

A BRIEF INTRODUCTION TO SOME SIMPLE STOCHASTIC PROCESSES

Benjamin Lindner

There is a variety of different sources of fluctuations which affect various variables in neural systems. Consequently, there are different types of stochastic processes relevant for a quantitative modelling of stochastic neural activity. Take, for instance, the behaviour of ion channels, which are the key players in the neural dynamics for action potential generation. Only many channels together generate the behaviour modelled in the deterministic Hodgkin–Huxley (HH) system – a single channel switches stochastically between open and closed states possibly going through a cascade of substates. In the simplest modelling approach this would be an example of a *two-state process* – a function that attains only two discrete values $\{\sigma_-, \sigma_+\}$ and switches stochastically (randomly) between them with rates r_- and r_+ according to the scheme



Taking into account the summed effect of many such channels and their effect on the potential difference of a small patch of neural membrane we come to another stochastic process: the membrane potential, which can vary continuously and is driven by synaptic noise (shot noise) and channel noise. Approximating the conductance fluctuations by current fluctuations and neglecting any spiking mechanism (voltage-dependent conductance), the membrane fluctuations obey the simple stochastic differential equation (formally equivalent to the Ornstein–Uhlenbeck process from statistical physics)

$$\tau_{\text{membrane}} \dot{V} = -(V - V_L) + \xi(t) \quad (1.2)$$

where the effective membrane time-scale is, in the simplest case, $\tau_{\text{membrane}} = C/g_L$ (with C and g_L being the capacitance and the leak conductance of the patch of membrane) or which may also include contributions from the synaptic conductance. All the driving fluctuations are lumped into $\xi(t)$ and assumed to be Gaussian and uncorrelated (white) noise. We will see how to extract statements about the statistics of $V(t)$ that are only based on this kind of knowledge.

Going one level further, noise and external signals (injected currents, evoked synaptic activity, etc.) together lead to the generation of stereotypical action

potentials which encode information in the spiking times only. In Eqn (1.2) this could be modelled by adding voltage-dependent conductances or simply a threshold for spike generation (leaky integrate-and-fire model). The generation of stereotypical spikes naturally leads to the notion of the *point process* – a random sequence of time instances $\{t_i\}$. The spike train can be represented by a series of δ functions

$$x(t) = \sum_{\{t_i\}} \delta(t - t_i) \quad (1.3)$$

with a certain (possibly time-dependent) rate r . The statistics of the point process can be based on the spike train statistics or on the statistics of the intervals between firing times (interval statistics) – both are related in a non-trivial way as we will see.

Basic concepts of the stochastic theory have already been introduced in the Introduction (p. x). Here, we present a detailed overview of methods to simulate, measure, and analytically determine the most common probability measures for the three simple processes introduced above and illustrated in Fig. 1.1. We start by introducing our notation for moments, probability densities, correlation functions, power spectra, and give definitions for the characteristics of a fluctuating function (correlation time, noise intensity). We then discuss simple continuous, two-state, and point processes with special emphasis on the relation between different statistics (e.g. power spectra and interval densities). Possibly unexpected relations among the three processes are given in the last section.

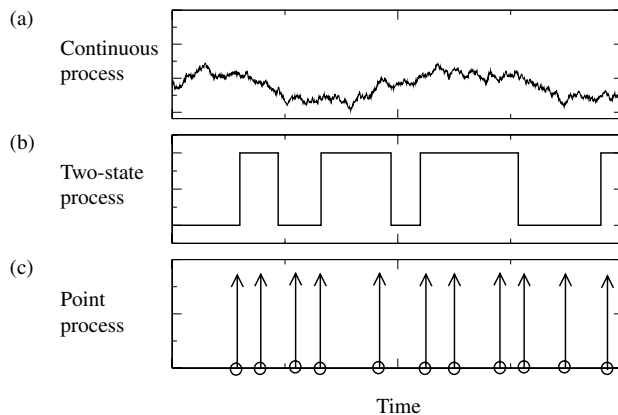


FIG. 1.1: The three classes of stochastic processes discussed in this chapter: (a) a continuous process (Ornstein–Uhlenbeck process according to Eqn (1.2), see Section 1.2); (b) a discrete process (random telegraph noise according to the scheme in Eqn (1.1), see Section 1.3); (c) a point process indicated by the circles and the associated δ spike train according to Eqn (1.3) shown by arrows (see Section 1.4).

This chapter can in no way replace a thorough introduction to probability theory and the theory of stochastic processes as given, for instance, in the following monographs: continuous systems are covered by Papoulis (1965), Stratonovich (1967), Risken (1984), Gardiner (1985), and van Kampen (1992). Gardiner (1985) and van Kampen (1992) also study discrete systems in more detail; useful results on point processes are covered in the books by Cox (Cox, 1962, Cox and Lewis, 1966) and Stratonovich (1967). Many useful results (some of which are rederived here) can be found in textbooks on stochastic processes in the neuroscience by Holden (1976), Ricciardi (1977), Tuckwell (1988), Tuckwell (1989), and more recently Gerstner and Kistler (2002).

1.1 Statistics of stochastic variables and stochastic processes

Averaging Having an ensemble of N independent samples $\{x_i\}$ of a stochastic variable x , we can estimate the average of any function $f(x)$ by

$$\langle f(x) \rangle \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (1.4)$$

which becomes exact for $N \rightarrow \infty$. In particular, we can ask for the mean of x

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.5)$$

and its variance

$$\langle \Delta x^2 \rangle = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \langle x \rangle^2. \quad (1.6)$$

If x is not only a variable but a stochastic function (or process) of time $x(t)$, we can ask for averages over functions of $x(t_i)$ at different times $\{t_i\}$ that are defined by

$$\langle f(x(t_1), x(t_2), \dots) \rangle = \frac{1}{N} \sum_{i=1}^N f(x_i(t_1), x_i(t_2), \dots). \quad (1.7)$$

The averages obviously depend on the ensemble from which we pick the values $x_i(t_k)$ at a certain time $t = t_k$ and they will depend (via the temporal evolution of this ensemble) on the time instants t_1, t_2, \dots . We may define a *stationary* ensemble, by requiring that all possible averages Eqn (1.7) depend only on the time differences $t_2 - t_1, t_3 - t_1, \dots$ but not on the absolute time. For other (weaker) kinds of stationarity, see Papoulis (1965).

Probability densities A very important average gives us the probability density of the stochastic variable or function. For a stochastic function $x(t)$ this density is given by

$$P(X, t) = \langle \delta(X - x(t)) \rangle. \quad (1.8)$$

If we know this density for a continuous process, $P(X, t)\Delta X$ with a small increment ΔX gives us the probability to observe $x(t) \in [X - \Delta X/2, X + \Delta X/2]$. Once we know the probability density, we can calculate the average of any function $f(x(t))$ by

$$\begin{aligned} \langle f(x(t)) \rangle &= \left\langle \int_{-\infty}^{\infty} dX f(X) \delta(x(t) - X) \right\rangle = \int_{-\infty}^{\infty} dX f(X) \langle \delta(x(t) - X) \rangle \\ &= \int_{-\infty}^{\infty} dX f(X) P(X, t) \end{aligned} \quad (1.9)$$

which still may depend on the time t as a parameter.

The definition of the probability density can be easily extended to values at different times yielding a multivariate distribution (joint probability density), formally defined by

$$P_n(X_1, t_1; X_2, t_2; \dots; X_n, t_n) = \langle \delta(X_1 - x(t_1)) \delta(X_2 - x(t_2)) \dots \delta(X_n - x(t_n)) \rangle. \quad (1.10)$$

The stochastic process is described in more and more detail by a hierarchy of probability densities of increasing order n . However, it is generally also more complicated to measure or calculate densities of higher order.

A conditional probability density assumes a certain knowledge about values of the process at time instants t_1, \dots, t_k (say, we know, for instance, that $x(t_1) = X_1$) and gives us the probability of values at other times under this condition. The conditional density (everything following the bar is the condition) is defined by

$$P(X_{k+1}, t_{k+1}; \dots; X_n, t_n | X_1, t_1; \dots; X_k, t_k) = \frac{P_n(X_1, t_1; \dots; X_k, t_k; \dots; X_n, t_n)}{P_k(X_1, t_1; \dots; X_k, t_k)}. \quad (1.11)$$

Markov process An important class of processes is defined by the property that the evolution of their probability density depends only on the present state but not on the past. Thus if we deal with a density $P_n(X_1, t_1; \dots; X_n, t_n)$ (where $t_1 < t_2 < \dots < t_n$) and ask for the lower-order probability conditioned on k values, then only the condition at the latest instant in time will matter

$$P(X_{k+1}, t_{k+1}; \dots; X_n, t_n | X_1, t_1; \dots; X_k, t_k) = P(X_{k+1}, t_{k+1}; \dots; X_n, t_n | X_k, t_k). \quad (1.12)$$

Knowledge about $x(t_k) = X_k$ at $t = t_k$ (which is the present time) determines the density at later times $t_{k'} > t_k$; knowledge about values in the past at $t_{k''} < t_k$ does not improve our statistical knowledge about the future.

For a Markov process, any multivariate probability density can be expressed by one specific conditional probability density $P(X, t|X_0, t_0)$ which is called the transition probability and is the central quantity of interest for a Markov process. *Correlation function* In order to characterize the stationary time-dependent features of the dynamics, one can use the autocorrelation function (or short correlation function)

$$C(\tau) = \lim_{t \rightarrow \infty} [\langle x(t)x(t+\tau) \rangle - \langle x(t) \rangle^2], \quad (1.13)$$

telling us essentially how much two values of the trajectory which are lagged by an interval τ have in common. The second term in Eqn (1.13) takes care of what we may expect for statistically independent values and the limit $t \rightarrow \infty$ is taken in order to achieve independence on the initial value $x(0)$. The correlation function for vanishing lag corresponds to the stationary variance of the random variable $x(t)$.

If we know the two-times probability density $P(X_0, t_0; X_1, t_1)$ or, equivalently, the conditional probability density $P(X_1, t_1|X_0, 0)$ together with the steady-state density $P_0(X)$ we can express the correlation function as follows:

$$\begin{aligned} C(\tau) &= \lim_{t \rightarrow \infty} \int_{-\infty}^{\infty} dX_0 \int_{-\infty}^{\infty} dX_1 X_0 X_1 [P(X_0, t; X_1, t+\tau) - P_0(X_0)P_0(X_1)] \\ &= \int_{-\infty}^{\infty} dX_0 \int_{-\infty}^{\infty} dX_1 X_0 X_1 P_0(X_0)[P(X_1, \tau|X_0, 0) - P_0(X_1)]. \end{aligned} \quad (1.14)$$

This formula is also useful if we know a differential equation for $P(X_1, \tau|X_0, 0)$: multiplying this equation by $X_0 X_1$ and integrating yields a potentially useful relation for $C(\tau)$.

Power spectrum An alternative way to quantify the fluctuations and their dynamics is to ask how the variance (in electrical systems proportional to the power of the process) is distributed with respect to frequency. The latter appears naturally by Fourier analysis. Defining the Fourier transform of $x(t)$ by

$$\tilde{x}(f) = \int_0^T dt x(t) e^{2\pi i f t} \quad (1.15)$$

we obtain a new (complex-valued) random variable that depends on the frequency f as a parameter. For a stationary time series $x(t)$ the mean value at finite frequency will vanish, that is $\langle \tilde{x}(f > 0) \rangle = 0$. The variance grows with the simulation time T ; the factor of proportionality is given by the *power spectrum*

$$S(f) = \lim_{T \rightarrow \infty} \frac{\langle \tilde{x} \tilde{x}^* \rangle}{T} \quad (1.16)$$

where \tilde{x}^* denotes the complex conjugate of \tilde{x} . In order to see that this has anything to do with the variance in the time domain let us consider the latter relation in more detail (assuming $\langle x \rangle = 0$, for simplicity):

$$\begin{aligned}
\frac{\langle \tilde{x}\tilde{x}^* \rangle}{T} &= \frac{1}{T} \int_0^T dt \int_0^T dt' e^{2\pi i f(t-t')} \langle x(t)x(t') \rangle \\
&= \frac{1}{T} \int_0^T dt \int_{t-T}^t d\tau e^{2\pi i f\tau} C(\tau) \\
&= \frac{1}{T} \int_{-T}^0 d\tau e^{2\pi i f\tau} C(\tau) \int_0^{T+\tau} dt + \frac{1}{T} \int_0^T d\tau e^{2\pi i f\tau} C(\tau) \int_{\tau}^T dt \\
&= \int_{-T}^T d\tau e^{2\pi i f\tau} C(\tau) - \int_{-T}^T d\tau e^{2\pi i f\tau} C(\tau) \frac{|\tau|}{T}. \tag{1.17}
\end{aligned}$$

Here, we have introduced a new variable τ , used the autocorrelation function $C(\tau)$, and exchanged the order of integration. For most processes of interest, the correlation function $C(\tau)$ decays sufficiently strongly at large times that the integral $\int d\tau C(\tau)\tau$ remains finite. Hence, for $T \rightarrow \infty$ the second term in the last line vanishes and we thus obtain the important relation

$$S(f) = \int_{-\infty}^{\infty} d\tau e^{2\pi i f\tau} C(\tau) \tag{1.18}$$

called the Wiener–Khinchin theorem (Risken, 1996, Gardiner, 1985). The relation is sometimes used as a definition of the spectrum: the power spectrum is the Fourier transform of the autocorrelation function. As for any Fourier transform, the reverse is also true: The correlation function is the (inverse) Fourier transform of the spectrum:

$$C(\tau) = \int_{-\infty}^{\infty} df e^{-2\pi i f\tau} S(f). \tag{1.19}$$

In particular for $\tau = 0$ we obtain

$$C(0) = \langle \Delta x^2 \rangle = \int_{-\infty}^{\infty} df S(f). \tag{1.20}$$

Thus as promised the integrated spectrum gives the variance (the power) and the spectrum shows how this power is distributed over frequencies.

The Wiener–Khinchin theorem offers a fast numerical way to estimate the correlation function: One measures or simulates many independent realizations of the process $x_i(t)$ for a sufficiently long time window T , calculates the fast Fourier transform, and determines the frequency-dependent variance of the resulting ensemble. Dividing by the time window T yields an estimate of the power spectrum and applying an additional Fourier back-transformation on the function gives the autocorrelation function $C(\tau)$.

Correlation time In order to estimate the time over which a trajectory is correlated, different definitions of a correlation time can be used. If the process' correlation function does not show oscillations and remains positive we may use the integral over the normalized autocorrelation function that is also simply related to the power spectrum at vanishing frequency:

$$\tau_{\text{corr}} = \int_0^\infty d\tau C(\tau)/C(0) = \frac{S(0)}{2C(0)}. \quad (1.21)$$

In other cases an integral over the absolute value (see, for instance, Hänggi and Jung 1995) or the square of the autocorrelation function yields a more meaningful estimate of a correlation time.

Noise intensity The intensity or strength of the process is *not* just given by the variance of its steady-state density. A widely used definition of the noise intensity of a process with non-negative correlation function is as follows:

$$D = \int_0^\infty d\tau C(\tau) = \frac{S(0)}{2} = \langle \Delta x^2 \rangle \tau_{\text{corr}}. \quad (1.22)$$

In the last step we have related the noise intensity to the variance and the correlation time of the process using Eqns (1.20) and (1.21). This illustrates that it matters not only *how large* typical amplitudes of the noise are (as quantified by the variance) but also for *how long* the noise acts with roughly the same value (as quantified by the correlation time). As is clear from the discussion of the correlation time, the definition of the noise intensity is also meaningful for processes with monotonically decaying correlation but does not apply to processes with a strongly oscillating correlation function.

In order to illustrate these rather dry definitions we turn to our standard examples. We will also use this opportunity to introduce simple simulation algorithms for the three classes of stochastic systems.

1.2 The Ornstein–Uhlenbeck process

As an example of a continuous stochastic process we consider the Ornstein–Uhlenbeck process (or OU process) which is also known in the mathematical

literature as a mean-reverting process. It appears in a number of chapters in this book; see Chapters 4, 5 and 9 for a conductance model, and Chapter 10. It obeys the Langevin equation (Uhlenbeck and Ornstein, 1930, Risken, 1996)

$$\dot{v} = -\gamma v + \sqrt{2D}\xi(t) \quad (1.23)$$

where $\xi(t)$ is a Gaussian white noise with average zero and a δ correlation in time

$$\langle \xi(t) \rangle = 0, \quad \langle \xi(t)\xi(t') \rangle = \delta(t - t'). \quad (1.24)$$

The intensity of the driving noise $\sqrt{2D}\xi(t)$ is (according to Eqn (1.22)) D , and its correlation time is zero as expected for an uncorrelated process. The power spectrum is flat, $S(f) = 2D$, which led to the name ‘white’ noise in analogy with the spectrum of white light. The fact that this *driving* noise has infinite variance and vanishing correlation time (unlike any function in the real world) is a mathematical abstraction leading to the very useful Markov property of the *driven* process. We note that the Gaussian property is used when dealing with increments of the process (which are also Gaussian).

We have seen that the voltage across a passive membrane obeys the same dynamics as Eqn (1.23) (with $\tau_{\text{membrane}} = 1/\gamma$ and a proper rescaling of the noise strength). In the above form, however, the model describes the velocity of a Brownian particle of unit mass by Newton’s law with the acceleration term on the left-hand side and a sum of a friction force (Stokes friction with coefficient γ) and a random stochastic force $\sqrt{2D}\xi(t)$ on the right-hand side. It was introduced 100 years ago by Langevin (Langevin, 1908) and later in detail studied by Uhlenbeck and Ornstein (Uhlenbeck and Ornstein, 1930). In neuroscience it has immense importance because, supplemented with a fire-and-reset condition, Eqn (1.23) is mathematically equivalent to the often-used leaky integrate-and-fire model.

The OU process is a Markov process. This might not be obvious since although we deal with a first-order equation in Eqn (1.23) that is (in the autonomous case) determined only by its initial condition (i.e. by the present time), we have time-dependent driving (the noise). The noise is, however, uncorrelated and thus does not introduce a statistical dependence on the past that would contradict the Markov property in Eqn (1.12).

Simulation algorithm A simple algorithm for this process is obtained by iterating its discretized version in a simple Euler procedure

$$v_{i,j+1} = v_{i,j}(1 - \gamma\Delta t) + \sqrt{2D\Delta t}\eta_{i,j}, \quad i = 1, \dots, N \quad (1.25)$$

where the first index i denotes the realization and the second index j the time $t = j\Delta t$; $\eta_{i,j}$ are independent Gaussian random numbers (see Section 1.6) with

zero mean and unit variance¹ $\langle \eta_{i,j} \eta_{i',j'} \rangle = \delta_{i,i'} \delta_{j,j'}$ (where we have used the Kronecker symbol).

In order to measure an example of a time-dependent probability density $P(V, t)$, we start every realization at $v_{i,0} = v_0$ and so, of course, the initial probability density would be $P(V, t = 0) = \delta(V - v_0)$; $P(V, t)$ is then the transition probability of the process. Performing NK iterations of Eqn (1.25) we may find an estimate of the density $P(V, t = K\Delta t)$ at time $t = K\Delta t$ by making a histogram of the N values $v_{i,K}$ ($i = 1, \dots, N$). The whole procedure is shown in Fig. 1.2. An estimate of the probability density is given by

$$P(V, t = K\Delta t) \approx \frac{1}{N\Delta v} \sum_{i=1}^N \Theta(V + \Delta v/2 - v_{i,K}) \Theta(v_{i,K} - (V - \Delta v/2)) \quad (1.26)$$

where the product of Heaviside functions $\Theta(\cdot)$ ensures that $v_{i,K}$ only contributes to the estimate if it falls into the interval $[V - \Delta v/2, V + \Delta v/2]$. Clearly, the density estimated by Eqn (1.26) depends on the time. In the example we have chosen there is a drift of the whole probability density towards the origin – for short times the histogram is biased towards the initial value (which is positive in our numerical example). For longer times ($t \gg \tau$) the probability density becomes independent of time and reaches a stationary limit $P(V, t) \rightarrow P_0(V)$ which means that the initial value (or the initial distribution) is forgotten. In many cases of importance (ergodic processes) this steady-state density would coincide with a time-average of the process itself. This means that instead of a large ensemble of trajectories we may just average a single trajectory over a longer time (bottom panel in Fig. 1.2b). Indeed, this yields the same density as the ensemble average at long times.

We can calculate the probability density analytically and compare it to our simulation result. For the Langevin equation there exists a corresponding Fokker–Planck equation that governs the evolution of the probability density

$$\partial_t P(V, t) = \partial_V [\gamma V + D \partial_V] P(V, t). \quad (1.27)$$

The first term on the right-hand side is the *drift term* (resulting from the friction term in the Langevin equation) and the second one is the *diffusion term* (resulting from the stochastic driving). The correct boundary conditions for a freely evolving Ornstein–Uhlenbeck process are natural boundary conditions

$$\lim_{V \rightarrow \infty} P(\pm V, t) = \lim_{V \rightarrow \infty} \partial_V P(\pm V, t) = 0. \quad (1.28)$$

¹Please note the uncommon scaling of the stochastic term with $\sqrt{\Delta t}$ which can be understood as follows: integrating Eqn (1.23) over the time step Δt , we get the increment $\Delta W = \int_t^{t+\Delta t} dt' \xi(t')$ of a Wiener process which is Gaussian, has vanishing mean, and a variance $\langle \Delta W^2 \rangle = \int_t^{t+\Delta t} \int_t^{t+\Delta t} dt' dt'' \langle \xi(t') \xi(t'') \rangle = \Delta t$; hence the standard deviation is $\sim \sqrt{\Delta t}$.

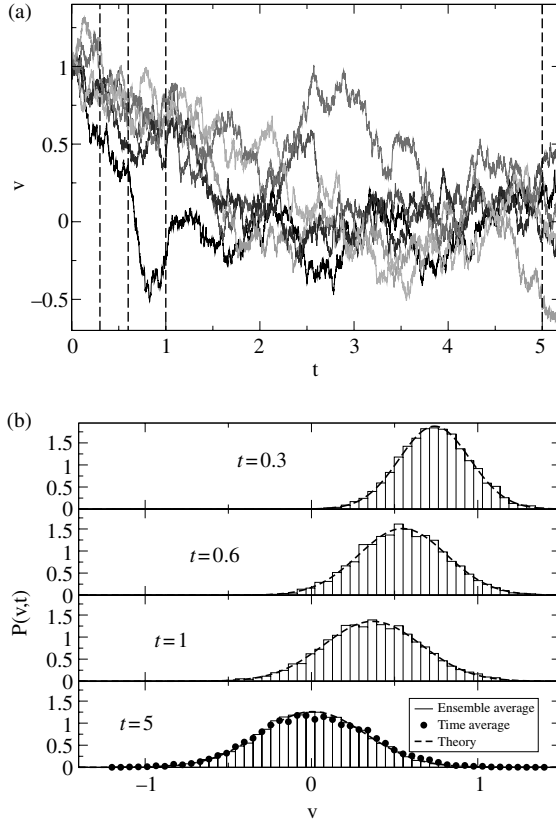


FIG. 1.2: Time-dependent probability densities of an Ornstein–Uhlenbeck (OU) process for $D = 0.1$ and $\gamma = 1$. $N = 10^4$ trajectories were started at $v(t = 0) = 1$ and simulated according to the scheme Eqn (1.25) (time step $\Delta t = 0.01$) and the density was estimated via Eqn (1.26). In (a) the first six trajectories are shown; the dashed lines indicate the instants at which snapshots of the density are shown in (b): the density is initially centred around a positive value (top panel in (b)), for later times it is centred around zero (bottom panel in (b)). Also shown by dots is the density obtained from a long time average ($T = 10^3$) which agrees nicely with the long-time ensemble average (bottom panel in (b)). Theory (dashed lines) is given in Eqn (1.29).

We note in passing that a Fokker–Planck equation can also be found when there is a nonlinear drift term and state-dependent noise (also called multiplicative noise) and that it can also be generalized to more than one variable; for derivation(s) of the Fokker–Planck equation and further applications see Risken (1996). For the OU process it is possible to find the full time-dependent solution with the

initial condition $P(V, 0) = \delta(V - v_0)$ and natural boundary conditions, yielding the transition probability (see Risken 1996)

$$P(V, t) = \frac{1}{\sqrt{2\pi\langle\Delta v^2(t)\rangle}} \exp\left[-\frac{(V - \langle v(t)\rangle)^2}{2\langle\Delta v^2(t)\rangle}\right] \quad (1.29)$$

where the time-dependent mean and variance read

$$\langle v(t)\rangle = v_0 e^{-\gamma t}, \quad (1.30)$$

$$\langle\Delta v^2(t)\rangle = \frac{D}{\gamma}[1 - e^{-2\gamma t}]. \quad (1.31)$$

Equation (1.29) has been plotted in Fig. 1.2; the shape of the histograms reproduce this formula quite well.

For long times ($t \rightarrow \infty$) the mean and variance approach $\langle v\rangle \rightarrow 0$ and $\langle\Delta v^2\rangle \rightarrow D/\gamma = k_B T$ (the latter relation is called the Einstein relation (Risken, 1996)) and in this limit the steady-state density is a so-called Maxwell distribution (Risken, 1996)

$$P_0(V) = \frac{e^{-V^2/(2k_B T)}}{\sqrt{2\pi k_B T}}. \quad (1.32)$$

The time-dependent density (either simulated or calculated) could be used to calculate the autocorrelation function of the OU process. However, we want to illustrate the calculation of the second-order statistics in different ways. First of all, we can determine the correlation function via the power spectrum and fast Fourier transform from simulations. Secondly, we can analytically calculate the correlation function using (i) the Fokker–Planck equation; (ii) the formal solution of Eqn (1.23); and (iii) the Fourier transform of the power spectrum which is obtained by Rice’s method.

Simulations are shown in Fig. 1.3: a single trajectory $v(t)$ (a) is Fourier transformed to \tilde{v} (b); many such realizations are used to estimate the power spectrum (variance of \tilde{v} divided by simulation time) shown in (d); a back-transformation into the time domain yields the autocorrelation function shown in (c). For an Ornstein–Uhlenbeck process the correlation function decays with a single rate (the friction coefficient) and the power spectrum displays a Lorentzian shape with a corner frequency determined by the friction coefficient, that is the power spectrum attains half its maximal value at $2\pi f_{\text{corner}} = \gamma$ (see below, Eqn (1.37)).

Calculation of the correlation function using the Fokker–Planck equation First, it is instructive to calculate the stationary variance $\langle v^2\rangle$ from the Fokker–Planck equation (1.27), that is the correlation function at vanishing lag $C(0)$. To this end, we set the time derivative on the left-hand side to zero (implying that we

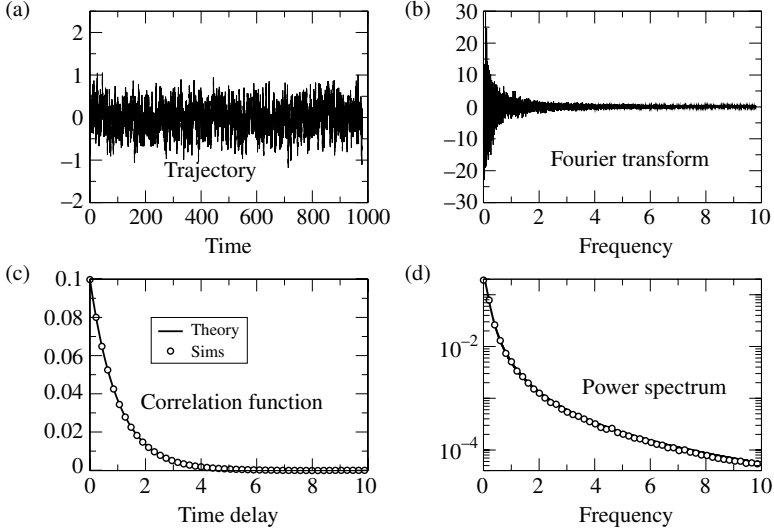


FIG. 1.3: Second-order statistics of the Ornstein–Uhlenbeck process. A single realization shown in (a) is Fourier transformed according to Eqn (1.15); the real part of this complex-valued stochastic quantity is shown as a function of frequency in (b). From many such realizations ($N = 1000$) one obtains the variance of the Fourier transform and from the latter one can determine the power spectrum (d) via Eqn (1.16). The Fourier back-transformation of the power spectrum yields, through Eqn (1.19), the correlation function (c) which shows an exponential decay. Theory is given in Eqn (1.35) (correlation function) and Eqn (1.37) (power spectrum). Numerical parameters: $D = 0.1, \gamma = 1, \Delta t = 0.001, T = 1000$.

now deal with the stationary density $P_0(V)$, multiply both sides by $V^2/2$, and integrate V over the real axis. Multiple integrations by parts (using Eqn (1.28)) yield

$$\begin{aligned}
 0 &= -\gamma \int_{-\infty}^{\infty} dV V^2 P_0(V) + D \int_{-\infty}^{\infty} dV P_0(V) = -\gamma \langle v^2 \rangle + D \\
 \Rightarrow \langle v^2 \rangle &= C(0) = \frac{D}{\gamma}.
 \end{aligned} \tag{1.33}$$

In order to derive an equation for the correlation function we proceed as follows. We multiply the Fokker–Planck equation (1.27) by $V V_0 P_0(V_0)$ and perform a double integral over V_0 and V . Using (i) the relation between correlation function and time-dependent density Eqn (1.14); (ii) the fact that the stationary mean value vanishes, that is $\langle v \rangle = \int dV_0 V_0 P_0(V_0) = 0$; (iii) all probability densities

and their derivatives vanish at infinity, one obtains after multiple integrations by parts on the right-hand side, the equation

$$\frac{d}{dt}C(t) = -\gamma C(t), \quad t > 0 \quad (1.34)$$

(here the partial derivative with respect to time turns into an ordinary derivative). From the initial condition in Eqn (1.33) and the fact that the correlation function is an even function in τ , we find

$$C(\tau) = \frac{D}{\gamma} e^{-\gamma|\tau|}. \quad (1.35)$$

This formula shows excellent agreement with our simulation data in Fig. 1.3(c). *Calculation of the correlation function using the formal solution* If we know the value of $v(t)$ at $t = 0$, i.e. $v(t = 0) = v_0$, we can formally solve the Langevin equation (it is a linear inhomogeneous first-order differential equation). We obtain

$$v(t) = v_0 e^{-\gamma t} + \sqrt{2D} \int_0^t dt' e^{-\gamma(t-t')} \xi(t'). \quad (1.36)$$

It is easily seen that the stationary mean value must vanish:

$$\langle v \rangle_{\text{st}} = \lim_{t \rightarrow \infty} \left[v_0 e^{-\gamma t} + \sqrt{2D} \int_0^t dt' e^{-\gamma(t-t')} \langle \xi(t') \rangle \right] = 0$$

where we have used the fact that the white noise has a zero mean at all times. The autocorrelation function is obtained as follows. For $t \rightarrow \infty$ any term involving the initial condition (the first term on the right-hand side in Eqn (1.36)) decays to zero; dropping the respective terms we obtain for positive increment $\tau > 0$

$$\begin{aligned} \langle v(t)v(t+\tau) \rangle &= 2D \int_0^t dt' \int_0^{t+\tau} dt'' e^{-\gamma(2t+\tau-t'-t'')} \langle \xi(t')\xi(t'') \rangle \\ &= 2D \int_0^t dt' e^{-\gamma(2t+\tau-2t')} = \frac{D}{\gamma} [e^{-\gamma\tau} - e^{-\gamma(2t+\tau)}]. \end{aligned}$$

Doing the same calculation with negative τ , we can generalize this formula and recover in the limit $t \rightarrow \infty$ the autocorrelation function (1.35).

Analytical calculation of the power spectrum using Rice's method It suffices to perform a Fourier transformation of the Langevin equation as follows:

$$\begin{aligned} \int_0^T dt e^{2\pi i f t} \dot{v} &= -\gamma \int_0^T dt e^{2\pi i f t} v + \sqrt{2D} \int_0^T dt e^{2\pi i f t} \xi(t) \\ v(T)e^{2\pi i f T} - v_0 - 2\pi i f \tilde{v} &= -\gamma \tilde{v} + \sqrt{2D} \tilde{\xi} \\ \Rightarrow \tilde{v} &= \frac{\sqrt{2D} \tilde{\xi} - v(T)e^{2\pi i f T} + v_0}{\gamma - 2\pi i f}. \end{aligned}$$

Multiplying by the complex conjugated Fourier transform \tilde{v}^* and dividing by T , the last two terms in the numerator vanish and only the white-noise spectrum remains ($\langle \tilde{\xi} \tilde{\xi}^* \rangle / T = S_\xi = 1$). By averaging we obtain

$$S(f) = \frac{2D}{\gamma^2 + (2\pi f)^2} \quad (1.37)$$

which agrees with the Fourier transform of the correlation function (1.35) in full accordance with the Wiener–Khinchin theorem (1.18).

Noise intensity and correlation time The noise intensity of the OU process is

$$D_{\text{OU}} = D\gamma^{-2}. \quad (1.38)$$

Its correlation time is easily calculated via Eqn (1.21):

$$\tau_{\text{corr}} = S(0)/[2C(0)] = \gamma^{-1} \quad (1.39)$$

which is independent of the parameter D . For the OU process we can nicely separate and control the intensity and the correlation time of the noise, e.g. by scaling $D = Q\gamma^2$; then Q and γ^{-1} are the intensity and the correlation time, respectively. This is in general not possible for processes generated by nonlinear stochastic differential equations (SDEs).

1.3 Two-state process

A switching between two states can result from complicated dynamics and is in general characterized by the sequence of residence times (the random periods spent in the two possible states). A very simple process is the Markovian two-state process in which transitions from one state to the other are entirely determined by a switching rate and do not depend on the whole history of previous switchings. This was the case illustrated in Eqn (1.1). We start with the statistics of such a simple process and then present the slightly more general case in which the residence times are drawn from two arbitrary probability densities.

1.3.1 Markovian telegraph noise (dichotomous noise)

Simulation algorithm A simple algorithm for simulating a two-state Markov process with rates r_+ (for leaving the state $x = \sigma_+$) and r_- (for leaving the state $x = \sigma_-$) would consist of the following steps:

1. draw a uniformly distributed random number $a_j \in [0, 1]$;
2. for $x(t) = \sigma_{\pm}$:

if $a_j < r_{\pm}\Delta t$	$\Rightarrow x(t + \Delta t) = \sigma_{\mp}$
else	$\Rightarrow x(t + \Delta t) = \sigma_{\pm}$

update time $t \rightarrow t + \Delta t$ and return to 1.

Here we have to use a time step $\Delta t \ll r_{\pm}^{-1}$ in order to keep the probability of multiple transitions within one time step negligibly small.

Probability density and master equation For the telegraph noise above with states $x \in \{\pm 1\}$ the ‘density’ in this case reduces to

$$P(X, t) = p_+(t)\delta(X - \sigma_+) + p_-(t)\delta(X - \sigma_-) \quad (1.40)$$

where $p_{\pm}(t)$ are governed by the master equation

$$\partial_t p_{\pm}(t) = r_{\mp}p_{\mp} - r_{\pm}p_{\pm}. \quad (1.41)$$

On the right-hand side we have two terms for gain and loss of probability or in other words, for influx and efflux of probability. These fluxes are proportional to the switching rate and the probability of the state that is left. The term r_+p_+ , for instance, gives the efflux of probability from the plus state (the trajectories switching from σ_+ to σ_-) and is, of course, equal to the influx to the minus state.

Residence times Before we come to the full solution of Eqn (1.41) we may answer a simple question: What is the distribution of residence times in one of the states? To answer this we just have to set the initial probability say in the σ_+ state to 1 and the influx term in the master equation to zero; then the efflux of probability describes the fraction of realizations that leave σ_+ at time t for the first time (leaving the gain term in the equation the same efflux term would describe the fraction of all the realizations that leave σ_+ at time t). Without the gain term, the equations reduce to only one and we obtain for the residence time density $\rho_+(t)$

$$\dot{p}_+ = -r_+p_+ \text{ with } p_+(0) = 1 \Rightarrow p_+(t) = e^{-r_+t} \Rightarrow \rho_+(t) = r_+e^{-r_+t} \quad (1.42)$$

and by a completely equivalent calculation $\rho_-(t) = r_-e^{-r_-t}$. The mean residence time is easily calculated and reads $\tau_{\pm} := \int_0^{\infty} dt t \rho_{\pm}(t) = r_{\pm}^{-1}$. The waiting time densities permit an alternative simulation algorithm: instead of drawing random numbers in each time step and asking whether a transition has occurred, we can draw random residence times from the respective exponential distribution – in between these switching times, the process is simply a constant (for the generation of exponentially distributed random numbers see Section 1.6).

Probability density Turning again to the two-state process, we can use Eqn (1.41) to calculate the time-dependent probability density and from it any desired statistics. Since Eqn (1.41) is a simple linear differential equation, we obtain by standard methods an exponential solution that can (for arbitrary initial conditions $p_+(0), p_-(0) = 1 - p_+(0)$) be written as

$$p_{\pm}(t) = p_{\pm}^0 \pm [p_+(0) - p_+^0] e^{-(r_+ + r_-)t}. \quad (1.43)$$

The first term is the steady-state solution that is approached for long times

$$p_{\pm}^0 = \lim_{t \rightarrow \infty} p_{\pm}(t) = \frac{r_{\mp}}{r_+ + r_-}. \quad (1.44)$$

The other terms in Eqn (1.43) describe the decay of the initial condition. In particular, from $p_+(0) = 1$ and $p_+(0) = 0$ we can obtain the four conditional probabilities to be in two specific states at time zero and at time t , i.e. the transition probabilities of the random telegraph noise

$$p_{\pm|+} = p_{\pm}^0 \pm p_-^0 e^{-(r_- + r_+)t}, \quad (1.45)$$

$$p_{\pm|-} = p_{\pm}^0 \mp p_+^0 e^{-(r_- + r_+)t}. \quad (1.46)$$

These conditional probabilities multiplied by the respective steady-state probability give the two-times probability needed for calculating the autocorrelation function via Eqn (1.14).

Correlation function As for the OU process there are different ways to calculate the correlation function. Here we use Eqn (1.14) (but for the discrete-state process the integrals turn into sums over the two states). We obtain

$$C(\tau) = \sum_{i,j=+,-} \sigma_i \sigma_j p_j^0 [p_{i|j}(\tau) - p_i^0]. \quad (1.47)$$

It is a simple exercise to calculate from Eqns (1.44)–(1.47) the autocorrelation function

$$C(\tau) = \frac{r_- r_+ (\sigma_+ - \sigma_-)^2}{(r_- + r_+)^2} e^{-(r_- + r_+)|\tau|}. \quad (1.48)$$

So, as for the Ornstein–Uhlenbeck process, we obtain a purely exponential correlation function for the Markovian telegraph noise. Hence the power spectrum is again a Lorentzian as for the OU process:

$$S(f) = \frac{2(\sigma_+ - \sigma_-)^2 / (r_+^{-1} + r_-^{-1})}{(r_+ + r_-)^2 + (2\pi f)^2}. \quad (1.49)$$

Noise intensity and correlation time The intensity of the two-state fluctuations is given by

$$D_{\text{dicho}} = \frac{S(0)}{2} = \frac{(\sigma_+ - \sigma_-)^2}{(r_+^{-1} + r_-^{-1})(r_+ + r_-)^2}. \quad (1.50)$$

The noise intensity vanishes, in particular, if one of the rates goes to infinity or zero (with the other rate and the amplitudes σ_{\pm} being fixed). In the symmetric case ($r_+ = r_- = r$), the intensity is $D_{\text{dicho}} = (\sigma_+ - \sigma_-)^2 / (8r) = \langle \Delta \sigma^2 \rangle / (2r)$.

The decay rate of the correlation function (1.48) is simply the sum of the two transition rates r_{\pm} . The correlation time of the process is

$$\tau_{\text{corr}} = \frac{1}{r_- + r_+} = \frac{1}{1/\tau_+ + 1/\tau_-}. \quad (1.51)$$

Here we have used the residence (or waiting) times $\tau_{\pm} = r_{\pm}^{-1}$ in the two states. It is amusing to note that the correlation time τ_{corr} is dominated by the *smaller* of the two times. So if we have a very asymmetric two-state process, i.e. a pulse train, then the correlation time will be close to the pulse width rather than to the interpulse interval. This makes sense: for very small pulse width this two-state process approaches the limit of a Poisson spike train which has no temporal correlations at all.

1.3.2 Renewal two-state processes

In general two-state processes do not follow Markovian dynamics. Switching rates, for instance, do depend on the past and not only on the current state of the system. In some comparably simple cases we can still relate statistics of different kinds. For instance, if the system is ergodic and stationary, we can relate by time-averaging the mean residence times τ_{\pm} of the two states and the steady-state probabilities

$$p_{\pm}^0 = \frac{\tau_{\pm}}{\tau_+ + \tau_-} \quad (1.52)$$

which is true in particular for the random telegraph noise as can be checked by Eqn (1.44). Also, the mean and variance of the process can be expressed in a simple way using τ_{\pm} and σ_{\pm} :

$$\langle \sigma \rangle = \frac{\tau_+ \sigma_+ + \tau_- \sigma_-}{\tau_+ + \tau_-}, \quad \langle \Delta \sigma^2 \rangle = \frac{\tau_+ \tau_-}{(\tau_+ + \tau_-)^2} (\sigma_+ - \sigma_-)^2. \quad (1.53)$$

In the following we consider a comparably simple kind of non-Markovian process: the renewal state process. For such a process the switching probability depends only on the time that has passed since the last switching event. The Markovian

telegraph noise is included as a special case for which the residence time densities are purely exponential.

Simulation of a renewal two-state process If we know that the residence times in each of the two states are independent of each other and distributed according to two waiting time densities $w_+(t), w_-(t)$ with mean waiting times τ_\pm , we can simulate the process by simulating alternating piecewise constant parts of the process $x(t) = \sigma_\pm$, the lengths of which are drawn from $w_\pm(t - t_i)$ (here t_i denotes the instant of last switching).

Power spectrum and residence time densities Knowing the Fourier transforms of the residence time densities \tilde{w}_\pm , we can calculate the power spectrum of the two-state process with the Stratonovich formula (cf. Stratonovich 1967, Vol. I, Eqn (6.121)):

$$S(f) = \frac{2(\sigma_+ - \sigma_-)^2}{(\tau_+ + \tau_-)(2\pi f)^2} \operatorname{Re} \left[\frac{(1 - \tilde{w}_+)(1 - \tilde{w}_-)}{1 - \tilde{w}_- \tilde{w}_+} \right]. \quad (1.54)$$

For the special case of Markovian telegraph noise, the waiting times are exponential $w_\pm(t) = \tau_\pm^{-1} \exp[-t/\tau_\pm]$ and their Fourier transforms read $\tilde{w}_\pm = 1/(1 - 2\pi i f \tau_\pm)$ from which we obtain the Lorentzian spectrum (1.49).

More interestingly is a non-Markovian case that is more regular (more periodic) than random telegraph noise. Suppose, for instance, equal residence time statistics in both states with the single time being a sum of a fixed dead time τ_D in which no transition is possible and an exponentially distributed time τ . Clearly, this setup contains a symmetric random telegraph noise as the limit case $\tau_D = 0$. Dead times, also known as refractory periods, are quite common in neural systems, for instance, as the finite duration of the spike (~ 1 ms). The Fourier transform of the residence time is now $\tilde{w} = \exp(2\pi i f \tau_D)/[1 - 2\pi i f \tau]$ and the power spectrum according to Eqn (1.54) reads

$$S(f) = \frac{1}{2} \times \frac{(\sigma_+ - \sigma_-)^2 \tau^2 / (\tau + \tau_D)}{1 + \cos(2\pi f \tau_D) - 2\pi f \tau \sin(2\pi f \tau_D) + 2(\pi f \tau)^2}. \quad (1.55)$$

For $\tau_D \rightarrow 0$ this agrees with the Markovian case in Eqn (1.49). In the other limiting case of very small exponential waiting time τ , the process becomes very regular and consequently the spectrum approaches a series of δ peaks at $f = (2\tau_D)^{-1} + n/\tau_D$ ($n = 0, 1, \dots$). Sample trajectories, waiting time densities of the two states, and the power spectrum of the resulting two-state processes are illustrated in Fig. 1.4. As the refractory period increases we start seeing oscillatory features in the power spectrum (bottom panels).

We note that a trajectory with strong oscillatory component, as indicated by a pronounced spectral peak, will generally result in a correlation function with damped oscillations (not shown). Correlation time and noise intensity defined by integrals of the correlation function can no longer be used in such a case.

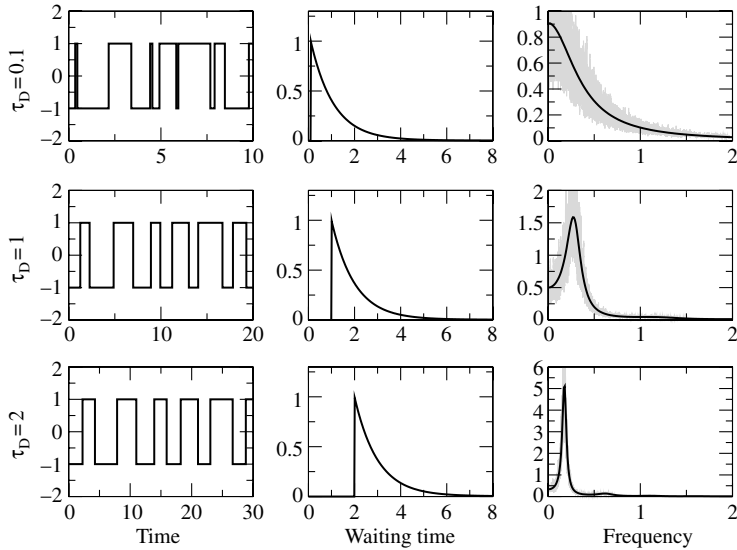


FIG. 1.4: Example of a two-state renewal process that is more regular than the Markovian telegraph process. The left column shows sample trajectories for the three different values of the refractory period τ_D (indicated at the left side): with increasing τ_D (top to bottom) the trajectory looks more regular. Middle column: probability densities which are exponential functions shifted to the right by the increasing values of the refractory period. Right column: power spectrum according to Eqn (1.55) which develops from an almost Lorentzian shape (top panel) to a pronounced peak roughly at the inverse of twice the refractory period (bottom panel). The peak in the power spectrum at a finite frequency indicates an almost regular, i.e. oscillatory, behaviour as also seen in the trajectory on the left. Parameters: $\sigma_{\pm} = \pm 1$, $\tau = 1$.

1.4 Point processes

Here we start with considerations that apply to general stationary point processes, continue with the simplest process, the Poisson process, and finally discuss renewal processes.

1.4.1 General spike train and interval statistics

Probability density as spike rate Consider the δ spike train

$$x(t) = \sum_{t_i} \delta(t - t_i) \quad (1.56)$$

associated with a point process $\{t_i\}$, e.g. as illustrated in Fig. 1.1. Does it make sense to study probability densities for this object? At first glance it does not,

since this pulse train attains only zero (with probability one) or infinity (with probability zero since the spike lasts only an infinitesimal period). However, this is not true in more than one sense. First of all, we can identify the probability of observing at least one spike in $[t - \Delta t/2, t + \Delta t/2]$ which gives us, after dividing by Δt , a probability density² – the well-known spike rate (of neural firing, for instance).

For a specific realization $x_i(t)$ we would get an indicator function over the interval by simply integrating $x_i(t)$ over the short interval Δt . In order to estimate the firing probability density we have to sum over an ensemble and divide by the number of realizations and by the interval, yielding

$$r(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{N\Delta t} \sum_{i=1}^N \int_{t-\Delta t/2}^{t+\Delta t/2} dt' x_i(t') = \langle x(t) \rangle. \quad (1.57)$$

The spike rate is thus the mean value of the spike train.

Another interpretation of a probability density can be achieved by considering the number of spikes in an interval $(0, t)$ given by the *spike count*

$$N(t) = \int_0^t dt' x(t') \quad (1.58)$$

with respect to which we may ask for a probability density $P(N, t)$. The latter is defined on a discrete space as is the telegraph noise. However, in the case of the spike count there are infinitely many discrete states and there is apparently no steady state (for a stationary spike train, the spike count never stops growing in time).

The third aspect in which probability density may be important for spike train statistics is when the spike train is passed through a linear filter and instead of the δ peaks we have a series of exponentially decaying pulses also referred to as shot noise. Synaptic noise is very often modelled in this way; in this case the variable y is proportional to a conductance change and obeys the dynamics

$$\tau_F \dot{y} = -y + \varepsilon x \quad \Rightarrow \quad y(t) = \frac{\varepsilon}{\tau_F} \sum \Theta(t - t_i) \exp \left[-\frac{(t - t_i)}{\tau_F} \right]. \quad (1.59)$$

The probability density $P(y, t)$ for this continuous variable is an important characteristic of the shot noise. In the problem of synaptic conductance it is essential in order to understand the membrane fluctuations which are caused by synaptic input.

²We emphasize that this probability density is a density with respect to time which was only a parameter in the previous cases.

Correlation function Higher order probability densities (always with respect to time!) are obtained from averaged products of the spike train. These in turn are related to correlation functions of the spike train:

$$r_2(t_1, t_2) = \lim_{\Delta t \rightarrow 0} \frac{1}{N\Delta t} \sum_{i=1}^N \int_{t-\Delta t/2}^{t+\Delta t/2} dt'_1 \int_{t-\Delta t/2}^{t+\Delta t/2} dt'_2 x_i(t'_1)x_i(t'_2) = \langle x(t_1)x(t_2) \rangle. \quad (1.60)$$

Furthermore, $r_2(t_1|t_0) = r_2(t_0, t_1)/r(t_0)$ gives us the probability of observing a spike at t_1 if we know that we had a spike at $t = t_0$. The correlation function of a stationary spike train with constant rate r can be expressed as

$$C(\tau) = r[r_2(\tau|0) - r]. \quad (1.61)$$

Interval statistics and its general relation to the power spectrum We can also base our characterization of the point process on the intervals between events (spikes) see Fig. 1.5. The best known is the interspike interval (ISI), the interval between adjacent spikes. Intervals that consist of the sum of n adjacent ISIs are known as n th-order intervals T_n . If we know the probability densities $\rho_n(T_n)$ of all n th-order intervals, the statistics of the associated stationary spike train is completely determined. For instance, the conditional probability density can be expressed as follows (see, for instance, Holden, 1976)

$$r_2(\tau|0) = \delta(\tau) + \sum_{n=1}^{\infty} \rho_n(\tau). \quad (1.62)$$

The first term reflects the sure event that we have a spike at $\tau = 0$ (which is our condition); the other terms sum over the probabilities to have the n th spike at finite τ . Using Eqn (1.62) in Eqn (1.61) relates the correlation function to the n th-order interval density, i.e. relates spike train statistics to interval statistics.

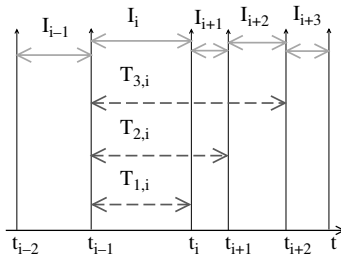


FIG. 1.5: Interval statistics. A spike train indicated by arrows (pointing up); the intervals between adjacent spikes, i.e. the interspike intervals ($\{I_i\}$) are shown with solid lines; the sum of n adjacent ISIs form the n th order intervals $T_{n,i}$ (dashed lines).

More useful is this relation in the Fourier domain relating the power spectrum to the Fourier transform of the n th-order interval density $\tilde{\rho}_n$ as follows (see Holden, 1976)

$$\begin{aligned} S(f) &= 2\text{Re} \int_0^\infty d\tau e^{2\pi i f \tau} r \left[\delta(\tau) + \sum_{n=1}^\infty \rho_n(\tau) - r \right] \\ \Rightarrow S(f) &= r \left[1 - r\delta(f) + \sum_{n=1}^\infty \tilde{\rho}_n(f) + \tilde{\rho}_n^*(f) \right]. \end{aligned} \quad (1.63)$$

A general remark on the characteristics of the shot noise in Eqn (1.56) is that since a δ spike train has infinite variance, the correlation time defined by Eqn (1.21) is always zero. This is so even if there is some memory in the spiking, i.e. for firing with a strong refractory period.

1.4.2 Poisson process

The simplest process is obtained if we say that the spiking depends only on the spike rate $r(t)$ and not on the past. Let us assume that this rate is a constant.

Three different simulation algorithms The Poisson statistics suggest three approaches. (i) In each time step we draw a random number $a_i \in [0, 1]$; if $a_i < r\Delta t$ a spike is assigned to this time step, i.e. $x_i(t_i) = 1/(\Delta t)$; (ii) starting at time t_0 , we draw exponentially distributed intervals I_i with mean $\langle I \rangle = 1/r$ and obtain the spike times recursively from $t_i = t_{i-1} + I_i$; (iii) we take a large interval $[0, T]$ and distribute $N \approx rT$ points randomly and uniformly on this interval; for a much shorter interval $T' \ll T$, the points form a Poisson process in $[0, T']$.

Probability density The probability density $P(N, t) = \sum p_n(t)\delta(N - n)$ of the spike count $N(t)$ obeys the master equation

$$\dot{p}_n = rp_{n-1} - rp_n, \quad p_0(0) = 1. \quad (1.64)$$

Again we deal with a gain term (from realizations with $n - 1$ spikes up to time t and one additional spike occurring with probability $r\Delta t$) and a loss term (having spiked n times, an additional spike will result in leaving the state $x = n$). We can determine the interspike interval density by asking what the fraction of probability is that leaves $n = 0$ at t by determining the probability current rp_0 (the efflux out of the state). With $p_{-1}(t) \equiv 0$ we obtain – not surprisingly – a simple exponential decay $\rho(I) = r \exp[-rI]$ as we did for the two-state residence time density of one single state. The mean is given by the inverse rate ($\langle I \rangle = 1/r$) and the coefficient of variation CV of the ISI (which is the relative standard deviation of the ISI and a measure for the randomness of the interval) is one, $\text{CV} = \sqrt{\langle \Delta I^2 \rangle} / \langle I \rangle = 1$.

We can also determine the general solution of the spike count's probability (using $p_n(0) = 0$ for $n \geq 1$):

$$p_n(t) = r \int_0^t dt' e^{-r(t-t')} p_{n-1}(t') + \delta_{n,0} e^{-rt} = \frac{(rt)^n}{n!} e^{-rt} \quad (1.65)$$

(the latter relation can be proven by induction). This is the famous Poisson distribution. We obtain from it the n th-order interval density by calculating the current from the $(n-1)$ th to the n th state:

$$\rho_n(T_n) = r p_{n-1}(T_n) = r \frac{(rT_n)^{(n-1)}}{(n-1)!} e^{-rT_n} \Rightarrow \tilde{\rho}_n(f) = \frac{1}{(1 - 2\pi i f/r)^n}. \quad (1.66)$$

Correlation function and power spectrum of the spike train From the independence of the spike generation on the past it follows that the conditional distribution is simply $p(\tau|0) = \delta(\tau) + r$ and thus we obtain (using Eqn (1.61)) for the Poisson process' correlation function and its Fourier transform, i.e. the power spectrum

$$C(\tau) = r\delta(\tau) \Rightarrow S(f) = r. \quad (1.67)$$

The spectrum is flat and does not depend on frequency at all. The same result is obtained when using the Fourier transform of the n th-order interval density (1.66) in Eqn (1.63).

1.4.3 More general renewal process

Non-Poissonian firing is observed for many neurons. One reason is the refractory period of neural firing which makes neural firing more regular than Poissonian and results in an unimodal ISI density with a peak at finite ISI. Some other neurons show bursting, which can lead to a bimodal ISI density (indicating the most probable interburst and intraburst intervals) and correlations among interspike intervals. Ignoring the latter, we can describe and model more experimental data by just assuming independent intervals given by a (generally non-exponential) density $\rho(I)$.

Simulation algorithm Such a renewal process can be simulated by drawing independent random numbers I_i according to the given ISI density $\rho(I)$. The spiking times are then $t_i = t_{i-1} + I_i$.

Formula for the power spectrum For a renewal spike train, the n th-order interval density is just the probability for a sum of identically distributed *independent* random numbers which is given by the n -fold convolution of the ISI density. Even more convenient, its Fourier transform turns into the n -fold product of the Fourier transform $\tilde{\rho}(f)$ of the ISI density $\rho(T)$. In Eqn (1.63) we then get a

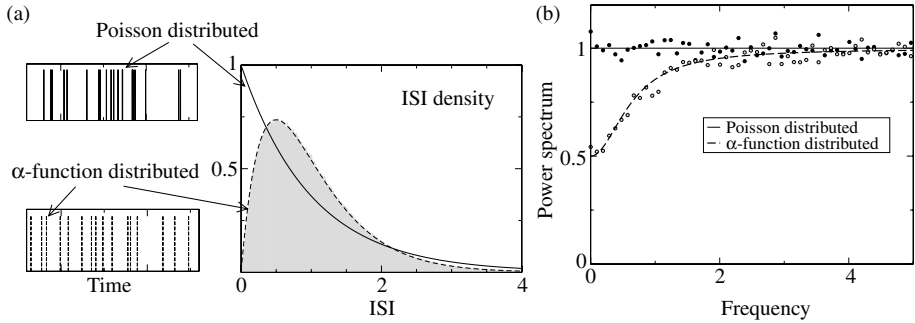


FIG. 1.6: Different renewal spike trains as indicated. ISI densities (a) and power spectra of spike trains (b). Symbols in (b) indicate results of numerical simulations. Both spike trains have a rate of $r = 1$.

simple geometric series that can be summed and yields

$$S(f) = r \frac{1 - |\tilde{\rho}(f)|^2}{|1 - \tilde{\rho}(f)|^2}. \quad (1.68)$$

A simple example As the ISI density we choose a so-called alpha function

$$\rho(I) = 4r^2 I \exp(-2rI) \quad (1.69)$$

which has the realistic feature of having a relative refractory period, as also observed in the firing of many neurons: very short intervals are unlikely. The ISI density starts at zero and attains its maximum at a finite ISI, unlike the exponential ISI density in the Poissonian case. We can generate intervals by adding up two exponentially distributed numbers³ with mean $(2r)^{-1}$. Samples of a Poissonian and an α -function spike train together with the ISI probability densities of these two processes are shown in Fig. 1.6(a); the respective power spectra are displayed in Fig. 1.6(b). An increased regularity in the spiking is already visible in the α -function spike train. This increased regularity leads to a drop of power at low frequencies in the power spectrum. The power spectrum can be calculated from the Fourier transform of the ISI density via Eqn (1.68) and reads

$$S(f) = r \left[1 - \frac{2r^2}{4r^2 + (\pi f)^2} \right]. \quad (1.70)$$

³The density (1.69) is the convolution of two exponential densities each with a rate $2r$; it agrees with the n th-order interval density of the Poisson process (1.66) for $n = 2$ and r replaced by $2r$.

Choosing an even more peaked ISI density, the spectrum will show even more reduced power at low frequencies and, additionally, a finite width peak around the inverse mean ISI and possibly smaller peaks at multiples of this frequency. Clustering of spikes, at the other extreme, will result in increased power at low frequency.

1.5 Relating the three processes

It is instructive to see formal mathematical relations between the three processes we have studied in this chapter. This also gives good intuition about many approximation schemes for stochastic systems.

Two-state process \rightarrow *spike train* Consider an asymmetric two-state process with $\sigma_- = 0, \sigma_+ = \tau_+^{-1}$, and fixed pulse width, i.e. $w_+(\tau) = \delta(\tau - \tau_+)$ ($\tilde{w}_+ = \exp(i\omega\tau_+)$); let the pulse width τ_+ go to zero. Then we obtain a spike train that clearly has an associated point process. If we neglect one of the time-scales in the two-state problem, we obtain the simpler point process. In this limit the formula for the power spectrum of the two-state process (Eqn (1.55)) turns into that for the spectrum of a renewal spike train (Eqn (1.68)).

Two-state process \rightarrow *Ornstein-Uhlenbeck process* It is also possible to obtain the Ornstein-Uhlenbeck process by means of the telegraph noise. If we add up many independent symmetric processes ($\langle \sigma_i(t) \rangle = 0$)

$$y_N = \frac{1}{N} \sum_{i=1}^N \sigma_i(t) \quad (1.71)$$

then for large N (as a consequence of the central limit theorem) the resulting sum will have approximately Gaussian (normal) statistics. A little more surprising perhaps is that the temporal structure of the correlations is maintained as the following simple calculation shows

$$\begin{aligned} C_Y(\tau) &= \langle y_N(t+\tau)y_N(t) \rangle = \frac{1}{N^2} \sum_{i,j=1}^N \langle \sigma_i(t)\sigma_j(t+\tau) \rangle \\ &= \frac{1}{N^2} \sum_{i=1}^N \langle \sigma_i(t)\sigma_i(t+\tau) \rangle = \frac{1}{N} C(\tau) \end{aligned} \quad (1.72)$$

where we have used the fact that $\sigma_i(t)$ and $\sigma_j(t)$ are independent and thus uncorrelated ($\langle \sigma_i(t)\sigma_j(t+\tau) \rangle = 0$ if $i \neq j$). Thus, the correlation function of the sum is just $1/N$ times the single process' correlation function. The resulting process has Gaussian statistics and has exponential correlation – it also approximates more and more accurately a continuous process since a single step changes the sum by only $1/N$. The only continuous process with Gaussian density and exponential correlation function is the Ornstein-Uhlenbeck process according to Doob's theorem. Numerical studies show that for many purposes $N \sim 12$ already yields statistics quite close to normal.

Poisson spike train \rightarrow white Gaussian noise Increasing the number of spikes and subtracting the mean of the spike train, we obtain a rapidly oscillating ‘spiky’ function, which is largely similar to that of Gaussian white noise. So if we take the original Ornstein–Uhlenbeck process and drive it with white Poissonian noise instead of white Gaussian noise we find

$$\dot{v} = -\gamma v + \varepsilon \left(\sum \delta(t - t_i) - r \right) \quad (1.73)$$

where we have subtracted the mean of the spike train, which is the steady-state spike rate ($\langle \sum \delta(t - t_i) \rangle = r$), and ε is an amplitude. We may interpret the equation by imagining that the particle is kicked around at random times t_i with each kick having the same impact ε . The density equation for this problem approaches the Fokker–Planck equation of an Ornstein–Uhlenbeck process if we set $\varepsilon = \sqrt{2D/r}$ and let r go to infinity. In this sense, the Poissonian shot noise with high rate and white Gaussian noise are very similar. Since the Fokker–Planck equation is a generalized diffusion equation, replacing the shot noise at a finite rate by Gaussian white noise is called the diffusion approximation. This is employed in neuroscience when dealing with subthreshold membrane fluctuations that are caused by synaptic shot noise. The latter can be, to a certain extent, approximated by white Gaussian noise.

1.6 Generation of random numbers

General idea Suppose we have a simple random number generator that provides uniformly distributed numbers x_i between 0 and 1 with $p_{\text{uni}}(x) = 1$ and we want to generate numbers y_i according to a distribution $p(y)$ (in general on $(-\infty, \infty)$). It should be possible to substitute the x_i into a nonlinear function, the corresponding values of which are then distributed with density $p(y)$. To get this function we first find a relationship between the two densities p_{uni} and $p(y)$ via the normalization integral

$$1 = \int_0^1 dx p_{\text{uni}} = \int_{-\infty}^{\infty} dy \left| \frac{dx}{dy} \right| p_{\text{uni}}(x(y)) = \int_{-\infty}^{\infty} dy p(y) \quad (1.74)$$

from which we get

$$\frac{dx}{dy} = p(y) \quad \Rightarrow \quad x(y) = F(y) = \int_{-\infty}^y dz p(z) \quad (1.75)$$

where $F(y)$ is the cumulative distribution telling us how probable a value *below* y is. By inverting (either numerically or analytically) the latter relation we obtain a way to generate the random numbers y_i from the uniformly distributed numbers x_i

$$y(x) = F^{-1}(x). \quad (1.76)$$

Exponentially distributed numbers Suppose we want to generate numbers according to $p(y) = r \exp(-ry)$. In this case it is easy to obtain $F(y) = 1 - \exp(-ry)$ and furthermore to invert this relation to obtain

$$y = -r^{-1} \ln(1 - x). \quad (1.77)$$

Equivalently, we could use $y = -r^{-1} \ln(x)$ because of the symmetry of x .

Gaussian numbers The formulas above can be generalized to multidimensional densities, which is particularly useful for generating Gaussian numbers. In this case we need *two* independent uniformly distributed random numbers x_1 and x_2 in order to obtain *two* independent Gaussian numbers y_1 and y_2 . The transformation is called the Box–Müller formula and reads

$$y_1 = \sqrt{2 \ln(x_1)} \cos(2\pi x_2) \quad (1.78)$$

$$y_2 = \sqrt{2 \ln(x_1)} \sin(2\pi x_2). \quad (1.79)$$

If these variables are normally distributed, the Jacobian (which replaces the simple derivative in the normalization integral) should yield a two-dimensional Gaussian, as can be checked by inserting

$$\left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| = \frac{1}{\sqrt{2\pi}} e^{-y_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-y_2^2/2} = p(y_1, y_2). \quad (1.80)$$

References

- Cox, D.R. (1962). *Renewal Theory*. Methuen, London.
- Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen, London.
- Gardiner, C. W. (1985). *Handbook of Stochastic Methods*. Springer-Verlag, Berlin.
- Gerstner, W. and Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.
- Hänggi, P. and Jung, P. (1995). Colored noise in dynamical systems. *Adv. Chem. Phys.*, **89**, 239.
- Holden, A. V. (1976). *Models of the Stochastic Activity of Neurones*. Springer-Verlag, Berlin.
- Langevin, P. (1908). Sur la theorie du mouvement brownien. *Comptes Rendus Acad. Sci. (Paris)*, **146**, 530–533.
- Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York.
- Ricciardi, L. (1977). *Diffusion Processes and Related Topics in Biology*. Springer.
- Risken, H. (1984). *The Fokker–Planck Equation*. Springer, Berlin.
- Risken, H. (1996). *The Fokker–Planck Equation: Methods of Solutions and Applications*. Springer-Verlag, New York.

- Stratonovich, R. L. (1967). *Topics in the Theory of Random Noise*. Routledge, New York.
- Tuckwell, H. (1989). *Stochastic Processes in the Neurosciences*. SIAM, Philadelphia.
- Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology*. Cambridge University Press, Cambridge.
- Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Phys. Rev.*, **36**, 823.
- van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam.