Maciej Lisicki & Miłosz Panfil
October 29, 2022

5 Stochastic processes and their description

We will now define stochastic processes and discuss their properties and methods of describing them. First, we consider a random variable X and a real parameter t. For convenience, we will refer to t as time, but this can be a different variable relevant to the process at hand. We define a family of random variables Y, generated by a mapping f such that

$$Y(t) = Y_X(t) = f(X, t). \tag{5.1}$$

We will call $Y(t)$ a *random function* or a *stochastic process*. If the value of the parameter $t = t_0$ is fixed, then $Y(X, t_0)$ is a random variable. If a particular realisation of X is chosen, $x \in X$, then $Y(x, t) = f(x, t)$ is a *trajectory* or *realisation* of the process. Thus, a stochastic process may be regarded as an ensemble of its realisations.

The main question in this Chapter will be – How can we describe such an ensemble? How can we characterise a stochastic process? We will present four different approaches below.

The underlying distribution   The first case is a convenient situation when we know the mapping f and the distribution $\rho(x)$ of the underlying random variable X. Then we know everything about the process and any quantity can be calculated directly. It is, however, a rare possibility in the context of natural sciences, where we typically have access to some observables of a process.

Averages of the process   The second scenario is that we know the averages, and generally moments of the distribution of Y. The basic quantity is the mean

$$\langle Y(t) \rangle = \int Y_X(t)\rho(x)\mathrm{d}x = \int f(x, t)\rho(x)\mathrm{d}x, \tag{5.2}$$

and higher-order moments, defined as

$$\langle Y(t_1) \cdots Y(t_n) \rangle = \int f(x, t_1) \cdots f(x, t_n)\mathrm{d}x. \tag{5.3}$$

Out of those, particularly important is the two-point average, called the *autocorrelation function* $\kappa(t_1, t_2)$, defined as

$$\kappa(t_1, t_2) = \langle\langle Y(t_1)Y(t_2)\rangle\rangle = \langle Y(t_1)Y(t_2) \rangle - \langle Y(t_1) \rangle \langle Y(t_2) \rangle, \tag{5.4}$$

where the double brackets $\langle\langle \cdot \rangle\rangle$ denote average of the quantity shifted by its mean. Note that $\kappa(t_1, t_2) = \kappa(t_2, t_1)$. For $t_1 = t_2$, autocorrelation function reduces to the variance,

$$\langle\langle Y^2(t) \rangle\rangle = \sigma^2(t). \tag{5.5}$$

The hierarchy of distribution functions Assume that we describe the trajectory of a stochastic process by measuring the probability of the trajectories passing through a number of gates' placed at r points in the (y, t) space. In other words, we are looking for the probability that at time t_i the value of the process was between y_i and $y_i + \mathrm{d}y_i$, with $i = 1, 2, \ldots, r$. The probability of such an event can be written using the joint probability distribution

$$P_r(y_1, t_1; \ldots; y_r, t_r)\mathrm{d}y_1 \cdots \mathrm{d}y_r. \tag{5.6}$$

We can relate this probability distribution to the underlying distribution for the variable X by writing it as an average using Eq. (2.45). Thus it reads

$$P_r(y_1, t_1; \ldots; y_r, t_r) = \langle \delta(y_1 - f(x, t_1)) \cdots \delta(y_r - f(x, t_r)) \rangle_X, \tag{5.7}$$

where the last average is calculated w.r.t. the density $\rho(X)$. We can now write the averages, such as that in Eq. (5.3), as

$$\langle Y(t_1) \cdots Y(t_n) \rangle = \int y_1 \cdots y_r P_r(y_1, t_1; \ldots; y_r, t_r)\mathrm{d}y_1 \cdots \mathrm{d}y_r. \tag{5.8}$$

The hierarchy of joint probability distributions,

$$\{P_1, P_2, P_3, \ldots\},$$

can thus be used to calculate all possible averages. These functions have to satisfy four *consistency conditions*, holding for all n

(a) $P_n \geq 0$,

(b) P_n does not change on swapping two pairs (y_k, t_k) and (y_l, t_l),

(c) $\int P_r(y_1, t_1; \ldots; y_r, t_r)\mathrm{d}y_r = P_{r-1}(y_1, t_1; \ldots; y_{r-1}, t_{r-1})$,

(d) $\int P_1(y_1, t_1)\mathrm{d}y_1 = 1$.

Kolmogorov showed that a set of functions satisfying (a)-(d) is a complete specification of a stochastic process. In fact, it is over-complete, since we can omit any finite number of P_m without losing information, as they can always be recovered from higher distributions thanks to the property (c).

Conditional distributions Imagine now that we take a different approach and look only at trajectories which passed through the point (y_1, t_1). The probability that a realisation of the process takes the value of y_2 at t_2, given that its value at t_1 was y_1, is described by the *conditional* probability distribution

$$P_{1|1}(y_2, t_2|y_1, t_1)\mathrm{d}y_2. \tag{5.9}$$

To compute it, we select from the ensemble of trajectories a sub-ensemble of only those that pass through y_1 at t_1. The fraction of the sub-ensemble that passes through y_2 at t_2

is given exactly by Eq. (5.9). Clearly, if we sum over all possible targets, we count in all the trajectories of the sub-ensemble, so

$$\int P_{1|1}(y_2, t_2|y_1, t_1)\mathrm{d}y_2 = 1. \tag{5.10}$$

More generally, we can specify more conditions and more targets, leading to multidimensional conditional distributions, defined by

$$P_{r|k}(y_{k+1}, t_{k+1}; \ldots; y_{r+k}, t_{r+k}|y_1, t_1; \ldots; y_k, t_k) = \frac{P_{r+k}(y_1, t_1; \ldots; y_{r+k}, t_{r+k})}{P_k(y_1, t_1; \ldots; y_k, t_k)}, \tag{5.11}$$

for which an elegant interpretation is that the joint probability of being at points $(y_1, t_1; \ldots; y_{r+k}, t_{r+k})$ is the probability of being at points $(y_1, t_1; \ldots; y_r, t_r)$ given that we were at points $(y_{r+1}, t_{r+1}; \ldots; y_{r+k}, t_{r+k})$ times the probability of being at these points, so simply $P_{r+k} = P_{r|k}P_k$. In particular, for $r = 1$ we get

$$P_{k+1}(y_1, t_1; \ldots; y_{k+1}, t_{k+1}) = P_{1|k}(y_{k+1}, t_{k+1}|y_1, t_1; \ldots; y_k, t_k)P_k(y_1, t_1; \ldots; y_k, t_k), \tag{5.12}$$

from which we see that if we know the set

$$\left\{P_1, P_{1|1}, P_{1|2}, \ldots\right\}$$

so all subsequent conditional probabilities $P_{1|k}$, we can again fully describe a stochastic process. The interpretation is the following – all we need to know is to consider one 'gate' and values of the process at specific points.

For further considerations, we will also need to define a *stationary process* that is one not affected by a shift of time. For such a process, for any τ we have

$$\langle Y(t_1 + \tau) \cdots Y(t_n + \tau)\rangle = \langle Y(t_1) \cdots Y(t_n)\rangle. \tag{5.13}$$

In particular, it follows that for a stationary process

$$\langle Y(t)\rangle = \text{const}, \tag{5.14}$$

is time-independent, and the correlation function

$$\kappa(t_1, t_2) = \kappa(t_1 - t_2) = \kappa(|t_1 - t_2|), \tag{5.15}$$

where the last equality follows from the symmetry $\kappa(t) = \kappa(-t)$. In a typical situation the autocorrelation function of a stationary process decays with time, so there exists a characteristic time τ_c, called the correlation time, such that for $t > \tau_c$ we have $\kappa(t) \approx 0$. Thus, for times $t \gg \tau_c$ two random variables $Y(X, t_1)$ and $Y(X, t_1 + \tau_c)$ are uncorrelated (but not independent).