Stochastic Processes

David Nualart
nualart@mat.ub.es

1 Stochastic Processes

1.1 Probability Spaces and Random Variables

In this section we recall the basic vocabulary and results of probability theory. A *probability space* associated with a *random experiment* is a triple $(\Omega, \mathcal{F}, P)$ where:

(i) Ω is the set of all possible outcomes of the random experiment, and it is called the *sample space*.

(ii) $\mathcal{F}$ is a family of subsets of Ω which has the structure of a σ-field:

a) $\emptyset \in \mathcal{F}$

b) If $A \in \mathcal{F}$, then its complement A^c also belongs to $\mathcal{F}$

c) $A_1, A_2, \ldots \in \mathcal{F} \Longrightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

(iii) P is a function which associates a number $P(A)$ to each set $A \in \mathcal{F}$ with the following properties:

a) $0 \leq P(A) \leq 1$,

b) $P(\Omega) = 1$

c) For any sequence $A_1, A_2, \ldots$ of disjoints sets in $\mathcal{F}$ (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$),

$$\boxed{P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)}$$

The elements of the σ-field $\mathcal{F}$ are called *events* and the mapping P is called a *probability measure.* In this way we have the following interpretation of this model:

$$\boxed{P(F)=\text{“probability that the event } F \text{ occurs”}}$$

The set $\emptyset$ is called the *empty event* and it has probability zero. Indeed, the additivity property (iii,c) implies

$$P(\emptyset) + P(\emptyset) + \cdots = P(\emptyset).$$

The set Ω is also called the *certain set* and by property (iii,b) it has probability one. Usually, there will be other events $A \subset \Omega$ such that $P(A) = 1$. If a statement holds for all ω in a set A with $P(A) = 1$, then it is customary

to say that the statement is true *almost surely*, or that the statement holds for almost all $\omega \in \Omega$.

The axioms a), b) and c) lead to the following basic rules of the probability calculus:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \text{ if } A \cap B = \emptyset \\ P(A^c) &= 1 - P(A) \\ A &\subset B \Longrightarrow P(A) \leq P(B). \end{aligned}$$

Example 1 Consider the experiment of flipping a coin once.

$$\begin{aligned} \Omega &= \{H, T\} \text{ (the possible outcomes are "Heads" and "Tails")} \\ \mathcal{F} &= \mathbb{P}(\Omega) \text{ ($\mathcal{F}$ contains all subsets of Ω)} \\ P(\{H\}) &= P(\{T\}) = \frac{1}{2} \end{aligned}$$

Example 2 Consider an experiment that consists of counting the number of traffic accidents at a given intersection during a specified time interval.

$$\begin{aligned} \Omega &= \{0, 1, 2, 3, \ldots\} \\ \mathcal{F} &= \mathbb{P}(\Omega) \text{ ($\mathcal{F}$ contains all subsets of Ω)} \\ P(\{k\}) &= e^{-\lambda} \frac{\lambda^k}{k!} \text{ (Poisson probability with parameter } \lambda > 0) \end{aligned}$$

Given an arbitrary family $\mathcal{U}$ of subsets of Ω, the smallest σ-field containing $\mathcal{U}$ is, by definition,

$$\sigma(\mathcal{U}) = \cap \{\mathcal{G}, \mathcal{G} \text{ is a } \sigma\text{-field}, \mathcal{U} \subset \mathcal{G}\}.$$

The σ-field $\sigma(\mathcal{U})$ is called the σ-field generated by $\mathcal{U}$. For instance, the σ-field generated by the open subsets (or rectangles) of $\mathbb{R}^n$ is called the Borel σ-field of $\mathbb{R}^n$ and it will be denoted by $\mathcal{B}_{\mathbb{R}^n}$.

Example 3 Consider a finite partition $\mathcal{P} = \{A_1, \ldots, A_n\}$ of Ω. The σ-field generated by $\mathcal{P}$ is formed by the unions $A_{i_1} \cup \cdots \cup A_{i_k}$ where $\{i_1, \ldots, i_k\}$ is an arbitrary subset of $\{1, \ldots, n\}$. Thus, the σ-field $\sigma(\mathcal{P})$ has 2^n elements.

Example 4 We pick a real number at random in the interval $[0, 2]$. $\Omega = [0, 2]$, $\mathcal{F}$ is the Borel σ-field of $[0, 2]$. The probability of an interval $[a, b] \subset [0, 2]$ is

$$P([a, b]) = \frac{b - a}{2}.$$

Example 5 Let an experiment consist of measuring the lifetime of an electric bulb. The sample space Ω is the set $[0, \infty)$ of nonnegative real numbers. $\mathcal{F}$ is the Borel σ-field of $[0, \infty)$. The probability that the lifetime is larger than a fixed value $t \geq 0$ is

$$P([t, \infty)) = e^{-\lambda t}.$$

A *random variable* is a mapping

$$\begin{array}{c} \Omega \xrightarrow{X} \mathbb{R} \\ \omega \to X(\omega) \end{array}$$

which is $\mathcal{F}$-measurable, that is, $X^{-1}(B) \in \mathcal{F}$, for any Borel set B in $\mathbb{R}$. The random variable X assigns a value $X(\omega)$ to each outcome ω in Ω. The measurability condition means that given two real numbers $a \leq b$, the set of all outcomes ω for which $a \leq X(\omega) \leq b$ is an event. We will denote this event by $\{a \leq X \leq b\}$ for short, instead of $\{\omega \in \Omega : a \leq X(\omega) \leq b\}$.

- A random variable defines a σ-field $\{X^{-1}(B), B \in \mathcal{B}_{\mathbb{R}}\} \subset \mathcal{F}$ called the σ-field generated by X.
- A random variable defines a probability measure on the Borel σ-field $\mathcal{B}_{\mathbb{R}}$ by $P_X = P \circ X^{-1}$, that is,

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega : X(\omega) \in B\}).$$

  The probability measure P_X is called the *law* or the *distribution* of X.

We will say that a random variable X has a *probability density* f_X if $f_X(x)$ is a nonnegative function on $\mathbb{R}$, measurable with respect to the Borel σ-field and such that

$$P\{a < X < b\} = \int_a^b f_X(x)dx,$$

for all $a < b$. Notice that $\int_{-\infty}^{+\infty} f_X(x)dx = 1$. Random variables admitting a probability density are called *absolutely continuous*.

We say that a random variable X is *discrete* if it takes a finite or countable number of different values x_k. Discrete random variables do not have densities and their law is characterized by the *probability function*:

$$p_k = P(X = x_k).$$

Example 6 In the experiment of flipping a coin once, the random variable given by

$$X(H) = 1, X(T) = -1$$

represents the earning of a player who receives or loses an euro according as the outcome is heads or tails. This random variable is discrete with

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

Example 7 If A is an event in a probability space, the random variable

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{if} \quad \omega \notin A \end{cases}$$

is called the indicator function of A. Its probability law is called the *Bernoulli* distribution with parameter $p = P(A)$.

Example 8 We say that a random variable X has the *normal law* $N(m, \sigma^2)$ if

$$P\left(a < X < b\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

for all $a < b$.

Example 9 We say that a random variable X has the *binomial law* $B(n, p)$ if

$$P\left(X = k\right) = \binom{n}{k} p^k (1-p)^{n-k},$$

for $k = 0, 1, \ldots, n$.

The function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$\boxed{F_X(x) = P\left(X \leq x\right) = P_X\left((-\infty, x]\right)}$$

is called the *distribution function* of the random variable X.

- The distribution function F_X is non-decreasing, right continuous and with

$$\begin{aligned} \lim_{x\to-\infty} F_X(x) &= 0, \\ \lim_{x\to+\infty} F_X(x) &= 1. \end{aligned}$$

- If the random variable X is absolutely continuous with density f_X, then,

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy,$$

and if, in addition, the density is continuous, then $F'_X(x) = f_X(x)$.

The *mathematical expectation (*or *expected value)* of a random variable X is defined as the integral of X with respect to the probability measure P:

$$E(X) = \int_\Omega X dP.$$

In particular, if X is a discrete variable that takes the values $\alpha_1, \alpha_2, \ldots$ on the sets $A_1, A_2, \ldots$, then its expectation will be

$$E(X) = \alpha_1 P(A_1) + \alpha_1 P(A_1) + \cdots .$$

Notice that $E(\mathbf{1}_A) = P(A)$, so the notion of expectation is an extension of the notion of probability.

If X is a non-negative random variable it is possible to find discrete random variables X_n, $n = 1, 2, \ldots$ such that

$$X_1(\omega) \leq X_2(\omega) \leq \cdots$$

and

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

for all ω. Then $E(X) = \lim_{n\to\infty} E(X_n) \leq +\infty$, and this limit exists because the sequence $E(X_n)$ is non-decreasing. If X is an arbitrary random variable, its expectation is defined by

$$E(X) = E(X^+) - E(X^-),$$

where $X^+ = \max(X, 0)$, $X^- = -\min(X, 0)$, provided that both $E(X^+)$ and $E(X^-)$ are finite. Note that this is equivalent to say that $E(|X|) < \infty$, and in this case we will say that X is integrable.

A simple computational formula for the expectation of a non-negative random variable is as follows:

$$E(X) = \int_0^\infty P(X > t)dt.$$

In fact,

$$\begin{aligned} E(X) &= \int_\Omega X dP = \int_\Omega \left(\int_0^\infty \mathbf{1}_{\{X>t\}} dt \right) dP \\ &= \int_0^{+\infty} P(X > t)dt. \end{aligned}$$

The expectation of a random variable X can be computed by integrating the function x with respect to the probability law of X:

$$E(X) = \int_\Omega X(\omega)dP(\omega) = \int_{-\infty}^\infty x dP_X(x).$$

More generally, if $g : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function and $E(|g(X)|) < \infty$, then the expectation of $g(X)$ can be computed by integrating the function g with respect to the probability law of X:

$$E(g(X)) = \int_\Omega g(X(\omega))dP(\omega) = \int_{-\infty}^\infty g(x)dP_X(x).$$

The integral $\int_{-\infty}^\infty g(x)dP_X(x)$ can be expressed in terms of the probability density or the probability function of X:

$$\int_{-\infty}^\infty g(x)dP_X(x) = \begin{cases} \int_{-\infty}^\infty g(x)f_X(x)dx, & f_X(x) \text{ is the density of } X \\ \sum_k g(x_k)P(X = x_k), & X \text{ is discrete} \end{cases}$$

Example 10 If X is a random variable with normal law $N(0,\sigma^2)$ and λ is a real number,

$$\begin{aligned} E(\exp(\lambda X)) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{\lambda x} e^{-\frac{x^2}{2\sigma^2}}\,dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{\sigma^2\lambda^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\sigma^2\lambda)^2}{2\sigma^2}}\,dx \\ &= e^{\frac{\sigma^2\lambda^2}{2}}. \end{aligned}$$

Example 11 If X is a random variable with Poisson distribution of parameter $\lambda > 0$, then

$$E(X) = \sum_{n=0}^{\infty} n\frac{e^{-\lambda}\lambda^n}{n!} = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{e^{-\lambda}\lambda^{n-1}}{(n-1)!} = \lambda.$$

The *variance* of a random variable X is defined by

$$\sigma_X^2 = \mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - [E(X)]^2,$$

provided $E(X^2) < \infty$. The variance of X measures the deviation of X from its expected value. For instance, if X is a random variable with normal law $N(m,\sigma^2)$ we have

$$\begin{aligned} P(m - 1.96\sigma &\leq X \leq m + 1.96\sigma) = P(-1.96 \leq \frac{X-m}{\sigma} \leq 1.96) \\ &= \Phi(1.96) - \Phi(-1.96) = 0.95, \end{aligned}$$

where Φ is the distribution function of the law $N(0,1)$. That is, the probability that the random variable X takes values in the interval $[m-1.96\sigma, m+1.96\sigma]$ is equal to 0.95.

If X and Y are two random variables with $E(X^2) < \infty$ and $E(Y^2) < \infty$, then its covariance is defined by

$$\begin{aligned} \mathrm{cov}(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

A random variable X is said to have a finite moment of order $p \geq 1$, provided $E(|X|^p) < \infty$. In this case, the pth moment of X is defined by

$$\boxed{m_p = E(X^p).}$$

The set of random variables with finite pth moment is denoted by $L^p(\Omega, \mathcal{F}, P)$.

The *characteristic function* of a random variable X is defined by

$$\varphi_X(t) = E(e^{itX}).$$

The moments of a random variable can be computed from the derivatives of the characteristic function at the origin:

$$m_n = \frac{1}{i^n}\varphi_X^{(n)}(t)|_{t=0},$$

for $n = 1, 2, 3, \ldots$.

We say that $X = (X_1, \ldots, X_n)$ is an n-dimensional *random vector* if its components are random variables. This is equivalent to say that X is a random variable with values in $\mathbb{R}^n$.

The mathematical expectation of an n-dimensional random vector X is, by definition, the vector

$$\boxed{E(X) = (E(X_1), \ldots, E(X_n))}$$

The *covariance matrix* of an n-dimensional random vector X is, by definition, the matrix $\Gamma_X = (\mathrm{cov}(X_i, X_j))_{1 \le i,j \le n}$. This matrix is clearly symmetric. Moreover, it is non-negative definite, that means,

$$\sum_{i,j=1}^{n} \Gamma_X(i,j) a_i a_j \ge 0$$

for all real numbers $a_1, \ldots, a_n$. Indeed,

$$\sum_{i,j=1}^{n} \Gamma_X(i,j) a_i a_j = \sum_{i,j=1}^{n} a_i a_j \mathrm{cov}(X_i, X_j) = \mathrm{Var}(\sum_{i=1}^{n} a_i X_i) \ge 0$$

As in the case of real-valued random variables we introduce the law or distribution of an n-dimensional random vector X as the probability measure defined on the Borel σ-field of $\mathbb{R}^n$ by

$$P_X(B) = P(X^{-1}(B)) = P(X \in B).$$

We will say that a random vector X has a *probability density* f_X if $f_X(x)$ is a nonnegative function on $\mathbb{R}^n$, measurable with respect to the Borel σ-field and such that

$$P\left\{a_i < X_i < b_i, i = 1, \ldots, n\right\} = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_X(x_1, \ldots, x_n) dx_1 \cdots dx_n,$$

for all $a_i < b_i$. Notice that

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_X(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1.$$

We say that an n-dimensional random vector X has a multidimensional *normal law* $N(m, \Gamma)$, where $m \in \mathbb{R}^n$, and Γ is a symmetric positive definite matrix, if X has the density function

$$f_X(x_1, \ldots, x_n) = (2\pi \det \Gamma)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i,j=1}^n (x_i - m_i)(x_j - m_j)\Gamma_{ij}^{-1}}.$$

In that case, we have, $m = E(X)$ and $\Gamma = \Gamma_X$.

If the matrix Γ is diagonal

$$\Gamma = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

then the density of X is the product of n one-dimensional normal densities:

$$f_X(x_1, \ldots, x_n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-m_i)^2}{2\sigma_i^2}} \right).$$

There exists degenerate normal distributions which have a singular covariance matrix Γ. These distributions do not have densities and the law of a random variable X with a (possibly degenerated) normal law $N(m, \Gamma)$ is determined by its characteristic function:

$$E\left(e^{it'X}\right) = \exp\left(it'm - \frac{1}{2}t'\Gamma t\right),$$

where $t \in \mathbb{R}^n$. In this formula t' denotes a row vector ($1 \times n$ matrix) and t denoted a column vector ($n \times 1$ matrix).