

Lecture 4: Classifications and Numerical Characteristics of Stochastic Processes

Lecturer: Ziqiao Wang

In this lecture,¹ we move from a finite collection of random variables to a random quantity indexed by time. We will distinguish sample paths from distributions, classify processes by their time and state spaces, and describe their dependence through mean, correlation, and covariance functions.

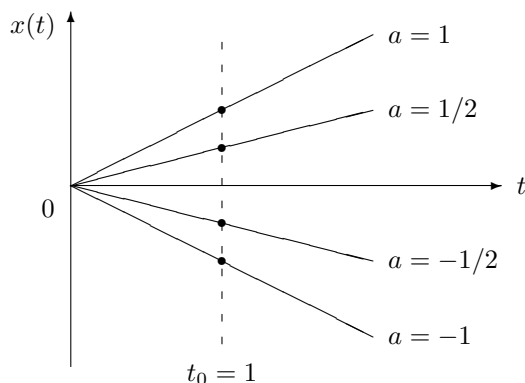
1 Random Processes and Sample Functions

Definition 1.1 (Stochastic process). A stochastic process is a family of random variables $\{X(t) : t \in \mathcal{T}\}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The set I is the *index set*; a set S containing the possible values is the *state space*.

The full notation $X(t, \omega)$ displays two arguments with different roles. The randomness comes from the experiment's outcome ω ; time tells us which random variable we observe.

What is fixed?	What remains?
$t = t_0$	$\omega \mapsto X(t_0, \omega)$: a random variable.
$\omega = \omega_0$	$t \mapsto X(t, \omega_0)$: a sample path or realization.
Both t_0 and ω_0	$X(t_0, \omega_0)$: a number.

Example 1 (A random slope). Draw A uniformly from $[-1, 1]$ *once*, and set $X(t) = At$, $t \geq 0$. If the draw is $A = a$, the entire observed path is the line $x(t) = at$. At time $t = 1$, the observed value is the random variable $X(1) = A$; at time $t = 2$, it is $X(2) = 2A$. These two values share the same random slope.



One line: fix the outcome.

Vertical slice: fix time.

Four representative paths
from a continuum of paths.

An observed trajectory can be deterministic once the outcome is fixed, even though the process is random before that outcome is known. A list of possible curves alone is not a complete probabilistic model: we must also say how the curves are selected.

2 Classification by Time and State Space

The time index and the observed values answer different questions: *when can we observe the process*, and *what values can it take*? The standard four cases are summarized below. Discrete means a finite or countably infinite set; continuous time usually means an interval of the real line.

¹Reading: Leon-Garcia, Sections 9.1–9.3. Section 9.6 is a reading preview for the later lectures on second-order processes.

Time	State space	Example
Discrete	Discrete	$N_n = \sum_{j=1}^n B_j$, $B_j \in \{0, 1\}$: a count after n trials.
Discrete	Continuous	$U_n \sim \text{Uniform}(0, 1)$: a sequence of real-valued measurements.
Continuous	Discrete	$J(t) = \mathbf{1}_{\{T \leq t\}}$: whether an arrival has occurred by time t .
Continuous	Continuous	$X(t) = At$, $A \sim \text{Uniform}[-1, 1]$: a random slope.

For discrete time we write X_n rather than $X(n)$, with $n \in \{0, 1, 2, \dots\}$ or $n \in \mathbb{Z}$. Both notations refer to the same kind of indexed family.

Example 2. Let $T \sim \text{Uniform}(0, 1)$ and $J(t) = \mathbf{1}_{\{T \leq t\}}$, $t \geq 0$. A realization $T = u$ produces a path that is zero before u and one from u onward. Thus J is continuous-time and discrete-state, even though its paths jump. The same arrival time T determines all time samples.

Other classifications. Independent samples, independent increments, stationarity, and the Markov property describe additional structure. These labels may overlap. In particular, a counting process can have independent increments without having independent values at different times.

3 Finite-Dimensional Distributions

Knowing the distribution at each time does not tell us how observations at different times are related. To ask about several observations, we sample a random vector from the process.

Definition 3.1 (Finite-dimensional distributions). For any positive integer k and any times $t_1, \dots, t_k \in \mathcal{T}$, the joint CDF is

$$F_{X(t_1), \dots, X(t_k)}(x_1, \dots, x_k) = \mathbb{P}(X(t_1) \leq x_1, \dots, X(t_k) \leq x_k). \quad (1)$$

The collection of these joint laws over all finite time selections is called the family of finite-dimensional distributions.

For discrete states we can instead use joint PMFs. Joint PDFs may be used *if those joint distributions have densities*; real-valued samples alone do not guarantee this. The finite-dimensional laws specify all probabilities involving finitely many time samples.

Example 3. Let B take -1 and 1 with equal probability. On $n \in \{0, 1, 2, \dots\}$, define two processes

$$X_n = B, \quad Y_n = (-1)^n B.$$

Both X_n and Y_n are uniform on $\{-1, 1\}$ at every time. However,

$$\mathbb{P}(X_0 = 1, X_1 = 1) = \frac{1}{2}, \quad \mathbb{P}(Y_0 = 1, Y_1 = 1) = 0.$$

The first process stays at its initial value, whereas the second alternates. Their one-dimensional distributions coincide, but their two-dimensional distributions do not.

Marginalizing a joint law must recover the law of the retained samples; for example,

$$\lim_{y \rightarrow \infty} F_{X(s), X(t)}(x, y) = F_{X(s)}(x).$$

For an i.i.d. sequence with common CDF F , the joint CDF at *distinct* indices is $\prod_{j=1}^k F(x_j)$.

4 Mean, Correlation, and Covariance Functions

Finite-dimensional distributions give a detailed description, but they can be cumbersome. Numerical characteristics summarize their first and second moments. For all second-order calculations below, we use real-valued processes with

$$\mathbb{E}[X(t)^2] < \infty \quad \text{for every } t \in \mathcal{T}.$$

This finite-second-moment assumption guarantees that the expectations below exist. Defining a mean alone only requires $\mathbb{E}[|X(t)|] < \infty$.

Definition 4.1 (First- and second-order characteristics). The mean and variance functions are

$$m_X(t) = \mathbb{E}[X(t)], \tag{2}$$

$$\text{Var}[X(t)] = \mathbb{E}[(X(t) - m_X(t))^2]. \tag{3}$$

The autocorrelation and autocovariance functions are

$$R_X(s, t) = \mathbb{E}[X(s)X(t)], \tag{4}$$

$$C_X(s, t) = \mathbb{E}[(X(s) - m_X(s))(X(t) - m_X(t))]. \tag{5}$$

These characteristics are deterministic functions of time. The prefix “auto” indicates that both samples come from the same process.

Expanding the centered product in Eq.(5) gives the useful identity

$$C_X(s, t) = R_X(s, t) - m_X(s)m_X(t). \tag{6}$$

In addition,

$$R_X(t, t) = \mathbb{E}[X(t)^2], \quad C_X(t, t) = \text{Var}[X(t)] = R_X(t, t) - m_X(t)^2.$$

Thus autocorrelation includes the contribution of the mean, while autocovariance describes fluctuations around the mean. They coincide when the mean function is identically zero.

How to calculate them. At fixed times, use the ordinary expectation of the sampled variables. For example, when the relevant densities exist,

$$m_X(t) = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx, \quad R_X(s, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X(s), X(t)}(x, y) dx dy.$$

Note that $R_X(s, t)$ is an unnormalized second moment. The normalized coefficient is

$$\rho_X(s, t) = \frac{C_X(s, t)}{\sqrt{C_X(s, s)C_X(t, t)}}, \quad C_X(s, s) > 0, \quad C_X(t, t) > 0.$$

It lies in $[-1, 1]$; $R_X(s, t)$ has no such universal range. The coefficient is undefined if either sample has zero variance.

Two processes. For real second-order processes X, Y on the same probability space, define

$$R_{XY}(s, t) = \mathbb{E}[X(s)Y(t)], \quad C_{XY}(s, t) = R_{XY}(s, t) - m_X(s)m_Y(t).$$

They are called *orthogonal* if $R_{XY}(s, t) = 0$ for every s, t , and *uncorrelated* if $C_{XY}(s, t) = 0$ for every s, t . Independent processes have zero cross-covariance, but need not be orthogonal when their means are nonzero. Here independence means that every finite vector of X -samples is independent of every finite vector of Y -samples.

5 From a covariance function to a covariance matrix

Sample the process at $t_1, \dots, t_k$ and write

$$\mathbf{X} = (X(t_1), \dots, X(t_k))^T, \quad \mathbf{m} = (m_X(t_1), \dots, m_X(t_k))^T.$$

The matrices introduced in Lecture 3 are obtained by evaluating the two-time functions:

$$[\mathbf{R}_X]_{ij} = R_X(t_i, t_j), \quad [\mathbf{K}_X]_{ij} = C_X(t_i, t_j), \quad \mathbf{K}_X = \mathbf{R}_X - \mathbf{m}\mathbf{m}^T.$$

For any real coefficients $a_1, \dots, a_k$,

$$\begin{aligned} \sum_{i,j=1}^k a_i a_j C_X(t_i, t_j) &= \text{Var} \left(\sum_{i=1}^k a_i X(t_i) \right) \geq 0, \\ \sum_{i,j=1}^k a_i a_j R_X(t_i, t_j) &= \mathbb{E} \left[\left(\sum_{i=1}^k a_i X(t_i) \right)^2 \right] \geq 0. \end{aligned} \tag{7}$$

Thus every sampled covariance or correlation matrix is *positive semidefinite*. This requirement must hold for every finite selection of times and coefficients. It does not require every entry $C_X(s, t)$ to be nonnegative.

Uncertainty of a change. Applying Eq. (7) to the difference of two samples gives

$$\text{Var}[X(t) - X(s)] = C_X(t, t) + C_X(s, s) - 2C_X(s, t). \tag{8}$$

For fixed marginal variances, positive covariance reduces uncertainty in the difference; negative covariance increases it.

Q: Could two time samples have covariance matrix $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$?

A: No. Eq. (8) would give variance $1 + 1 - 2 \cdot 2 = -2$. Equivalently, the matrix is not positive semidefinite. Checking only its nonnegative diagonal entries is insufficient.

The mean and covariance functions summarize the first two moments; in general they do not determine all finite-dimensional distributions.

Example 4 (A sinusoid with random phase). Let

$$Y(t) = b \cos(\nu t + \Theta), \quad t \in \mathbb{R},$$

where the amplitude $b > 0$ and angular frequency $\nu > 0$ are fixed, and $\Theta \sim \text{Uniform}[0, 2\pi)$ is drawn once for the entire path. Averaging over the phase gives

$$m_Y(t) = \frac{b}{2\pi} \int_0^{2\pi} \cos(\nu t + \theta) d\theta = 0.$$

Using $\cos u \cos v = \frac{1}{2}[\cos(u - v) + \cos(u + v)]$, we obtain

$$\begin{aligned} R_Y(s, t) &= \frac{b^2}{2\pi} \int_0^{2\pi} \cos(\nu s + \theta) \cos(\nu t + \theta) d\theta \\ &= \frac{b^2}{2} \cos(\nu(t - s)) + \frac{b^2}{4\pi} \int_0^{2\pi} \cos(\nu(s + t) + 2\theta) d\theta \\ &= \frac{b^2}{2} \cos(\nu(t - s)). \end{aligned}$$

Consequently,

$$C_Y(s, t) = R_Y(s, t), \quad \text{Var}[Y(t)] = \frac{b^2}{2}, \quad \rho_Y(s, t) = \cos(\nu(t - s)).$$

Here the mean and variance are constant, and covariance depends only on the lag $t - s$. This contrasts with the increasing variance of the random slope.

The samples are uncorrelated when $\nu(t - s) = \pi/2 + k\pi$, $k \in \mathbb{Z}$. This still need not mean independence: for $h = \pi/(2\nu)$, the quarter-period samples satisfy $Y(s)^2 + Y(s + h)^2 = b^2$. They share the same random phase.