[46]

$$\begin{aligned}\lambda\int_0^\infty dt\, e^{-\lambda t}\,\frac{e^{-c^2/2t}}{\sqrt{2\pi t}}\left\langle\!\!\left\langle \exp\Big(-s\ell[B_{t(0\to c)}(t');d]-s'\ell[B_{t(0\to c)}(t');d']\Big)\right\rangle\!\!\right\rangle_{B_{t(0\to c)}(t')} \\ = \sqrt{\frac{\lambda}{2}}\,e^{-\sqrt{2\lambda}|c|}\left(1-\frac{s\left[\left(\sqrt{2\lambda}+s'\right)e^{-\sqrt{2\lambda}|d|}-s'e^{-\sqrt{2\lambda}(|d'|+|d-d'|)}\right]e^{-\sqrt{2\lambda}(|c-d|-|c|)}}{\left(\sqrt{2\lambda}+s\right)\left(\sqrt{2\lambda}+s'\right)-ss'e^{-2\sqrt{2\lambda}|d-d'|}}\right. \\ \left.-\frac{s'\left[\left(\sqrt{2\lambda}+s\right)e^{-\sqrt{2\lambda}|d'|}-se^{-\sqrt{2\lambda}(|d|+|d-d'|)}\right]e^{-\sqrt{2\lambda}(|c-d'|-|c|)}}{\left(\sqrt{2\lambda}+s\right)\left(\sqrt{2\lambda}+s'\right)-ss'e^{-2\sqrt{2\lambda}|d-d'|}}\right),\end{aligned} \tag{7.338}$$

where the Gaussian factor on the left-hand side indicates this expectation value was taken with respect to a joint distribution for the local times and for the stopping point of $B(t)$.

Problem 7.5

(a) Using the representation

$$\delta'(x)=\lim_{a\to 0^+}\frac{\delta(x+a/2)-\delta(x-a/2)}{a} \tag{7.339}$$

and the results of Problem 7.4, derive the following formula for the expectation value

$$\lambda\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\Big(-s\ell'[W(t);d]\Big)\right\rangle\!\!\right\rangle_{W(t)} = 1-e^{-\sqrt{2\lambda}|d|}, \tag{7.340}$$

giving the moment-generating function for the local-time derivative $\ell'[W(t);d]$ and $\ell[W(t);d']$, given that the process $W(t)$ is stopped exponentially at rate λ.

(b) Invert the λ and s Laplace transforms in this result to obtain an expression for the probability density $f_{\ell'}(x)$ of $\ell'[W(t);d]$. You should obtain something a bit funny; what gives?

Problem 7.6

Consider closed Brownian bridges $B_T(t)$ [i.e., such that $B_T(T)=0$].

(a) Using the probability method of Eqs. (7.155), show that

$$\left\langle\!\!\left\langle \delta[B_T(\tau)-c]\,F[B_T(t)]\right\rangle\!\!\right\rangle = \Phi[c,\tau(1-\tau/T)]\left\langle\!\!\left\langle F\big[B_T(t)\big]\right\rangle\!\!\right\rangle_{B_T(\tau)=c}, \tag{7.341}$$

[45] Andrei N. Borodin and Paavo Salminen, *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. (Birkhäuser, 2002), p. 177, Eq. (1.18.1).

[46] This is similar to, but not quite the same as, Borodin and Salminen, *op. cit.*, p. 177, Eq. (1.18.5).

where the conditional ensemble average on the right-hand side refers to Brownian bridges pinned to c at time τ $(0 < \tau < T)$, and

$$\Phi(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/2\sigma^2} \tag{7.342}$$

is the centered Gaussian distribution with variance σ^2.

(b) Prove the same result by adapting the explicit-integration method of Section 7.2.4.

Chapter 8

Bessel Processes and Brownian Recurrence

8.1 Bessel Processes

A **Bessel process** in d dimensions is simply the ''radius'' or Euclidian norm of a vector Wiener process in d dimensions:

$$R_d := \left\|\mathbf{W}(t)\right\|_d = \sqrt{\sum_{\alpha=1}^{d} [W_\alpha(t)]^2}\ . \tag{8.1}$$
(Bessel process)

As we will see, this process is useful for characterizing the motion of a vector Wiener process, particularly in the sense of whether or not it returns to its starting point. Also, note that this is the multidimensional generalization of the reflecting Wiener process of Section 7.3.4.1, to which this reduces for $d = 1$.

To begin, we can try to work out an SDE for R_d. First, writing out the differential for R_d,

$$dR_d = \frac{1}{2R_d} \sum_{\alpha=1}^{d} \left(2W_\alpha\, dW_\alpha + dW_\alpha^2\right) = \frac{d}{2R_d}\, dt + \frac{1}{R_d} \sum_{\alpha=1}^{d} W_\alpha\, dW_\alpha. \tag{8.2}$$

Unfortunately, it is difficult to proceed directly from this point. However, we can also work more simply with the *square* of the Bessel process, and write out the differential:

$$dR_d^2 = \sum_{\alpha=1}^{d} \left(2W_\alpha\, dW_\alpha + dW_\alpha^2\right) = d\, dt + 2 \sum_{\alpha=1}^{d} W_\alpha\, dW_\alpha. \tag{8.3}$$

On the other hand, we can write out the differential as

$$dR_d^2 = 2R_d\, dR_d + (dR_d)^2. \tag{8.4}$$

From Eq. (8.2), we can find

$$(dR_d)^2 = \frac{1}{R_d^2} \sum_{\alpha=1}^{d} W_\alpha^2\, (dW_\alpha)^2 = dt, \tag{8.5}$$

where we used $dW_\alpha\, dW_\beta = 0$ if $\alpha \neq \beta$ for independent Wiener processes. Then using this result and (8.3) in Eq. (8.4), we find

$$dR_d = \frac{d-1}{2R_d}\, dt + \frac{1}{R_d} \sum_{\alpha=1}^{d} W_\alpha\, dW_\alpha. \tag{8.6}$$

What remains is to analyze the last term, which turns out to be equivalent to a Wiener process. Defining

$$d\tilde{W}(t) := \frac{1}{R_d} \sum_{\alpha=1}^{d} W_\alpha \, dW_\alpha, \tag{8.7}$$

it is easy to see that $\langle\!\langle d\tilde{W} \rangle\!\rangle = 0$ (since $\langle\!\langle f(W)\, dW \rangle\!\rangle = 0$ in Itō calculus), and also that

$$\langle\!\langle d\tilde{W}^2 \rangle\!\rangle = \frac{1}{R_d^2} \sum_{\alpha=1}^{d} W_\alpha^2 \, (dW_\alpha)^2 = dt. \tag{8.8}$$

The higher-order moments of $d\tilde{W}$ vanish also, so we can see that $\tilde{W}$ is another representation of a Wiener process. Then dropping the tilde, we have the SDE

$$dR_d = \frac{d-1}{2R_d}\, dt + dW \tag{8.9}$$
(Bessel-process SDE)

to directly generate the dynamics of the Bessel process. Note that because of the divergence at $R_d = 0$, we should be careful to specify that the transformation to this equation is only valid provided $R_d \neq 0$. That is, if the Bessel process hits the origin, then this equation becomes invalid afterwards. [For example, in $d = 1$, we are missing the delta-function term that arises in Eq. (7.288).] We will soon return to the question of whether or not this actually happens.

8.1.1 Generator

The term "Bessel process" comes specifically for the generator. Recall that an SDE of the form $dx = \alpha\, dt + \beta\, dW$ implies a diffusion-type equation, which can be summarized by the generator [see Eq. (3.59)]

$$\mathcal{G} := \alpha(x,t)\, \partial_x + \frac{\beta^2(x,t)}{2}\, \partial_x^2. \tag{8.10}$$

For the Bessel-process SDE (8.9), the generator is

$$\mathcal{G} = \frac{d-1}{2x}\partial_x + \frac{1}{2}\partial_x^2, \tag{8.11}$$

and the generator-eigenvalue equation $\mathcal{G}u(x) = \gamma u(x)$ has the form

$$\frac{d-1}{2x}u' + \frac{1}{2}u'' = \gamma u. \tag{8.12}$$

[Note that this can be viewed as a Laplace transform of the backward Kolmogorov equation (3.67).] The idea is that this differential equation is equivalent to a Bessel equation, as we will now see.

First, changing variables according to

$$v = x^m u, \tag{8.13}$$

so that $v' = mx^{m-1}u + x^m u'$ and $v'' = m(m-1)x^{m-2}u + 2mx^{m-1}u' + x^m u''$, we will need to solve these for the derivatives of u to give $u' = x^{-m}v' - mx^{-1}u = x^{-m}v' - mx^{-m-1}v$ and $u'' = x^{-m}v'' - m(m-1)x^{-2}u - 2mx^{-1}u' = x^{-m}v'' - m(m-1)x^{-m-2}v - 2mx^{-m-1}v' + 2m^2x^{-m-2}v$. Thus, Eq. (8.12) becomes

$$\frac{1}{2}x^{-m}v'' + \frac{d-2m-1}{2}x^{-m-1}v' - \frac{m(d-m-2)}{2}x^{-m-2}v = \gamma x^{-m}v. \tag{8.14}$$

Then multiplying through by $2x^{m+2}$,

$$x^2v'' + (d-2m-1)xv' - \big[2\gamma x^2 + m(d-m-2)\big]v = 0. \tag{8.15}$$

Now we can scale out the eigenvalue by letting $\tilde{x} := \sqrt{2\gamma}\,x$, and choosing m such that $(\mathsf{d} - 2m - 1) = 1$, leading to $m = (\mathsf{d}-2)/2$ and $m(\mathsf{d} - m - 2) = [(\mathsf{d}-2)/2]^2$, we have

$$\tilde{x}^2 v'' + \tilde{x} v' - \left[\tilde{x}^2 + m^2\right] v = 0. \tag{8.16}$$

This is the modified Bessel equation, and the solutions are the modified Bessel functions $I_m(\tilde{x})$ and $K_m(\tilde{x})$. Note that for small z, $I_m(z)\tilde{z}^m = z^{(d/2)-1}$, while $K_m(z)\tilde{z}^{-m} = z^{1-(d/2)}$. Also, for large z, $I_m(z) \longrightarrow \infty$ as $e^z/\sqrt{z}$, while $K_m(z) \longrightarrow 0$. So the $K_m(z)$ represent physical solutions in this problem.

8.1.2 Brownian Recurrence

One useful question we can address with Bessel processes is, given a Wiener path in d dimensions starting at the origin and running for arbitrarily large times, how far away does it go, and does it return to the origin?

8.1.2.1 One Dimension

In $\mathsf{d} = 1$, we have already answered this question by analyzing boundary-crossing probabilities. That is, from Eq. (6.9) we know that as $t \longrightarrow \infty$, the probability for crossing a boundary at distance d away converges to unity for any value of d. This means that the path wanders arbitrarily far away from the origin, and once it has wandered any given distance away, it is guaranteed to return to the origin. Thus Brownian motion in $\mathsf{d} = 1$ dimension is said to be **recurrent**.

8.1.2.2 Two Dimensions

The $\mathsf{d} = 2$ case turns out to be *marginally* recurrent, as we will see. We already know that R_2 will wander arbitrarily far from the origin, because the projection onto one dimension already does this, as we know from the $\mathsf{d} = 1$ case. However, the question of returning to the origin is somewhat different.

To analyze this case, it is convenient to define

$$L := \log R_2. \tag{8.17}$$

Then the SDE (8.9) for this case,

$$dR_2 = \frac{1}{2R_2}\,dt + dW, \tag{8.18}$$

will simplify as follows. Expanding the logarithm to second order gives

$$dL = \frac{dR_2}{R_2} - \frac{(dR_2)^2}{2R_2^2}, \tag{8.19}$$

which with Eq. (8.18) becomes

$$dL = \frac{1}{2R_2^2}\,dt + \frac{1}{R_2}dW - \frac{dW^2}{2R_2^2} = \frac{dW}{R_2}. \tag{8.20}$$

Then since $R_2 = e^L$,

$$dL = e^{-L}\,dW. \tag{8.21}$$

This transformation simplifies things by removing the drift, at the expense of multiplicative noise. However, we will use the martingale property of this noise, which means that the ensemble average vanishes:

$$d\langle\!\langle L \rangle\!\rangle = 0. \tag{8.22}$$

This means that $\langle\!\langle L \rangle\!\rangle = \langle\!\langle \log R_2 \rangle\!\rangle$ is a constant of the motion. Since we know that L will take on arbitrarily large values, this essentially means that L will also have to take on arbitrarily negative values to keep the mean constant. This means that R_2 will come arbitrarily close to 0.

To make this argument more formal,[1] consider two positive radii $R_<$ and $R_>$ that bound the initial value R_0 of R_2: $R_< \leq R_0 \leq R_>$. Since $\langle\langle L \rangle\rangle$ is constant,

$$\log R_0 = \left\langle\left\langle \log R_2(t) \right\rangle\right\rangle. \tag{8.23}$$

The reasoning that led to Eq. (6.112) applies in a similar way here. Let τ be the first time at which $R_2(t)$ achieves either $R_<$ or $R_>$. Then we can consider a paths up to its stopping time τ, at which point L takes the value $\log R_2(\tau)$; when averaging over all continuations of the path up to time $t > \tau$, the value $\log R(\tau)$ is unchanged. Now extending the average over all paths up to the stopping time, we see that it is equivalent to write Eq. (8.23) in terms of the hitting time as

$$\log R_0 = \left\langle\left\langle \log R_2(\tau) \right\rangle\right\rangle \tag{8.24}$$

after taking the limit $t \longrightarrow \infty$. Now writing the expectation value in terms of probabilities for first hitting $R_<$ or $R_>$,

$$\log R_0 = (\log R_<)\, P[R_2(\tau) = R_<] + (\log R_>)\, P[R_2(\tau) = R_>]. \tag{8.25}$$

Solving for $P[R(\tau) = R_>]$ gives

$$P[R_2(\tau) = R_>] = \frac{\log R_0 - (\log R_<)\, P[R_2(\tau) = R_<]}{\log R_>}. \tag{8.26}$$

Now fixing R_0 and $R_< \leq R_0$, we can let $R_> \longrightarrow \infty$ to obtain

$$P[R_2(\tau) = \infty] = 0. \tag{8.27}$$

That is, the probability for $R_2(t)$ to "hit infinity before hitting $R_<$" (which means *never* hitting $R_<$) is zero. Thus, for any inner bound $R_> > 0$, $R_2(t)$ is guaranteed to hit it in finite time. However, the same argument with fixed $R_>$ and $R_< \longrightarrow 0$ gives that $R_2(t)$ will never hit the origin in finite time. Thus, so to speak, the origin is nonrecurrent, but any disc surrounding the origin is recurrent.

8.1.2.3 Three and More Dimensions

In three and more dimensions, it turns out that Brownian motion is nonrecurrent, and in fact the path wanders "far away" in the sense that $R_d(t) \longrightarrow \infty$. It is sufficient to show this in three dimensions, again because any projection of a d-dimensional motion onto three dimensions will satisfy these conditions. However, it isn't much more difficult to treat any $d \geq 3$ directly.

In this case, the simplifying transformation for the SDE (8.9) has the form

$$P := aR_d^b, \tag{8.28}$$

for constants a and b to be determined. Computing the differential of P (and expanding to second order),

$$dP = abR_d^{b-1}\, dR_d + \frac{ab(b-1)}{2} R_d^{b-2}\, (dR_d)^2, \tag{8.29}$$

and then using Eq. (8.9) for dR_d,

$$dP = \frac{ab(b+d-2)}{2} R_d^{b-2}\, dt + abR_d^{b-1}\, dW. \tag{8.30}$$

We can then force the drift term to vanish by setting

$$b = 2 - d, \tag{8.31}$$

[1] See Ioannis Karatzas and Steven E. Shreve, *Brownian Motion and Stochastic Calculus* (Springer, 1988), pp. 161-3 (doi: 10.1007/978-1-4612-0949-2).

in which case

$$dP = a(2-\mathsf{d})R_d^{1-\mathsf{d}}\,dW. \tag{8.32}$$

Now since $P = aR_d^{2-d}$, we have $R_\mathsf{d} = (P/a)^{1/(2-d)}$, and thus

$$dP = a^{1-c}(2-\mathsf{d})P^c\,dW, \tag{8.33}$$

where

$$c := \frac{1-\mathsf{d}}{2-\mathsf{d}} = \frac{\mathsf{d}-1}{\mathsf{d}-2}. \tag{8.34}$$

Finally, if we choose

$$a = (\mathsf{d}-2)^{1/(1-c)}, \tag{8.35}$$

we end up with the SDE

$$dP = -P^c\,dW, \tag{8.36}$$

which again has the martingale property $\langle\langle dP\rangle\rangle = 0$. In what follows, it is useful to remember that for $\mathsf{d} > 2$, P goes as an *inverse* power of R_d.

As in the previous section, we may take $R_\mathsf{d}(0) = R_0$ to be the initial condition, and define inner and outer boundaries satisfying $R_< \le R_0 \le R_>$, with τ the first passage time through either boundary. Then since

$$P_0 = \big\langle\big\langle P(\tau)\big\rangle\big\rangle, \tag{8.37}$$

this becomes

$$R_0^{2-\mathsf{d}} = \Big\langle\Big\langle R_\mathsf{d}^{2-\mathsf{d}}(\tau)\Big\rangle\Big\rangle \tag{8.38}$$

directly in terms of the radial coordinate. We can rewrite the expectation value as before in terms of the probabilities for first crossing the inner and outer boundaries as

$$R_0^{2-\mathsf{d}} = R_<^{2-\mathsf{d}}\,P[R_\mathsf{d}(\tau) = R_<] + R_>^{2-\mathsf{d}}\,P[R_\mathsf{d}(\tau) = R_>]. \tag{8.39}$$

Now if $R_> \longrightarrow \infty$, the last term vanishes, and

$$P[R_\mathsf{d}(\tau) = R_<] = \left(\frac{R_<}{R_0}\right)^{\mathsf{d}-2}.$$

(probability for achieving inner bound $R_<$) (8.40)

This gives the probability for the path, starting from $R_0 > 0$, to hit $R_< \le R_0$. Note that this probability converges to zero as $R_< \longrightarrow 0$. Thus, the probability for a path starting at $R_0 > 0$ to hit the origin is zero (and thus the origin is nonrecurrent, since once a path moves away from the origin, it will not return). Each path starting at R_0 has a (pathwise) minimum radius, and the probability density for this minimum is given by differentiating (8.40), with the result

$$f(r) = \frac{(\mathsf{d}-2)\,r^{\mathsf{d}-3}}{R_0^{\mathsf{d}-2}}.$$

(probability density for minimum radius) (8.41)

Now to show that $R_\mathsf{d} \longrightarrow \infty$ for $\mathsf{d} \ge 3$, first fix some upper limit $R_>$. We want to show that after some time $t_>$, $R_\mathsf{d}(t) > R_>$ for all $t > t_>$ with probability 1. To demonstrate this, pick a larger boundary $R_{\gg} > R_>$. We know that R_d will hit $R_{\gg}$ in finite time $t_>$. From Eq. (8.40), the probability that the path crosses below $R_>$ is $(R_>/R_{\gg})^{\mathsf{d}-2}$. But this probability converges to zero as $R_{\gg} \longrightarrow \infty$, in which case we are guaranteed that the path does not return below $R_>$.

Chapter 9

Differentiation of Stochastic Processes

9.1 Differentiation of Stochastic Quantities

We have so far studied stochastic processes and ensemble averages over stochastic problems. In cases where we compute quantities like the sojourn time or the local time, analytic expressions are available for the ensemble average, generating function, and so on. But for more complicated statistics, analytic expressions may not be available, and numerical simulations are needed to compute such averages.

Another important class of ensemble-average problems that we have not yet discussed is the *differentiation* of ensemble-averaged statistics. Since Wiener processes are nondifferentiable, the differentation of numeric quantities with underlying Wiener processes can be tricky. Recall, for example, that the sojourn time can be differentiated to yield the local time [see Eq. (7.213)]; this differentiation then carries over to the mean sojourn and local times, but differentiation of the underlying Wiener process causes larger fluctuations in the ensemble-average computation on a pathwise basis [see Eq. (7.219), due to the variation in the number of intersections and the possibility of small values of ΔW_j]. In a further differentiation of the local-time process, the derivative of the ensemble average is well-defined, but the pathwise derivative has arbitrarily large fluctuations in the continuum limit [see Eq. (7.223) and the following discussion].

Thus, it is useful to consider some general approaches to handling derivatives of stochastic processes, which can be handled efficiently in certain cases with a variety of tricks. As a byproduct, we will also briefly study the derivative of a stochastic process with respect to a stochastic quantity, which gives the generalizes variational calculus to stochastic processes. These techniques are widely employed in financial mathematics, where ensemble averages over stochastic trajectories are used to price financial derivatives (e.g., financial options). In such pricing, derivatives of the price with respect to parameters such as the volatility or starting price yields the ''sensitivity'' of the price with respect to parameter variations. These techniques also apply in optimization problems, where the ''payoff'' to optimize is estimated via Monte-Carlo simulations.

9.1.1 Parameter Differentiation and Likelihood Estimation

Consider an ensemble average of the form

$$A(\lambda) := \left\langle\!\!\left\langle \Phi(\mathbf{Z}) \right\rangle\!\!\right\rangle_{\mathbf{Z}}, \qquad (9.1)$$

(model ensemble average)

where $\mathbf{Z} = (Z_1, \ldots, Z_n)$ is a set of stochastic variables, $\Phi(\mathbf{Z})$ is some scalar ''payoff'' function of the stochastic quantities that we wish to average, and the λ on the left-hand side indicates a parameter dependence of the ensemble average. In some important cases we can arrange things such that the only dependence on the right-hand side on the parameter λ is in the probability distribution for $\mathbf{Z}$. That is, we can write the ensemble average as

$$A(\lambda) = \int d\mathbf{z}\, \Phi(\mathbf{z})\, f_\lambda(\mathbf{z}), \qquad (9.2)$$

where $f_\lambda(\mathbf{z})$ is the probability density for $\mathbf{Z}$, and the λ subscript emphasizes the dependence on λ. This approach can work even in some cases where it seems that the payoff function should represent the parameter dependence. For example, suppose we differentiate the sojourn time of Wiener paths with respect to the boundary distance d, which is a part of the functional rather than the paths. However, this quantity can be computed equivalently by differentiating with respect to the starting point of the paths.

Now the derivative can be written simply as

$$\partial_\lambda A(\lambda) = \int d\mathbf{z}\, \Phi(\mathbf{z})\, \partial_\lambda f_\lambda(\mathbf{z}), \tag{9.3}$$

and by multiplying and dividing the integrand by $f_\lambda(\mathbf{z})$, we can rewrite this as an ensemble average:

$$\partial_\lambda A(\lambda) = \left\langle\!\!\left\langle \Phi(\mathbf{Z})\, \frac{\partial_\lambda f_\lambda(\mathbf{Z})}{f_\lambda(\mathbf{Z})} \right\rangle\!\!\right\rangle_{\mathbf{Z}} = \left\langle\!\!\left\langle \Phi(\mathbf{Z})\, \partial_\lambda \log f_\lambda(\mathbf{Z}) \right\rangle\!\!\right\rangle_{\mathbf{Z}}.$$

(likelihood-ratio derivative estimator) (9.4)

Of course, this generalizes readily to higher derivatives,

$$\partial_\lambda^m A(\lambda) = \left\langle\!\!\left\langle \Phi(\mathbf{Z})\, \frac{\partial_\lambda^m f_\lambda(\mathbf{Z})}{f_\lambda(\mathbf{Z})} \right\rangle\!\!\right\rangle_{\mathbf{Z}},$$

(likelihood ratio mth derivative estimator) (9.5)

although of course without the logarithmic form. Thus, the derivative here appears simply as a reweighted version of the original ensemble average (9.1). Of course, ''good'' behavior of this ensemble average is not guaranteed, but the hope is that the logarithmic weight will not cause much in the way of fluctuations. In particular, this approach should give an advantage when the probability density is smooth, whereas the payoff function is not (e.g., it may have a discontinuity or singularity as in the sojourn or local time, such that a small change in the boundary location d can produce a large change in the payoff value). One particularly nice property of this expression is that the weight is universal for *any* payoff function $\Phi(\mathbf{Z})$, because the parameter dependence lies entirely with the probability distribution.

9.1.1.1 Likelihood

Now for a brief digression to explain some terminology. The **likelihood** of a parameter λ, given some observed outcome $\mathbf{Z}$, is defined as the probability of the outcome given the particular parameter value:

$$\mathcal{L}(\lambda|\mathbf{Z}) := P(\mathbf{Z}|\lambda). \tag{9.6}$$

In our ensemble average, the likelihood is just the probability density $f_\lambda(\mathbf{Z})$. Then the ratio of two likelihoods $f_{\lambda'}(\mathbf{Z})/f_\lambda(\mathbf{Z})$ is a common quantity in statistics to evaluate the relative plausibility of two models given observed data. Then the likelihood-ratio derivative $\partial_{\lambda'}[f_{\lambda'}(\mathbf{Z})/f_\lambda(\mathbf{Z})]$ appears in the ensemble average (9.4), so a Monte-Carlo evaluation of Eq. (9.4) is known as a **likelihood-ratio estimator** for the derivative.[1]

9.1.1.2 Application: Differentiation of Brownian-Bridge Path Averages

When computing statistics related to sojourn or local times, or also boundary-crossing and escape probabilities, typically path-averages functional of the form

$$A(x_0) := \left\langle\!\!\left\langle \Phi[V(x)] \right\rangle\!\!\right\rangle_{x(t)}, \tag{9.7}$$

(model ensemble average)

[1] Martin I. Reiman and Alan Weiss, ''Sensitivity analysis via likelihood ratios,'' *Proceedings of the 1986 Winter Simulation Conference*, J. Wilson, J. Henriksen, and S. Roberts, Eds., p. 285 (1986) (doi: 10.1145/318242.318450). P. W. Glynn, ''Stochastic approximation for Monte Carlo optimization,'' *Proceedings of the 1986 Winter Simulation Conference*, J. Wilson, J. Henriksen, and S. Roberts, Eds., p. 356 (1986) (doi: 10.1145/318242.318459). Reuven Y. Rubinstein, ''Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models,'' *Operations Research*, **37**, 72 (1989) (doi: 10.1287/opre.37.1.72). P. W. Glynn, ''Likelihood Ratio Derivative Estimators for Stochastic Systems,'' *Proceedings of the 1989 Winter Simulation Conference*, E. A. MacNair, K. J. Musselman, and P. Heidelberger, Eds., p. 374 (1989) (doi: 10.1109/WSC.1989.718702).

where $x(t)$ are stochastic paths over an evolution time t. In this form, for example, $V(x) = \Theta(x - d)$ for the sojourn time, and $\Phi(x) = e^{-sx}$ computes a moment-generating function or $\Phi(x) = x$ computes a simple average. As noted before, derivatives with respect to a boundary position d can be regarded as derivatives with respect to the initial point x_0 of the path. Thus, we will specifically consider Brownian-bridge paths

$$x(t') = x_0 + B_t(t'), \tag{9.8}$$

such that x_0 represents both the source point, $B_t(0) = x_0$, and the path terminus, $B_t(t) = x_0$. In this case, if we consider the time-sliced path in N steps of duration $\Delta t = t/N$, the probability density of the paths in the average (9.7) have the x_0-dependent factor

$$f[x(t)] \propto e^{-(x_1 - x_0)^2/2\Delta t}\, e^{-(x_0 - x_{N-1})^2/2\Delta t}, \tag{9.9}$$

where recall that by convention we are identifying $x_N = x_0$. Then the derivative (9.4) becomes

$$\frac{\partial_{x_0} f[x(t)]}{f[x(t)]} = \frac{(x_1 - x_0)}{\Delta t} - \frac{(x_0 - x_{N-1})}{\Delta t} = \frac{2(\bar{x}_1 - x_0)}{\Delta t}, \tag{9.10}$$

where we are defining the shorthand

$$\bar{x}_1 := \frac{x_1 + x_{N-1}}{2}. \tag{9.11}$$

Thus, the first derivative of the functional (9.7) becomes

$$\partial_{x_0} A(x_0) = \left\langle\!\!\left\langle \Phi[V(x)]\, \frac{2(\bar{x}_1 - x_0)}{\Delta t} \right\rangle\!\!\right\rangle_{x(t)}.$$

(differentiated ensemble average) (9.12)

Note that $(\bar{x}_1 - x_0) = (\Delta x_1 - \Delta x_{N-1})/2 = O(\Delta t^{1/2})$, so overall the average (9.12) includes an overall weight of magnitude $O(\Delta t^{-1/2}) = O(N^{1/2})$ compared to the undifferentiated expression (9.7).

It is also straightforward to derive an expression for high-order derivatives. First, rewriting Eq. (9.9) as

$$f[x(t)] \propto e^{-(x_0 - \bar{x}_1)^2/\Delta t}\, e^{-(x_{N-1} - x_1)^2/4\Delta t} \tag{9.13}$$

by completing the square, we can define $y := (x_0 - \bar{x}_1)/\sqrt{\Delta t}$, and use the Hermite-polynomial definition

$$H_m(y) := (-1)^m e^{y^2} \partial_y^m e^{-y^2} \tag{9.14}$$

we can write out the derivative

$$\frac{\partial_{x_0}^m f[x(t)]}{f[x(t)]} = (\Delta t)^{-m/2}\, \frac{\partial_y^m f(y)}{f(y)} = (-1)^m\, (\Delta t)^{-m/2}\, H_m\!\left(\frac{x_0 - \bar{x}_1}{\sqrt{\Delta t}}\right) = (\Delta t)^{-m/2}\, H_m\!\left(\frac{\bar{x}_1 - x_0}{\sqrt{\Delta t}}\right). \tag{9.15}$$

Then the derivative of the functional (9.7) becomes

$$\partial_{x_0}^m A(x_0) = (\Delta t)^{-m/2} \left\langle\!\!\left\langle \Phi[V(x)]\, H_m\!\left(\frac{\bar{x}_1 - x_0}{\sqrt{\Delta t}}\right) \right\rangle\!\!\right\rangle_{x(t)}.$$

(differentiated ensemble average) (9.16)

Of course, since $H_1(y) = 2y$, this expression reduces to Eq. (9.12) for $m = 1$. But remarkably, high-order derivatives may be computed merely by reweighting via x_1 and x_{N-1}. Of course, each derivative supplies an additional factor $\Delta t^{-1/2} \propto N^{1/2}$, making the fluctuations larger for each successively higher derivative. Note that the expression (9.16) also applies to ordinary Wiener processes if Δt is replaced by $2\Delta t$, and $\bar{x}_1$ is replaced by x_1.

Although in principle the expression (9.16) applies to arbitrarily high derivatives, it may only practical for relatively small m, depending on the nature of the solution to the path average. The factor of $(\Delta t)^{-m/2}$

is something we can put aside for the moment, since this problem can at least be tamed. Recall that the harmonic-oscillator eigenfunctions have the (normalized) form

$$\psi_m(x) = \frac{1}{\sqrt{2^m m! \sqrt{\pi}}} H_m(x)\, e^{-x^2/2}, \tag{9.17}$$

which grow in width as $\sqrt{m}$. Roughly, this means that the weighting factor in Eq. (9.16) favors increasingly large mean steps $|\bar{x}_1|$ with increasing m. This width can match the original probability density quite poorly for large m, in which case it could be advantageous to absorb the Hermite polynomial into the probability measure for the paths—this avoids rare events in the Gaussian tails from making large contributions. However, doing so introduces a normalization factor that grows exponentially with m. It is only in cases where the ensemble-average derivative grows similarly that the relative accuracy in a Monte-Carlo calculation does not suffer (an average that depends on x_0^{-s} for $s > 0$ is one case where this is okay).

9.1.1.3 Variance Reduction

Now to deal with the factor of $(\Delta t)^{-m/2}$ that causes Eq. (9.16) to become numerically inefficient for larger derivatives. Again, this is because, generally speaking, simulations of path integrals and path averages are accurate in the limit $N \longrightarrow \infty$, but the fluctuations of the path average grow as $N^{m/2}$. This obviously complicates accurate simulation of the path averages difficult.

The key idea is that under certain conditions, the "payoff" functional $\Phi[V(x)]$ is approximately independent of the path samples near the beginning $(x_1, x_2, \ldots)$ and end $(x_{N-1}, x_{N-2}, \ldots)$ of the path. This can happen, for example, if $\Phi[V(x)]$ represents the sojourn-time functional, where the path at times $t' \ll d^2$ (and $t - t' \ll d^2$, where t is the total running time of the bridge) has a very small probability of crossing the boundary. These parts of the path thus contributes very little to the mean sojourn time. Under these conditions, we can successively integrate over the early coordinates $(x_1, x_2, \ldots)$ and the late coordinates $(x_{N-1}, x_{N-2}, \ldots)$ to perform a "partial average" over paths. More specifically, we can hold x_0 and x_2 fixed, while averaging over all possible values of x_1. Then while holding x_3 fixed, average over all possible values of x_2, and so on until the error associated with the averaging is no longer negligible.

To arrive at the appropriate, partially averaged version of Eq. (9.16), consider the integral

$$I := \int dx_j\, H_m\!\left(\frac{x_0 - \bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right) \frac{e^{-(x_0 - x_{N-k})^2/2k\Delta t}}{\sqrt{2\pi k\Delta t}} \frac{e^{-(x_j - x_0)^2/2j\Delta t}}{\sqrt{2\pi j\Delta t}} \frac{e^{-(x_{j+1} - x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}}, \tag{9.18}$$

which consists of the Gaussian and Hermite-polynomial weights for a step of size $j\Delta t$ at the beginning of the path (from x_0 to x_j), and of size $k\Delta t$ (from x_{N-k} to $x_N = x_0$) at the end of the path; a Gaussian weight for the Gaussian step of size Δt from x_j to x_{j+1}; and the partial-averaging integral over x_j. The mean displacement of the first and last steps here is

$$\bar{x}_{jk} := \frac{kx_j + jx_{N-k}}{j+k}. \tag{9.19}$$

Then completing the squares on the first product of Gaussian factors and then the second product, we can

rewrite this as

$$\begin{aligned}
I &= \int dx_j\, H_m\!\left(\frac{x_0-\bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right)\frac{e^{-(x_0-\bar{x}_{jk})^2(j+k)/2jk\Delta t}}{\sqrt{2\pi jk\Delta t/(j+k)}}\,\frac{e^{-(x_{N-k}-x_j)^2/2(j+k)\Delta t}}{\sqrt{2\pi(j+k)\Delta t}}\,\frac{e^{-(x_{j+1}-x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}}\\
&= \int dx_j\, H_m\!\left(\frac{x_0-\bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right)\frac{e^{-(x_0-\bar{x}_{jk})^2(j+k)/2jk\Delta t}}{\sqrt{2\pi jk\Delta t/(j+k)}}\,\frac{e^{-(x_j-\check{x}_{jk})^2(j+k+1)/2(j+k)\Delta t}}{\sqrt{2\pi(j+k)\Delta t/(j+k+1)}}\\
&\qquad\qquad\times\frac{e^{-(x_{N-k}-x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{2\pi(j+k+1)\Delta t}}\\
&= \frac{e^{-(x_{N-k}-x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{8\pi^3jk^3\Delta t^3/(j+k)^2}}\int d\bar{x}_{jk}\, H_m\!\left(\frac{x_0-\bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right)e^{-(x_0-\bar{x}_{jk})^2(j+k)/2jk\Delta t}\\
&\qquad\qquad\times e^{-(x_j-\check{x}_{jk})^2(j+k+1)/2(j+k)\Delta t}\\
&= \frac{e^{-(x_{N-k}-x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{8\pi^3jk^3\Delta t^3/(j+k)^2}}\int d\bar{x}_{jk}\, H_m\!\left(\frac{x_0-\bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right)e^{-(x_0-\bar{x}_{jk})^2(j+k)/2jk\Delta t}\\
&\qquad\qquad\times e^{-(\bar{x}_{jk}-\bar{x}_{(j+1)k})^2(j+k)(j+k+1)/2k^2\Delta t},
\end{aligned}\tag{9.20}$$

where we have defined the offset

$$\check{x}_{jk} := \frac{(j+k)x_{j+1}+x_{N-k}}{j+k+1},\tag{9.21}$$

and we used

$$\begin{aligned}
x_j-\check{x}_{jk} &= \frac{(j+k)\bar{x}_{jk}-jx_{N-k}}{k}-\frac{(j+k)x_{j+1}+x_{N-k}}{j+k+1}\\
&= \frac{(j+k)}{k}\left(\bar{x}_{jk}-\frac{(j+1)}{(j+k+1)}x_{N-k}-\frac{kx_{j+1}}{j+k+1}\right)\\
&= \frac{(j+k)}{k}\left(\bar{x}_{jk}-\bar{x}_{(j+1)k}\right),
\end{aligned}\tag{9.22}$$

where now

$$\bar{x}_{(j+1)k} := \frac{kx_{j+1}+(j+1)x_{N-k}}{j+1+k}.\tag{9.23}$$

Then using the convolution formula[2]

$$e^{-x^2/\alpha^2} * \left[H_n\!\left(\frac{x}{\beta}\right)e^{-x^2/\beta^2}\right] = \frac{\sqrt{\pi}\,\alpha\beta^{n+1}}{(\alpha^2+\beta^2)^{(n+1)/2}}\,H_n\!\left(\frac{x}{\sqrt{\alpha^2+\beta^2}}\right)e^{-x^2/(\alpha^2+\beta^2)},\tag{9.24}$$

with

$$\alpha=\sqrt{\frac{2k^2\Delta t}{(j+k)(j+k+1)}},\qquad \beta=\sqrt{\frac{2jk\Delta t}{(j+k)}}\tag{9.25}$$

such that

$$\alpha^2+\beta^2=\frac{2k(j+1)}{(j+k+1)}\Delta t,\qquad \frac{\beta}{\sqrt{\alpha^2+\beta^2}}=\sqrt{\frac{j(j+k+1)}{(j+1)(j+k)}},\tag{9.26}$$

we find

$$\begin{aligned}
I = \frac{e^{-(x_{N-k}-x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{4\pi^2(j+1)k\Delta t^2}}&\left[\frac{j(j+k+1)}{(j+1)(j+k)}\right]^{m/2}\\
&\times H_m\!\left(\frac{x_0-\bar{x}_{(j+1)k}}{\sqrt{2(j+1)k\Delta t/(j+1+k)}}\right)e^{-(x_0-\bar{x}_{(j+1)k})^2(j+1+k)/2(j+1)k\Delta t}.
\end{aligned}\tag{9.27}$$

[2]Daniel A. Steck, *Classical and Modern Optics* (2006), available online at `http://steck.us/teaching`.

Rearranging to make the recursion more clear, we have

$$\begin{aligned}\left[\frac{j+k}{jk}\right]^{m/2} I &= \left[\frac{j+k+1}{(j+1)k}\right]^{m/2} H_m\!\left(\frac{x_0-\bar{x}_{(j+1)k}}{\sqrt{2(j+1)k\Delta t/(j+1+k)}}\right) \\ &\qquad\qquad \times \frac{e^{-(x_0-\bar{x}_{(j+1)k})^2(j+1+k)/2(j+1)k\Delta t}}{\sqrt{2\pi(j+1)k\Delta t/(j+1+k)}}\,\frac{e^{-(x_{N-k}-x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{2\pi(j+1+k)\Delta t}} \\ &= \left[\frac{j+k+1}{(j+1)k}\right]^{m/2} H_m\!\left(\frac{x_0-\bar{x}_{(j+1)k}}{\sqrt{2(j+1)k\Delta t/(j+1+k)}}\right) \\ &\qquad\qquad \times \frac{e^{-(x_0-x_{N-k})^2/2k\Delta t}}{\sqrt{2\pi k\Delta t}}\,\frac{e^{-(x_{j+1}-x_0)^2/2(j+1)\Delta t}}{\sqrt{2\pi(j+1)\Delta t}}\end{aligned} \tag{9.28}$$

where we are basically "unravelling" the first square that we completed in Eq. (9.20). The same argument may be made for integrating over x_{N-k} by the time-symmetry of the Brownian bridge. Thus the expression that satisfies this recursion and reduces to Eq. (9.16) for $j = k = 1$ is

$$\begin{aligned}\partial_{x_0}^m A(x_0) &\approx \left[\frac{j+k}{2jk\Delta t}\right]^{m/2} \left\langle\!\!\left\langle \Phi[V(x)]\, H_m\!\left(\frac{\bar{x}_{jk}-x_0}{\sqrt{2jk\Delta t/(j+k)}}\right)\right\rangle\!\!\right\rangle_{x(t)} \\ &= \left[\frac{t_1+t-t_2}{2t_1(t-t_2)}\right]^{m/2} \left\langle\!\!\left\langle \Phi[V(x)]\, H_m\!\left(\frac{\bar{x}_{jk}-x_0}{\sqrt{2t_1(t-t_2)/(t_1+t-t_2)}}\right)\right\rangle\!\!\right\rangle_{x(t)},\end{aligned}$$

(differentiated ensemble average, with partial average) (9.29)

as we have just proven by induction, where again

$$\bar{x}_{jk} := \frac{kx_j + jx_{N-k}}{j+k} = \frac{(t-t_2)x(t_1) + t_1 x(t-t_2)}{t_1+t-t_2} \tag{9.30}$$

in both discrete and continuous notation, where $t_1 = j\Delta t$ and $t - t_2 = k\Delta t$, with a total running time t of the Brownian bridge. Again, the expression here is approximate because we are assuming that $\partial_{x_i}\Phi[V(x)] \approx 0$ for all $0 < i < j$ and $N - k < j < N$. In the simpler, symmetric case where $j = k$, we have

$$\begin{aligned}\partial_{x_0}^m A(x_0) &\approx (j\Delta t)^{-m/2} \left\langle\!\!\left\langle \Phi[V(x)]\, H_m\!\left(\frac{\bar{x}_{jj}-x_0}{\sqrt{j\Delta t}}\right)\right\rangle\!\!\right\rangle_{x(t)} \\ &= (t_1/t)^{-m/2} \left\langle\!\!\left\langle \Phi[V(x)]\, H_m\!\left(\frac{\bar{x}_{jj}-x_0}{\sqrt{t_1/t}}\right)\right\rangle\!\!\right\rangle_{x(t)},\end{aligned}$$

(differentiated ensemble average, with symmetric partial average) (9.31)

where

$$\bar{x}_{jj} = \frac{x_j + x_{N-j}}{2} = \frac{x(t_1) + x(t-t_1)}{2}. \tag{9.32}$$

The advantage here is clear, especially in the last expression. In Eq. (9.16), the statistical fluctuations in the path average grew as $N^{m/2}$. But since typically j and k are chosen for some fixed times j and k (e.g., as in the sojourn-time example), the fluctuations here are *independent* of N. Of course, since $t_1 < t/2$, the fluctuations still increase by a factor of $(t/t_1)^{1/2}$ for each derivative, but this is a huge improvement when working with large N.

One remaining detail is to examine the statistics of $\bar{x}_{jk}$ in the Hermite polynomial, to make sure it does not cause any problem. For a discrete Brownian bridge B_n of unit running time, the covariance is given by [Problem 5.6]

$$\langle\!\langle B_n\, B_m \rangle\!\rangle = \frac{[N - \max(m,n)]\min(m,n)}{N^2}, \tag{9.33}$$

which gives

$$\mathrm{Var}(x_j) = \mathrm{Var}(x_{N-j}) = \frac{j(N-j)\Delta t}{N} \tag{9.34}$$

and

$$\mathrm{Cov}(x_j, x_{N-j}) = \frac{j^2\Delta t}{N} \tag{9.35}$$

for the symmetric case. Then setting $x_0 = 0$, we thus have

$$\begin{aligned}\mathrm{Var}[\bar{x}_{jj}] &= \frac{1}{4}\Big[\mathrm{Var}(x_j) + \mathrm{Var}(x_{N-j}) + 2\,\mathrm{Cov}(x_j, x_{N-j})\Big] \\ &= \frac{\Delta t}{2N}\Big[j(N-j) + j^2\Big] \\ &= \frac{j\Delta t}{2}.\end{aligned} \tag{9.36}$$

Thus the typical $\bar{x}_{jj}$ is of the order $\sqrt{j\Delta t}$, which is precisely what we see in the denominator of the Hermite polynomial.

Again, to use this partially averaged path integral as a numerical technique, it can be helpful to incorporate the Hermite polynomial into the sampling distribution for the path itself. To simplify the math somewhat, we will stick to the symmetric expression (9.31) for the path integral. We know from our discussion of the Brownian bridge above that we may choose the path coordinates at times $j\Delta t = t_1$ and $(N-j)\Delta t = t - t_1$, since the path at either time is Gaussian with variance $j(N-j)\Delta t^2 = t_1(t-t_1)$. However, they are correlated because they are part of a bridge, with covariance $j^2\Delta t^2 = t_1^2$. We can decouple this correlation by working with $\bar{x}_{jj}$ and

$$\delta x_{jj} := \frac{x_j - x_{N-j}}{2} = \frac{x(t_1) - x(t-t_1)}{2}, \tag{9.37}$$

such that $x_j = \bar{x}_{jj} + \delta x_{jj}$ and $x_{N-j} = \bar{x}_{jj} - \delta x_{jj}$. We already computed $\mathrm{Var}[\bar{x}_{jj}]$, and we also need

$$\begin{aligned}\mathrm{Var}[\delta x_{jj}] &= \frac{1}{4}\Big[\mathrm{Var}(x_j) + \mathrm{Var}(x_{N-j}) - 2\,\mathrm{Cov}(x_j, x_{N-j})\Big] \\ &= \frac{\Delta t}{2N}\Big[j(N-j) - j^2\Big] \\ &= \frac{j(N-2j)\Delta t}{2N}.\end{aligned} \tag{9.38}$$

Of course, the simplification comes because of the vanishing covariance:

$$\mathrm{Cov}(\bar{x}_{jj}, \delta x_{jj}) = \frac{1}{4}\Big[\mathrm{Var}(x_j) - \mathrm{Var}(x_{N-j})\Big] = 0. \tag{9.39}$$

Thus the ensemble average in the path integral (9.31) implies a probability measure

$$P[x(t)] = \left[\frac{e^{-(\bar{x}_{jj}-x_0)^2/j\Delta t}}{\sqrt{\pi j\Delta t}}\frac{e^{-(\delta x_{jj})^2/j(1-2j/N)\Delta t}}{\sqrt{\pi j(1-2j/N)\Delta t}}\right]\frac{e^{-(x_{j+1}-x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}}\cdots\frac{e^{-(x_{N-j}-x_{N-j-1})^2/2\Delta t}}{\sqrt{2\pi\Delta t}}, \tag{9.40}$$

where the bracketed factor gives the partially averaged first and last steps, and the rest of the factors give the small, in-between steps. Note that there is a factor of 1/2 that is now omitted from the bracketed factor, owing to the choice of $\bar{x}_{jj}$ and δx_{jj} as path variables (before transforming to these variables the corresponding normalization factors had a factor of 2 associated with each π); this factor of 1/2 is absorbed into the integration measure via the Jacobian derivative in the variable transformation. Then the Hermite polynomial may be grouped with the first Gaussian factor

$$\begin{aligned}P_{H_m}[x(t)] = \left\{\left[\frac{\eta_m}{\sqrt{\pi j\Delta t}}\left|H_m\left(\frac{\bar{x}_{jj}-x_0}{\sqrt{j\Delta t}}\right)\right|e^{-(\bar{x}_{jj}-x_0)^2/j\Delta t}\right]\frac{e^{-(\delta x_{jj})^2/j(1-2j/N)\Delta t}}{\sqrt{\pi j(1-2j/N)\Delta t}}\right\} \\ \times\frac{e^{-(x_{j+1}-x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}}\cdots\frac{e^{-(x_{N-j}-x_{N-j-1})^2/2\Delta t}}{\sqrt{2\pi\Delta t}},\end{aligned} \tag{9.41}$$

where the factor

$$\eta_m^{-1} := \frac{1}{\sqrt{\pi j\Delta t}} \int_{-\infty}^{\infty} dx \left| H_m\!\left(\frac{x}{\sqrt{j\Delta t}}\right)\right| e^{-x^2/j\Delta t} = \frac{2}{\sqrt{\pi}} \int_0^{\infty} dx \; |H_m(x)| \; e^{-x^2} \tag{9.42}$$

normalizes the Hermite–Gaussian probability density for $\bar{x}_{jj}$. Then to summarize, we should take

$$\bar{x}_{jj} = x_0 + \sqrt{j\Delta t}\,\bar{z}, \qquad \delta x_{jj} = \sqrt{\frac{j(1-2j/N)\Delta t}{2}}\, z, \tag{9.43}$$

where z is standard normal, and $\bar{z}$ is chosen from the density

$$f(\bar{z}) = \frac{\eta_m}{\sqrt{\pi}} \; |H_m(x)| \; e^{-x^2}. \tag{9.44}$$

Then with the path measure (9.41), the path average (9.31) becomes

$$\partial_{x_0}^m A(x_0) \approx \eta_m^{-1} (j\Delta t)^{-m/2} \left\langle\!\!\left\langle \Phi[V(x)] \operatorname{sgn}\big[H_m(\bar{z})\big] \right\rangle\!\!\right\rangle_{H_m},$$

(differentiated ensemble average, with Hermite–Gauss path) (9.45)

where note that we kept the sign of the Hermite polynomial in the path-average argument rather than the measure, and we have included the factor of $1/2$ that we discussed after Eq. (9.40). Then with the new path measure, x_j and x_{N-j} can be calculated via $x_j = \bar{x}_{jj} + \delta x_{jj}$ and $x_{N-j} = \bar{x}_{jj} - \delta x_{jj}$, while the rest of the path can be constructed as a Brownian bridge connecting x_j to x_{N-j} in time $t - 2t_1$.

9.1.2 Stochastic Differentiation

For a stochastic process $y(t)$ satisfying the SDE

$$dy = \alpha[y(t)]\,dt + \beta[y(t)]\,dW(t), \tag{9.46}$$

the notion of a **stochastic derivative** arises by considering the effect on $y(t)$ of a perturbation to $dW(t_0)$ at some earlier time t_0, where $0 \le t_0 \le t$. The corresponding derivative is typically written $D_{t_0} y(t)$, which for our purposes can be written as the partial derivative[3]

$$D_{t_0} y(t) = \frac{\partial y(t)}{\partial[dW(t_0)]}. \tag{9.47}$$

(stochastic derivative)

Another common, related notation is Dy, which is essentially the collection of all stochastic derivatives $D_{t_0} y(t)$. That is, considering t to be fixed, Dy is a function of t_0, with $0 \le t_0 \le t$.

This derivative generalizes the notion of a functional derivative to stochastic processes. Thus, Dy is essentially the functional derivative $\delta y/\delta(dW)$, regarding $\delta(dW)(t)$ as a perturbation to $dW(t)$. And while Dy acts as a gradient, $D_{t_0} y$ acts as a directional derivative in the "direction" $\delta(t - t_0)$, in the same sense as the functional derivative via a perturbation only at time t_0: $\delta y/\delta(dW)(t_0)$.

9.1.2.1 Example: Stochastic Derivative of the Wiener Process

As a first example of stochastic differentiation, we will begin by computing $D_{t_0} W(t)$. Since

$$W(t) = \int_0^t dW(t'), \tag{9.48}$$

[3]See Arturo Kohatsu–Higa and Miquel Montero, "Malliavin Calculus in Finance," in *Handbook of Computational and Numerical Methods in Finance* Svetlozar T. Rachev, Ed. (Springer, 2004), pp. 111-74, especially p. 130. This derivative is more generally known as the **Malliavin derivative**. For a more rigorous and sophisticated development, see David Nualart, *The Malliavin Calculus and Related Topics*, 2nd ed. (Springer, 2006), Section 1.2, p. 24.

the function to differentiate is a simple sum of increments $dW(t')$ and we are differentiating with respect to one of them. Then we can use

$$\frac{\partial[dW(t)]}{\partial[dW(t')]} = dt\,\delta(t-t'), \tag{9.49}$$

as a consequence of the independence of $dW(t)$ at different times: that is, the derivative is unity if $t = t'$, and zero otherwise. Thus,

$$D_{t_0}W(t) = \int_0^t D_{t_0}dW(t') = \int_0^t dt'\,\delta(t_0-t'), \tag{9.50}$$

or simply

$$D_{t_0}W(t) = 1, \tag{9.51}$$

provided, of course, $0 \le t_0 \le t$. Otherwise this derivative vanishes.

9.1.2.2 Example: Stochastic Derivative of Additive Diffusion

As a slightly more sophisticated example, let's now compute Dy, where

$$dy(t) = \beta(t)\,dW(t). \tag{9.52}$$

We can do this by generalizing the solution from the previous section, which leads straightforwardly to

$$D_{t_0}y(t) = \beta(t_0). \tag{9.53}$$

Thus, a simple relabeling gives the result[4] $Dy = \beta$.

9.1.2.3 Variation Process

Going back to the stochastic derivative $D_{t_0}y(t)$, where $y(t)$ satisfies the prototype SDE (9.46), we can think of the derivative *itself* as a stochastic process (considering t to be the time variable, with fixed t_0). We can refer to this as the **derivative process** or the **variation process**. This describes the modified evolution of $y(t)$ as a linearized displacement, given a perturbation to $dW(t_0)$. That is, the perturbed evolution is given as $y(t) = y_0(t) + [D_{t_0}y_0(t)]\,\delta(dW)(t_0)$, where $y_0(t)$ is the unperturbed solution, and $\delta(dW)(t_0)$ is the perturbation to $dW(t_0)$.

Taking the stochastic derivative of Eq. (9.46) by applying the D_{t_0} operator, we find

$$d[D_{t_0}y(t)] = \alpha'[y(t)]\,[D_{t_0}y(t)]\,dt + \beta'[y(t)]\,[D_{t_0}y(t)]\,dW(t) + \beta[y(t)]\,D_{t_0}dW(t). \tag{9.54}$$

Using Eq. (9.49) again in the last term, and dividing through by $D_{t_0}y(t)$, we arrive at the SDE

$$\frac{d[D_{t_0}y(t)]}{D_{t_0}y(t)} = \alpha'[y(t)]\,dt + \beta'[y(t)]\,dW(t) + \frac{\beta[y(t_0)]}{D_{t_0}y(t)}\,\delta(t-t_0)\,dt, \tag{9.55}$$

(SDE for variation process)

which yields the evolution of the variation processs.

To write down an integral solution for this SDE, the idea is now to integrate Eq. (9.55) from $t' = 0$ to t, which gives the perturbation to $y(t)$ due to the perturbation to $dW(t_0)$ (where $0 \le t_0 \le t$). Of course, there is no effect for $t' \le t_0$, so we may take $D_{t_0}y(t') = 0$ for all $t' < t_0$. The delta function induces a jump discontinuity in $D_{t_0}y(t')$ at $t' = t_0$ of height $\beta[y(t_0)]$. Thus, the task is to integrate the homogeneous differential equation

$$\frac{d[D_{t_0}y(t)]}{D_{t_0}y(t)} = \alpha'[y(t)]\,dt + \beta'[y(t)]\,dW(t) \tag{9.56}$$

[4] Compare to the equivalent statement $DW(h) = h$ in the cryptic notation of Nualart, *op. cit.*, p. 25, just after Eq. (1.29), where it is not obvious that $W(h)$ can be interpreted as the integral of $h\,dW$, but see Example 1.1 on p. 3 in the corresponding lecture notes at `http://www.math.wisc.edu/~kurtz/NualartLectureNotes.pdf`; for a more direct comparison, see Giulia Di Nunno, Bernt Øksendal, and Frank Proske, *Malliavin Calculus for Lévy Processes with Applications to Finance* (Springer, 2009), p. 29, Eq. (3.8).

from $t' = t_0$ to t, with initial condition $D_{t_0} y(t' = t_0) = \beta[y(t_0)]$. Rewriting the left-hand side of the homogeneous SDE,

$$d\log[D_{t_0} y(t)] + \frac{\left\{d[D_{t_0} y(t)]\right\}^2}{2[D_{t_0} y(t)]^2} = \alpha'[y(t)]\, dt + \beta'[y(t)]\, dW(t), \tag{9.57}$$

where from the second-order expansion $d\log x = \log(x + dx) - \log x = dx/x - (dx)^2/2x^2$, it follows that $dx/x = d\log x + (dx)^2/2x^2$. Using the SDE (9.56) to evaluate the second-order term, we find

$$d\log[D_{t_0} y(t)] = \alpha'[y(t)]\, dt - \frac{\beta'^2[y(t)]}{2}\, dt + \beta'[y(t)]\, dW(t). \tag{9.58}$$

Integrating from t_0 to t,

$$\log[D_{t_0} y(t)] - \log \beta[y(t_0)] = \int_{t_0}^{t} \left(\alpha'[y(t')] - \frac{\beta'^2[y(t')]}{2}\right) dt' + \int_{t_0}^{t} \beta'[y(t')]\, dW(t'). \tag{9.59}$$

Exponentiating this equation leads to the expression

$$D_{t_0} y(t) = \beta[y(t_0)] \exp\left[\int_{t_0}^{t} \left(\alpha'[y(t')] - \frac{\beta'^2[y(t')]}{2}\right) dt' + \int_{t_0}^{t} \beta'[y(t')]\, dW(t')\right], \tag{9.60}$$

(solution for variation process)

which solves the SDE (9.55).

9.1.3 Integration by Parts

The notion of a stochastic derivative also gives a kind of integration by parts, which can be helpful in transforming certain path averages. As an example, consider the model path average

$$A(c) = \left\langle\!\!\left\langle \delta[W(t) - c] \right\rangle\!\!\right\rangle_{W(t)}. \tag{9.61}$$

(model ensemble average)

That is, we are averaging over Wiener paths $W(t)$, which only "score" when the endpoint $W(t)$ matches the desired ending point c. Since we know the explicit probability density for $W(t)$, we can compute this explicitly as

$$A(c) = \int_{-\infty}^{\infty} dx\, \frac{e^{-x^2/2t}}{\sqrt{2\pi t}}\, \delta(x - c) = \frac{e^{-c^2/2t}}{\sqrt{2\pi t}}. \tag{9.62}$$

However, as we will see, the general idea here will apply to more general stochastic processes, where we may not have an analytic expression for the density of paths.

From the standpoint of numerical simulation, Eq. (9.61) is very difficult, because only a subset of paths of zero measure contribute at all to the average, each with an infinite "score." In a numerical simulation, the delts function could, for example, be changed into a function of finite height and width, but the variance of a sample path average would still be very large for any function narrow enough to give an accurate result. Of course, one can also change explicitly to Brownian bridges that run to $W(t) = c$ as in Eq. (7.155), but this presumes that we know the distribution of $W(t)$. In this case we do, but we would like to explore more general methods for cases where the distribution is not known.

The approach we will try here is to rewrite the path functional as

$$A(c) = -\frac{1}{2}\partial_c \left\langle\!\!\left\langle \mathrm{sgn}[W(t) - c] \right\rangle\!\!\right\rangle_{W(t)}, \tag{9.63}$$

which removes the delta function at the expense of introducing a derivative. Now switching to shifted Brownian bridges

$$x(t) = x_0 + W(t) \tag{9.64}$$

with source point x_0, we can write the path functional as

$$A(c) = -\frac{1}{2}\partial_c \Big\langle\!\Big\langle \mathrm{sgn}[x(t) - c] \Big\rangle\!\Big\rangle_{x(t)}\Big|_{x_0=0} = \frac{1}{2}\partial_{x_0} \Big\langle\!\Big\langle \mathrm{sgn}[x(t) - c] \Big\rangle\!\Big\rangle_{x(t)}\Big|_{x_0=0}, \tag{9.65}$$

Then using the likelihood-ratio estimator (9.4),

$$A(c) = \frac{1}{2}\Big\langle\!\Big\langle \mathrm{sgn}[x(t) - c]\, \partial_{x_0} \log f[x(t)] \Big\rangle\!\Big\rangle_{x(t)}\Big|_{x_0=0}, \tag{9.66}$$

where $f[x(t)]$ is the probability density of the paths $x(t)$, which we regard as containing all the dependence on x_0. For Wiener paths,

$$\log f[x(t)] = -\frac{[x(t) - x_0]^2}{2t} - \log\sqrt{2\pi t}, \tag{9.67}$$

and thus

$$\partial_{x_0} \log f[x(t)] = -\frac{x(t) - x_0}{t} = -\frac{W(t)}{t}. \tag{9.68}$$

Then Eq. (9.66) becomes

$$A(c) = \frac{1}{2t}\Big\langle\!\Big\langle \mathrm{sgn}[W(t) - c]\, W(t) \Big\rangle\!\Big\rangle_{W(t)}. \tag{9.69}$$
(alternative path average)

We can verify by direct integration with respect to the probability desnity that this gives the correct answer. However, unlike the original path functional (9.61), *every* path $W(t)$ contributes value to the average. The ''cost'' is a weight of $W(t)/2t$ in addition to the integral of the delta function, but everything is well-behaved here from a numerical standpoint. Note that other antiderivatives of the delta function are possible: the Heaviside function is an obvious choice, but then half of the paths contribution no ''information'' to the path average.

However, in deriving Eq. (9.69), we again used the explicit probability density for $W(t)$, as an illustration of the method. To perform a more general calculation, we cannot do this. Consider now the model path integral with more general stochastic processes:

$$\begin{aligned} A(c) &= \Big\langle\!\Big\langle \delta[y(t) - c] \Big\rangle\!\Big\rangle_{y(t)} \\ dy &= \alpha[y(t)]\, dt + \beta[y(t)]\, dW(t). \end{aligned} \tag{9.70}$$
(model ensemble average)

To replace the delta function here, first we will compute the chain-rule derivative of the step function:

$$\frac{1}{2} D_{t_0} \mathrm{sgn}[y(t) - c] = \delta[y(t) - c]\, D_{t_0} y(t). \tag{9.71}$$

Solving for the delta function and integrating over time gives

$$\delta[y(t) - c] = \frac{1}{2t}\int_0^t dt_0\, \frac{D_{t_0}\mathrm{sgn}[y(t) - c]}{D_{t_0} y(t)}. \tag{9.72}$$

Now if we include an explicit probability measure for $dW(t_0)$, we can integrate by parts with respect to $dW(t_0)$ to write

$$\int dW(t_0)\, \frac{e^{-dW^2(t_0)/2dt}}{\sqrt{2\pi dt}} D_{t_0}\mathrm{sgn}[y(t) - c] = \int dW(t_0)\, \frac{e^{-dW^2(t_0)/2dt}}{\sqrt{2\pi dt}} \frac{dW(t_0)}{dt} \mathrm{sgn}[y(t) - c]. \tag{9.73}$$

Thus, *under an ensemble average* integration by parts amounts to writing

$$D_{t_0}\mathrm{sgn}[y(t) - c] = \frac{dW(t_0)}{dt}\, \mathrm{sgn}[y(t) - c]. \tag{9.74}$$

If we then introduce an expectation value in Eq. (9.72) and perform this integration by parts, we find

$$\Big\langle\!\Big\langle \delta[y(t)-c] \Big\rangle\!\Big\rangle = \frac{1}{2t}\int_0^t dt_0 \left\langle\!\!\left\langle \frac{D_{t_0}\mathrm{sgn}[y(t)-c]}{D_{t_0}y(t)} \right\rangle\!\!\right\rangle = \frac{1}{2t}\int_0^t dt_0 \left\langle\!\!\left\langle \frac{dW(t_0)}{dt}\frac{\mathrm{sgn}[y(t)-c]}{D_{t_0}y(t)} \right\rangle\!\!\right\rangle, \tag{9.75}$$

noting that $D_{t_0}y(t)$ is independent of $dW(t_0)$. Simplifying, we have[5]

$$A(c) = \frac{1}{2t}\left\langle\!\!\left\langle \mathrm{sgn}[y(t)-c]\int_0^t \frac{dW(t_0)}{D_{t_0}y(t)} \right\rangle\!\!\right\rangle, \tag{9.76}$$
(alternate path functional)

where the derivative in the denominator is given by Eq. (9.60). Note that in the limit $\alpha = 0$ and $\beta = 1$, the derivative reduces to $D_{t_0}y(t) = D_{t_0}W(t) = 1$, and the path integral here reduces to the simpler one in Eq. (9.69). This expression also simplifies somewhat for the case of constant diffusion ($\beta = 1$), in the sense that the derivative (9.60) in the denominator of the integral reduces to the simpler form

$$D_{t_0}y(t) = \exp\left[\int_{t_0}^t dt'\,\alpha'[y(t')]\right], \tag{9.77}$$

so that Eq. (9.76) becomes

$$A(c) = \frac{1}{2t}\left\langle\!\!\left\langle \mathrm{sgn}[y(t)-c]\int_0^t dW(t_0)\,\exp\left[-\int_{t_0}^t dt'\,\alpha'[y(t')]\right] \right\rangle\!\!\right\rangle, \tag{9.78}$$

such that a drifting path still introduces a nontrivial weighting function that measures the drift-induced focusing or divergence of paths.

9.1.3.1 Digression: Faddeev–Popov Approach to Conditional Averages

An alternative approach, based on the method of Faddeev and Popov, can work to give regularized expressions for the conditional path averages in the previous section. To start with the simpler case of Eq. (9.61),

$$A(c) = \Big\langle\!\Big\langle \delta[W(t)-c] \Big\rangle\!\Big\rangle_{W(t)}, \tag{9.79}$$
(model ensemble average)

as a first step, let's write out the probability measure explicitly in discrete form with N path steps:

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2}\int d\Delta W_0 \cdots d\Delta W_{N-1}\,\exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\big(\Delta W_j\big)^2\right]\delta\big[W(t)-c\big]. \tag{9.80}$$

Note that we can regard $W(t)$ here as an explicit sum:

$$W(t) = \sum_{j=0}^{N-1}\Delta W_j. \tag{9.81}$$

Now the key is to note that because we are integrating over all possible values of ΔW_j, the integral $A(c)$ is invariant if we shift the variable ΔW_j. In particular, suppose that we introduce a scalar quantity ω, and shift each coordinate according to

$$\Delta W_j \longrightarrow \Delta W_j - \frac{\omega}{N}. \tag{9.82}$$

[5] A more general version of this argument is given in Peter K. Friz, "Malliavin Calculus in Finance," Section 7.1, available at `http://www.math.nyu.edu/phd_students/frizpete/finance/my_case_lecture/malliavin_lecture.pdf`. See also Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, and Pierre-Louis Lions, "Applications of Malliavin calculus to Monte-Carlo methods in finance. II," *Finance and Stochastics* **5**, 201 (2001), Section 4.1 (doi: 10.1007/PL00013529).

Then $A(c)$ is independent of ω, and we can regard this shift as a kind of *gauge transformation* with gauge parameter ω. Implementing this gauge transformation, Eq. (9.80) becomes

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{\omega}{N}\right)^2\right] \delta\big[W(t) - c - \omega\big]. \tag{9.83}$$

Now suppose that we introduce some function $g(\omega)$, with the only requirement that it be normalized such that

$$\int d\omega \, g(\omega) = 1. \tag{9.84}$$

Then we don't change the value of Eq. (9.83) by multiplying by $g(\omega)$ and integrating, thus taking a linear combination of the same value. The delta function disappears in the result:

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{W(t) - c}{N}\right)^2\right] g\big[W(t) - c\big]. \tag{9.85}$$

In the exponentiated summand, we can multiply out the quadratic factor to obtain

$$\left(\Delta W_j - \frac{W(t) - c}{N}\right)^2 = \big(\Delta W_j\big)^2 - 2\Delta W_j \frac{\big[W(t) - c\big]}{N} + \frac{\big[W(t) - c\big]^2}{N^2}, \tag{9.86}$$

and thus when summed over j, we find

$$\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{W(t) - c}{N}\right)^2 = \sum_{j=0}^{N-1}\big(\Delta W_j\big)^2 + \frac{c^2}{N} - \frac{\big[W(t)\big]^2}{N}. \tag{9.87}$$

In this case the path integral (9.85) becomes

$$A(c) = \left\langle\!\!\left\langle \exp\left[\frac{\big[W(t)\big]^2 - c^2}{2t}\right] g\big[W(t) - c\big] \right\rangle\!\!\right\rangle_{W(t)} \tag{9.88}$$

after again hiding the path measure in the ensemble average. At this point we are free to choose $g(\omega)$ in order to simplify or improve the statistical behavior of this ensemble average. For example, if we choose $g(\omega)$ to be a Gaussian of variance t,

$$g(\omega) = \frac{1}{\sqrt{2\pi t}} \, e^{-\omega^2/2t}, \tag{9.89}$$

then the path average (9.88) becomes

$$A(c) = \frac{e^{-c^2/t}}{\sqrt{2\pi t}} \left\langle\!\!\left\langle e^{cW(t)/t} \right\rangle\!\!\right\rangle_{W(t)}, \tag{9.90}$$

(alternate path integral)

which attains the correct value of $e^{-c^2/2t}/\sqrt{2\pi t}$. On the other hand, if we take $g(\omega)$ to be centered at $-c$ instead of at 0,

$$g(\omega) = \frac{1}{\sqrt{2\pi t}} \, e^{-(\omega+c)^2/2t}, \tag{9.91}$$

the path integral (9.88) takes on the average value directly, with zero variance. This illustrates the advantage of adapting $g(\omega)$ to the problem as well as can possibly be done.

Of course, this approach becomes more complicated for the more general diffusion problem (9.70):

$$\begin{aligned} A(c) &= \left\langle\!\!\left\langle \delta[y(t) - c] \right\rangle\!\!\right\rangle_{y(t)} \\ dy &= \alpha[y(t)]\, dt + \beta[y(t)]\, dW(t). \end{aligned} \tag{9.92}$$

(model ensemble average)

The general approach, however, will be the same. Writing out the details of the path integral, we have

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j\right)^2\right] \delta\big[y(t) - c\big] \tag{9.93}$$

when integrating in terms of the unshifted path variables, where

$$y(t) = \int_0^t dt' \, \alpha\big[y(t')\big] + \int_0^t dW(t') \, \beta\big[y(t')\big]. \tag{9.94}$$

Implementing the same gauge transformation (9.82), the result is

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{\omega}{N}\right)^2\right] \delta\left[y(t) - c - \frac{\omega}{N}\sum_{j=0}^{N-1}\frac{\partial y(t)}{\partial \Delta W_j}\right], \tag{9.95}$$

where in the delta function we have expanded to $O(N^{-1})$. Then we can define the shorthand for the last term in the argument of the delta function,

$$\mathcal{A} := \frac{1}{N}\sum_{j=0}^{N-1}\frac{\partial y(t)}{\partial \Delta W_j} = \frac{1}{t}\int_0^t dt_0 \, D_{t_0} y(t), \tag{9.96}$$

in both discrete and continuous notation. This quantity is written more explicitly using the expression (9.60) for the Malliavin derivative, which in the case of constant diffusion ($\beta = 1$), for example, simplifies via Eq. (9.77) to give

$$\mathcal{A} = \frac{1}{t}\int_0^t dt_0 \, \exp\left[\int_{t_0}^t dt' \, \alpha'[y(t')]\right]. \tag{9.97}$$

Then we can proceed by introducing the function $g(\omega)$ and integrating over all ω, with the result

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{y(t)-c}{N|\mathcal{A}|}\right)^2\right] \frac{1}{|\mathcal{A}|}\, g\left[\frac{y(t)-c}{|\mathcal{A}|}\right], \tag{9.98}$$

where we have used $\delta(ax) = \delta(x)/|a|$. Multiplying out the exponentiated factor and hiding the path measure again gives

$$A(c) = \left\langle\!\!\left\langle \frac{1}{|\mathcal{A}|} \, \exp\left[\frac{[y(t)-c]}{2t|\mathcal{A}|}\left(2W(t) - \frac{[y(t)-c]}{|\mathcal{A}|}\right)\right] g\left[\frac{y(t)-c}{|\mathcal{A}|}\right] \right\rangle\!\!\right\rangle_{y(t)}. \tag{9.99}$$

Again, we can choose the form of $g(\omega)$ to simplify the problem, but for a general drift function $\alpha[y]$, we cannot choose it to arrive directly at the answer. Choosing the gaussian form (9.89) again gives the result

$$A(c) = \frac{1}{\sqrt{2\pi t}}\left\langle\!\!\left\langle \frac{1}{|\mathcal{A}|} \, \exp\left[\frac{[y(t)-c]}{t|\mathcal{A}|}\left(W(t) - \frac{[y(t)-c]}{|\mathcal{A}|}\right)\right] \right\rangle\!\!\right\rangle_{y(t)}, \tag{9.100}$$

which does not make the path integral particularly simple, but at least it casts the path contribution in terms of a direct difference between $W(t)$ and path $y(t)/|A|$ in the case $c = 0$. Note that in the case $\alpha = 0$, then $y(t)$ reduces to $W(t)$ and $A = 1$, so that Eq. (9.100) reduces to Eq. (9.90). Of course, we can also make the probably better choice (9.91), which leads to

$$A(c) = \frac{e^{-c^2/2t}}{\sqrt{2\pi t}}\left\langle\!\!\left\langle \frac{1}{|\mathcal{A}|} \, \exp\left[\frac{[y(t)-c]}{t|\mathcal{A}|}\left([W(t)-c] - \frac{[y(t)-c]}{|\mathcal{A}|}\right)\right] \right\rangle\!\!\right\rangle_{y(t)}, \tag{9.101}$$

(alternate path integral)

which again reduces directly to the mean value (with no variance among the paths) in the case $\alpha = 0$. Note in the path integrals how $y(t)$ is accompanied by a factor of $\mathcal{A}$; intuitively, from the discussion in Section 9.1.2.3, this factor gives a time-averaged measure of the divergence of trajectories in the vicinity of $y(t)$, and thus accounts for any focusing or defocusing effects due to gradients in $\alpha(y)$.

Problem 9.1

Work out an analytic expression for the path integral (9.92) in the special case $\alpha(y) = -ay$,

$$\begin{aligned} A(c) &= \left\langle\!\!\left\langle \delta[y(t) - c] \right\rangle\!\!\right\rangle_{y(t)} \\ dy &= -a\, y(t)\, dt + dW(t), \end{aligned} \tag{9.102}$$

under the initial condition $y(0) = 0$, and verify numerically that the Faddeev–Popov form (9.101) of the path integral converges to the same value.

Chapter 10

Gaussian Numerical Methods

Now we will consider the numerical solution to *stochastic* differential equations (SDEs) of the form

$$dy = \alpha(y,t)\,dt + \beta(y,t)\,dW, \tag{10.1}$$

where $dW(t)$ is the Wiener process as usual (see Chapter 3). We will stick to considering only Itō SDEs for simplicity. Because of the unusual and singular nature of $dW(t)$, standard methods that apply to *deterministic* ordinary differential equations (ODEs) do not work well here. Thus, specialized methods for SDEs are needed.[1]

10.1 Stochastic Euler Method

The simplest numerical method for solving SDEs is the **Euler method**. This is also the simplest method for solving deterministic ODEs as well, and for the SDE (10.1) when $\beta = 0$, the deterministic update equation is

$$y_{n+1} = y_n + \alpha(y_n, t_n)\,\Delta t + O(\Delta t^2), \tag{10.2}$$

where again the solution is evolved in finite time steps of Δt, and we are using the condensed notation $y_n := y(n\Delta t)$ and $t_n := n\Delta t$. We can try extending this method to the SDE (10.1) by taking the same linear approximation to the stochastic term, to arrive at the **stochastic Euler method** (often called the **Euler–Maruyama method**):[2]

$$y_{n+1} = y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\,\Delta W_n. \tag{10.3}$$
(stochastic Euler method)

We are defining the time increment as before, and the Wiener increment ΔW_n is defined in an analogous way:

$$\begin{aligned} \Delta t_n &:= \int_{t_n}^{t_{n+1}} dt' \\ \Delta W_n &:= \int_{t_n}^{t_{n+1}} dW(t'). \end{aligned} \tag{10.4}$$
(time increments)

Of course, $\Delta t_n \equiv \Delta t$ is independent of n, so we need not indicate explicit time dependence. The key, however, is that the approximation works when α and β vary slowly over the interval $[t_n, t_{n+1}]$, and thus

[1] An excellent references on the subject is Peter E. Kloeden and Eckhard Platen, *Numerical Solution of Stochastic Differential Equations*, 3rd ed. (Springer, 2000).

[2] Gisirō Maruyama, ‘‘Continuous Markov Processes and Stochastic Equations,’’ *Rendiconti del Circolo Matematico di Palermo* **4**, 48 (1955).

when we try to compute the solution

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} dt'\, \alpha(y(t'), t') + \int_{t_n}^{t_{n+1}} dW(t')\, \beta(y(t'), t'), \tag{10.5}$$

we can, to first approximation, treat α and β as constants and pull them out of their respective integrals, which precisely yields Eq. (10.3).

10.1.1 Truncation Error

Recall the Itō chain rule (3.118) for a function $f(y)$, where $y(t)$ satisfies the SDE (10.1):

$$df(y) = \left[f'(y)\, \alpha(y,t) + \frac{1}{2} f''(y)\, \beta^2(y,t) \right] dt + f'(y)\, \beta(y,t)\, dW. \tag{10.6}$$

Integrating this expression from $t' = t_0$ to t,

$$f(y(t)) = f(y(t_0)) + \int_{t_0}^{t} \left[f'(y)\, \alpha(y,t') + \frac{1}{2} f''(y)\, \beta^2(y,t') \right] dt' + \int_{t_0}^{t} f'(y)\, \beta(y,t')\, dW(t'). \tag{10.7}$$

Now we can let $f(y) \longrightarrow \alpha(y,t)$, $f(y) \longrightarrow \beta(y,t)$, and $t_0 \longrightarrow t_n$, and then put the two resulting expressions in Eq. (10.5) to obtain

$$\begin{aligned} y_{n+1} = y_n &+ \alpha(y_n, t_n)\, \Delta t + \beta(y_n, t_n)\, \Delta W_n \\ &+ \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dt'' \left[\alpha'(y,t'')\, \alpha(y,t'') + \frac{1}{2} \alpha''(y,t'')\, \beta^2(y,t'') \right] + \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dW(t'')\, \alpha'(y,t'')\, \beta(y,t'') \\ &+ \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dt'' \left[\beta'(y,t'')\, \alpha(y,t'') + \frac{1}{2} \beta''(y,t'')\, \beta^2(y,t'') \right] \\ &+ \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t'')\, \beta'(y,t'')\, \beta(y,t''). \end{aligned} \tag{10.8}$$

The primes here are equivalent to the partial derivatives ∂_y. Again, we can pretend that the α and β functions are constant over the short time interval Δt. Then we have simple integrals of the form

$$\int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dt''; \qquad \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dW(t''); \qquad \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dt''; \qquad \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t''), \tag{10.9}$$

to deal with. The first integral is just $\Delta t^2/2$, or for our purposes, simply $O(\Delta t^2)$. Similarly, if we proceed with counting each ΔW on average as equivalent to $\Delta t^{1/2}$, the second and third integrals are $O(\Delta t^{3/2})$, and the last integral is $O(\Delta t^1)$. Clearly, any error in treating α and β as constants will result in higher-order errors, so the truncation error here is $O(\Delta t)$. Thus, we may again write the stochastic Euler method as

$$y_{n+1} = y_n + \alpha(y_n, t_n)\, \Delta t + \beta(y_n, t_n)\, \Delta W_n + O(\Delta t). \tag{10.10}$$
(stochastic Euler method)

The local truncation error here is, as we see, much worse than the deterministic Euler method (10.2). To estimate the *global* truncation error, again suppose we use this method to evolve the solution out to a fixed final time t in N steps of $\Delta t = t/N$. The local truncation error on each step is $O(\Delta t)$, due essentially to the stochastic term. The dominant errors will thus be random and uncorrelated, so we consider the error in a random walk of N steps of mean size of order t/N, which will scale as $\sqrt{N}(t/N) = t/\sqrt{N} = O(\Delta t^{1/2})$. Thus the global error of the stochastic Euler scheme converges very badly, as $O(\Delta t^{1/2})$. This is indicative of the fact that stochastic differential equations are much more difficult to solve than the deterministic counterparts. In any case, as in the ODE case, we call this method an $O(\Delta t^{1/2})$ method, because the local truncation is correct to this order, or equivalently, the global error is $O(\Delta t^{1/2})$.

10.2 Milstein Method

To construct the next, relatively simple, higher-order method, we note that the only term we neglected in Eq. (10.8) that was $O(\Delta t)$ was the last one, involving the double integral

$$\begin{aligned}\int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t'') &= \int_{t_n}^{t_{n+1}} dW(t')\,[W(t') - W(t_n)] \\ &= \frac{1}{2}\int_{t_n}^{t_{n+1}} \left\{d[W(t')]^2 - dt\right\} - W(t_n)\int_{t_n}^{t_{n+1}} dW(t') \\ &= \frac{1}{2}\left[W^2(t_{n+1}) - W^2(t_n)\right] - \frac{\Delta t}{2} - W(t_n)\Delta W(t_n) \\ &= \frac{1}{2}\left[W(t_{n+1}) + W(t_n)\right]\Delta W(t_n) - \frac{\Delta t}{2} - W(t_n)\Delta W(t_n) \\ &= \frac{1}{2}\left[W(t_{n+1}) - W(t_n)\right]\Delta W(t_n) - \frac{\Delta t}{2} \\ &= \frac{1}{2}\Big(\Delta W_n^2 - \Delta t\Big),\end{aligned} \tag{10.11}$$

where we used the Itō rule $d(W^2) = 2W\,dW + dW^2 = 2W\,dW + dt$ in the second step. The idea behind the **Milstein method**[3] is to keep this correction term, keeping in mind that the factor we just worked out is multiplied by $\beta(y,t_n)\,\beta'(y,t_n)$, which we are treating as constant to this order of approximation:

$$y_{n+1} = y_n + \alpha(y_n,t_n)\,\Delta t + \beta(y_n,t_n)\,\Delta W_n + \frac{1}{2}\beta(y_n,t_n)\,\beta'(y_n,t_n)\left(\Delta W_n^2 - \Delta t\right) + O(\Delta t^{3/2}). \tag{10.12}$$

(Milstein method)

Again, the prime refers to partial differentiation with respect to y_n. Clearly the local truncation error is $O(\Delta t^{3/2})$, since we have ignored the two integrals in Eqs. (10.9) of the same order. By the same argument as for the Euler method, the global error is a factor of $\Delta t^{1/2}$ worse, or $O(\Delta t)$, which is the same global error as the *deterministic* Euler method.

10.2.1 Multiplicative vs. Additive Noise

Note that the correction term in the Milstein method (10.12) is of the form $\beta\beta'$, and thus is only required for *multiplicative* noise (where β is a function of y). For *additive* noise (β independent of y), the correction term vanishes, and the stochastic Euler method has $O(\Delta t)$ global error. In general, we expect additive noise to be easier to solve numerically than multiplicative noise, and one strategy to improve the accuracy of numerical methods is to use the Lamperti transform (Section 5.1.4.2) to change a multiplicative-noise process to an additive process, assuming that the transform can be inverted to give the original solution.

10.3 Stochastic Taylor Expansion

What we have written down in Eq. (10.8) is something like a Tayler expansion for y_{n+1} in terms of y_n. Again, treating α and β as constants, essentially what we have is a Taylor expansion, neglecting $O(\Delta t^{3/2})$ terms. The full Taylor expansion is given by iterating the above procedure, for example using Eq. (10.7) to replace the functions α and β evaluated at time t''. This procedure obviously introduces triple stochastic integrals; the next iteration introduces quadruple integrals, and so forth.

[3] G. N. Mil'shtein, "Approximate Integration of Stochastic Differential Equations," *Theory of Probability and its Applications* **19**, 557 (1974).

10.3.1 Single and Double Integrals

The first thing that is clear is that we will need to employ a hierarchy of stochastic integrals. The ones we have considered so far are the single integrals

$$\begin{aligned}(I_0)_n &:= \int_{t_n}^{t_{n+1}} dt' = \Delta t \\ (I_1)_n &:= \int_{t_n}^{t_{n+1}} dW(t') = \Delta W_n,\end{aligned} \tag{10.13}$$

as well as the double integrals

$$\begin{aligned}(I_{00})_n &:= \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dt'' = \frac{\Delta t^2}{2} \\ (I_{10})_n &:= \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dW(t'') \\ (I_{01})_n &:= \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dt'' = \Delta t\, \Delta W_n - (I_{10})_n = (I_0)_n (I_1)_n - (I_{10})_n \\ (I_{11})_n &:= \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t'') = \frac{\Delta W_n^2}{2} - \frac{\Delta t}{2}.\end{aligned} \tag{10.14}$$

The first and the last integral we worked out before, and the third integral follows from

$$\begin{aligned}(I_{01})_n &= \int_{t_n}^{t_{n+1}} dW(t')\,(t' - t_n) \\ &= \int_{t_n}^{t_{n+1}} dW(t')\,t' - t_n \Delta W_n \\ &= [t' W(t')]_{t_n}^{t_{n+1}} - \int_{t_n}^{t_{n+1}} dt'\, W(t') - t_n \Delta W_n \\ &= t_{n+1} W(t_{n+1}) - t_n W(t_n) - \int_{t_n}^{t_{n+1}} dt'\,[W(t') - W(t_n)] - \Delta t\, W(t_n) - t_n \Delta W_n \\ &= t_{n+1} W(t_{n+1}) - t_n W(t_{n+1}) - (I_{10})_n - \Delta t\, W(t_n) \\ &= \Delta t\, W(t_{n+1}) - (I_{10})_n - \Delta t\, W(t_n) \\ &= \Delta t\, \Delta W_n - (I_{10})_n,\end{aligned} \tag{10.15}$$

where we integrated by parts in the third step. Note that all the double integrals can be reduced to expressions in terms of single integrals and $(I_{10})_n$; however, this last double integral is irreducible in the sense that it cannot be written only in terms of single integrals. We can thus characterize it more completely. It is clearly Gaussian, as from its definition it is a sum over independent, Gaussian random variables. It has mean, variance, and covariance with ΔW_n given by

$$\begin{aligned}\left\langle\!\left\langle (I_{10})_n \right\rangle\!\right\rangle &= 0 \\ \left\langle\!\left\langle (I_{10})_n^2 \right\rangle\!\right\rangle &= \frac{\Delta t^3}{3} \\ \left\langle\!\left\langle (I_{10})_n\, \Delta W_n \right\rangle\!\right\rangle &= \frac{\Delta t^2}{2}.\end{aligned} \tag{10.16}$$

The mean is obvious, as $(I_{10})_n$ is again the sum over independent, Gaussian random variables of zero mean. The variance and covariance we leave as an exercise (Problem 10.1). For the purposes of simulation, given two independent, standard-normal random numbers z_1 and z_2 (i.e., variance 1 and mean 0), it is not hard to verify that the linear combinations

$$\begin{aligned} \Delta W_n &= \sqrt{\Delta t}\, z_1 \\ (I_{10})_n &= \frac{\Delta t^{3/2}}{2}\left(z_1 + \frac{z_2}{\sqrt{3}}\right) \end{aligned} \tag{10.17}$$

have the correct statistics for the two desired quantities. However, we will return to more useful strategies for computing these numbers below.

10.3.2 Higher-Order Multiple Itō Integrals

The notation here generalizes readily to further multiple integrals. In the case of I_0, the zero subscript indicates a simple integration of dt over the time interval; for I_1, the unit subscript indicates a simple integration of $dW(t)$ instead. For the double-integrals, the subscripts have the same meaning, but the result of integrating according to the first index is integrated again according to the second integral; thus, I_{10} means to integrate dW, and then integrate the result with a the differential dt. In the general case, with $I_{j_1 j_2 j_3 \ldots j_n}$, where $j_\alpha \in \{0, 1\}$, we again integrate dt or dW according to the value of j_1, then integrate the result over dt or dW according to the value of j_2, and so on.

Clearly, when we iterate the procedure leading to Eq. (10.8) to generate the stochastic Taylor expansion, we are generating on the very next iteration triple integrals, such as I_{110}. Roughly speaking, when counting order we should count each 1 as a 1/2 order, while a 0 is a full extra order. Thus, the Taylor expansion to order Δt involves I_0, I_1, and I_{11}. To order $\Delta t^{3/2}$, the expansion also involves I_{10}, I_{01}, and I_{111}; while to order Δt^2, the expansion also involves I_{00}, I_{110}, I_{101}, I_{011}, and I_{1111}. It is not hard to see that the stochastic Taylor expansion is much more complicated than the regular Taylor expansion: the regular Taylor expansion only involves I_0, I_{00}, I_{000}, and so on. In terms of actual calculations, there is the additional complication that at higher orders, new random quantities are introduced that cannot be expressed entirely in terms of lower-order quantities. Again, these integrals may be written in terms of other integrals, although this becomes complicated where many indices are involved; however, it can be shown that[4]

$$(\Delta W)^j I_{j_1 \ldots j_n} = \sum_{i=0}^{n} I_{j_1 \ldots j_{i-1} j j_{i+1} \ldots j_n} + \sum_{i=1}^{n} j\, \delta_{j j_i} \Delta W\, I_{j_1 \ldots j_{i-1} 0 j_{i+1} \ldots j_n}, \tag{10.18}$$

where recall that j and all j_i are either 0 or 1.

It is precisely the existence of extra terms that causes traditional numerical methods for ODEs to fail in general for SDEs. In certain special cases (such as additive noise), ordinary methods may provide better performance (since certain terms in the stochastic Taylor expansion will vanish). However, in general different methods must be developed to handle SDEs at "high" order.[5]

10.3.2.1 Statistics and Simulation

As a reference for using and simulating higher-order multiple Itō integrals, we will list some statistics and simulation formulae. First of all, the variances of $W(t)$ and its integrals with respect to t are

$$\left\langle\!\left\langle \Delta W^2 \right\rangle\!\right\rangle = \Delta t, \qquad \left\langle\!\left\langle (I_{10})^2 \right\rangle\!\right\rangle = \frac{\Delta t^3}{3}, \qquad \left\langle\!\left\langle (I_{100})^2 \right\rangle\!\right\rangle = \frac{\Delta t^5}{20}, \qquad \left\langle\!\left\langle (I_{1000})^2 \right\rangle\!\right\rangle = \frac{\Delta t^7}{252}, \tag{10.19}$$

[4] Kloeden and Platen, *op. cit.*, Proposition 5.2.3, p. 170.

[5] see Kevin Burrage, Pamela Burrage, Desmond J. Higham, Peter E. Kloeden, and Eckhard Platen, "Comment on 'Numerical methods for stochastic differential equations,' " *Physical Review E* **74**, 068701 (2006) (doi: 10.1103/PhysRevE.74.068701), which is a comment on the use of standard Runge–Kutta methods for SDEs by Joshua Wilkie, "Numerical methods for stochastic differential equations," *Physical Review E* **70**, 017701 (2004) (doi: 10.1103/PhysRevE.70.017701).

where all the integrals here are taken with respect to the time interval $[t, t+\Delta t]$. These variances are derived by noting that the increments $dW(t)$ are independent and of variance dt. Then in considering the integrals of $dW(t)$ from 0 to t, the effect of the variance dt of each increment dW must be multiply counted. Thus, the variance of I_{10} follows from integrating $t' \cdot t'$ over the time interval, the variance of I_{110} follows from integrating $(t'^2/2) \cdot (t'^2/2)$ over the time interval, and so on.

Covariances of the stochastic integrals with ΔW are similarly straightforward to calculate:

$$\left\langle\!\left\langle (I_{10})\Delta W \right\rangle\!\right\rangle = \frac{\Delta t^2}{2}, \qquad \left\langle\!\left\langle (I_{100})\Delta W \right\rangle\!\right\rangle = \frac{\Delta t^3}{6}, \qquad \left\langle\!\left\langle (I_{1000})\Delta W \right\rangle\!\right\rangle = \frac{\Delta t^4}{24}. \tag{10.20}$$

Using the same argument as for the variances, for example, the average of $(I_{10})\Delta W$ is computed by integrating $t' \cdot 1$ from 0 to t, the average of $(I_{100})\Delta W$ by integrating $(t'^2/2) \cdot 1$, and so on. The other covariances are similarly

$$\left\langle\!\left\langle (I_{100})(I_{10}) \right\rangle\!\right\rangle = \frac{\Delta t^4}{8}, \qquad \left\langle\!\left\langle (I_{1000})(I_{10}) \right\rangle\!\right\rangle = \frac{\Delta t^5}{30}, \qquad \left\langle\!\left\langle (I_{1000})(I_{100}) \right\rangle\!\right\rangle = \frac{\Delta t^6}{72}. \tag{10.21}$$

Again, the average of $(I_{100})(I_{10})$ is computed by integrating $(t'^2/2) \cdot t'$, and so on.

In a simulation, the above multiple integrals over the nth time interval $[t, t+\Delta t]$ can be simulated as

$$\begin{aligned}
\Delta W_n &= \sqrt{\Delta t}\, z_1 \\
(I_{10})_n &= \frac{\Delta t^{3/2}}{2}\left(z_1 + \frac{z_2}{\sqrt{3}}\right) \\
(I_{100})_n &= \frac{\Delta t^{5/2}}{6}\left(z_1 + \frac{\sqrt{3}\, z_2}{2} + \frac{z_3}{2\sqrt{5}}\right) \\
(I_{1000})_n &= \frac{\Delta t^{7/2}}{24}\left(z_1 + \frac{3\sqrt{3}\, z_2}{5} + \frac{z_3}{\sqrt{5}} + \frac{z_4}{5\sqrt{7}}\right),
\end{aligned} \tag{simulation of multiple integrals} \quad (10.22)$$

in terms of independent, standard-normal random deviates z_1, z_2, z_3, and z_4. These are easy to derive (Problem 10.7) by considering the multiple integrals to be linear combinations of these deviates (note that they are all Gaussian random variables), and then fixing the superposition coefficients by demanding that they reproduce the above covariances.

Other multiple integrals of the same order (involving only one stochastic integral) may be obtained from these simulated values using relations such as

$$\begin{aligned}
\Delta t\, \Delta W &= I_{10} + I_{01} \\
\Delta t\, I_{10} &= I_{010} + 2I_{100} \\
\Delta t\, I_{01} &= 2I_{001} + I_{010} \\
\Delta t\, I_{100} &= I_{0100} + 3I_{1000} \\
\Delta t\, I_{010} &= 2(I_{0010} + I_{0100}) \\
\Delta t\, I_{001} &= 3I_{0001} + I_{0010},
\end{aligned} \tag{10.23}$$

which all follow from Eq. (10.18), and which can be combined to give useful relations. For example, I_{001} can be computed as

$$I_{001} = I_{100} - \Delta t\, I_{10} + \frac{1}{2}\Delta t^2\, \Delta W \tag{10.24}$$

directly from the simulated values (10.22).

Handling the multiple integrals involving more than one stochastic integral is somewhat different. For example, from Eq. (10.11), the simplest double stochastic integral is

$$(I_{11})_n = \frac{\Delta W_n^2 - \Delta t}{2}. \tag{10.25}$$

But ΔW^2 has a gamma distribution (Section 10.8.3.4),

$$f(x;\gamma,\beta) = \frac{\beta^\gamma}{\Gamma[\gamma]}\, x^{\gamma-1}e^{-\beta x} \qquad (x>0,\ \gamma>0,\ \beta>0), \tag{10.26}$$

with $\gamma = 1/2$ and $\beta = 1/2\Delta t$. So I_{11} is not Gaussian-distributed, and neither are its temporal integrals. Nevertheless it is straightforward to compute second-order moments, for example the variances

$$\left\langle\!\!\left\langle (I_{11})^2 \right\rangle\!\!\right\rangle = \frac{\Delta t^2}{2}, \qquad \left\langle\!\!\left\langle (I_{110})^2 \right\rangle\!\!\right\rangle = \frac{\Delta t^4}{12}, \qquad \left\langle\!\!\left\langle (I_{1100})^2 \right\rangle\!\!\right\rangle = \frac{\Delta t^6}{120}. \tag{10.27}$$

The derivation of the latter two variances is less straightforward than for Eqs. (10.19), so we will leave it as an exercise to derive these and other related moments (Problem 10.8). Correlations of these doubly stochastic integrals with ΔW vanish because these are effectively odd moments of ΔW:

$$\left\langle\!\!\left\langle (I_{11})\Delta W \right\rangle\!\!\right\rangle = 0, \qquad \left\langle\!\!\left\langle (I_{110})\Delta W \right\rangle\!\!\right\rangle = 0, \qquad \left\langle\!\!\left\langle (I_{1100})\Delta W \right\rangle\!\!\right\rangle = 0. \tag{10.28}$$

The same covariances with I_{10}, I_{100}, etc. vanish also for the same reason. There are correlations between the doubly stochastic integrals, however:

$$\left\langle\!\!\left\langle (I_{110})(I_{11}) \right\rangle\!\!\right\rangle = \frac{\Delta t^3}{3!}, \qquad \left\langle\!\!\left\langle (I_{1100})(I_{11}) \right\rangle\!\!\right\rangle = \frac{\Delta t^4}{4!}, \qquad \left\langle\!\!\left\langle (I_{1100})(I_{110}) \right\rangle\!\!\right\rangle = \frac{\Delta t^5}{40}. \tag{10.29}$$

The derivation of these correlations is also left as an exercise (Problem 10.8)

$$\begin{aligned} (I_{11})_n &= \frac{\Delta t}{2}\left(z_1^2 - 1\right) \\ (I_{110})_n &= \frac{\Delta t^2}{6}\left(z_1^2 - 1 + z_{11}\right) \end{aligned} \qquad \text{(simulation of multiple integrals)} \tag{10.30}$$

blah

10.3.3 Expression for the Taylor Expansion

As we have indicated thus far, the stochastic Taylor expansion is quite complicated. We will simply state the result here for the Itō–Taylor expansion[6] for the solution $y(t)$ to the scalar SDE (10.1), which is

$$y(t) = y(t_0) + \sum_{n=1}^{\infty} \sum_{j_1 \ldots j_n = 0}^{1} I_{j_1 \ldots j_n} f_{j_1 \ldots j_n}, \tag{10.31}$$

(Itō–Taylor expansion)

where the multiple stochastic integrals $I_{j_1 \ldots j_n}$ are defined over the time interval from t_0 to t, and the coefficient functions are defined recursively in terms of the functions with one index removed:

$$f_{j_1 \ldots j_n} = L^{j_1} f_{j_2 \ldots j_n}, \tag{10.32}$$

The lowest-order case of f (with all indices removed) is defined by $f = y$, and the operators L^j are defined by

$$\begin{aligned} L^0 &= \partial_t + \alpha\partial_y + \frac{1}{2}\beta^2\partial_y^2 \\ L^1 &= \beta\partial_y. \end{aligned} \tag{10.33}$$

[6] Kloeden and Platen, *op. cit.*, Section 5.5, p. 181. The Itō–Taylor expansion given in this reference is much more rigorous than what we have written here; here, we have written what amounts to an infinite number of iterations of the Itō integral formula. Truncations of the expansion should in principe be done carefully.

Thus, for example,

$$\begin{aligned} f_0 &= L^0 y = \alpha \\ f_1 &= L^1 y = \beta \\ f_{11} &= L^1 f_1 = \beta \partial_y \beta \\ f_{01} &= L^0 f_1 = \partial_t \beta + \alpha \partial_y \beta + \frac{1}{2}\beta^2 \partial_y^2 \beta. \end{aligned} \tag{10.34}$$

Clearly, these coefficient functions become much more complicated as the order increases. However, it should also be reasonably clear how this expansion comes about by the iterated application of the Itō integral forumula. This is one case where Stratonovich calculus is fairly nice: the corresponding Stratonovich formulae for the multiple integrals and coefficients are simplified, for example, by the absence of the final term in the expression for L^0.

10.3.4 Multiple Wiener Processes

As you can imagine, a *vector* SDE driven by multiple Wiener processes dW_k is *way* more complicated than the scalar case. In particular, the multiple integrals generalize to indices beyond 0 and 1, so that there are many more terms in the Taylor expansion, and the differential operators must be similarly generalized. In any case, some extensions of the methods we show below to vector Wiener processes are given by Kloeden and Platen[7]

10.4 Stochastic Runge–Kutta Methods

The Milstein method (10.12) is the simplest example of ''Taylor'' methods that explicitly cancel the higher-order terms in the Itō–Tayler expansion. The problem with these methods is that they require not only specifying the functions α and β in the SDE (10.1), but also their derivatives. In the Milstein case, we have to specify β', but for higher-order methods, more and more derivatives must be specified. This is not a problem in principle, but merely for convenience of the user.

Instead, we can try an approach following the ODE case of Runge–Kutta methods, where multiple function evaluations with different arguments are used to estimate the information that would otherwise be supplied by the derivatives. Again, though, these are not the same Runge–Kutta methods as in the deterministic case, but different methods designed to cancel error terms in the Itō–Taylor expansion above. We will simply quote two Itō methods here.[8] An $O(\Delta t)$ method (i.e., a stochastic Runge–Kutta method of comparable accuracy to the Milstein method) is

$$\begin{aligned} d_n &= y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\sqrt{\Delta t} \\ y_{n+1} &= y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\,\Delta W_n + \frac{1}{2\sqrt{\Delta t}}[\beta(d_n, t_{n+1}) - \beta(y_n, t_n)][\Delta W_n^2 - \Delta t]. \end{aligned}$$

(stochastic Runge–Kutta method, order 1) (10.35)

[7] Kloeden and Platen, *op. cit.*

[8] Kloeden and Platen, *op. cit.*; P. E. Kloeden and E. Platen, ''Higher-Order Implicit Strong Numerical Schemes for Stochastic Differential Equations,'' *Journal of Statistical Physics* **66**, 283 (1992) (doi: 10.1007/BF01060070).

An $O(\Delta t^{3/2})$ method is

$$
\begin{aligned}
d_{1\pm} &= y_n + \alpha(y_n,t_n)\,\Delta t \pm \beta(y_n,t_n)\sqrt{\Delta t} \\
d_{2\pm} &= d_{1+} \pm \beta(d_{1+},t_{n+1})\sqrt{\Delta t} \\
y_{n+1} &= y_n + \beta(y_n,t_n)\,\Delta W_n + \frac{1}{2\sqrt{\Delta t}}\left[\alpha(d_{1+},t_{n+1}) - \alpha(d_{1-},t_{n-1})\right](I_{10})_n \\
&\quad + \frac{1}{4}\left[\alpha(d_{1+},t_{n+1}) + 2\alpha(y_n,t_n) + \alpha(d_{1-},t_{n-1})\right]\Delta t \\
&\quad + \frac{1}{4\sqrt{\Delta t}}\left[\beta(d_{1+},t_{n+1}) + \beta(d_{1-},t_{n-1})\right]\left(\Delta W_n^2 - \Delta t\right) \\
&\quad + \frac{1}{2\Delta t}\left[\beta(d_{1+},t_{n+1}) - 2\beta(y_n,t_n) + \beta(d_{1-},t_{n-1})\right]\left[\Delta W_n\,\Delta t - (I_{10})_n\right] \\
&\quad + \frac{1}{4\Delta t}\left[\beta(d_{2+},t_{n+1}) - \beta(d_{2-},t_{n-1}) - \beta(d_{1+},t_{n+1}) + \beta(d_{1-},t_{n-1})\right]\left(\frac{\Delta W_n^2}{3} - \Delta t\right)\Delta W_n.
\end{aligned}
$$

(stochastic Runge–Kutta method, order 3/2) (10.36)

At the time of writing, there seems to have been no true $O(\Delta t^2)$ methods that have been reported for the general (multiplicative noise) case.

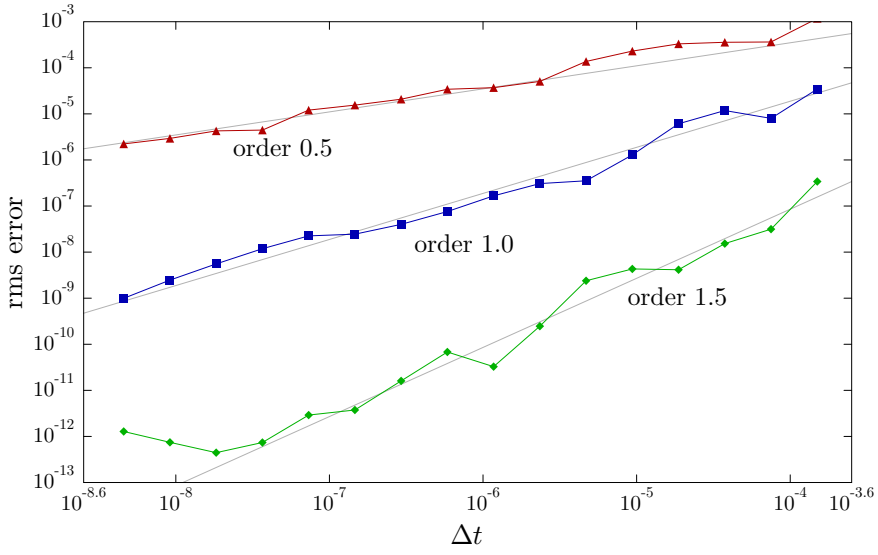
To see the relative performance of the schemes we have presented thus far, we test them on the sample problem[9]

$$dy = -(a + b^2 y)(1 - y^2)\,dt + b(1 - y^2)\,dW, \tag{10.37}$$

where a and b are constants. This is a multiplicative SDE with the analytic solution

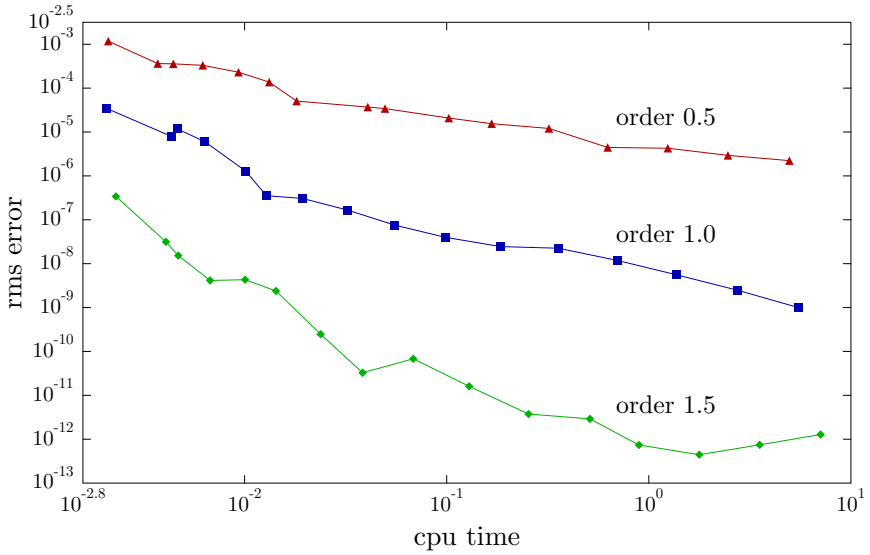
$$y(t) = \frac{(1 + y_0)\exp[-2at + 2bW(t)] + y_0 - 1)}{(1 + y_0)\exp[-2at + 2bW(t)] + 1 - y_0)}. \tag{10.38}$$

For this calculation we took $a = b = 1$ and $y_0 = 0$. The base step size was $\Delta t = 0.00015$, with the other runs having the same base step size reduced by powers of 2, down to $\Delta t = 0.00015/2^{15}$. The solutions were computed from $t = 0$ out to $t = 0.045$ (300 of the largest steps), and the rms error was computed for a single trajectory in each case. Every calculation was performed on the same Brownian path $W(t)$ (see below). The following plot compares the accuracy of the stochastic Euler method (10.3), the $O(\Delta t)$ Runge–Kutta method (10.35), and the $O(\Delta t^{3/2})$ Runge–Kutta method (10.36).



[9]K. Burrage and P. M. Burrage, "High strong order explicit Runge–Kutta methods for stochastic ordinary differential equations," *Applied Numerical Mathematics* **22**, 81 (1996); Peter E. Kloeden and Eckhard Platen, *Numerical Solution of Stochastic Differential Equations*, 3rd ed. (Springer, 2000).

The scaling behavior in each case is clear, though again the convergence is slow compared to ODEs, and the step sizes required to achieve high accuracy are quite small. Note that in the $O(\Delta t^{3/2})$ case, rounding dominates the error at the smallest step sizes. In terms of cpu time, the higher order methods are again superior in this problem.



In this case, each method takes a similar amount of time for a given time step Δt. This is because the test SDE is very simple, and the cpu effort is dominated by random-number generation and construction of the Brownian path. For very complicated SDEs, there will obviously be more of a difference in run times between the different methods. However, it is clear that higher-order methods still provide substantial benefit despite the added complexity.

10.5 Implicit Schemes

Implicit methods—where the update rule involves the solution at the new time—have some advantages in certain classes of deterministic ODEs. It is therefore natural to try to construct similar implicit methods for SDEs. Unfortunately, we can quickly run into problems. Consider the simple SDE

$$dy = -ay\,dt - by\,dW \tag{10.39}$$

with constant coefficients a and b. The stochastic Euler method for this SDE reads

$$y_{n+1} = y_n - ay_n\,\Delta t - by_n\,\Delta W_n. \tag{10.40}$$

The obvious generalization of the impicit Euler method for deterministic ODEs is to evaluate the functions on the right-hand side at time t_{n+1} instead of t:

$$y_{n+1} = y_n - ay_{n+1}\,\Delta t - by_{n+1}\,\Delta W_n. \tag{10.41}$$

Note that ΔW_n is unchanged, because it must be chosen according to the convention for Itō SDEs. Solving this equation for y_{n+1},

$$y_{n+1} = \frac{y_n}{1 + a\,\Delta t + b\,\Delta W_n}. \tag{10.42}$$

Clearly there is a problem here: ΔW_n is a stochastic variable that can take on essentially any value with nonzero probability, and typically takes on values of the order of $\sqrt{\Delta t}$. For small Δt it may thus happen that the denominator can come very close to zero, when ΔW_n takes on values near $-(a/b)\Delta t$. Thus we lose the stability properties we gained in the deterministic case, due just to the stochastic nature of the Wiener process. The problem is apparently just with the *stochastic* part of the equation. There is no problem, for

example, if we choose a hybrid scheme where the deterministic step is taken implicitly, but the stochastic parts are treated explicitly, as in

$$y_{n+1} = y_n - a y_{n+1}\,\Delta t - b y_n\,\Delta W_n \tag{10.43}$$

for the example problem, which becomes

$$y_{n+1} = \frac{y_n}{1 + a\,\Delta t} - b y_n\,\Delta W_n. \tag{10.44}$$

which has no special problems with small denominators.

Again, we will simply quote two implicit Itō methods here,[10] corresponding to the explicit $O(\Delta t)$ and $O(\Delta t^{3/2})$ methods above. An $O(\Delta t)$ method (i.e., a stochastic Runge–Kutta method of comparable accuracy to the Milstein method) is

$$\begin{aligned} d_n &= y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\sqrt{\Delta t} \\ y_{n+1} &= y_n + \frac{1}{2}\left[\alpha(y_n, t_n) + \alpha(y_{n+1}, t_{n+1})\right]\,\Delta t + \beta(y_n, t_n)\,\Delta W_n \\ &\quad + \frac{1}{2\sqrt{\Delta t}}[\beta(d_n, t_{n+1}) - \beta(y_n, t_n)][\Delta W_n^2 - \Delta t]. \end{aligned}$$

(implicit stochastic Runge–Kutta method, order 1) (10.45)

Here, we have taken a simple average of $\alpha(y_n, t_n)$ and $\alpha(y_{n+1}, t_{n+1})$. Of course, any weighted average could be taken, interpolating between explicit and fully implicit. Similarly, an $O(\Delta t^{3/2})$ method is

$$\begin{aligned} d_{1\pm} &= y_n + \alpha(y_n, t_n)\,\Delta t \pm \beta(y_n, t_n)\sqrt{\Delta t} \\ d_{2\pm} &= d_{1+} \pm \beta(d_{1+}, t_{n+1})\sqrt{\Delta t} \\ y_{n+1} &= y_n + \beta(y_n, t_n)\,\Delta W_n + \frac{1}{2}\left[\alpha(y_{n+1}, t_{n+1}) + \alpha(y_n, t_n)\right]\,\Delta t \\ &\quad + \frac{1}{4\sqrt{\Delta t}}\left[\beta(d_{1+}, t_{n+1}) + \beta(d_{1-}, t_{n-1})\right]\,\left(\Delta W_n^2 - \Delta t\right) \\ &\quad + \frac{1}{2\Delta t}\left[\beta(d_{1+}, t_{n+1}) - 2\beta(y_n, t_n) + \beta(d_{1-}, t_{n-1})\right]\,\left[\Delta W_n\,\Delta t - (I_{10})_n\right] \\ &\quad + \frac{1}{2\sqrt{\Delta t}}\left[\alpha(d_{1+}, t_{n+1}) - \alpha(d_{1-}, t_{n-1})\right]\left[(I_{10})_n - \frac{\Delta W_n\,\Delta t}{2}\right] \\ &\quad + \frac{1}{4\Delta t}\left[\beta(d_{2+}, t_{n+1}) - \beta(d_{2-}, t_{n-1}) - \beta(d_{1+}, t_{n+1}) + \beta(d_{1-}, t_{n-1})\right]\,\left(\frac{\Delta W_n^2}{3} - \Delta t\right)\Delta W_n. \end{aligned}$$

(stochastic Runge–Kutta method, order 3/2) (10.46)

This form is specific to the choice of a half ''degree of implicitness'' (that is, an average of α at the present and advanced times). When comparing these methods to the corresponding explicit methods in the last section on the example problem (10.37), the performance is about the same (with a slight advantage in the order 1 case) for a given time step, and the cpu time is again about the same since the calculation is dominated by the random-number generation, not by the implementation of the finite-difference formulae or by the functional iteration in the implicit schemes.

Of course, similar tricks can be done with any of the schemes we have presented so far: the deterministic step can be taken with high-order deterministic methods (such as fourth-order Runge–Kutta), so long as the stochastic parts are treated according to SDE-adapted methods as presented here.

The formulae here and in the last section have been implemented in a publicly-available Fortran 90 module.[11] This module includes the facilities described below to generate consistent Brownian paths with different step sizes, and a sample code implementing the test problem above is included.

[10] Kloeden and Platen, *op. cit.*; P. E. Kloeden and E. Platen, ''Higher-Order Implicit Strong Numerical Schemes for Stochastic Differential Equations,'' *Journal of Statistical Physics* **66**, 283 (1992) (doi: 10.1007/BF01060070).

[11] `SDERK90`, available online at `http://atomoptics.uoregon.edu/~dsteck/computer.html`.

10.6 Strong and Weak Convergence

Up until now, we have been considering a particular kind of convergence of solutions of SDEs, **strong convergence**. This means we are considering *pathwise* convergence of solutions. If $\tilde{y}(t;\Delta t)$ is a finite-difference approximation to the true solution y(t) to the SDE (10.1), then the method for generating $\tilde{y}(t;\Delta t)$ is of **strong order** γ if at fixed t,

$$\langle\langle y(t) - \tilde{y}(t;\Delta t)\rangle\rangle = O(\Delta t^{\gamma}). \qquad (10.47)$$

(strong convergence condition)

The order here again refers to the scaling behavior of the *global* error. Another less-demanding convergence criterion refers to convergence of ensemble means. The same numerical method is said to have **weak order** δ if for every polynomial $g(y)$,

$$\langle\langle g(y(t))\rangle\rangle - \langle\langle g(\tilde{y}(t;\Delta t))\rangle\rangle = O(\Delta t^{\delta}) \qquad (10.48)$$

(weak convergence condition)

at fixed time t. Strong convergence at order γ implies weak convergence of at least the same order. For example, we argued that the stochastic Euler method has strong order 1/2. This is due to an error term proportional to I_{11}, which has zero mean; when considering expectation values, this error term is wiped out, and the stochastic Euler scheme is actually of weak order 1.[12] The Milstein method turns out to have both strong and weak order 1.[13] Since we are normally thinking about solutions to SDEs as individually interesting objects (modeling individual realizations of a continuous measurement process, for example), we will generally be concerned with the more difficult case of strong convergence.

10.7 Consistent Brownian Paths

In considering the strong convergence of the solution to an SDE, it is in general easy to check the convergence where the exact solution is known, as in the test problem above. However, this is obviously more difficult in the generic case where the exact solution is *not* known. In deciding whether or not to accept a particular numerical solution, in the deterministic case you would usually just run the integration again, but with a different step size (say half the original step size). If the two solutions match to within some tolerance, then you can accept the solution. But this is trickier with SDEs. Suppose that you use a random number generator to generate a sequence of Wiener increments

$$\Delta W_0,\ \Delta W_1,\ \Delta W_2,\ \ldots, \qquad (10.49)$$

where $\langle\langle \Delta W_n^2\rangle\rangle = \Delta t$. To cut the step size in half and rerun things, first of all, the same set of random numbers must be used, otherwise the two runs will not in general be well correlated. But even if we use the same set of random numbers to generate the new Wiener increments,

$$\Delta W_0^{(1/2)},\ \Delta W_1^{(1/2)},\ \Delta W_2^{(1/2)},\ \Delta W_3^{(1/2)},\ \Delta W_4^{(1/2)},\ \Delta W_5^{(1/2)},\ \ldots, \qquad (10.50)$$

where $\langle\langle (\Delta W_n^{(1/2)})^2\rangle\rangle = \Delta t/2$, we will *still* have problems, because the increments don't line up: the first random number in the two cases generated ΔW_0 and $\Delta W_0^{(1/2)}$, while the second random number generated ΔW_1 and $\Delta W_1^{(1/2)}$. However, ΔW_1 and $\Delta W_1^{(1/2)}$ don't correspond to the same absolute time. In fact, for consistency, what we require is

$$\Delta W_0 = \Delta W_0^{(1/2)} + \Delta W_1^{(1/2)};\quad \Delta W_1 = \Delta W_2^{(1/2)} + \Delta W_3^{(1/2)};\quad \Delta W_3 = \Delta W_4^{(1/2)} + \Delta W_5^{(1/2)}, \qquad (10.51)$$

and so on. These conditions allow both sequences to correspond to time integrals $(I_1)_n$ of the same ideal, particular realization of the Wiener process $W(t)$, which we will refer to as a particular **Brownian path**. Thus, what we require is a procedure for constructing sequences of Wiener increments with different step sizes, but corresponding to different paths.

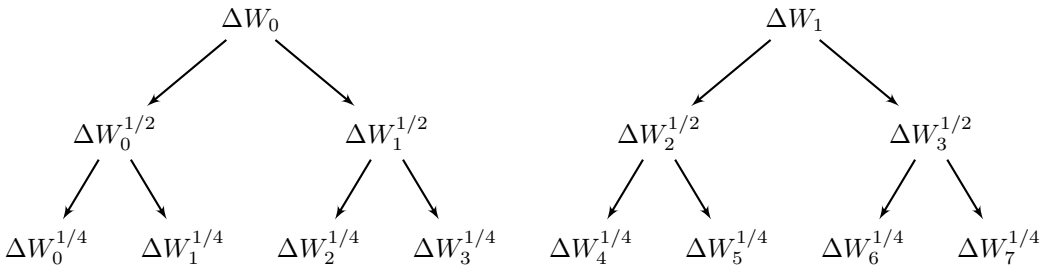
[12] G. N. Milshtein, "A Method of Second-Order Accuracy Integration of Stochastic Differential Equations," *Theory of Probability and its Applications* **23**, 396 (1978) (doi: 10.1137/1123045).

[13] G. N. Milshtein, *op. cit.*

One convenient method for doing this is to start with the sequence of Wiener increments ΔW_n on the coarser time step Δt, and then to *refine* it consistently onto the finer time grid of step $\Delta t/2$.[14] That is, given a Gaussian Wiener increment ΔW_n with variance Δt, we will use another Gaussian random variable z of unit variance (i.e., standard normal) to generate two new, uncorrelated Gaussian random numbers $\Delta W_n^{(1)}$ and $\Delta W_n^{(2)}$, each of variance $\Delta t/2$ and satisfying $\Delta W_n^{(1)} + \Delta W_n^{(2)} = \Delta W_n$. These conditions are clearly satisfied if we choose[15]

$$\begin{aligned} \Delta W_n^{(1)} &= \frac{1}{2}\Delta W_n + \frac{\sqrt{\Delta t}}{2}\, z \\ \Delta W_n^{(2)} &= \frac{1}{2}\Delta W_n - \frac{\sqrt{\Delta t}}{2}\, z. \end{aligned} \qquad \begin{array}{r} (10.52) \\ \text{(refinement of Brownian path)} \end{array}$$

This procedure may obviously be iterated to obtain consistent Brownian paths with yet smaller time steps (by powers of two), creating a tree structure of Wiener increments, with different levels corresponding to different step sizes.



Descending one level on this Brownian tree involves generating a new random number for each Wiener increment on the coarser level, and then combining them as in Eqs. (10.52) to generate the Wiener increments on the finer level. This procedure may be continued indefinitely to consistently realize Brownian paths with arbitrarily fine steps.

This procedure allows you, for example, to run one simulation with one time step Δt and another $\Delta t/2$ on the same Brownian path, and then compare the two runs to estimate the numerical error. Or you can run many different step sizes and examine the convergence behavior of the solution. Consistency of the Brownian path is also an essential first step in implementing adaptive time-step methods for SDEs, which are much more difficult than for ODEs. (The stochastic Euler method, for example, as a basis for an adaptive method can lead to convergence to a wrong solution![16])

10.7.1 Consistent Iterated Integrals

One remaining point to address is that for the order 3/2 Runge–Kutta methods, we also need to refine $(I_{10})_n$ consistently onto finer time grids. One conceptually simple way to generate the $(I_{10})_n$ in the first place is to generate the Wiener increments on a *finer* scale than needed for the integration. That is, if we intend to generate a finite-difference solution to an SDE with time step Δt, Wiener increments ΔW_n, and iterated integrals $(I_{10})_n$, then we should start by generating *finer* Wiener increments δW_n for step size $\delta t < \Delta t$. A reasonable choice in practice would be, say, $\delta t = \Delta t/20$ (although this also works if $\delta t = \Delta t/2$). Then compute ΔW_n by the simple sum,

$$\Delta W_n = \sum_{j=1}^{N_\delta} \delta W_j, \tag{10.53}$$

where N_δ is the number of temporal substeps per integration step (20 or 2, as we mentioned above). This expression is exact, and hasn't gotten us anything because we could have just picked ΔW_n directly. The

[14] J. G. Gaines and T. J. Lyons, "Variable Step Size Control in the Numerical Solution of Stochastic Differential Equations," *SIAM Journal on Applied Mathematics* **57**, 1455 (1997) (doi: 10.1137/S0036139995286515).

[15] Paul Lévy, *Processus Stochastiques et Mouvement Brownien* (Gauthier–Villars, 1948).

[16] Gaines and Lyons, *op. cit.*

point is that we may also compute $(I_{10})_n$ via the deterministic Euler method

$$(I_{10})_n = \int_{t_n}^{t_{n+1}} dt'\, W(t') \approx \delta t \sum_{j=2}^{N_\delta} \left(\sum_{k=1}^{j} \delta W_k \right). \tag{10.54}$$

We would normally count the global error of this approximation as $O(\delta t) = O(\Delta t)$ if we keep N_δ fixed. However, $(I_{10})_n$ itself is $O(\Delta t^{3/2})$, and so accounting for the *coefficient* of the global error term, the error in the approximation (10.54) is $O(\Delta t^{3/2}\delta t) = O(\Delta t^{5/2})$. Thus, this approximation is easily adequate for the order 1.5 methods (10.36) and (10.46). Furthermore, from the last section we know how to refine the increments δW_n onto a finer time grid, and thus we can consistently generate the $(I_{10})_n$ also with finer time steps on the same Brownian path.

10.7.1.1 Lévy Areas

A better approximation to the integral $(I_{10})_n$ is to compute the **Lévy area**[17]

$$(A_{ij})_n := \frac{1}{2}\left(\int_{t_n}^{t_{n+1}} dW^{(j)}(t') \int_t^{t'} dW^{(i)}(t'') - \int_{t_n}^{t_{n+1}} dW^{(i)}(t') \int_t^{t'} dW^{(j)}(t'') \right) = \frac{1}{2}\left[(I_{ij})_n - (I_{ji})_n \right], \tag{10.55}$$

(Lévy area)

where in the notation here, $dW^0 \equiv dt$ and $dW^1 \equiv dW$. We can also approximate this by the (stochastic) Euler scheme (which is equivalent to the Milstein scheme for these integrals),

$$(A_{ij})_n \approx \frac{1}{2}\left(\sum_{p=2}^{N_\delta} \delta W_p^{(j)} \sum_{q=1}^{p-1} \delta W_q^{(i)} - \sum_{p=2}^{N_\delta} \delta W_p^{(i)} \sum_{q=1}^{p-1} \delta W_q^{(j)} \right). \tag{10.56}$$

In particular, the Lévy area that we want to approximate is

$$\begin{aligned} (A_{10})_n &\approx \frac{1}{2}\left(\sum_{p=2}^{N_\delta} \delta t \sum_{q=1}^{p-1} \delta W_q - \sum_{p=2}^{N_\delta} \delta W_p \sum_{q=1}^{p-1} \delta t \right) \\ &= \frac{\delta t}{2}\left(\sum_{p=2}^{N_\delta} \sum_{q=1}^{p-1} \delta W_q - \sum_{p=2}^{N_\delta} \delta W_p (p-1) \right). \end{aligned} \tag{10.57}$$

Then since

$$(A_{10})_n = \frac{1}{2}\left[(I_{10})_n - (I_{01})_n \right] = (I_{10})_n - \frac{1}{2}\Delta t\, \Delta W_n, \tag{10.58}$$

where we used Eqs. (10.14), we can compute $(I_{10})_n$ based on this approximation to the Lévy area.

To see why the Lévy area is better despite the extra complication of the formula, we can compute the variance of the numerical approximant (10.54) to $(I_{10})_n$:

$$\begin{aligned} \langle\!\langle (\hat{I}_{10})_n^2 \rangle\!\rangle &= \delta t^2 \left[(N_\delta - 1)^2 \langle\!\langle \delta W_1^2 \rangle\!\rangle + (N_\delta - 1)^2 \langle\!\langle \delta W_2^2 \rangle\!\rangle + (N_\delta - 2)^2 \langle\!\langle \delta W_3^2 \rangle\!\rangle + \cdots + (1)^2 \langle\!\langle \delta W_{N_\delta}^2 \rangle\!\rangle \right] \\ &= \delta t^3 \left[(N_\delta - 1)^2 + \sum_{j=1}^{N_\delta - 1} j^2 \right] \\ &= \delta t^3 \left[\frac{N_\delta^3}{3} + \frac{N_\delta^2}{2} - \frac{11 N_\delta}{6} + 1 \right] \\ &= \frac{\Delta t^3}{3}\left[1 + \frac{3}{2}\frac{\delta t}{\Delta t} - \frac{11}{2}\left(\frac{\delta t}{\Delta t} \right)^2 + 3\left(\frac{\delta t}{\Delta t} \right)^3 \right]. \end{aligned} \tag{10.59}$$

[17] Gaines and Lyons, *op. cit.*

Here, we used $N_\delta\,\delta t = \Delta t$. The variance *should* be $\Delta t^3/3$, as we wrote in Eq. (10.16). Thus, this numerical approximation gets variance of the generated $(I_{10})_n$'s wrong by an $O(\delta t/\Delta t)$ bias.

For the Lévy area, we can also compute the variance of the numerical approximant:

$$\begin{aligned}
\langle\!\langle (\hat{A}_{10})_n^2 \rangle\!\rangle &= \frac{\delta t^2}{4}\Big[(N_\delta - 1)^2 \langle\!\langle \delta W_1^2 \rangle\!\rangle + (N_\delta - 1 - 1)^2 \langle\!\langle \delta W_2^2 \rangle\!\rangle + (N_\delta - 2 - 2)^2 \langle\!\langle \delta W_3^2 \rangle\!\rangle \\
&\qquad + \cdots + (1 - (N_\delta - 1))^2 \langle\!\langle \delta W_{N_\delta}^2 \rangle\!\rangle\Big] \\
&= \frac{\delta t^3}{4}\left[(N_\delta - 1)^2 + \sum_{j=1}^{N_\delta - 1} [j - (N_\delta - j)]^2\right] \\
&= \frac{\delta t^3}{4}\left[\frac{N_\delta^3}{3} - \frac{4N_\delta}{3} + 1\right] \\
&= \frac{\Delta t^3}{12}\left[1 - 4\left(\frac{\delta t}{\Delta t}\right)^2 + 3\left(\frac{\delta t}{\Delta t}\right)^3\right].
\end{aligned} \tag{10.60}$$

The true variance of the Lévy area is

$$\langle\!\langle (A_{10})_n^2 \rangle\!\rangle = \langle\!\langle (I_{10})_n^2 \rangle\!\rangle + \frac{\Delta t^2}{4}\langle\!\langle \Delta W_n^2 \rangle\!\rangle - 2\frac{\Delta t}{2}\langle\!\langle (I_{10})_n \Delta W_n \rangle\!\rangle = \frac{\Delta t^3}{3} + \frac{\Delta t^2}{4}\Delta t - 2\frac{\Delta t}{2}\frac{\Delta t^2}{2} = \frac{1}{12}\Delta t^3, \tag{10.61}$$

and thus the approximant biases the variance by an error of $O[(\delta t/\Delta t)^2]$. Since we can compute the integral $(I_{10})_n$ directly from the Lévy area by adding a quantity that can be computed exactly, the bias to the variance of $(I_{10})_n$ by this method is only $O[(\delta t/\Delta t)^2]$, which is one order better than the previous method.

10.7.1.2 Direct Refinement

Another method for directly refining ΔW_n and $(I_{10})_n$ onto a finer temporal grid, in the spirit of Eqs. (10.52), is to combine them directly with *two* independent, standard-normal random numbers z_1 and z_2, according to the linear transformation[18]

$$\begin{bmatrix} \Delta W_n^{(1)} \\ (I_{10})_n^{(1)} \\ \Delta W_n^{(2)} \\ (I_{10})_n^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{\sqrt{\Delta t}}{4} & -\frac{1}{4} & \frac{3}{2\Delta t} \\ \frac{\Delta t^{3/2}}{8\sqrt{3}} & 0 & -\frac{\Delta t}{8} & \frac{1}{2} \\ 0 & \frac{\sqrt{\Delta t}}{4} & \frac{5}{4} & -\frac{3}{2\Delta t} \\ -\frac{\Delta t^{3/2}}{8\sqrt{3}} & \frac{\Delta t^{3/2}}{8} & \frac{\Delta t}{4} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \Delta W_n \\ (I_{10})_n \end{bmatrix}. \tag{10.62}$$

(refinement of Brownian path)

Here, the $\Delta W_n^{(1)}$ and $\Delta W_n^{(2)}$ are the Wiener increments for the two subintervals of duration $\Delta t/2$, and $(I_{10})_n^{(1)}$ and $(I_{10})_n^{(2)}$ are the corresponding double integrals. This method obviously has the advantage of fewer extra random numbers generated and fewer arithmetic operations for each refinement, compared to the Lévy-area method above. This method also has no bias in terms of the variance of the refined variables.

To see where this method comes from, consider the *inverse* problem, which is that if you have a Wiener path defined on a fine time grid, how do you define the Wiener path on a coarser subgrid? The transformation for the coarse increments ΔW are easy to compute by summing the finer increments, but the transformations for the multiple integrals are more complicated. Consider the multiple integrals over

[18] Kevin Burrage, Pamela Burrage, and Taketomo Mitsui, "Numerical solutions of stochastic differential equations — implementations and stability issues," *Journal of Computational and Applied Mathematics* **125**, 171 (2000) (doi: 10.1016/S0377-0427(00)00467-2). Actually, these authors give a more general transformation to two time intervals of *nonequal* duration.

the time interval $[0,t]$, which is divided into N subintervals of time $\Delta = t/N$. Then the multiple integrals on the full interval in terms of the fine-step integrals may be written (Problem 10.4)

$$
\begin{aligned}
W(t) &= \sum_{n=0}^{N-1} \Delta W(n\Delta t) \\
(I_{10})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{10})_{n\Delta t}^{(n+1)\Delta t} + (N-1-n)\Delta t\, \Delta W(n\Delta t) \Big] \\
(I_{100})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{100})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^2}{2}(N-1-n)^2 \Delta W(n\Delta t) \Big] \\
(I_{1000})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{1000})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{100})_{n\Delta t}^{(n+1)\Delta t} \\
&\qquad + \frac{\Delta t^2}{2}(N-1-n)^2 (I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^3}{3!}(N-1-n)^3 \Delta W(n\Delta t) \Big]
\end{aligned}
$$

(coarser Brownian path) (10.63)

and

$$
\begin{aligned}
(I_{11})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
(I_{110})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{110})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n) \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
&\qquad + (I_{10})_{n\Delta t}^{(n+1)\Delta t} W(n\Delta t) \Big] \\
(I_{1100})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{1100})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n) \Big((I_{110})_{n\Delta t}^{(n+1)\Delta t} + (I_{10})_{n\Delta t}^{(n+1)\Delta t}\, W(n\Delta t) \Big) \\
&\qquad + \frac{\Delta t^2}{2}(N-1-n)^2 \Big((I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big) + (I_{100})_{n\Delta t}^{(n+1)\Delta t} W(n\Delta t) \Big].
\end{aligned}
$$

(coarser Brownian path) (10.64)

Here, the explicit time ranges are specified for each multiple integral for clarity, so that for example

$$
(I_{10})_{t_1}^{t_2} := \int_{t_1}^{t_2} dt \int_{t_1}^{t} dW(t'). \tag{10.65}
$$

Then the transformations (10.62) follow from writing each refined variable as a linear combination of the coarse variables and the extra random variables, and enforcing consistency of Eqs. (10.63) as well as the proper variances and covariances of the refined variables.

10.8 Random Numbers

The ability to generate quality ''random'' numbers on a computer is obviously an important part of simulating stochastic systems. We will thus spend a bit of time surveying some useful techniques towards these goals.

10.8.1 Uniform Distribution

The workhorse of stochastic simulations is a random-number generator that generates numbers uniformly distributed on the interval $[0,1)$. These numbers can then be transformed to whatever distribution needed by various methods. Most compilers and canned software packages already include such a generator, so why bother thinking about it? First, you might not want to *trust* just any random-number generator that was

handed to you, since there have historically been problems found with random-number generators, some subtle.[19] Second, it is useful to have a *portable* random-number generator, which can generate the exact same sequence of ''random'' numbers under any compiler/architecture, which is impossible using ''built-in'' generators. Doing this generally requires implementing the generator in integer arithmetic to avoid floating-point rounding variations on different architectures. This is very handy when porting your code to a new platform: the easiest test is just to run it and verify that you get essentially the same output for a particular run. This won't happen if the random numbers are different, however. Finally, having several known methods available allows you to switch methods to make sure you aren't seeing any artifacts due to a particular method.

So just what is a random-number algorithm on a computer? Obviously, it is a *deterministic algorithm*, and ultimately, with the finite precision of any computer, the algorithm will be periodic. The idea is to come up with an algorithm that (1) produces the correct (uniform) distribution, (2) has a period much larger than the number of pseudorandom numbers needed for simulations, and (3) shows no detectable correlations over sets of pseudorandom numbers large enough for useful simulations. The algorithm must generally be chosen carefully to meet these criteria.[20] Useful algorithms may be ''seeded'' with an initial value to produce the same set of random numbers each time. Also, it is desireable that different seeds correspond to different random-number sequences (that is, they should start the generator with initial conditions such that the sequences do not overlap for many iterations. Batteries of statistical tests are available to check the quality of random-number generators, such as Marsaglia's ''Diehard Battery''[21] and the battery of tests described by Knuth.[22] We point out three algorithms here that meet all these criteria, plus some methods for improving the random numbers further.[23]

10.8.1.1 L'Ecuyer's Multiple Recursive Generator

A class of computationally very simple methods go by the name of **linear congruential generators** (LCGs),[24] and implement the recurrence

$$x_{n+1} = (ax_n + c) \mod m \tag{10.66}$$

(linear conguential generator)

in integer arithmetic. These algorithms, implemented with real numbers, would clearly be simple chaotic systems, where the state x_n is stretched by a, shifted by c, and then folded by m. Of course, in integer arithmetic, the sequence is periodic, and the constants here must be chosen carefully to give decent performance (clearly $a = 1$, $c = 1$, and $m = 10$ would be a bad choice!). These methods are popular due to their simplicity, but they can have problems, such as a period that may be short in the context of modern processors, and they can have some problems with statistical tests[25]

A **multiple recursive generator** (MRG) improves on LCGs at the expense of added complexity by expanding the depth of the recurrence. An MRG of order k has the form

$$x_{n+1} = (a_1 x_n + a_2 x_{n-1} + \cdots + a_k x_{n-k+1} + c) \mod m.$$

(multiple recursive generator) (10.67)

[19] Alan M. Ferrenberg, D. P. Landau, and Y. Joanna Wong, ''Monte Carlo simulations: Hidden errors from 'good' random number generators,'' *Physical Review Letters* **69**, 3382 (1992) (doi: 10.1103/PhysRevLett.69.3382). The authors found that certain Monte-Carlo methods combined with certain random-number generators (which otherwise passed standard statistical tests) produced clear systematic errors in the fourth or fifth decimal places of calculated quantities in the 2D Ising model.

[20] For a particularly amusing discussion, see Donald E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd ed. (Addison Wesley, 1998), Chapter 3.

[21] George Marsaglia and Wai Wan Tsang, ''Some difficult-to-pass tests of randomness,'' *Journal of Statistical Software* **7**, No. 3, 1 (2002).

[22] Donald E. Knuth, *op. cit.*, Section 3.3, p. 41.

[23] All of the algorithms described here are implemented in the Fortran 90 module/library `RANDOM_PL`, available online at `http://atomoptics.uoregon.edu/~dsteck/computer.html`. Also implemented are the shuffling and combination algorithms for any combination of the three generators, as well as facilities for checkpointing, generating vectors of numbers, generating Gaussian numbers, and running multiple equivalent (but differently seeded) generators in parallel. These algorithms are also implemented in the `SDERK` module mentioned above.

[24] see Donald E. Knuth, *op. cit.*, Sections 3.2 and 3.3 for a detailed discussion of LCGs.

[25] Donald E. Knuth, *op. cit.*, Sections 3.3.3 and 3.3.4.

L'Ecuyer's combined multiple recursive generator[26] uses *two* MRGs, where the first generator hsa coefficients

$$\begin{aligned} m_1 &= 2\ 147\ 483\ 647 = 2^{31} - 1 \\ a_{11} &= 0 \\ a_{12} &= 63\ 308 \\ a_{13} &= -183\ 326 \\ b_1 &= 0 \end{aligned} \tag{10.68}$$

while the second has

$$\begin{aligned} m_2 &= 2\ 145\ 483\ 479 \\ a_{21} &= 86\ 098 \\ a_{22} &= 0 \\ a_{23} &= -539\ 608 \\ b_2 &= 0. \end{aligned} \tag{10.69}$$

Some of the coeffiecients are zero as a compromise between speed and quality. The idea then is to run *both* generators simultaneously and *combine* their outputs via

$$x_n = (x_n^{(1)} + x_n^{(2)}) \mod m_1. \tag{10.70}$$

Since the first random-number generator is uniform over the positive range $[0, m_1 - 1]$, we can think of it being a uniform distribution on the circle. Adding the second random number corresponds to a random rotation of the circle, which doesn't affect its distribution. This combination further breaks up any correlations or problems that might occur with each individual generator. This combined generator has a period of about 2^{185}, and outputs integers in the range $[0, 2^{31} - 2]$, which can then be divided by 2^{31} to produce uniform random numbers on $[0, 1)$.

This generator can be seeded in an elegant way that guarantees that different seeds will produce sequences that don't overlap those of other seeds for a maximally long time. First, six initial numbers are needed to start the recurrence, so we can just pick them to be some fixed but otherwise arbitrary, "random" numbers. For concreteness, we can choose

$$\begin{aligned} x_1^{(1)} &= 1\ 852\ 689\ 663 \\ x_2^{(1)} &= 1\ 962\ 642\ 687 \\ x_3^{(1)} &= 580\ 869\ 375 \\ x_1^{(2)} &= 2\ 039\ 711\ 750 \\ x_2^{(2)} &= 1\ 671\ 394\ 257 \\ x_3^{(2)} &= 879\ 888\ 250. \end{aligned} \tag{10.71}$$

Now the idea is that with the MRGs, we can efficiently skip ahead to *anywhere* in the random sequence via a divide-and-conquer algorithm.[27] To understand this, first note that the MRGs can be implemented as the linear transformation

$$\begin{bmatrix} x_{n+1} \\ x_n \\ x_{n-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_n \\ x_{n-1} \\ x_{n-2} \end{bmatrix} \mod m =: \mathbf{A} \begin{bmatrix} x_n \\ x_{n-1} \\ x_{n-2} \end{bmatrix} \mod m \tag{10.72}$$

on the three-dimensional state of the generator. Then advancing the generator forward n times is equivalent to instead applying the matrix $(\mathbf{A}^n) \bmod m$ Then since the period is just slightly more than 2^{184}, and if we

[26] Pierre L'Ecuyer, "Combined Multiple Recursive Random Number Generators," *Operations Research* **44**, 816 (1996), specifically the generator in Example 4.

[27] This algorithm is described in depth by Pierre L'Ecuyer, "Random Numbers for Simulation," *Communications of the ACM* **33**, 85 (1990) (doi: 10.1145/84537.84555).

seed with a single 32-bit integer, there are 2^{32} possible seed values. Thus, we can start different seeds 2^{184-32} or 5.7×10^{45} iterations apart. Thus if s is the seed value, we must apply the matrices $(\mathbf{A}^{2^{184-32}})^s \bmod m$ to the initial state of the random number generators, where s is possibly of the order of 2^{32}. This seems like an absurdly large amount of matrix multiplication just to get started!

The trick behind the divide and conquer algorithm is to compute the large power of the matrix $\mathbf{A}$ *recursively*, according to

$$(\mathbf{A}^n) \bmod m = \begin{cases} \mathbf{A} & (n = 1) \\ (\mathbf{A}\mathbf{A}^{n-1}) \bmod m & (n > 1,\ n \text{ odd}) \\ (\mathbf{A}^{n/2}\mathbf{A}^{n/2}) \bmod m & (n > 1,\ n \text{ even}). \end{cases}$$

(divide-and-conquer algorithm) (10.73)

In this way, $\mathbf{A}^n$ may be computed in only $O(\log_2 n)$ matrix multiplications. Specifically, $(\mathbf{A}^{2^{184-32}}) \bmod m$ may be computed in about 152 matrix multiplications. Then the matrix $(\mathbf{A}^{2^{184-32}})^s \bmod m$ needed to compute the seed can be computed in at most about 32 extra matrix multiplications, which is certainly quite feasible.

10.8.1.2 Knuth's Lagged-Fibonacci Generator

The second generator we mention is Knuth's subtractive, lagged-fibonacci generator.[28] This generator implements the integer recurrence

$$X_j = (X_{j-100} - X_{j-37}) \bmod 2^{30},$$

(lagged–Fibonacci generator) (10.74)

which is obviously a multiple recursive generator. However, to improve the random numbers here, 1009 numbers can be generated at a time, but only the first 100 numbers used (the rest to be discarded). Knuth provides an algorithm for initializing the first block of values so that the recurrence can take over; this is done in such a way that each seed in the range $[0, 2^{30} - 3 = 1\ 073\ 741\ 821]$ gives a distinct sequence for roughly at least the first 2^{70} numbers (but take away a factor of 10 to account for discarding numbers). This generator has a period of $2^{29} \cdot (2^{100} - 1)$, or 10^{38} numbers, which is quite good (again, take away a factor of 10 to account for discarding numbers). The generator gives a 32-bit integer output in the range $[0, 2^{31} - 2]$ (note that the least significant bit is always 0, so the number is always odd), which can then be divided by 2^{31} to produce a uniform random number on $[0, 1)$.

10.8.1.3 Mersenne Twister

The Mersenne Twister generator of Matsumoto and Nishimura is rather more complicated than the other algorithms, being of much higher dimension. Thus we do not describe the algorithm, but refer to the original reference, where C source code is included.[29] The reason we mention it is that it has an incredible period of $2^{19937} - 1$, and so by this standard is an extremely good generator, while still being fairly computationally efficient. The state of this generator is represented by 624 32-bit integers, so the same number of initial values are required. This method can be seeded by a single number simply by using it as the first integer, and using a simple LCG method to seed the rest. The period is so long that it is exceedingly unlikely that two seeds will produce overlapping sequences on any reasonable scale.

10.8.1.4 Randomizing Random Numbers

There are a few methods to try to improve the quality of random-number generators like the ones above. The first method is to ''shuffle'' or ''scramble'' the output of one random-number generator, based on the output of another.[30] The idea is to maintain a ''buffer'' table of, say, 64 output numbers filled by the

[28] Donald E. Knuth, *op. cit.*, p. 186, Eq. (2) and the following discussion.

[29] Makoto Matsumoto and Takuji Nishimura, ''Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator,'' *ACM Transactions on Modeling and Computer Simulation* **8**, 3 (1998) (doi: 10.1145/272991.272995).

[30] Donald E. Knuth, *op. cit.*, p. 33, Algorithm M.

primary generator. Then a random number is chosen from the second generator, and based on this, one of the numbers in the buffer table is chosen to be the output number, and it is replaced by the primary generator. This procedure extends the period of the shuffled sequence to be the least common multiple of the two parent sequences, which for the above methods is to say incredibly long. An alternative is to to a similar shuffle of a random sequence using its own randomness to choose which element of the shuffle table to output. This is the **Bays–Durham shuffle**[31] Then for example, the last output random number is used to select which element of the scramble table to output. Even though only one generator is involved, the method still improves the ''randomness'' of the parent generator. Both of these shuffling methods break up short-range correlations that may exist in the methods above.

The other method for combining two random-number generators is the subtraction we mentioned above for the L'Ecuyer algorithm. Combining in a second generator can't make the first one worse, but may help if it has some subtle defects. Combining three random number generators via subtraction and scrambling gives the ultimate security for the paranoid compuational scientist.

10.8.2 Gaussian Distribution

To pick random deviates from a *standard normal* distribution, or a Gaussian distribution with zero mean and unit variance, with probability density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{10.75}$$

it turns out to be convenient to do so by generating *two* uniform deviates at a time, and then transform them to Gaussian deviates. The first algorithm we mention for doing this is the **Box–Muller method**,[32] To understand it, we first write the joint distribution for two standard normal deviates as

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \tag{10.76}$$

which we can transform to polar coordinates by setting $f(x,y)\,dx\,dy = f_r(r,\theta)\,dr\,d\theta$ to obtain

$$f_r(r,\theta) = \frac{1}{2\pi} r\, e^{-r^2/2} \tag{10.77}$$

via the usual polar transformation. This is still separable, as we may regard the polar distribution to be the product of

$$f_\theta(\theta) = \frac{1}{2\pi}, \qquad f_r(r) = r\, e^{-r^2/2}, \tag{10.78}$$

where the first distribution is obviously uniform over $[0, 2\pi)$. For the second, we may set

$$du = f_r(r)\,dr = r\, e^{-r^2/2}\,dr = -d\left(e^{-r^2/2}\right), \tag{10.79}$$

then it is consistent to identify

$$u = 1 - \left(e^{-r^2/2}\right), \tag{10.80}$$

or solving for r,

$$r = \sqrt{-2\log(1-u)}. \tag{10.81}$$

Of course, $1-u$ has the same distribution as u, so we may as well use it instead:

$$r = \sqrt{-2\log u}. \tag{10.82}$$

[31] Carter Bays and S. D. Durham, ''Improving a Poor Random Number Generator,'' *ACM Transactions on Mathematical Software* **2**, 59 (1976) (doi: 10.1145/355666.355670); Donald E. Knuth, *op. cit.*, p. 34, Algorithm B.

[32] G. E. P. Box, Mervin E. Muller, ''A Note on the Generation of Random Normal Deviates,'' *Annals of Mathematical Statistics* **29**, 610 (1958).

What this means is that if we take u to be a uniform deviate and use this transformation to find r, then r will have the distribution $f_r(r)$ in Eqs. (10.78). The boundary conditions are that $u = 1$ corresponds to $r = 0$ and $u = 0$ corresponds to $r \longrightarrow \infty$, with the transformation $r(u)$ being a monotonic (and one-to-one) function.

So, to summarize the Box–Muller algorithm, we choose U_1 and U_2 to be uniform deviates on $[0, 1)$, and set $U = U_1$ and $\Theta = 2\pi U_2$. Then $r = \sqrt{-2\log u}$, and since $x = r\cos\theta$ and $y = r\sin\theta$, then the two transformed numbers

$$Z_1 = \sqrt{-2\log(1-U_1)}\,\cos(2\pi U_2), \qquad Z_2 = \sqrt{-2\log(1-U_1)}\,\sin(2\pi U_2), \tag{10.83}$$

(Box–Muller transformation)

are two independent, standard-normal deviates. Note here that we were careful to compute the logarithm of $1 - U_1$ rather than U_1 here—if U_1 is uniform on $[0, 1)$ (as is the case for most random-number generators on computers) it is best to avoid the (rare but problematic) possibility of computing $\log(0)$, whereas $\log(1)$ is no problem at all.

An variation on the Box–Muller method is the **polar Marsaglia method**.[33] Suppose w_1 and w_2 define the coordinates for a uniformly distributed random number inside the *unit circle*. Then let $R^2 = W_1^2 + W_2^2$. From the usual transformation to polar coordinates, the probability density for R is $2R$ (the area element at R has an area that scales as R, and the 2 is fixed by normalization). Then since $d(R^2) = 2R\,dR$, we can identify the probability density for R^2 as being *uniform* on $[0, 1)$. Thus, we may transform R to the radius r in the 2D normal distribution via Eq. (10.82) (identifying R with u). Finally, since $\sin\theta = w_2/R$ and $\cos\theta = w_1/R$, we may compute the trigonometric functions in Eq. (10.83) by these simple ratios, so that

$$Z_1 = W_1\sqrt{-\frac{2\log R^2}{R^2}}, \qquad Z_2 = W_2\sqrt{-\frac{2\log R^2}{R^2}}. \tag{10.84}$$

(polar Marsaglia method)

The advantage here is that the sin and cos functions are eliminated, and they may potentially be computationally expensive. Now the remaining issue is, how to pick W_1 and W_2? Simply pick the usual uniform deviates U_1 and U_2 on $[0, 1)$, and take $W_{1,2} = 2U_{1,2} - 1$ to create a uniform distribution on the unit box from $[-1, 1)$ in both directions. Now compute $R^2 = W_1^2 + W_2^2$. If $R^2 < 1$, then proceed with the transformation; otherwise pick new deviates U_1 and U_2 and try again until you end up with a pair inside the unit circle. The probability on one attempt of succeeding by ending up in the unit circle is $\pi/4 \approx 78.5\%$; on average, the random numbers will need to be picked $(1)(\pi/4) + (2)(1-\pi/4)(\pi/4) + (3)(1-\pi/4)^2(\pi/4) + \cdots = 4/\pi \approx 1.27$ times to successfully generate the two normal numbers.

In principle, the polar Marsaglia method should be faster than the Box–Muller method because the sin and cos functions can be expensive to evaluate. However, this is tempered by the presence of a division and the generation of extra random numbers. To test this, 10^8 normal deviates were generated on a relatively modern computer (2.16 GHz Intel Core Duo) in Fortran 90. Using Knuth's lagged Fibonacci generator, the polar Marsaglia method is indeed faster by about 10-15%. However, using a more complicated generator (the L'Ecuyer combined MRG subtractively mixed with the Mersenne twister, then scrambled by Knuth's generator), the two methods were about equally fast. Thus, either method is acceptable, and the Box–Muller method may even be preferable in that the generator will advance by a known number of iterations when generating normal deviates.

10.8.3 Other Distributions

While not directly useful in simulations of the type of SDEs that we considered at the start of the chapter, the methods so far for generating random numbers are useful in generating random deviates with other distributions useful in quantum optics, so it is worth considering these briefly here.

[33] G. Marsaglia, "Improving the polar method for generating a pair of random variables," Boeing Scientific Research Laboratory report D1-82-0203 (1962); Donald E. Knuth, *op. cit.*, Algorithm P (p. 122).

10.8.3.1 Angular Distributions

One useful class of random-deviate distributions are *angular* distributions, as for choosing random spontaneous-emission directions. A simple idea here is to simply use the rejection method, as we mentioned above in the polar Marsaglia method to convert a uniform variate on the unit box to a uniform variate on the unit circle. In that case, we simply rejected any deviate that fell outside the desired circle. Thus, suppose we have a distribution function $f(\theta, \phi)$ that we wish to simulate. Our procedure will be a ''double rejection'' as follows:

1. Choose three uniform deviates U_1, U_2, and U_3 on $[0, 1)$.
2. Convert them to obtain a uniform deviate in the unit cube (that is, from $[-1, 1)$ in all directions) by setting $X = 2U_1 - 1$, $Y = 2U_2 - 1$, and $Z = 2U_3 - 1$.
3. Obtain a uniform deviate in the unit sphere by rejection: Compute $R^2 = X^2 + Y^2 + Z^2$ and continue if $R^2 \leq 1$ (and also compute $R = \sqrt{R^2}$); go back to step 1 otherwise.
4. Convert to spherical coordinates by computing

$$\begin{aligned} \Theta &= \tan^{-1} \frac{\sqrt{X^2 + Y^2}}{Z} \\ \Phi &= \tan^{-1} \frac{Y}{X}. \end{aligned} \tag{10.85}$$

   Note that for computing Φ, the arctangent must be defined such that it can distinguish between arguments such as $(X, Y) = (1, 1)$ and $(X, Y) = (-1, -1)$, which would naively give the same answer. This is handled in Fortran and other languages by the `atan2` function.
5. Obtain a deviate with the desired angular distribution by testing to see if

$$U_4 \leq \frac{f(\Theta, \Phi)}{\sup\{f(\theta, \phi), \theta \in [0, \pi], \phi \in [0, 2\pi]\}}, \tag{10.86}$$

   where U_4 is another uniform deviate on $[0, 1)$. To save from running the random-number generator again, we can alternately use R^3 in place of U_4, since it is likewise a uniform deviate on $[0, 1)$, and it is independent of the angles. We normalize by the maximum possible value of $f(\theta, \phi)$ to maximize the efficiency of the method. If the condition is true, we're done; if not, start over at step 1.

For example, if we were generating a dipole distribution

$$f(\theta, \phi) = \frac{3}{8\pi} \sin^2 \theta, \tag{10.87}$$

in step 5 we would check to see if

$$R^3 \leq \sin^2 \Theta, \tag{10.88}$$

and if so, keep the generated angles. In this example, the success probability for each triple of uniform deviates $U_{1,2,3}$ to generate a uniform deviate in the unit sphere is $\pi/6 \approx 52\%$. Once this happens, the probability of generating a deviate with the dipole distribution pattern is $2/3 \approx 67\%$, for a combined success probability of $\pi/9 \approx 34.9\%$. The first probability is given by the volume ratio of the unit sphere and cube, while the second is given by the volume under the ''unwrapped'' surface $\sin^3 \theta$ (i.e., *not* the volume inside the boundary $r = \sin^2 \theta$), relative to the same volume under the surface $\sin \theta$, or simply 4π.

Of course, the normalized values X/R, Y/R, and Z/R give the projections of the unit vector pointing in the (Θ, Φ) direction along the x, y, and z directions.

A slightly more efficient method generates the angle values more directly. The normalized probability density for equally-likely emission in any direction is

$$f(\theta, \phi) = \frac{\sin \theta}{4\pi}, \tag{10.89}$$

with normalization convention

$$\int_0^\pi d\theta \int_0^{2\pi} d\phi\, f(\theta,\phi) = 1. \tag{10.90}$$

That is, the ϕ variable is already uniform, so we can handle it by picking a uniform deviate u_2 and rescaling: $\phi = 2\pi u_2$. Thus, we need only consider the θ variable, with normalized density

$$f(\theta) = \frac{1}{2}\sin\theta. \tag{10.91}$$

If we pick a second uniform deviate u_1 and equate probability densities via $du_1 = (1/2)\sin\theta\, d\theta$, note that this is equivalent to equating the cumulative probability functions. For the uniform deviate, the cumulative distribution is simply u_1, while for θ, the cumulative distribution is

$$F(\theta) = \frac{1}{2} - \frac{1}{2}\cos\theta = \sin^2\theta. \tag{10.92}$$

Equating the cumulative distributions, we obtain $\theta = \sin^{-1}\sqrt{u_1}$. Thus, to summarize, we should choose two uniform deviates U_1 and U_2, and then compute the angles via

$$\Theta = \sin^{-1}\sqrt{U_1}, \qquad \Phi = 2\pi U_2. \tag{10.93}$$

(transformation from uniform to angular variables)

This saves one uniform random deviate compared to the rejection method (plus any rejected deviates), although it requires evaluating the more complicated arcsine function.

10.8.3.2 Exponential Distributions

Another simple example is the exponential distribution,

$$f(x) = \gamma\, e^{-\gamma x}, \tag{10.94}$$

defined on the nonnegative half of the real line. The cumulative density is an exponential rise

$$F(x) = 1 - e^{-\gamma x}, \tag{10.95}$$

and the inverse of this function gives x in terms of a uniform deviate U:

$$X = -\frac{1}{\gamma}\log(1-U). \tag{10.96}$$

(exponential-deviate generation)

Of course, the result is always a positive real number. Also, U can be substituted for U in the logarithm, but if the random-number generator gives $U \in [0,1)$, the above form guards against the possibility of crashing the code by computing $\log(0)$.

10.8.3.3 Power-Law Distributions

Suppose we have a power-law probability density defined between a and b:

$$f(x) = \eta x^\gamma, \qquad 0 < a \le x \le b. \tag{10.97}$$

We have chosen the bounds "safely" here, but in some cases they may be negative, too. Here, η is a normalization factor, given by

$$\eta = \begin{cases} \dfrac{\gamma+1}{b^{\gamma+1} - a^{\gamma+1}} & \text{if } \gamma \neq -1 \\ \dfrac{1}{\log b - \log a} & \text{if } \gamma = -1. \end{cases} \tag{10.98}$$

Note that we can safely take $a \longrightarrow 0$ if $\gamma > -1$, otherwise we need $a > 0$. Similarly, we can safely take $b \longrightarrow \infty$ if $\gamma < -1$, otherwise we need $b < \infty$. If $\gamma = 1$, we strictly need $0 < a < b < \infty$.

First, in the case $\gamma \neq -1$, we have a cumulative distribution

$$F(x) = \frac{x^{\gamma+1} - a^{\gamma+1}}{b^{\gamma+1} - a^{\gamma+1}} \qquad (\gamma \neq -1). \tag{10.99}$$

Setting this equal to a cumulative distribution u for a uniform deviate, and solving, we find the algorithm

$$X = \sqrt[\gamma+1]{a^{\gamma+1} + (b^{\gamma+1} - a^{\gamma+1})U} \;=\; \sqrt[\gamma+1]{a^{\gamma+1}(1-U) + b^{\gamma+1}U} \qquad (\gamma \neq -1)$$

(power-law-deviate generation) (10.100)

for a power-law deviate X in terms of a uniform deviate U. However, if $\gamma = -1$, then we have the cumulative distribution

$$F(x) = \frac{\log x - \log a}{\log b - \log a} \qquad (\gamma = -1). \tag{10.101}$$

This leads to the transformation

$$X = \exp\big[\log a + (\log b - \log a)U\big] = \exp\big[(1-U)\log a + U \log b\big] \qquad (\gamma = -1).$$

(power-law-deviate generation) (10.102)

In both cases, the transformation is a weighted average of the endpoints (according to the value U), within a nonlinear transformation (power-law or logarithm).

One context in which the power-law distribution comes up in the Monte-Carlo evaluation of an integral of the form

$$I = \int_0^\infty dx\, x^\gamma f(x) \qquad (\gamma < -1), \tag{10.103}$$

where the restriction on γ is just to simplify the following analysis. Clearly, this integral only converges if $f(x)$ converges to a finite constant as $x \longrightarrow \infty$, and if $f(x)$ converges quickly enough to zero as $x \longrightarrow 0$. For the power-law distribution to do us any good here we must have a lower bound $a > 0$. We will thus further assume that $f(x) = 0$ for all $x < x_{\min}$ (or at least that the region $x < x_{\min}$ contributes negligibly to the integral—we can always find some appropriate boundary that satisfies this condition to any desired precision). Then

$$I = \int_{x_{\min}}^\infty dx\, x^\gamma f(x) \qquad (\gamma < -1), \tag{10.104}$$

and thus we can take advantage of the power-law distribution to turn this into an ensemble average,

$$I = \eta^{-1} \big\langle\!\big\langle f(x) \big\rangle\!\big\rangle_x = \frac{-x_{\min}^{\gamma+1}}{\gamma+1} \big\langle\!\big\langle f(x) \big\rangle\!\big\rangle_x \qquad (\gamma < -1), \tag{10.105}$$

where from Eq. (10.100), we generate the values of x according to

$$X = x_{\min}(1-U)^{1/(\gamma+1)} \qquad (\gamma < -1). \tag{10.106}$$

Note again that we could replace $1 - U$ by U to simplify the formula here, but it is safer if U is defined on $[0, 1)$ to leave it in this form, so that the (rare) occurrence of the value $U = 0$ does not pose any problem.

10.8.3.4 Gamma Distribution

A more complicated distribution, something of a combination of power-law and exponential distributions, is the **gamma distribution**

$$f(x; \gamma, \beta) = \frac{\beta^\gamma}{\Gamma[\gamma]}\, x^{\gamma-1} e^{-\beta x} \qquad (x > 0,\ \gamma > 0,\ \beta > 0), \tag{10.107}$$

named thusly due to the gamma function in the normalization factor. In the case of $\gamma = 1$, this is simply an exponential distribution, as in Eq. (10.94).

One useful interpretation of the gamma distribution comes from the property that if $X_1 \sim f(x;\gamma_1,\beta)$ (i.e., X_1 is gamma distributed with parameters γ_1 and β) and $X_2 \sim f(x;\gamma_2,\beta)$, then (Problem 10.2)

$$X = X_1 + X_2 \sim f(x;\gamma_1+\gamma_2,\beta). \tag{10.108}$$

Thus, for example, if γ is an integer, the gamma-distributed deviate X can be interpreted as the sum of γ, independent exponential deviates. This also means that as γ becomes large, $f(x;\gamma,\beta)$ should converge to a Gaussian distribution via the central limit theorem. It is also fairly simple to choose a gamma-distributed deviate if $\gamma = a$ is a relatively small integer, since the combination of exponential deviates

$$X = -\frac{1}{\beta}\sum_{j=1}^{a}\log(1-U_j) = -\frac{1}{\beta}\log\left(\prod_{j=1}^{a}(1-U_j)\right)$$

(gamma-distributed deviate, integer γ) (10.109)

works, for a uniform deviates U_j. The latter form is more efficient on a computer, since it avoids the calculation of multiple (expensive) logarithms.

Another useful special case of the gamma distribution occurs for $\gamma = 1/2$:

$$f(x;{}^1\!/_2,\beta) = \sqrt{\frac{\beta}{\pi x}}\,e^{-\beta x} \qquad (x>0,\ \beta>0). \tag{10.110}$$

Changing variables via $x = z^2/2$ and $dx = z\,dz = \sqrt{2x}\,dz$ leads to

$$f(z;{}^1\!/_2,\beta) = \sqrt{\frac{2\beta}{\pi}}\,e^{-\beta z^2/2} \qquad (z>0,\ \beta>0). \tag{10.111}$$

This is a Gaussian distribution (or at least the positive half) with variance $1/\beta$. Thus, it is also easy to pick a gamma-distributed deviate for any *half*-integer value of γ. That is, if

$$X = \frac{\beta}{2}\,Z^2 - \frac{1}{\beta}\sum_{j=1}^{a}\log(1-U_j) = \frac{\beta}{2}\,Z^2 - \frac{1}{\beta}\log\left(\prod_{j=1}^{a}(1-U_j)\right),$$

(gamma-distributed deviate, half-integer γ) (10.112)

where a is a positive integer, Z is a standard-normal deviate, and the U_j are uniform deviates, then X is distributed according to $f(x;\gamma = a + {}^1\!/_2,\beta)$.

By the same coordinate change, the $\gamma = 3/2$ case,

$$f(x;{}^3\!/_2,\beta) = \frac{2\beta^{3/2}}{\sqrt{\pi}}\,\sqrt{x}\,e^{-\beta x} \qquad (x>0,\ \beta>0), \tag{10.113}$$

becomes

$$f(z;{}^3\!/_2,\beta) = \sqrt{\frac{\beta^3}{2\pi}}\,z^2\,e^{-\beta z^2/2} \qquad (z>0,\ \beta>0). \tag{10.114}$$

Notice that this is the same as a Gaussian distribution of variance $1/\beta$ for three variables x_1, x_2, and x_3, but transformed into spherical coordinates by $z^2 = x_1^2 + x_2^2 + x_3^2$. Thus, for $\gamma = 3/2$, it is possible to create a deviate by summing the squares of three normal deviates (with appropriate scaling), or to add the square of a normal deviate to an exponential deviate as in Eq. (10.112).

For more general γ (i.e., neither integer nor half-integer), it is in general necessary to use more complicated methods, based on various forms of rejection sampling.[34]

[34] Luc Devroye, *Non-Uniform Random Variate Generation* (Springer, 1986), Section IX.3 (ISBN: 0387963057), `http://www.nrbook.com/devroye/`.

10.9 Exercises

Problem 10.1

Let us define ΔW and I_{10} over the same time interval Δt as usual as

$$\begin{aligned} \Delta W &:= \int_0^{\Delta t} dW(t) \\ I_{10} &:= \int_0^{\Delta t} dt \int_0^t dW(t'). \end{aligned} \tag{10.115}$$

Show that the variance of I_{10} and its covariance with ΔW are given by Eq. (10.16),

$$\begin{aligned} \langle\!\langle I_{10}^2 \rangle\!\rangle &= \frac{\Delta t^3}{3} \\ \langle\!\langle I_{10}\, \Delta W \rangle\!\rangle &= \frac{\Delta t^2}{2} \end{aligned} \tag{10.116}$$

as follows. View I_{10}, defined over a time step Δt, as the solution of the differential equation

$$dy = W(t)\, dt, \tag{10.117}$$

with initial condition $y(0) = 0$, such that $I_{10} = y(\Delta t)$. Then consider a finite-difference approximation to the solution by integrating in N steps of duration $\delta t = \Delta t/N$ according to the Euler method (which becomes exact as $N \longrightarrow \infty$).

Problem 10.2

Prove the addition property of the gamma distribution [Eq. (10.108)]: if $X_1 \sim f(x; \gamma_1, \beta)$ (i.e., X_1 is gamma distributed with parameters γ_1 and β) and $X_2 \sim f(x; \gamma_2, \beta)$, where the gamma probability density is

$$f(x; \gamma, \beta) = \frac{\beta^\gamma}{\Gamma[\gamma]}\, x^{\gamma-1} e^{-\beta x} \qquad (x > 0,\ \gamma > 0,\ \beta > 0), \tag{10.118}$$

then the sum is also gamma-distributed:

$$X = X_1 + X_2 \sim f(x; \gamma_1 + \gamma_2, \beta). \tag{10.119}$$

Problem 10.3

Consider the probability density

$$f(\mathcal{T}) = \frac{(2d^2)^{(D+m)/2}}{\Gamma\left[\frac{D+m}{2}\right]} \frac{e^{-2d^2/\mathcal{T}}}{\mathcal{T}^{1+(D+m)/2}}. \tag{10.120}$$

Show that this can be transformed to a gamma distribution, and describe the generation of deviates $\mathcal{T}$ for $D = 4$ and integer $m \geq 0$.

Problem 10.4

Derive the equations (10.63) relating the multiple integrals on the full time interval $[0, t]$ to the multiple

integrals defined on subintervals $\Delta t = t/N$, including the expressions with one stochastic integral,

$$
\begin{aligned}
W(t) &= \sum_{n=0}^{N-1} \Delta W(n\Delta t) \\
(I_{10})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{10})_{n\Delta t}^{(n+1)\Delta t} + (N-1-n)\Delta t\, \Delta W(n\Delta t) \Big] \\
(I_{100})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{100})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^2}{2}(N-1-n)^2 \Delta W(n\Delta t) \Big] \\
(I_{1000})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{1000})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{100})_{n\Delta t}^{(n+1)\Delta t} \\
&\qquad + \frac{\Delta t^2}{2}(N-1-n)^2 (I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^3}{3!}(N-1-n)^3 \Delta W(n\Delta t) \Big],
\end{aligned}
\tag{10.121}
$$

and two stochastic integrals:

$$
\begin{aligned}
(I_{11})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
(I_{110})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{110})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n) \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
&\qquad + (I_{10})_{n\Delta t}^{(n+1)\Delta t} W(n\Delta t) \Big] \\
(I_{1100})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{1100})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n) \Big((I_{110})_{n\Delta t}^{(n+1)\Delta t} + (I_{10})_{n\Delta t}^{(n+1)\Delta t}\, W(n\Delta t) \Big) \\
&\qquad + \frac{\Delta t^2}{2}(N-1-n)^2 \Big((I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big) + (I_{100})_{n\Delta t}^{(n+1)\Delta t} W(n\Delta t) \Big].
\end{aligned}
\tag{10.122}
$$

Problem 10.5

Consider the stochastic integral

$$
(I)_t := \int_0^t dW(t')\, e^{a(t-t')} \tag{10.123}
$$

for $a \in \mathbb{C}$. Show that this integral may be expanded in terms of multiple Itō integrals as

$$
(I)_t = W(t) + f'(I_{10})_0^t + f'^2(I_{100})_0^t + f'^3(I_{1000})_0^t + \cdots . \tag{10.124}
$$

Problem 10.6

Consider the stochastic integral

$$
(I')_t := \int_0^t dt'\, (I_{11})_0^{t'}\, e^{a(t-t')} \tag{10.125}
$$

for $a \in \mathbb{C}$. Show that this integral may be expanded in terms of multiple Itō integrals as

$$
(I')_t = (I_{110})_0^t + a(I_{1100})_0^t + a^2(I_{11000})_0^t + \cdots . \tag{10.126}
$$

Problem 10.7

Derive the simulation formulae (10.22)

$$\begin{aligned}
\Delta W_n &= \sqrt{\Delta t}\, z_1 \\
(I_{10})_n &= \frac{\Delta t^{3/2}}{2}\left(z_1 + \frac{z_2}{\sqrt{3}}\right) \\
(I_{100})_n &= \frac{\Delta t^{5/2}}{6}\left(z_1 + \frac{\sqrt{3}\, z_2}{2} + \frac{z_3}{2\sqrt{5}}\right) \\
(I_{1000})_n &= \frac{\Delta t^{7/2}}{24}\left(z_1 + \frac{3\sqrt{3}\, z_2}{5} + \frac{z_3}{\sqrt{5}} + \frac{z_4}{5\sqrt{7}}\right),
\end{aligned} \tag{10.127}$$

in terms of standard-normal random deviates z_1, z_2, z_3, and z_4 by writing each multiple integral as a linear combination of these deviates and demanding that they have the correct variances and covariances.

Problem 10.8

Derive the moments and correlations listed below [see Eqs. (10.27) and (10.29)] for the multiple Itō integrals from 0 to t:

$$\begin{aligned}
\left\langle\!\left\langle (I_{11})^2 \right\rangle\!\right\rangle &= \frac{t^2}{2}, & \left\langle\!\left\langle (I_{110})^2 \right\rangle\!\right\rangle &= \frac{t^4}{12}, & \left\langle\!\left\langle (I_{1100})^2 \right\rangle\!\right\rangle &= \frac{t^6}{120} \\
\left\langle\!\left\langle (I_{110})(I_{11}) \right\rangle\!\right\rangle &= \frac{t^3}{3!}, & \left\langle\!\left\langle (I_{1100})(I_{11}) \right\rangle\!\right\rangle &= \frac{t^4}{4!}, & \left\langle\!\left\langle (I_{1100})(I_{110}) \right\rangle\!\right\rangle &= \frac{t^5}{40} \\
\left\langle\!\left\langle (I_{11})^3 \right\rangle\!\right\rangle &= t^3, & \left\langle\!\left\langle (I_{110})^3 \right\rangle\!\right\rangle &= \frac{t^6}{120}, & \left\langle\!\left\langle (I_{1100})^3 \right\rangle\!\right\rangle &= \frac{t^9}{480} \\
\left\langle\!\left\langle (I_{11})^4 \right\rangle\!\right\rangle &= \frac{15t^3}{4}, & \left\langle\!\left\langle (I_{110})^4 \right\rangle\!\right\rangle &= \frac{57t^8}{560}, & \left\langle\!\left\langle (I_{1100})^4 \right\rangle\!\right\rangle &= \frac{53t^{12}}{52800}.
\end{aligned} \tag{10.128}$$

Problem 10.9

Derive the simulation formulae (10.30)

$$\begin{aligned}
(I_{11})_n &= \frac{\Delta t}{2}\left(z_1^2 - 1\right) \\
(I_{110})_n &= \frac{\Delta t^2}{6}\left(z_1^2 - 1 + z_{11}\right)
\end{aligned} \tag{10.129}$$

in terms of standard-normal random deviates z_1, z_{11}, and z_{12}, where $\Delta W_n = \sqrt{\Delta t}\, z_1$ as in Problem 10.7. For the higher-order multiple integrals, assume that they are linear combinations of I_{11} and the other normal deviates, and verify that the formulae generate the appropriate fourth moments.

Part III

Jumps and Lévy Processes

Chapter 11

The Poisson Process

11.1 Poisson Distribution

The **Poisson distribution** is a one-parameter probability distribution defined on the non-negative integers. Intuitively, the Poisson distribution gives the probability for a particular number of events to occur, given the possibility that the number of *independent* random events that may occur is unbounded. With parameter λ, the explicit form of the Poisson distribution for a Poisson random variable N is

$$P(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}, \tag{11.1}$$
(Poisson distribution)

where n is a nonnegative integer. The Poisson distribution is particularly useful in modeling independent (but identical in character) random events that occur in a given interval of time. Classic examples include the number of cars that arrive at an intersection (in the limit of low traffic density) or the number of decay events in an ensemble of radioactive or optically excited atoms, again in the low-density limit. Poisson random variation is also responsible for **shot noise**, which occurs as noise in electrical current due to random fluctuations in the rate at which electrons flow through a device, or as noise in the detected intensity of classical light due to the random detection times of individual photons.

11.1.1 Moments

The moment-generating function for the Poisson distribution is

$$\left\langle\!\left\langle e^{sN} \right\rangle\!\right\rangle = \sum_{n=0}^{\infty} P(n)\, e^{sn} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda\, e^s)^n}{n!}, \tag{11.2}$$

and recognizing the series as the expansion of the exponential function, we find

$$\left\langle\!\left\langle e^{sN} \right\rangle\!\right\rangle = \exp[\lambda(e^s - 1)]. \tag{11.3}$$
(moment-generating function, Poisson distribution)

Then the computation of moments from this result is straightforward; the first moment is

$$\langle\!\langle N \rangle\!\rangle = \partial_s \left\langle\!\left\langle e^{sN} \right\rangle\!\right\rangle \Big|_{s=0} = \lambda\, e^s \exp[\lambda(e^s - 1)]\Big|_{s=0} = \lambda, \tag{11.4}$$

while the second moment is

$$\langle\!\langle N^2 \rangle\!\rangle = \partial_s^2 \left\langle\!\left\langle e^{sN} \right\rangle\!\right\rangle \Big|_{s=0} = (\lambda + \lambda^2)\, e^s \exp[\lambda(e^s - 1)]\Big|_{s=0} = \lambda + \lambda^2. \tag{11.5}$$

This latter result gives the variance as the difference $\langle\langle N^2\rangle\rangle - \langle\langle N\rangle\rangle^2$. We can summarize the mean and variance as

$$\langle\langle N\rangle\rangle = \lambda, \qquad \mathrm{Var}[N] = \lambda. \tag{11.6}$$

(mean and variance, Poisson distribution)

This shows the significance of the parameter λ: it gives both the mean number of events as well as the variance. Of course, it is straightforward enough to to compute these statistics directly from the probability distribution, without going through the moment-generating function (Problem 11.2).

11.1.2 Limiting Case of the Binomial Distribution

To gain a little more insight into the Poisson distribution, we can consider N random, independent ''trials,'' where each trial is a success with probability p and a failure with probability $1-p$. The probability to have n successes ($0 \le n \le N$) is

$$P(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}. \tag{11.7}$$

(binomial distribution)

To explain this probability, note that the probability for the first n trials to succeed is p^n, and the remaining trials to fail is $(1-p)^{N-n}$. There are $N!$ ways to reorder N trials, but $n!$ of the rearrangements of the successful trials are redundant, and $(N-n)!$ rearrangements of the failed trials are redundant, leaving

$$\binom{N}{n} := \frac{N!}{n!(N-n)!} \tag{11.8}$$

(binomial coefficeint)

independent rearrangements of the N trials, given exactly n successes. This counting factor is important, and commonly called the **binomial coefficient** (the symbol is colloquially read as ''N choose n'').

The Poisson distribution is similar, but applies in the limit $N \longrightarrow \infty$, where the probability of success in each trial becomes vanishingly small, such that the probability of a fixed number of successes converges to a constant. In this spirit, we can take $p = \lambda/N$ in the binomial distribution, so that

$$P(n) = \frac{N!}{n!(N-n)!} \frac{\lambda^n}{N^n} \left(1 - \frac{\lambda}{N}\right)^{N-n}. \tag{11.9}$$

The last factor here converges to e^λ as $N \longrightarrow \infty$, and we can also use Stirling's approximation

$$N! \sim \sqrt{2\pi N}\, N^N e^{-N} \tag{11.10}$$

for the asymptotic behavior of the factorials, with the result

$$P(n) = \frac{N^N e^{-N}}{n!(N-n)^{N-n} e^{-(N-n)}} \frac{\lambda^n}{N^n} e^{-\lambda} = \frac{e^{-n}}{(1-n/N)^{N-n}} \frac{\lambda^n}{n!} e^{-\lambda}. \tag{11.11}$$

Again, in the limit $N \longrightarrow \infty$, the first factor converges to unity, and we thus recover the Poisson distribution in the large-N limit. This emphasizes a key aspect of the Poisson distribution: there are many possibilities for an event to occur, but no event occurs in the vast majority of the possbilities.

This was a kind of ''top-down'' derivation of the Poisson process. The analogous ''bottom-up'' derivation is coming up once we define the Poisson process, and treat explicitly a vanishingly small probability of success for each of many trials.

11.2 Poisson Process

In general, we can speak of a *rate* at which events occur by setting $\lambda = \Gamma\,\Delta t$ for finite time interval Δt, where Γ is the mean rate of occurence (also called the **intensity** of the Poisson process). Then

$$P(n) = \frac{e^{-\Gamma\,\Delta t}(\Gamma\,\Delta t)^n}{n!}. \tag{11.12}$$

Note that the Poisson distribution implies an exponential waiting time for the first event, because the probability for the event to occur *after* waiting a time Δt is given by setting $n = 0$ in the above probability function:

$$P(0) = e^{-\Gamma\,\Delta t}. \tag{11.13}$$

Then according to our interpretation, this probability is related to the probability density $P'(t)$ for the time of first occurence by

$$e^{-\Gamma\,\Delta t} = \int_{\Delta t}^{\infty} P'(t)\,dt, \tag{11.14}$$

so that

$$P'(t) = \Gamma e^{-\Gamma t}. \tag{11.15}$$

Thus, Poisson random variables are intimately connected with exponential-decay processes, such as spontaneous emission from an atom prepared in the excited state.

In the infinitesimal limit $\Delta t \longrightarrow dt$, all the probabilities for $n \geq 2$ becomes negligible (of order dt^2 or higher). The probability for a single event occuring during an interval of duration dt thus becomes $\Gamma\,dt$, with no events occuring otherwise. We can denote this by the infinitesimal random variable $dN(t)$—the **Poisson process**—which has an ensemble mean

$$\langle\!\langle dN(t) \rangle\!\rangle = \Gamma\,dt. \tag{11.16}$$
(Poisson process: ensemble mean)

In the standard Poisson process $N(t)$, the intensity Γ is a constant that characterizes the process. Thus, in general, when writing down a Poisson process, you must always also specify the intensity, which is *not* specified in the notation dN in contrast to the Wiener process dW (where there is no freedom to specify the moments). In generalizations of the Poisson process to time- or state-dependent intensities (see Section 11.3), an explicit specification of the intensity is even more critical.

Again, as an integer-valued differential random variable, dN can take on only the values 0 and 1, where the value of 1 occurs with probability equal to the mean. Because $dN(t) \in \{0, 1\}$, it immediately follows that

$$dN^2 = dN, \tag{11.17}$$
(Poisson-process property)

so that the ensemble-averaged variance is equal to the ensemble mean,

$$\langle\!\langle dN^2 \rangle\!\rangle = \langle\!\langle dN \rangle\!\rangle = \Gamma\,dt, \tag{11.18}$$

as we expect for a Poisson-distributed random variable. Note that the infinitesimal variance here is just the second moment, since the square of the mean is $O(dt^2)$.

In another view, note that $dN(t)/dt$ is zero except in isolated intervals of length dt, where the value is $1/dt$. Thus, we can write this form of the Poisson process as the sum of delta functions,

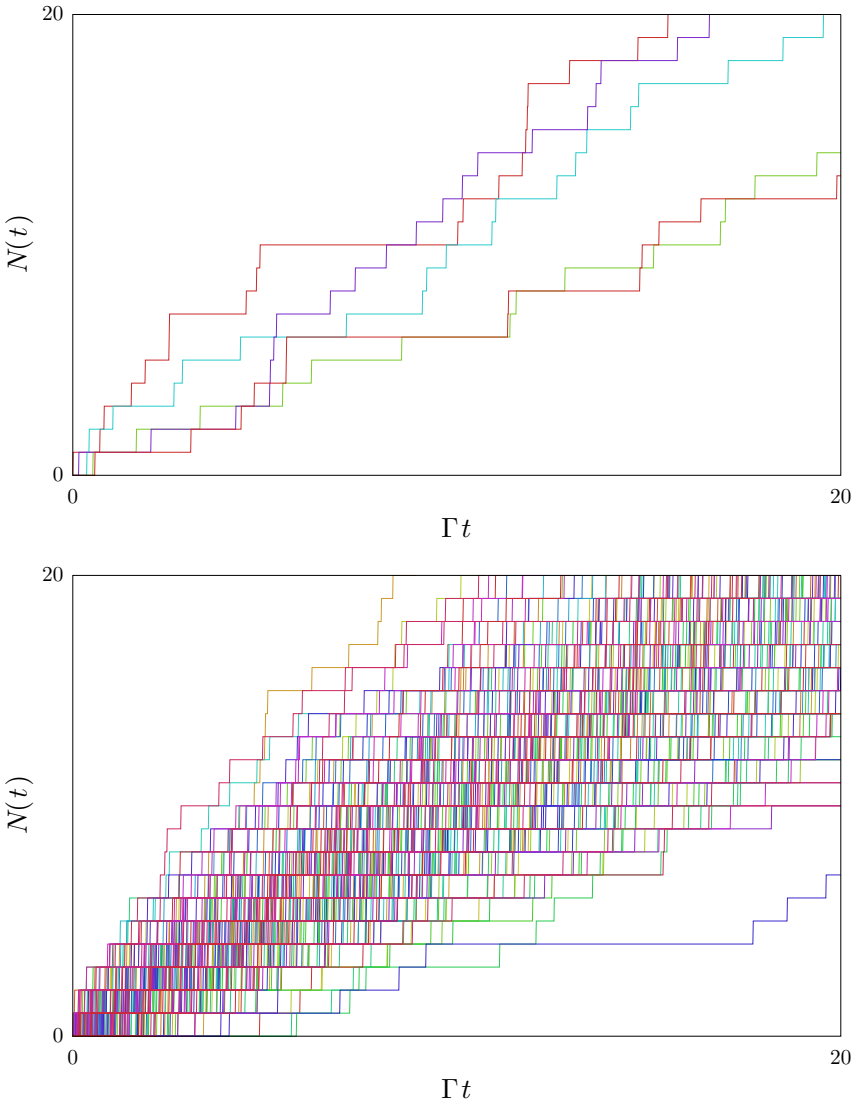
$$\frac{dN(t)}{dt} = \sum_j \delta(t - t_j), \tag{11.19}$$

if the events occur at times t_j. In view of our discussion above, $\Delta t_j := t_{j+1} - t_j$ is a random variable with probability density

$$P(\Delta t_j) = \Gamma e^{-\Gamma\,\Delta t_j}, \tag{11.20}$$

since the waiting time until the next event is given by the exponential distribution. Intuitively, these delta functions represent, say, the clicks of a Geiger counter.

Shown below are plots of 5 (first plot) and 200 (second plot) Poisson processes.



The derivative $dN(t)/dt$ of each of these processes is again just a bunch of delta functions, located at each of the discontinuities of $N(t)$.

11.2.1 The Poisson Process Implies the Poisson Distribution

If we take the Poisson process $dN(t)$ of intensity Γ (i.e., of mean $\Gamma\, dt$) as the fundamental object, we should also be able to derive the Poisson distribution for the frequency of events in finite time intervals. That is, we can show that the Poisson distribution arises if there is a constant probability per unit time of a *single* event occuring during an arbitrarily short time interval. Specifically, defining the time integral of the Poisson process,

$$\Delta N := \int_t^{t+\Delta t} dN(t') \tag{11.21}$$

we can ask, what is the probability distribution for ΔN? For a given value n of ΔN, this means that during exactly n infinitesimal intervals, $dN(t)$ took on the value unity, while it took on the value of zero during the remaining intervals. The probability of doing so is the product of three factors:

1. The probability for having exactly n such events, one in each of n particular time intervals: $(\Gamma\, dt)^n$.

2. The number of ways to distribute the n events among all such intervals. In the time interval $[t, t+\Delta t)$, there are $\Delta t/dt$ such time intervals, and so there are $(\Delta t/dt)^n$ ways to distribute n events among all possible time intervals. But we divide by $n!$ since we take the n events to be indistinguishable, so we don't overcount, so the total factor is $(\Delta t/dt)^n/n!$.

3. The probability of having zero events in *all other* time intervals. Again, there are $(\Delta t/dt)$ total time intervals, and the probability of zero events in a given time interval is $(1-\Gamma\,dt)$, so the total probability is

$$(1-\Gamma\,dt)^{\Delta t/dt} = \lim_{M\to\infty}\left(1-\frac{\Gamma\,\Delta t}{M}\right)^M = e^{-\Gamma\,\Delta t}. \tag{11.22}$$

We are being cavalier in slinging around factors of dt, but our manipulations here are equivalent to the "correct" approach of using finite, small subintervals δt, where we neglect n compared to $\Delta t/\delta t$, and we neglect the probability that two events end up in the same subinterval. Both of these approximations are appropriate (and exact) in the continuum limit. The total probability is thus

$$\begin{aligned} P(\Delta N = n) &= (\Gamma\,dt)^n\left(\frac{\Delta t}{dt}\right)^n\frac{1}{n!}e^{-\Gamma\,\Delta t} \\ &= \frac{e^{-\Gamma\,\Delta t}(\Gamma\,\Delta t)^n}{n!} \\ &= \frac{e^{-\lambda}\lambda^n}{n!}, \end{aligned} \tag{11.23}$$

which is the Poisson distribution, where again the mean is $\lambda = \Gamma\,\Delta t$. Thus, the Poisson distribution results when there are many "trials" (short time intervals), where there is a vanishingly small probability of "success" (an event occurrence) in each trial.

More mathematically, the Poisson process $N(t)$ is typically defined[1] for $t \geq 0$ in such a way that $N(t)$ is piecewise constant, taking on only constant, nonnegative integer values in intervals of the form $[t_j, t_{j+1})$, and with "jumps" of unit size occurring at the times t_j, so that $N(t)$ is also nondecreasing. Further, $N(t)$ must follow a Poisson distribution (with mean Γt in the homogeneous case, where Γ is the intensity). A more general process that relaxes the requirement of the Poisson distribution is called a **counting process**, and a **jump process** additionally relaxes the requirement of unit jumps (and therefore of monotonicity, if negative jumps are allowed).

11.3 Inhomogeneous Poisson Process and State Dependence

In general, the rate Γ of event occurence may depend on time, either explicitly or via dependence on the state $y(t)$ of the system. We can handle this by noting that according to our construction for the Poisson distribution above, then if X_1 and X_2 are Poisson-distributed random variables with means λ_1 and λ_2, then $X_1 + X_2$ is a Poisson random variable of mean $\lambda_1 + \lambda_2$. This statement amounts to agglomerating two adjacent time intervals of different duration in the above derivation of the Poisson distribution. Then if Γ is time-dependent, we can subdivide the interval $[t, t+\Delta t)$ into sufficiently fine increments such that Γ is constant over each increment, and sum them to find that the number of events occuring in the interval $[t, t+\Delta t)$ is still a Poisson variable with mean

$$\bar{\lambda} = \int_t^{t+\Delta t}\Gamma(t')\,dt'. \tag{11.24}$$

(inhomogeneous Poisson process: mean)

The variance is also of course just $\bar{\lambda}$.

However, while we can define the inhomogeneous Poisson process as above, a generalization to a process with state-dependent intensity $\Gamma(y)$, where $y(t)$ is some process driven by $dN(t)$, is *not* a Poisson process:

[1] see, e.g., Rama Cont and Peter Tankov, *Financial Modelling with Jump Processes* (Chapman & Hall/CRC, 2004), Section 2.5.3, p. 48.

the argument above for the inhomogeneous process does not apply, because $dN(t)$ is no longer statistically independent at different times.[2] Since in this case $N(t)$ is no longer Poisson-distributed, it is more proper to refer to it as a *counting process*, as we defined in Section 11.2.1.

11.4 White-Noise Limit

Consider the scaled Poisson process

$$dy = \frac{dN}{\sqrt{\Gamma}}, \tag{11.25}$$

where

$$\langle\langle dN \rangle\rangle = \Gamma\, dt. \tag{11.26}$$

It may be that in a given system, the rate Γ of Poisson events is much faster than the processes of physical interest. In such a case, we can ignore the discreteness of the events, and coarse-grain the dynamics to approximate the Poisson events by white noise. Note in particular that the mean of dy is

$$\langle\langle dy \rangle\rangle = \frac{\langle\langle dN \rangle\rangle}{\sqrt{\Gamma}} = \sqrt{\Gamma}\, dt, \tag{11.27}$$

while the variance is

$$\langle\langle (dy)^2 \rangle\rangle = \frac{\langle\langle dN^2 \rangle\rangle}{\Gamma} = \frac{\langle\langle dN \rangle\rangle}{\Gamma} = dt. \tag{11.28}$$

Thus, if events occur rapidly on time scales of interest—that is, we only measure Δy over time intervals $\Delta t \gg 1/\Gamma$, by the central limit theorem, we may effectively regard dy as a *Gaussian* random variable of mean $\sqrt{\Gamma}\, dt$ and variance dt. In particular, the Poisson process corresponds to a random walk of steps of length $\sqrt{\Gamma}$, in one direction only, taken at random times, as we can see by writing

$$y(t) = \frac{1}{\sqrt{\Gamma}} \int_0^t dN(t') = \frac{1}{\sqrt{\Gamma}} \int_0^t \frac{dN(t')}{dt'}\, dt', \tag{11.29}$$

and recalling from Eq. (11.19) that dN/dt is a sum of delta functions. After many events, the result is the same as a biased random walk, and the central limit theorem again guarantees an asymptotic Gaussian probability density. In this limit, it thus is a good approximation to write

$$dy = \sqrt{\Gamma}\, dt + dW. \tag{11.30}$$

Thus, in the limit where Poisson events occur at a very large rate Γ, we can make the formal replacement

$$dN \longrightarrow \Gamma\, dt + \sqrt{\Gamma}\, dW, \tag{11.31}$$

(white-noise limit of Poisson process)

to approximate the Poisson process with a mean drift plus white noise.

11.4.1 Shot Noise

As an example, let's consider shot noise of an electrical current. Let Q denote the total charge that has crossed a certain point along a wire. The current is given by

$$I = \frac{dQ}{dt}, \tag{11.32}$$

so that

$$dQ = I\, dt. \tag{11.33}$$

[2] Note, however, that this is *still* sometimes referred to as a "state-dependent Poisson process." See, e.g., Edoardo Daly and Amilcare Porporato, "Intertime jump statistics of state-dependent Poisson processes," *Physical Review E* **75**, 011119 (2007) (doi: 10.1103/PhysRevE.75.011119).

We will model the current as a stream of *independent* electrons of charge $-e$ with Poisson arrival times, so that

$$dQ = -e\,dN, \tag{11.34}$$

with $\langle\langle dN \rangle\rangle = \Gamma dt$ as usual. Then equating the two expression for dQ, we find

$$I\,dt = -e\,dN. \tag{11.35}$$

The mean current is then given by taking the ensemble average of this relation, so that

$$\langle\langle I \rangle\rangle = -e\Gamma. \tag{11.36}$$

Frequencies of interest for measuring electrical currents generally range from ns to s, whereas $\Gamma \sim 10^{19}$ s^{-1}. Thus, the white-noise approximation is quite appropriate, and thus

$$\begin{aligned} dQ &\approx -e\Gamma\,dt - e\sqrt{\Gamma}\,dW \\ &= \big\langle\!\big\langle I \big\rangle\!\big\rangle dt + \frac{\langle\langle I \rangle\rangle}{\sqrt{\Gamma}}\,dW \\ &= \big\langle\!\big\langle I \big\rangle\!\big\rangle dt - \sqrt{|e\langle\langle I \rangle\rangle|}\,dW. \end{aligned} \tag{11.37}$$

Thus, we see that due simply to the discreteness of charge, a mean current $\langle\langle I \rangle\rangle$ is accompanied by white noise of amplitude

$$\frac{|\langle\langle I \rangle\rangle|}{\sqrt{\Gamma}} = \sqrt{|e\langle\langle I \rangle\rangle|}. \tag{11.38}$$

Note that the SI units of the current noise amplitude are in $\mathrm{A}/\sqrt{\mathrm{Hz}}$, since when multiplied by dW/dt, which has dimensions $1/\sqrt{\mathrm{s}}$, the noise amplitude takes the dimensions of current. The alternate way to view this is that physically, the noise is always bandwidth-limited (as in the *RC* model of the Ornstein–Uhlenbeck process), and so the filtered noise amplitude is given by multiplying the above noise amplitude by the square root of the circuit bandwidth. More explicitly, the above white noise corresponds to a uniform spectral density of signal power. According to Eq. (3.89), an input signal $\alpha\xi(t)$ corresponds to rms fluctutations of $\alpha\sqrt{\gamma/2}$ at the output of a low-pass filter. Thus, the rms current fluctuation through a low-pass filter due to shot noise is given by

$$\delta I_{\mathrm{rms}} = \sqrt{|e\langle\langle I \rangle\rangle|}\sqrt{\frac{\gamma}{2}}, \tag{11.39}$$

where $\gamma = 1/RC$ is the angular cutoff frequency for the low-pass filter. The **equivalent-power bandwidth** $\Delta\nu$ is defined as the bandwidth of the ''brick wall'' filter (with flat response up to a sudden cutoff at frequency $\Delta\nu$, where $\Delta\nu$ is a frequency in Hz, *not* an angular frequency). The low-pass filter function for power transmission is (up to a constant factor, set by requiring the dc transmission to unity) the Fourier transform of the Ornstein–Uhlenbeck correlation function $e^{-\gamma\tau}$, so that we may write the transmission function as

$$T(\omega) = \frac{\gamma^2}{\gamma^2 + \omega^2}. \tag{11.40}$$

Obviously, γ is the (angular) ''3 dB'' frequency, or the frequency where the transmission drops to 1/2 the dc value:

$$f_{3\,\mathrm{dB}} = \frac{\gamma}{2\pi}. \tag{11.41}$$

Since

$$\int_0^\infty T(\omega)\,d\omega = \frac{\pi}{2}\gamma = 2\pi\Delta\nu, \tag{11.42}$$

where the second result applies to the brick-wall filter, we can write

$$\Delta\nu = \frac{\gamma}{4} = \frac{\pi}{2} f_{3\,\mathrm{dB}}, \tag{11.43}$$

and thus the shot-noise magnitude is

$$\delta I_{\rm rms} = \sqrt{2\Delta\nu |e\langle\langle I\rangle\rangle|}. \tag{11.44}$$
(shot-noise amplitude)

This expression applies to filters beyond the low-pass filter, so long as the appropriate equivalent-power bandwidth is used in this relation. Thus, a 1 A average current detected in a 1 MHz bandwidth has an rms noise current of 0.57 μA, a fluctuation at under the ppm level. Shot noise clearly gets much worse for smaller currents: for the same bandwidth, an average current of 1 μA has fluctuations of 0.57 nA rms, or 0.057% relative noise, and an average current of 1 pA has fluctuations of 0.57 pA rms, or 57% relative noise. Note that this model assumes the independence of electrons, and gives an appropriate result, e.g., for semiconductor junctions, but not in metallic-wire circuits, where long-range correlations between electrons suppress shot noise.[3] Essentially, this is just because of Coulomb interactions between electrons, which causes them to antibunch: in a conducting, crystalline lattice, it is energetically favorable to have two electrons in *different* lattice sites, as compared to having them occupy the *same* lattice site. Probabilities of seeing more than one electron pass in a metallic wire in a short time interval are thus suppressed compared to the Poissonian expectation.

Of course, we can adapt this result to the case of optical shot noise. Instead of an electrical current, we have a detected optical power P. We can treat the photon arrival times as independent in the case of coherent light (recall that in a coherent state the photon-number occupation probabilities are Poisson-distributed). The rms fluctuations for a mean optical power $\langle\langle P\rangle\rangle$ detected in an equivalent-power bandwidth $\Delta\nu$ are then given by

$$\delta P_{\rm rms} = \sqrt{2\Delta\nu\,\hbar\omega\langle\langle P\rangle\rangle|}, \tag{11.45}$$
(optical shot-noise amplitude)

where the photon energy $\hbar\omega$ plays the role of the electron charge. For 780 nm light in a 1 MHz detection bandwidth, a 1 W power has fluctutations of 0.71 μW, less than the ppm level. For the same bandwidth, a 1 μW mean power has fluctuations of 0.71 nW (0.071%), and a 1 pW mean power has fluctuations of 0.71 pW (71%). Of course, the relative fluctuations will be even larger for thermal (bunched) light, and smaller for antibunched light (nonclassical light, as for resonance fluorenscence of a two-level atom).

11.5 Compound Poisson Process

Looking ahead towards our study of Lévy processes, it will be useful to generalize the Poisson process to form a more general class of random walks. First, a simple generalization is to allow jumps of a size other than unity. Calling the jump size δ, we can define the δ-jump Poisson process as

$$N_\delta(t) = N(t)\,\delta \tag{11.46}$$

in terms of the standard Poisson process $N(t)$. Recall that the intensity Γ is an implicit part of the definition here. Still, though, this is just a rescaled version of the regular Poisson process. A manifestly more general process is the **compound Poisson process**, which is simply a sum over different Poisson processes of different jump sizes:

$$N_{\rm C}(t) = \sum_j N_{\delta_j}(t)\,\delta_j. \tag{11.47}$$

The jump sizes δ_j could be positive or negative, so this is not necessarily monotonic in the sense of the standard Poisson process. Also, the constituent Poisson processes could have different intensities:

$$\left\langle\left\langle dN_{\delta_j}(t)\right\rangle\right\rangle = \Gamma_j\,dt. \tag{11.48}$$

Yet more generally, a compound Poisson process could be a sum over a continuum of Poisson processes,

$$N_{\rm C}(t) = \int d\delta\, N_\delta(t). \tag{11.49}$$
(compound Poisson process)

[3] Paul Horowitz and Winfield Hill, *The Art of Electronics*, 2nd ed. (Cambridge, 1989), pp. 431-2.

To check that this is sensible, the compound jump intensity is

$$\langle\!\langle dN_{\text{C}}(t) \rangle\!\rangle = \int d\delta \, \langle\!\langle dN_\delta(t) \rangle\!\rangle = dt \int d\delta \, \Gamma(\delta). \tag{11.50}$$

That is, the intensity distribution $\Gamma(\delta)$ here represents a density—the vast majority of the continuum of Poisson processes should not have activated at any given time, otherwise the compound process would diverge. As an alternative way to view the compound process, we could define the total intensity

$$\Gamma := \int d\delta \, \Gamma(\delta) \tag{11.51}$$

as the rate at which jumps occur in the compound process, and then $\Gamma(\delta)/\Gamma$ acts as a probability density for the size of any particular jump.

11.5.1 Characteristic Function

It will likewise also be useful to have the characteristic function in hand for the compound Poisson process. Starting with the standard Poisson process, the exponentiated form of the increment is (Problem 11.1)

$$e^{dN} = 1 + (e-1)\, dN. \tag{11.52}$$

The two terms here correspond to the values of e^{dN} when $dN = 0$ or 1, respectively. The expectation value of this expression is

$$\left\langle\!\!\left\langle e^{dN} \right\rangle\!\!\right\rangle = 1 + (e-1)\, \Gamma \, dt. \tag{11.53}$$

To compute a characteristic function, we should instead compute the same exponentiated dN, but including a factor of ik,

$$\left\langle\!\!\left\langle e^{ik\, dN} \right\rangle\!\!\right\rangle = 1 + (e^{ik} - 1)\, \Gamma \, dt = \exp\left[\Gamma \, dt \left(e^{ik} - 1\right)\right], \tag{11.54}$$

where the last expressions are equivalent to order dt. Now since the Poisson process $N(t)$ is a sum over the increments $dN(t)$, the characteristic function is simply the product over the characteristic functions of the increments:

$$\left\langle\!\!\left\langle e^{ik\, N(t)} \right\rangle\!\!\right\rangle = \exp\left[\Gamma t \left(e^{ik} - 1\right)\right]$$

(Poisson process, characteristic function) (11.55)

Generalizing this result to the δ-step process (11.46) amounts to rescaling k:

$$\left\langle\!\!\left\langle e^{ik\, N_\delta(t)} \right\rangle\!\!\right\rangle = \exp\left[\Gamma t \left(e^{ik\delta} - 1\right)\right]. \tag{11.56}$$

For the compound Poisson process, we can again take advantage of the observation that it is the sum over δ-step processes. The characteristic function is thus a product over the individual functions, which means we should integrate over the possible δ inside the exponent:

$$\left\langle\!\!\left\langle e^{ik\, N_{\text{C}}(t)} \right\rangle\!\!\right\rangle = \exp\left[t \int d\delta \, \Gamma(\delta) \left(e^{ik\delta} - 1\right)\right] = \exp\left[\Gamma t \int d\delta \, f(\delta) \left(e^{ik\delta} - 1\right)\right].$$

(compound Poisson process, characteristic function) (11.57)

In the last expression, we set the jump density $f_\delta(\delta) := \Gamma(\delta)/\Gamma$ and wrote the characteristic function in terms of the total jump rate Γ.

11.5.2 Compensated Poisson Process

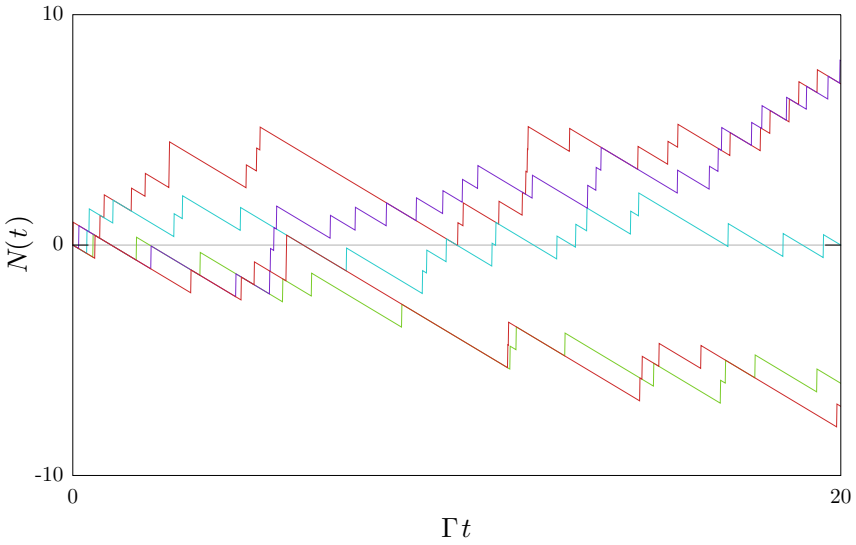
The ordinary Poisson process (with unit jumps, rate Γ) has a mean of

$$\langle\!\langle N(t) \rangle\!\rangle = \Gamma t. \tag{11.58}$$

Often, it is useful to transform into the "rest frame" of this process by defining a **compensated Poisson process** via

$$M(t) := N(t) - \Gamma t, \tag{11.59}$$
(compensated Poisson process)

such that $\langle\langle M(t) \rangle\rangle = 0$. As an illustration, 5 trajectories of a compound Poisson process with $\Gamma = 1$ are shown below.



The same idea applies to the δ-step process (where the compensation term is $\delta\Gamma t$), or to the compound process, where the compensated version is

$$M_{\rm C}(t) := N_{\rm C}(t) - \langle\delta\rangle\,\Gamma t, \tag{11.60}$$
(compensated Poisson process)

where Γ is defined by Eq. (11.51), and

$$\langle\delta\rangle := \int d\delta\,\delta\,f(\delta). \tag{11.61}$$

The characteristic function has the same form as in Eq. (11.57), but with the mean set to zero:

$$\left\langle\!\!\left\langle e^{ik\,M_{\rm C}(t)} \right\rangle\!\!\right\rangle = \exp\left[\Gamma t \int d\delta\, f(\delta)\left(e^{ik\delta} - 1 - ik\delta\right)\right].$$
(compensated compound Poisson process, characteristic function) (11.62)

A compound Poisson process may of course already have a zero mean, by virtue of the vanishing of $\langle\delta\rangle$, in which case the drift term in Eq. (11.60) has no effect. However, a different "compensation" implemented by shifting away the mean of $f(\delta)$ is of course not the same as superimposing the drift as in Eq. (11.60).

11.6 Exercises

Problem 11.1

By formally resumming the Taylor series expansion, compute e^{dN}.

Problem 11.2

Recall the Poisson distribution in terms of the single parameter λ has the form

$$P(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \tag{11.63}$$

where n is a nonnegative integer. Also, suppose that N is a Poisson-distributed random variable. By direct calculation from the probability distribution:

(a) Show that $P(n)$ is normalized.

(b) Show that $\langle N \rangle = \lambda$.

(c) Show that $\mathrm{Var}[N] = \lambda$.

Chapter 12

Lévy Processes

The notion of a Lévy process is a quite wide-ranging one, and it encompasses and generalizes both the Wiener and Poisson processes that we have studied before. An important application of Lévy processes is towards modeling **heavy-tailed statistics**, or stochastic processes with large jumps (''extreme events'') that can't otherwise be modeled with only a Gaussian diffusion. In some sense, as we will see, in general Lévy processes act something like combinations of Wiener and Poisson processes.

12.1 Definition

To more specifically pin down the mathematical idea, a **Lévy process** is a continuous-time stochastic process $L(t)$ satisfying the following properties:[1]

1. Lévy processes have **independent increments**, meaning that $L(t) - L(t_0)$ is independent of $L(t') - L(t'_0)$, provided the time intervals $[t_0, t)$ and $[t'_0, t')$ do not overlap. This, for example, rules out state-dependent stochastic processes, such as the Ornstein–Uhlenbeck process.

2. Lévy processes have **stationary increments**, meaning that the probability distribution for the increment $\Delta L(t) := L(t + \Delta t) - L(t)$ over the same fixed time Δt is independent of t. This, for example, rules out time dependence like $dy = \beta(t)\, dW$ for $\beta'(t) \neq 0$.

3. Lévy processes have **stochastic continuity**, meaning that for any $\varepsilon > 0$, the probability that $|\Delta L(t)| = |L(t + \Delta t) - L(t)| > \varepsilon$ converges to zero as $\Delta t \longrightarrow 0$. Note, however, that this does not imply strict continuity: finite jump discontinuities in $L(t)$ are permitted, so long as the *probability* of the jump happening goes to zero in a vanishingly short time interval. In particular, jumps at fixed (nonrandom) times are not permissible.

4. A sort of technical requirement is that Lévy processes are **cadlag**, which means that at every t, $L(t)$ is equal to its right limit $L(t + 0^+)$. This is automatically true for a continuous section of $L(t)$, but it means that at the time t_0 of a discontinuity, the value of $L(t_0)$ corresponds to the *future* value just *after* the jump. When necessary, though, for clarity it is useful to explicitly distinguish the states just before and just after the jump. Although we haven't mentioned this requirement before, the cadlag property also applies to Poisson and compound Poisson processe.

5. Lévy processes have **infinite divisibility**, which means that the probability density for $L(t)$ is such that $L(t)$ can be represented as a sum of N independent, identically distributed random variables, for any integer $N \geq 2$. Intuitively, these random variables correspond to the Lévy increments $\Delta L(t) = L(t + \Delta t) - L(t)$, if we take $\Delta t = t/N$. Thus, infinite divisibility of Lévy processes already follows form the first and second properties above. However, it also follows that the characteristic function

[1] Rama Cont and Peter Tankov, *Financial Modelling with Jump Processes* (Chapman & Hall/CRC, 2004), Section 3.1, p. 68.

(Section 1.4.1) has a certain form. Since the characteristic function of $L(t)$ is the product of the characteristic functions of the N increments. Letting $\tilde{f}(k;t)$ denote the characteristic function for $L(t)$ [i.e., the Fourier transform of the corresponding density $f(x)$],

$$\tilde{f}(k;t) := \left\langle\!\left\langle e^{ikL(t)} \right\rangle\!\right\rangle, \tag{12.1}$$

we can write this in terms of the increment characteristics as

$$\tilde{f}(k;t) = \prod_{j=0}^{N-1} \tilde{f}(k;\Delta t_j) = \tilde{f}^N(k;\Delta t), \tag{12.2}$$

where we used the independence and stationarity of Lévy processes. This means that the log of the characteristic function satisfies

$$\log \tilde{f}(k;t) = N \log \tilde{f}(k;\Delta t), \tag{12.3}$$

which in turn means that the log of the characteristic function is proportional to time. Thus, a Lévy characteristic function must have the from

$$\tilde{f}(k;t) = e^{t\varphi(k)} \tag{12.4}$$

for a function $\varphi(k)$. In fact, recall that $t\varphi(k)$ is essentially just the cumulant-generating function for $L(t)$. The concept of infinite divisibility obviously also applies to a probability distribution as well, not just a stochastic process. For such a distribution, one should be able to find an nth root of the characteristic function and still obtain a sensible characteristic function. The Gaussian distribution is infinitely divisible, as is the entire family of α-stable distributions that we will introduce below. The uniform distribution is not infinitely divisible—the problem relates to the negative values in the characteristic function.

That the Wiener process satisfies the above properties should be reasonably obvious by this point. The characteristic function

$$\tilde{f}_W(k;t) = e^{-tk^2/2} \tag{12.5}$$

has the expected infinitely divisible form. That the standard Poisson process and compound Poisson process also satisfies these properties is reasonably obvious; the stochastic continuity follows from the jump probability $\Gamma\, dt$ in time t. (Of course, the inhomogeneous Poisson process is not a Lévy process.) The characteristic function (11.57) for the compound Poisson process also has the expected divisible form.

12.1.1 Stable Processes

Suppose that we generalize the characteristic function of the Wiener process (12.5) as

$$\tilde{f}_W(k;t) = e^{-t\sigma^2 k^2}, \tag{12.6}$$

where the conventional choice is $\sigma = 1/\sqrt{2}$. Since this characteristic function corresponds to the probability density

$$f_W(x;t) = \frac{1}{\sqrt{4\pi\sigma^2 t}}\, e^{-x^2/4\sigma^2 t}, \tag{12.7}$$

it is clear that σ is a width-scaling parameter, although it is *not* the standard deviation of this scaled $W(t)$; rather, the standard deviation is $\sigma\sqrt{2t}$. Generalizing the characteristic function by modifying the power in the exponent to

$$\tilde{f}_\alpha(k;t) = \left\langle\!\left\langle e^{ikS_\alpha(t)} \right\rangle\!\right\rangle = e^{-t\sigma^\alpha |k|^\alpha}.$$

(characteristic function, α-stable Lévy process) (12.8)

This two-parameter family of stochastic processes $S_\alpha(t)$, defined by this family of characteristic functions, goes by the name of the ***α*-stable Lévy processes**, or also the the **Paretian processes**. The corresponding probability density can be readily expressed in terms of the inverse Fourier transform as

$$f_\alpha(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\, e^{-t\sigma^\alpha |k|^\alpha} = \frac{1}{\pi}\int_0^\infty dk\, \cos(kx)\, e^{-t\sigma^\alpha k^\alpha},$$

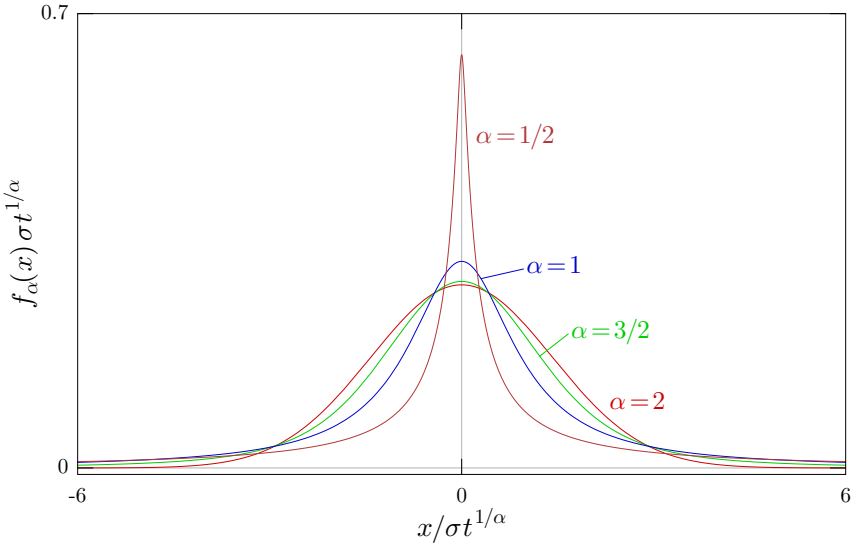
(probability density, α-stable Lévy process) (12.9)

but this integral has simple closed forms only in certain special cases, as we will discuss below.[2] Nevertheless, it has been shown that the stable Lévy distribution functions are bell-shaped, in the sense that the nth derivative changes sign exactly n times (in particular, the curvature changes sign exactly twice).[3]

The α-stable processes make sense only for $0 < \alpha \leq 2$. The requirement $\alpha > 0$ is necessary for the defining integral (12.9) to make sense. The case $\alpha = 2$ is the Gaussian (Wiener) process that we have already studied in detail; that $\alpha \leq 2$ is basically a consequence of the central-limit theorem, and we will return to this point in Section 12.1.4. Also, as we will see in Section 12.1.4, the probability densities for the α-stable processes all have power-law tails (with the exception of the Gaussian process $\alpha = 2$). Since power-law distributions are called **Pareto distributions**, this explains the alternate name of Paretian processes for the α-stable Lévy processes.

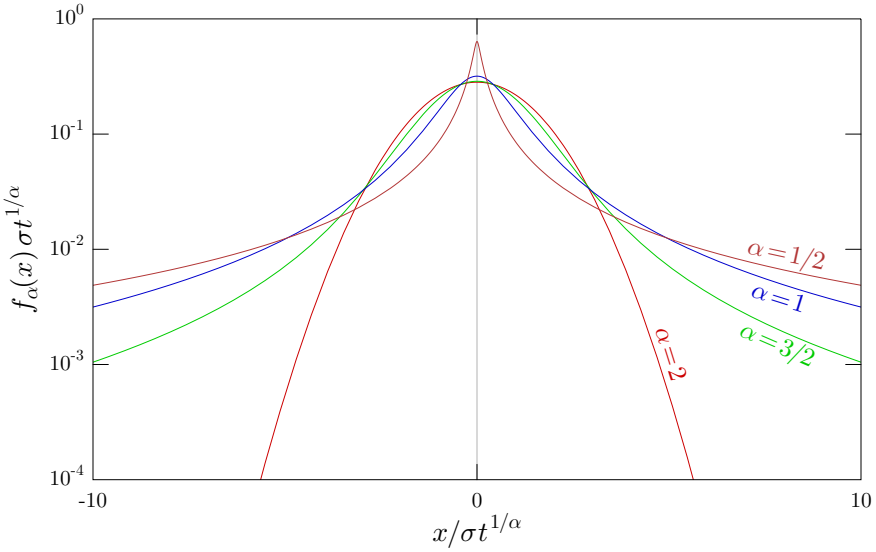
We will furthermore see that for any α-stable process other than $\alpha = 2$, the variance is undefined at any $t > 0$. For $\alpha \leq 1$, even the *mean* of the distribution is undefined (although for the symmetric distributions we are considering here, we can regard the distributions as having a zero mean in the sense of a Cauchy principal value for the mean integral). For this reason, the range $1 < \alpha < 2$ is probably the most reasonable in terms of modeling physical processes with heavy-tailed distributions.

To illustrate some of these Paretian distributions, a few of them are plotted below. The Gaussian ($\alpha = 2$) case stands out in particular with the logarithmic vertical scale, since it lacks the power-law tails of the other distributions.

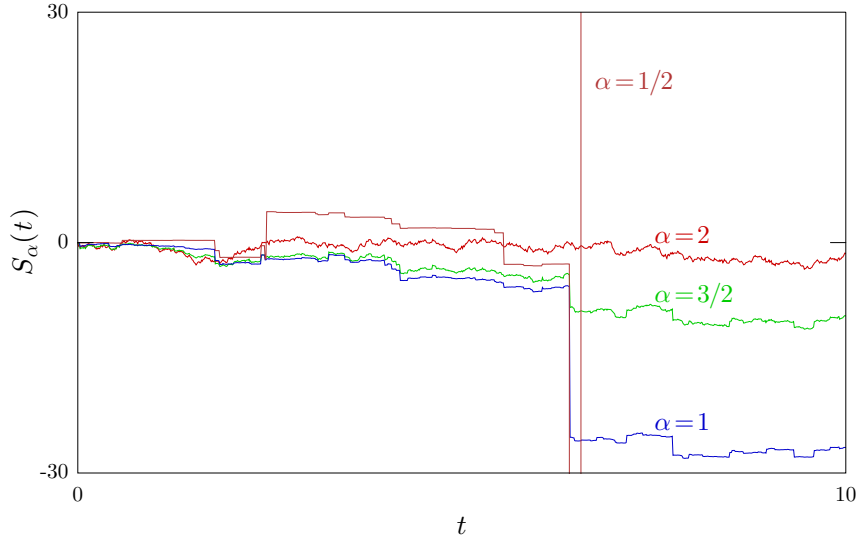


[2]Note that the α-stable Lévy distributions can be written in terms of the Fox H-function; see W. R. Schneider, "Stable distributions: Fox function representation and generalization," in *Stochastic Processes in Classical and Quantum Systems*, S. Albeverio, G. Casati, and D. Merlini, Eds. (Springer, 1985), p. 497 (doi: 10.1007/3540171665_92). The distributions can also be expressed in ters of Meijer G-functions; see V. M. Zolotarev, "On Representation of Densities of Stable Laws by Special Functions," *Theory of Probability and Its Applications* **39** 354 (1994) (doi: 10.1137/1139025). Though it's interesting to have such expressions, the special functions are such a general objects that it isn't obvious that they are a big improvement over the Fourier integral (12.9).

[3]Mateusz Kwaśnicki, "A new class of bell-shaped functions," *Transactions of the American Mathematical Society* **373**, 2255 (2020) (doi: 10.1090/tran/7825). This result also applies to the asymmetric stable probability density functions (with $\beta \neq 0$) in Section 12.3.



Also, to illustrate the consequences of the tail, a simulated trajectory is plotted below for each of the distributions in the above plots. The trajectories in each case are comparable in that the same set of random numbers is transformed to make the trajectories. Jumps are apparent in each case except for the Gaussian, and the jumps tend to be larger for smaller α.



12.1.2 Width Scaling and Addition Rules

A common alternative notation for the α-stable Lévy characteristic function (12.8) and probability density (12.9) corresponds to setting $t = 1$ in the above expressions, giving a characteristic function

$$\tilde{f}_\alpha(k) = e^{-\sigma^\alpha |k|^\alpha},$$

(characteristic function, α-stable Lévy distribution) (12.10)

and a probability density

$$f_\alpha(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx}\, e^{-\sigma^\alpha |k|^\alpha} = \frac{1}{\pi}\int_0^{\infty} dk\, \cos(kx)\, e^{-\sigma^\alpha k^\alpha}.$$

(α-stable Lévy density) (12.11)

In this case, the distribution here does not necessarily refer to a time-dependent stochastic process, but rather just to a probability density for a random variable.

With this simplification we can focus for a bit on the interpretation of σ. Since x only enters the Fourier integrals in the combination kx, and σ scales k is the stretched-Gaussian factor, σ acts as a width for the distribution. We can see this more explicitly by changing $k \longrightarrow k/\sigma$ and $x \longrightarrow x\sigma$ in Eqs. (12.11) to write

$$f_\alpha(x) = \frac{1}{2\pi\sigma}\int_{-\infty}^{\infty} dk\, e^{-ik(x/\sigma)}\, e^{-|k|^\alpha} = \frac{1}{\pi\sigma}\int_0^\infty dk\, \cos[k(x/\sigma)]\, e^{-k^\alpha}. \tag{12.12}$$

In this form the x-dependence is of the form x/σ, so σ characterizes the width of the distribution (while the leading factor of σ^{-1} preserves the normalization of the variable-width distribution).

Now suppose X_1 and X_2 are two α-stable-distributed random variables with width parameters σ_1 and σ_2, respectively. Recalling that the characteristic function for $X_1 + X_2$ is the product of the individual characteristic functions, the sum also has an α-stable distribution with width parameter given by

$$\sigma = \left(\sigma_1^\alpha + \sigma_2^\alpha\right)^{1/\alpha}. \tag{12.13}$$

For the Gaussian case $\alpha = 2$, this reduces to the familiar addition in quadrature of the Gaussian widths.

Restoring the time dependence as in the expressions (12.8) and (12.9) for the α-stable processes, the main feature to notice is that time enters in the combination $t\sigma^\alpha$. Thus the same arguments that apply to σ also apply to $t^{1/\alpha}$. In particular, the width of the probability density scales with time as $t^{1/\alpha}$. Again, for the Gaussian ($\alpha = 2$) case, this corresponds to diffusive growth as $\sqrt{t}$. Any smaller value of α corresponds to growth with a larger power of t; this is called **anomalous diffusion** or **superdiffusion**. This also, for example, implies the time-rescaling relation

$$S_\alpha(t) = T^{1/\alpha} S_\alpha(t/T), \tag{12.14}$$

which relates a time rescaling to a width rescaling of the α-stable process—a longer evolution time results in a wider exploration of space, but with a power-law form that depends on α. This is an example of **scale invariance**. The scale invariance or self-similarity is sometimes described in terms of the **Hurst exponent** H, where in this case a time dilation by factor T is accompanied by a spatial dilation factor T^H. Thus, stable Lévy processes have a Hurst exponent $H = 1/2\alpha$.

12.1.3 Special Cases

For the symmetric α-stable distributions that we have considered so far, there are only a few cases with simple analytic forms, and we will briefly review these here. There are more known forms in the case of the asymmetric generalization to the stable distributions, which we will examine later in Section 12.3.

12.1.3.1 Gaussian Process ($\alpha = 2$)

As we already pointed out, beginning with Eq. (12.7), the case $\alpha = 2$ is a Gaussian distribution, with density function

$$f_2(x;t) = \frac{1}{\sqrt{4\pi\sigma^2 t}}\, e^{-x^2/4\sigma^2 t}, \tag{12.15}$$

(Gaussian stable density)

with standard deviation $\sqrt{2t}\sigma$. Thus, in the stable-process notation, $S_2(t) = \sqrt{2}\sigma W(t)$ in terms of the standard Wiener process $W(t)$.

12.1.3.2 Cauchy Process ($\alpha = 1$)

In the case $\alpha = 1$, Eq. (12.12) becomes an integral in terms of ordinary exponentials, and is readily solved, with solution

$$f_1(x) = \frac{\sigma t}{\pi[(\sigma t)^2 + x^2]}. \tag{12.16}$$

Then the distribution clearly has $|x|^{-2}$ power-law tails; furthermore, σt is the half-width at half maximum of the distribution. This is the **Cauchy distribution** with width parameter σt. We encountered this distribution before in Section 2.3.1 as an example of the failure of the central limit theorem.

12.1.3.3 A Heavier-Tailed Example ($\alpha = 2/3$)

As a somewhat more exotic example, it is possible to work out the case $\alpha = 2/3$; the result is[4]

$$f_{2/3}(x) = \frac{\sigma t^{3/2}}{\sqrt{12\pi}|x|} \exp\left(\frac{2\sigma^2 t^3}{27x^2}\right) W_{-1/2,1/6}\left(\frac{4\sigma^2 t^3}{27x^2}\right), \tag{12.17}$$

where the Whittaker function is defined by[5]

$$W_{k,m}(z) := \frac{e^{-z/2}\, z^k}{\Gamma(1/2 - k + m)} \int_0^\infty t^{-k-1/2+m} \left(1 + \frac{t}{z}\right)^{k-1/2+m} e^{-t}\, dt. \tag{12.18}$$

Since $W_{-1/2,1/6}(x) \approx \Gamma(1/3)\, x^{1/3}/\Gamma(7/6)$ for small x, the tails of the distribution scale as $|x|^{-5/3}$.

12.1.4 Power-Law Tails

Now to demonstrate explicitly the power-law tails in the general case of the α-stable distributions (12.9), let's begin with the Fourier integral

$$f_\alpha(x;t) = \frac{1}{\pi} \int_0^\infty dk\, \cos(kx)\, e^{-t\sigma^\alpha k^\alpha}. \tag{12.19}$$

Since we want the asymptotic, large-x behavior, this is encoded in the small-k region of the characteristic function. To proceed with an asymptotic analysis, we can simply use a small-k expansion for the characteristic function:

$$f_\alpha(x;t) = \frac{1}{\pi} \int_0^\infty dk\, \cos(kx) \sum_{n=0}^\infty \frac{1}{n!}\Big(-t\sigma^\alpha k^\alpha\Big)^n. \tag{12.20}$$

Interchanging the sum and integral here is problematic, because the resulting termwise integrals diverge. Nevertheless, since we are interested in the large-x behavior of the distribution, and the problem here is a divergence at large-k (corresponding to high spatial frequencies and thus fine structure of the distribution), we can "cure" the problematic behavior by inserting a convergence factor of $e^{-\lambda k}$, carrying out the integral, and letting $\lambda \longrightarrow 0$ at the end of the calculation. Note that the $n = 0$ term simply leads to a delta function at $x = 0$, which does not contribute to the asymptotic behavior; we have thus

$$f_\alpha(x;t) \sim \frac{1}{\pi} \sum_{n=1}^\infty \frac{(-t)^n \sigma^{n\alpha}}{n!} \lim_{\lambda\to 0} \int_0^\infty dk\, \cos(kx)\, k^{n\alpha}\, e^{-\lambda k}. \tag{12.21}$$

Evaluating the integral and taking the limit, we have[6]

$$f_\alpha(x;t) \sim -\frac{1}{\pi} \sum_{n=1}^\infty \frac{(-1)^n \Gamma(n\alpha+1)\sin(n\pi\alpha/2)}{n!} (\sigma t^{1/\alpha})^{n\alpha}\, |x|^{-(n\alpha+1)}. \tag{12.22}$$

The leading contribution at large $|x|$ comes from the $n = 1$ term. Dropping the remaining terms, the asymptotic expression becomes

$$f_\alpha(x;t) \sim \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi} \big(\sigma t^{1/\alpha}\big)^\alpha\, |x|^{-(\alpha+1)} \qquad \Big(|x| \gg \sigma t^{1/\alpha}\Big), \tag{12.23}$$

(asymptotic tails, α-stable process)

[4] W. R. Schneider, *op. cit.*, Eq. (3.3).

[5] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th reprinted ed. (Cambridge, 1935), p. 340.

[6] For an error-bounded treatment of the asymptotics of integrals more general than (12.19), see F. W. J. Olver, "Error Bounds for Stationary Phase Approximations," *SIAM Journal on Mathematical Analysis* **5**, 19 (doi: 10.1137/0505003).

Fixing $x > 0$, the asymptotic probability to be in the tail beyond x is

$$\begin{aligned} P[S_\alpha(t) > x] &\sim \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\left(\sigma t^{1/\alpha}\right)^\alpha \int_x^\infty dx'\, x'^{-(\alpha+1)} \\ &= \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\alpha\pi}\left(\sigma t^{1/\alpha}\right)^\alpha x^{-\alpha}. \end{aligned} \tag{12.24}$$

Then using $\alpha\Gamma(\alpha) = \Gamma(\alpha+1)$, the resulting probability is

$$P[S_\alpha(t) > x] \sim \frac{\Gamma(\alpha)\sin(\pi\alpha/2)}{\pi}\left(\frac{x}{\sigma t^{1/\alpha}}\right)^{-\alpha}.$$

(asymptotic probability, α-stable process) (12.25)

Clearly, we must require $\alpha > 0$ for this probability to decay to zero as $x \longrightarrow \infty$.

At this point, we can see explicitly that the variance integral, which has tails of the form $|x|^{-(\alpha-1)}$, diverges for $\alpha < 2$ (i.e., for anything non-Gaussian), since a power-law tail of an integral must decay *faster* than $1/x$ to converge. Note that this argument doesn't apply to the case of $\alpha = 2$, which is Gaussian and thus lacks power-law tails. Additionally, the mean integral, which has a tail of the form $|x|^{-\alpha}$, diverges for $\alpha \leq 1$. Now we can also see explicitly that $\alpha \leq 2$: Remember that according to the central-limit theorem, the only stable distribution with finite variance is the Gaussian distribution. Beyond this, a larger value of the index α implies a more rapid power-law decay in the probability-density tails.

12.1.5 Series Representations

Revisiting the series (12.22) that we derived in, let's analyze its convergence. Recall the ratio test, which says that if a series has terms a_n, then we can consider the magnitude of the ratio a_{n+1}/a_n; if this magnitude converges to anything less than unity, then the series converges, if the ratio converges to something more than unity (or diverges), then the series diverges, and if the magnitude converges to unity then the test is inconclusive. Thus we can consider the ratio

$$\left|\frac{a_{n+1}}{a_n}\right| = \frac{(n+1)\Gamma[(n+1)\alpha+1]\sin[(n+1)\pi\alpha/2]}{\Gamma(n\alpha+1)\sin(n\pi\alpha/2)}\,(\sigma t^{1/\alpha})^\alpha\,|x|^{-\alpha}. \tag{12.26}$$

If we focus just on the scaling with n, then we can consider

$$\left|\frac{a_{n+1}}{a_n}\right| \sim \frac{\Gamma[(n+1)\alpha+1]}{(n+1)\Gamma(n\alpha+1)}. \tag{12.27}$$

Note that in the case of $\alpha = 1$, the n-dependent factor here is just unity, because $(n+1)\Gamma(n\alpha+1) = \Gamma[(n+1)\alpha+1]$. If α is any smaller, then the factor of $(n+1)$ wins, and the series converges. Thus, we can write the series (12.22) as the series representation

$$f_\alpha(x;t) = \frac{1}{\pi}\sum_{n=1}^\infty \frac{(-1)^{n-1}\Gamma(n\alpha+1)\sin(n\pi\alpha/2)}{n!}\,(\sigma t^{1/\alpha})^{n\alpha}\,|x|^{-(n\alpha+1)},$$

(density series representation, $0 < \alpha < 1$) (12.28)

which converges for $0 < \alpha < 1$. The series is also convergent in the borderline case, provided $|x| > \sigma t^{1/\alpha}$.

In the case $1 < \alpha < 2$, or $\alpha = 1$ with $|x| < \sigma t^{1/\alpha}$, the series (12.28) still functions as an *asymptotic* series. If $(\sigma t^{1/\alpha})^\alpha\,|x|^{-\alpha} \ll 1$, then it may take a number of terms in the series for the factor $\Gamma(n\alpha+1)/n!$ to grow enough for the nth term in the series to become of order unity. Sticking to terms below this point, the first terms in the series can still give a good approximation for the true density (in this case, out in the tails of the distribution).

An alternate series representation arises if we return to the Fourier integral (12.19),

$$f_\alpha(x;t) = \frac{1}{\pi}\int_0^\infty dk\,\cos(kx)\,e^{-t\sigma^\alpha k^\alpha}, \tag{12.29}$$

and work with the series expansion of the cosine:

$$\begin{aligned} f_\alpha(x;t) &= \frac{1}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{(2n)!}\int_0^\infty dk\, k^{2n}\, e^{-t\sigma^\alpha k^\alpha} \\ &= \frac{1}{\pi\alpha}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}\frac{x^{2n}}{(t^{1/\alpha}\sigma)^{2n+1}}\,\Gamma\!\left(\frac{2n+1}{\alpha}\right). \end{aligned} \tag{12.30}$$

Now we can apply the ratio test again,

$$\left|\frac{a_{n+1}}{a_n}\right| = \frac{1}{(2n+2)(2n+1)}\frac{\Gamma[(2n+3)/\alpha]}{\Gamma[(2n+1)/\alpha]}\frac{x^2}{(t^{1/\alpha}\sigma)^2}, \tag{12.31}$$

where we see that the n-dependent factor is exactly unity in the case $\alpha = 1$. Convergence happens then for $1 < \alpha \le 2$, or in the borderline case $\alpha = 1$ provided $|x| < \sigma t^{1/\alpha}$. Thus, we have the alternate series representation

$$f_\alpha(x;t) = \frac{1}{\pi\alpha}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}\,\Gamma\!\left(\frac{2n+1}{\alpha}\right)\frac{x^{2n}}{(t^{1/\alpha}\sigma)^{2n+1}}.$$

(density series representation, $1 < \alpha \le 2$) (12.32)

Again, the borderline case $\alpha = 1$ is divided between Eqs. (12.28) and (12.32). Also, again for $0 < \alpha < 1$, although this series diverges, it serves as an asymptotic expansion in the range $|x| \ll t^{1/\alpha}\sigma$.

Note that it is fairly simple to verify the convergence of the series (12.32) directly in the case $\alpha = 2$,

$$\begin{aligned} f_{\alpha=2}(x;t) &= \frac{1}{2\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}\,\Gamma(n+{}^1\!/_2)\frac{x^{2n}}{(t^{1/2}\sigma)^{2n+1}} \\ &= \frac{1}{2\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}\frac{(2n-1)!!}{2^n}\frac{x^{2n}}{(t^{1/2}\sigma)^{2n+1}} \\ &= \frac{1}{2\sqrt{\pi}t^{1/2}\sigma}\sum_{n=0}^{\infty}\frac{(2n-1)!!}{(2n)!}\left(-\frac{x^2}{2t\sigma^2}\right)^n \\ &= \frac{1}{2\sqrt{\pi}t^{1/2}\sigma}\sum_{n=0}^{\infty}\frac{1}{2^n n!}\left(-\frac{x^2}{2t\sigma^2}\right)^n \\ &= \frac{1}{2\sqrt{\pi}t^{1/2}\sigma}\exp\!\left(-\frac{x^2}{4t\sigma^2}\right)^n, \end{aligned} \tag{12.33}$$

yielding the Gaussian distribution, where we used the gamma-function identity

$$\Gamma(n/2) = \frac{\sqrt{\pi}(n-2)!!}{2^{(n-1)/2}}, \tag{12.34}$$

with $n!! = n(n-2)(n-4)\cdots 1$. Likewise, in the borderline case $\alpha = 1$,

$$f_{\alpha=1}(x;t) = \frac{1}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}\,\Gamma(2n+1)\frac{x^{2n}}{(t\sigma)^{2n+1}} = \frac{1}{\pi t\sigma}\sum_{n=0}^{\infty}\left(-\frac{x^2}{(t\sigma)^2}\right)^n = \frac{1}{\pi t\sigma[1+(x/t\sigma)^2]}, \tag{12.35}$$

yielding the correct Cauchy density.

12.2 Generalized Central Limit Theorem

One of the major reasons that the stable distributions are so important is that they act as limiting distributions for random walks in cases where the usual central limit theorem fails. We will demonstrate this in a

fairly informal way that parallels the proof of the central limit theorem in Section 2.2.[7] In particular, suppose we have independent, identically distributed random variables $X_1, \ldots, X_N$, with a probability density function having power-law tails,

$$f(x) \sim \frac{c}{|x|^{\alpha+1}}, \qquad |x| \longrightarrow \infty,\ 0 < \alpha < 2. \tag{12.36}$$

Note that the variance of such a distribution diverges, and hence the central limit theorem fails. We are also considering only symmetric tails in the density for the moment; we will return to the asymmetric case later.

In analogy to the central limit theorem, consider the scaled sum statistic

$$S_N := \frac{1}{N^{1/\alpha}} \sum_{j=1}^{N} X_j. \tag{12.37}$$

Our goal is to compute the characteristic function for this statistic. But first, we require the characteristic function for the individual steps,

$$\tilde{f}(k) = \int_0^\infty dx\, f(x)\, \cos(kx). \tag{12.38}$$

We know only the asymptotic behavior in the tails of $f(x)$. Thus, let's split the integral at some point $d > 0$,

$$\begin{aligned} \tilde{f}(k) &= 2\int_0^d dx\, f(x)\cos(kx) + \int_d^\infty dx\, f(x)\cos(kx) \\ &\approx 2\int_0^d dx\, f(x)\cos(kx) + \int_d^\infty dx\, \frac{c}{x^{\alpha+1}}\cos(kx), \end{aligned} \tag{12.39}$$

where we assume d is sufficiently large that the approximation in terms of the asymptotic power law is good for $|x| > d$ (note that we can make it as accurate as we like by increasing d). Rescaling the second integral by $x \longrightarrow x/|k|$, we have

$$\tilde{f}(k) \approx 2\int_0^d dx\, f(x)\cos(kx) + 2c|k|^\alpha \int_{|k|d}^\infty \frac{dx}{x^{\alpha+1}}\cos x, \tag{12.40}$$

and then the second integral can be evaluated as[8]

$$\tilde{f}(k) \approx \tilde{f}_{\text{C}}(k) + c|k|^\alpha \Big(i^\alpha \Gamma(-\alpha, i|k|d) + \text{c.c.} \Big) \tag{12.41}$$

in terms of the incomplete gamma function

$$\Gamma(a, x) := \int_x^\infty t^{a-1} e^{-t}\, dt. \tag{12.42}$$

We also named the first integral as a characteristic function $\tilde{f}_{\text{C}}(k)$ for the original distribution truncated beyond $|x| > d$.

Now using the expansion

$$\Gamma(-\alpha, ix) = \Gamma(-\alpha) + x^{-\alpha}\left(\frac{(i\,\text{sgn}x)^{-\alpha}}{\alpha} + i\frac{(i\,\text{sgn}x)^{-\alpha}x}{1-\alpha} + \frac{(i\,\text{sgn}x)^{-\alpha}x^2}{2(2-\alpha)} + O(x^3) \right), \tag{12.43}$$

[7] For the formal treatment, see B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, English translation by K. L. Chung and J. L. Doob (Addison-Wesley, 1954).

[8] The integral is reasonably obvious in terms of the incomplete gamma function, but it is also listed for example in I. S. Gradstein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, English translation 7th ed., A. Jeffrey and D. Zwillinger, Eds. (Academic Press, 2007), p. 436, Formula 3.761.7.

valid for $x \in \mathbb{R}$ and $0 < \alpha < 2$ (except $\alpha = 1$, which will require more careful attention later), we can write

$$\begin{aligned}\tilde{f}(k) &\approx \eta + ik\mu + O(k^2) + 2c|k|^\alpha \cos(\pi\alpha/2)\Gamma(-\alpha) + \frac{2c}{\alpha d^\alpha} + O(k^2) \\ &= 1 + ik\mu - \sigma^\alpha |k|^\alpha + O(k^2)\end{aligned} \tag{12.44}$$

where we also expanded the first distribution $\tilde{f}_{\rm C}(k)$—this characteristic function corresponds to a distribution with finite mean and variance, and so has a more normal series expansion in k. Importantly, the k^α term here is independent of the cutoff d of the integral. In the same term we have also defined σ by

$$\sigma^\alpha := -2c\cos(\pi\alpha/2)\Gamma(-\alpha) = \frac{2\pi c\cos(\pi\alpha/2)}{\alpha\sin(\pi\alpha)\Gamma(\alpha)} = \frac{\pi c}{\sin(\pi\alpha/2)\Gamma(\alpha+1)}. \tag{12.45}$$

where we used the gamma-function identity

$$\Gamma(x)\Gamma(-x) = -\frac{\pi}{x\sin(\pi x)}, \tag{12.46}$$

the double-angle formula $\sin(2\theta) = 2\sin\theta\cos\theta$, and the identity $z\Gamma(z) = \Gamma(z+1)$. In the last form, it is clear that $\sigma^\alpha > 0$, justifying the sign of the definition. In Eqs. (12.44) we have also used η for the normalization of $\tilde{f}_{\rm C}(k)$; since the original density must be normalized, it follows that the k-independent terms must sum to unity, and we have already implemented this. Further, the constant μ is analogous to the mean, although the mean is not necessarily defined here; it merely characterizes an asymmetry of the center part of the distribution.

Now the characteristic function for the scaled sum S_N is the product of N of the individual characteristic functions, rescaled according to $k \longrightarrow k/N^{1/\alpha}$:

$$\tilde{f}_{S_N}(k) = \left[1 + \frac{ik\mu}{N^{1/\alpha}} - \frac{\sigma^\alpha|k|^\alpha}{N} + O\left(\frac{k^2}{N^{2/\alpha}}\right)\right]^N. \tag{12.47}$$

In the limit $N \longrightarrow \infty$, clearly the $O(k^2)$ component becomes irrelevant, because $N/N^{2/\alpha} \longrightarrow 0$ for $0 < \alpha < 2$. Also for $0 < \alpha < 1$, the μ term vanishes in the $N \longrightarrow \infty$ limit. If $1 \le \alpha < 2$ this term stays; in this case it corresponds to a shift of the density of S_N by μ on each step. That is, the mean of S_N is $N\mu$ (conveniently, μ only matters when the mean is well-defined, though the marginal $\alpha = 1$ Cauchy case is only slightly trickier). If μ actually matters it can be removed by shifting the one-step distributions by μ without upsetting the asymptotic tails of the distribution, in the spirit of the proof of the central limit theorem. Doing so, we may then take the limit of many steps to find

$$\lim_{N\to\infty} \tilde{f}_{S_N}(k) = e^{-\sigma^\alpha|k|^\alpha}. \tag{12.48}$$

This is just the characteristic function (12.10) for the α-stable Lévy density.

12.2.1 Convergence Rate

We have seen that the (rescaled) distribution of a sum of random steps with heavy tails will converge to a stable distribution. It is interesting to ask, how rapid is this convergence? There are fortunately some nice results describing the decay of the difference $f_{S_N}(x) - f_\alpha(x)$ with N.[9] The convergence of the difference to zero is uniform, and is $O(1/N)$ for $0 < \alpha \le 1$, and $O(N^{(2-\alpha)/\alpha})$ for $1 < \alpha < 2$ [the latter scaling is suggested by the rate at which the $O(k^2)$ term vanishes in Eq. (refgencltcharfunc)]. Note that these results also hold for asymmetric stable distributions introduced below.

[9] Joseph B. Keller and Rachel Kuske, "Rate of Convergence to a Stable Law," *SIAM Journal on Applied Mathematics* **61**, 1308 (2001) (doi: 10.1137/S0036139998342715).

Interestingly, the Gaussian case $\alpha = 2$ can converge quite slowly, as $O(1/\log N)$. This ultraslow convergence to the Gaussian occurs, for example, with one-step distribution that have truncated power-law tails.[10] From the proof of the central limit theorem (Section 2.2), the decay of the $O(k^3)$ term in the characteristic function [Eq. (2.32)] suggests a Gaussian convergence as $O(1/\sqrt{N})$ if the one-step distribution has a nonvanishing third moment, and faster if the third moment vanishes. But this assumes that the higher moments are well behaved; the $O(1/\log N)$ Gaussian convergence applies when the tails of the one-step distribution decay as $1/|x|^3$ so that the high moments do not exist.

12.3 Asymmetric Stable Distributions

A generalization of the α-stable distributions to antisymmetric versions arises if we assume that the one-step distribution has one-step tails of the form

$$f(x) \sim \frac{c_+\Theta(x)}{|x|^{\alpha+1}} + \frac{c_-\Theta(-x)}{|x|^{\alpha+1}}, \qquad |x| \longrightarrow \infty,\ 0 < \alpha < 2, \tag{12.49}$$

with different weights c_+ and c_-. Note that it is sufficient to assume the same power for each tail. In the case where the tails scale differently, this corresponds to taking the coefficient for the ''lighter'' tail to be zero, as this tail will effectively disappear in the renormalized-sum statistic. As in Eqs. (12.39), we can split the characteristic function integral so that beyond $|x| > d$ (for fixed $d > 0$), we can use the asymptotic form of the distribution:

$$\begin{aligned}
\tilde{f}(k) &= 2\int_0^d dx\, f(x)\cos(kx) + \int_d^\infty dx\, f(x)\, e^{ikx} + \int_{-\infty}^{-d} dx\, f(x)\, e^{ikx} \\
&\approx 2\int_0^d dx\, f(x)\cos(kx) + \int_d^\infty dx\, \frac{c_+}{x^{\alpha+1}}\, e^{ikx} + \int_{-\infty}^{-d} dx\, \frac{c_-}{|x|^{\alpha+1}}\, e^{ikx} \\
&\approx 2\int_0^d dx\, f(x)\cos(kx) + c_+|k|^\alpha \int_{|k|d}^\infty \frac{dx}{x^{\alpha+1}}\, e^{ix\,\mathrm{sgn}k} + c_-|k|^\alpha \int_{|k|d}^\infty \frac{dx}{x^{\alpha+1}}\, e^{-ix\,\mathrm{sgn}k}.
\end{aligned} \tag{12.50}$$

The development of the characteristic function parallels the symmetric case, and the manipulations in Eqs. (12.41) become (note that the sign of k in the incomplete gamma functions is now significant)

$$\begin{aligned}
\tilde{f}(k) &\approx \tilde{f}_{\mathrm{C}}(k) + |k|^\alpha\Big((-i\,\mathrm{sgn}\,k)^\alpha c_+\Gamma(-\alpha,-ikd) + (i\,\mathrm{sgn}\,k)^\alpha c_-\Gamma(-\alpha,ikd)\Big) \\
&= \tilde{f}_{\mathrm{C}}(k) + c|k|^\alpha\Big((-i\,\mathrm{sgn}\,k)^\alpha\Gamma(-\alpha,-ikd) + \text{c.c.}\Big) + c\beta|k|^\alpha\Big((-i\,\mathrm{sgn}\,k)^\alpha\Gamma(-\alpha,-ikd) - \text{c.c.}\Big),
\end{aligned} \tag{12.51}$$

where we have defined the conventional parameters

$$c := \frac{c_+ + c_-}{2}, \qquad \beta := \frac{c_+ - c_-}{c_+ + c_-}. \tag{12.52}$$

Since $c_+ \geq 0$ and $c_- \geq 0$ with $c_+ + c_- > 0$ for a nontrivial distribution, we must have $-1 \leq \beta \leq 1$. Again using the expansion (12.43), Eq. (12.44) becomes

$$\begin{aligned}
\tilde{f}(k) &\approx \eta + ik\mu' + 2c|k|^\alpha\cos(\pi\alpha/2)\Gamma(-\alpha) + \frac{2c}{\alpha d^\alpha} \\
&\qquad - i2c\beta|k|^\alpha\sin(\pi\alpha/2)\,\mathrm{sgn}(k)\Gamma(-\alpha) - i\frac{2c\beta k}{(1-\alpha)d^{\alpha-1}} + O(k^2) \\
&= 1 + ik\mu + 2c|k|^\alpha\cos(\pi\alpha/2)\Gamma(-\alpha)\big[1 - i\beta\,\mathrm{sgn}(k)\tan(\pi\alpha/2)\big] + O(k^2) \\
&= 1 + ik\mu - \sigma^\alpha|k|^\alpha\big[1 - i\beta\,\mathrm{sgn}(k)\tan(\pi\alpha/2)\big] + O(k^2)
\end{aligned} \tag{12.53}$$

[10] Jun-ichi Inoue and Naoya Sazuka, ''Crossover between Lévy and Gaussian regimes in first-passage processes,'' *Physical Review E* **76**, 021111 (2007) (doi: 10.1103/PhysRevE.76.021111); Rosario N. Mantegna and H. Eugene Stanley, ''Stochastic Process with Ultraslow Convergence to a Gaussian: The Truncated Lévy Flight,'' *Physical Review Letters* **73**, 2946 (1994) (doi: 10.1103/PhysRevLett.73.2946).

where σ is defined as in Eq. (12.45), and we must handle the $\alpha = 1$ case (where the gamma function diverges and the cosine vanishes) separately. Also, we have written μ' as the first-order coefficient of $\tilde{f}_{\text{C}}(k)$, and then we combined this shift with the shift resulting from the tail asymmetry, defining

$$\mu := \mu' - \frac{2c\beta}{(1-\alpha)d^{\alpha-1}}. \tag{12.54}$$

Note that while μ' depends on d (being a moment of a truncated distribution), μ depends on the entire distribution, and is d-independent (provided d is taken to be sufficiently large). Taking the limit of many steps in the scaled-sum statistic goes through as before, with a modified coefficient of $|k|^\alpha$:

$$\lim_{N\to\infty} \tilde{f}_{S_N}(k) = \exp\Big(-\sigma^\alpha |k|^\alpha \big[1 - i\beta\,\mathrm{sgn}(k)\tan(\pi\alpha/2)\big]\Big). \tag{12.55}$$

In the case of $\alpha = 1$, the treatment above must be modified somewhat. Specifically, Eq. (12.43) should be replaced by the expansion

$$\Gamma(-1, ix) = -\frac{i}{x} - 1 + \gamma + \frac{i\pi\,\mathrm{sgn}(x)}{2} + \log|x| - \frac{ix}{2} - \frac{x^2}{12} + O(x^3) \tag{12.56}$$

where γ is Euler's constant. Then repeating the above manipulations, the one-step characteristic function from Eqs. (12.53) becomes

$$\begin{aligned}
\tilde{f}(k) &\approx \eta + ik\mu' + \frac{2c}{d} - \pi c|k| - i2(\gamma-1)c\beta k - i2c\beta k\log(|k|d) + O(k^2) \\
&= 1 + ik\mu - \pi c|k| - i2c\beta k\log|k| + O(k^2) \\
&= 1 + ik\mu - \pi c|k|\big[1 + i(2\beta/\pi)\,\mathrm{sgn}(k)\log|k|\big] + O(k^2) \\
&= 1 + ik\mu - \sigma|k|\big[1 + i(2\beta/\pi)\,\mathrm{sgn}(k)\log|k|\big] + O(k^2).
\end{aligned} \tag{12.57}$$

Then the rescaled density after N steps is

$$\begin{aligned}
\tilde{f}_{S_N}(k) &= \left[1 + \frac{ik\mu}{N} - \frac{\sigma|k|\big[1 + i(2\beta/\pi)\,\mathrm{sgn}(k)\log|k/N|\big]}{N} + O\left(\frac{k^2}{N^2}\right)\right]^N \\
&= \left[1 + \frac{ik\big[\mu + (2\sigma\beta/\pi)\log N\big]}{N} - \frac{\sigma|k|\big[1 + i(2\beta/\pi)\,\mathrm{sgn}(k)\log|k|\big]}{N} + O\left(\frac{k^2}{N^2}\right)\right]^N.
\end{aligned} \tag{12.58}$$

The second term corresponds to the magnification of the shift in the one-step distribution, though what is different now is that for $\beta \neq 0$, the shift is magnified after N steps by a part logarithmic in N, compared to what we would otherwise expect. Nevertheless, this term can be eliminated by appropriately shifting the one-step distributions, and so the limiting distribution for $\alpha = 1$ becomes

$$\lim_{N\to\infty} \tilde{f}_{S_N}(k) = \exp\Big(-\sigma|k|\big[1 + i(2\beta/\pi)\,\mathrm{sgn}(k)\log|k|\big]\Big). \tag{12.59}$$

The limiting form here only really differs from the $\alpha \neq 1$ expression in the asymmetric case $\beta \neq 0$.

12.3.1 Summary and Gallery

To summarize our findings here, we can assemble Eqs. (12.55) and (12.59), as a generalized form of the α-stable characteristic functions (12.8), but now with an asymmetry parameter β:

$$\begin{aligned}
&\tilde{f}_{\alpha,\beta,\sigma,\mu}(k;t) = \exp\Big(-t\sigma^\alpha|k|^\alpha\big[1 - i\beta\,\mathrm{sgn}(k)\tan(\pi\alpha/2)\big] + itk\mu\Big), && \big(0<\alpha<2,\ \alpha\neq 1,\ -1\le\beta\le 1\big) \\
&\tilde{f}_{1,\beta,\sigma,\mu}(k;t) = \exp\Big(-t\sigma|k|\big[1 + i(2\beta/\pi)\,\mathrm{sgn}(k)\log|k|\big] + itk\mu\Big) && \big(\alpha = 1,\ -1\le\beta\le 1\big).
\end{aligned}$$

(characteristic functions, generalized α-stable densities) (12.60)

Here, note that we have put in a factor of t so that these correspond to stochastic processes. We have also included a shift in the distribution of $t\mu$, as is consistent with the common general convention. The corresponding probability densities are adapted from Eqs. (12.9) as

$$f_{\alpha,\beta,\sigma,\mu}(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\, \tilde{f}_{\alpha,\beta,\sigma,\mu}(k;t).$$

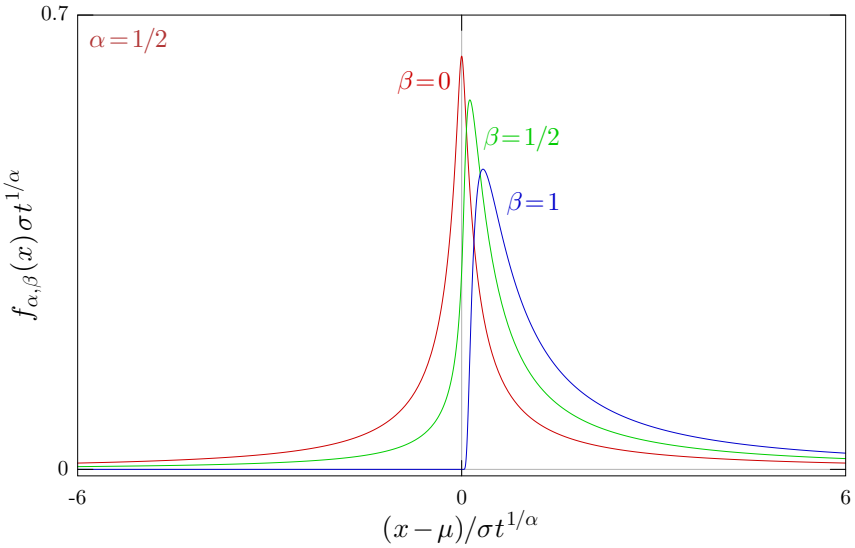
(probability density, α-stable Lévy process) (12.61)

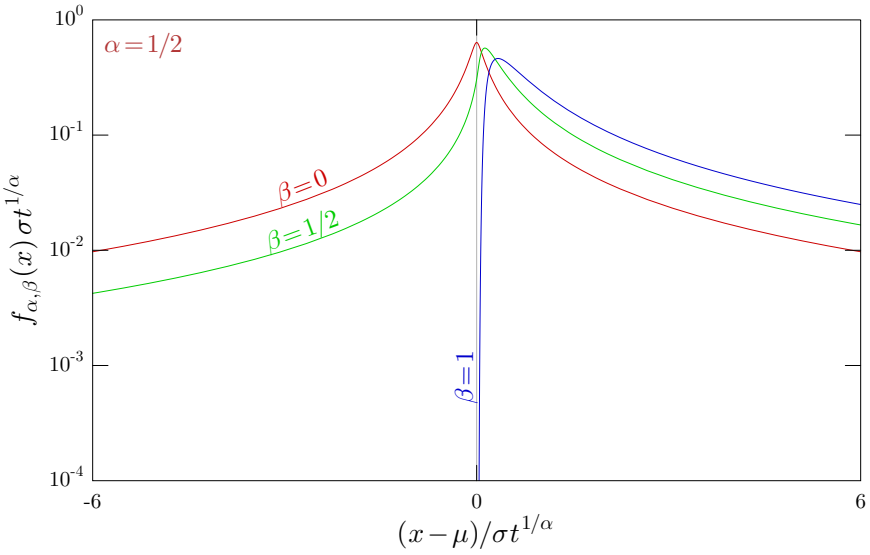
These densities are still called α-stable distributions, even though they now involve the asymmetry parameter β, because only the α parameter determines the temporal scaling of the distribution (or the scaling of the width under convolutions).

In the symmetric case, the distribution mean is always zero (where μ was missing from the parameterization), although this is only true in the sense of a regularized integral for $\alpha \leq 1$. In the more general case, we may read off the mean directly from the characteristic function (12.60) to see that it is μt for $1 < \alpha \leq 2$ (and undefined otherwise).

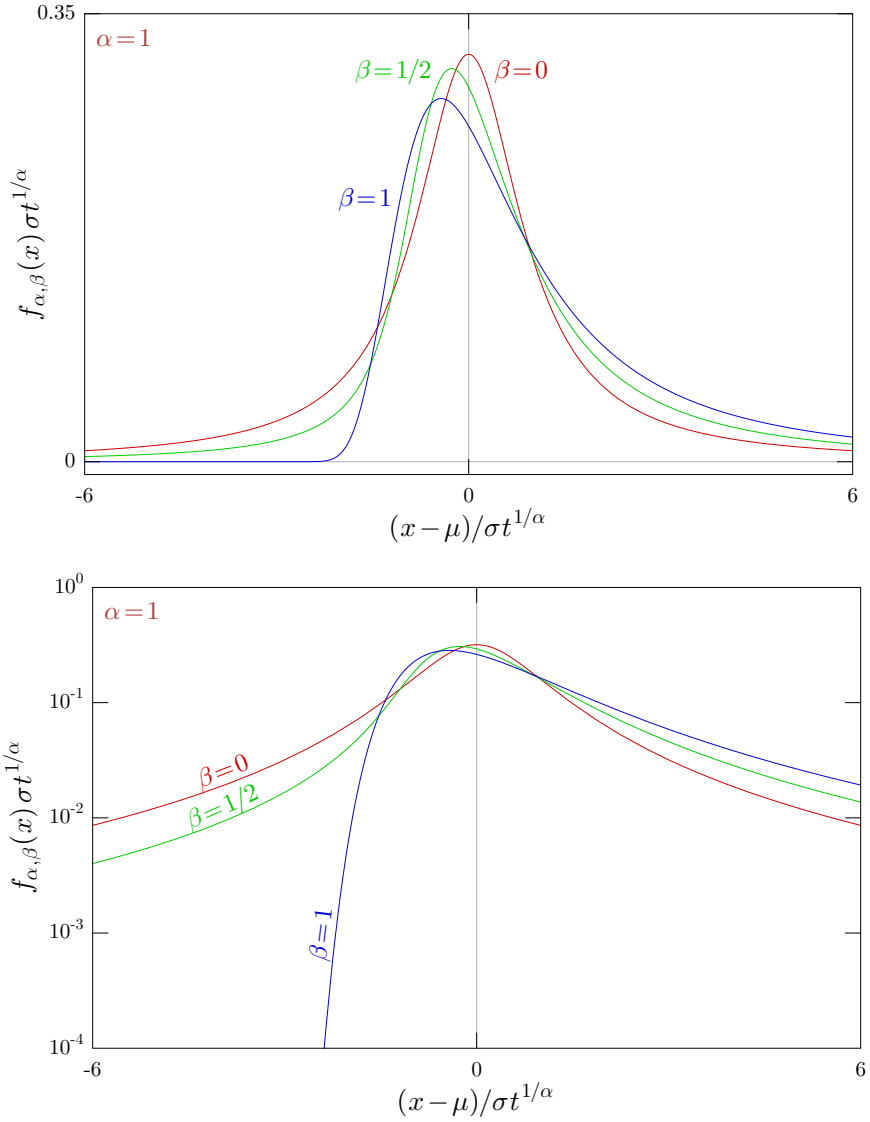
It also turns out that the support of the stable densities (i.e., the domain over which they are nonvanishing) is the entire real line $\mathbb{R}$ whenever $\alpha \geq 1$ or $\beta \neq \pm 1$. If $\beta = 1$ and $\alpha < 1$ then the support is $[\mu, \infty)$ (i.e., the positive real line in the absence of a shift μ), and $(-\infty, \mu]$ for $\beta = -1$.

Plotted below are stable densities for $\alpha = 1/2$ and three value of β, corresponding to symmetric ($\beta = 0$), partially asymmetric ($\beta = 1/2$), and maximally asymmetric ($\beta = 1$). The common scaling of the tails is more apparent in the logarithmic plot. Note that in the completely asymmetric case, the density completely vanishes for $x < 0$. The densities for $\beta < 0$ are of course the same as those for $\beta > 0$, but reflected about the $x = \mu$.

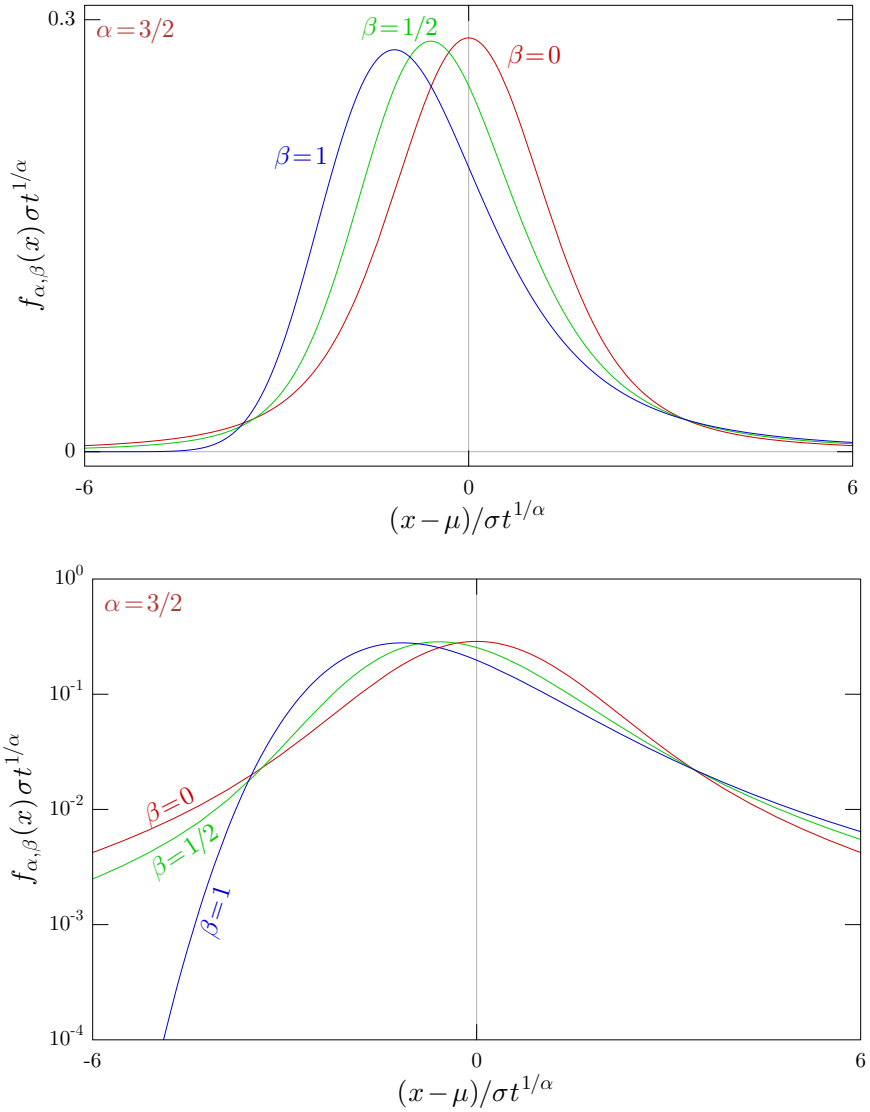



10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
$\alpha = 1/2$
$\beta = 0$
$\beta = 1/2$
$\beta = 1$
$f_{\alpha,\beta}(x)\,\sigma t^{1/\alpha}$
-6
0
6
$(x-\mu)/\sigma t^{1/\alpha}$

Plotted here are distributions for $\alpha = 1$, but the same values of β. Note that the asymmetry is manifest in a shift to the left, rather than to the right as in the $\alpha = 1/2$ case. Note also that the $\beta = 1$ distribution is not one-sided, but rather has an exponential-type decay for $x < 0$.



Finally, shown below are distributions for $\alpha = 3/2$, and again the same values of β. The behavior of the asymmetry is qualitatively the same as in the $\alpha = 1$ case, but with overall slower decay in the tails.



12.3.2 Asymptotics

To obtain the asymptotic tails of the distribution, note that we already set up the tails in Eq. (12.49), repeated here:

$$f(x) \sim \frac{c_+\Theta(x)}{|x|^{\alpha+1}} + \frac{c_-\Theta(-x)}{|x|^{\alpha+1}}, \qquad |x| \longrightarrow \infty,\ 0 < \alpha < 2. \tag{12.62}$$

We then worked out the form of the characteristic function in terms of the variables (12.52)

$$c := \frac{c_+ + c_-}{2}, \qquad \beta := \frac{c_+ - c_-}{c_+ + c_-}. \tag{12.63}$$

Solving for the tail coefficients, we have

$$c_\pm = (1 \pm \beta)c. \tag{12.64}$$

Using Eq. (12.45) to eliminate c, we find (with $\mu = 0$)

$$f_{\alpha,\beta,\sigma}(x;t) \sim \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\big(1+\operatorname{sgn}(x)\beta\big)\big(\sigma t^{1/\alpha}\big)^{\alpha}\,|x|^{-(\alpha+1)} \qquad \Big(|x| \gg \sigma t^{1/\alpha}\Big)$$

(asymptotic tails, asymmetric α-stable process) (12.65)

for the tails of the asymmetric Lévy densities, in the case $|\beta| \neq 1$. Dependence on μ can of course be restored by shifting $x \longrightarrow x - \mu$.

The situation is somewhat more complicated in the maximally asymmetric case $|\beta| = 1$, because the distributions only have one power-law tail (to which the above expression applies), and one "short" tail. Fortunately the form of the short tail is still known.[11] There are two cases to discuss for $\beta = 1$ (where the $\beta = -1$ case follows from mirror symmetry). The first case is for $\alpha \in [1, 2)$, in which case the short tail occurs as $x \longrightarrow -\infty$. The second case is for $\alpha \in (0, 1)$, where the support of the density function is $(0, \infty)$, and the short tail occurs as $x \longrightarrow 0$. In both cases, the expressions (for $\sigma = t = 1$ and $\mu = 0$)

$$f_{\alpha\neq 1,\beta}(x) \sim \frac{|x/\alpha|^{(1-\alpha/2)/(\alpha-1)}}{\sqrt{2\pi\alpha|1-\alpha|}}\,\exp\!\left(-|1-\alpha|\left|\frac{x}{\alpha}\right|^{\alpha/(\alpha-1)}\right)$$
$$f_{\alpha=1,\beta}(x) \sim \frac{1}{\sqrt{2\pi}}\,\exp\!\left(\frac{|x|-1}{2} - e^{|x|-1}\right)$$

(short asymptotic tails) (12.66)

apply. Again, dependence other parameters can be restored by rescaling $f(x)\,dx$ according to $x \longrightarrow x/\sigma t^{1/\alpha}$ and then shifting $x \longrightarrow x - \mu$.

12.3.3 Convolution Rules

Given a generalized set of densities, it is useful to also generalize the convolution relations from the symmetric case of Section 12.1.2. Again, setting $t = 1$ in Eqs. (12.60) gives the α-stable densities that are not specific to stochastic processes:

$$\tilde{f}_{\alpha,\beta,\sigma,\mu}(k) = \exp\Big(-\sigma^{\alpha}|k|^{\alpha}\big[1 - i\beta\operatorname{sgn}(k)\tan(\pi\alpha/2)\big] - ik\mu\Big), \qquad \big(0<\alpha<2,\ \alpha\neq 1,\ -1\leq\beta\leq 1\big)$$
$$\tilde{f}_{1,\beta,\sigma,\mu}(k) = \exp\Big(-\sigma|k|\big[1 + i(2\beta/\pi)\operatorname{sgn}(k)\log|k|\big] - ik\mu\Big) \qquad \big(\alpha = 1,\ -1\leq\beta\leq 1\big).$$

(characteristic functions, generalized α-stable densities) (12.67)

As before, suppose X_1 and X_2 are two α-stable-distributed random variables with parameter triples $(\sigma_1, \beta_1, \mu_1)$ and $(\sigma_2, \beta_2, \mu_2)$, respectively. (Note that α must be the same for both variables, otherwise the resulting density is not stable.) The characteristic function for $X_1 + X_2$ is the product of the individual characteristic functions, where we denote the parameters of the sum density by (σ, β, μ). From the form of the characteristic functions, the relation (12.13) clearly still holds:

$$\sigma = \Big(\sigma_1^{\alpha} + \sigma_2^{\alpha}\Big)^{1/\alpha}. \qquad (12.68)$$

(width parameter addition rule)

The β parameters combine in a similar way, since they appear with σ in the combination $\beta\sigma^{\alpha}$; the addition rule is then

$$\beta = \frac{\beta_1\sigma_1^{\alpha} + \beta_2\sigma_2^{\alpha}}{\sigma_1^{\alpha} + \sigma_2^{\alpha}}. \qquad (12.69)$$

(asymmetry parameter addition rule)

Finally, since the shift parameter appears as a separate term in the exponential, we have

$$\mu = \mu_1 + \mu_2. \qquad (12.70)$$

(shift parameter addition rule)

[11] Vladimir V. Uchaikin and Vladimir M. Zolotarev, *Chance and Stability. Stable Distributions and their Applications* (VSP, 1999), p. 127 (ISBN: 9067643017).

as the simple rule for combining shifts. However, there is an extra subtlety here. If X is an α-stable variable with $\alpha \neq 1$ and parameter triple (σ, β, μ), the shifted, scaled version $aX + b$ has the parameter triple $(|a|\sigma, \mathrm{sgn}(a)\beta, a\mu + b)$. On the other hand, because of the form of the characteristic function, the parameter triple in the case $\alpha = 1$ is $(|a|\sigma, \mathrm{sgn}(a)\beta, a\mu + b - (2\beta/\pi)a \log|a|)$. That is, the asymmetry parameter induces some shift on a width rescaling of the distribution. Thus, in the case $\alpha = 1$, if a summing over N random variables with $\alpha = 1$ and parameters $(\sigma/N, \beta, \mu = 0)$, corresponding to N steps each over time $\Delta t = 1/N$, the sum of the steps has parameters $(\sigma, \beta, (2\beta/\pi)\log N)$, which has an overall shift parameter despite the unshifted steps.

12.3.4 Lévy-Density Process ($\alpha = 1/2$, $\beta = 1$)

Now that we have introduced the asymmetric stable densities, there is one more well known case where the density integral

$$f_{\alpha,\beta,\sigma,\mu}(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\, \tilde{f}_{\alpha,\beta,\sigma,\mu}(k;t) \tag{12.71}$$

can be evaluated in a simple closed form: for $\alpha = 1/2$ and $\beta = 1$:

$$f_{1/2,1,\sigma,\mu}(x;t) = \sqrt{\frac{\sigma}{2\pi}}\,\frac{1}{(x-\mu)^{3/2}}\, e^{-\sigma/2(x-\mu)}\,\Theta(x-\mu). \tag{12.72}$$
(Lévy distribution)

This density is called the **Lévy distribution**. Of course, the case $\alpha = 1/2$ and $\beta = -1$, the ''reflected Lévy distribution,'' has the same form but reflected about $x = \mu$. We have encountered this distribution function before; it arose in as the first-passage density for a Wiener process crossing a boundary at d, with the identifications $\mu = 0$ and $\sigma = d^2$ [Eq. (6.11)].

12.4 Continuity of Sample Paths

We mentioned in the defintion of Lévy processes (p. 227) that they must have the property of stochastic continuity, which says that for any $\varepsilon > 0$, the probability that $|\Delta L(t)| = |L(t + \Delta t) - L(t)| > \varepsilon$ converges to zero as $\Delta t \longrightarrow 0$. In terms of the conditional density, this condition reads

$$\lim_{\Delta t \to 0} \int_{|x - x_0| > \varepsilon} dx\; f(x, t + \Delta t | x_0, t_0) = 0 \tag{12.73}$$

A Poisson process is stochastically continuous, because the probability of jump is infinitesimal in a time dt. More concretely, if $\varepsilon < 1$, the above integral takes on the value $\Gamma\,\Delta t$ for a standard Poisson process, to lowest order in Δt. Although it satisfies the condition for stochastic continuity, the sample paths are clearly *not* continuous: the Poisson process is a prototype example of a jump process.

If we are interested in whether the sample paths of a stochastic process are continuous functions of time, it is necessary to come up with a different (stronger) continuity condition. Thinking about an ordinary function $g(t)$, continuity means that given some $\varepsilon > 0$, there is some small time perturbation Δt such that $g(t + \Delta t)$ falls within ε of $g(t)$. This is enough to guarantee that $g(t + \Delta t) - g(t)$ will decay at least as quickly as Δt as $\Delta t \longrightarrow 0$. The problem with a stochastic process is that the *probability* can also fade like Δt as $\Delta t \longrightarrow$, without the possible increments of the process also contracting to zero.

Thus, for continuous sample paths we will require the condition[12]

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x - x_0| > \varepsilon} dx\; f(x, t + \Delta t | x_0, t_0) = 0,$$

(condition for continuous sample paths) (12.74)

[12] As in Crispin Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer, 2009) (ISBN: 9783540707127), Section 3.3.1, p, 46.

so that the probability for $|\Delta L(t)| = |L(t+\Delta t) - L(t)| > \varepsilon$ still converges to zero as $\Delta t \longrightarrow 0$, but it must do so *more quickly than* Δt. This explicitly excludes the simple Poisson process, for which the limit on the left-hand side is Γ for any $\varepsilon < 1$.

A Wiener process satisfies this condition, so that a Wiener process has continuous sample paths (with unit probability). Specifically, the transition density is

$$f(x, t+\Delta t|x_0, t_0) = \frac{1}{\sqrt{2\pi\Delta t}}\, e^{-(x-x_0)^2/2\Delta t} \tag{12.75}$$

with integral

$$\int_{|x-x_0|>\varepsilon} dx\; f(x, t+\Delta t|x_0, t_0) = \operatorname{erfc}\left(\frac{\varepsilon}{\sqrt{2\Delta t}}\right). \tag{12.76}$$

For small Δt, this function cuts off exponentially (as $\Delta t^{1/2} e^{-\varepsilon^2/2\Delta t}$), so this converges to zero even with the extra factor of Δt^{-1}. The same conclusion of course holds for an $\alpha = 2$ stable Lévy process. A Cauchy process, on the other hand, is discontinuous, as we saw before in numerical plots. From the distribution (12.16), we have (with $\sigma = 1$)

$$f(x, t+\Delta t|x_0, t_0) = \frac{\Delta t}{\pi[(\Delta t)^2 + (x-x_0)^2)}, \tag{12.77}$$

with integral

$$\int_{|x-x_0|>\varepsilon} dx\; f(x, t+\Delta t|x_0, t_0) = 1 - \frac{2}{\pi}\tan^{-1}\left(\frac{\varepsilon}{\Delta t}\right). \tag{12.78}$$

Since $\tan^{-1}(1/x) \approx \pi/2 - x$ for large $x > 0$, this result goes as Δt to leading order, and so with the extra factor of Δt^{-1}, this integral does not converge to zero. As we have seen, Cauchy processes have jumps. In fact, with the transition density of the α-stable processes having tails (12.23) that scale as $dt\,|x|^{-(\alpha_1)}$ ($0 < \alpha < 2$), the integral has a component that scales as $\Delta t/\alpha\varepsilon^\alpha$ (for large ε), and so any α-stable process besides $\alpha = 2$ fails the continuity test.

12.5 Master Equation

Ensembles of stochastic processes evolve according to partial differential equations—Gaussian processes, for example, correspond to the diffusion equation, which is the simplest instance of the Fokker–Planck equation (Section 3.3.5.1). Here we will consider partial differential equations for Lévy processes in several useful forms. Since Lévy processes are Markov processes, which in turn satisfy the Chapman–Kolmogorov equation [Eq. (3.17)], the equations here will correspond to a differential version of the Chapman–Kolmogorov equation—the differential form is also called the **master equation**.

12.5.1 Fractional Diffusion

First, starting with only the symmetric case (12.8) of the characteristic function for α-stable processes, we can write the characteristic function for evolution from t_0 to t as

$$\tilde{f}_\alpha(k; t|x_0, t_0) = e^{-(t-t_0)\sigma^\alpha|k|^\alpha + ikx_0}, \tag{12.79}$$

including a shift factor for the initial position x_0. A simple time derivative gives the differential equation.

$$\partial_t \tilde{f}_\alpha(k; t|x_0, t_0) = -\sigma^\alpha|k|^\alpha\, \tilde{f}_\alpha(k; t|x_0, 0). \tag{12.80}$$

A Fourier transform of this equation to position space gives the

$$\partial_t f_\alpha(x, t|x_0, t_0) = \sigma^\alpha \frac{\partial^\alpha}{\partial|x|^\alpha} f_\alpha(x, t|x_0, t_0), \tag{12.81}$$

(fractional diffusion equation)

where the **fractional derivative** is defined by the Fourier transform

$$\frac{\partial^\alpha}{\partial|x|^\alpha} f(x) := \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,\big(-|k|^\alpha\big)\tilde{f}(k)\,e^{ikx} = \frac{1}{\pi}\int_0^{\infty} dk\,\big(-|k|^\alpha\big)\tilde{f}(k)\cos(kx)$$

(fractional Laplace operator) (12.82)

(in this form, $\partial^\alpha/\partial|x|^\alpha$ is often called the **Riesz–Feller derivative**), by analogy to the ordinary second derivative

$$\frac{\partial^2}{\partial x^2} f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,\big(-k^2\big)\tilde{f}(k)\,e^{ikx}. \tag{12.83}$$

Unlike the ordinary second derivative (or ordinary first derivative), the fractional derivative is a nonlocal operator via the Fourier transform—roughly speaking this is how it effects long tails and superdiffusive behavior. Another common notation for this fractional derivative in multiple dimensions is as $\partial^\alpha/\partial|x|^\alpha \equiv -(-\triangle)^{\alpha/2}$, where $\triangle \equiv \nabla^2$ is the ordinary Laplacian operator (the sign $-\nabla^2$ is chosen here to select a positive definite operator, so that the fractional power doesn't introduce any complex eigenvalues). Thus, the derivative $\partial^\alpha/\partial|x|^\alpha$ in the form here is also called the **fractional Laplacian** (see Section 12.6.2 for more details). The fractional derivative operator is called a **pseudo-differential operator**, since it generalizes the concept of differentiation without the usual construction of a derivative in terms of ratios of infinitesimal differences.

Another common representation of the fractional derivative arises by considering the inverse-Fourier-transform relation

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,\big(-|k|^\alpha\big)\,e^{ikx} = \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\,|x|^{-(\alpha+1)}. \tag{12.84}$$

Of course, this integral is nonsense for the range of interest for α (i.e., $\alpha \in (0,2]$ for the α-stable processes). For now, we can simply note that this relation can be justified by inserting a convergence factor $e^{-|k|\lambda}$ in the integral, and letting $\lambda \longrightarrow 0$ at the end of the day. Physically, ignoring large $|k|$ amounts to ignoring arbitrarily large velocities associated with α-stable processes, which is physically reasonable, although we see that the exact cutoff is not important. Now using the convolution theorem (2.14), applied to the product $(-|k|^\alpha)\,\tilde{f}(k)$ in Eqs. (12.82), we have

$$\frac{\partial^\alpha}{\partial|x|^\alpha}\,f(x) = \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\int_{-\infty}^{\infty} dx'\,\frac{f(x')}{|x-x'|^{\alpha+1}}. \tag{12.85}$$

Again, this expression is divergent for the relevant range of α. However, the integrals here and in Eq. (12.84) *do* make sense for $\alpha \in (-1,0]$—in which case the expression is a fractional *integration* rather than a derivative. Thus, to treat the expressions here more carefully, for example in the case of $\alpha \in (1,2)$ we can factor out a second derivative by rewriting Eq. (12.82) as

$$\frac{\partial^{\alpha-2}}{\partial|x|^{\alpha-2}} f''(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,\big(|k|^{\alpha-2}\big)(-k^2)\tilde{f}(k)\,e^{ikx}, \tag{12.86}$$

with $-k^2\tilde{f}(k)$ as the Fourier transform of $f''(x)$. That is, the fractional part of the derivative is

$$\frac{\partial^{\alpha-2}}{\partial|x|^{\alpha-2}} f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,\big(|k|^{\alpha-2}\big)\tilde{f}(k)\,e^{ikx}. \tag{12.87}$$

In this case we can adapt Eq. (12.85) with $\alpha \longrightarrow \alpha - 2$ for the fractional part $(\alpha-2)$ of the derivative, tacking on an explicit second derivative, in order to write

$$\frac{\partial^\alpha}{\partial|x|^\alpha}\,f(x) = \frac{\Gamma(\alpha-1)\sin(\pi\alpha/2)}{\pi}\,\frac{d^2}{dx^2}\int_{-\infty}^{\infty} dx'\,\frac{f(x')}{|x-x'|^{\alpha-1}}, \qquad (1<\alpha<2). \tag{12.88}$$

For the case $\alpha \in (0,1)$, we can instead differentiate Eq. (12.85),

$$\frac{\partial^{\alpha+1}}{\partial|x|^{\alpha+1}}\,f(x) = -\frac{\Gamma(\alpha+2)\sin(\pi\alpha/2)}{\pi}\int_{-\infty}^{\infty} dx'\,\frac{f(x')\,\mathrm{sgn}(x'-x)}{|x-x'|^{\alpha+2}}. \tag{12.89}$$

Then letting $\alpha \longrightarrow \alpha - 2$, and introducing an explicit derivative to write

$$\frac{\partial^\alpha}{\partial|x|^\alpha} f(x) = \frac{\Gamma(\alpha)\sin(\pi\alpha/2)}{\pi} \frac{d}{dx} \int_{-\infty}^{\infty} dx' \, \frac{f(x')\,\mathrm{sgn}(x'-x)}{|x-x'|^\alpha}, \qquad (0 < \alpha < 1). \tag{12.90}$$

Note that the integral here is sensible for the desired range of α. Finally, in the case of $\alpha = 1$, we can let $\alpha \longrightarrow 1$ in Eq. (12.90) to obtain

$$\frac{\partial^1}{\partial|x|^1} f(x) = \frac{1}{\pi} \frac{d}{dx} \int_{-\infty}^{\infty} dx' \, \frac{f(x')}{x'-x}, \tag{12.91}$$

where the resulting integral—which has the form of a Hilbert transform—should now be interpreted in terms of the Cauchy principal value due to the $1/x'$ divergence. This integral is fine, but we can obtain another expression by introducing a derivative and a compensating integration:

$$\frac{\partial^1}{\partial|x|^1} f(x) = \frac{1}{\pi} \frac{d^2}{dx^2} \int_{-\infty}^{\infty} dx' \, f(x') \, \log \frac{1}{|x-x'|}, \tag{12.92}$$

This is also a convergent integral, provided $f(x)$ decays to zero sufficiently quickly as $|x| \longrightarrow \infty$. To summarize this convolution representation of the fractional Laplacian, combining Eqs. (12.88), (12.90), (12.91) and (12.92), we have

$$\frac{\partial^\alpha}{\partial|x|^\alpha} f(x) = \begin{cases} \dfrac{\Gamma(\alpha-1)\sin(\pi\alpha/2)}{\pi} \dfrac{d^2}{dx^2} \displaystyle\int_{-\infty}^{\infty} dx' \, \frac{f(x')}{|x-x'|^{\alpha-1}} & (1 < \alpha < 2) \\ \dfrac{1}{\pi} \dfrac{d^2}{dx^2} \displaystyle\int_{-\infty}^{\infty} dx' \, f(x') \, \log \frac{1}{|x-x'|} = \frac{1}{\pi} \frac{d}{dx} \int_{-\infty}^{\infty} dx' \, \frac{f(x')}{x'-x} & (\alpha = 1) \\ \dfrac{\Gamma(\alpha)\sin(\pi\alpha/2)}{\pi} \dfrac{d}{dx} \displaystyle\int_{-\infty}^{\infty} dx' \, \frac{f(x')\,\mathrm{sgn}(x'-x)}{|x-x'|^\alpha} & (0 < \alpha < 1). \end{cases}$$

(fractional Laplacian, integral form) (12.93)

Of course the expressions here also serve as forms for the infinitesimal generators [recall the discussion around Eq. (3.59)] of the symmetric α-stable processes, in terms of which the master equation (12.81) becomes an integro-differential equation. Although these are the most relevant derivatives for the α-stable processes, higher-order derivatives can be constructed from these expressions by applying two derivatives at a time.

The above (brief) treatment relied on a Fourier transform to define a fractional derivative, but this is not the only possible definition. In particular, derivatives can be defined in such a way as to reduce to an ordinary single derivative in the case $\alpha = 1$. Additionally, the derivatives here are only valid on an unbounded domain; working with such operators on a bounded domain is in fact quite tricky, for reasons we will see later.[13] In fact, the forms (12.93) are the ones that generalize straightforwardly, since the integration limits can be modified to match a bounded domain.[14]

12.5.2 General Form: Master Equation

In the spirit of the derivation of the Fokker–Planck equation of Section 3.3.5.1, let's start by computing the derivative of an expectation value of some function $g(x)$, with respect to a conditional density $f(x,t|x_0,t_0)$:

$$\partial_t \int dx \, g(x) \, f(x,t|x_0,t_0) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int dx \, g(x) \Big[f(x,t+\Delta t|x_0,t_0) - f(x,t|x_0,t_0) \Big]. \tag{12.94}$$

[13] B. Dybiec, E. Gudowska-Nowak, and P. Hänggi, "Lévy-Brownian motion on finite intervals: Mean first passage time analysis," *Physical Review E* **73**, 046104 (2006) (doi: 10.1103/PhysRevE.73.046104); this paper has an interesting comment on other papers that misapplied the boundary condition to give suspect results. See also A. Zoia, A. Rosso, and M. Kardar, "Fractional Laplacian in bounded domains," *Physical Review E* **76**, 021116 (2007) (doi: 10.1103/PhysRevE.76.021116).

[14] Shinzō Watanabe (渡辺 信三), "On stable processes with boundary conditions," *Journal of the Mathematical Society of Japan* **14**, 170 (1962) (doi: 10.2969/jmsj/01420170).

Applying the Chapman–Kolmogorov equation (3.17),

$$\partial_t \int dx\, g(x)\, f(x,t|x_0,t_0) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \int dx\, g(x) \Big[\int f(x,t+\Delta t|x',t)\, f(x',t|x_0,t_0)\, dx' - f(x,t|x_0,t_0) \Big]. \tag{12.95}$$

We can rewrite this as

$$\partial_t \int dx\, g(x)\, f(x,t|x_0,t_0) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \int dx \int dx'\, g(x) \Big[f(x,t+\Delta t|x',t)\, f(x',t|x_0,t_0) - f(x',t+\Delta t|x,t)\, f(x,t|x_0,t_0) \Big], \tag{12.96}$$

where the new factor in the last term integrates to unity, because it is a normalized density in x'. Now we can identify $\Delta t^{-1} f(x,t+\Delta t|x',t)$ as the transition rate from x' to x for a small (infinitesimal) time step Δt. Since this is an important quantity, which only depends on $x-x'$, suppose that we define the transition rate

$$W(x-x') := \lim_{\Delta t\to 0} \frac{1}{\Delta t}\, f(x,t+\Delta t|x',t). \tag{12.97}$$

Then Eq. (12.96) becomes

$$\partial_t \int dx\, g(x)\, f(x,t|x_0,t_0) = \int dx \int dx'\, g(x) \Big[W(x-x')\, f(x',t|x_0,t_0) - W(x'-x)\, f(x,t|x_0,t_0) \Big], \tag{12.98}$$

and if we drop the expectation value with respect to the arbitrary function, we arrive at

$$\partial_t f(x,t|x_0,t_0) = \int dx' \Big[W(x-x')\, f(x',t|x_0,t_0) - W(x'-x)\, f(x,t|x_0,t_0) \Big], \tag{12.99}$$

(master equation, Lévy process)

which is the **master equation** for any Lévy process. The first term here counts all the transitions from all possible source points x' into x in time dt; the second term counts the transitions *away from* x to all possible destinations x'.

Now consider the Fourier transform of the master equation (12.99), in terms of the characteristic function:

$$\partial_t \tilde{f}(k,t|x_0,t_0) = \tilde{W}(k)\, \tilde{f}(k,t|x_0,t_0) - \gamma\, \tilde{f}(k,t|x_0,t_0). \tag{12.100}$$

In the first term, the convolution theorem allows us to write the product of the Fourier transform of $W(x)$ with the characteristic function, and in the second term, we used the integral

$$\gamma := \int dx\, W(x), \tag{12.101}$$

which reflects the total rate of transitions defined by $W(x)$. This equation now has the simple exponential solution

$$\tilde{f}(k,t|x_0,t_0) = \tilde{f}(k,t_0|x_0,t_0)\, \exp\Big[(t-t_0) \big[\tilde{W}(k) - \gamma \big] \Big]. \tag{12.102}$$

The dependence on x_0 and t_0 amounts to a space-time shift; thus we may rewrite $\tilde{f}(k,t_0|x_0,t_0)$ as an exponential shift factor:

$$\tilde{f}(k,t|x_0,t_0) = \exp\Big[(t-t_0) \big[\tilde{W}(k) - \gamma \big] + ikx_0 \Big]. \tag{12.103}$$

Assuming a transition rate with power-law tails of the form

$$W(x) \sim c|x|^{-(\alpha+1)}, \tag{12.104}$$

and assuming for simplicity a symmetric transition rate $W(x)$, we can adapt Eq. (12.44) by noting that $W(x)/\gamma$ acts as a probability density, with result

$$\tilde{W}(k) = \gamma - \gamma\sigma^\alpha |k|^\alpha + O(k^2). \tag{12.105}$$

Here σ is still defined by Eq. (12.45). Note that the leading term in Eq. (12.105) changed from 1 to γ to reflect the proper normalization of the transition rate, and we ignored any shift μ via the assumption of symmetry of $W(x)$. If we put the leading part of this expression into the solution (12.103) to find the solution

$$\tilde{f}(k,t|x_0,t_0) = \exp\Big[-\gamma(t-t_0)\sigma^\alpha|k|^\alpha + ikx_0\Big]. \tag{12.106}$$

This is just a shifted form of the symmetric α-stable characteristic function (12.10), with the time shifted and scaled by the transition rate γ. Again, this represents an asymptotic solution since the correction terms that we ignored in Eq. (12.105) become less important compared to the $|k|^\alpha$ term as time progresses.

12.5.3 Differential Chapman–Kolmogorov Equation

A modification to the above derivation of the master equation leads to an important equation of motion for Lévy processes.[15] Starting with Eq. (12.96), suppose that we take $\varepsilon > 0$ (where later $\varepsilon \longrightarrow 0^+$), and divide the integrals into sections where $|x-x'|$ is either larger or smaller than ε:

$$\begin{aligned}
&\partial_t \int dx\, g(x)\, f(x,t|x_0,t_0) \\
&\quad = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \iint\limits_{|x-x'|<\varepsilon} dx\, dx'\, g(x)\Big[f(x,t+\Delta t|x',t)\, f(x',t|x_0,t_0) - f(x',t+\Delta t|x,t)\, f(x,t|x_0,t_0)\Big] \\
&\qquad + \lim_{\Delta t\to 0} \frac{1}{\Delta t} \iint\limits_{|x-x'|\geq\varepsilon} dx\, dx'\, g(x)\Big[f(x,t+\Delta t|x',t)\, f(x',t|x_0,t_0) - f(x',t+\Delta t|x,t)\, f(x,t|x_0,t_0)\Big].
\end{aligned} \tag{12.107}$$

We can handle the second integral in essentially the same way as before by defining the transition rate density

$$W_\varepsilon(x-x') := \lim_{\Delta t\to 0} \frac{1}{\Delta t}\, f(x,t+\Delta t|x',t) \tag{12.108}$$

in place of $W(x-x')$ in Eq. (12.97), where the difference now is that the convergence to $W_\varepsilon(x-x')$ is assumed to be *uniform* for $|x-x'| \geq \varepsilon$. Colloquially, uniform convergence means that the function is getting closer *everywhere* to its limit as $\Delta t \longrightarrow 0$. (A example of nonuniform convergence is $\exp[-n^2(x-1/n)^2]$ on $(0,\infty)$ as $n \longrightarrow \infty$, which is a Gaussian function of unit height, peak location $1/n$, and width of order $1/n$; this function has the limit zero everywhere, but for any n the Gaussian is unity somewhere, so the convergence is not uniform.) Note that the exclusion of small $|x-x'|$ in this definition will allow for a divergence at $x = x'$ without upsetting the uniform convergence. Thus, Eq. (12.109) becomes

$$\begin{aligned}
&\partial_t \int dx\, g(x)\, f(x,t|x_0,t_0) \\
&\quad = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \iint\limits_{|x-x'|<\varepsilon} dx\, dx'\, g(x)\Big[f(x,t+\Delta t|x',t)\, f(x',t|x_0,t_0) - f(x',t+\Delta t|x,t)\, f(x,t|x_0,t_0)\Big] \\
&\qquad + \iint\limits_{|x-x'|\geq\varepsilon} dx\, dx'\, g(x)\Big[W_\varepsilon(x-x')\, f(x',t|x_0,t_0) - W_\varepsilon(x'-x)\, f(x,t|x_0,t_0)\Big].
\end{aligned} \tag{12.109}$$

In the first integral we will make the expansion

$$g(x') = g(x) + (x'-x)g'(x) + \frac{1}{2}(x'-x)^2 g''(x) + O[(x'-x)^3], \tag{12.110}$$

[15] Here we are more or less following C. Gardiner, *op. cit.*, Section 3.4.1, p. 48. See this reference for more rigor and generality. This is also the source of the term "differential Chapman–Kolmogorov equation" for Eq. (12.116).

and we will assume we have chosen $g(x)$ such that this expansion is sensible over $|x - x'| < \varepsilon$. To go along with the expansion coefficients we will also define conditional moments; the first moment

$$A(x) := \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|x'-x|<\varepsilon} dx' \, (x' - x) \, f(x', t + \Delta t | x, t) + O(\varepsilon) \tag{12.111}$$

will characterize the drift, while the second moment

$$B(x) := \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|x'-x|<\varepsilon} dx' \, (x' - x)^2 \, f(x', t + \Delta t | x, t) + O(\varepsilon) \tag{12.112}$$

will characterize diffusion, just as in the Fokker–Planck equation. These are moments of a truncated (transition) density and therefore should exist provided $f(x', t + \Delta t|x, t)$ is physically sensible. We do not need to consider higher moments, though, as on the bounded interval they will be at most $O(\varepsilon)$ compared to $B(x)$. Thus, taking Eq. (12.109), exchanging the labels x and x' in the first term of the first integral, using the expansion (12.110), and then writing the results in terms of the above moments, we find

$$\begin{aligned} \partial_t \int dx \, g(x) \, f(x, t|x_0, t_0) = \int dx \left[A(x) \, g'(x) + \frac{1}{2} B(x) \, g''(x) \right] f(x, t|x_0, t_0) \\ + \iint\limits_{|x-x'|\geq\varepsilon} dx \, dx' \, g(x) \Big[W_\varepsilon(x - x') \, f(x', t|x_0, t_0) - W_\varepsilon(x' - x) \, f(x, t|x_0, t_0) \Big], \end{aligned} \tag{12.113}$$

where we have already dropped any $O(\varepsilon)$ contributions in anticipation of the upcoming limit. Now taking the limit $\varepsilon \longrightarrow 0$, the only remaining change is in the last integral, which we can write

$$\begin{aligned} \partial_t \int dx \, g(x) \, f(x, t|x_0, t_0) = \int dx \, g(x) \Big[- \partial_x \big[A(x) f(x, t|x_0, t_0) \big] + \frac{1}{2} \partial_x^2 \big[B(x) f(x, t|x_0, t_0) \big] \Big] \\ + \int dx \, g(x) \fint dx' \Big[W_0(x - x') \, f(x', t|x_0, t_0) - W_0(x' - x) \, f(x, t|x_0, t_0) \Big], \end{aligned} \tag{12.114}$$

where we have defined the Cauchy principle value integral via

$$\fint dx' \, h(x', x) := \lim_{\varepsilon \to 0} \int\limits_{|x-x'|\geq\varepsilon} dx' \, h(x', x) \tag{12.115}$$

for a test function $h(x', x)$, which will allow the resulting integral to be defined in case of an antisymmetric singular part of $h(x', x)$ at $x' = x$. We have also integrated by parts in the first integral in order to factor out the test function $g(x)$. The remaining integration with respect to $g(x)$ can now be discarded, to obtain the result

$$\begin{aligned} \partial_t f(x, t|x_0, t_0) = -a \, \partial_x f(x, t|x_0, t_0) + \frac{1}{2} b^2 \, \partial_x^2 f(x, t|x_0, t_0) \\ + \fint dx' \Big[W_0(x - x') \, f(x', t|x_0, t_0) - W_0(x' - x) \, f(x, t|x_0, t_0) \Big], \end{aligned}$$

(differential Chapman–Kolmogorov equation) (12.116)

where we are assuming constants $a = A(x)$ and $b^2 = B(x)$. This equation, called the **differential Chapman–Kolmogorov equation**, generalizes the Fokker–Planck equation (Section 3.3.5.1) with the inclusion of the final integral term. These will induce longer range jumps that are not included in the drift or diffusion actions of the other terms. Note also that the drift and diffusion coefficients (not to mention W_0) can depend on time, but the idea here is to model Lévy process, which are by definition stationary—this is the reason that we discarded the explicit spatial dependence of the drift and diffusion coefficients. Notice that this equation is equivalent to the master equation (12.99), differing only in how the contributions to the density evolution are written out. Indeed, setting the drift and diffusion to zero simply recovers the master equation, though the form here allows for a singular part of the transition density $W_0(x)$.

12.5.4 Lévy–Khintchine Representation

At this we can consider a Fourier transform to the characteristic function in order to obtain a solution to Eq. (12.116). In this equation, we can begin by shifting the x' integration by x and then also reversing the sign of the integration in the first term, to obtain

$$\begin{aligned}\partial_t f(x,t|x_0,t_0) = -a\,\partial_x f(x,t|x_0,t_0) + \frac{1}{2}b^2\,\partial_x^2 f(x,t|x_0,t_0) \\ + \fint dx'\,W_0(x')\Big[f(x-x',t|x_0,t_0) - f(x,t|x_0,t_0)\Big],\end{aligned} \tag{12.117}$$

Multiplying by e^{ikx} and integrating over x, we have

$$\begin{aligned}\partial_t \tilde{f}(k,t|x_0,t_0) = ika\,\tilde{f}(k,t|x_0,t_0) - \frac{k^2}{2}b^2\,\tilde{f}(k,t|x_0,t_0) \\ + \fint dx'\,W_0(x')\Big(e^{ikx'}-1\Big)\tilde{f}(k,t|x_0,t_0),\end{aligned} \tag{12.118}$$

where we used the usual shifted transform

$$\tilde{f}(k)\,e^{ikx'} = \int dx\,e^{ikx}\,f(x-x'). \tag{12.119}$$

Integrating this equation from t_0 to t,

$$\tilde{f}(k,t|x_0,t_0) = \exp\left[(t-t_0)\left(ika - \frac{b^2}{2}k^2 + \fint dx'\,W_0(x')\Big(e^{ikx'}-1\Big)\right)\right], \tag{12.120}$$

where again the initial condition $\tilde{f}(k,t_0|x_0,t_0) = 1$ for the result to be a sensible characteristic function. Note that the a term represents a drift in the center of the distribution, whereas the b^2 term represents Gaussian diffusion. The remaining term bears further discussion.

We must now be a little more careful with $W_0(x)$, which we said could diverge at $x = 0$. It is most useful for the moment to think of the $e^{ikx'}$ factor in terms of its power-series expansion. The zeroth-order term has already been removed, which is good since it could otherwise affect the normalization of the distribution. The first-order term (linear in k) represents an additional drift of the distribution. This is not a problem unless the principal-value integral of $x'W_0(x')$ diverges, in which case we would be stuck with an infinite drift. To explicitly rule out this case, we will ''excise'' this problematic term in some region around $x' = 0$, say $|x'| < 1$,

$$\tilde{f}(k,t|x_0,t_0) = \exp\left[(t-t_0)\left(ika - \frac{b^2}{2}k^2 + \int dx'\,W_0(x')\Big(e^{ikx'}-1-ikx'\Theta(1-x'^2)\Big)\right)\right], \tag{12.121}$$

(Lévy–Khintchine representation)

where the Heaviside function $\Theta(1-x^2)$ prevents this term from acting outside $|x'| < 1$. Again, this prevents an infinite drift, and any non-problematic component of this part of the integral can just be absorbed into the existing drift a.[16] Note that we don't need to extend this excision procedure to higher-order terms, as we have already assumed the integral of $x^2W_0(x)$ in any finite interval around zero to converge—an assumption that amounted to a finite value of β^2. Since we have thus cured any problematic part of the integral at the origin, we no longer write it as a principal-value integral.

The form (12.121) is an important general representation for the characterstic function, and it is called the **Lévy–Khintchine representation** of the characteristic function.

12.5.5 Lévy–Itō Decomposition

The importance of the (12.121) Lévy–Khintchine representation is that it shows that an arbitrary Lévy process may be decomposed into other, simple processes. First, the $ik\alpha$ component represents a drift with

[16] For a more detailed form of this argument, see C. Gardiner, *op. cit.*, Sections 10.3.5-6, pp. 251-2.

rate α. The rest of the characteristic function is the product of the characteristic function for a Gaussian process [$\exp(-\beta^2 k^2/2)$] and the characteristic function for a compound Poisson process. In fact, the Poisson-process factor should be considered as a product of two Poisson factors, corresponding to the $|x'| < 1$ and $|x'| \geq 1$ parts of the integral.

Since the characteristic function for a sum of a random variables is the product of the individual characteristic functions, this means that we can write a Lévy process as

$$L(t) = at + bW(t) + N_{\text{C}}(t) + M_{\text{C}}(t),$$

(Lévy–Itō decomposition) (12.122)

or that is, the sum of a drift, scaled Gaussian process, and two compound Poisson processes. This separation of a Lévy process into component parts is called the **Lévy–Itō decomposition**. The process $N_{\text{C}}(t)$ is a compound Poisson process with jump distribution $W_0(x)\,[1 - \Theta(1 - x^2)]$, representing long jumps (of size larger than 1), while the process $M_{\text{C}}(t)$ is a *compensated* compound Poisson process with a distribution of short (size less than 1) jumps $W_0(x)\,\Theta(1-x^2)$. We already discussed the possibility that $W_0(x)$ could diverge at $x = 0$, provided the divergence is not too extreme. This corresponds to an arbitrarily large rate of jumps as the jump size goes to zero. Such a process is said to have **infinite activity**.

Because of this decomposition, Lévy processes are commonly characterized by the **Lévy triplet** or **characteristic triplet** (b^2, W_0, a), although of course conventions for the order of the triplet abound.[17] The drift rate a and diffusion rate b^2 are both constant scalars (or constant drift vector and diffusion tensor in the case of a multidimensional process), and the transition density or **Lévy measure** $W_0(x)$ characterizes the jump components of the process.

The transition density corresponding to the α-stable densities is just Eq. (12.49), except without the caveat of being an asymptotic expression:

$$W_0(x) = \frac{c_+\Theta(x)}{|x|^{\alpha+1}} + \frac{c_-\Theta(-x)}{|x|^{\alpha+1}}.$$

(jump distriubution for α-stable Lévy processes) (12.123)

Note that the integral of $x^2 W_0(x)$ converges over $|x| < 1$, as required, and that all α-stable processes other than $\alpha = 2$ have infinite activity. The argument for this transition density leading to the characteristic function for the α-stable densities is essentially that of Section 12.3 (corresponding to the limit where $d \longrightarrow 0$). Note that Eqs. (12.52) still hold to define β and c in terms of $c\pm$, and c in turn defines σ via Eq. (12.45). Thus, for example, the explicit jump distribution is

$$W_0(x) = \frac{\sin(\pi\alpha/2)\Gamma(\alpha+1)\sigma^\alpha}{\pi|x|^{\alpha+1}} \tag{12.124}$$

in the case of a stable process. Equation (12.54) still defines the drift μ for $1 < \alpha < 2$, but with $\mu' = 0$ and $d = 1$ (note that if we let $d \longrightarrow 0$, the drift would diverge).

12.6 Multiple Dimensions

Thus far, we have restricted our discussion of Lévy processes to one dimension, but of course the generalizations to multiple dimensions are of equal utility.

12.6.1 Characteristic Function

12.6.2 Fractional Laplacian

The Laplacian operator $\triangle \equiv \nabla^2$ is diagonal in the momentum representation, because we have

$$\triangle\, e^{i\mathbf{k}\cdot\mathbf{r}} = -k^2\, e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{12.125}$$

[17] The convention here follows Rama Cont and Peter Tankov, *op. cit.*, p. 80.

Then in analogy to the fractional derivative (12.87), we can define a fraction $\alpha/2$ of the Laplacian by

$$-(-\triangle)^{\alpha/2}\, e^{i\mathbf{k}\cdot\mathbf{r}} := \left(-|k|^\alpha\right) e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{12.126}$$

for $\alpha \in (0, 2]$. Again, this operator is defined in terms of a power of $-\triangle$ because it is positive definite. As in one dimension, the case $\alpha = 2$ recovers the standard Laplacian, and fractional derivatives outside this range can be defined by the recursion $-(-\triangle)^{(\alpha+2)/2} = -(-\triangle)^{\alpha/2}\triangle$.

To work out an expression for the fractional Laplacian in $\boldsymbol{d}$ dimensions, we should first lay out the explicit Fourier-transform convention

$$\tilde{f}(\mathbf{k}) = \int d^{\boldsymbol{d}}r\, f(\mathbf{r})\, e^{-i\mathbf{k}\cdot\mathbf{r}}, \qquad f(\mathbf{r}) = \frac{1}{(2\pi)^{\boldsymbol{d}}} \int d^{\boldsymbol{d}}k\, \tilde{f}(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{12.127}$$

The convolution theorem (2.14) still reads

$$\mathscr{F}[f * g] = \mathscr{F}[f]\mathscr{F}[g]. \tag{12.128}$$

in this convention. Now the action of the fractional derivative on a general function $f(\mathbf{r})$ is given by integrating to complete the Fourier transform of Eq. (12.126):

$$-(-\triangle)^{\alpha/2}\, f(\mathbf{r}) := -\frac{1}{(2\pi)^{\boldsymbol{d}/2}} \int d^{\boldsymbol{d}}k\, \tilde{f}(\mathbf{k})\, (-\triangle)^{\alpha/2}\, e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{1}{(2\pi)^{\boldsymbol{d}/2}} \int d^{\boldsymbol{d}}k \left(-|k|^\alpha\right) \tilde{f}(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{12.129}$$

The aim is now to employ the convolution theorem (12.128) to invert the Fourier integral on the right-hand side.

To proceed, we will need to find the Fourier counterpart to $|k|^\alpha$, which generalizes the transform (12.84) in one dimension. However, directly approaching the inverse transform in $\boldsymbol{d}$ dimensions is not a simple integral, because $|k|^\alpha$ doesn't factorize into different dimensions (unless $\alpha = 2$), and the integration in spherical coordinates is cumbersome beyond $\boldsymbol{d} = 3$. The trick to work out this transform starts with the integral representation of the gamma function:

$$\Gamma(z) = \int_0^\infty ds\, s^{z-1}\, e^{-s}. \tag{12.130}$$

Changing the integration variable to $s = r^2\lambda^2$ gives

$$\Gamma(\gamma/2) = 2\int_0^\infty d\lambda\, r^\gamma \lambda^{\gamma-1}\, e^{-\lambda^2 r^2}, \tag{12.131}$$

and then setting $z = \gamma/2$ gives the integral formula

$$\frac{1}{r^\gamma} = \frac{2}{\Gamma(\gamma/2)} \int_0^\infty d\lambda\, \lambda^{\gamma-1}\, e^{-\lambda^2 r^2}. \tag{12.132}$$

The advantage of this expression is that the dependence on $\mathbf{r}$ is Gaussian on the right-hand side, and is thus a separable function in $\boldsymbol{d}$ dimensions. The Fourier transform of this formula reads

$$\int d^{\boldsymbol{d}}r\, \frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{r^\gamma} = \frac{2}{\Gamma(\gamma/2)} \int_0^\infty d\lambda\, \lambda^{\gamma-1} \int d^{\boldsymbol{d}}r\; e^{-\lambda^2 r^2}\, e^{-i\mathbf{k}\cdot\mathbf{r}}. \tag{12.133}$$

This integral can be factored into a product of $\boldsymbol{d}$ Gaussian integrals in one dimension; carrying these out

gives

$$\begin{aligned}\int d^d r\, \frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{r^\gamma} &= \frac{2}{\Gamma(\gamma/2)} \int_0^\infty d\lambda\, \lambda^{\gamma-1} \frac{1}{(2\lambda^2)^{d/2}}\, e^{-k^2/4\lambda^2} \\ &= \frac{2^{1-d/2}}{\Gamma(\gamma/2)} \int_0^\infty d\lambda\, \lambda^{\gamma-1-d}\, e^{-k^2/4\lambda^2} \\ &= \frac{2^{1-d/2}}{\Gamma(\gamma/2)} \int_0^\infty d\xi\, \xi^{d-1-\gamma}\, e^{-k^2\xi^2/4} \\ &= \frac{2^{1-d/2}}{\Gamma(\gamma/2)} 2^{d-1-\gamma} k^{\gamma-d}\, \Gamma\left(\frac{d-\gamma}{2}\right) \\ &= \frac{2^{d/2-\gamma}\Gamma[(d-\gamma)/2]}{\Gamma(\gamma/2)}\, k^{\gamma-d},\end{aligned} \tag{12.134}$$

after changing the integration variable to $\xi = 1/\lambda$ (so that $d\lambda = -d\xi/\xi^2$) to obtain a Gaussian-moment integral. The ξ integration requires $d - \gamma > 0$. Setting $\gamma = \alpha + d$ (so that $\alpha < 0$), we obtain

$$\frac{1}{(2\pi)^{d/2}} \int d^d r\, \frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{r^{\alpha+d}} = \frac{2^{-d/2-\alpha}\Gamma(-\alpha/2)}{\Gamma[(\alpha+d)/2]}\, k^\alpha. \tag{12.135}$$

Inverting the transform gives

$$\frac{2^{d/2+\alpha}\Gamma[(\alpha+d)/2]}{\Gamma(-\alpha/2)} \frac{1}{r^{\alpha+d}} = \frac{1}{(2\pi)^{d/2}} \int d^d k\; k^\alpha\, e^{i\mathbf{k}\cdot\mathbf{r}} \tag{12.136}$$

The range of allowed α is then $-d - 1 < \alpha < 0$, with the lower limit needed to guard against a divergence in the Fourier integrals at $\mathbf{r} = 0$ or $\mathbf{k} \longrightarrow \infty$.

Finally, we can apply the convolution theorem, which amounts to plugging the inverse transforms into the convolution integral. The result is

$$-(-\triangle)^{\alpha/2}\, f(\mathbf{r}) = -\frac{2^\alpha \Gamma[(\alpha+d)/2]}{\pi^{d/2}\Gamma(-\alpha/2)} \int d^d r'\, \frac{f(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^{\alpha+d}} = \frac{2^\alpha \Gamma[(\alpha+d)/2]}{\pi^{d/2}|\Gamma(-\alpha/2)|} \int d^d r'\, \frac{f(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^{\alpha+d}}. \tag{12.137}$$

The last expression is written for the region $\alpha \in (0, 2)$, where the modulus of $\Gamma(-\alpha/2)$ acts because $\Gamma(x) < 0$ for $x \in (-1, 0)$.

12.6.2.1 Cauchy Principal Value and Hadamard Finite Part

The integral expression (12.137) is divergent for the parameters of interest, $\alpha \in (0, 2)$ and $d \geq 1$. And so we will discuss common method for regularizing such integrals. The more famous regularization deals with $1/x$ singularities. Suppose that $f(x)$ is smooth and finite but nonzero at $x = 0$; then one definition of the **Cauchy principal value** integral of $f(x)/x$ is

$$\fint_{-\infty}^{\infty} \frac{f(x)}{x}\, dx := \lim_{\epsilon\to 0^+} \left[\int_\epsilon^\infty \frac{f(x)}{x}\, dx + \int_{-\infty}^{-\epsilon} \frac{f(x)}{x}\, dx\right] = \lim_{\epsilon\to 0^+} \int_{\mathbb{R}-(-\epsilon,\epsilon)} \frac{f(x)}{x}\, dx. \tag{12.138}$$

(Cauchy principal value)

Essentially, the idea is to excise the symmetric neighborhood $(-\epsilon, \epsilon)$ around the problematic singular point from the range of the integral, and then compute the limit as this region shrinks symmetrically to zero size. So for example, if $f(x) = 1$, then the Cauchy principal value integral gives a zero result, because the counterparts of opposite sign of $1/x$ on either size of $x = 0$ cancel in this definition. It is a kind of removable singularity of the integral, because there is cancellation of the singular part that must be allowed to happen in just the right way. (Note that for other singularities that don't have this symmetry, it may be appropriate to define the excision region in an asymmetric way.)

For other singularities, as for example a $1/x^2$ singularity, such a cancellation can't be relied upon, and other means of regularization are necessary. A more general approach consists of excising the problematic

region as before, and analyzing the behavior of the resulting integral as $\epsilon \longrightarrow 0$. More concretely, suppose $g(x)$ is a function that is singular at $x = 0$; then excising a neighborhood of the singularity in the integral, the result of the integral can be decomposed as

$$\int_{\mathbb{R}-(-\epsilon,\epsilon)} g(x)\,dx = G_0(\epsilon) + G_\infty(\epsilon), \tag{12.139}$$

where $G_\infty(\epsilon)$ diverges as $\epsilon \longrightarrow 0$, and where $G_0(\epsilon)$ remains finite. Then the **Hadamard finite part** of the integral is defined as the limit of the finite part of the integral:

$$⨍_{-\infty}^{\infty} g(x)\,dx := \lim_{\epsilon\to 0} G_0(\epsilon). \tag{12.140}$$
(Hadamard finite part)

Obviously, the decomposition (12.139) must be performed in some unique way for this definition to be valid. For example, half of the finite part could be lumped in to the divergent part, which would still be divergent, so the divergent part must be defined carefully to be the ‘‘minimal’’ divergent part.

As a first example, suppose that $f(x)$ is a smooth function that decays to zero at large $|x|$ sufficiently rapidly that any boundary terms at infinity from integration by parts may be discarded. This will allow us to treat a prototype for handling a $1/x^2$ singularity, where the excised form of the integral reads

$$\int_{\mathbb{R}-(-\epsilon,\epsilon)} \frac{f(x)}{x^2}\,dx = \int_{\mathbb{R}-(-\epsilon,\epsilon)} \frac{f'(x)}{x}\,dx + \frac{f(\epsilon)+f(-\epsilon)}{\epsilon} \tag{12.141}$$

after integration by parts, because the boundary terms at infinity vanish by assumption (only the boundary terms associated with the excised region remain). This is a clear decomposition into finite and diverging parts as $\epsilon \longrightarrow 0$; the first integral simply becomes a Cauchy principal value in this limit, so that

$$⨍_{-\infty}^{\infty} \frac{f(x)}{x^2}\,dx = \fint_{-\infty}^{\infty} \frac{f'(x)}{x}\,dx. \tag{12.142}$$

In fact, the Cauchy principal value integral (12.138) has the form of a Hadamard-type regularization, where the divergent part is always zero (i.e., the Hadamard finite part is identical to the Cauchy principal value whenever it is defined), so that we could equally well write

$$⨍_{-\infty}^{\infty} \frac{f(x)}{x^2}\,dx = ⨍_{-\infty}^{\infty} \frac{f'(x)}{x}\,dx, \tag{12.143}$$

and more generally, integration by parts works under the Hadamard finite part via

$$⨍_{-\infty}^{\infty} \frac{f(x)}{x^n}\,dx = \frac{1}{(n-1)} ⨍_{-\infty}^{\infty} \frac{f'(x)}{x^{n-1}}\,dx \tag{12.144}$$

for integer powers $n \neq -1$. This idea generalizes naturally for noninteger powers $a \neq 1$,

$$⨍_{-\infty}^{\infty} \frac{f(x)}{|x|^a}\,dx = \frac{1}{(a-1)} ⨍_{-\infty}^{\infty} \frac{x f'(x)}{|x|^a}\,dx, \tag{12.145}$$

and in the case of $a = 1$,

$$\fint_{-\infty}^{\infty} \frac{f(x)}{x}\,dx = ⨍_{-\infty}^{\infty} \frac{f(x)}{x}\,dx = ⨍_{-\infty}^{\infty} \log|x|\, f'(x)\,dx = \int_{-\infty}^{\infty} \log|x|\, f'(x)\,dx. \tag{12.146}$$

where terms logarithmic in ϵ are discarded (and uniqueness is trickier, since for example[18] the difference between $\log\epsilon$ and $\log 2\epsilon$ is an additive constant).

[18] Lars Hörmander, *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis* (Springer-Verlag, 1983), Section 3.2 (ISBN: 9783642614972) (doi: 10.1007/978-3-642-61497-2).

Another way to interpret Eq. (12.141) is that the Hadamard finite part with the $1/x^2$ singularity can be expressed by subtracting the divergent boundary terms from the excised integral:

$$⨍_{-\infty}^{\infty} \frac{f(x)}{x^2}\,dx = \lim_{\epsilon\to 0}\left[\int_{\mathbb{R}-(-\epsilon,\epsilon)} \frac{f(x)}{x^2}\,dx - \frac{2f(0)}{\epsilon}\right]. \tag{12.147}$$

Re-expressing the boundary terms as an integral,

$$⨍_{-\infty}^{\infty} \frac{f(x)}{x^2}\,dx = \lim_{\epsilon\to 0}\left[\int_{\mathbb{R}-(-\epsilon,\epsilon)} \frac{f(x)}{x^2}\,dx - \int_{\mathbb{R}-(-\epsilon,\epsilon)} \frac{f(0)}{x^2}\,dx\right], \tag{12.148}$$

and thus the Hadamard finite part

$$⨍_{-\infty}^{\infty} \frac{f(x)}{x^2}\,dx = \int_{-\infty}^{\infty} \frac{f(x)-f(0)}{x^2}\,dx \tag{12.149}$$

as an alternate expression for the regularized form of the integral. This expression is of course equivalent to Eq. (12.142), as can be seen by integration by parts.

In the case of a singularity of a general power-law exponent a, the regularization (12.149) also carries through as

$$⨍_{-\infty}^{\infty} \frac{f(x)}{|x|^a}\,dx = ⨍_{-\infty}^{\infty} \frac{f(x)-f(0)}{|x|^a}\,dx, \tag{12.150}$$

although the interpretation of the right-hand side depends on the exact value of a. On the left-hand side, the $1/|x|^a$ singularity is no problem (i.e., absolutely integrable) if $a < 1$; such a kernel is called a **weakly singular kernel**. In the case $a \geq 1$ (and $1/x^a$ for $a > 1$), there is definitely a problem with the integral diverging unless some regularization like the Hadamard finite part is employed. Such a kernel is called a **hypersingular kernel**. (In d dimensions, a hypersingular kernel has $a > d$.) On the right-hand side of Eq. (12.150), the numerator has the form $xf'(0)$ to leading order, which tempers the singularity. Thus the integral is absolutely integrable if $a < 2$ (in which case the double lines on the integral may be dropped), and the integral makes sense in terms of a Cauchy principal value if $a = 2$. In fact, for $a < 3$, the integral can still be sensibly defined in terms of a Cauchy principal value, since the same cancelation works: Thinking of $[f(x)-f(0)]/|x|^a = f'(0)\,x/|x|^a + f''(0)\,x^2/(2|x|^a) + \cdots$, the first term vanishes under the Cauchy principal value, while the second and higher terms are absolutely integrable. For $a \geq 3$, further regularization is needed (by subtracting $xf'(0)$, $x^2f''(0)/2$, $x^3f^{(3)}(0)/3!$, and so on) until the integral is absolutely integrable, and then the double lines may be dropped.

The Hadamard finite part may seem somewhat suspect, as it seems like just a method of defining away the problematic part of an integral. This is, of course, true, but it's important to note why this procedure is sensible. Physical calculations are supposed to begin with some sensible, well defined expression. In the case of the fractional Laplacian, this was the Fourier transform (12.129), which is well defined if the Fourier spectrum $\tilde{f}(\mathbf{k})$ decays sufficiently quickly with k, even in the range of interesting α. However, the resulting integral expression (12.137), obtained through standard manipulations, is divergent in the range of interesting α (although it is okay for other ranges of α). The divergence here is then spurious. The regularization gives a way of associating the ill-defined integral expressions with well-defined ones, to facilitate arriving at a useful answer.

12.6.2.2 Regularization of the Fractional Laplacian

The integral expression (12.137) is divergent for $\alpha \in (0,2)$ in the same way as is commonly regularized as the one-dimensional integral (12.150), and the same regularization applies. Thus, the integral in Eq. (12.150) should be interpreted in terms of the Hadamard finite part, and the same regularization in Eq. (12.137) applies:

$$-(-\triangle)^{\alpha/2}\,f(\mathbf{r}) = \frac{2^\alpha\Gamma[(\alpha+d)/2]}{\pi^{d/2}|\Gamma(-\alpha/2)|} ⨍ d^dr'\,\frac{f(\mathbf{r}')-f(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|^{\alpha+d}}. \tag{12.151}$$

(regularized fraction Laplacian)

Again, this integral is absolutely integrable for $\alpha \in (0,1)$, but should still be interpreted as a Cauchy principal value for $\alpha \in [1,2)$. In one dimension, the rules (12.145) and (12.146) reproduce the explicitly regularized integrals in Eq. (12.93). It is exactly these rules—that the Hadamard finite part preserves the rules of integration by parts—which justify the manipulations leading up to those one-dimensional expressions (see also Problem 12.3).

12.7 Stochastic Calculus for Stable Lévy Processes

The stable processes, and Lévy processes, represent a generalized form of noise. Just as in the Gaussian case in Section 3.3, stochastic differential equations may be driven by Lévy noise.[19] For example, we could consider an SDE of the form

$$dy = a(y,t)\,dt + b(y,t)\,dS_\alpha, \tag{12.152}$$

in analogy to the Itō SDE (3.24) driven by dW. In this case dS_α is an α-stable Lévy increment with probability density of the form (12.9), with the replacements $t \longrightarrow dt$ and $\sigma \longrightarrow 1$. (To keep things reasonably simple we'll assume $\beta = 0$.)

A simple example of a linear SDE has a linear damping action driven by additive noise,

$$dy = -\gamma y\,dt + dS_\alpha, \tag{12.153}$$

in a generalization of the Ornstein–Uhlenbeck process (3.73). Because this equation is linear, we can simply solve the SDE as before by rewriting it in terms of $ye^{\gamma t}$, and then integrating from 0 to t. The result is

$$y(t) = y_0\,e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} dS_\alpha(t'). \tag{12.154}$$

The generalization of the increment addition rule (12.13),

$$\sigma = \left(\sigma_1^\alpha + \sigma_2^\alpha\right)^{1/\alpha}, \tag{12.155}$$

to the present solution says that we can think of the noisy integral in Eq. (12.154) as an α-stable random variable with density of the form (12.9) at time t, and with

$$\sigma(t) = \left[\int_0^t e^{-\alpha\gamma(t-t')} dS_\alpha(t')\right]^{1/\alpha}. \tag{12.156}$$

Of course, this is also true for the Gaussian case. Like the Gaussian case, it is simple to characterize the average solution,

$$\left\langle\!\left\langle y(t) \right\rangle\!\right\rangle = y_0\,e^{-\gamma t}, \tag{12.157}$$

because the probability density of the stochastic term in the solution (12.154) is symmetric. However, characterizing the correlation function is problematic if $\alpha < 2$, because $\left\langle dS_\alpha^2 \right\rangle$ diverges. Furthermore, nonlinear transformations of Lévy-drive SDEs are tricky for the same reason: the Itō rule $dW^2 = dt$ breaks down for stable Lévy noise.

12.7.1 Chain Rule for General Lévy Processes

To generalize the Itō chain rule (3.118) to the case of Lévy processes, consider a general Lévy-noise-driven SDE of the form[20]

$$dy = a(y,t)\,dt + b(y,t)\,dW + dN_{\mathrm{C}}, \tag{12.158}$$

[19] Here we are following Kurt Jacobs, *Stochastic Processes for Physicists: Understanding Noisy Systems* (Cambridge, 2010), Section 9.2.1 (ISBN: 9780511815980) (doi: 10.1017/CBO9780511815980).

[20] Again we are following Kurt Jacobs, *op. cit.*, Section 9.4, and Rama Cont and Peter Tankov, *op. cit.*, Section 8.3.2.

where $N_{\rm C}(t)$ is a compound Poisson process. That the Lévy-noise driving can be written as a combination of dt, dW, and $dN_{\rm C}$ follows from the Lévy–Itō decomposition (12.122), though for the moment we will assume $N_{\rm C}(t)$ to have only finite activity, and return to the infinite-activity case later. Note that according to the definition of Lévy processes, $y(t)$ is not itself a Lévy process unless $a(y,t)$ and $b(y,t)$ are constants.

Since we are assuming that the jump process does not have infinite activity, we can assume jumps occur at discrete times t_j, with jump sizes $\Delta N_j := \Delta N_{\rm C}(t_j)$. We can then consider the evolution *between* jumps, according to the modification

$$dy_{\rm c} = a(y,t)\,dt + b(y,t)\,dW, \tag{12.159}$$

which is Eq. (12.158) except for the jump term. The increments dy and $dy_{\rm c}$ are equivalent between jumps, but the chain rule for $z = f(y)$ is more straightforward for $dy_{\rm c}$, being just the Itō rule (3.118):

$$\begin{aligned} dz = df(y_{\rm c}) &= f'(y)\,dy + \frac{1}{2}f''(y)(dy)^2 \\ &= \left[f'(y)\,a + \frac{1}{2}f''(y)\,b^2\right]dt + f'(y)\,b\,dW \\ &= f'(y)\,dy_{\rm c} + \frac{b^2}{2}f''(y)\,dt. \end{aligned} \tag{12.160}$$

At the jump times, the evolution is discontinuous,

$$dz(t_j) = z(y_j + \Delta N_j) - z(y_j) = f(y_j + \Delta N_j) - f(y_j), \tag{12.161}$$

where y_j is $y(t)$ just before the jump at t_j. Then the *total* evolution combines the continuous and discontinuous evolutions (12.160) and (12.161), respectively:

$$dz(t) = f'(y)\,dy_{\rm c} + \frac{b^2}{2}f''(y)\,dt + \Big(z[y(t)] - z[y(t^-)]\Big). \tag{12.162}$$

The last term here represents the possible jump discontinuity; remember that the notation is consistent with the cadlag assumption, where $y(t) = \lim_{t\to 0^+} y(t)$, and $y(t^-) \equiv y(t-0^+)$ is the state of the Lévy process just before the jump discontinuity at time t. Now remember that $y_{\rm c}$ was introduced as a sort of crutch to let us apply the Itō chain rule. So, we should eliminate it. Comparing Eqs. (12.158) and (12.159) gives $dy_{\rm c} = dy - dN_{\rm C}$. Then remembering that at the discontinuity times t_j of $N_{\rm C}(t)$, $dN_{\rm C}(t)$ acts like $\Delta N_j\,\delta(t-t_j)\,dt$, so that

$$dz(t) = f'[y(t^-)]\,dy(t) + \frac{b^2}{2}f''[y(t)]\,dt + \Big(f[y(t)] - f[y(t^-)] - f'[y(t^-)]\,\Delta y(t)\Big), \tag{12.163}$$

(Lévy-process chain rule)

where $\Delta y(t) = y(t) - y(t^-)$ is a possible discontinuity if t is one of the discontinuity times t_j. Clearly, the last term vanishes in a time interval where no jump occurs. Also note the time argument in the first term is t^- and not t, due to the transformation away from $y_{\rm c}$ (whose presence denoted the ignoring of the jumps).

Now it turns out that the chain-rule formula (12.162) is also true if the jump component of the Lévy process is infinite activity (as in the α-stable Lévy processes). This is not obvious, because in any finite time interval, the last term in (12.162) must be summed over an infinity of discontinuities, which may or may not converge. Again referring to the Lévy–Itō decomposition, we can separate the jump process into a component with long, discrete jumps, and an infinite-activity component of short jumps. The dividing line in terms of the jump magnitude between these two components is arbitrary; let's call it ϵ, with the idea that ϵ will be nonzero but small. In this case, let's return to the last term in (12.162). We've already handled this for the long, discrete jumps, so let's handle it in the infinite-activity case. Since the jumps Δy are small, and assuming $f'(y)$ is bounded,

$$f[y(t)] - f[y(t^-)] - f'[y(t^-)]\,\Delta y(t) = O(\Delta y^2) = O(\epsilon^2). \tag{12.164}$$

Thus, the integral of this quantity over a finite integral converges absolutely (note that we don't need to worry about multiple jumps in any short time interval dt because of the assumption of stochastic continuity).

Furthermore, this bound says that we can define a modification $y_\epsilon(t)$ to $y(t)$ by ignoring the infinite-activity component of the jumps smaller in magnitude than ϵ. Then $f(y_\epsilon)$ differs from $f(y)$ by $O(\epsilon^2)$. As $\epsilon \longrightarrow 0$, then, $f(y_\epsilon) \longrightarrow f(y)$. But the chain rule (12.162) applies to $f(y_\epsilon)$ for any $\epsilon > 0$, so it applies also to $f(y)$, so it works just fine in the infinite-activity case.

12.7.1.1 Example: Linear Damping

Returning to the example of the linearly damped SDE (12.153), suppose that we consider a linear, multiplicative-noise version, driven by a general Lévy process $L(t)$,[21]

$$dy = -\gamma y\, dt + y\, dL, \tag{12.165}$$

where we can decompose $dL(t)$ as

$$dL = a\, dt + b\, dW + dN_{\mathrm{C}}, \tag{12.166}$$

with dN_{C} a compound Poisson process. To solve this equation, consider the transformation

$$z = \log y. \tag{12.167}$$

Using the chain rule (12.163), we have

$$\begin{aligned} dz &= \frac{dy}{y} - \frac{b^2}{2}\, dt + \log[y(t)] - \log[y(t^-)] - \frac{\Delta y(t)}{y(t^-)} \\ &= -\gamma\, dt + dL - \frac{b^2}{2}\, dt + \log\left(1 + \frac{\Delta y(t)}{y(t^-)}\right) - \frac{\Delta y(t)}{y(t^-)}, \end{aligned} \tag{12.168}$$

where again $\Delta y(t) = y(t^-)\, dN_{\mathrm{C}}(t)$ is the (possible) jump at time t. Thus,

$$dz = -\left(\gamma + \frac{b^2}{2}\right) dt + dL + \log(1 + dN_{\mathrm{C}}) - dN_{\mathrm{C}}. \tag{12.169}$$

Integration from 0 to t is now straightforward:

$$z(t) = z(0) - \left(\gamma + \frac{b^2}{2}\right) t + L(t) - L(0) - N_{\mathrm{C}}(t) + N_{\mathrm{C}}(0) + \sum_j \log\Big(1 + \Delta N_{\mathrm{C}}(t_j)\Big). \tag{12.170}$$

Finally, transforming back to $y(t)$ amounts to exponentiating everything, and we obtain

$$y(t) = e^{z(t)} = y(0) \exp\left[-\left(\gamma + \frac{b^2}{2}\right) t + \big[L(t) - L(0)\big] - \big[N_{\mathrm{C}}(t) - N_{\mathrm{C}}(0)\big]\right] \prod_j \Big(1 + \Delta N_{\mathrm{C}}(t_j)\Big) \tag{12.171}$$

as the final solution.

12.7.2 Stochastic Exponential

Geometric Brownian motion (Section 4.1) is a Gaussian model for the stochastic price of something like a stock or other financial instrument. In terms of a Wiener process $W(t)$, the solution is exponentiated as

$$S(t) = S_0 e^{(\mu - \sigma^2/2)t + \sigma W(t)}, \tag{12.172}$$
(geometric Brownian motion)

where the constant parameter μ is a steady growth (representing inflation or profit), and the constant parameter σ is a stochasticity parameter modeling volatility. It is the solution of the Itō SDE

$$dS = \mu S\, dt + \sigma S\, dW. \tag{12.173}$$
(SDE for geometric Brownian motion)

[21] Kurt Jacobs, *op. cit.*, Section 9.4.1.

However, the major lesson from extreme financial events like stock-market crashes is that Gaussian models are inadequate in accounting for such events. Processes involving large jumps—such as Lévy processes—are needed instead.

Then how can we generalize geometric Brownian motion to a Lévy process? A naïve attempt would simply write down a modified Black–Scholes SDE (12.173), driving it with Lévy noise instead of Wiener noise:

$$dS = S\,dL. \tag{12.174}$$

For the Lévy process, we will assume the form

$$dL = \mu\,dt + \sigma\,dW + dN_{\mathrm{C}}, \tag{12.175}$$

or that is, a process with Lévy triple $(\sigma^2, W_0, \mu+\sigma^2/2)$. To solve this SDE, we can simply adapt the solution (12.171) with $\gamma = 0$, $b = \sigma$, $L(0) = N_{\mathrm{C}}(0) = 0$ to write

$$S(t) = S(0)\exp\left[L(t) - \frac{\sigma^2}{2}t - N_{\mathrm{C}}(t)\right]\prod_j \Big(1 + \Delta N_{\mathrm{C}}(t_j)\Big). \tag{12.176}$$

Additionally, we can write the discrete jumps $\Delta N_{\mathrm{C}}(t)$ as $\Delta S(t)$, and then identify the jump process $N_{\mathrm{C}}(t)$ with $S(t)$ as the sum of the jumps, to arrive at the solution

$$S(t) = S(0)\,e^{L(t)-\sigma^2 t/2}\prod_j \Big(1+\Delta S(t_j)\Big)\,e^{-\Delta S(t_j)} = S(0)\,e^{(\mu-\sigma^2/2)t+\sigma W(t)}\prod_j \Big(1+\Delta S(t_j)\Big).$$

(Doléans–Dade exponential) (12.177)

This result is called the **Doléans-Dade exponential**[22] or **stochastic exponential**. Note that the last expression is technically only valid when the product does not diverge; it turns out that a sufficient condition for this is for $\int_{-1}^{1} |x|\,W_0(x)\,dx$ to be finite for the jump process. Now a really inconvenient feature of this solution should be clear. If the jumps ΔS are unbounded (as they are for the stable processes), then $1+\Delta S$ can be negative for some jump. This means that this kind of "exponentiation" of a process can lead to negative solutions! For mathematical modeling (as in the stock market, this is typically not considered to be a workable solution.[23]

12.7.2.1 Exponential of a Lévy Process

If the Doléans-Dade exponential admits negative solutions, what kind of model is more appropriate? Consider a generalized Lévy process

$$dL(t) = \mu(t)\,dt + \sigma(t)\,dW(t) + dN_{\mathrm{C}}(t) \tag{12.178}$$

(which is not a Lévy process of course unless μ and σ are constant). Then consider the exponentiated form

$$S(t) = S_0 e^{L(t)} = S_0\exp\left[\int_0^t \mu(t')\,dt' + \int_0^t \sigma(t')\,dW(t') + \int_0^t dN_{\mathrm{C}}(t')\right], \tag{12.179}$$

which is now inherently positive. To connect this to an underlying SDE, we can apply the chain rule (12.163) for $z = e^y$ (or in this case $S = S_0 e^L$), with the result

$$dS(t) = S(t^-)\,dL(t) + \frac{\sigma^2(t)}{2}S(t)\,dt + S(t) - S(t^-) - S(t^-)\,\Delta L(t). \tag{12.180}$$

[22] C. Doléans-Dade, "Quelques applications de la formule de changement de variable pour les semi-martingales," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **16**, 181 (1970) (doi: 10.1007/BF00534595); C. Gardiner, *op. cit.*, Section 10.5.3, p. 259; Rama Cont and Peter Tankov, *op. cit.*, Section 8.4.2, p. 284.

[23] Note, however, that there is precedent for negative security prices; this happened to oil-futures prices during the COVID-19 pandemic due to low demand for oil and nowhere to store the excess. See Julia Horowitz, "American oil crashes below $0 a barrel – a record low" *CNN Business*, `https://web.archive.org/web/20200420193610/https://www.cnn.com/2020/04/20/investing/premarket-stocks-trading/index.html`

Now writing this in terms of $L(t)$ and its constituent parts, we obtain

$$\begin{aligned} dS(t) &= S(t^-)\,dL(t) + \frac{\sigma^2(t)}{2}S(t)\,dt + S(t^-)\Big(e^{\Delta L(t)} - 1 - \Delta L(t)\Big) \\ &= \left(\mu(t) + \frac{\sigma^2(t)}{2}\right) S(t)\,dt + S(t)\,\sigma(t)\,dW(t) + S(t^-)\,dN_{\mathrm{C}}(t) + S(t^-)\Big(e^{\Delta N_{\mathrm{C}}(t)} - 1 - \Delta N_{\mathrm{C}}(t)\Big) \end{aligned}$$

(SDE for exponentiated process) (12.181)

as a relatively unwieldy SDE corresponding to the solution (12.179). Although this seems to be a distinctly different animal compared to the Doléans-Dade exponential, it turns out that they give rise to the same class of stochastic processes, at least in the case where the underlying process $L(t)$ is a Lévy process. That is, given a *positive* Doléans-Dade exponential, there is a corresponding exponentiated Lévy process [with a different underlying $L(t)$], and vice versa.[24]

12.7.3 Continuity and Differentiability of the Stable Lévy Processes

With some tools in hand for handling stochastic calculus for Lévy processes, we can return to the question of continuity of the stable Lévy processes. In the case of infinite-activity Lévy processes, as in the stable Lévy processes, the divergent jump rate makes it somewhat tricky to discuss their continuity.[25] We have already discussed this idea fairly precisely in Section 12.4, but it is useful to take another, less formal look at continuity of the stable Lévy processes. Recall that the Wiener process (Section 3.1) is everywhere continuous but nondifferentiable (all of the statements here are, of course, ''almost surely'' or with unit probability). The Lévy–Khintchine representation (12.121), along with the jump transition density (12.123), says that the α-stable processes with $\alpha < 2$ are yet more pathological, because all increments are jumps that strictly cannot be thought of as Gaussian increments. The divergence of the transition density says that the stable processes are infinite activity, so that there is an infinity of jump discontinuities in any finite time interval. Of course, as the time interval shrinks, the typical size of the discontinuities shrinks. This is consistent with the fractal nature of the sample paths. However, the discontinuities are dense—at any given point in time, there is a jump discontinuity arbitrarily close by.

To make the above statements more specific, in any time step dt, from Eq. (12.25) the asymptotic probability of having a jump larger than x is

$$P[dS_\alpha(t) > x] \sim \frac{\Gamma(\alpha)\sin(\pi\alpha/2)}{\pi}\left(\frac{x}{\sigma}\right)^{-\alpha} dt. \tag{12.182}$$

Note that this statement is no longer asymptotic because of the infinitesimal time interval. Thus, there is a finite rate of occurrence for even an arbitrarily large jump discontinuity [except for $\alpha = 2$, for which $\sin(\pi\alpha/2) = 0$]. These large jump discontinuities are called **Lévy flights**.

In a short time Δt, the *typical* increment of an α-stable process is of order $\Delta x \sim \Delta t^{1/\alpha}$, according to the argument leading up to Eq. (12.16). This typical scale always shrinks to zero as $\Delta t \longrightarrow 0$, provided $\alpha > 0$. Then in an attempt to analyze the differentiability of a sample path, $\Delta x/\Delta t \sim \Delta t^{(1-\alpha)/\alpha}$, which only diverges for $\alpha > 1$ as $\Delta t \longrightarrow 0$. Thus, we have a rather strange picture of the sample paths. They are everywhere continuous, except on a dense subset of discontinuities if $\alpha < 2$. For $\alpha \leq 1$, the sample paths are furthermore everywhere differentiable (except on the dense set of discontinuities). This seems crazy, since these processes would seem to be more extreme in having longer-tailed paths, but the differentiability is a statement about the smallest increments, not the jumps way out in the tails.

But now we are stuck with sample paths that are continuous and possibly differentiable at almost every point, despite having discontinuities at an arbitrarily short time away. So it's worth considering some examples of (pathological) functions that exhibit some similar behavior. To give some examples:

- $\sin(1/x)$ is a classic example of a function that is continuous everywhere except at $x = 0$: the limit as $x \longrightarrow 0$ does not exist because of the arbitrarily rapid oscillations.

[24] Rama Cont and Peter Tankov, *op. cit.*, Section 8.4.3, p. 286.

[25] The discussion here follows the rather cheeky discussion in Vladimir V. Uchaikin and Vladimir M. Zolotarev, *op. cit.*, Section 17.2.

- Squaring off the first example, $\operatorname{sgn}[\sin(1/x)]$ is a function that is still discontinuous at $x = 0$, with other discontinuities arbitrarily close to $x = 0$ (in particular, at $x = \pm 1/n\pi$ for $n \in \mathbb{Z}$).
- Now tempering the previous example a little, $x \operatorname{sgn}[\sin(1/x)]$ is now *continuous* at $x = 0$, but has a discontinuity arbitrarily close to this same point. The function is also nondifferentiable at $x = 0$ (why?).
- Finally, $x^2 \operatorname{sgn}[\sin(1/x)]$ is both continuous and differentiable at $x = 0$, with a discontinuity arbitrarily close by.

12.8 Exercises

Problem 12.1

Show that the uniform probability distribution is not infinitely divisible.

Problem 12.2

Show that the gamma distribution (Section 10.8.3.4)

$$f(x;\gamma,\beta) = \frac{\beta^\gamma}{\Gamma[\gamma]}\, x^{\gamma-1} e^{-\beta x} \qquad (x>0,\ \gamma>0,\ \beta>0), \tag{12.183}$$

is infinitely divisible.

Problem 12.3

Show that the expression (12.151)

$$-(-\triangle)^{\alpha/2}\, f(\mathbf{r}) = \frac{2^\alpha \Gamma[(\alpha+d)/2]}{\pi^{d/2}|\Gamma(-\alpha/2)|} \int d^d r'\, \frac{f(\mathbf{r}') - f(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|^{\alpha+d}} \tag{12.184}$$

for the fractional Laplacian reduces for $d=1$ to the explicitly regular forms (12.93)

$$\frac{\partial^\alpha}{\partial|x|^\alpha} f(x) = \begin{cases} \dfrac{\Gamma(\alpha-1)\sin(\pi\alpha/2)}{\pi} \dfrac{d^2}{dx^2} \displaystyle\int_{-\infty}^{\infty} dx'\, \frac{f(x')}{|x-x'|^{\alpha-1}} & (1<\alpha<2) \\ \dfrac{1}{\pi}\dfrac{d^2}{dx^2} \displaystyle\int_{-\infty}^{\infty} dx'\, f(x') \log\frac{1}{|x-x'|} = \frac{1}{\pi}\frac{d}{dx} \fint_{-\infty}^{\infty} dx'\, \frac{f(x')}{x'-x} & (\alpha=1) \\ \dfrac{\Gamma(\alpha)\sin(\pi\alpha/2)}{\pi} \dfrac{d}{dx} \displaystyle\int_{-\infty}^{\infty} dx'\, \frac{f(x')\,\mathrm{sgn}(x'-x)}{|x-x'|^\alpha} & (0<\alpha<1), \end{cases} \tag{12.185}$$

observing the rules for the Hadamard finite part. (Note that commuting a derivative with an integral is allowed provided both possible expressions are well defined.)

Problem 12.4

Suppose that $X_1,\ldots,X_N$ are independent, identically distributed random variables chosen from an α-stable distribution with parameters β_1, σ_1, and μ_1. If we define the sample mean

$$M_N := \frac{X_1+\cdots+X_N}{N}, \tag{12.186}$$

then what is the distribution of M_N?

Chapter 13

Lévy Processes Crossing Boundaries

Boundary-crossing problems for Lévy processes are much more difficult to handle yet more subtle and interesting than the Gaussian counterparts that we studied in Chapter 6. Before discussing the *when* of the first passage, it will make things a little simpler to start with the *where*—a question that has much more ''depth'' than in the Gaussian case.

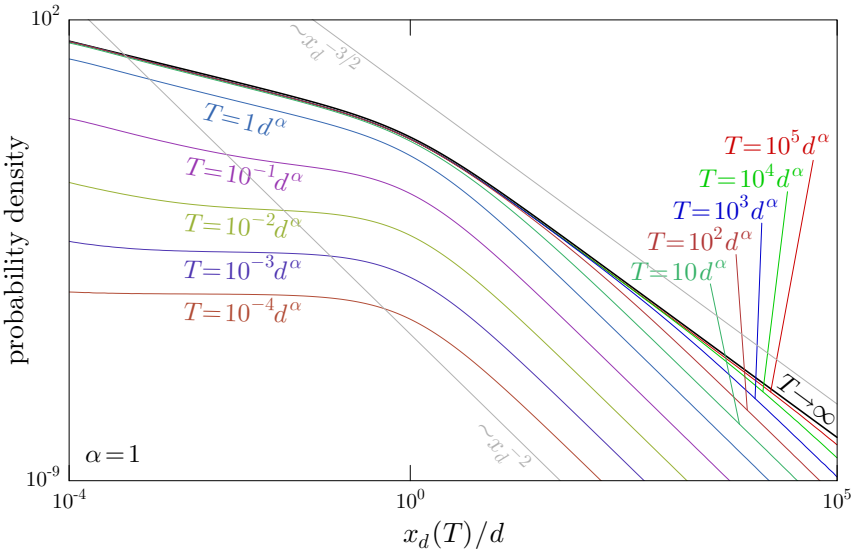
13.1 Leapover at First Passage

First-passage problems involving Lévy processes are somewhat different than their Gaussian counterparts. Gaussian, continuous-time stochastic processes are continuous in position, but Lévy processes are in general not. Thus, while a Gaussian process will cross continuously over a boundary, a Lévy process can jump discontinuously over it. Thus, the position of a Lévy process just after crossing a boundary $x = d$ (with $d > 0$) is called the **first-passage position**, or **leapover position** $x_d > 0$, and has a density in the case of symmetric α-stable Lévy processes of[1]

$$f_{x_d}(x) = \frac{\sin(\pi\alpha/2)}{\pi} \frac{d^{\alpha/2}}{x^{\alpha/2}(x+d)}, \qquad (13.1)$$
(leapover density)

to be derived below. Somewhat remarkably, this leapover density is independent independent of the width-rescaling parameter σ. This density corresponds to the case where we don't care about the *time* at which the first crossing occurred. However, the leapover distribution varies with time. The plot below shows the evolution of the leapover density for a Cauchy process ($\alpha = 1$), where first crossings are only counted up to a maximum time T. The $T \longrightarrow \infty$ curve corresponds to Eq. (13.1); the other curves are obtained from simulations.

[1] Daniel Ray, ''Stable Processes with an Absorbing Barrier,'' *Transactions of the American Mathematical Society* **89**, 16 (1958) (doi: 10.2307/1993130); Tal Koren, Michael A. Lomholt, Aleksei V. Chechkin, Joseph Klafter, and Ralf Metzler, ''Leapover Lengths and First Passage Time Statistics for Lévy Flights,'' *Physical Review Letters* **99**, 160602 (2007) (doi: 10.1103/PhysRevLett.99.160602); Bartłomiej Dybiec, Ewa Gudowska-Nowak, and Aleksei Chechkin, ''To hit or to pass it over—remarkable transient behavior of first arrivals and passages for Lévy flights in finite domains,'' *Journal of Physics A: Mathematical and Theoretical* **49**, 504001 (2016) (doi: 10.1088/1751-8113/49/50/504001).



The densities here are unnormalized, the integrated area giving the total fraction of trajectories that crossed the boundary by time T. The corresponding integrated crossing probabilities at each final time T are given in the table below.

T	probability
10^{-4}	0.003%
10^{-3}	0.03%
10^{-2}	0.3%
10^{-1}	3.4%
1	32.9%
10	79.0%
10^2	93.6%
10^3	98.0%
10^4	99.4%
10^5	99.8%

13.1.1 Leapover in the Escape from an Interval

In order to derive the leapover density above, we will consider a slightly different problem,[2] from which (13.1) will emerge as a limit. In particular, we will consider the escape of a symmetric, stable Lévy process from an interval. To simplify notation, we will consider a Lévy process starting at $x_0 \in (-1,1)$ at $t=0$, and then escaping from this interval. Also, we will use x_+ to denote the first-exit position (i.e., the leapover), which is the first position outside $(-1,1)$.

A useful way to think about the escape problem is in terms of perfectly absorbing boundary conditions, where absorption occurs everywhere outside $(-1,1)$. As we have seen, in view of the leapover effect, a Lévy process can end up anywhere in the absorbing region on its first exit, in contrast to Gaussian processes, which are absorbed at one of the boundary points at ± 1. To set up the escape problem, the fractional diffusion equation (12.81) says that in a time interval dt, the change in the density is given in terms of the fractional derivative as

$$f_\alpha(x,t+dt|x_0,0) = f_\alpha(x,t|x_0,0) + \sigma^\alpha\, dt\, \frac{\partial^\alpha}{\partial|x|^\alpha} f(x,t|x_0,0). \tag{13.2}$$

[2]The following derivation, including the inspired contour integration, follows that of Shinzō Watanabe (渡辺 信三), "On stable processes with boundary conditions," *Journal of the Mathematical Society of Japan* **14**, 170 (1962) (doi: 10.2969/jmsj/01420170); the multi-dimensional version of this escape problem was considered by R. M. Blumenthal, R. K. Getoor and D. B. Ray, "On the Distribution of First Hits for the Symmetric Stable Processes," *Transactions of the American Mathematical Society* **99**, 540 (1961) (doi: 10.1090/S0002-9947-1961-0126885-4).

where again $f(x,t|x_0,0)$ is the evolving probability density given the initial condition x_0 at $t=0$. Since the fractional derivative is nonlocal for $\alpha \in (0,2)$, this means that some of the probability density will be swept into the absorbing region, in which case it will vanish before the next time step dt. This density that is getting absorbed is exactly what we want. Let's denote the density of the first-passage position by $f_{x_+}(x,T)$ for $x_+ \notin (-1,1)$, which applies to all trajectories that have escaped $(-1,1)$ by time $T>0$—that is, the first-passage time $\tau_+ \in (0,T]$. What we will end up calculating is the density of the first-passage position for *any* escape time, which we will denote by

$$f_{x_+}(x) = \lim_{T\to\infty} f_{x_+}(x,T). \tag{13.3}$$

From Eq. (13.2) we can then infer that between times t and $t+dt$, the leapover density accumulates escaped trajectories according to

$$f_{x_+}(x,t+dt) = f_{x_+}(x,t) + \sigma^\alpha\, dt\, \frac{\partial^\alpha}{\partial|x|^\alpha} f_\alpha(x,t|x_0,0). \tag{13.4}$$

Integration up to time T then yields

$$f_{x_+}(x,T) = \sigma^\alpha \int_0^T dt\, \frac{\partial^\alpha}{\partial|x|^\alpha} f_\alpha(x,t|x_0,0). \tag{13.5}$$

Now for the fractional derivative, we can use Eq. (12.93). But noting that there is no problem with the singularity because we are only interested in the region where $f_\alpha(x,t|x_0,0)=0$, we can carry out the regular derivatives, which amounts to using the derivative-free expression (12.85) for the fractional Laplacian:

$$\frac{\partial^\alpha}{\partial|x|^\alpha} f(x) = \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi} \int_{-1}^1 dx'\, \frac{f(x')}{|x-x'|^{\alpha+1}}. \tag{13.6}$$

This is of course valid under the assumption that $f(x)=0$ for $|x|>1$. Putting this expression into Eq. (13.5) gives

$$f_{x_+}(x,T) = \frac{\sigma^\alpha\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi} \int_{-1}^1 \frac{dx'}{|x-x'|^{\alpha+1}} \int_0^T dt\, f(x',t|x_0,0). \tag{13.7}$$

The integration limits for x' now only cover the interval $x' \in (-1,1)$, because the probability density vanishes outside this region. Thus we have an integral expression for the leapover distribution, but to simplify it, we will proceed by computing its fractional derivative with respect to the initial point:

$$\begin{aligned} \frac{\partial^\alpha}{\partial|x_0|^\alpha} f_{x_+}(x,T) &= \frac{\sigma^\alpha\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi} \int_{-1}^1 \frac{dx'}{|x-x'|^{\alpha+1}} \int_0^T dt\, \frac{\partial^\alpha}{\partial|x_0|^\alpha} f(x',t|x_0,0) \\ &= \frac{\sigma^\alpha\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi} \int_{-1}^1 \frac{dx'}{|x-x'|^{\alpha+1}} \int_0^T dt\, \frac{\partial^\alpha}{\partial|x'|^\alpha} f(x',t|x_0,0). \end{aligned} \tag{13.8}$$

The replacement here of $\partial^\alpha/\partial|x_0|^\alpha$ by $\partial^\alpha/\partial|x|^\alpha$ may not be obvious because of the presence of absorbing boundaries—a variation of x changes where the density is sampled, but the corresponding variation of x_0 changes the source point, which may not be the same if something like a boundary breaks the translational symmetry of the problem. One way show that the substitution is valid start with the equation (13.4) for the differentials, the last term has the form of a derivative with respect to the density of *free* Lévy walkers, and the change in derivatives may be made at that stage; then time integration leads to an equivalent expression to Eq. (13.5) for $f_{x_+}(x,T)$, but with the x_0 derivative. This is exactly the substitution that we made here in Eq. (13.8). To see the equivalence directly in Eq. (13.8), note that the absorbing boundaries are enforced by the x' integration; inside the integral, $f(x',t|x_0,0)$ acts as a density for free Lévy walkers, justifying the switcheroo in the derivative. Now we can proceed by using the fractional diffusion equation (12.81) to switch

to the time derivative, and then carry out the time integration:

$$\begin{aligned}\frac{\partial^\alpha}{\partial|x_0|^\alpha} f_{x_+}(x,T) &= \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\int_{-1}^{1}\frac{dx'}{|x-x'|^{\alpha+1}}\int_0^T dt\,\partial_t f(x',t|x_0,0)\\ &= \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\int_{-1}^{1}\frac{dx'}{|x-x'|^{\alpha+1}}\Big[f(x',T|x_0,0)-f(x',0|x_0,0)\Big]\\ &= \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\int_{-1}^{1}\frac{dx'}{|x-x'|^{\alpha+1}}\Big[f(x',T|x_0,0)-\delta(x'-x_0)\Big]\\ &= -\frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\left[\frac{1}{|x-x_0|^{\alpha+1}}-\int_{-1}^{1}\frac{dx'}{|x-x'|^{\alpha+1}}f(x',T|x_0,0)\right].\end{aligned} \tag{13.9}$$

Note that the expression here is independent of σ, but still involves the probability density at time T. In the limit $T\longrightarrow\infty$, under the assumption that all trajectories eventually escape—an assumption to be justified, but one that is expected since this is already true for the $\alpha=2$ Gaussian case, the unescaped density vanishes: $f(x,T\to\infty|x_0,0)\longrightarrow 0$. Thus, the resulting derivative of the leapover position becomes

$$\frac{\partial^\alpha}{\partial|x_0|^\alpha} f_{x_+}(x) = -\frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\frac{1}{|x-x_0|^{\alpha+1}}, \tag{13.10}$$

which is conveniently independent of the unescaped state.

To proceed, consider the following integral identity for $x>1$ and $x_0\in(-1,1)$,

$$\frac{\sin(\pi\alpha/2)}{\pi}\int_{-1}^{1}\frac{dx'}{x-x'}\frac{(1-x'^2)^{\alpha/2}}{|x'-x_0|^{\alpha-1}} = [x+(\alpha-1)x_0]-\frac{(x^2-1)^{\alpha/2}}{(x-x_0)^{\alpha-1}}, \tag{13.11}$$

to be justified later. Differentiating both side twice with respect to x_0,

$$\frac{\sin(\pi\alpha/2)}{\pi}\int_{-1}^{1}\frac{dy}{x-x'}\frac{(1-x'^2)^{\alpha/2}}{|x'-x_0|^{\alpha+1}} = -\frac{(x^2-1)^{\alpha/2}}{(x-x_0)^{\alpha+1}}, \tag{13.12}$$

and then rearranging a bit,

$$\frac{\sin(\pi\alpha/2)}{\pi}\int_{-1}^{1}\frac{dx'}{x-x'}\left(\frac{1-x'^2}{x^2-1}\right)^{\alpha/2}\frac{1}{|x'-x_0|^{\alpha+1}} = -\frac{1}{(x-x_0)^{\alpha+1}}. \tag{13.13}$$

Identifying the integral here as the derivative from Eq. (13.6),

$$\frac{\partial^\alpha}{\partial|x_0|^\alpha}\left(\frac{1-x_0^2}{x^2-1}\right)^{\alpha/2}\frac{1}{x-x_0} = -\frac{\Gamma(\alpha+1)}{(x-x_0)^{\alpha+1}}, \tag{13.14}$$

we can then compare this to Eq. (13.10), to identify the leapover density

$$f_{x_+}(x) = \frac{\sin(\pi\alpha/2)}{\pi}\left(\frac{1-x_0^2}{x^2-1}\right)^{\alpha/2}\frac{1}{|x-x_0|}, \quad \text{(leapover density from } (-1,1)) \tag{13.15}$$

where the case $x<-1<x_0$ is covered by reflection symmetry of the problem. That we can make this identification uniquely follows from the form of Eq. (13.10), which is the steady-state form of a (fractional) diffusion equation with a source term. Again, a noteworthy aspect of this result is the absence of the width parameter σ, although it was present in the fractional diffusion equation. Another useful conclusion is that because the integral of (13.15) over $x\notin(-1,1)$ is unity; thus the probability of a trajectory being absorbed right on the border of the escape region is zero.

In deriving the above solution, we have essentially applied the inverse of the operator $\partial^\alpha/\partial|x_0|^\alpha$ to the function $1/(x-x_0)^{\alpha-1}$ on the right-hand side of Eq. (13.10). We can thus apply the same transformation

to Eq. (13.9), with the result[3]

$$f_{x_+}(x,T) = \frac{\sin(\pi\alpha/2)}{\pi}\left[\left(\frac{1-x_0^2}{x^2-1}\right)^{\alpha/2}\frac{1}{|x-x_0|} - \int_{-1}^{1}dx'\left(\frac{1-x'^2}{x^2-1}\right)^{\alpha/2}\frac{1}{|x-x'|}\,f(x',T|x_0,0)\right].$$

(finite-time leapover density from $(-1,1)$) (13.16)

The disadvantage in this expression, of course, is the presence of the unescaped density $f(x',T|x_0,0)$, which is not known in general, except at very short and very long times.

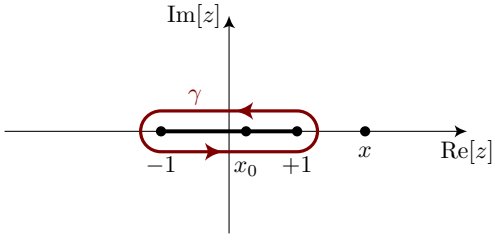
Now to finish up, we should go back to derive Eq. (13.11). To do so, consider the function

$$g(z) := \frac{(z^2-1)^{\alpha/2}}{(z-x_0)^{\alpha-1}}\frac{1}{x-z}, \tag{13.17}$$

where as before $\alpha \in (0,2)$, $x_0 \in (-1,1)$, and $x > 1$. Because of the logarithmic nature of the function

$$\frac{(z^2-1)^{\alpha/2}}{(z-x_0)^{\alpha-1}} = (z-1)^{\alpha/2}(z+1)^{\alpha/2}(z-x_0)^{1-\alpha}, \tag{13.18}$$

there needs to be branch cuts associated with the three branch points $z = \pm1$ and $z = x_0$, and it is convenient to combine them into a single cut running from $z = -1$ to $+1$. Then consider the integral of this function around the contour shown below.



First, you can see that the integrand is single-valued around the contour (thus justifying the choice of branch cut) by noting that in one round trip, the three factors conspire to contribute a net phase of $(e^{2\pi i})^{\alpha/2}(e^{2\pi i})^{\alpha/2}(e^{2\pi i})^{1-\alpha} = e^{2\pi i} = 1$. The branch cut still represents a discontinuity in the value of the integrand just above vs. just below the cut. When the contour is contracted to hug the branch cut,

$$\frac{(z^2-1)^{\alpha/2}}{(z-x_0)^{\alpha-1}} = e^{\pm i\pi\alpha/2}(1-x')^{\alpha/2}(1+x')^{\alpha/2}(x'-x_0)^{1-\alpha}, \tag{13.19}$$

where $z \longrightarrow x' \in (-1,1)$, and $e^{\pm i\pi\alpha/2}$ applies to the top side of the branch cut. The $(1\pm x')^{\alpha/2}$ factors are now manifestly real and positive, but the $(x'-x_0)^{1-\alpha}$ factor is more complicated. For now, take $x' > x_0$ so that this factor is also positive. In this case, the portions of the contour integral running along the branch cut from x_0 to 1 contribute a factor of the form

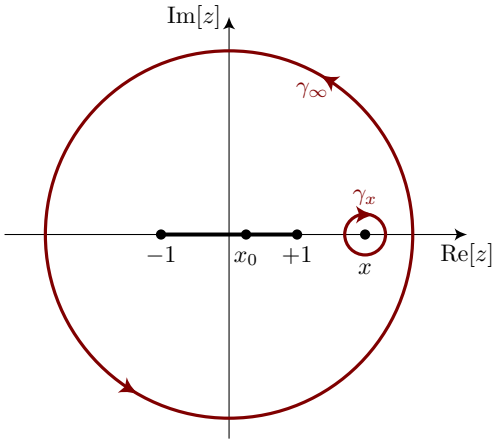
$$e^{+i\pi\alpha/2}\int_1^{x_0}dx'\,\frac{(1-x'^2)^{\alpha/2}}{|x'-x_0|^{\alpha-1}} + e^{-i\pi\alpha/2}\int_{x_0}^{1}dx'\,\frac{(1-x'^2)^{\alpha/2}}{|x'-x_0|^{\alpha-1}} = -2i\sin(\pi\alpha/2)\int_{x_0}^{1}dx'\,\frac{(1-x'^2)^{\alpha/2}}{|x'-x_0|^{\alpha-1}}, \tag{13.20}$$

In the portions of the contour integral running along the branch cut from -1 to x_0, there are extra factors associated with $(x'-x_0)^{1-\alpha} = e^{\pm i\pi(1-\alpha)}|x'-x_0|^{1-\alpha} = -e^{\mp i\pi\alpha}|x'-x_0|^{1-\alpha}$. Combined with the existing factor $e^{\pm i\pi\alpha/2}$, the net result for these sections has the same form as for the sections from x_0 to 1, except for the different integration limits, because the minus sign undoes the effect of exchanging the $e^{\pm i\pi\alpha/2}$ factors. Thus, assembling all the pieces, we have

$$\frac{1}{2\pi i}\oint_\gamma g(z)\,dz = -\frac{\sin(\pi\alpha/2)}{\pi}\int_{-1}^{1}\frac{dx'}{x-x'}\,\frac{(1-x'^2)^{\alpha/2}}{|x'-x_0|^{\alpha-1}}. \tag{13.21}$$

Now an equivalent contour, or rather pair of contours, is shown in the diagram below.

[3]This equation and Eq. (13.15) were also derived by Harold Widom, "Stable Processes and Integral Equations," *Transactions of the American Mathematical Society* **98**, 430 (1961) (doi: 10.2307/1993340).



For this we can use **Cauchy's integral formula**; recall that this says that

$$f(a) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z-a}\, dz, \tag{13.22}$$
(Cauchy integral formula)

under the assumptions that $f(z)$ (the **residue**) is an analytic function in a region encompassing the contour γ, a is a point interior to the closed contour γ, and that the integration proceeds once counterclockwise around the contour (a clockwise integration implies an extra minus sign, and multiple passes around a would imply an integer multiplying factor on the left-hand side). Thus because the contour γ_x encloses a simple pole at x, Cauchy's integral formula gives

$$\frac{1}{2\pi i} \oint_{\gamma_x} g(z)\, dz = \frac{(x^2-1)^{\alpha/2}}{(x-x_0)^{\alpha-1}}. \tag{13.23}$$

Finally, the contour γ_∞ encloses the pole at infinity. That is, we will make the variable change $z \longrightarrow 1/z$ (with $d(1/z) = -dz/z^2$), and thus apply the Cauchy integral formula at $z = 0$ (that is, $1/z = 0$) to

$$-\frac{g(1/z)}{z^2} = -\frac{(z^{-2}-1)^{\alpha/2}}{(1/z-x_0)^{\alpha-1}} \frac{1}{z^2(x-1/z)} = -\frac{(1-z^2)^{\alpha/2}}{(1-zx_0)^{\alpha-1}} \frac{1}{z^2(zx-1)}. \tag{13.24}$$

Note that under the variable change, the direction of the contour integration reverses too (because of the inversion of the phase factor $e^{i\varphi}$. The modified function here doesn't have a simple pole at $z = 0$, however, so we will need a generalization of the integral formula (13.22):

$$f^{(n)}(a) = \frac{n!}{2\pi i} \oint_\gamma \frac{f(z)}{(z-a)^{n+1}}\, dz. \tag{13.25}$$
(Cauchy differentiation formula)

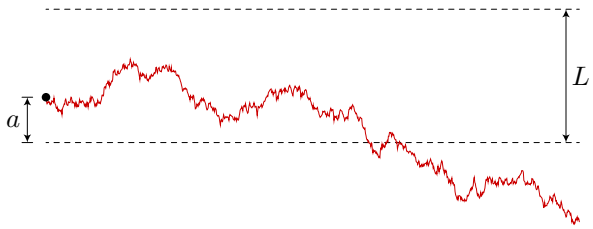
This formula is basically the same as before, but poles of higher order (larger inverse powers of $z-a$) yield derivatives of the residue. The net result is that $-g(1/z)/z^2$ should be differentiated once with respect to z, then z set to zero, with an extra minus sign from the reversal of the contour direction. The final result is

$$\frac{1}{2\pi i} \oint_{\gamma_\infty} g(z)\, dz = -[x + (\alpha-1)x_0]. \tag{13.26}$$

Now equating Eq. (13.21) with the sum of Eqs. (13.23) and (13.26) yields the desired integral identity (13.11).

13.1.1.1 Limit of an Absorbing Half-Space

In our previous notation of Section 6.2, we considered the escape of Gaussian processes from an interval of width L, starting a distance a from the lower edge, as shown below.



We can adapt Eq. (13.15) to this situation by replacing $x \longrightarrow 2x/L - 1$ and $x_0 \longrightarrow 2a/L - 1$ (with $dx \longrightarrow 2dx/L$, so an extra factor of $2/L$ must be included since we are transforming a density) so that $a \in (0, L)$ and $x \notin (0, L)$. After simplifying we obtain

$$f_{x_+}(x)\,dx = \frac{\sin(\pi\alpha/2)}{\pi}\left(\frac{a(L-a)}{x(x-L)}\right)^{\alpha/2}\frac{1}{|x-a|}.$$

(leapover density from interval of width L) (13.27)

To return to the one-sided case, of the first-passage location over a single barrier, we can simply take the limit $L \longrightarrow \infty$,

$$f_{x_a}(x)\,dx = \frac{\sin(\pi\alpha/2)}{\pi}\left(-\frac{a}{x}\right)^{\alpha/2}\frac{1}{|x-a|}, \tag{13.28}$$

for $x < 0$. For this version of the problem, it is more natural to have a positive leapover so to connect with our previous notation, we can replace $x \longrightarrow -x$ and $a \longrightarrow d$ to obtain

$$f_{x_d}(x) = \frac{\sin(\pi\alpha/2)}{\pi}\frac{d^{\alpha/2}}{x^{\alpha/2}(x+d)}.$$

(13.29)
(leapover density)

which recovers the leapover distribution (13.1). Applying the same transformations to Eq. (13.30), we obtain

$$f_{x_d}(x,T) = \frac{\sin(\pi\alpha/2)}{\pi}\left[\frac{d^{\alpha/2}}{x^{\alpha/2}(x+d)} - \frac{1}{x^{\alpha/2}}\int_{-\infty}^{0} dx'\,\frac{|x'|^{\alpha/2}}{x-x'}\,f(x',T|-d,0)\right]$$

(finite-time leapover density) (13.30)

for the leapover density at finite times, with $x > 0$ and the unescaped density $f(x', T|-d, 0)$ nonzero for $x' < 0$.

13.2 First-Passage Time

13.2.1 Sparre Andersen Theorem

Now we come to an important result that, although it applies to *discrete* random walks, it reflects more general asymptotic behavior in boundary-crossing problems in continuous-time random walks. The **Sparre Andersen theorem**,[4] which gives the probability for a discrete random walk starting at $x = 0$ to first enter the region $x < 0$ on the nth step. Of particular interest is the asymptotic scaling, and that the result is *independent of the one-step probability density*, requiring only that it be symmetric.

Let's denote the coordinates of the random-walk by x_n, with $x_0 = 0$, and let

$$f_n(x) := \text{Prob}[x_j > 0\ (\forall_{1<j<n}) \wedge x \le x_n < x + dx] \tag{13.31}$$

[4]The original proof is by Erik Sparre Andersen, "On the Fluctuations of Sums of Random Variables," *Mathematica Scandinavica* **1**, 263 (1953) (doi: 10.7146/math.scand.a-10385). Other, simpler standard derivations are given by William Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, 2nd ed. (Wiley, 1971), Section XII.7 (ISBN: 0471257095); and J. Klafter and I. M. Sokolov, *First Steps in Random Walks: From Tools to Applications* (Oxford, 2011) (ISBN: 9780199234868) (doi: 10.1093/acprof:oso/9780199234868.001.0001). We are following the calculation as given by Uriel Frisch and Hélène Frisch, "Universality of Escape from a Half-space for Symmetrical Random Walks," in *Lévy Flights and Related Topics in Physics: Proceedings of the International Workshop Held at Nice, France, 27– 30 June 1994*, Micheal F. Shlesinger, George M. Zaslavsky, and Uriel Frisch, Eds. (Springer-Verlag, 1995), p. 262 (doi: 10.1007/3-540-59222-9).

be the probability density for x_n and having never visited $x \leq 0$ on any step up to the nth (not including the initial condition). In particular, $f_1(x)$ is the one-step density, which we assume to be an even function. Then these densities satisfy the recursion relation

$$f_n(x) = \int_0^\infty dx'\, f_1(x - x')\, f_{n-1}(x') \tag{13.32}$$

because $f_1(x)$ is the one-step distribution, and successive steps in the random walk are given by convolutions with the one-step density [with the negative half-line excluded according to the definition of $f_n(x)$].

Then let

$$p_n := \int_0^\infty dx\, f_n(x) \qquad (n \geq 1) \tag{13.33}$$

be the probability of having never visited $x \leq 0$ through the nth step. This is slightly different than the probability that we mentioned at the outset, but the idea will be to derive a recursion relation for p_n. To do this, define the quantity

$$\mathscr{Q}_n := \sum_{\substack{\ell+m=n \\ \ell\geq 1, m\geq 1}} \int_0^\infty dx\, f_\ell(x) \int_0^x dx'\, f_m(x'). \tag{13.34}$$

Then since in general

$$\int_0^\infty dx \int_0^x dy\, f(x,y) = \int_0^\infty dy \int_y^\infty dx\, f(x,y), \tag{13.35}$$

and the expression (13.34) is invariant under exchange of ℓ and m (and thus of x and x', we then have

$$\mathscr{Q}_n = \frac{1}{2} \sum_{\substack{\ell+m=n \\ \ell\geq 1, m\geq 1}} \int_0^\infty dx\, f_\ell(x) \int_0^\infty dx'\, f_m(x') = \frac{1}{2} \sum_{\substack{\ell+m=n \\ \ell\geq 1, m\geq 1}} p_\ell p_m. \tag{13.36}$$

On the other hand, putting the recursion (13.32) into the definition (13.34) for $f_\ell(x)$ gives

$$\mathscr{Q}_n = \sum_{\substack{\ell+m=n \\ \ell\geq 1, m\geq 1}} \int_0^\infty dx \int_0^x dx'\, f_m(x') \int_0^\infty dx''\, f_1(x - x'')\, f_{\ell-1}(x''). \tag{13.37}$$

Shifting $\ell \longrightarrow \ell + 1$ gives

$$\mathscr{Q}_n = \sum_{\substack{\ell+m=n-1 \\ \ell\geq 0, m\geq 1}} \int_0^\infty dx \int_0^x dx'\, f_m(x') \int_0^\infty dx''\, f_1(x - x'')\, f_\ell(x''), \tag{13.38}$$

and separating the $\ell = 0$ term in the sum gives

$$\mathscr{Q}_n = \int_0^\infty dx \int_0^x dx'\, f_{n-1}(x')\, f_1(x) + \sum_{\substack{\ell+m=n-1 \\ \ell\geq 1, m\geq 1}} \int_0^\infty dx \int_0^x dx'\, f_m(x') \int_0^\infty dx''\, f_1(x - x'')\, f_\ell(x''), \tag{13.39}$$

after using $f_0(x') = \delta(x')$. Now working with the first integral, we have

$$\begin{aligned}
I_n &:= \int_0^\infty dx\, f_1(x) \int_0^x dx'\, f_{n-1}(x') \\
&= \int_0^\infty dx\, f_1(x) \int_0^\infty dx'\, f_{n-1}(x') - \int_0^\infty dx\, f_1(x) \int_x^\infty dx'\, f_{n-1}(x') \\
&= \int_0^\infty dx'\, f_{n-1}(x') \int_0^\infty dx\, f_1(x) - \int_0^\infty dx'\, f_{n-1}(x') \int_0^{x'} dx\, f_1(x) \\
&= \int_0^\infty dx'\, f_{n-1}(x') \int_{x'}^\infty dx\, f_1(x) \\
&= \int_0^\infty dx'\, f_{n-1}(x') - \int_0^\infty dx'\, f_{n-1}(x') \int_{-x'}^\infty dx\, f_1(x) \\
&= p_{n-1} - \int_0^\infty dx'\, f_{n-1}(x') \int_0^\infty dx\, f_1(x - x') \\
&= p_{n-1} - \int_0^\infty dx\, f_n(x) \\
&= p_{n-1} - p_n,
\end{aligned} \tag{13.40}$$

where we used the symmetry of $f_1(x)$ and the recursion (13.32). Working with the symmetrized second term in Eq. (13.39), we have

$$\begin{aligned}
I'_n &:= \int_0^\infty dx \int_0^x dx'\, f_m(x') \int_0^\infty dx''\, f_1(x - x'')\, f_\ell(x'') + (\ell \longleftrightarrow m) \\
&= \left[\int_0^x dx'\, f_m(x') \int_0^\infty dx''\, F_1(x - x'')\, f_\ell(x'') \right]_{x=0}^\infty - \int_0^\infty dx\, f_m(x) \int_0^\infty dx''\, F_1(x - x'')\, f_\ell(x'') + (\ell \longleftrightarrow m)
\end{aligned} \tag{13.41}$$

after an integration by parts, where $f_1(x) = F_1'(x)$ and $F_1(0) = 0$ and $(\ell \longleftrightarrow m)$ means to repeat the previous expression with ℓ and m interchanged. Notice that the last term is symmetric under exchange of ℓ and m, but *anti*symmetric under exchange of x and x'' (owing to the asymmetry of the antiderivative $F_1(x)$, which in turn follows from the symmetry of $f_1(x)$. Thus, only the first term can be nonvanishing. In the remaining term, we can take the limit $x \longrightarrow \infty$ while holding x'' fixed, and use $\lim_{x\to\infty} F_1(x) = 1/2$ to write

$$\begin{aligned}
I'_n &= \frac{1}{2} \int_0^\infty dx'\, f_m(x') \int_0^\infty dx''\, f_\ell(x'') + (\ell \longleftrightarrow m) \\
&= \int_0^\infty dx\, f_\ell(x) \int_0^\infty dx'\, f_m(x') = p_\ell p_m.
\end{aligned} \tag{13.42}$$

Now putting Eqs. (13.40) and (13.42) into Eq. (13.39),

$$Q_n = p_{n-1} - p_n + \frac{1}{2} \sum_{\substack{\ell+m=n-1 \\ \ell\geq 1, m\geq 1}} p_\ell p_m, \tag{13.43}$$

and equating this result with Eq. (13.36) yields the recursion

$$p_n = p_{n-1} + \frac{1}{2} \sum_{\substack{\ell+m=n-1 \\ \ell\geq 1, m\geq 1}} p_\ell p_m - \frac{1}{2} \sum_{\substack{\ell+m=n \\ \ell\geq 1, m\geq 1}} p_\ell p_m \tag{13.44}$$

with initial condition $p_1 = 1/2$. Note that, at this point, this recursion *only* involves the noncrossing probabilities p_n, without any reference to the details of the probability densities [other than the assumption of symmetry of $f_1(x)$]. Thus we are already seeing the emergence of universal behavior in this problem.

To solve the recurrence (13.44), a handy tool is the generating function (Section 1.4.4)

$$p(z) := \sum_{n=1}^{\infty} p_n z^n \tag{13.45}$$

for the probabilities p_n, where $p_0 = 0$. Multiplying Eq. (13.44) by z^n and sum n from 2 to ∞, the recursion becomes

$$\sum_{n=2}^{\infty} p_n z^n = z \sum_{n=2}^{\infty} p_{n-1} z^{n-1} + \frac{z}{2} \sum_{n=2}^{\infty} \sum_{\substack{\ell+m=n-1 \\ \ell\geq 1, m\geq 1}} p_\ell z^\ell p_m z^m - \frac{1}{2} \sum_{n=2}^{\infty} \sum_{\substack{\ell+m=n \\ \ell\geq 1, m\geq 1}} p_\ell z^\ell p_m z^m. \tag{13.46}$$

Adding $p_1 z$ to both sides allows the sum on the left-hand side to be completed, while shifting an index in the second term on the right-hand side lets us combine the double summations:

$$p(z) = zp(z) + zp_1 + \frac{(z-1)}{2} \sum_{n=2}^{\infty} \sum_{\substack{\ell+m=n \\ \ell\geq 1, m\geq 1}} p_\ell z^\ell p_m z^m. \tag{13.47}$$

Now recognizing the remaining sum as a product of two series, we arrive at an equation determining $p(z)$:

$$(1-z)p(z) = zp_1 + \frac{1}{2}(z-1)p^2(z). \tag{13.48}$$

Since $p_1 = 1/2$ due to the symmetry of $f_1(x)$, as we have assumed multiple times throughout, we can solve the quadratic relation (13.48) to find

$$p(z) = -\frac{1 - z \pm \sqrt{1-z}}{1-z} = \mp \frac{1}{\sqrt{1-z}} - 1. \tag{13.49}$$

Only the positive solution makes sense, and so

$$p(z) = \frac{1}{\sqrt{1-z}} - 1. \tag{13.50}$$

Now using the series expansion for the inverse square root,[5] we have

$$p(z) = \sum_{n-1}^{\infty} \frac{(2n-1)!!}{(2n)!!} z^n, \tag{13.51}$$

and thus

$$p_n = \frac{(2n-1)!!}{(2n)!!} \tag{13.52}$$

where the double factorial is $n!! := n(n-2)\cdots 1$. We can simplify to regular factorials using $(2n)!! = 2^n n!$ and $(2n-1)!! = (2n)!/(2n)!!$, with the explicit result

$$p_n = \frac{(2n)!}{2^{2n}(n!)^2} \quad \text{(probability to not visit } n \leq 0 \text{ by } n\text{th step)} \tag{13.53}$$

for the non-crossing probability. In particular, using the Stirling approximation for large n

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \tag{13.54}$$

[5] I. S. Gradstein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, English translation 7th ed., A. Jeffrey and D. Zwillinger, Eds. (Academic Press, 2007), p. 25, Formula 1.112.4.

the noncrosssing probability has a universal, long-time scaling

$$p_n \sim \frac{1}{\sqrt{\pi n}}. \qquad (13.55)$$
(tail of noncrossing distribution)

The fact that this probability converges to zero as $n \longrightarrow \infty$ says that a symmetric random walk will eventually cross the origin, independent of the details of the step distribution.

Again, p_n is not *quite* the probability we set out to calculate; rather we wanted the probability P_n of the *first* crossing to be step n, which means we want the probability of noncrossing on the $(n-1)$th step, and then exclude the subset that also does not cross on the nth step:

$$P_n = p_{n-1} - p_n. \qquad (13.56)$$

Using Eq. (13.53), we have

$$P_n = \frac{(2n-3)!!}{(2n-2)!!} - \frac{(2n-1)!!}{(2n)!!} = \frac{(2n-1)!!}{(2n)!!}\left(\frac{2n}{2n-1} - 1\right), \qquad (13.57)$$

so that

$$P_n = \frac{(2n)!}{(2n-1)2^{2n}(n!)^2} \qquad (13.58)$$
(first-crossing probability on nth step)

for the first-crossing probability. Again computing the large-n asymptotic scaling, we find

$$P_n \sim \frac{1}{\sqrt{4\pi n^3}}. \qquad (13.59)$$
(tail of first-crossing)

The results (13.53), (13.55), (13.58), and (13.59) are all forms of the **Sparre Andersen theorem**. Again, that not only the asymptotics, but the full distributions may be worked out under only the assumption of a symmetric one-step density $f_1(x)$ is pretty amazing. The density need not even have a finite variance, so this result applies to Lévy-type discrete processes as well. (Note that similar asymptotic scalings apply to walks on a discrete lattice, say, where the lattice sites are not integers, although the exact probabilities and the asymptotic coefficient are different—can you see where the above treatment breaks down?)

Sparre Andersen's name is used in particular to refer to the universal $n^{-3/2}$ scaling of the first-crossing distribution, even in cases where this result does not strictly apply. The most important example is for continuous-time stochastic processes, where an infinity of steps occur in any finite interval. The error, of course, in applying these results to the continuous-time case is that there is no ''checking'' for crossings between the discrete times (and taking the time interval between steps to zero amounts to $n \longrightarrow \infty$ and thus $p_n \longrightarrow 0$). However, in the long-time tail, this distinction does not matter so much, since a coarse-graining of this distribution is perfectly appropriate; thus we expect a $t^{-3/2}$ tail to the first-crossing distribution for continuous-time processes. For this to make sense, we have to assume that the process has ''broken away'' from the origin at some short time, so heuristically this result should also apply to the first crossing of a barrier away from $d = 0$ for continuous-time processes. We have already seen that this is the case for Wiener processes [see Eq. (6.11)], and this result, remarkably, *also* holds for continuous-time Lévy processes.

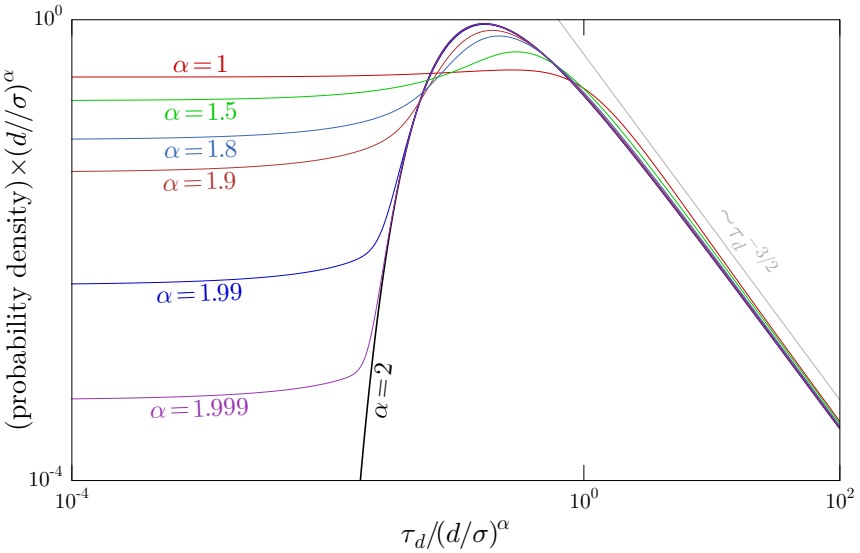
13.2.2 First-Passage Time of Stable Lévy Processes

Here we will summarize some of the known results for the boundary-crossing problem in the case of the stable Lévy processes. The Sparre Andersen theorem still applies in the case of symmetric $\beta = 0$ processes, and this pins down the scaling of the asymptotic tail of the probability density of first-passage times. However, the coefficient of the $\tau^{-3/2}$ dependence has also been calculated, and the result is[6]

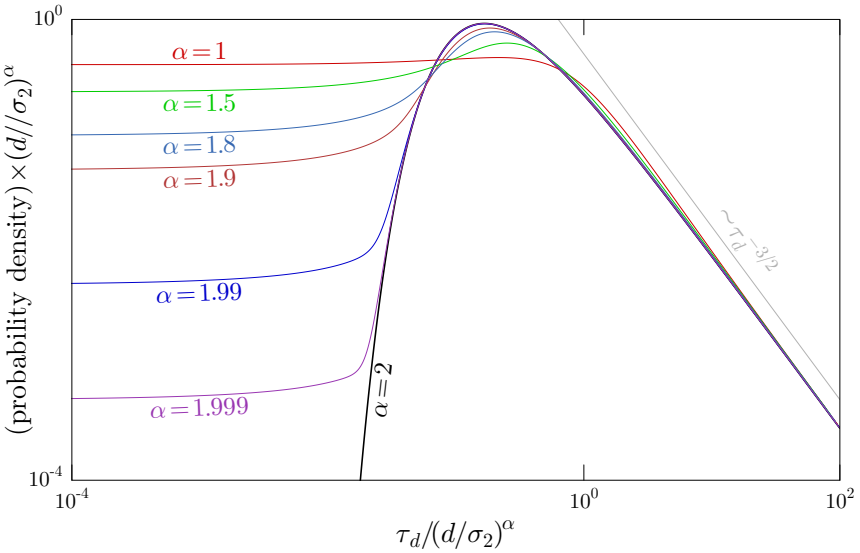
$$f_{\tau_d}(\tau) \sim \frac{(d/\sigma)^{\alpha/2}}{\alpha\sqrt{\pi}\Gamma(\alpha/2)}\,\tau^{-3/2}. \qquad (13.60)$$
(asymptotic first-passage density)

[6]Tal Koren, Michael A. Lomholt, Aleksei V. Chechkin, Joseph Klafter, and Ralf Metzler, ''Leapover Lengths and First Passage Time Statistics for Lévy Flights,'' *Physical Review Letters* **99**, 160602 (2007) (doi: 10.1103/PhysRevLett.99.160602).

This behavior is illustrated in the plots below, which show numerically calculated first-passage distributions.



The universal Sparre Andersen scaling is clearly visible in the tail of the distributions. The other thing to notice is how the distributions for various α tend to "hug" the Gaussian ($\alpha = 2$) distribution at long times; with the breakaway point moving to the left as $\alpha \longrightarrow 2$. This is slightly more apparent if we rescale the distributions to match the asymptotic behavior (13.60) among the various α, as shown below.



To understand this rescaling of the distributions, first note that schematically, we have a distribution tail of the form

$$f(\tau) \sim c\tau^{-3/2}. \tag{13.61}$$

Suppose that we rescale the time by $\tau \longrightarrow s\tau$, preserving normalization of the distribution:

$$f(\tau) \sim c(\alpha)\,\tau^{-3/2}s^{-3/2}s = c(\alpha)\,\tau^{-3/2}s^{-1/2} \tag{13.62}$$

To eliminate dependence of the tail on $c(\alpha)$, we should choose $s \propto c^2(\alpha)$, in particular $s = c^2(\alpha)/c^2(2)$, or

$$s = \frac{(d/\sigma)^{\alpha-2}}{\Gamma^2(\alpha/2+1)}. \tag{13.63}$$

This corresponds to, for example, choosing σ to vary with α according to

$$\sigma_\alpha^\alpha = \frac{\sigma_2^2}{\Gamma^2(1+\alpha/2)\, d^{2-\alpha}} \tag{13.64}$$

(with the Gaussian value σ_2 as a reference), which makes for more sensible comparisons in this case than leaving σ constant.

In some cases more is known. For example, the full first-passage density for a Cauchy process is[7]

$$f_{\tau_d}(t) = \frac{d}{\pi\sigma t^2 \left(1+\dfrac{d^2}{\sigma^2 t^2}\right)^{3/4}} \left(\frac{d}{\sigma t}\right)^{-(1/2)-(1/\pi)\tan^{-1}(d/\sigma t)} \exp\left[\frac{d}{4\pi\sigma t}\Phi(-d^2/\sigma^2 t^2, 2, 1/2)\right],$$

(first-passage density, Cauchy process) (13.65)

where the Lerch transcendent is defined by

$$\Phi(z,s,\alpha) := \sum_{n=0}^{\infty} \frac{z^n}{(n+\alpha)^s} \tag{13.66}$$

As $t \longrightarrow \infty$, $\Phi(-d^2/t^2, 2, 1/2) \longrightarrow 4$, so the exponential factor becomes unity at long times; the remaining long-time asymptotic scaling is

$$f_{\tau_d}(t \to \infty) \sim \frac{d}{\pi\sigma^2 t^2}\left(\frac{d}{\sigma t}\right)^{-(1/2)} = \frac{\sqrt{d}}{\pi(\sigma t)^{3/2}}. \tag{13.67}$$

At short times, the dependence of the Lerch transcendent is not so straightforward via the series definition; however, the asymptotics can be determined by the identification

$$x^{-(1/\pi)\tan^{-1} x} \exp\left[\frac{x}{4\pi}\Phi(-x^2, 2, 1/2)\right] = \exp\left[-\frac{1}{\pi}\int_0^x \frac{\log w}{1+w^2}\, dw\right], \tag{13.68}$$

where the right hand side reduces to unity as $x \longrightarrow \infty$ (and $x \longrightarrow 0$, as we already determined). Then the short-time limit of the first-passage density is

$$f_{\tau_d}(t \to 0) \sim \frac{d}{\pi t^2 \left(\dfrac{d^2}{\sigma^2 t^2}\right)^{3/4}} \left(\frac{d}{\sigma t}\right)^{-(1/2)} = \frac{\sigma}{\pi d}. \tag{13.69}$$

This can also be inferred from the jump probability (12.182)

$$P[dS_\alpha(t) > x] \sim \frac{\Gamma(\alpha)\sin(\pi\alpha/2)}{\pi}\left(\frac{x}{\sigma}\right)^{-\alpha} dt, \tag{13.70}$$

with $\alpha = 1$ and $x = d$.

Additionally, the full first-passage density is known in the case of one-sided ($\beta = \pm 1$) stable processes,[8]

$$f_{\tau_d}(\tau) = \frac{\sigma^\alpha}{\cos(\pi\alpha/2)\, d^\alpha} \sum_{n=0}^{\infty} \frac{(-\sigma\tau)^{\alpha n}}{n!\Gamma(1-\alpha-\alpha n)\cos^n(\alpha\pi/2)\, d^{\alpha n}},$$

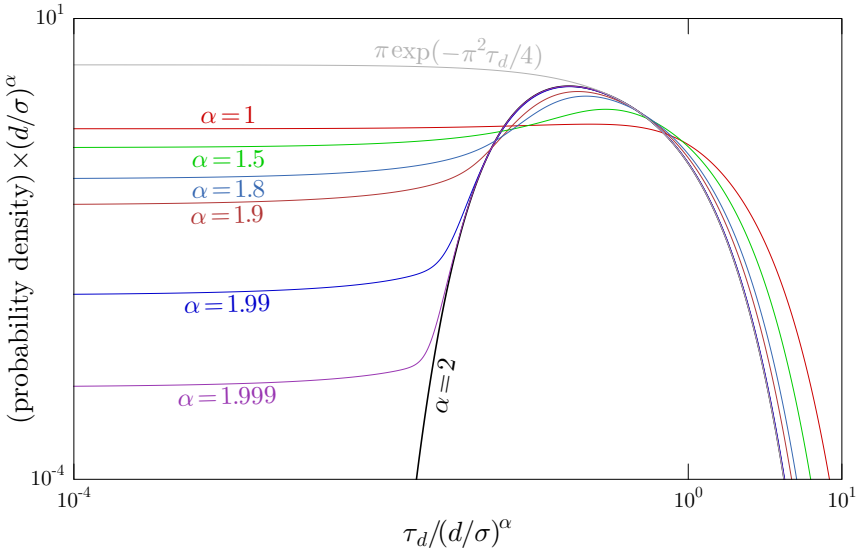
(first-passage density, one-sided, $\alpha < 1$) (13.71)

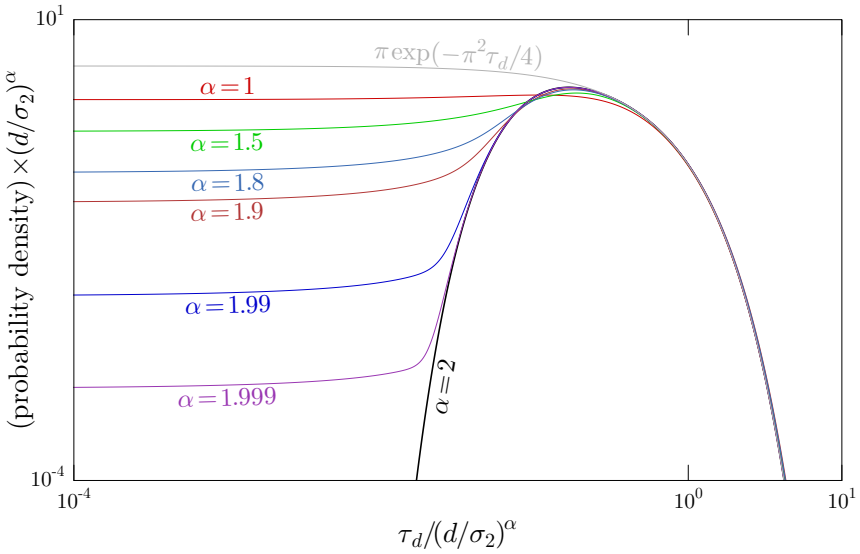
but only for with $\alpha < 1$.

[7] D. A. Darling, "The Maximum of Sums of Stable Random Variables," *Transactions of the American Mathematical Society* **83**, 164 (1956) (doi: 10.1090/S0002-9947-1956-0080393-6) (doi: 10.2307/1992908).

[8] Tal Koren *et al.*, *op. cit.*

13.3 First Escape of Stable Lévy Processes





13.4 Exercises

Problem 13.1

Find the generating function corresponding to the first-crossing probabilities (13.56)

$$P_n = \frac{(2n)!}{(2n-1)2^{2n}(n!)^2} \qquad (n \geq 1), \tag{13.72}$$

and then show directly from the generating function that the distribution is normalized.

Hint: start by working with the definition (13.56)

$$P_n = p_{n-1} - p_n. \tag{13.73}$$

Chapter 14

Lévy Numerical Methods

14.1 Simulation of Symmetric Stable Processes

In simulations involving α-stable Lévy processes, it is necessary to generate random deviates according to the probability density (12.61):

$$f_{\alpha,\beta,\sigma,\mu}(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\, \tilde{f}_{\alpha,\beta,\sigma,\mu}(k;t). \tag{14.1}$$

It is somewhat simpler to discuss the generation of the symmetric processes ($\beta = 0$) according to the probability density (12.9)

$$f_\alpha(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\, e^{-t\sigma^\alpha |k|^\alpha} = \frac{1}{\pi}\int_0^{\infty} dk\, \cos(kx)\, e^{-t\sigma^\alpha k^\alpha}. \tag{14.2}$$

In this case it turns out that the appropriate deviates may be generated according to[1]

$$\Delta S_\alpha = \sigma(\Delta t)^{1/\alpha}\frac{\sin(\alpha u)}{(\cos u)^{1/\alpha}}\left(\frac{\cos[(1-\alpha)u]}{w}\right)^{(1-\alpha)/\alpha}$$
(random deviate for symmetric α-stable Lévy process) (14.3)

where u is a uniformly distributed random deviate on $(-\pi/2, \pi/2)$, and w is a random deviate with distribution $\exp(-x)$, generated by $w = -\log(u')$, where u' is a uniformly distributed random deviate on $(0, 1)$. In the case of a Cauchy process ($\alpha = 1$), this formula simplifies considerably to

$$\Delta C = \Delta S_{\alpha=1} = \sigma\Delta t\, \tan u, \tag{14.4}$$
(random deviate for Cauchy process)

where again u is a uniformly distributed random deviate on $(-\pi/2, \pi/2)$. In the $\alpha = 2$ (Gaussian) case, this method reduces to

$$\Delta S_2 = \sigma\sqrt{2\Delta t}\sin(u)\sqrt{2w}.$$
(random deviate for Gaussian process) (14.5)

This is essentially the Box–Muller transformation [Eq. (10.83)], except for the scaling of $2\Delta t$ instead of Δt (as expected due to the scaling convention for α-stable processes), and that the complementary deviate using cos is omitted.

[1] This is a special case of a more general method that includes the asymmetric stable processes by J. M. Chambers, C. L. Mallows, and B. W. Stuck, ''A Method for Simulating Stable Random Variables,'' *Journal of the American Statistical Association* **71**, 340 (1976) (doi: 10.2307/2285309). See also Rama Cont and Peter Tankov, *Financial Modelling with Jump Processes* (Chapman & Hall/CRC, 2004), Algorithm 6.6. For an explicit proof and an overview of references and errors in the literature, see Rafał Weron, ''On the Chambers-Mallows-Stuck method for simulating skewed stable random variables,'' *Statistics & Probability Letters* **28**, 165 (1996) (doi: 10.1016/0167-7152(95)00113-1).

While the formula (14.3) is fine in principle, it is neither particularly efficient nor accurate for the purposes of numerical computation. A better adapted version is[2]

$$\begin{aligned}
\Delta S_\alpha &= \sigma(\Delta t)^{1/\alpha}\big[1+(1-\alpha)d\big]\frac{2(a-b)(1+ab)}{(1-a^2)(1+b^2)} \\
a &:= \tan\left(\frac{u}{2}\right) \\
b &:= \tan\left(\frac{(1-\alpha)u}{2}\right) \\
d &:= \log(z)\frac{D_2[(1-\alpha)\log(z)/\alpha]}{\alpha} \\
z &:= \frac{(1+a^2)(1-b^2)}{w(1-a^2)(1+b^2)} \\
D_2(x) &:= \frac{e^x-1}{x}.
\end{aligned} \tag{14.6}$$

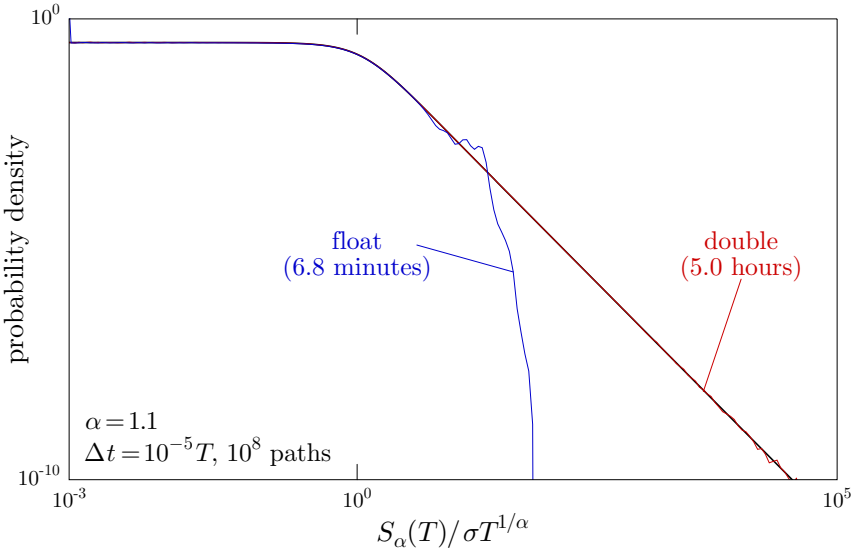
The intermediate quantities a, b, d, and z are evaluated numerically and assembled in the final result; care should be taken to evaluate w, a, $(1-a^2)$, b, and $(1-b^2)$ accurately (e.g., in double precision). Also, the function $D_2(x)$ can be computed accurately in terms of a rational-function approximation to avoid accuracy problems near $x=0$, but this can be computed in reduced precision.

As an illustration, these methods are implemented in a simulation of stable Lévy processes ($\alpha = 1.1$, $\beta = 0$), using the "naïve" method (14.3). A C (actually, OpenCL) function that implements the equation is shown below.

```
double levyinc_naive(uint *rstate) {
  double u1, u2;
  const double pi = 3.1415926535897932385;
  u1 = (drand(rstate)-0.5)*pi; u2 = drand(rstate);
  return sin($alpha*u1) / pow(cos(u1),(double)$onebyalpha)
           * pow(-cos($onemalpha*u1)/log(u2), $onemalphabyalpha);
}
```

Here the variables `$alpha`, `$onebyalpha`, `$onemalpha`, and `$onemalphabyalpha` are the values α, $1/\alpha$, $1-\alpha$, and $(1-\alpha)/\alpha$, which are pre-computed in double precision and inserted into the code (in essentially the same way as `#define` preprocessor directives) just before compilation, to ensure these quantities are not recalculated at each iteration. Also `drand` is a function that generates a random `double` in $(0,1)$, with full 53 randomized mantissa bits. The numerical results are shown below, for simulations that ran over 10^5 steps to a final time $T=1$; the plots are histograms of 10^8 trajectories, compared to the expected distribution.

[2]This again is a special case of the algorithm in J. M. Chambers, C. L. Mallows, and B. W. Stuck, *op. cit.*



The "double" case uses the function as written, while the "float" case replaces the `double` declarations by `float` (and a `float` version `frand` of the random-number generator). The float case is clearly much faster, but at the expensive of accuracy (the float results are not even particularly usable). The times indicated are run times for an OpenCL implementation running on an NVIDIA Titan Xp GPU. On this GPU, double-precision operations run slower than single-precision operations by a factor of 24, illustrating why one might want to worry about the precision of individual operations.

For comparison, we will present the same simulations are shown, but using the more careful implementation of Eqs. (14.6), showing their improvements in performance and accuracy. The Lévy-increment function is shown below, this time coded so that only critical steps are computed in `double`, while less critical operations are done in `float` for speed. The variables `$alpha`, `$onebyalpha`, `$onemalpha`, and `$onemalphabyalpha` are defined as before, and the `$iscauchy` variable is a flag to use the simpler calculation for $\alpha = 1$, with the proper code selected at compile time to avoid an execution branch.

```
float levyinc(uint *rstate) {
  float onema2, onemb2, onepb2, d, z, logz, aw, bw;
  double u, phiby2, a, b;
  const double piby2 = 1.5707963267948966192;
  u = drand(rstate);
  phiby2 = piby2*(u-0.5); /* scaled uniform deviate */
  a = tanapx18(phiby2); //  a = tan(phiby2);
  aw = (float)a;
  onema2 = 1-a*a; /* a's precision is critical here */
  /* save time in Cauchy case */
  #if $iscauchy
    return 2*aw / onema2;
  #else
    /* precision of tangent is critical here */
    b = tanapx18($onemalpha*phiby2); //  b = tan((1-alpha)*phiby2);
    bw = (float)b;
    onemb2 = 1-b*b; /* b's precision critical here */
    onepb2 = 2.-onemb2;
    z = (2.-onema2)*(onemb2) / (fexpdev(rstate)*onema2*onepb2);
    logz = log(z);
    d = D2(logz*(float)$onemalphabyalpha) * logz*(float)$onebyalpha;
    return (1+(float)$onemalpha*d) * 2 * (float)(aw-bw)*(float)(1+aw*bw)
             / (onema2*onepb2);
  #endif
}
```

This method requires some support functions; the function `tanapx18` implements $\tan(x)/x$ to high accuracy, while `D2` implements $(e^x - 1)/x$ to somewhat lower accuracy.

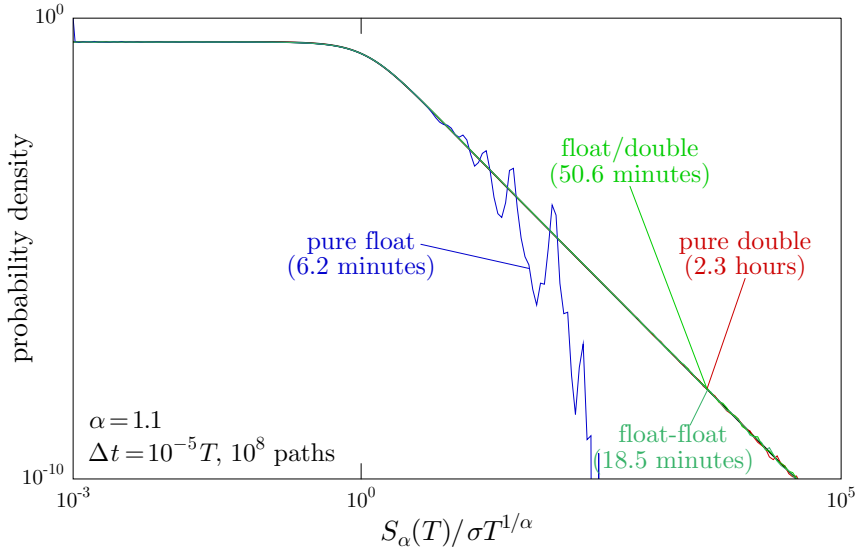
```
/* tan(x), Chebyshev polynomial approx of tan(x)/x */
/*   18th order approx for tan(x)/x good to better than 8.e-12 */
/*   only defined for abs(x)<pi/4 */
double tanapx18(double x) {
  const double a0 = 0.99999999999368;
  const double a1 = 0.333333335356791;
  const double a2 = 0.133333226831938;
  const double a3 = 0.053970419358616;
  const double a4 = 0.0218472580312769;
  const double a5 = 0.0089932295021841;
  const double a6 = 0.0031353885035028;
  const double a7 = 0.00242692527355665;
  const double a8 = -0.00060300844921943;
  const double a9 = .00096153489803288;
  double x2, y;
  x2 = x*x;
  y = fma(x2, a9, a8);
  y = fma(x2, y, a7);
  y = fma(x2, y, a6);
  y = fma(x2, y, a5);
  y = fma(x2, y, a4);
  y = fma(x2, y, a3);
  y = fma(x2, y, a2);
  y = fma(x2, y, a1);
  y = fma(x2, y, a0);
  return x*y;
}

/* D2(x) = (exp(x)-1)/x, from Chambers76 */
float D2(float z) {
  const float p1 = 0.840066852536483239e3;
  const float p2 = 0.200111141589964569e2;
  const float q1 = 0.168013370507926648e4;
  const float q2 = 0.18013370407390023e3;
  const float q3 = 1.0;
  float zz, pv;
  if (fabs(z)>0.1) {
    return expm1(z)/z;
  }
  zz = z*z;
  pv = p1+zz*p2;
  return 2*pv/(q1+zz*(q2+zz*q3)-z*pv);
}
```

Finally, the generation of the exponential deviate requires some care. The `double` implementation below is the naïve one (a logarithmic transformation of a uniform deviate) that works acceptably with the high precision. In a `float` implementation, accuracy is a concern when the uniform deviate is close to 0 or 1. In this case, a 32-bit random unsigned integer is generated directly by `irand`; 23 bits are used to generate the uniform deviate, while one extra bit acts as a switch between two cases that accurately handle a uniform deviate in $(0, 0.5)$ and $(0.5, 1)$. The trick of adding `FLOAT_RAND_FAC/2` ensures that the `float` deviate never returns a `0.0f`.

```
/*** exponential deviates, careful evaluation near 0 and infinity ***/
/* pure double precision version */
#define dexpdev(rstate) (-log(drand(rstate)))
```

```
/* pure single precision version */
#define FLOAT_RAND_FAC        (float)1.1920928955078125000e-7   // 2^-23
float fexpdev(uint *rstate) {
  uint ui, sw;
  float u;
  ui = irand(rstate);
  sw = ui >> 31; // select (0,0.5) or (0.5,1) on unused bit
  u = (ui & 0x007fffff)*FLOAT_RAND_FAC + (FLOAT_RAND_FAC/2);
  return -log(0.5f*u)*sw - log1p(-0.5f*u)*(1-sw);
}
```

The "float/double" case carries out all the computations in single precision except for the critical ones in double, as shown in the code above. The "pure float" and "pure double" cases are pure single-precision and double-precision versions of the same code, respectively. The pure-double histogram overlaps the analytic curve almost perfectly.



Of course, the pure float computation is still the fastest, but with only marginal improvement in accuracy over the naîve algorithm. The pure double algorithm has excellent accuracy, and is a significant improvement in speed (factor of 2) over the naïve calculation. The float/double algorithm saves again more than a factor of 2 by speeding up noncritical calculations, while maintaining excellent accuracy. Note in all of these cases, the *single*-step histograms are visually identical; the deviations from the expected curve in the pure float case represent spurious correlations induced by roundoff error after many steps. The run time in the float/double case can be reduced even further (down to 18.5 minutes) by using "float–float" arithmetic[3] in place of double-precision operations, because the large difference is float vs. double speed on the GPU device. The algorithm emulates about 44 bits of mantissa accuracy using a pair of `float` variables, which is comparable to the 53 mantissa bits for a `double` variable.

[3]see, e.g., Guillaume Da Graça and David Defour, "Implementation of float-float operators on graphics hardware," (arXiv: cs/0603115) (2006).

14.1.1 Simulation of Asymmetric Stable Processes

For asymmetric stable densities, the transformation to obtain a deviate from a uniform deviate ϕ on $(-\pi/2, \pi/2)$ and an exponential deviate w is[4]

$$S_{\alpha\neq 1,\beta} = \frac{\sin\left(\alpha\phi + \tilde{\beta}\right)}{(\cos\tilde{\beta}\cos\phi)^{1/\alpha}} \left(\frac{\cos\left[(1-\alpha)\phi + \tilde{\beta}\right]}{w}\right)^{(1-\alpha)/\alpha}$$
$$\tilde{\beta} = \tan^{-1}\left(\beta \tan\frac{\pi\alpha}{2}\right)$$

(random deviate for asymmetric α-stable Lévy process) (14.7)

in the case of $\alpha \neq 1$, and

$$S_{\alpha=1,\beta} = \left[\left(1 + \frac{2\beta}{\pi}\phi\right)\tan\phi - \frac{2\beta}{\pi}\log\left(\frac{w\cos\phi}{1 + 2\beta\phi/\pi}\right)\right]$$

(random deviate for asymmetric $\alpha = 1$-stable Lévy process) (14.8)

in the case $\alpha = 1$. [5]

(put in terms of tan)

14.2 Calculation of Stable Densities

It is useful to be able to numerically compute the values of the stable Lévy densities, particularly because they typically don't have analytic expressions. Codes to perform this calculation both generally and accurately rely typically on numerical quadrature of integral representations of the distribution functions.[6]

As a simple example, in the symmetric ($\beta = 0$) case, the integral expression for the probability density (12.9)

$$f_\alpha(x;t) = \frac{1}{\pi}\int_0^\infty dk\, \cos(kx)\, e^{-t\sigma^\alpha k^\alpha} \tag{14.9}$$

applies. Below is a simple Julia code that implements this numerical integration.

```
using Plots
using SpecialFunctions
using QuadGK

const α = 1.5;
const rtl=1e-7;
const ord=9;
const mxeval=10^9;

### function to handle numerical integration
function f(x, α)
  TOL=1e-5;
  c = sin(pi*α/2)*gamma(α+1)/pi;
  asympt = c*abs(x)^(-α-1);
  if (asympt<TOL)
    return asympt;
  else
```

[4] Rafał Weron, *op. cit.*; John P. Nolan, *Univariate Stable Distributions: Models for Heavy Tailed Data* (Springer, 2020), p. 20 (ISBN: 9783030529154).

[5] Rafał Weron, *op. cit.*; John P. Nolan, *op. cit.*. Note that the formulas in these two references differ slightly, apparently due to a minor typo in the former.

[6] Aurel Wintner, "Stable Distributions and Laplace Transforms," *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze 3e série* **10**, 127 (1956) http://www.numdam.org/article/ASNSP_1956_3_10_3-4_127_0.pdf; John P. Nolan, "Numerical calculation of stable densities and distribution functions," *Communications in Statistics. Stochastic Models*, **13**, 759 (1997) (doi: 10.1080/15326349708807450).

```
    integ, err = quadgk(k -> exp(-abs(k)^α) * cos(k*x),
                        0, Inf, rtol=rtl, order=ord, maxevals=mxeval);
    return integ/pi;
  end #if
end #function f

# call function on position array to calculate density
dx = 0.1;
xarr = collect( 1:1000 ) * dx;
dist = f.(xarr, α);

### make plot
gr(); display(plot(xarr, dist, xaxis=:log, yaxis=:log));
```

The asymmetric case is similar, and for example we can use the integral definition (12.61)

$$\begin{aligned} f_{\alpha\neq1,\beta,\sigma,\mu}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\, \exp\Big(-\sigma^\alpha|k|^\alpha\big[1-i\beta\,\mathrm{sgn}(k)\tan(\pi\alpha/2)\big]+ik\mu\Big) \\ f_{\alpha=1,\beta,\sigma,\mu}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ikx}\, \exp\Big(-\sigma|k|\big[1+i(2\beta/\pi)\,\mathrm{sgn}(k)\log|k|\big]+ik\mu\Big) \end{aligned} \tag{14.10}$$

as the basis for the numerical method. It is somewhat more efficient to rewrite this integral as

$$\begin{aligned} f_{\alpha\neq1,\beta,\sigma,\mu}(x) &= \frac{1}{\pi}\int_{0}^{\infty} dk\, e^{-\sigma^\alpha k^\alpha}\cos\Big(\beta\tan(\pi\alpha/2)\,\sigma^\alpha k^\alpha - k(x-\mu)\Big) \\ f_{\alpha=1,\beta,\sigma,\mu}(x) &= \frac{1}{\pi}\int_{0}^{\infty} dk\, e^{-\sigma k}\cos\Big((2\beta/\pi)\sigma k\log k + k(x-\mu)\Big) \end{aligned} \tag{14.11}$$

in order to eliminate the imaginary parts of the integral, which must ultimately vanish. This integral (with $\sigma = 1$ and $\mu = 0$) is implemented in the Mathematica code below.

```
f[a_?NumericQ, b_?NumericQ, x_?NumericQ] := (1/Pi) Abs[
    NIntegrate[
     Exp[-Abs[k]^a] Cos[b Tan[Pi a/2] k^a - k x], {k, 0, Infinity},
     Method -> Oscillatory]];
alpha = 3/2;
beta = 1;
Plot[f[alpha, beta, x], {x, -10, 10}, {PlotRange -> All, PlotPoints -> 400}]
```

This code grumbles about numerical convergence problems, yet produces accurate results until the tails have decayed to quite small values.

The main problem with the above methods is the oscillatory nature of the integrals on an unbounded domain. An alternate integral representation of the stable densities can be written[7]

$$f_{\alpha\neq1,\beta}(x) = \begin{cases} \dfrac{\alpha|x|^{1/(\alpha-1)}}{\pi|\alpha-1|}\displaystyle\int_{-\theta_0}^{\pi/2} d\theta\, V_{\alpha,\beta}(\theta)\, \exp\Big(-|x|^{\alpha/(\alpha-1)}\, V_{\alpha,\beta}(\theta)\Big) & (x\neq0) \\ \Gamma\left(\dfrac{\alpha+1}{\alpha}\right)\cos\theta_0\big(\cos\alpha\theta_0\big)^{1/\alpha} & (x=0), \end{cases} \tag{14.12}$$

where

$$\begin{aligned} V_{\alpha\neq1,\beta}(\theta) &= \big(\cos\alpha\theta_0\big)^{1/(\alpha-1)}\left(\frac{\cos\theta}{\sin\alpha(\theta_0+\theta)}\right)^{\alpha/(\alpha-1)}\frac{\cos[\alpha\theta_0+(\alpha-1)\theta]}{\cos\theta} \\ \theta_0 &= \frac{1}{\alpha}\tan^{-1}\Big[\mathrm{sgn}(x)\beta\tan\frac{\pi\alpha}{2}\Big] \end{aligned} \tag{14.13}$$

[7] John P. Nolan, *Univariate Stable Distributions: Models for Heavy Tailed Data* (Springer, 2020), pp. 67-71, in particular Eqs. (3.22) and (3.24), and Theorem 3.3 (ISBN: 9783030529154).

for the $\alpha \neq 1$ cases. In the case of $\alpha = 1$, the integral representation is

$$f_{\alpha=1,\beta}(x) = \begin{cases} \dfrac{1}{2|\beta|}\, e^{-\pi x/2\beta} \displaystyle\int_{-\pi/2}^{\pi/2} d\theta\, V_{\alpha=1,\beta}(\theta)\, \exp\Big(-e^{-\pi x/2\beta}\, V_{\alpha=1,\beta}(\theta)\Big) & (\beta \neq 1) \\ \dfrac{1}{\pi(1+x^2)} & (\beta = 0), \end{cases} \tag{14.14}$$

where

$$V_{\alpha=1,\beta\neq 0}(\theta) = \left(\frac{1+(2\beta/\pi)\theta}{\cos\theta}\right) \exp\left[\frac{\pi}{2\beta}\left(1+\frac{2\beta}{\pi}\theta\right)\tan\theta\right]. \tag{14.15}$$

Note that these integrands may be sharply peaked. Since they have the form $g(\theta)e^{-}g(\theta)$, the integrand maxima may be found by solving $g(\theta) = 1$. The integral can be done in terms of the intervals above and below this solution, to ensure that it is not missed by a numerical quadrature algorithm.

14.3 Fractional Diffusion

Grünwald-Letnikov scheme (see Zoia for refs)

14.4 Eigenfunctions of the Fractional Laplacian

Recalling again the fractional diffusion equation (12.81),

$$\partial_t f(x,t) = \sigma^\alpha \frac{\partial^\alpha}{\partial|x|^\alpha} f(x,t), \tag{14.16}$$

an alternative to studying evolution via direct propagation as in Section 14.3 is to consider evolution as a superposition of eigenfunctions. This is particularly the case if the evolution is on a finite domain, with perfectly absorbing boundaries, as in the problem of escape from an interval (Section 13.1.1). This is analogous to the evolution of a particle in an infinite square well in quantum mechanics as a superposition of eigenfunctions. In terms of ordinary diffusion, this is the generalization of the eigenfunction approach to the escape problem in Section 6.3.3.

For definiteness, we will consider the domain of the evolution to be $x \in [-1,1]$, as in the leapover/escape problem. The eigenfunction $f_\lambda(x)$ with eigenvalue λ of the fractional-derivative operator is defined by

$$-\lambda f_\lambda(x) = \sigma^\alpha \frac{\partial^\alpha}{\partial|x|^\alpha} f_\lambda(x), \qquad \text{(eigenvalue problem, fractional Laplacian)} \tag{14.17}$$

and according to the evolution equation (14.16), the time evolution of an eigenfunction has the form

$$f_\lambda(x,t) = f_\lambda(x,0)\, e^{-\lambda t}. \tag{14.18}$$

The eigenvalues and eigenfunctions of course differ for differing values of α. On a finite interval we expect a countable set of eigenvalues, and so the evolution of an initial distribution $f(x,0)$ can be represented

$$f(x,t) = \sum_n c_n f_{\lambda_n}(x)\, e^{-\lambda_n t}. \tag{14.19}$$

Since we are dealing with eigenfunctions of a Hermitian operator, we can take the eigenfunctions to form an orthonormal basis, with

$$\int_{-1}^{1} dx\, f_{\lambda_m}(x)\, f_{\lambda_n}(x) = \delta_{mn}, \tag{14.20}$$

where we further have real eigenfunctions (and real, positive eigenvalues, as in the $\alpha = 2$ case of normal diffusion). The coefficients appearing in Eq. (14.19) are then given by the projection integral

$$c_n = \int_{-1}^{1} dx\, f(x,0)\, f_{\lambda_n}(x). \tag{14.21}$$

Except in the case $\alpha = 2$, the eigenfunctions and eigenvalues must be computed numerically. However, note that the estimate[8]

$$\lambda_n = \sigma^\alpha \left(\frac{n\pi}{2} - \frac{(2-\alpha)\pi}{8} \right)^\alpha + O(1/n) \tag{14.22}$$

is available, though this is something of a crude approximation. More accurate eigenvalues must be computed numerically.

As an example of the utility of this expansion, the survival probability for the escape problem is given by

$$P_{\text{surv}}(t) = \int_{-1}^{1} dx\, f(x,t) = \sum_n c_n A_n\, e^{-\lambda_n t}, \tag{14.23}$$

where

$$A_n := \int_{-1}^{1} dx\, f_{\lambda_n}(x) \tag{14.24}$$

is the area of the nth eigenfunction. In particular, if λ_1 is the smallest eigenvalue, then its term dominates the others at long times, and thus

$$P_{\text{surv}}(t) \sim c_1 A_1\, e^{-\lambda_1 t} \tag{14.25}$$

gives the asymptotic behavior at long times. This expansion works out just as in the Gaussian case (Section 6.3.3), if we identify $\lambda_n = n^2\pi^2\sigma^2/L^2$, $A_n = [1-(-1)^n]\sqrt{2L}/n\pi$, and $c_n = [1-(-1)^n]/\sqrt{2L}$ for x_0 in the middle of the interval. With $L = 2$ in the present problem, we have the asymptotic decay of the survival probability

$$P_{\text{surv}}(t) \sim \frac{4}{\pi}\, e^{-\pi^2\sigma^2 t/L^2} = \frac{4}{\pi}\, e^{-\pi^2\sigma^2 t/4}, \tag{14.26}$$

with the escape-time distribution having an asymptotic tail of the same form:

$$f_{\text{escape}}(t) = -\frac{d}{dt} P_{\text{surv}}(t) \sim \frac{4\pi\sigma^2}{L^2}\, e^{-\pi^2\sigma^2 t/L^2} = \pi\sigma^2\, e^{-\pi^2\sigma^2 t/4}. \tag{14.27}$$

Note that this is consistent with the previous calculation of the escape-time distribution (6.84) in the Gaussian case, provided we identify $\sigma = 1/\sqrt{2}$ for a Wiener process.

14.4.1 Numerical Diagonalization

To find the eigenfunctions and eigenvalues numerically, a convenient approach is to expand the eigenfunctions in the $\alpha = 2$ eigenfunctions (i.e., in a Fourier series).[9] It is most convenient to treat the even- and odd-parity eigenfunctions separately. The expansion on the even and odd subspaces read

$$f_\lambda(x) = \sum_{n=1}^{\infty} a_n \phi_n(x), \qquad f_\lambda(x) = \sum_{n=1}^{\infty} b_n \varphi_n(x), \tag{14.28}$$

using the eigenfunctions that vanish on the boundaries:

$$\phi_n(x) := \cos\left(\frac{(2n-1)\pi x}{2} \right), \qquad \varphi_n(x) := \sin(n\pi x). \tag{14.29}$$

[8] Mateusz Kwaśnicki, ''Eigenvalues of the fractional Laplace operator in the interval,'' *Journal of Functional Analysis*, **262**, 2379 (2012), (doi: 10.1016/j.jfa.2011.12.004); Bartłomiej Dybiec, Ewa Gudowska-Nowak, Eli Barkai, and Alexander A. Dubkov, ''Lévy flights versus Lévy walks in bounded domains,'' *Physical Review E* **95**, 052102 (2017) (doi: 10.1103/PhysRevE.95.052102).

[9] Elena V. Kirichenko, Piotr Garbaczewski, Vladimir Stephanovich, and Mariusz Żab, ''Lévy flights in an infinite potential well as a hypersingular Fredholm problem,'' *Physical Review E* **95**, 052110 (2016) (doi: 10.1103/PhysRevE.93.052110).

These functions are orthonormal in the sense

$$\int_{-1}^{1} dx\,\phi_m(x)\,\phi_n(x) = \delta_{mn}, \qquad \int_{-1}^{1} dx\,\varphi_m(x)\,\varphi_n(x) = \delta_{mn}, \qquad \int_{-1}^{1} dx\,\phi_m(x)\,\varphi_n(x) = 0. \tag{14.30}$$

Putting the expansions (14.28) into Eq. (14.17), multiplying by the appropriate mth eigenfunction, and integrating over x gives

$$\lambda a_m = -\sigma^\alpha \sum_n \int_{-1}^{1} dx\,\phi_m(x) \frac{\partial^\alpha}{\partial|x|^\alpha} a_n \phi_n(x) =: \sum_n D_{mn} a_n \tag{14.31}$$

in the even-parity case with the analogous form in the odd-parity case. A convenient form for the fractional derivative is the Fourier-space expression (12.82),

$$\frac{\partial^\alpha}{\partial|x|^\alpha} f(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\,\big(-|k|^\alpha\big)\tilde{f}(k)\,e^{ikx} = \frac{1}{\pi}\int_0^\infty dk\,\big(-|k|^\alpha\big)\tilde{f}(k)\cos(kx). \tag{14.32}$$

Using

$$\begin{aligned} \tilde{\phi}_n(k) &= \int_{-1}^{1} dx\,\phi_n(x)\,e^{-ikx} = \operatorname{sinc}[k-(n-1/2)\pi] + \operatorname{sinc}[k+(n-1/2)\pi] \\ \tilde{\varphi}_n(k) &= \int_{-1}^{1} dx\,\varphi_n(x)\,e^{-ikx} = -i\operatorname{sinc}(k-n\pi) + i\operatorname{sinc}(k+n\pi), \end{aligned} \tag{14.33}$$

then in the even case we have the matrix

$$\begin{aligned} D_{mn} &= \int_{-1}^{1} dx\,\phi_m(x) \int_{-\infty}^{\infty} dk\,\frac{\sigma^\alpha|k|^\alpha}{2\pi}\Big(\operatorname{sinc}[k-(n-1/2)\pi] + \operatorname{sinc}[k+(n-1/2)\pi]\Big)e^{ikx} \\ &= \frac{\sigma^\alpha}{2\pi}\int_{-\infty}^{\infty} dk\,|k|^\alpha\Big(\operatorname{sinc}[k-(m-1/2)\pi] + \operatorname{sinc}[k+(m-1/2)\pi]\Big) \\ &\qquad\qquad \times\Big(\operatorname{sinc}[k-(n-1/2)\pi] + \operatorname{sinc}[k+(n-1/2)\pi]\Big), \end{aligned} \tag{14.34}$$

and in the odd case,

$$\begin{aligned} D_{mn} &= \int_{-1}^{1} dx\,\varphi_m(x) \int_{-\infty}^{\infty} dk\,\frac{\sigma^\alpha|k|^\alpha}{2\pi i}\Big(\operatorname{sinc}(k-n\pi) - \operatorname{sinc}(k+n\pi)\Big)e^{ikx} \\ &= \frac{\sigma^\alpha}{2\pi}\int_{-\infty}^{\infty} dk\,|k|^\alpha\Big(\operatorname{sinc}(k-m\pi) - \operatorname{sinc}(k+m\pi)\Big)\Big(\operatorname{sinc}(k-n\pi) - \operatorname{sinc}(k+n\pi)\Big). \end{aligned} \tag{14.35}$$

The integrals here can be computed via numerical quadrature, and the resulting matrix (truncated to some maximum dimension) diagonalized numerically.

A simple code, written in Julia, to implement this algorithm for even-parity eigenfunctions is shown below.

```
using Printf; using QuadGK; using LinearAlgebra

function main()
α = 1; # Lévy index
N = 1200; # max matrix dimension

# D matrix elements for diagonalization
#   Note julia's sinc(x) is sin(π*x)/(π*x)
@inline sincn(x) = sinc(x/π);
function evenD(n, m)
  integ, err = quadgk(
    k -> abs(k)^α * ( sincn(k-(m-0.5)*π) + sincn(k+(m-0.5)*π) )
```

```
                * ( sincn(k-(n-0.5)*π) + sincn(k+(n-0.5)*π) ),
    0, Inf, rtol=1e-6);
  return integ/π;
end #function evenD

D = zeros(Float64, N, N);
for m=1:N
  for n=1:m
    D[m,n] = evenD(m,n); D[n,m] = D[m,n];
  end #for n
end #for m

# get sorted eigenvalues and corresponding eigenvectors
eig=eigen(D);
λ=sort(eig.values);
y=eig.vectors[:, sortperm(eig.values)];
@printf("α = %f, smallest eigenvalue = %f\n", α, λ[1]);
@printf("c(0) for smallest eigenvalue = %f\n", abs(sum(y[:,1])));

end; main()
```

However, note that the integrals in (14.34) are slowly decaying in k, and not even absolutely convergent. They are thus fairly difficult to handle numerically; the above code as written takes almost 7 hours to execute on 8 cores on an i9-9900ks computer.

14.4.2 Convolution Algorithm

An alternate approach is to return to the integral definition of the D matrix:

$$D_{mn} = -\sigma^\alpha \int_{-1}^{1} dx\, \phi_m(x) \frac{\partial^\alpha}{\partial|x|^\alpha} \phi_n(x). \tag{14.36}$$

Using the (divergent) convolution expression (12.85)

$$\frac{\partial^\alpha}{\partial|x|^\alpha}\, \phi_n(x) = A_\alpha \int_{-1}^{1} dx'\, \frac{\cos[(n-1/2)\pi x']}{|x-x'|^{\alpha+1}} \tag{14.37}$$

for the fractional Laplacian, where for brevity the numerical coefficient has the shorthand

$$A_\alpha := \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}. \tag{14.38}$$

Integration of Eq. (12.85) by parts gives

$$\frac{\partial^\alpha}{\partial|x|^\alpha}\, \phi_n(x) = \frac{\pi_n A_\alpha}{\alpha} \int_{-1}^{1} dx'\, \frac{\sin \pi_n x'}{(x-x')^\alpha}, \tag{14.39}$$

where here, the fractional power $(x-x')^{-\alpha}$ is shorthand for $\mathrm{sgn}(x-x')\,|x-x'|^{-\alpha}$, and $\pi_n := (n-1/2)\pi$. This expression is still divergent, however; proceeding by changing variables,

$$\begin{aligned}
\frac{\partial^\alpha}{\partial|x|^\alpha}\, \phi_n(x) &= -\frac{\pi_n A_\alpha}{\alpha} \int_{-1}^{1} dx'\, \frac{\sin \pi_n x'}{(x+x')^\alpha} \\
&= -\frac{\pi_n A_\alpha}{\alpha} \int_{x-1}^{x+1} dx'\, \frac{\sin[\pi_n(x'-x)]}{(x')^\alpha} \\
&= \frac{\pi_n A_\alpha}{\alpha} \left[\sin \pi_n x \int_{x-1}^{x+1} dx'\, \frac{\cos \pi_n x'}{(x')^\alpha} - \cos \pi_n x \int_{x-1}^{x+1} dx'\, \frac{\sin \pi_n x'}{(x')^\alpha} \right] \\
&= \frac{\pi_n A_\alpha}{\alpha} \left[\sin \pi_n x \left(\int_{x-1}^{x+1} dx'\, \frac{\cos \pi_n x' - 1}{(x')^\alpha} + \int_{x-1}^{x+1} \frac{dx'}{(x')^\alpha} \right) - \cos \pi_n x \int_{x-1}^{x+1} dx'\, \frac{\sin \pi_n x'}{(x')^\alpha} \right],
\end{aligned} \tag{14.40}$$

where we have used $\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$. There are now three integrals; the first has a removable singularity at $x'=0$, the second integral has a singularity that can be handled analytically in the sense of a Cauchy principal value:

$$\fint_{x-1}^{x+1} \frac{dx'}{(x')^\alpha} = \begin{cases} \dfrac{1}{(1-\alpha)}\left[(x+1)^{1-\alpha} - (x-1)^{1-\alpha}\right] & (\alpha \neq 1) \\ \log\left|\dfrac{x+1}{x-1}\right| & (\alpha = 1). \end{cases} \tag{14.41}$$

The third integral in Eq. (14.40) is still singular at $x'=0$, but the singularity can be integrated (generally speaking, the integral should be split into two parts, with the singularity on the boundary of each part, for numerical algorithms to handle it efficiently).

While the algorithm above can proceed simply by computing the other integral in Eq. (14.36), the process may be accelerated considerably by noting that this equation has the form of a Fourier cosine transform. Specifically, comparing to the discrete cosine transform (specifically, the `REDFT11` algorithm[10]), which has the form

$$Y_m = 2\sum_{n=0}^{n_x-1} X_n \cos[\pi(n+1/2)(m+1/2)/n_x], \tag{14.42}$$

we can see that Eq. (14.36) has essentially the same form. Recalling the convention for discrete Fourier transforms, we chooise choose N_x to be even integer, so that the grid spacing in x is $\Delta x = 2/N_x$, with grid values $0, \Delta x, \ldots, (N_x-1)\Delta x$. The function to be transformed is discretely sampled, with samples $X_n = X(n\Delta x)$. Since the function is assumed to be periodic (with period 2), $X_{N_x} = X_0$; furthermore the sampled function is assumed to be even, so only the first n_x samples are stored, where $N_x = 2(n_x-1)$ or $n_x = N_x/2+1$. (Note that X_{n_x} is cut off anyway by the zero of the cosine.) So, the samples of the function (14.40) are calculated on this grid (the negative part of the grid is redundant); a discrete cosine transform gives the nth column of the D_{mn} matrix, which has dimension $n_x \times n_x$.

The odd case is analogous. The matrix to diagonalize is

$$D_{mn} = -\sigma^\alpha \int_{-1}^{1} dx\, \varphi_m(x) \frac{\partial^\alpha}{\partial|x|^\alpha}\varphi_n(x). \tag{14.43}$$

where the partial derivative may be written in regularized form as

$$\begin{aligned} \frac{\partial^\alpha}{\partial|x|^\alpha}\varphi_n(x) &= A_\alpha \int_{-1}^{1} dx' \frac{\sin(n\pi x')}{|x-x'|^{\alpha+1}} \\ &= -\frac{\pi_n A_\alpha}{\alpha}\int_{-1}^{1} dx' \frac{\cos\pi_n x'}{(x-x')^\alpha} \\ &= -\frac{\pi_n A_\alpha}{\alpha}\int_{-1}^{1} dx' \frac{\cos\pi_n x'}{(x+x')^\alpha} \\ &= -\frac{\pi_n A_\alpha}{\alpha}\int_{x-1}^{x+1} dx' \frac{\cos[\pi_n(x'-x)]}{(x')^\alpha} \\ &= -\frac{\pi_n A_\alpha}{\alpha}\left[\cos\pi_n x \int_{x-1}^{x+1} dx' \frac{\cos\pi_n x'}{(x')^\alpha} + \sin\pi_n x \int_{x-1}^{x+1} dx' \frac{\sin\pi_n x'}{(x')^\alpha}\right] \\ &= -\frac{\pi_n A_\alpha}{\alpha}\left[\cos\pi_n x \left(\int_{x-1}^{x+1} dx' \frac{\cos\pi_n x' - 1}{(x')^\alpha} + \fint_{x-1}^{x+1} \frac{dx'}{(x')^\alpha}\right) + \sin\pi_n x \int_{x-1}^{x+1} dx' \frac{\sin\pi_n x'}{(x')^\alpha}\right], \end{aligned} \tag{14.44}$$

where now $\pi_n = n\pi$, and we used $\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$. The integral (14.43) now has the form of a Fourier sine transform; specifically it is comparable to the `RODFT00` algorithm[11]),

$$Y_m = 2\sum_{n=0}^{n_x-1} X_n \sin[\pi(n+1)(m+1)/(n_x+1)], \tag{14.45}$$

[10] FFTW documentation, `http://www.fftw.org/fftw3_doc/1d-Real_002deven-DFTs-_0028DCTs_0029.html`.
[11] FFTW documentation, `http://www.fftw.org/fftw3_doc/1d-Real_002dodd-DFTs-_0028DSTs_0029.html`.

where $N_x = 2(n_x + 1)$. Again, this transform will yield the nth column of the D_{mn} matrix, which has dimension $n_x \times n_x$.

Another simple code, written in Julia, to implement the even-parity version of the algorithm is shown below.

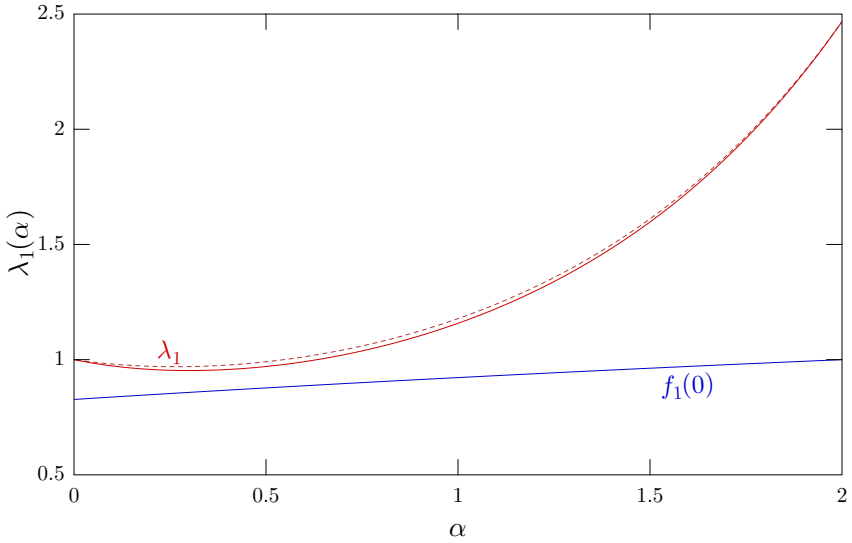
```
using Printf; using QuadGK; using LinearAlgebra;
using SpecialFunctions; using FFTW; function main()

α = 1; # Lévy index
nx = 1200; # matrix dimension
tol = 1e-6; # integration tolerance
mxeval = 10^7; # max quadgk evaluations

function pow(x,n)  # equiv to abs(x)^n * sign(x);
  candidate = abs(x)^n;
  if (candidate>1e150)
    return 1e150*sign(x) # clamp overflow
  else
    return candidate*sign(x);
  end
end #function pow

function cospim1xpma(nm,x) # (cos(nm*pi*x) - 1)*x^-α, accurate near zeros
  nmx = nm*x;
  if (abs(nmx)>.03)
    return (cospi(nmx)-1)*pow(x,-α);
  else # near x=0, use Taylor expansion and explicit cancellation
    pn = π*nm; pn2 = pn*pn;
    x2 = x*x; x2a = pow(x,2-α);
    return -pn2*x2a/2*(1 - pn2*x2/12*(1 - pn2*x2/30));
  end
end #function cospim1xpma

function sinpixpma(nm,x) # sin(nm*pi*x)*x^-α, accurate for small x
  if (abs(nm*x)>.03)
    return sinpi(nm*x)*pow(x,-α);
  else
    pn = π*nm; pn2 = pn*pn;
    x2 = x*x; xa = pow(abs(x),1-α);
    return pn*xa*(1 - pn2*x2/6*(1 - pn2*x2/20));
  end
end #function sinpixpma

function cpvthing(x)
  if (α == 1)
    return log(abs((1+x)/(1-x)));
  else
    return (pow(1+x,1-α)-pow(1-x,1-α))/(1-α);
  end
end #function cpvthing

function int1(xl,xh,nm) # nm = n-0.5
  integ, err = quadgk(
    xp -> cospim1xpma(nm,xp),
    xl, xh, rtol=tol, maxevals=mxeval);
  return integ;
end #function int1
```

```
function int2(xl,xh,nm) # nm = n-0.5
  if ((xl==0) | (xh==0)) # handle integral from 0 to dx analytically
    dx = xh-xl;
    pn = π*nm; pn2 = pn*pn; dx2 = dx*dx;
    return pn*dx^(2-α)*(1/(2-α)-pn2*dx2/6*(1/(4-α)+pn2*dx2/20/(6-α)));
  else
    integ, err = quadgk(
      xp -> sinpixpma(nm,xp),
      xl, xh, rtol=tol, maxevals=mxeval);
    return integ;
  end
end #function int2

A = gamma(α+1)*sin(π*α/2)/π;
dx = 1/nx;
xg = (collect(0:(nx-1)).+0.5)*dx; # x grid, dx/2, 3dx/2, ..., (nx-1/2)*dx
xd = vcat(0, xg, 1, (xg.+1)); # grid of integral boundaries, length 2nx+2
pfft = FFTW.plan_r2r!(xg, FFTW.REDFT11)

# D matrix elements for diagonalization
D = zeros(Float64, nx, nx);
for n = 1:nx
  fd = zeros(Float64, nx); int1p=fd; int1m=fd; int2p=fd; int2m=fd;
  int1a = zeros(Float64, 2*nx+1); int2a = int1a;
  nm = n-0.5;
  # compute int1 and int2 over intervals from x to x+dx, for x=1 to 2-dx
  #   and from x-dx to x for x=0 to -1+dx
  int1a = broadcast(int1, xd[1:(2*nx+1)], xd[2:(2*nx+2)], nm);
  int2a = broadcast(int2, xd[1:(2*nx+1)], xd[2:(2*nx+2)], nm);
  int1a = cumsum(int1a); int2a = cumsum(int2a); # now from x=0 to xg
  # exploit (anti)symmetry to get integral sections <0 or >0
  int1p = int1a[(nx+2):(2*nx+1)]; int2p = int2a[(nx+2):(2*nx+1)];
  int1m = -int1a[nx:-1:1];    int2m =  int2a[nx:-1:1];
  # now fractional derivative array
  @.fd = sinpi(nm*xg)*(int1p+int1m+cpvthing(xg)) - cospi(nm*xg)*(int2p+int2m);
  pfft * fd;
  fd *= -π*nm*A/(α*nx); # 1/nx is the dx in the integral
  D[:,n] = fd;
end# for n

# get sorted eigenvalues and corresponding eigenvectors
D = 0.5*(D + transpose(D)); # symmetrize, because asymmetry = discretization err
eig=eigen(D);
λ=sort(eig.values);
y=eig.vectors[:, sortperm(eig.values)];
@printf("α = %f, smallest eigenvalue = %f\n", α, λ[1]);
@printf("c(0) for smallest eigenvalue = %f\n", abs(sum(y[:,1])));

end; main()
```

As written, this algorithm runs much more quickly than the previous one, finishing in about 2 seconds on the same computer (with multithreading, and with about 75% of that execution time being startup overhead). The code is considerably more complicated, but the integrals are performed in many small chunks which are assembled to avoid redundant integration. At this matrix size, about 5 digits of accuracy are obtained, however; it is better to run with `nx=10^4`.

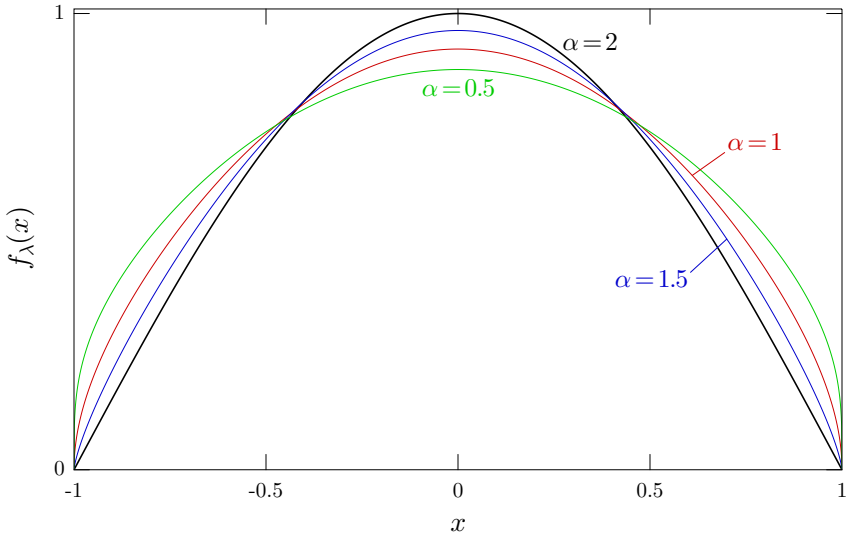
14.4.3 Numerical Results

The results of the above calculation are shown below, using `nx=10^4`. The λ_1 curve is the smallest eigenvalue (the "ground-state energy"), which varies from 1 in the limit $\alpha \longrightarrow 0$ to $\pi^2/4$ in the Gaussian limit $\alpha = 2$ [cf. Eq. (6.74) with $L = \sqrt{2}$]. The accompanying dashed line indicates the value of the estimate (14.22).



The curve labeled $f_1(0)$ is the value of the corresponding eigenfunction at $x = 0$. This varies quite slowly, rising to the expected value of 1 at $\alpha = 2$, where the eigenfunctions ϕ_n in Eq. (14.31) apply. Assuming an initial condition localized at $x = 0$, $f_1(0)$ gives the fraction of the total population associated with the λ_1 eigenvalue.

The eigenfunctions corresponding to λ_1 are shown in the plot below for various α, with the usual cosinusoidal form for $\alpha = 2$.



Note that the function becomes steeper at the boundaries as α decreases, and generically, the eigenfunctions for $\alpha \neq 2$ decay nonlinearly to zero at the boundaries (not just the "ground-state" eigenfunctions. It is conjectured[12] that the decay near the boundaries vanish has the form $(1 \pm x)^{\alpha/2}$.

[12] The conjecture is based on numerical evidence and an analytic approximation at $\alpha = 1$. See Elena V. Kirichenko *et al.*, *op. cit.*

Index