Stochastic Models (Lecture #2)

Thomas Verdebout

Université libre de Bruxelles (ULB)

Summary

In the first lecture, we learned about the concepts of

- random experiments,
- probability measures, and
- random variables.

We also

- defined the distribution of a rv and
- considered two main types of distributions
  (discrete and absolutely continuous).

Outline of the course

1. A short introduction.
2. $\boxed{\text{2. Basic probability review.}}$
3. Martingales.
4. Markov chains.
5. Markov processes, Poisson processes.
6. Brownian motions.

Today

Today, our goal is

- to define integration of rv,
- to define the usual moments/parameters that summarize the distribution of a rv,

and

- to consider random vectors,
- to define the concept of conditional probability.

Integration, moments, and parameters

Let X be a rv on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

We know the distribution of X is characterized by (e.g.) its cdf F. But this is quite complicated, and we wonder whether it could be possible to summarize this distribution by one (or several) well chosen number(s)...

If one had to use a single number, what should it be?
How to define a quantity that would best represent the "most typical value" of X?

A weighted average (involving weights from the probability distribution of X) seems to be a good choice...

$$\rightsquigarrow \text{ the expectation of } X : \quad \mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega).$$

How to define this integral ?

Integration, moments and parameters

Our plan is the following:

- defining (rigorously) $\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$,
- studying how to compute such quantities in practice (both for discrete and absolutely continuous rv's.), and
- defining the standard parameters (based on such integrals) that are used to summarize distributions.

Rigorous definition of $\int_\Omega X \, d\mathbb{P}$

The definition proceeds in 3 steps.

A. Assume first that X is a step function, i.e., that $X = \sum_{i=1}^{n} a_i I_{A_i}$ (with $a_i \in \mathbb{R}$), where $A_1, \dots, A_n \in \mathcal{A}$ is a partition of Ω (meaning that the A_i's are pairwise disjoint with $\cup_i A_i = \Omega$).

In other words,

$$X(\omega) = \begin{cases} a_1 & \text{if} \quad \omega \in A_1 \\ & \vdots \\ a_n & \text{if} \quad \omega \in A_n. \end{cases}$$

Then, we define

$$\mathbb{E}[X] = \int_\Omega X(\omega) \, d\mathbb{P}(\omega) = \sum_{i=1}^{n} a_i \mathbb{P}[A_i].$$

Rigorous definition of $\int_\Omega X\, d\mathbb{P}$

This is extended to any rv in the following way:

B. Assume that $X \geq 0$. Then, we define

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \lim_{n\to\infty} \int_\Omega X_n(\omega)\, d\mathbb{P}(\omega),$$

where the X_n's are step functions such that $(X_n) \nearrow X$.

C. For a general X, we define

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int_\Omega X^+(\omega)\, d\mathbb{P}(\omega) - \int_\Omega X^-(\omega)\, d\mathbb{P}(\omega),$$

provided that both integrals are finite (we then say that X is integrable).

Remark: if X is integrable, $|X| = X^+ + X^-$ is also integrable.

Some properties of $\int_\Omega X \, d\mathbb{P}$

- $\mathbb{E}[\alpha_1 X_1 + \alpha_2 X_2] = \alpha_1 \mathbb{E}[X_1] + \alpha_2 \mathbb{E}[X_2]$, for all $\alpha_1, \alpha_2 \in \mathbb{R}$.
- $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
- If X integrable, then $\mathbb{P}[A] = 0 \Rightarrow \int_A X(\omega) \, d\mathbb{P}(\omega) = 0$ here, $\int_A X(\omega) \, d\mathbb{P}(\omega) := \int_\Omega X(\omega) \, I_A(\omega) \, d\mathbb{P}(\omega)$.
- If $A_1, A_2, \ldots \in \mathcal{A}$ are pairwise disjoint, $\int_{\cup_i A_i} X(\omega) \, d\mathbb{P}(\omega) = \sum_i \int_{A_i} X(\omega) \, d\mathbb{P}(\omega)$.
- If X_n is integrable for all n and $X_n \nearrow X$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ (monotone convergence theorem).
- If $|X_n| \leq Y$, where Y is integrable, and if $\mathbb{P}[\{\omega \mid X_n(\omega) \to X(\omega)\}] = 1$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ (dominated convergence theorem).

Integration, moments, and parameters

Let X be an integrable rv on $(\Omega, \mathcal{A}, \mathbb{P})$.

Assume that X is discrete, say with distribution

distribution of X			
values	x_1	x_2	$\ldots$
probabilities	p_1	p_2	$\ldots$

Then

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \sum_i x_i p_i.$$

Similarly, if $g(X)$ is an integrable rv,

$$\mathbb{E}[g(X)] = \int_\Omega g(X(\omega))\, d\mathbb{P}(\omega) = \sum_i g(x_i) p_i.$$

Integration, moments, and parameters

Let X be a $\mathbb{R}$-valued and integrable rv on $(\Omega, \mathcal{A}, \mathbb{P})$.

Assume that X is absolutely continuous, with pdf f, say.

Then

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

Similarly, if $g(X)$ is an integrable rv,

$$\mathbb{E}[g(X)] = \int_\Omega g(X(\omega))\, d\mathbb{P}(\omega) = \int_\mathbb{R} g(x)\, f(x)\, dx.$$

Before considering some examples... our plan is defining the standard parameters (based on such integrals) that are used to summarize distributions.

Integration, moments, and parameters

Let X be an integrable rv on $(\Omega, \mathcal{A}, \mathbb{P})$.

The expectation (the “mean value") of X is $\mu_X := \mathbb{E}[X]$.
This is a measure of location.

The variance of X is $\sigma_X^2 = \mathrm{Var}[X] := \mathbb{E}[(X - \mu_X)^2]$ (if $\mathbb{E}[X^2] < \infty$). This is a measure of dispersion.

Remark: $\mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2 - 2\mu_X X + \mu_X^2] = \mathbb{E}[X^2] - \mu_X^2$

The standard deviation of X is $\sigma_X = \sqrt{\mathrm{Var}[X]}$
(advantage over the variance: it has the same unit as X itself).

These quantities try to summarize the distribution of X.
But of course, much information is lost...

Integration, moments, and parameters

More generally, one defines

- The moment of order k of X as $\mathbb{E}[X^k]$.
- The absolute moment of order k of X as $\mathbb{E}[|X|^k]$.
- The centered moment of order k of X as $\mathbb{E}[(X - \mu_X)^k]$.

Remarks:

- Assumptions of finite moments...
- If $\mathbb{E}[|X|^k] < \infty$ (for some $k > 0$), then $\mathbb{E}[|X|^j] < \infty$ for all $j \leq k$ (if particular, $\mathrm{Var}[X] < \infty \Rightarrow \mathbb{E}[|X|] < \infty$).

It is time for examples...

Example (discrete):
assume you bet b euros on “even" in one game of roulette...

$\mathcal{E}$: roulette, $\Omega = \{0, 1, 2, \ldots, 36\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}[A] = \frac{\#A}{\#\Omega}$.

Your (random) gain X is given by

$$X(\omega) = \begin{cases} -b & \text{if} \quad \omega \in \{0, 1, 3, \ldots, 35\} \\ b & \text{if} \quad \omega \in \{2, 4, \ldots, 36\}, \end{cases}$$

whose distribution is

distribution of X		
values	$-b$	b
probabilities	$\frac{19}{37}$	$\frac{18}{37}$

Therefore,

your expected winning is

$$\mathbb{E}[X] = \sum_i x_i p_i = (-b) \times \frac{19}{37} + b \times \frac{18}{37} = \frac{-b}{37}.$$

Similarly,

$$\mathbb{E}[X^2] = \sum_i (x_i)^2 p_i = (-b)^2 \times \frac{19}{37} + b^2 \times \frac{18}{37} = b^2$$

so that

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = b^2 - \left(\frac{-b}{37}\right)^2 = \left(1 - \frac{1}{37^2}\right) b^2.$$

Integration, moments, and parameters

Exercises: check that

- if $X \sim \mathrm{Bern}(p)$,

$$\mathbb{E}[X] = p \quad \text{and} \quad \mathrm{Var}[X] = p(1-p);$$

- if $X \sim \mathrm{Bin}(n, p)$,

$$\mathbb{E}[X] = np \quad \text{and} \quad \mathrm{Var}[X] = np(1-p);$$

- if $X \sim \mathcal{P}(\lambda)$,

$$\mathbb{E}[X] = \lambda \quad \text{and} \quad \mathrm{Var}[X] = \lambda.$$

Integration, moments, and parameters

Example (absolutely continuous):
assume that $X \sim \mathrm{Unif}(a, b)$, i.e., that X has the pdf

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\, f(x)\, dx = \int_a^b x\, \frac{1}{b-a}\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{a+b}{2}.$$

Similarly (exercise),

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2\, f(x)\, dx = \int_a^b x^2\, \frac{1}{b-a}\, dx = \ldots = \frac{a^2+ab+b^2}{3},$$

so that $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \ldots = \frac{(b-a)^2}{12}$.

Integration, moments, and parameters

Exercises: check that

- if $X \sim \mathrm{Exp}(\lambda)$,

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{and} \quad \mathrm{Var}[X] = \frac{1}{\lambda^2};$$

- if $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}[X] = \mu \quad \text{and} \quad \mathrm{Var}[X] = \sigma^2.$$

Random vectors

A further step towards sp ...

So far we have considered a single rv

$\rightsquigarrow$ Here we will consider jointly several rv $X_1, \ldots, X_k$, collected in a so-called random vector (rv!!!)

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}.$$

Remarks:

- Below, we will only treat the case $k = 2$;
- however, the extension to the general case is trivial.

Random vectors

Consider again a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Formally,

A bivariate (i.e., $k = 2$) random vector X is a mapping

$$\begin{aligned} X : \quad \Omega \quad &\to \quad \mathbb{R}^2 \\ \omega \quad &\mapsto \quad X(\omega) = \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \end{pmatrix}, \end{aligned}$$

which satisfies $X^{-1}(B) = \{\omega \in \Omega \,|\, X(\omega) \in B\} \in \mathcal{A}$, for all $B \in \mathcal{B}^2$. $\quad (*)$

Remarks:

- (*) is still referred as the measurability property.
- (*) still allows for speaking of $\mathbb{P}[X \in B]$ for all $B \in \mathcal{B}^2$.
- In (*), $\mathcal{B}^2$ denotes the two-dimensional Borel σ-algebra, that is,

$$\sigma\Big(\{(a_1, b_1] \times (a_2, b_2], a_1 < b_1, a_2 < b_2\}\Big).$$

Random vectors

Terminology: X_1 and X_2 are the marginals of X. They are rv.

The distribution of X is the probability measure $\mathbb{P}_X$ defined on the measurable space $(\mathbb{R}^2, \mathcal{B}^2)$ by

$$\mathbb{P}_X[B] := \mathbb{P}[X \in B], \quad B \in \mathcal{B}^2.$$

$\rightsquigarrow$ Again, there are “essentially" two types of multivariate distributions: discrete ones and absolutely continuous ones.

Random vectors (discrete case)

The distribution $\mathbb{P}_X$ of X is discrete if there exists a (finite or infinite) sequence

$$x_i = \begin{pmatrix} (x_i)_1 \\ (x_i)_2 \end{pmatrix}, \quad i = 1, 2, \ldots$$

of real couples such that

$$\mathbb{P}_X[B] = \sum_{i \,|\, x_i \in B} \mathbb{P}\left[X = x_i\right].$$

Again, such a distribution is characterized by the (vector) values x_i that X can assume along with the corresponding probabilities p_i.

distribution of X			
(vector) values	x_1	x_2	$\ldots$
probabilities	p_1	p_2	$\ldots$

Random vectors (discrete case)

If each marginal is integrable, one can obtain the expectation of X componentwise:

$$\mathbb{E}[X] = \sum_i x_i p_i = \sum_i \begin{pmatrix} (x_i)_1 \\ (x_i)_2 \end{pmatrix} p_i = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{pmatrix}$$

If $g : \mathbb{R}^2 \to \mathbb{R}$ is such that $g(X)$ is an integrable random variable, we have

$$\mathbb{E}[g(X)] = \sum_i g(x_i) p_i.$$

Random vectors (abs. continuous case)

The distribution $\mathbb{P}_X$ of X is absolutely continuous if, for all $B \in \mathcal{B}^2$ such that $m_2[B] = 0$, we have $\mathbb{P}_X[B] = 0$.

It can be shown (Radon-Nikodym theorem) that, for such X, there exists a so-called probability density function (pdf) $f : \mathbb{R}^2 \to \mathbb{R}$ such that, for all $B \in \mathcal{B}^2$,

$$\mathbb{P}_X[B] = \int_B f(x)\, dx.$$

Properties of a pdf: $f(x) \geq 0$ and $\int_{\mathbb{R}^2} f(x)\, dx = 1$.

distribution of X	
(vector) values	x
pdf	$f(x)$

Random vectors (abs. continuous case)

If each marginal is integrable, one can obtain the expectation of X componentwise:

$$\mathbb{E}[X] = \int_{\mathbb{R}^2} x\, f(x)\, dx = \int_{\mathbb{R}^2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} f(x)\, dx = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{pmatrix}$$

If $g : \mathbb{R}^2 \to \mathbb{R}$ is such that $g(X)$ is an integrable random variable, we have

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^2} g(x)\, f(x)\, dx.$$

Random vectors

In both cases, one can define the cumulative distribution function (cdf) F by

$$F(x) = \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2].$$

- The marginal distribution of X_1 is obtained via $\lim_{x_2 \to \infty} \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2]$.
- If $X = (X_1, X_2)$ has density $f^{(X_1, X_2)}(x, y)$ then the marginal density of X_1 is

$$f^{X_1}(x) = \int_{\mathbb{R}} f^{(X_1, X_2)}(x, y) dy.$$

- If X is discrete and takes values (x_i, y_i) with probabilities $p^{(X_1, X_2)}(x_i, y_i)$, then the marginal law of X_1 is

$$p^{X_1}(x_i) = \sum_k p^{(X_1, X_2)}(x_i, y_k).$$

Random vectors

A concept that is specific to the multivariate case is the covariance between two rv

$$\mathrm{Cov}[X_1, X_2] = \mathbb{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]$$

(interpretation).

Properties:

- $\mathrm{Cov}[X_1, X_2] = \mathrm{Cov}[X_2, X_1]$.
- $\mathrm{Cov}[X_1, X_2] = \mathbb{E}[X_1 X_2] - \mu_{X_1}\mu_{X_2}$.
  (as for the variance: bad for interpretation, good for computations). (Exercise).
- $|\mathrm{Cov}[X_1, X_2]| \leq \sqrt{\mathrm{Var}[X_1]}\sqrt{\mathrm{Var}[X_2]}$ (from Cauchy-Schwarz).
  Hence, the correlation $\rho = \mathrm{Corr}[X_1, X_2] := \frac{\mathrm{Cov}[X_1, X_2]}{\sqrt{\mathrm{Var}[X_1]}\sqrt{\mathrm{Var}[X_2]}}$ satisfies $-1 \leq \rho \leq 1$.

Random vectors

In the general case (k-variate case), the location and dispersion parameters of some random vector

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

are collected in the mean vector

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_k] \end{pmatrix}$$

Random vectors

and in the variance-covariance matrix

$$\operatorname{Var}[X] = \Big(\operatorname{Cov}[X_i, X_j]\Big)_{i,j=1,\dots,k}$$

$$= \begin{pmatrix} \operatorname{Var}[X_1] & \dots & \operatorname{Cov}[X_1, X_k] \\ \vdots & \ddots & \vdots \\ \operatorname{Cov}[X_1, X_k] & \dots & \operatorname{Var}[X_k] \end{pmatrix},$$

respectively.

Conditional probability

$\mathcal{E}$: dice, $\Omega = \{1, 2, \ldots, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}[A] = \frac{\#A}{\#\Omega}$.

Consider the events $A = \{6\}$ and $B = \{4, 5, 6\}$.

Assume that you know that B occurred.
What is then the probability that A occurs?

$\rightsquigarrow$ we will write $\mathbb{P}[A|B] = \frac{1}{3}$.

Before defining $\mathbb{P}[A|B]$, let us consider another example...

Conditional probability

E: dice, $\Omega = \{1, 2, \ldots, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}[A] = \frac{\#A}{\#\Omega}$.

Consider the events $A =$“obtain an even result" and $B = \{4, 5, 6\}$.
What would be the value of $\mathbb{P}[A|B]$ here?

(discussion). $\rightsquigarrow \mathbb{P}[A|B] = \frac{2}{3} = \frac{\#(A \cap B)}{\#B} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$.

This leads to the following definition:
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $A, B \in \mathcal{A}$, with $\mathbb{P}[B] \neq 0$.

$\rightsquigarrow$ We then define $\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$. (remark: $\mathbb{P}[A|\Omega] = \mathbb{P}[A]$).

Conditional probability

Two important formulas:

(A) The total probability formula.

(B) The Bayes formula.

Conditional probability

Two important formulas:

(A) The total probability formula:
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.
Let $A \in \mathcal{A}$. Let $B_1, B_2, \ldots, B_n \in \mathcal{A}$, such that $\{B_1, B_2, \ldots, B_n\}$ is a partition of Ω with $\mathbb{P}[B_i] \neq 0$ for all i.

Then $\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A|B_i]\mathbb{P}[B_i]$ (proof).

Example:

3 types of machines in a factory (M_1, M_2, M_3). Out of 100 machines, there are 50 M_1, 30 M_2, and 20 M_3. Products made with M_1, M_2, M_3 are “good" with probability $0.7, 0.8, 0.9$, respectively. Then the probability that *some* product is good is
$\mathbb{P}[\text{good}] = \mathbb{P}[\text{good}|M_1]\mathbb{P}[M_1] + \mathbb{P}[\text{good}|M_2]\mathbb{P}[M_2] + \mathbb{P}[\text{good}|M_3]\mathbb{P}[M_3]$
$= 0.7 \times \frac{50}{100} + 0.8 \times \frac{30}{100} + 0.9 \times \frac{20}{100} = 0.77.$

Conditional probability

(B) The Bayes formula:
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.
Let $A \in \mathcal{A}$, with $\mathbb{P}[A] \neq 0$. Let $B_1, B_2, \ldots, B_n \in \mathcal{A}$, such that $\{B_1, B_2, \ldots, B_n\}$ is a partition of Ω with $\mathbb{P}[B_i] \neq 0$ for all i.

Then

$$\mathbb{P}[B_j|A] = \frac{\mathbb{P}[A|B_j]\mathbb{P}[B_j]}{\sum_{i=1}^{n} \mathbb{P}[A|B_i]\mathbb{P}[B_i]}$$

(proof).

Example:

In the same factory as above, the probability that some good product was made by M_1 is

$$\mathbb{P}[M_1|\mathrm{good}] = \frac{\mathbb{P}[\mathrm{good}|M_1]\mathbb{P}[M_1]}{\mathbb{P}[\mathrm{good}|M_1]\mathbb{P}[M_1] + \mathbb{P}[\mathrm{good}|M_2]\mathbb{P}[M_2] + \mathbb{P}[\mathrm{good}|M_3]\mathbb{P}[M_3]}$$

$$= \frac{0.7 \times \frac{50}{100}}{0.7 \times \frac{50}{100} + 0.8 \times \frac{30}{100} + 0.9 \times \frac{20}{100}} \approx 0.45.$$

Conditional probability

Very important example: assume a drug test returns "+" if drug was taken in 99% of time and "-" if drug was not taken in 99% of time. Is this test reliable? IT DEPENDS!!

Assume on 0.5% of people take the drug. If test says yes, what is actual probability the person took the drug?

$$
\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} \\
&= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} = 0.33.
\end{aligned}
$$