8

Addition of Random Variables and Brownian Motion

In the following chapters we will be interested in the properties of sums (and products) of random matrices. Before embarking on this relatively new field, the present chapter will quickly review some classical results concerning sums of random *scalars*, and the corresponding continuous time limit that leads to the Brownian motion and stochastic calculus.

8.1 Sums of Random Variables

Let us thus consider $X = X_1+X_2$ where X_1 and X_2 are two random variables, independent, and distributed according to, respectively, $P_1(x_1)$ and $P_2(x_2)$. The probability that X is equal to x (to within $\mathrm{d}x$) is given by the sum over all combinations of x_1 and x_2 such that $x_1 + x_2 = x$, weighted by their respective probabilities. The variables X_1 and X_2 being independent, the joint probability that $X_1 = x_1$ and $X_2 = x - x_1$ is equal to $P_1(x_1)P_2(x - x_1)$, from which one obtains

$$P^{(2)}(x) = \int P_1(x')P_2(x - x')\,\mathrm{d}x'. \tag{8.1}$$

This equation defines the *convolution* between P_1 and P_2, which we will write $P^{(2)} = P_1 \star P_2$. The generalization to the sum of N independent random variables is immediate. If $X = X_1 + X_2 + \cdots + X_N$ with X_i distributed according to $P_i(x_i)$, the distribution of X is obtained as

$$P^{(N)}(x) = \int \prod_{i=1}^{N} \mathrm{d}x_i\, P_1(x_1)P_2(x_2)\ldots P_N(x_N)\delta\left(x - \sum_{i=1}^{N} x_i\right), \tag{8.2}$$

where $\delta(.)$ is the Dirac delta function. The analytical or numerical manipulations of Eqs. (8.1) and (8.2) are much eased by the use of Fourier transforms, for which convolutions become simple products. The equation $P^{(2)}(x) = [P_1 \star P_2](x)$, reads, in Fourier space,

$$\varphi^{(2)}(k) = \int \mathrm{e}^{\mathrm{i}k(x-x'+x')} \int P_1(x')P_2(x - x')\,\mathrm{d}x'\,\mathrm{d}x \equiv \varphi_1(k)\varphi_2(k), \tag{8.3}$$

https://doi.org/10.1017/9781108768900.011 Published online by Cambridge University Press

where $\varphi(k)$ denotes the Fourier transform of the corresponding probability density $P(x)$. It is often called its characteristic (or generating) function. Since the characteristic functions multiply, their logarithms add, i.e. the function $H(k)$ defined below is additive:

$$H(k) := \log \varphi(k) = \log \mathbb{E}[\mathrm{e}^{\mathrm{i}kX}]. \tag{8.4}$$

It allows one to recovers its so called *cumulants* c_n (provided they are finite) through

$$c_n := (-\mathrm{i})^n \left. \frac{\mathrm{d}^n}{\mathrm{d}z^n} H(k) \right|_{k=0}. \tag{8.5}$$

The cumulants c_n are polynomial combinations of the moments m_p with $p \leq n$. For example $c_1 = m_1$ is the mean of the distribution and $c_2 = m_2 - m_1^2 = \sigma^2$ its variance. It is clear that the mean of the sum of two random variables (independent or not) is equal to the sum of the individual means. The mean is thus additive under convolution. The same is true for the variance, but only for independent variables.

More generally, from the additive property of $H(k)$ all the cumulants of two independent distributions simply add. The additivity of cumulants is a consequence of the linearity of derivation. The cumulants of a given law convoluted N times with itself thus follow the simple rule $c_{n,N} = N c_{n,1}$, where the $\{c_{n,1}\}$ are the cumulants of the elementary distribution P_1.

An important case is when P_1 is a Gaussian distribution,

$$P_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}, \tag{8.6}$$

such that $\log \varphi_1(k) = \mathrm{i}mk - \sigma^2 k^2/2$. The Gaussian distribution is such that all cumulants of order ≥ 3 are zero. This property is clearly preserved under convolution: the sum of Gaussian random variables remains Gaussian. Conversely, one can always write a Gaussian variable as a sum of an arbitrary number of Gaussian variables: Gaussian variables are infinitely divisible.

In the following, we will consider infinitesimal Gaussian variables, noted $\mathrm{d}B$, such that $\mathbb{E}[\mathrm{d}B] = 0$ and $\mathbb{E}[\mathrm{d}B^2] = \mathrm{d}t$, where $\mathrm{d}t \to 0$ is an infinitesimal quantity, which we will interpret as an infinitesimal time increment. In other words, $\mathrm{d}B$ is a mean zero Gaussian random variable which has fluctuations of order $\sqrt{\mathrm{d}t}$.

8.2 Stochastic Calculus

8.2.1 Brownian Motion

The starting point of stochastic calculus is the Brownian motion (also called Wiener process) X_t, which is a Gaussian random variable of mean μt and variance $\sigma^2 t$. From the infinite divisibility property of Gaussian variables, one can always write

$$X_{t_k} = \sum_{\ell=0}^{k-1} \mu \delta t + \sum_{\ell=0}^{k-1} \sigma \delta B_\ell, \tag{8.7}$$

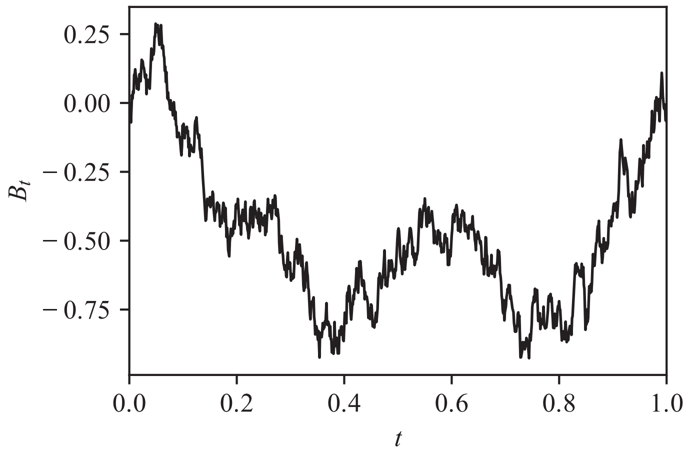

Figure 8.1 An example of Brownian motion.

where $t_k = kt/N$, $0 \leq k \leq N$, $\delta t = T/N$ and $\delta B_\ell \sim \mathcal{N}(0, \delta t)$ for each ℓ. By construction $X(t_N) = X_t$. In the limit $N \to \infty$, we have $\delta t \to \mathrm{d}t$, $\delta B_k \to \mathrm{d}B$ and X_{t_k} becomes a continuous time process with

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}B_t, \qquad X_0 = 0, \tag{8.8}$$

where $\mathrm{d}B_t$ are independent, infinitesimal Gaussian variables as defined above (see Fig. 8.1). The process X_t is continuous but nowhere differentiable. Note that X_t and $X_{t'}$ are not independent but their increments are, i.e. $X_{t'}$ and $X_t - X_{t'}$ are independent whenever $t' < t$.

Note the convention that X_{t_k} is built from *past increments* δB_ℓ for $\ell < k$ but does not include δB_k. This convention is called the Itô prescription.[1] Its main advantage is that X_t is independent of the equal-time $\mathrm{d}B_t$, but this comes at a price: the usual chain rule for differentiation has to be corrected by the so-called Itô term, which we now discuss.

8.2.2 Itô's Lemma

We now study the behavior of functions $F(X_t)$ of a Wiener process X_t. Because $\mathrm{d}B^2$ is of order $\mathrm{d}t$, and not $\mathrm{d}t^2$, one has to be careful when evaluating derivatives of functions of X_t.

Given a twice differentiable function $F(.)$, we consider the process $F(X_t)$. Reverting for a moment to a discretized version of the process, one has

$$F(X(t + \delta t)) = F(X_t) + \delta X\, F'(X_t) + \frac{(\delta X)^2}{2} F''(X_t) + o(\delta t), \tag{8.9}$$

[1] In the Stratonovich prescription, half of δB_k contributes to X_{t_k}. In this prescription, the Itô lemma is not needed, i.e. the chain rule applies without any correction term, but the price to pay is a correlation between X_t and $\mathrm{d}B(t)$. We will not use the Stratonovich prescription in this book.

https://doi.org/10.1017/9781108768900.011 Published online by Cambridge University Press

where

$$\delta X = \mu \delta t + \sigma \delta B, \tag{8.10}$$

and

$$(\delta X)^2 = \mu^2 (\delta t)^2 + \sigma^2 \delta t + \sigma^2 \left[(\delta B)^2 - \delta t\right] + 2\mu\sigma \delta t \delta B. \tag{8.11}$$

The random variable $(\delta B)^2$ has mean $\sigma^2 \delta t$ so the first and last terms are clearly $o(\delta t)$ when $\delta t \to 0$. The third term has standard deviation given by $\sqrt{2}\sigma^2 \delta t$; so the third term is also of order δt but of zero mean. It is thus a random term much like δB, but much smaller since δB is of order $\sqrt{\delta t} \gg \delta t$. The Itô lemma is a precise mathematical statement that justifies why this term can be neglected to first order in δt. Hence, letting $\delta t \to \mathrm{d}t$, we get

$$\mathrm{d}F_t = \frac{\partial F}{\partial X} \mathrm{d}X_t + \frac{\sigma^2}{2} \frac{\partial^2 F}{\partial^2 X} \mathrm{d}t. \tag{8.12}$$

When compared to ordinary calculus, there is a correction term – the Itô term – that depends on the second order derivative of F.

More generally, we can consider a general Itô process where μ and σ themselves depend on X_t and t, i.e.

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}B_t. \tag{8.13}$$

Then, for functions $F(X, t)$ that may have an explicit time dependence, one has

$$\mathrm{d}F_t = \frac{\partial F}{\partial X} \mathrm{d}X_t + \left[\frac{\partial F}{\partial t} + \frac{\sigma^2(X_t, t)}{2} \frac{\partial^2 F}{\partial^2 X}\right] \mathrm{d}t. \tag{8.14}$$

Itô's lemma can be extended to functions of several stochastic variables. Consider a collection of N independent stochastic variables $\{X_{i,t}\}$ (written vectorially as $\mathbf{X}_t$) and such that

$$\mathrm{d}X_{i,t} = \mu_i(\mathbf{X}_t, t)\mathrm{d}t + \mathrm{d}W_{i,t}, \tag{8.15}$$

where $\mathrm{d}W_{i,t}$ are Wiener noises such that

$$\mathbb{E}\left[\mathrm{d}W_{i,t}\mathrm{d}W_{j,t}\right] := C_{ij}(\mathbf{X}_t, t)\mathrm{d}t. \tag{8.16}$$

The vectorial form of Itô's lemma states that the time evolution of a function $F(\mathbf{X}_t, t)$ is given by the sum of three contributions:

$$\mathrm{d}F_t = \sum_{i=1}^{N} \frac{\partial F}{\partial X_i} \mathrm{d}X_{i,t} + \left[\frac{\partial F}{\partial t} + \sum_{i,j=1}^{N} \frac{C_{ij}(\mathbf{X}_t, t)}{2} \frac{\partial^2 F}{\partial X_i \partial X_j}\right] \mathrm{d}t. \tag{8.17}$$

The formula simplifies when all the Wiener noises are independent, in which case the Itô term only contains the second derivatives $\partial^2 F / \partial^2 X_i$.

https://doi.org/10.1017/9781108768900.011 Published online by Cambridge University Press

8.2.3 Variance as a Function of Time

As an illustration of how to use Itô's formula let us recompute the time dependent variance of X_t. Assume $\mu = 0$ and choose $F(X) = X^2$. Applying Eq. (8.12), we get that

$$\mathrm{d}F_t = 2X_t\mathrm{d}X_t + \sigma^2\mathrm{d}t \Rightarrow F(X_t) = 2\int_0^t \sigma X_s \mathrm{d}B_s + \sigma^2 t. \tag{8.18}$$

In order to take the expectation value of this equation, we scrutinize the term $\mathbb{E}[X_s\mathrm{d}B_s]$. As alluded to above, the random infinitesimal element $\mathrm{d}B_s$ does not contribute to X_s, which only depends on $\mathrm{d}B_{s'<s}$. Therefore $\mathbb{E}[X_s\mathrm{d}B_s] = 0$, and, as expected,

$$\mathbb{E}[X_t^2] = \mathbb{E}[F(X_t)] = \sigma^2 t. \tag{8.19}$$

The Brownian motion has a variance from the origin that grows linearly with time. The same result can of course be derived directly from the integrated form $X_t = \sigma B_t$, where B_t is a Gaussian random number of variance equal to t.

8.2.4 Gaussian Addition

Itô's lemma can be used to compute a special case of the law of addition of independent random variables, namely when one of the variables is Gaussian. Consider the random variable $Z = Y + X$, where Y is some random variable, and X is an independent Gaussian ($X \sim \mathcal{N}(\mu, \sigma^2)$). The law of Z is uniquely determined by its characteristic function:

$$\varphi(k) := \mathbb{E}[\mathrm{e}^{\mathrm{i}kZ}]. \tag{8.20}$$

We now let $Z \to Z_t$ be a Brownian motion with $Z_0 = Y$:

$$\mathrm{d}Z_t = \mu\mathrm{d}t + \sigma\mathrm{d}B_t, \quad Z_0 = Y. \tag{8.21}$$

Note that $Z_{t=1}$ has the same law as Z. The idea is now to study the function $F(Z_t) := \mathrm{e}^{\mathrm{i}kZ_t}$ using Itô's lemma, Eq. (8.14). Hence,

$$\mathrm{d}F_t = \mathrm{i}k\mathrm{e}^{\mathrm{i}kZ_t}\mathrm{d}Z_t - \frac{k^2\sigma^2}{2}\mathrm{e}^{\mathrm{i}kZ_t}\mathrm{d}t = \left(\mathrm{i}k\mu F - \frac{k^2\sigma^2}{2}F\right)\mathrm{d}t + \mathrm{i}kF\mathrm{d}B_t. \tag{8.22}$$

Taking the expectation value, writing $\varphi_t(k) = \mathbb{E}[F(t)]$, and noting that the differential d is a linear operator and therefore commutes with the expectation value, we obtain

$$\mathrm{d}\varphi_t(k) = \left(\mathrm{i}k\mu - \frac{k^2\sigma^2}{2}\right)\varphi_t(k)\mathrm{d}t, \tag{8.23}$$

or

$$\frac{1}{\varphi_t(k)}\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t(k) = \frac{\mathrm{d}}{\mathrm{d}t}\log\left(\varphi_t(k)\right) = \left(\mathrm{i}k\mu - \frac{k^2\sigma^2}{2}\right). \tag{8.24}$$

https://doi.org/10.1017/9781108768900.011 Published online by Cambridge University Press

From its solution at $t = 1$, we get

$$\log(\varphi_1(k)) = \log(\varphi_0(k)) + ik\mu - \frac{k^2\sigma^2}{2}. \tag{8.25}$$

Recognizing the last two terms in the right hand side as the characteristic function of a Gaussian variable, we recover the fact that the log-characteristic function is additive under the addition of independent random variables. Although the result is true in general, the calculation above using stochastic calculus is only valid if one of the random variable is Gaussian.

8.2.5 The Langevin Equation

We would like to construct a stochastic process for a variable X_t such that in the steady-state regime the values of X_t are drawn from a given probability distribution $P(x)$. To build our stochastic process, let us first consider the simple Brownian motion with unit variance per unit time:

$$\mathrm{d}X_t = \mathrm{d}B_t. \tag{8.26}$$

As revealed by Eq. (8.19) the variance of X_t grows linearly with time and the process never reaches a stationary state. To make it stationary we need a mechanism to limit the variance of X_t. We cannot 'subtract' variance but we can reduce X_t by scaling. If at every infinitesimal time step we replace $X_{t+\mathrm{d}t}$ by $X_{t+\mathrm{d}t}/\sqrt{1+\mathrm{d}t}$, the variance of X_t will remain equal to unity. We also know that the distribution of X_t is Gaussian (if the initial condition is Gaussian or constant). With this extra rescaling, X_t is still Gaussian at every step, so clearly this will describe the stationary state of our rescaled process. As a stochastic differential equation, we have, neglecting terms of order $(\mathrm{d}t)^{3/2}$

$$\mathrm{d}X_t = \mathrm{d}B_t + \frac{X_t}{\sqrt{1+\mathrm{d}t}} - X_t = \mathrm{d}B_t - \frac{1}{2}X_t\mathrm{d}t. \tag{8.27}$$

This stationary version of the random walk is the Ornstein–Uhlenbeck process (see Fig. 8.2). A physical interpretation of this equation is that of a particle located at X_t moving in a viscous medium subjected to random forces $\mathrm{d}B_t/\mathrm{d}t$ and a deterministic harmonic force ("spring") $-X_t/2$. The viscous medium is such that velocity (and not acceleration) is proportional to force.

We would like to generalize the above formalism to generate any distribution $P(x)$ for the distribution of X in the stationary state. One way to do so is to change the linear force $-X_t/2$ to a general non-linear force $F(X_t) := -V'(X_t)/2$, where we have written the force as the derivative of a potential V and introduced a factor of 2 which will prove to be convenient. If the potential is convex, the force will drive the particle towards the minimum of the potential while the noise $\mathrm{d}B_t$ will drive the particle away. We expect that this system will reach a steady state. Our stochastic equation is now

$$\mathrm{d}X_t = \mathrm{d}B_t + F(X_t)\mathrm{d}t. \tag{8.28}$$

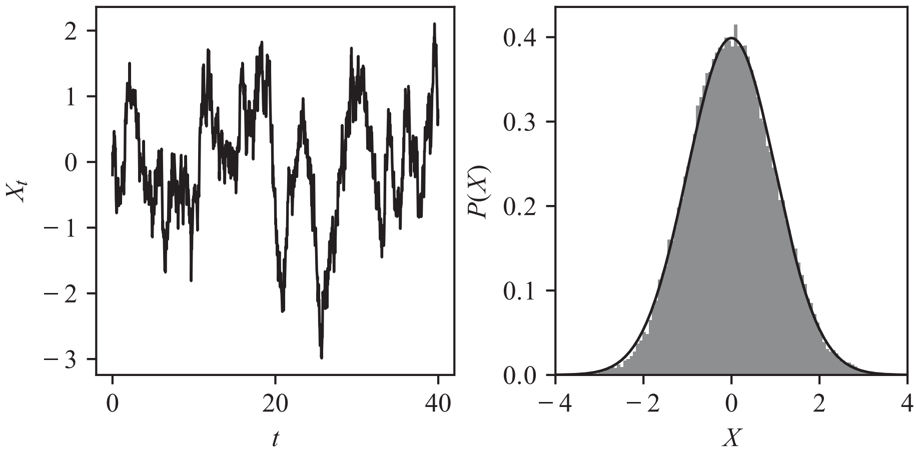

Figure 8.2 (left) A simulation of the Langevin equation for the Ornstein–Uhlenbeck process (8.27) with 50 steps per unit time. Note that the correlation time is $\tau_c = 2$ here, so excursions away from zero typically take 2 time units to mean-revert back to zero. Farther excursions take longer to come back. (right) Histogram of the values of X_t for the same process simulated up to $t = 2000$ and comparison with the normal distribution. The agreement varies from sample to sample as a rare far excursion can affect the sample distribution even for $t = 2000$.

What is the distribution of X_t in the steady state? To find out, let us consider an arbitrary test function $f(X_t)$ and see how it behaves in the steady state. Using Itô's lemma, Eq. (8.12), we have

$$\mathrm{d}f(X_t) = f'(X_t)\left[\mathrm{d}B_t - \frac{1}{2}V'(X_t)\mathrm{d}t\right] + \frac{1}{2}f''(X_t)\mathrm{d}t. \tag{8.29}$$

Taking the expectation value on both sides and demanding that $\mathrm{d}\mathbb{E}[f_t]/\mathrm{d}t = 0$ in the steady state we find

$$\mathbb{E}\left[f'(X_t)V'(X_t)\right] = \mathbb{E}\left[f''(X_t)\right]. \tag{8.30}$$

This must be true for any function f. In order to infer the corresponding stationary distribution $P(x)$, let us write $h(x) = f'(x)$ and write these expectation values as

$$\int h(x)V'(x)P(x)\mathrm{d}x = \int h'(x)P(x)\mathrm{d}x. \tag{8.31}$$

Since we want to relate an integral of h to one of h' we should use integration by parts:

$$\int h'(x)P(x)\mathrm{d}x = -\int h(x)P'(x)\mathrm{d}x = -\int h(x)\frac{P'(x)}{P(x)}P(x)\mathrm{d}x. \tag{8.32}$$

Since Eq. (8.31) is true for any function $h(x)$ we must have

https://doi.org/10.1017/9781108768900.011 Published online by Cambridge University Press

$$V'(x) = -\frac{P'(x)}{P(x)} \quad \Rightarrow \quad P(x) = Z^{-1}\exp[-V(x)], \tag{8.33}$$

where Z is an integration constant that fixes the normalization of the law $P(x)$.

To recapitulate, given a probability density $P(x)$, we can define a potential $V(x) = -\log P(x)$ (up to an irrelevant additive constant) and consider the stochastic differential equation

$$\mathrm{d}X_t = \mathrm{d}B_t - \frac{1}{2}V'(X_t)\mathrm{d}t. \tag{8.34}$$

The stochastic variable X_t will eventually reach a steady state. In that steady state the law of X_t will be given by $P(x)$. Equation (8.34) is called the Langevin equation. The strength of the Langevin equation is that it allows one to replace the average over the probability $P(x)$ by a sample average over time of a stochastic process.[2] Any rescaling of time $t \to \sigma^2 t$ would yield a Langevin equation with the same stationary state:

$$\mathrm{d}X_t = \sigma\mathrm{d}B_t - \frac{\sigma^2}{2}V'(X_t)\mathrm{d}t. \tag{8.35}$$

We have learned another useful fact from Eq. (8.30): the random variable $V'(X)$ acts as a derivative with respect to X under the expectation value. In that sense $V'(X)$ can be considered the conjugate variable to X.

It is very straightforward to generalize our one-dimensional Langevin equation to a set of N variables $\{X_i\}$ that are drawn from the joint law $P(\mathbf{x}) = Z^{-1}\exp[-V(\mathbf{x})]$. We get

$$\mathrm{d}X_i = \sigma\mathrm{d}B_i + \frac{\sigma^2}{2}\frac{\partial}{\partial X_i}\log P(\mathbf{x})\,\mathrm{d}t, \tag{8.36}$$

where we have dropped the subscript t for clarity.

Exercise 8.2.1 Langevin equation for Student's t-distributions

The family of Student's t-distributions, parameterized by the tail exponent μ, is given by the probability density

$$P_\mu(x) = Z_\mu^{-1}\left(1+\frac{x^2}{\mu}\right)^{-\frac{\mu+1}{2}} \quad \text{with } Z_\mu^{-1} = \frac{\Gamma\left(\frac{\mu+1}{2}\right)}{\sqrt{\mu\pi}\,\Gamma\left(\frac{\mu}{2}\right)}. \tag{8.37}$$

(a) What is the potential $V(x)$ and its derivative $V'(x)$ for these laws?

(b) Using Eq. (8.30), show that for a t-distributed variable x we have

$$\mathbb{E}\left[\frac{x^2}{x^2+\mu}\right] = \frac{1}{1+\mu}. \tag{8.38}$$

[2] A process for which the time evolution samples the entire set of possible values according to the stationary probability is called ergodic. A discussion of the condition for ergodicity is beyond the scope of this book.

(c) Write the Langevin equation for a Student's t-distribution. What is the $\mu \to \infty$ limit of this equation?

(d) Simulate your Langevin equation for $\mu = 3$, 20 time steps per unit time and run the simulation for 20 000 units of time. Make a normalized histogram of the sampled values of X_t and compare with the law for $\mu = 3$ given above.

(e) Compared to the Gaussian process (Ornstein–Uhlenbeck), the Student t-process has many more short excursions but the long excursions are much longer than the Gaussian ones. Explain this behavior by comparing the function $V'(x)$ in the two cases. Describe their relative small $|x|$ and large $|x|$ behavior.

8.2.6 The Fokker–Planck Equation

It is interesting to derive, from the Langevin equation Eq. (8.28), the so-called Fokker–Planck equation that describes the dynamical evolution of the time dependent probability density $P(x,t)$ of the random variable X_t. The trick is to use Eq. (8.29) again, with $f(x)$ an arbitrary test function. Taking expectations, one finds

$$\mathrm{d}\mathbb{E}[f(X_t)] = \mathbb{E}\left[f'(X_t)F(X_t)\mathrm{d}t\right] + \frac{1}{2}\mathbb{E}[f''(X_t)]\mathrm{d}t, \tag{8.39}$$

where we have used the fact that, in the Itô convention, $\mathbb{E}[f'(X_t)\mathrm{d}B_t] = 0$. But by definition of $P(x,t)$, one also has

$$\mathbb{E}[f(X_t)] := \int f(x)P(x,t)\mathrm{d}x. \tag{8.40}$$

Hence,

$$\int f(x)\frac{\partial P(x,t)}{\partial t}\mathrm{d}x = \int f'(x)F(x)P(x,t)\mathrm{d}x + \frac{1}{2}\int f''(x)P(x,t)\mathrm{d}x. \tag{8.41}$$

Integrating by parts the right-hand side leads to

$$\int f(x)\frac{\partial P(x,t)}{\partial t}\mathrm{d}x = -\int f(x)\frac{\partial F(x)P(x,t)}{\partial x}\mathrm{d}x + \frac{1}{2}\int f(x)\frac{\partial^2 P(x,t)}{\partial x^2}\mathrm{d}x. \tag{8.42}$$

Since this equation holds for an arbitrary test function $f(x)$, it must be that

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial F(x)P(x,t)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 P(x,t)}{\partial x^2}, \tag{8.43}$$

which is called the Fokker–Planck equation. We have reintroduced an arbitrary value of σ here, to make the equation more general. One can easily check that when $F(x) = -V'(x)/2$, the stationary state of this equation, such that the left hand side is zero, is

$$P(x) = Z^{-1}\exp[-V(x)/\sigma^2], \tag{8.44}$$

as expected from the previous section.

https://doi.org/10.1017/9781108768900.011 Published online by Cambridge University Press

Bibliographical Notes

- A general introduction to probability theory:
  - W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, 1968.
- For introductory textbooks on the Langevin and Fokker–Planck equation, see
  - C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, volume 13 of Springer Series in Synergetics. Springer-Verlag, Berlin, 2004,
  - N. V. Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, Amsterdam, 2007.
- The central limit theorem:
  - P. Lévy. *Théorie de l'addition des variables aléatoires*. Gauthier Villars, Paris, 1937–1954,
  - B. V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley Publishing Co., Reading, Mass., London, Don Mills., Ont., 1968.
- For a physicist discussion of the central limit theorem, see
  - J.-P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, 2nd edition, 2003.