APPENDIX C

Basic notions on stochastic processes

Similar to the previous one, this Appendix introduces further notions of probability theory, namely now some basic definitions and results on stochastic processes.

C.1 Definitions

Consider a random variable X with sample space Ω and probability distribution p_X. Let t denote an additional variable, which takes an infinite number (countable or not) of values in some set $\mathcal{I}$. Any function f on the product $\mathcal{I} \times \Omega$ defines an infinite number of stochastic variables

$$Y_X(t) = f(t, X) \tag{C.1}$$

labeled by t. Such a quantity is referred to as a *random function* of the variable t. In case the latter stands for time, $Y_X(t)$ is called a *stochastic process*.
Taking at every t a realization x of the random variable X, one obtains a *realization* of the process or *sample function*

$$Y_x(t) = f(t, x), \tag{C.2}$$

which is a function in the usual sense of analysis. In turn, fixing $t \in \mathcal{I}$, $Y_X(t)$ is a random variable in the sense of Appendix B.

The random function can be multidimensional, $\mathbf{Y}_X(t) = Y_X^1(t), Y_X^2(t), \ldots, Y_X^D(t)$. This is in particular often the case when the random variable itself is multidimensional, $\mathbf{X}$.

> In this Appendix, the random functions we shall consider will take their values in (subsets of) $\mathbb{R}$ (in the one-dimensional case) or $\mathbb{R}^D$ with $D > 1$. The results can be extended to further spaces, provided a product can be defined on them, so that e.g. the integrand of Eq. (C.4) makes sense.

In turn, there might be more than one additional variable, that is the random function is labeled by a multidimensional variable. In physics, this corresponds for instance to the case of random fields, whose value is a stochastic variable at each instant and at each point in space.

For the sake of brevity, we shall hereafter refer to the variable t as "time", and assume that it takes its values in (an interval of) $\mathbb{R}$: we thus consider *continuous-time stochastic processes*. The points of the t-axis will be referred to as "instants".

C.1.1 Averages and moments

Using the probability distribution $p_X(x)$ of the random variable, one easily defines averages as in Sec. B.2. For instance, the (single-time) *sample average* or *ensemble average* of the stochastic process $Y_X(t)$ is given by

$$\langle Y_X(t) \rangle \equiv \int_\Omega Y_x(t)\, p_X(x)\, \mathrm{d}x, \tag{C.3a}$$

where the integration runs over the sample space Ω.

It is sometimes helpful to view this sample average differently: let $Y_{x^{(r)}}(t)$ with $r = 1, 2, \ldots, N$ denote different realizations of the process at the same instant t, where the corresponding realizations $x^{(r)}$ of the random variable X are distributed according to p_X. Then the sample average is given by

$$\langle Y_X(t) \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{r=1}^{N} Y_{x^{(r)}}(t), \tag{C.3b}$$

in accordance with the definition of the probability distribution p_X.

More generally, one can define multi-time averages, or higher moments, as follows. Let $n \in \mathbb{N}^*$, and consider n (not necessarily different) values $t_1, t_2, \ldots, t_n$ of the time variable. The n-th moment is then defined as

$$\big\langle Y_X(t_1) Y_X(t_2) \cdots Y_X(t_n) \big\rangle \equiv \int_\Omega Y_x(t_1) Y_x(t_2) \cdots Y_x(t_n)\, p_X(x)\, \mathrm{d}x. \tag{C.4}$$

By combining first and second moments, one obtains the *autocorrelation function*

$$\kappa(t_1, t_2) \equiv \big\langle \big[Y_X(t_1) - \langle Y_X(t_1) \rangle \big] \big[Y_X(t_2) - \langle Y_X(t_2) \rangle \big] \big\rangle \tag{C.5a}$$

$$= \big\langle Y_X(t_1) Y_X(t_2) \big\rangle - \big\langle Y_X(t_1) \big\rangle \big\langle Y_X(t_2) \big\rangle. \tag{C.5b}$$

In case the random function is multidimensional, this autocorrelation function is replaced by the *correlation matrix*

$$\kappa^{ij}(t_1, t_2) \equiv \big\langle \big[Y_X^i(t_1) - \langle Y_X^i(t_1) \rangle \big] \big[Y_X^j(t_2) - \langle Y_X^j(t_2) \rangle \big] \big\rangle, \tag{C.6}$$

whose diagonal coefficients are *autocorrelations*, while the off-diagonal elements are referred to as *cross-correlations*.

Generalizing the concept of characteristic function for a random variable [see Eqs. (B.6a) and (B.17)], one defines for a given stochastic process $Y_X(t)$ the *characteristic functional*

$$G_{Y_X}[k(t)] \equiv \left\langle \exp\left[\mathrm{i} \int k(t)\, Y_X(t)\, \mathrm{d}t \right] \right\rangle, \tag{C.7}$$

where the integral runs over the domain in which the time variable t takes its values, while $k(t)$ is a test function defined over this domain. One easily checks that expanding this characteristic functional in powers of k yields the n-time averages (C.4) as functional derivatives.

C.1.2 Distribution functions

Consider a stochastic process $Y_X(t)$. The probability density that $Y_X(t)$ takes the value y at time t, also called *single-time density*, is trivially given by

$$p_{Y,1}(t, y) \equiv \int_\Omega \delta\big(y - Y_x(t)\big)\, p_X(x)\, \mathrm{d}x. \tag{C.8}$$

Introducing now *different* instants $t_1, t_2, \ldots, t_n$ with $n > 1$, the joint probability for Y_X to take the value y_1 at t_1, the value y_2 at t_2, ..., and the value y_n at t_n is given by

$$p_{Y,n}(t_1, y_1; t_2, y_2; \ldots; t_n, y_n) \equiv \int_\Omega \delta\big(y_1 - Y_x(t_1)\big)\, \delta\big(y_2 - Y_x(t_2)\big) \cdots \delta\big(y_n - Y_x(t_n)\big)\, p_X(x)\, \mathrm{d}x. \tag{C.9}$$

$p_{Y,n}$ is referred to as *n-time density* or *n-point density*. With its help, the n-th moment (C.4) can be rewritten as

$$\big\langle Y_X(t_1) Y_X(t_2) \cdots Y_X(t_n) \big\rangle = \int y_1 y_2 \cdots y_n\, p_{Y,n}(t_1, y_1; t_2, y_2; \ldots; t_n, y_n)\, \mathrm{d}y_1\, \mathrm{d}y_2 \cdots \mathrm{d}y_n, \tag{C.10}$$

where the integral runs over (n copies of the) space on which the realizations of $Y_X(t)$ take their values.

One easily checks that the n-point densities satisfy the following four properties:

- $p_{Y,n}(t_1, y_1; t_2, y_2; \ldots; t_n, y_n) \geq 0$ for every $n \geq 1$ and for all $(t_1, y_1), \ldots, (t_n, y_n)$. (C.11a)
- $p_{Y,n}$ is symmetric under the exchange or two pairs (t_j, y_j) and (t_k, y_k) for all j, k. (C.11b)
- The densities obey for all $m < n$ and for all $(t_1, y_1), \ldots, (t_m, y_m)$ the consistency conditions

$$p_{Y,m}(t_1, y_1; t_2, y_2; \ldots; t_m, y_m) = \int p_{Y,n}(t_1, y_1; \ldots; t_m, y_m; t_{m+1}, y_{m+1}; \ldots; t_n, y_n)\, \mathrm{d}y_{m+1} \cdots \mathrm{d}y_n. \tag{C.11c}$$

  That is, every $p_{Y,n}$ encompasses all information contained in all $p_{Y,m}$ with $m < n$.

- The single-time density $p_{Y,1}$ is normalized to unity: $\int p_{Y,1}(t, y)\, \mathrm{d}y = 1$. (C.11d)

Remarks:

∗ Property (C.11b) allows one to order the time arguments at will.

∗ The definition of the densities need not be extended to the case where two or more of the time arguments, say t_j and t_k, are equal, since it in that case, only $y_j = y_k$ is meaningful—the probability that the process takes two different values at the same instant is obviously zero.

∗ Relation (C.11c) expresses $p_{Y,m}$ as a marginal distribution of $p_{Y,n}$, cf. Eq. (B.20).

∗ Starting from the normalization (C.11d) and using Eq. (C.11c), one easily proves recursively that every n-point density $p_{Y,n}$ is normalized to unity as well

$$\int p_{Y,n}(t_1, y_1; t_2, y_2; \ldots; t_n, y_n)\, \mathrm{d}y_1\, \mathrm{d}y_2 \cdots \mathrm{d}y_n = 1. \tag{C.12}$$

Together with the positivity condition (C.11a), this means that the n-point densities possess the "good properties" (B.1) of probability distributions.

Conditional n-point densities

One also introduces *conditional probability densities*, by considering the probability density that Y_X takes the value y_1 at t_1, the value y_2 at t_2, ..., and the value y_m at t_m, knowing that it takes the value y_{m+1} at t_{m+1}, the value y_{m+2} at t_{m+2}, ..., and the value y_n at t_n:

$$p_{Y,m|n-m}(t_1, y_1; \ldots; t_m, y_m \,|\, t_{m+1}, y_{m+1}; \ldots; t_n, y_n) = \frac{p_{Y,n}(t_1, y_1; \ldots; t_m, y_m; t_{m+1}, y_{m+1}; \ldots; t_n, y_n)}{p_{Y,n-m}(t_{m+1}, y_{m+1}; \ldots; t_n, y_n)} \tag{C.13}$$

[cf. Bayes' theorem (B.22)].

Remarks:

∗ Working recursively, one finds that every n-point density can be expressed as the product of conditional probability densities $p_{Y,1|m}$, with m ranging from $n-1$ to 1, and of a single-time density:

$$\begin{aligned} p_{Y,n}(t_1, y_1; \ldots; t_n, y_n) = {} & p_{Y,1|n-1}(t_n, y_n \,|\, t_1, y_1; \ldots; t_{n-1}, y_{n-1}) \\ & \times p_{Y,1|n-2}(t_{n-1}, y_{n-1} \,|\, t_1, y_1; \ldots; t_{n-2}, y_{n-2}) \cdots p_{Y,1|1}(t_2, y_2 \,|\, t_1, y_1) \\ & \times p_{Y,1}(t_1, y_1), \end{aligned} \tag{C.14}$$

which is easily interpreted.

∗ Writing down the previous identity for $n = 3$ and integrating over y_2 under consideration of the consistency condition (C.11c), one finds

$$p_{Y,2}(t_1, y_1; t_3, y_3) = \int p_{Y,1|2}(t_3, y_3 | t_1, y_1; t_2, y_2)\, p_{Y,1|1}(t_2, y_2 | t_1, y_1)\, p_{Y,1}(t_1, y_1)\, \mathrm{d}y_2.$$

Dividing by $p_{Y,1}(t_1, y_1)$, which can readily be factorized out of the integral, then yields

$$p_{Y,1|1}(t_3, y_3 | t_1, y_1) = \int p_{Y,1|2}(t_3, y_3 | t_1, y_1; t_2, y_2)\, p_{Y,1|1}(t_2, y_2 | t_1, y_1)\, \mathrm{d}y_2. \tag{C.15}$$

Again, this identity has an intuitive meaning. Mathematically, it is an integral-functional equation for the conditional probability $p_{Y,1|1}$, involving the integration kernel $p_{Y,1|2}$. Similarly, one can write down an analogous equation for $p_{Y,1|2}$, with $p_{Y,1|3}$ as integration kernel; and more generally, a whole hierarchy of integral-functional relations, where the equation for $p_{Y,1|n}$ admits $p_{Y,1|n+1}$ as integration kernel.

C.2 Some specific classes of stochastic processes

The knowledge of all n-point probability densities $p_{Y,n}$ for a random function $Y_X(t)$ allows the computation of all n-point averages and thus replaces the knowledge of the probability density p_X. Accordingly, we shall from now on drop any reference to the random variable X and denote a stochastic process more simply as $Y(t)$, and its realizations as $y(t)$.

C.2.1 Centered processes

A stochastic process $Y(t)$ is called *centered* if its single-time average $\langle Y(t) \rangle$ is identically vanishing for any time t.

Given an arbitrary stochastic process $Y(t)$, the process $Z(t) \equiv Y(t) - \langle Y(t) \rangle$ is obviously centered. One checks at once that $Y(t)$ and the associated process $Z(t)$ share the same autocorrelation function $\kappa(t_1, t_2)$.

C.2.2 Stationary processes

A stochastic process $Y(t)$ is said to be *stationary* when all its moments are invariant under arbitrary shifts of the origin of times, that is when for all $n \in \mathbb{N}^*$, $\Delta t \in \mathbb{R}$ and n-uplets t_1, t_2, …, t_n, one has the identity

$$\big\langle Y(t_1 + \Delta t) Y(t_2 + \Delta t) \cdots Y(t_n + \Delta t) \big\rangle = \big\langle Y(t_1) Y(t_2) \cdots Y(t_n) \big\rangle. \tag{C.16a}$$

In particular, the single-time average $\langle Y(t) \rangle$ is time-independent, so that it is convenient to work with the associated centered process $Y(t) - \langle Y \rangle$.

Remark: An equivalent definition is that all n-point densities of the process are invariant under arbitrary time translations:

$$p_{Y,n}(t_1 + \Delta t, y_1; t_2 + \Delta t, y_2; \ldots; t_n + \Delta t, y_n) = p_{Y,n}(t_1, y_1; t_2, y_2; \ldots; t_n, y_n). \tag{C.16b}$$

The autocorrelation function $\kappa(t_1, t_2)$ of a stationary process only depends on the time difference $\tau \equiv t_2 - t_1$, and is an even function of τ (i.e. it only depends on $|\tau|$):

$$\kappa(\tau) = \big\langle Y(t) Y(t + \tau) \big\rangle - \big\langle Y \big\rangle^2. \tag{C.17}$$

A widespread case in physics is that of processes whose autocorrelation function only takes significant values over some scale $|\tau| \lesssim \tau_c$—the *autocorrelation time*—, and become negligible for $|\tau| \gg \tau_c$.

For a centered stationary multidimensional stochastic process $\mathbf{Y}(t)$ with components $Y^1(t)$, $Y^2(t)$, …, defining [cf. the correlation matrix (C.6)]

$$\kappa^{ij}(\tau) \equiv \big\langle Y^i(t) Y^j(t + \tau) \big\rangle, \tag{C.18a}$$

which may not be even in τ, one has the obvious property

$$\kappa^{ij}(\tau) = \kappa^{ji}(-\tau). \tag{C.18b}$$

Stationary processes are conveniently characterized by their spectral properties, which follow from considering their (discrete) Fourier transform. This idea will be further discussed in Sec. C.3.

C.2.3 Ergodic processes

A stationary stochastic process $Y(t)$ is called *ergodic*, when any single realization $y(t)$ contains all statistical information on the whole process, i.e. allows one to compute all possible n-point averages.

Let $y(t)$ denote a given realization. The time average of the stationary process over the finite interval $[t - \frac{\mathcal{T}}{2}, t + \frac{\mathcal{T}}{2}]$, where $\mathcal{T} > 0$, is defined as

$$\overline{Y(t)}^{\mathcal{T}} \equiv \frac{1}{\mathcal{T}} \int_{t-\frac{\mathcal{T}}{2}}^{t+\frac{\mathcal{T}}{2}} y(t')\,\mathrm{d}t'. \tag{C.19}$$

This average depends on t, $\mathcal{T}$ and the realization y. In the limit of large $\mathcal{T}$, the average becomes the *time average* $\overline{Y}$ of the process,

$$\overline{Y} = \lim_{\mathcal{T}\to+\infty} \overline{Y(t)}^{\mathcal{T}} \equiv \lim_{\mathcal{T}\to+\infty} \frac{1}{\mathcal{T}} \int_{t-\frac{\mathcal{T}}{2}}^{t+\frac{\mathcal{T}}{2}} y(t')\,\mathrm{d}t'. \tag{C.20}$$

As hinted at by the notation, $\overline{Y}$ no longer depends on t and $\mathcal{T}$, thanks to the assumed stationarity; yet it still depends on the specific realization of the process. If it is independent of the realization, then the time average $\overline{Y}$ is equal to the (time-independent, since $Y(t)$ is stationary) ensemble average $\langle Y \rangle$.

A stochastic process is ergodic when the identity between time average and sample average holds for all products of $Y(t)$ at different times, i.e. for all moments.

C.2.4 Gaussian processes

A stochastic process $Y(t)$ is called *Gaussian process* if all its n-point densities (C.9) are Gaussian distributions. Equivalently, for every nonnegative integer n and every choice of arbitrary instants $t_1, t_2, \ldots, t_n$, the n-dimensional random variable with components $Y(t_1), \ldots, Y(t_n)$ is Gaussian-distributed.

The corresponding characteristic functional reads

$$G_Y[k(t)] = \exp\left[\mathrm{i}\int k(t)\big\langle Y(t)\big\rangle \mathrm{d}t - \frac{1}{2}\int k(t_1)k(t_2)\,\kappa(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\right], \tag{C.21}$$

so that the process is entirely determined by its single-time average $\langle Y(t)\rangle$ and its autocorrelation function $\kappa(t_1,t_2)$—or equivalently, by the single- and two-time densities $p_{Y,1}$, $p_{Y,2}$. For instance, one can show[(82)] that for even n, the n-point moment is given by

$$\big\langle Y(t_1)Y(t_2)\cdots Y(t_n)\big\rangle = \sum \big\langle Y(t_j)Y(t_k)\big\rangle \cdots \big\langle Y(t_l)Y(t_m)\big\rangle,$$

where the sum runs over all possible pairings of the indices $1, 2, \ldots, n$, while the product for a given pairing involves all $n/2$ corresponding pairs.

If $Y(t)$ is a Gaussian process, then the associated centered process $Z(t) \equiv Y(t) - \langle Y(t)\rangle$ is also Gaussian, and all moments of odd order of $Z(t)$ vanish.

(82) This is (part of) the Isserlis[(bu)] theorem, better known in physics as Wick's[(bv)] theorem.

(bu) L. Isserlis, 1881–1966 (bv) G.-C. Wick, 1909–1992