Stochastic Processes and Stochastic Calculus - 1
Review of Basic Concepts in Probability

Prof. Maurizio Pratelli

Università degli Studi di Pisa

San Miniato - 12 September 2016

Overview

What is Probability?

Probability is a quantitative mathematical theory of reasoning and arguing about random (= uncertain) facts, called events.

Uncertainty may be due to incomplete information about these facts.

The starting points of mathematical probability are the following natural definitions.

- Ω: an abstract set collecting all the possible outcomes of the underlying experiments/facts that we want to study.
- A family $\mathscr{A}$ of subsets of Ω, which is an algebra of "events", i.e.
  1. $\emptyset \in \mathscr{A}$, $\Omega \in \mathscr{A}$
  2. if $A \in \mathscr{A}$, then $A^c \in \mathscr{A}$ (where $A^c = \Omega \setminus A$ is the "complement of A")
  3. if $A, B \in \mathscr{A}$, then $(A \cup B) \in \mathscr{A}$, $(A \cap B) \in \mathscr{A}$.

The family $\mathscr{A}$ represents all the possible statements about our facts to which we want to assign some measure of plausibility, i.e. a probability.

What is Probability?

Let Ω be a set, $\mathscr{A}$ be an algebra of events of Ω.

We can define a probability measure as a function

$$P : \mathscr{A} \to [0, 1], \quad A \mapsto P(A)$$

such that

$$P(\Omega) = 1, \qquad P(\emptyset) = 0, \qquad \text{and}$$

if $A, B \in \mathscr{A}$ are incompatible, i.e. $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

From this rules much more follows, e.g.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

for any two events $A, B \in \mathscr{A}$.

σ-fields and σ-additivity

For mathematical convenience (passage to limits), we require more on $\mathscr{A}$, i.e. being stable with respect to countable (σ-) operations

$$\text{If } (A_n)_{n\geq 1} \subseteq \mathscr{A}, \text{ then } \bigcup_{n=1}^{\infty} \in \mathscr{A}, \text{ and } \bigcap_{n=1}^{\infty} \in \mathscr{A}.$$

The family $\mathscr{A}$ is called σ-field.

At the same time, we strengthen our assumptions on P, requiring σ-additivity:

If $(A_n)_{n\geq 1} \subseteq \mathscr{A}$, are pairwise incompatible, i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_i).$$

These axioms are historically due to A. Kolmogorov.

Probabilities on $\mathbb{R}$

An interesting case is $\Omega = \mathbb{R}$. We have two types of probability measures.

Discrete probability

Fix a finite or countable set of points

$$x_1, x_2, \ldots, x_n, \ldots \in \mathbb{R}$$

and numbers

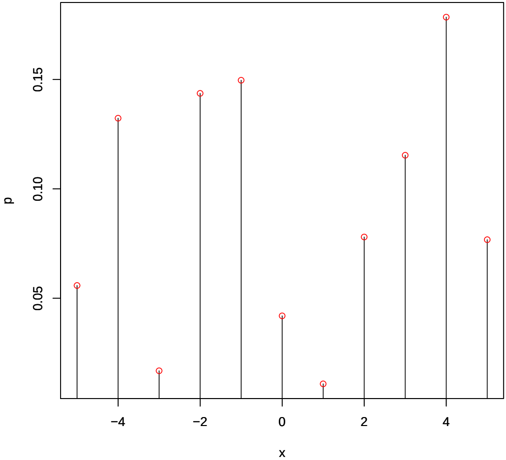
$$p(x_1), p(x_2), \ldots, p(x_n), \ldots \in [0,1], \quad \text{with} \quad \sum_{k=1}^{\infty} p(x_k) = 1.$$

Define for any $A \subseteq \mathbb{R}$,

$$P(A) := \sum_{x_k \in A} p(x_k).$$

Notice that P is defined an all subsets of $\mathbb{R}$, i.e. $\mathscr{A} = \mathcal{P}(\mathbb{R})$.

An example:



$$x = (-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)$$

$$p \sim (0.06, 0.13, 0.02, 0.14, 0.15, 0.04, 0.01, 0.08, 0.1, 0.17, 0.08)$$

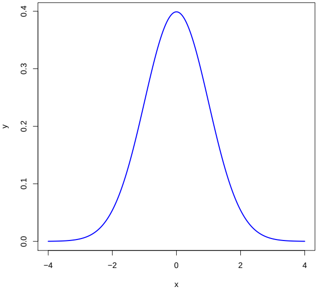
Probability defined by a density

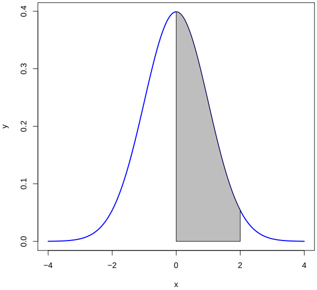
Fix a function $f : \mathbb{R} \to [0, \infty)$,

$$f(x) \geq 0, \qquad \int_{\mathbb{R}} f(x)dx = 1.$$

We want to define for $A \subseteq \mathbb{R}$

$$P(A) := \int_A f(x)dx$$





In general, P cannot be defined on all subsets of $\mathbb{R}$.

Borel subsets

For instance, $P(A) = \int_A f(x)dx$ can be defined on Borel subsets of $\mathbb{R}$, i.e. those that can be obtained via a countable number of operations of the type

$$A \cup B, \quad A \cap B, \quad \text{and} \quad (A)^c$$

starting from intervals.

Definition

Borel sets are the smallest σ-field containing all intervals.

All the sets that appear on practical computations turn out to be Borel, so we never worry again about this fact.

Usually we are not interested on a single event A, but we have a family of pairwise incompatible events (A_n), with $A_n \cap A_m = \emptyset$, and $\bigcup_n A_n = \Omega$.

We can use the notion random variable, i.e. functions $X : \Omega \to \mathbb{R}$.

Example

A toss of a dice can give results in the set $\{1, 2, 3, 4, 5, 6\}$, so we have 6 events

$$A_n = \{\text{the outcome is the face numbered } n\}, \quad n \in \{1, 2, 3, 4, 5, 6\}.$$

Alternatively, we may introduce the random variable

$$X : \Omega \to \{1, 2, 3, 4, 5, 6\}$$

$$X(\omega) = n \quad \text{on } A_n.$$

We recover the events A_n simply writing

$$A_n = \{\omega \in \Omega \,:\, X(\omega) = n\} = \{X = n\} = X^{-1}(n).$$

Laws of real random variables

We have two very related notions:

a) A random variable $X : \Omega \to \mathbb{R}$, i.e. such that, for every $x \in\in \mathbb{R}$,

$$\{X \leq x\} = X^{-1}(-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathscr{A}.$$

b) A probability distribution Q on Borel sets of $\mathbb{R}$.

The link is given by the formula, for $A \subseteq \mathbb{R}$ Borel,

$$Q(A) = P(X^{-1}(A)) = P(X \in A) = P(\{\omega \in \Omega \, : \, X(\omega) \in A\}).$$

We call Q the law of X and write P_X.

Identically distributed r.v.

Notice that two random variables can have the same law (we call them identically distributed) but be very different. Consider for example the outputs of two different tosses of a dice.

Cumulative distribution function (c.d.f.)

A useful way to describe the law of a r.v. X is through the function

$$F_X : \mathbb{R} \to [0,1], \quad F_X(x) = P(X \leq x) = P_X((-\infty, x]).$$

Correspondingly to the fact that we have probabilities that are discrete or defined via a density, we obtain

- Discrete random variables (when the law is discrete)

$$P(X \in A) = \sum_{x \in A} p(x),$$

- Random variables with density (also called, but not properly, continuous random variables)

$$P(X \in A) = \int_A f(x)dx.$$

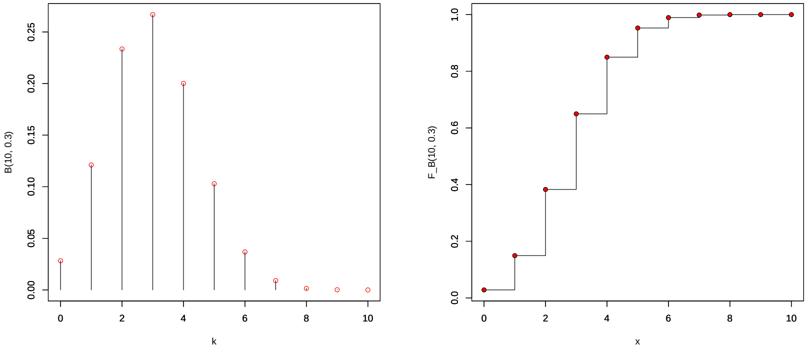
The function f is called probability density.

Examples

Binomial law

For $n \in \mathbb{N}$, $n \geq 1$, $p \in [0,1]$, we say that X is Binomial $B(n,p)$ if

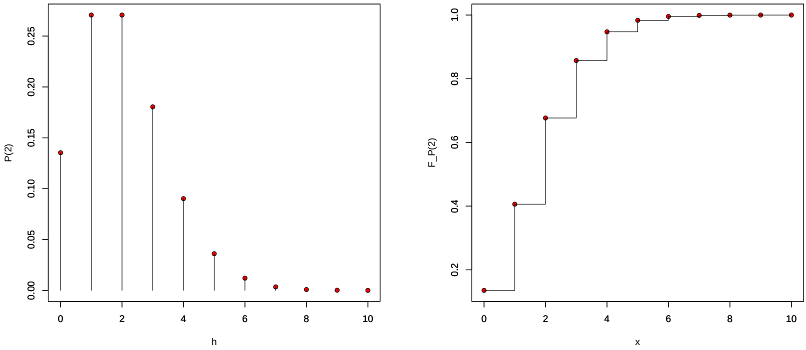
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k \in \{0, 1, \dots, n\}$$



Poisson law

For $\lambda > 0$, we say that X is Poisson $\mathcal{P}(\lambda)$ if

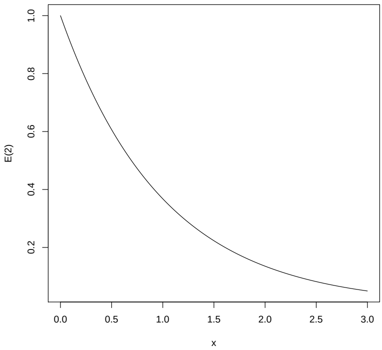
$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\frac{\lambda^k}{k\cdot(k-1)\cdot(k-2)\cdot\ldots\cdot2\cdot1} \quad \text{for } k \in \mathbb{N}$$

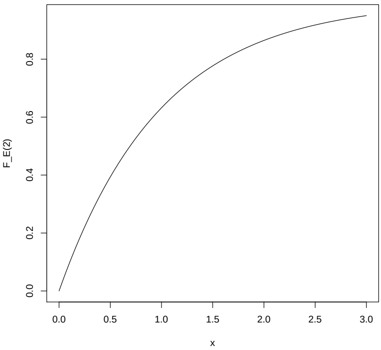


Exponential law

For $\lambda > 0$, we say that X is Exponential $\mathcal{E}(\lambda)$ if it has density given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

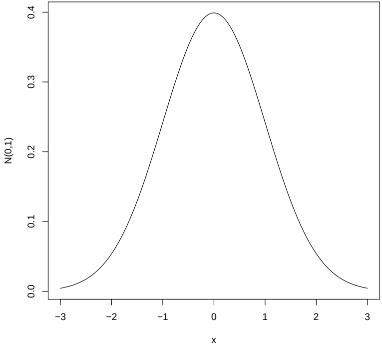


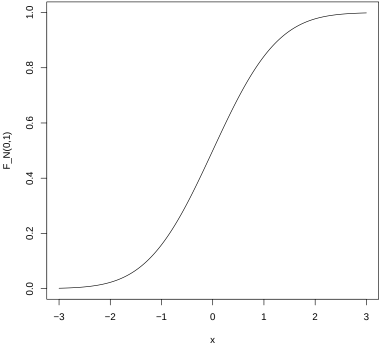


Gaussian law

For $\mu \in \mathbb{R}$, $\sigma^2 > 0$, we say that X is Gaussian (or normal) $\mathcal{N}(\mu, \sigma^2)$ if it has density given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}}.$$





Expectation

The mathematical expectation (if it exists) gives a first approximation of the average values assumed by a random variable X.

There is a general definition, let us recall first the two important cases:

- Discrete r.v. X

  if $\sum_k |x_k|\, p(x_k) < \infty$ then $E[X]$ exists and is given by

$$E[X] = \sum_k x_k p(x_k).$$

- Variables X with density f

  if $\int_{\mathbb{R}} |x|\, f(x) dx < \infty$ then $E[X]$ exists and is given by

$$E[X] = \int_{\mathbb{R}} x\, f(x) dx.$$

Expectation

The general formula, if it is defined, reads as follows

$$E[g(X)] = \int_{\Omega} g(X(\omega))dP(\omega) = \int_{\mathbb{R}} g(x)dP_X(x) = \begin{cases} (a) \\ (b) \end{cases}$$

Where

(a) $= \sum_k g(x_k)p(x_k)$ for X discrete

(b) $= \int_{\mathbb{R}} g(x)f(x)dx$ for X with density f.

Notice that we can take any (Borel) $g : \mathbb{R} \to \mathbb{R}$, not only $g(x) = x$.

Moments

For $k \in \mathbb{N}$, we define the k-th moment of a random variable X as

$$m_k = E\left[X^k\right]$$

if it exists, i.e. $E\left[|X|^k\right] < \infty$.

The expectation of X is the first moment of X.

Moments provide better information on the law of X. Related to the second moment is the variance of X

$$\sigma^2(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2$$

which roughly quantifies the dispersion of the law of X around the expectation $E[X]$. One writes an approximation $X \sim E[X] \pm \sigma(X)$. Mathematically,

$$P(|X - E[X]| > \lambda\sigma(X)) \leq \frac{1}{\lambda^2}$$

Examples

Binomial

If X is $B(n, p)$, then

$$E[X] = np \quad \text{and} \quad \sigma(X)^2 = np(1-p),$$

hence we could write informally

$$X \sim np \pm \sqrt{np(1-p)}$$

Gaussian

If X is $N(\mu, \sigma^2)$, then

$$E[X] = \mu \quad \text{and} \quad \sigma(X)^2 = \sigma^2,$$

hence we could write informally

$$X \sim \mu \pm \sigma$$

If $B \in \mathscr{A}$ is such that $P(B) > 0$, i.e. it is not a "null-set", or we think it has some chance to occur, we define

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

that we interpret as the new probability of A if we know that B occurred.

Example

Say that the probability that tomorrow rains is $P(A) = 0.3$. Should we change this probability if we know

$B =$ "the TV weather forecasts say it is going to be sunny"?

We say that A and B are independent if

$$P(A|B) = P(A) \quad \Leftrightarrow \quad P(B|A) = P(B) \quad \Leftrightarrow \quad P(A \cap B) = P(A)P(B),$$

We let the last identity be the definition of independence (it works even if $P(A)$ or $P(B) = 0$).

Families of independent events

How to generalize independence to larger families of events?

Independent family of events

We say that $A_1, A_2, \ldots, A_i, \ldots \in \mathscr{A}$ are independent if for every finite subset $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$, we have

$$P\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right) = P\left(A_{i_1}\right) \cdot P\left(A_{i_2}\right) \cdot \ldots \cdot P\left(A_{i_k}\right).$$

Vector-valued random variables

Usually we do not work with a single random variable X, but with many of them

$$X_1, X_2, \dots, X_n$$

We can collect them into a single vector-valued random variable

$$X = (X_1, X_2, \dots, X_n) : \Omega \to \mathbb{R}^n.$$

Example

Consider n repeated tosses of a dice, and collect each outcome in different variables X_i, for $i \in \{1, \dots, n\}$.

Sometimes we can even consider infinitely many random variables $(X_1, X_2, \dots, X_n, \dots)$.

Joint density of vector-valued random variables

Let us consider first the case of a couple $(X, Y) : \Omega \to \mathbb{R}^2$.
We say that (X, Y) has joint density $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ if, for every $A \subseteq \mathbb{R}^2$ Borel, we have

$$P_{X,Y}(A) = P((X, Y) \in A) = \int_A f_{X,Y}(x, y)dxdy$$

Borel sets of $\mathbb{R}^2$ are defined as Borel sets of $\mathbb{R}$, but starting from rectangles $A_1 \times A_2$.

If we know (and it exists) the joint density $f_{X,Y}$, we can obtain the densities of the laws of X and Y:

$$f_X(x) = \int_{\mathbb{R}} f(x, y)dy, \qquad f_Y(y) = \int_{\mathbb{R}} f(x, y)dx.$$

We call them marginal densities.

Remark

In general, we cannot obtain the joint density of (X, Y) knowing only the marginals f_X, f_Y: the joint density contains information on correlations.

Independent random variables

There is a special case when we can reconstruct the joint density knowing only f_X and f_Y.

Definition (Independent r.v.'s)

We say that $X, Y : \Omega \to \mathbb{R}$ are independent if, for any $A_1, A_2 \subseteq \mathbb{R}$ Borel

$$\text{the events } \{X \in A_1\} \text{ and } \{Y \in A_2\} \text{ are independent}$$

In words: X and Y are independent if any information about X does not influence our degree of uncertainty about Y (and viceversa).

If X and Y are independent, then $g(X)$, $h(Y)$ are independent as well.

$$E[XY] = E[X]\,E[Y] \quad \text{(akin to } P(A \cap B) = P(A)P(B).\text{)}$$

Theorem

Assume that X and Y have densities f_X and f_Y. Then

$$X \text{ and } Y \text{ are independent} \quad \Leftrightarrow \quad \text{the joint density of } (X, Y) \text{ is } f_X(x)f_Y(y).$$

Correlations

Lack of independence is not necessarily related to physical causation.

Example (Correlation is NOT necessarily causation)

The weather forecast for tomorrow and the actual weather of tomorrow are not independent, but none physically causes the other.

We look for some quantification of lack of independence. One is the covariance

$$\mathrm{Cov}(X, Y) := E\left[(X - E[X])\,(Y - E[Y])\right].$$

- If $\mathrm{Cov}(X, Y) > 0$, X, Y are positively correlated
- If $\mathrm{Cov}(X, Y) < 0$, X, Y are negatively correlated
- If $\mathrm{Cov}(X, Y) = 0$, X, Y are not correlated

Example

Assume that $Y = g(X)$, where $g : \mathbb{R} \to \mathbb{R}$ is increasing, i.e.

$$x_1 < x_2 \Rightarrow g(x_1) < g(x_2).$$

Then we have $\mathrm{Cov}(X, Y) > 0$. Positive correlation is a weaker condition!

Independence and absence of correlation

If X and Y are independent $\Rightarrow$ they are not correlated: $X - E[X]$ and $Y - E[Y]$ are independent, hence

$$E[(X - E[X])(Y - E[Y])] = E[X - E[X]]\, E[Y - E[Y]] = 0.$$

$$(X, Y) \text{ not correlated } \Leftrightarrow \mathrm{Cov}(X, Y) = 0 \Leftrightarrow \mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

In general,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

Independence is a stronger condition than being not correlated: there are r.v.'s not independent but with $\mathrm{Cov}(X, Y) = 0$.

Sums of independent r.v.'s – discrete case

Let X, Y be the outcome of two tosses of a dice. What is the law of $X+Y$?
We have $X+Y : \Omega \to \{2,3,\dots,12\}$. For example

$$P(X+Y=7) = \sum_{k=1}^{6} P(X=k, Y=7-k) = \sum_{k=1}^{6} P(X=k)P(Y=7-k)$$

because the two tosses are independent.

In general,

$$P(X+Y=n) = \sum_{k=1}^{6} P(X=k)P(Y=n-k),$$

we call the sum on the right a convolution in the discrete case.

Sums of independent r.v.'s – continuous case

Let X, Y be two real random variables with laws f_X, f_Y and independent
What is the law of $X + Y$?

We have the convolution formula for the density of $X + Y$:

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z - x)\, dx.$$

Fundamental example

Let X be $\mathcal{N}(\mu_1, \sigma_1^2)$, Y be $\mathcal{N}(\mu_2, \sigma_2^2)$, and independent. Then

$$X + Y \text{ is } \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$