Applied Stochastic Processes

Chapter 1 - Preliminaries

Peng Jin

August 30, 2024

Overview

In this chapter we collect some basic knowledge of probability theory and then give a brief introduction to stochastic processes.

Events and their Probabilities

Sample space and events

- The set of all possible outcomes of a random experiment is known as the sample space of the experiment and is denoted by Ω.
- An “event” is a property which can be observed either to hold or not to hold *after* the experiment is done.

Example 1.1

- If the experiment consists of tossing two coins, then the sample space consists of the following four outcomes:

$$\Omega = \{HH, HT, TH, TT\}.$$

- In the example of tossing two coins, if

$$E = \{HH, HT\},$$

then E is the event that a head appears on the first coin.

σ-Fields

- One collects "good" subsets of Ω, the events, in a class $\mathcal{F}$, say.
- In probability theory we require $\mathcal{F}$ to be a σ-field (also called σ-algebra). Such a class is supposed to contain all interesting events and thus have at least the following properties:
  - $\emptyset \in \mathcal{F}$;
  - If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, where A^c denotes the complement of A;
  - $\mathcal{A}$ is closed under countable unions: that is, if $A_1, A_2, A_3, \ldots$ is a countable sequence of events in $\mathcal{F}$, then $\cup_{i=1}^{\infty} A_i$ is also in $\mathcal{F}$.
- Rougly speaking, we would like that elementary operations such as $\cap$, $\cup$ and complement on the events of $\mathcal{F}$ should not lead outside the class $\mathcal{F}$. This is the intuitive meaning of a σ-field $\mathcal{F}$.

σ-Fields cont'd

Proposition. Suppose $\mathcal{F}$ is a σ-field, $A_1, A_2, \ldots$ are in $\mathcal{F}$, and $m \in \mathbb{N}$. Then each of the sets

$$\Omega,\ A_1 \backslash A_2,\ \bigcup_{j=1}^{m} A_j,\ \bigcap_{j=1}^{m} A_j,\ \bigcap_{j \in \mathbb{N}} A_j$$

also belongs to $\mathcal{F}$.

Probability measure

- To each event $A \in \mathcal{F}$ we assign a number $\mathbb{P}(A) \in [0, 1]$. This number is the expected fraction of occurrences of the event A in a long series of experiments where A or A^c are observed.
- A probability measure *defined on a σ-algebra $\mathcal{F}$ of Ω is a function* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *that satisfies*:
  - $\mathbb{P}(\Omega) = 1$;
  - $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ *whenever the A_i are in $\mathcal{F}$ and are pairwise disjoint (i.e. $A_n \cap A_m = \emptyset$ if $n \neq m$).*
- We call $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Some elementary properties of probability measures

For $A, B \in \mathcal{F}$ and $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$, we have:

1. (finite additivity) $\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ if $A_1, \ldots, A_n$ are pairwise disjoint.
2. $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$.
3. $\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$ if $A \subset B$.
4. $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.
5. $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, that is, $\mathbb{P}$ is monotone.
6. $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ if $A_n \uparrow A$. Here $A_n \uparrow A$ means that $A_1 \subset A_2 \subset \ldots$ and $\cup_{n=1} A_n = A$.
7. $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$ if $A_n \downarrow A$. Here $A_n \downarrow A$ means that $A_1 \supset A_2 \supset \ldots$ and $\cap_{n=1} A_n = A$.

Conditional probability and product rule

- Conditional probability:

$$\mathbb{P}(B \mid A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

- All basic formulas of probability remain true, conditionally, e.g.:

$$\mathbb{P}(B^c \mid A) = 1 - \mathbb{P}(B \mid A),$$

$$\mathbb{P}(B \cup C \mid A) = \mathbb{P}(B \mid A) + \mathbb{P}(C \mid A) - \mathbb{P}(B \cap C \mid A).$$

Product rule:

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \\
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B) \\
\mathbb{P}(A \cap B \cap C \cap D) &= \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B) \cdot \mathbb{P}(D \mid A \cap B \cap C) \\
&\vdots
\end{aligned}$$

Law of total probability

- A partition represents chopping the sample space into several smaller events, say $A_1, A_2, A_3, \ldots, A_k, \ldots$, so that they

  1. don't overlap (i.e. are all mutually exclusive): $A_i \cap A_j = \emptyset$ for any $i \neq j$
  2. cover the whole Ω (i.e. 'no gaps'): $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_k \cup \ldots = \Omega$.

Law of total probability: For any partition, and any event B, we have

$$\mathbb{P}(B) = \mathbb{P}(B \mid A_1) \cdot \mathbb{P}(A_1) + \mathbb{P}(B \mid A_2) \cdot \mathbb{P}(A_2) + \ldots + \mathbb{P}(B \mid A_k) \cdot \mathbb{P}(A_k) + \ldots.$$

- In this theorem, the number of events in the partition can be finite or countable.

Random Variables

Example: random variables

- A fair coin is tossed twice: $\Omega = \{\text{HH, HT, TH, TT}\}$. For $\omega \in \Omega$, let $X(\omega)$ be the number of heads, so that

$$X(\mathrm{HH}) = 2,\ X(\mathrm{HT}) = X(\mathrm{TH}) = 1,\ X(\mathrm{TT}) = 0.$$

- Now suppose that a gambler wagers his fortune of $\$1$ on the result of this experiment. He gambles cumulatively so that his fortune is doubled each time a head appears, and is annihilated on the appearance of a tail. His subsequent fortune W is a random variable given by

$$W(\mathrm{HH}) = 4,\ W(\mathrm{HT}) = W(\mathrm{TH}) = W(\mathrm{TT}) = 0.$$

Random variables and their distribution functions

- Rigorously, a **random variable** is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. Such a function is said to be $\mathcal{F}$**-measurable**.

- The function

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega : X(\omega) \leq x\}), \quad x \in \mathbb{R} = (-\infty, \infty),$$

is the **distribution function** F_X of X. It yields e.g. the probability that X belongs to the interval $(a, b]$. Indeed,

$$\mathbb{P}(\{\omega : a < X(\omega) \leq b\}) = F_X(b) - F_X(a), \quad a < b.$$

More on the distribution function

- We also obtain the probability that X is equal to a number:

$$\begin{aligned}
&\mathbb{P}(X = x) \\
&\quad = \mathbb{P}(\{\omega : X(\omega) = x\}) = \mathbb{P}(\{\omega : X \leq x\}) - \mathbb{P}(\{\omega : X < x\}) \\
&\quad = \mathbb{P}(\{\omega : X(\omega) \leq x\}) - \lim_{h \downarrow 0} \mathbb{P}(\{\omega : X(\omega) \leq x - h\}) \\
&\quad = F_X(x) - \lim_{h \downarrow 0} F_X(x - h)
\end{aligned}$$

- With these probabilities one can approximate the probability of the event $\{\omega : X(\omega) \in B\}$ for very complicated subsets B of $\mathbb{R}$.

Distribution of a random variable

- The collection of the probabilities

$$P_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega : X(\omega) \in B\})$$

for suitable subsets $B \subset \mathbb{R}$ is the distribution of X. "Suitable" subsets of $\mathbb{R}$ are the so-called Borel sets. They are obtained by a countable number of operations $\cap$, $\cup$ or ^{c} acting on the intervals; we will give a precise definition later in this chapter.

- To be more precise, P_X is a probability measure on the σ-field of all Borel sets.
- The distribution P_X and the distribution function F_X are equivalent notions in the sense that both of them can be used to calculate the probability of any event $\{X \in B\}$.

Discrete random variables

- A distribution function is continuous or it has jumps.
- We first consider the special case when the distribution function F_X is a pure jump function:

$$F_X(x) = \sum_{k:x_k \le x} p_k, \quad x \in \mathbb{R} \tag{1}$$

  where $0 \le p_k \le 1$ for all k and $\sum_{k=1}^{\infty} p_k = 1$.
- The distribution function (1) and the corresponding distribution are said to be discrete; a random variable with distribution function (1) is a discrete random variable.
- A discrete random variable only assumes a finite or countably infinite number of values $x_1, x_2, \ldots$, and $p_k = \mathbb{P}\left(X = x_k\right)$. In particular, the distribution function F_X has an upward jump of size p_k at $x = x_k$.

Continuous random variables

- In contrast to discrete distributions and random variables, the distribution function of a continuous random variable does not have jumps, hence $\mathbb{P}(X = x) = 0$ for all x, or equivalently,

$$\lim_{h \to 0} F_X(x + h) = F_X(x) \text{ for all } x. \tag{2}$$

- More precisely, we call X a *continuous random variable* if it has a density f_X :

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

where $f_X(x) \geq 0$ for every $x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f_X(y)dy = 1$.

Common distributions

(a) *Bernoulli*. A random variable is Bernoulli if $\mathbb{P}(X = 1) = p, \mathbb{P}(X = 0) = 1 - p$ for some $p \in [0, 1]$.

(b) *Binomial*. This is defined by $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, where n is a positive integer, $0 \leq k \leq n$, and $p \in [0, 1]$.

(c) *Geometric*. For $p \in (0, 1)$ we set $\mathbb{P}(X = k) = (1 - p)p^k$. Here k is a nonnegative integer.

(d) *Poisson*. For $\lambda > 0$ we set $\mathbb{P}(X = k) = e^{-\lambda}\lambda^k/k$! Again k is a nonnegative integer.

(e) *Uniform*. For some positive integer n, set $\mathbb{P}(X = k) = 1/n$ for $1 \leq k \leq n$.

(f) *Uniform on* $[a, b]$. Define $f(x) = (b - a)^{-1}\mathbf{1}_{[a,b]}(x)$. This means that if X has a uniform distribution, then

$$\mathbb{P}(X \in A) = \int_A \frac{1}{b - a}\mathbf{1}_{[a,b]}(x)dx.$$

Common distributions cont'd

(g) *Exponential*. For $x > 0$ let $f(x) = \lambda e^{-\lambda x}$ and otherwise $f(x) = 0$.

(h) *Standard normal*. Define $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. So

$$\mathbb{P}(X \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx.$$

(i) $N\left(\mu, \sigma^2\right)$. We shall see later that a standard normal has mean zero and variance one. If Z is a standard normal, then a $N\left(\mu, \sigma^2\right)$ random variable has the same distribution as $\mu + \sigma Z$. It is an exercise in calculus to check that such a random variable has density

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}. \tag{3}$$

(j) *Cauchy*. Here

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Expectation, variance and moments for continuous RVs

- The expectation or mean value of a random variable X with density f_X is given by

$$\mu_X = \mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x) dx,$$

  provided that $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$.
- The variance of X is defined as

$$\sigma_X^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx.$$

- The l th moment of X for $l \in \mathbb{N}$ is defined as

$$\mathbb{E}\left(X^l\right) = \int_{-\infty}^{\infty} x^l f_X(x) dx$$

- For a real-valued function g the expectation of $g(X)$ is given by $\mathbb{E}g(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$.

Expectation, variance and moments for discrete RVs

- The expectation or mean value of a discrete random variable X with probabilities $p_k = P(X = x_k)$ is given by

$$\mu_X = \mathbb{E}X = \sum_{k=1}^{\infty} x_k p_k,$$

  provided that $\sum_{k=1}^{\infty} |x_k| p_k < \infty$.
- The variance of X is defined as

$$\sigma_X^2 = \mathrm{Var}(X) = \sum_{k=1}^{\infty} (x_k - \mu_X)^2 p_k$$

- The l th moment of X for $l \in \mathbb{N}$ is defined as

$$\mathbb{E}\left(X^l\right) = \sum_{k=1}^{\infty} x_k^l p_k$$

- For a real-valued function g the expectation of $g(X)$ is given by $\mathbb{E}g(X) = \sum_{k=1}^{\infty} g(x_k) p_k$.

Interpretation of the variance

The spread or dispersion of the random values $X(\omega)$ around the expectation μ_X is described by the variance

$$\begin{aligned}\sigma_X^2 &= \mathrm{Var}(X) = \mathbb{E}\left(X - \mu_X\right)^2 \\ &= \mathbb{E}\left(X^2 - 2\mu_X X + \mu_X^2\right) = \mathbb{E}\left(X^2\right) - 2\mu_X^2 + \mu_X^2 \\ &= \mathbb{E}\left(X^2\right) - \mu_X^2\end{aligned}$$

and the standard deviation σ_X.

- Recall the normal density from (3). The parameter μ is the expectation μ_X and the parameter σ^2 is the variance σ_X^2 of a random variable X with density (3).

Random Vectors

From random vectors to stochastic processes

- In what follows we frequently make use of finite-dimensional and infinite-dimensional random structures.
- We commence with finite-dimensional random vectors as a first step toward the definition of a stochastic process.

Random vectors

- $\mathbf{X} = (X_1, \ldots, X_n)^\top$ is an n-dimensional random vector if its components $X_1, \ldots, X_n$ are one-dimensional real-valued random variables. Here and in what follows, $\mathbf{y}^\top$ denotes the transpose of $\mathbf{y}$, where $\mathbf{y}$ can be a matrix or a vector.
- If we interpret $t = 1, \ldots, n$ as equidistant instants of time, X_t can stand for the outcome of an experiment at time t. Such a time series may, for example, consist of BMW share prices X_t at n succeeding days.

Distribution and dependence structure for random vectors

- Analogously to one-dimensional random variables one can introduce the distribution function, the expectation, moments and the covariance matrix of a random vector in order to describe
  - its distribution
  - and its dependence structure.
- The latter aspect is a new one; dependence does not make sense when we talk just about one random variable.

Example 1.2

Toss a fair coin twice. We consider the four pairs $(H, H), (T, T), (H, T)$ and (T, H) (H=head, T = tail) as outcomes of the experiment "flipping a coin twice".

- These four pairs constitute the outcome space Ω.
- We assign 1 to H and 0 to T. In this way we obtain two random variables X_1 and X_2, and $\mathbf{X} = (X_1, X_2)^\top$ is a two-dimensional random vector.
- Notice that

$$\mathbf{X}(H, H) = (1, 1)^\top, \mathbf{X}(T, T) = (0, 0)^\top, \mathbf{X}(T, H) = (0, 1)^\top, \mathbf{X}(H, T) = (1, 0)^\top.$$

If the coin is indeed fair, we can assign the probability 0.25 to each of the four outcomes, i.e.

$$\mathbb{P}(\{\omega : \mathbf{X}(\omega) = (k, l)^\top\}) = 0.25, \quad k, l \in \{0, 1\}.$$

Distribution function of a random vector

As before, we consider a collection $\mathcal{F}$ of subsets of Ω and define a probability measure on it, i.e. we assign a number $\mathbb{P}(A) \in [0,1]$ to each $A \in \mathcal{F}$.

- The collection of the probabilities

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) =& \mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) \\ =& \mathbb{P}\left(\{\omega : X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n\}\right), \\ & \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \end{aligned} \tag{4}$$

  is the distribution function $F_{\mathbf{X}}$ of $\mathbf{X}$.

- It provides us with the probability of the event that $\mathbf{X}$ assumes values in the rectangle

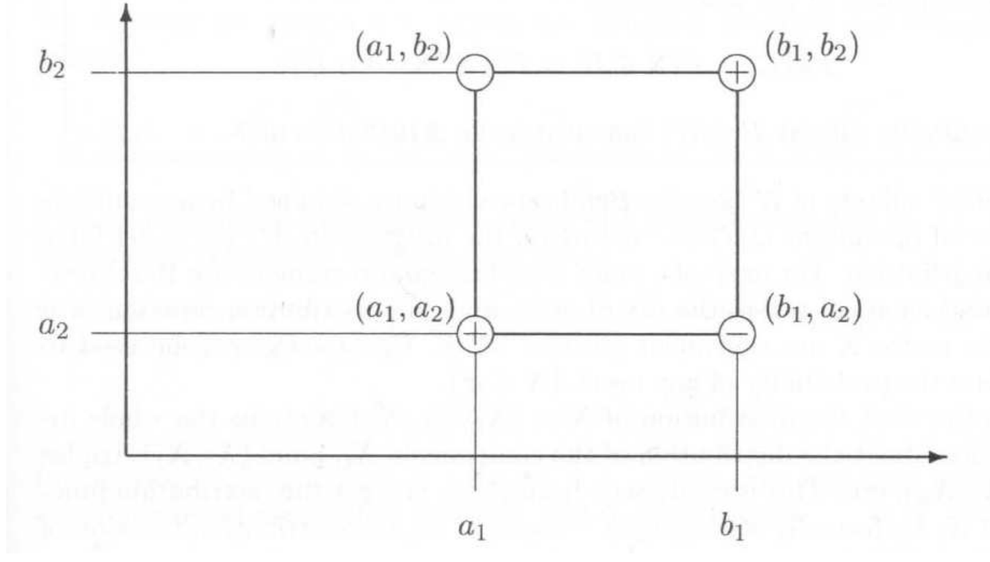
$$(\mathbf{a}, \mathbf{b}] = \{\mathbf{x} : a_i < x_i \leq b_i, \quad i = 1, \ldots, n\}.$$

Probability of falling into a rectangle

For example, if $\mathbf{X}$ is two-dimensional,

$$\mathbb{P}(\mathbf{X} \in (\mathbf{a}, \mathbf{b}]) = F_{\mathbf{X}}(b_1, b_2) + F_{\mathbf{X}}(a_1, a_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2),$$

which can be seen from the below Figure.



- As in the case of one-dimensional random variables, these probabilities approximate $\mathbb{P}(\mathbf{X} \in B)$ for very general sets B.

Distribution of a random vector

The collection of the probabilities

$$P_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B) = \mathbb{P}(\{\omega : \mathbf{X}(\omega) \in B\})$$

for suitable subsets $B \subset \mathbb{R}^n$ constitutes the distribution of $\mathbf{X}$.

- "Suitable" subsets of $\mathbb{R}^n$ are the Borel sets which are obtained by a countable number of operations $\cup$, $\cap$ or ^{c} acting on the intervals in $\mathbb{R}^n$; see below for a precise definition.
- For example, point sets, balls and rectangles are Borel sets. In a mathematical sense, the distribution and the distribution function of a random vector $\mathbf{X}$ are equivalent notions. Both, $F_{\mathbf{X}}$ and $P_{\mathbf{X}}$, can be used to calculate the probability of any event $\{\mathbf{X} \in B\}$.

Remark on the distribution of random vectors

- Notice that the distribution of $\mathbf{X} = (X_1, \ldots, X_n)^\top$ contains the whole information about the distribution of the components X_i, pairs $(X_i, X_j)^\top$, triples $(X_i, X_j, X_k)^\top$, etc.
- This is easily seen from (4): you get the distribution function of X_1 by formally setting $x_2 = \cdots = x_n = \infty$, the distribution function of $(X_1, X_2)^\top$ by setting $x_3 = \cdots = x_n = \infty$, etc.

Analogously to random variables one can introduce discrete and continuous random vectors and distributions. For our purposes, continuous random vectors with a density will be relevant, and so we mostly restrict our attention to them.

Continuous random vectors

If the distribution of a random vector $\mathbf{X}$ has density $f_{\mathbf{X}}$, one can represent the distribution function $F_{\mathbf{X}}$ of $\mathbf{X}$ as

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, \ldots, y_n)\, dy_1 \cdots dy_n$$
$$(x_1, \ldots, x_n) \in \mathbb{R}^n$$

where the density is a function satisfying

$$f_{\mathbf{X}}(\mathbf{x}) \geq 0 \quad \text{for every } \mathbf{x} \in \mathbb{R}^n$$

and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(y_1, \ldots, y_n)\, dy_1 \cdots dy_n = 1.$$

Martingal densities

- If a vector $\mathbf{X}$ has density $f_{\mathbf{X}}$, all its components X_i, the vectors of the pairs $(X_i, X_j)^\top$, triples $(X_i, X_j, X_k)^\top$, etc., have a density. They are called marginal densities.

Example: We consider the case $n = 3$. Then the marginal densities are obtained as follows:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})dx_2dx_3, \quad f_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})dx_3$$

$f_{X_2}(x_2)$ is obtained by integrating $f_{\mathbf{X}}(\mathbf{x})$ with respect to x_1 and x_3, f_{X_1,X_3} by integrating $f_{\mathbf{X}}(\mathbf{x})$ with respect to x_2, etc.

Gaussian random vector

A *Gaussian* or *normal* random vector has a Gaussian or normal distribution. The n-dimensional normal or Gaussian distribution is given by its density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\top}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x} \in \mathbb{R}^n \tag{5}$$

with parameters $\mu \in \mathbb{R}^n$ and Σ.

- The quantity Σ is a symmetric positivedefinite $n \times n$ matrix, Σ^{-1} is its inverse and $\det \Sigma$ its determinant.

Expectation and covariance matrix of random vectors

- The expectation of a random vector has a similar function as the mean value of a random variable. The values $\mathbf{X}(\omega)$ are concentrated around it.
- The expectation or mean value of a random vector $\mathbf{X}$ is given by

$$\mu_{\mathbf{X}} = \mathbb{E}\mathbf{X} = (\mathbb{E}X_1, \ldots, \mathbb{E}X_n)^\top$$

- The covariance matrix of $\mathbf{X}$ is defined as

$$\Sigma_{\mathbf{X}} = \left(\operatorname{cov}(X_i, X_j); i, j = 1, \ldots, n\right),$$

where

$$\begin{aligned}\operatorname{Cov}(X_i, X_j) &= \mathbb{E}\left[(X_i - \mu_{X_i})\left(X_j - \mu_{X_j}\right)\right] \\ &= \mathbb{E}(X_i X_j) - \mu_{X_i}\mu_{X_j}\end{aligned}$$

is the covariance of X_i and X_j. Notice that $\operatorname{Cov}(X_i, X_i) = \sigma^2_{X_i}$.

More on Gaussian random vectors

Recall from (5) the density of a multivariate Gaussian random vector $\mathbf{X}$. The parameter μ is the expectation $\mu_{\mathbf{X}}$ of $\mathbf{X}$, and Σ is its covariance matrix $\Sigma_{\mathbf{X}}$.

- Thus the density of a Gaussian vector (hence its distribution) is completely determined via its expectation and covariance matrix.
- In particular, if $\mu = \mathbf{0}$ and Σ is the n-dimensional identity matrix I_n, we have $\det I_n = 1$ and $\Sigma^{-1} = I_n$. The density f_{X} is then simply the product of n standard normal densities:

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \varphi(x_1) \cdots \varphi(x_n).$$

Linear transform of Gaussion random vectors

- We write $N(\mu, \Sigma)$ for the distribution of an n-dimensional Gaussian vector $\mathbf{X}$ with expectation μ and covariance matrix Σ.
- Such a vector has the appealing property that it remains Gaussian under linear transformations (recall that $\mu^\top$ and $A^\top$ denote the transposes of μ and A, respectively).

Theorem: Let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ have an $N(\mu, \Sigma)$ distribution and A be an $m \times n$ matrix. Then $A\mathbf{X}$ has an $N\left(A\mu, A\Sigma A^\top\right)$ distribution.

Correlation coefficient

- It is convenient to standardize covariances by dividing the corresponding random variables by their standard deviations.
- The resulting quantity

$$\begin{aligned}\operatorname{corr}(X_1, X_2) &= \frac{\operatorname{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \\ &= \frac{\mathbb{E}\left[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})\right]}{\sigma_{X_1}\sigma_{X_2}}\end{aligned}$$

  is the correlation coefficient of X_1 and X_2.
- As a result of this standardization, the correlation coefficient of two random variables is always between -1 and $+1$.

Independence and Dependence

Independence

Flip a fair coin twice and let the random numbers $X_1(\omega), X_2(\omega) \in \{0, 1\}$ be the corresponding outcomes of the first and second experiment, where $X_i(\omega) = 1$ means getting a head. It is easy to verify that

$$\mathbb{P}(X_1 = k, X_2 = l) = \mathbb{P}(X_1 = k)\,\mathbb{P}(X_2 = l), \quad k, l \in \{0, 1\}.$$

This property is called **independence** of the random variables X_1 and X_2.

- Intuitively, independence means that the first experiment does not influence the second one, and vice versa. For example, knowledge of X_1 does not allow one to predict the value of X_2, and vice versa.

Definition of independence

Below we recall some of the essential definitions and properties of independent events and independent random variables.

- Two events A_1 and A_2 are independent if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,.$$

- Two random variables X_1 and X_2 are independent if

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in B_1)\,\mathbb{P}(X_2 \in B_2) \tag{6}$$

for all suitable subsets B_1 and B_2 of $\mathbb{R}$.

Equivalent definition of independent RVs

Alternatively, one can define independence via distribution functions and densities.

- The random variables X_1 and X_2 are independent if and only if

$$F_{X_1,X_2}(x_1,x_2) = F_{X_1}(x_1)\,F_{X_2}(x_2)\,, \quad x_1,x_2 \in \mathbb{R}$$

- Assume that (X_1, X_2) has density f_{X_1,X_2} with marginal densities f_{X_1} and f_{X_2}. Then the random variables X_1 and X_2 are independent if and only if

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)\,f_{X_2}(x_2)\,, \quad x_1,x_2 \in \mathbb{R}.$$

Independence for multiple events or RVs

- The definition of independence can be extended to an arbitrary finite number of events and random vectors. Notice that independence of the components of a random vector implies the independence of each pair of its components, but the converse is in general not true.
- The events $A_1, \ldots, A_n$ are independent if, for every choice of indices $1 \leq i_1 < \cdots < i_k \leq n$ and integers $1 \leq k \leq n$

$$\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) = \mathbb{P}\left(A_{i_1}\right) \cdots \mathbb{P}\left(A_{i_k}\right).$$

- The random variables $X_1, \ldots, X_n$ are independent if, for every choice of indices $1 \leq i_1 < \cdots < i_k \leq n$, integers $1 \leq k \leq n$ and all suitable subsets $B_1, \ldots, B_n$ of $\mathbb{R}$,

$$\mathbb{P}\left(X_{i_1} \in B_{i_1}, \ldots, X_{i_k} \in B_{i_k}\right) = \mathbb{P}\left(X_{i_1} \in B_{i_1}\right) \cdots \mathbb{P}\left(X_{i_k} \in B_{i_k}\right)$$

  This means that the events $\left\{X_1 \in B_1\right\}, \ldots, \left\{X_n \in B_n\right\}$ are independent.

Random vector with independent components

- The random variables $X_1, \ldots, X_n$ are independent if and only if their joint distribution function can be written as follows:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n$$

- If the random vector $\mathbf{X} = (X_1, \ldots, X_n)^\top$ has density $f_{\mathbf{X}}$, then $X_1, \ldots, X_n$ are independent if and only if

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{7}$$

Independent components of a Gaussian random vector

- Recall from (5) the density of an n-dimensional Gaussian vector $\mathbf{X}$.
- It is easily seen from the form of the density that its components are independent if and only if the covariance matrix Σ is diagonal.
- This means that $\operatorname{corr}(X_i, X_j) = \operatorname{Cov}(X_i, X_j) = 0$ for $i \neq j$, and so we can write the density of $\mathbf{X}$ in the form (7).
- Thus, in the Gaussian case, uncorrelatedness and independence are equivalent notions. This statement is wrong for non-Gaussian random vectors.

Properties of independent RVs

An important consequence of the independence of random variables is the following property: If $X_1, \ldots, X_n$ are independent, then for any real-valued functions $g_1, \ldots, g_n$,

$$\mathbb{E}\left[g_1\left(X_1\right) \cdots g_n\left(X_n\right)\right] = \mathbb{E}g_1\left(X_1\right) \cdots \mathbb{E}g_n\left(X_n\right),$$

provided the considered expectations are well defined.

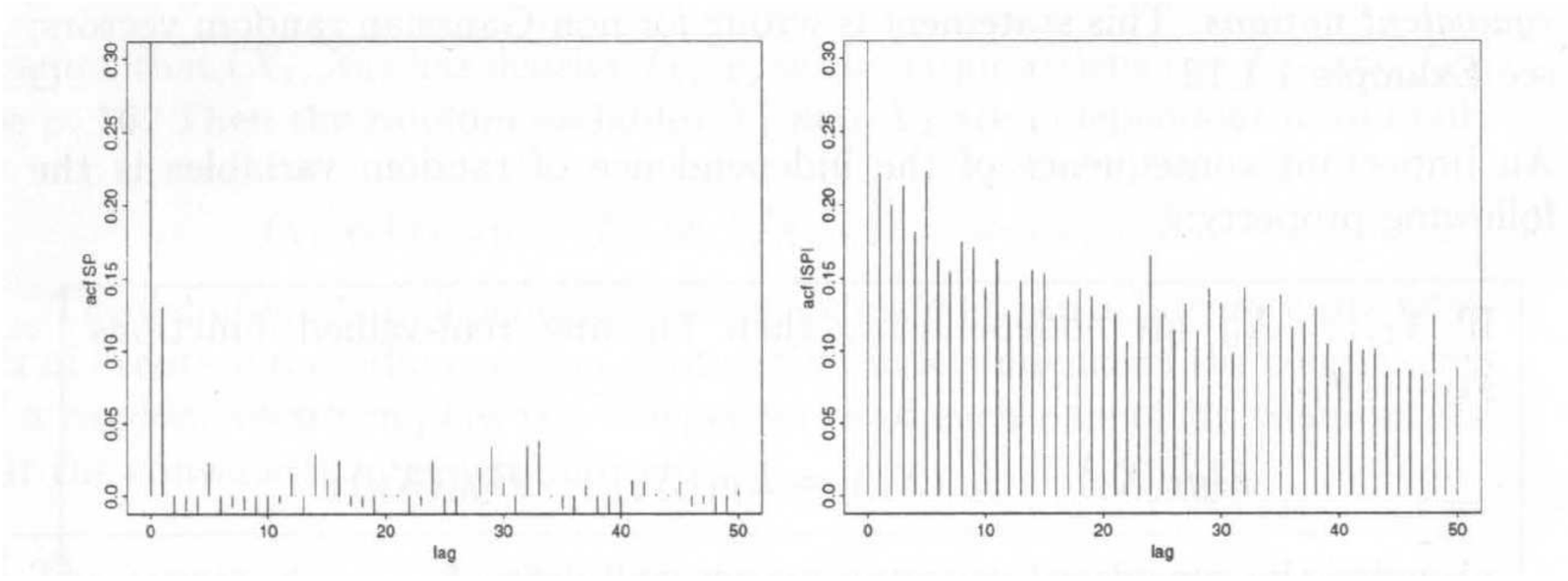
- In particular, we may conclude that the independent random variables X_1 and X_2 are uncorrelated, i.e. $\operatorname{corr}\left(X_1, X_2\right) = \operatorname{Cov}\left(X_1, X_2\right) = 0$.
- The converse is in general not true.

Autocorrelations of a time series

- A time series is simply a sequence of random variables. E.g., denote today's closing price of Tesla by X_0, the closing price in following days by $X_1, X_2, \ldots, X_n, \ldots$, then we get a time series $(X_t, t = 0, 1, 2, \ldots)$.
- For a time series $X_0, X_1, X_2, \ldots$ the autocorrelation at lag h is defined by $\operatorname{corr}(X_0, X_h),\ h = 0, 1, \ldots.$

Example: the series of daily log-returns of the $\mathrm{S\&P}$ index

- A claim which can frequently be found in the literature is that financial time series (derived from stock indices, share prices, exchange rates, etc.) are nearly uncorrelated.
- This is supported by the sample autocorrelations of the daily log-returns X_t of the $\mathrm{S\&P}$ index; see the below Figure.
- In contrast to this observation, the estimated autocorrelations of the absolute values $|X_t|$ are different from zero even for large lags h. This indicates that there is dependence in this time series.



First step towards stochastic processes

- In what follows, we will often deal with infinite collections $(X_t, t \in T)$ of random variables X_t, i.e. T is an infinite index set.
- In this set-up, we may also introduce independence: The collection of random variables $(X_t, t \in T)$ is independent if for every choice of distinct indices $t_1, \ldots, t_n \in T$ and $n \geq 1$ the random variables $X_{t_1}, \ldots, X_{t_n}$ are independent.
- This collection is independent and identically distributed (i.i.d) if it is independent and all random variables X_t have the same distribution.

Stochastic Processes

Example

- We suppose that the exchange rate NZ$/US$ at every fixed instant t between 9 a.m. and 10 a.m. this morning is random.
- Therefore we can interpret it as a realization $X_t(\omega)$ of the random variable X_t, and so we observe $X_t(\omega)$, $9 \leq t \leq 10$.
- In order to make a guess at 10 a.m. about the exchange rate $X_{11}(\omega)$ at 11 a.m. it is reasonable to look at the whole evolution of $X_t(\omega)$ between 9 a.m. and 10 a.m.
- This is also a demand of the high standard technical devices which provide us with almost continuous information about the process considered.

Definition of stochastic processes

- A mathematical model for describing such a phenomenon is called a stochastic process.
- A stochastic process X is a collection of random variables

$$(X_t, t \in T) = (X_t(\omega), t \in T, \omega \in \Omega)\,,$$

  defined on some space Ω.
- For our purposes, T is often an interval, for example $T = [a, b], [a, b)$ or $[a, \infty)$ for $a < b$. Then we call X a continuous-time process in contrast to discrete-time processes. In the latter case, T is a finite or countably infinite set.
- For obvious reasons, the index t of the random variable X_t is frequently referred to as time, and we will follow this convention.

Stochastic process as a function

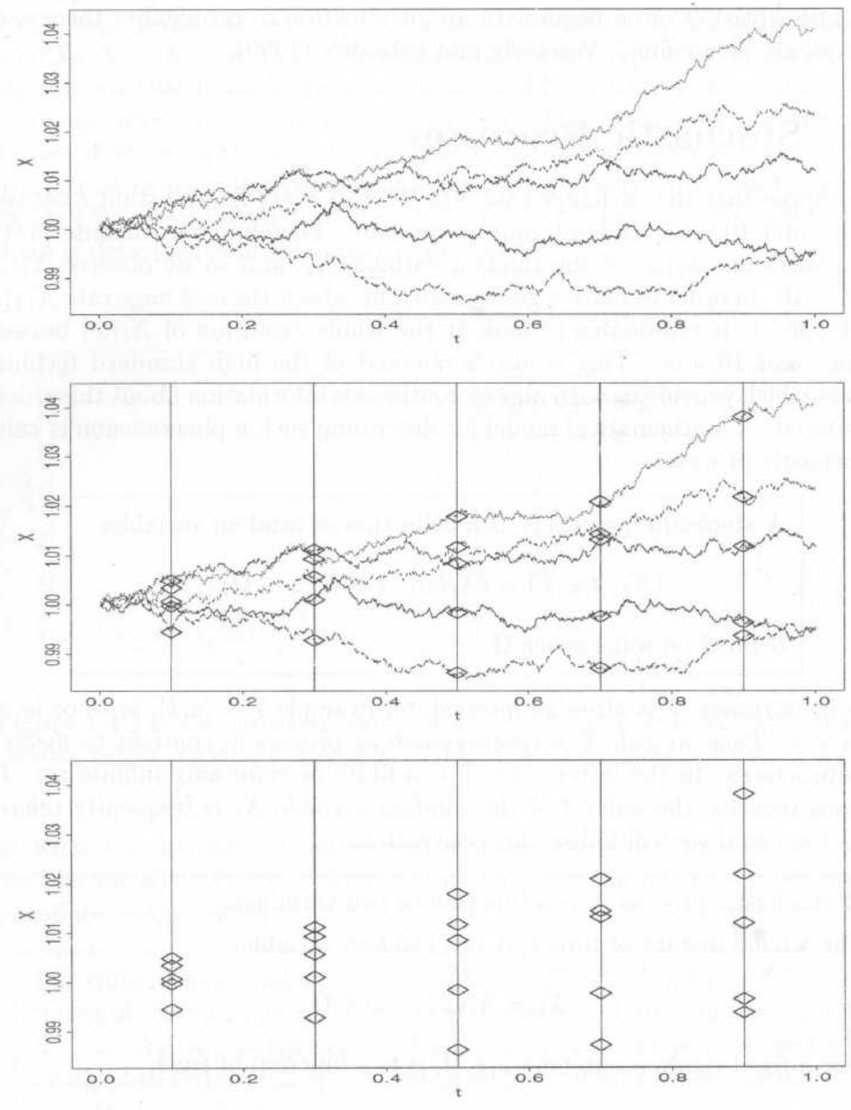
A stochastic process X is a function of two variables. For a fixed instant of time t, it is a random variable:

$$X_t = X_t(\omega), \quad \omega \in \Omega.$$

For a fixed random outcome $\omega \in \Omega$, it is a function of time:

$$X_t = X_t(\omega), \quad t \in T.$$

This function is called a realization, a trajectory or a sample path of the process X.



5 sample paths of a stochastic process $(X_t, t \in [0,1])$. Top: every path corresponds to a different $\omega \in \Omega$. Middle and bottom: the values on the vertical lines at $t = 0.1, \ldots, 0.9$ visualize the random variables $X_{0.1}, \ldots, X_{0.9}$; they occur as the projections of the sample paths on the vertical lines.

Comparison of RVs and stochastic processes

- We see that the concepts of a random variable X and of a stochastic process $(X_t, t \in T)$ are not so much different.
- Both have random realizations, but the realization $X(\omega)$ of a random variable is a number, whereas the realization $X_t(\omega), t \in T$, of a stochastic process is a function on T.
- So we are completely correct if we understand a stochastic process to be a "random element" taking functions as values. Moreover, we can interpret a random variable and a random vector as special stochastic processes with a finite index set T.

Finite-dimensional distributions of a stochastic process

We want to introduce non-random characteristics of a stochastic process such as its distribution, expectation, etc. and describe its dependence structure. This is a task much more complicated than the description of a random vector.

- The key observation is that a stochastic process can be interpreted as a collection of random vectors.
- The *finite-dimensional distributions* (fidis) of the stochastic process X are the distributions of the finite-dimensional vectors

$$(X_{t_1}, \ldots, X_{t_n}), \quad t_1, \ldots, t_n \in T$$

  for all possible choices of times $t_1, \ldots, t_n \in T$ and every $n \geq 1$.
- Fortunately, the fidis completely determine the distributional nature of X. In this sense, we refer to the collection of the fidis as the distribution of the stochastic process.

Gaussian processes

Stochastic processes can be classified according to different criteria. One of them is the kind of fidis.

- Recall from (5) the definition of an n-dimensional Gaussian density.
- A stochastic process is called Gaussian if all its fidis are multivariate Gaussian. We learnt that the parameters μ and Σ of a Gaussian vector are its expectation and covariance matrix, respectively. Hence the distribution of a Gaussian stochastic process is determined only by the collection of the expectations and covariance matrices of the fidis.
- **Example**: The simplest Gaussian process on $T = [0, 1]$ consists of i.i.d $N(0, 1)$ random variables. In this case the fidis are characterized by the distribution functions

$$\mathbb{P}\left(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n\right) = \mathbb{P}\left(X_{t_1} \leq x_1\right) \cdots P\left(X_{t_n} \leq x_n\right) = \Phi\left(x_1\right) \cdots \Phi\left(x_n\right),$$

where $0 \leq t_1 \leq \cdots \leq t_n \leq 1, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Sample paths of a Gaussian process

The sample paths of the process in last example are very irregular. See Figure below for an illustration.

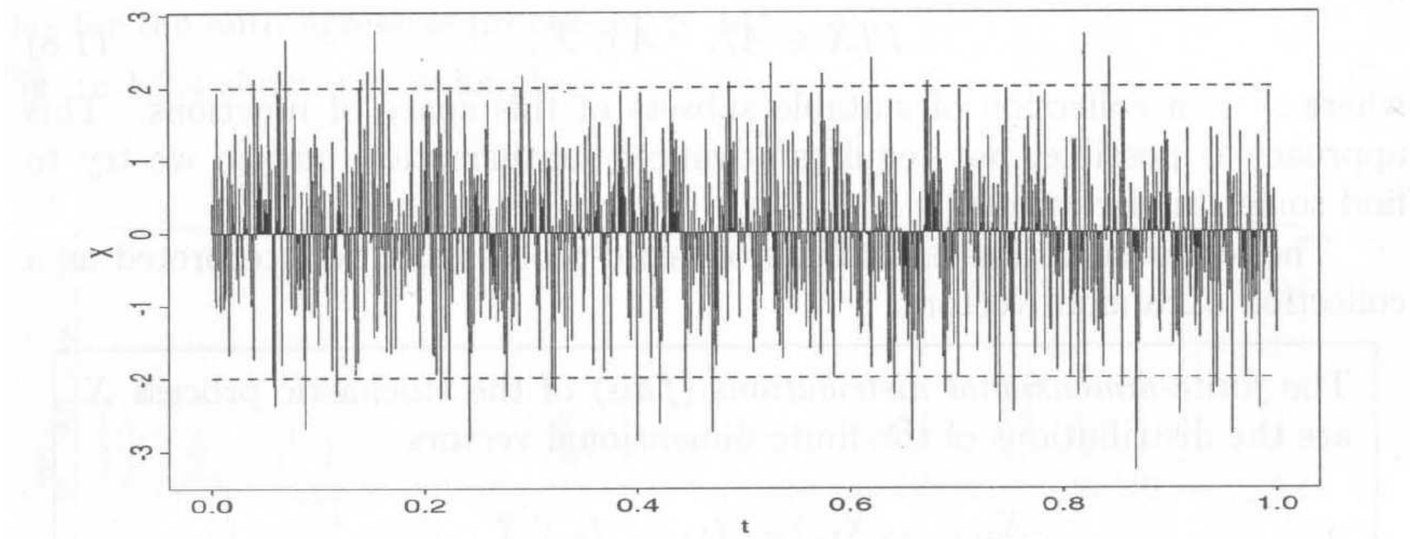


Figure 2: A sample path of the Gaussian process $(X_t, t \in [0,1])$, where the X_t's are iid $N(0,1)$. The expectation function is $\mu_X(t) = 0$ and the dashed lines indicate the curves $\pm 2\sigma_X(t) = \pm 2$.

Mean, variance and covariance function

Consider a stochastic process $X = (X_t, t \in T)$.

- The expectation function of X is given by

$$\mu_X(t) = \mu_{X_t} = \mathbb{E}X_t, \quad t \in T.$$

- The covariance function of X is given by

$$c_X(t,s) = \mathrm{Cov}\left(X_t, X_s\right) = \mathbb{E}\left[\left(X_t - \mu_X(t)\right)\left(X_s - \mu_X(s)\right)\right], \quad t,s \in T.$$

- The variance function of X is given by

$$\sigma_X^2(t) = c_X(t,t) = \mathrm{var}\left(X_t\right), \quad t \in T.$$

Example

Consider the Gaussian process $(X_t, t \in [0,1])$ of iid $N(0,1)$ random variables X_t. Its expectation and covariance functions are given by

$$\mu_X(t) = 0 \quad \text{and} \quad c_X(t,s) = \begin{cases} 1 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

- We learnt that Gaussian processes are determined only via their expectation and covariance functions. This is **not** correct for a non-Gaussian process.

Mean, variance and covariance function cont'd

- As for a random vector, the expectation function $\mu_X(t)$ is a deterministic quantity around which the sample paths of X are concentrated.
- The covariance function $c_X(t, s)$ is a measure of dependence in the process X.
- The variance function $\sigma_X^2(t)$ can be considered as a measure of spread of the sample paths of X around $\mu_X(t)$.

Dependence Structure

- We have already introduced Gaussian processes by specifying their fidis as multivariate Gaussian.
- Another way of classifying stochastic processes consists of imposing a special dependence structure.

Example

We randomly select playing cards from an ordinary deck. We are interested in whether the card is in black (denoted by B) or red color (denoted by R). Let's calculate the chance of observing the sequence RRB using two different sampling methods.

(a) With replacement

$$\begin{aligned}
&\mathbb{P}\left(X_0 = R, X_1 = R, X_2 = B\right) \\
=&\mathbb{P}\left(X_0 = R\right)\mathbb{P}\left(X_1 = R \mid X_0 = R\right)\mathbb{P}\left(X_2 = B \mid X_1 = R, X_0 = R\right) \\
=&\frac{26}{52}\frac{26}{52}\frac{26}{52} \\
\approx&.12500
\end{aligned}$$

In this case the process (X_0, X_1, X_2) shows a structure of independence.

Example cont'd

(b) Without replacement

$$\begin{aligned}
&\mathbb{P}\left(X_0 = R, X_1 = R, X_2 = B\right) \\
=&\mathbb{P}\left(X_0 = R\right)\mathbb{P}\left(X_1 = R \mid X_0 = R\right)\mathbb{P}\left(X_2 = B \mid X_1 = R, X_0 = R\right) \\
=&\frac{26}{52}\frac{25}{51}\frac{26}{50} \\
\approx&.12745
\end{aligned}$$

In this case, (X_0, X_1, X_2) shows some dependence structure.

Example cont'd

Note that

$$\begin{aligned}
&\mathbb{P}\left(X_2 = B \mid X_1 = R\right) \\
&\quad = \frac{\mathbb{P}\left(X_1 = R, X_2 = B\right)}{\mathbb{P}\left(X_1 = R\right)} \\
&\quad = \frac{\mathbb{P}\left(X_0 = R, X_1 = R, X_2 = B\right) + \mathbb{P}\left(X_0 = B, X_1 = R, X_2 = B\right)}{\mathbb{P}\left(X_0 = R\right)\mathbb{P}\left(X_1 = R \mid X_0 = R\right) + \mathbb{P}\left(X_0 = B\right)\mathbb{P}\left(X_1 = R \mid X_0 = B\right)} \\
&\quad = \frac{\frac{26}{52}\frac{25}{51}\frac{26}{50} + \frac{26}{52}\frac{26}{51}\frac{25}{50}}{\frac{26}{52}\frac{25}{51} + \frac{26}{52}\frac{26}{51}} = \frac{13}{51},
\end{aligned}$$

which is different from

$$\mathbb{P}\left(X_2 = B \mid X_1 = R, X_0 = R\right) = \frac{13}{25}.$$

This means, that the behavior of X_2 not only depends on X_1, but also on X_0. In other words, the process is not *memoryless*.

Processes with stationary and independent increments

Let $X = (X_t, t \in T)$ be a stochastic process and $T \subset \mathbb{R}$ be an index set.

- X is said to have stationary increments if $X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h}$ for all $t, s \in T$ (here $\stackrel{d}{=}$ stands for the identity of the distributions) and h with $t + h, s + h \in T$.
- X is said to have independent increments if for every choice of $t_i \in T$ with $t_0 < \cdots < t_n$ and $n \geq 1$,

$$X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$$

are independent random variables.

Random walk

One of the prime examples of processes with independent, stationary increments is the **random walk**.

- Let $e_1, e_2, \ldots$ be a sequence of independent, identically distributed random variables each with zero mean and variance σ_e^2.
- The random walk $\{Y_t : t = 1, 2, \ldots\}$, is constructed as follows:

$$\left.\begin{aligned} Y_1 &= e_1 \\ Y_2 &= e_1 + e_2 \\ &\vdots \\ Y_t &= e_1 + e_2 + \cdots + e_t \end{aligned}\right\} \tag{8}$$

Random walk cont'd

- If the e's are interpreted as the sizes of the "steps" taken (forward or backward) along a number line, then Y_t is the position of the "random walker" at time t.
- $t_i \in \mathbb{Z}_+ := \{0, 1, 2, \ldots\}$ with $t_0 < \cdots < t_n$ and $n \geq 1$,

$$Y_{t_1} - Y_{t_0}, \ldots, Y_{t_n} - Y_{t_{n-1}} = e_{t_0+1} + \cdots + e_{t_1}, \ldots, e_{t_{n-1}+1} + \cdots + e_{t_n}$$

  are indeed independent random variables.
- Moreover,

$$Y_t - Y_s = e_{s+1} + \cdots + e_t \stackrel{d}{=} e_{s+h+1} + \cdots + e_{t+h} = X_{t+h} - X_{s+h}$$

  for all $s < t$ and h.
- So the the random walk has independent, stationary increments.

Random walk cont'd

- From Equation (8), we obtain the mean function

$$\begin{aligned}\mu_t &= E(Y_t) = E(e_1 + e_2 + \cdots + e_t) = E(e_1) + E(e_2) + \cdots + E(e_t) \\ &= 0 + 0 + \cdots + 0\end{aligned}$$

so that $\mu_t = 0$ for all t.

- We also have

$$\begin{aligned}\mathrm{Var}(Y_t) &= \mathrm{Var}(e_1 + e_2 + \cdots + e_t) = \mathrm{Var}(e_1) + \mathrm{Var}(e_2) + \cdots + \mathrm{Var}(e_t) \\ &= \sigma_e^2 + \sigma_e^2 + \cdots + \sigma_e^2\end{aligned}$$

so that

$$\mathrm{Var}(Y_t) = t\sigma_e^2$$

Random walk cont'd

Notice that the process variance increases linearly with time. To investigate the covariance function, suppose that $1 \leq t \leq s$. Then we have

$$\gamma_{t,s} = \mathrm{Cov}\left(Y_t, Y_s\right) = \mathrm{Cov}\left(e_1 + e_2 + \cdots + e_t, e_1 + e_2 + \cdots + e_t + e_{t+1} + \cdots + e_s\right)$$

We have

$$\gamma_{t,s} = \sum_{i=1}^{s} \sum_{j=1}^{t} \mathrm{Cov}\left(e_i, e_j\right)$$

- However, these covariances are zero unless $i = j$, in which case they equal $\mathrm{Var}\left(e_i\right) = \sigma_e^2$.
- There are exactly t of these so that

$$\gamma_{t,s} = t\sigma_e^2.$$

Random walk cont'd

Since $\gamma_{t,s} = \gamma_{s,t}$, this specifies the autocovariance function for all time points t and s and we can write

$$\gamma_{t,s} = t\sigma_e^2 \quad \text{for } 1 \leq t \leq s$$

The autocorrelation function for the random walk is now easily obtained as

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} = \sqrt{\frac{t}{s}} \quad \text{for } 1 \leq t \leq s.$$

About σ-Fields

Some examples of a σ-field

Consider an arbitrary sample space Ω. Check that the following collections of subsets of Ω are σ-fields:

$$\mathcal{F}_1 = \{\emptyset, \Omega\}$$
$$\mathcal{F}_2 = \{\emptyset, \Omega, A, A^c\} \text{ for some } A \neq \emptyset \text{ and } A \neq \Omega,$$
$$\mathcal{F}_3 = \mathcal{P}(\Omega) = \{A : A \subset \Omega\}$$

$\mathcal{F}_1$ is the smallest σ-field on Ω, and $\mathcal{F}_3$, the power set of Ω, is the biggest one, as it contains all possible subsets of Ω.

Generated σ-fields

- Now suppose that $\mathcal{C}$ is a collection of subsets of Ω, but not necessarily a σ-field.
- By adding more sets to $\mathcal{C}$, one can always obtain a σ-field, for example the power set $\mathcal{P}(\Omega)$. However, there are mathematical reasons showing that $\mathcal{P}(\Omega)$ is in general too big.
- But one can also prove that, for a given collection $\mathcal{C}$ of subsets of Ω, there exists a **smallest** σ-field $\sigma(\mathcal{C})$ on Ω containing $\mathcal{C}$. We call $\sigma(\mathcal{C})$ the *σ-field generated by $\mathcal{C}$*.

Remark: Recall the σ-fields on the previous slide. In fact, $\mathcal{F}_i = \sigma\left(\mathcal{C}_i\right)$, where

$$\mathcal{C}_1 = \{\emptyset\}, \quad \mathcal{C}_2 = \{A\}, \quad \mathcal{C}_3 = \mathcal{F}_3.$$

The Borel sets

Take $\Omega = \mathbb{R}$ and

$$\mathcal{C} = \{(a, b] : -\infty < a < b < \infty\}$$

The σ-field $\mathcal{B}(\mathbb{R}) = \sigma\left(\mathcal{C}\right)$ contains very general subsets of $\mathbb{R}$. It is called the Borel σ-field, its elements are the Borel sets.

- A normal human being cannot imagine the large variety of Borel sets. For example, it is a true fact, but not easy to verify, that $\mathcal{B}_1$ is a genuine subset of the power set $\mathcal{P}(\mathbb{R})$.
- In order to show that a given subset $C \subset \mathbb{R}$ is a Borel set, it is necessary to obtain C by a countable number of operations $\cap, \cup, ^c$ acting on the rectangles. For example, show that every point set $\{a\}, a \in \mathbb{R}$, is a Borel set. Also check that the intervals $(a, b), [a, b)(-\infty, a), (b, \infty)$ are Borel sets.

σ-Field generated by a RV

Let Y be a random variable. We define the σ-fields $\sigma(Y)$ as the smallest σ-field containing all sets of the form

$$\{Y \in (a, b]\} = \{\omega : a < Y(\omega) \leq b\}, \quad -\infty < a < b < \infty.$$

- We call $\sigma(Y)$ the σ-field generated by the random variable Y.
- The σ-field $\sigma(Y)$ generated by Y contains the essential information about the structure of the random variable Y as a function of $\omega \in \Omega$. It contains all sets of the form $\{\omega : Y(\omega) \in C\}$ for all Borel sets $C \subset \mathbb{R}$.

σ-Field generated by a family of RVs

Let I be an index set. For a family of random variables $(Y_i, i \in I)$, the σ-field $\sigma(Y_i, i \in I)$ is the smallest σ-field containing all sets of the form

$$A_{i,C} = \{\omega : X_i(\omega) \in C\}, \quad i \in I, C \text{ is a Borel set.}$$

- This is the smallest σ-field containing the essential information about the family of random variables $(Y_i, i \in I)$.
- From the mathematical point of view, the σ-field $\sigma(Y_i, i \in I)$ is the smallest σ-field for which all Y_i, $i \in I$ become measurable.