Markov Processes and Martingales

Károly Simon

Department of Stochastics
Institute of Mathematics
Technical University of Budapest
www.math.bme.hu/~simonk

B File
Interlude: how to compute the conditional expectation?

Review of a simple situation

Let X, Y be r.v. on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume they have joint density $f_{X,Y}(x,y)$. Then to compute $\mathbb{E}[X|Y]$ as first we determine the marginal and then the conditional densities

$$f_Y(y) := \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \text{ and } f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Let $g(y) := \mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)dx$. Then we get

$$\mathbb{E}[X|Y] = g(Y). \tag{1}$$

Independence Lemma

Lemma 1.1 (Independence Lemma)

Let $\mathbf{X} = (X_1, \dots, X_k)$ and $\mathbf{Y} := (Y_1, \dots, Y_\ell)$, where $X_1, \dots, X_k, Y_1, \dots, Y_\ell$ are r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. We assume that

- *$X_1, \dots, X_k \in \mathcal{G}$*
- *$Y_1, \dots, Y_\ell$ are independent of $\mathcal{G}$.*

Let ϕ be a bounded Borel function. Let $f_\phi : \mathbb{R}^k \to \mathbb{R}$, $f_\phi(x_1, \dots, x_k) := \mathbb{E}[\phi(x_1, \dots, x_k, \mathbf{Y})]$. Then

$$\mathbb{E}[\phi(\mathbf{X}, \mathbf{Y})|\mathcal{G}] = f_\phi(\mathbf{X}). \tag{2}$$

Independence Lemma (cont.)

The proof of the Lemma We follow the line of the proof in Resnik's book We present the main steps of the proof here for the case $k = \ell = 1$. It is a homework to fill the gaps.
Step 1. Let $K, L \in \mathcal{R}$ (that is K, L are Borel subsets of $\mathbb{R}$). Let $\phi := \mathbb{1}_J$ where $J = K \times L$. Then we say that J is a measurable rectangle.

$$\begin{aligned}\mathbb{E}[\phi(\mathbf{X}, \mathbf{Y})|\mathcal{G}] &= \mathbb{P}(X \in K, Y \in L|\mathcal{G}) \\ &= \mathbb{1}\{X \in K\}\mathbb{P}(Y \in L|\mathcal{G}) \\ &= \mathbb{1}\{X \in K\}\mathbb{P}(Y \in L) = f_{\mathbb{1}_{K \times L}}(X).\end{aligned}$$

Independence Lemma (cont.)

Step 2. We write RECTS for the family of measurable rectangles (like J above). Let

$$\mathcal{C} := \left\{J \in \mathcal{R}^2 : (2) \text{ holds for } \phi = \mathbb{1}_J\right\}.$$

Then $\mathrm{RECTS} \subset \mathcal{C}$. Now we verify that $\mathcal{C}$ is a λ-system. (To recall the definition, see the File "Some basic facts from probability theory" Definition **??**.) That is

(a) $\mathbb{R}^2 \in \mathcal{C}$. This holds because $\mathbb{R}^2 \in \mathrm{RECTS}$.
(b) $J \in \mathcal{C}$ implies $J^c \in \mathcal{C}$. This is so because

$$\begin{aligned}\mathbb{P}((X,Y) \in J^c|\mathcal{G}) &= 1 - \mathbb{P}((X,Y) \in J|\mathcal{G}) \\ &1 - f_{\mathbb{1}_J}(X) = f_{\mathbb{1}_{J^c}}(X).\end{aligned}$$

Independence Lemma (cont.)

(c) If $A_n \in \mathcal{C}$ and A_n are disjoint then $\bigcup_n A_n \in \mathcal{C}$.

We do not prove (c) here. By definition, (a), (b) and (c) implies that

- $\mathcal{C}$ is a λ-system and
- $\mathcal{C} \supset \mathrm{RECTS}$.

Using that RECTS is a π-system it follows from File "Some basic facts from probability theory" Theorem **??** that

$$\mathcal{C} \supset \sigma(\mathrm{RECTS}) = \mathcal{R}^2. \tag{3}$$

Independence Lemma (cont.)

So, we have indicated that (2) holds when ϕ is an indicator function of Borel subsets of the plane.
Step 3. We could prove that (2) also holds when ϕ is a simple function. We say that a Borel function ϕ is a simple function if its range is finite. That is if there exist a k and a partition $J_1, \dots, J_k$ of $\mathbb{R}^2$, $J_k \in \mathcal{R}$ and real numbers $c_1, \dots, c_k$ such that

$$\phi = \sum_{i=1}^{k} c_i \mathbb{1}_{J_i}. \tag{4}$$

Independence Lemma (cont.)

Step 4. Then we represent $\phi = \phi^+ - \phi^-$ and we can find sequences of simple functions $\{\phi_n^+\}$ and $\{\phi_n^-\}$ such that

$$\phi_n^+ \uparrow \phi^+ \text{ and } \phi_n^- \uparrow \phi^-.$$

Then using Conditional Monotone Convergence Theorem (see File "Some basic facts from probability theory" slide #??) we conclude the proof. ■

Monotone Class Theorem

We could have used in the previous proof the so called Monotone Class Theorem (for the proof see [6, p. 235])

Theorem 1.2 (Monotone Class Theorem)

Let $\mathcal{A}$ be a π-system with $\Omega \in \mathcal{A}$ and let $\mathcal{H}$ be a family of real valued function defined on Ω with the following three properties:

(a) $\mathbb{1}_A \in \mathcal{H}$ *whenever* $A \in \mathcal{A}$.

(b) $f, g \in \mathcal{H} \Longrightarrow f + g \in \mathcal{H}$ *further,* $\forall c \in \mathbb{R}$: $c \cdot f \in \mathcal{H}$

Monotone Class Theorem cont.

Theorem 1.3 (Monotone Class Theorem cont)

(c) *If* $f_n \in \mathcal{H}$ *satisfying*
- $f_n \geq 0$ *and*
- $f_n \uparrow f$

then $f \in \mathcal{H}$

The $\mathcal{H}$ contains all bounded functions measurable w.r.t. $\sigma(\mathcal{A})$.

Review

Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and $A \in \mathcal{F}$. Acording to the definition of the conditional probability w.r.t. sub-σ-algebra:

$$\mathbb{P}(A|\mathcal{G}) := \mathbb{E}[\mathbb{1}_A|\mathcal{G}]. \tag{5}$$

This is not a probability in the worst case but it is a probability (for almost all ω) under mild assumptions. Our aim here is to get a better understanding of this notion. We start with a very easy example from Billingley's book.

Review cont

Example 2.1

Consider a Poisson process $N(t) \sim \mathrm{Poisson}(\lambda)$ which is, say the number of telephone calls to a call center during time interval $[0, t]$.
(Remember: this means that the number of calls until time t_0 has distribution $\mathrm{Poi}(\lambda \cdot t_0)$.) Fix $0 < s < t$.
$A := \{N(s) = 0\}$ and $B_i := \{N(t) = 0\}$, $i = 0, 1, 2, \dots$.
Then $\{B_i\}_{i=0}^{\infty}$ is a partition of Ω. So, it generates a σ-algebra which we call $\mathcal{G}$.

Review cont.

Example 2.2 (Example cont.)

It is immediate that

$$\mathbb{P}(A|B_i) = \frac{\mathbb{P}(N(s)=0) \cdot \mathbb{P}(N(t)-N(s)=i)}{\mathbb{P}(N(t)=i)} = \left(1 - \frac{s}{t}\right)^i. \tag{6}$$

Then the random variable

$$\mathbb{P}(N(s)=0|\mathcal{G}) = \left(1 - \frac{s}{t}\right)^{N(t)} \tag{7}$$

Review cont.

This is so, because we learned that in general

Lemma 2.3

Let $\Omega_1, \Omega_2, \dots$ be a partition of Ω and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then

$$\mathbb{E}[X|\mathcal{G}](\omega) = \frac{\mathbb{E}[X;\Omega_i](\omega)}{\mathbb{P}(\Omega_i)} \tag{8}$$

If we apply Lemma 2.3 with $X = \mathbb{1}_A$:

$$\mathbb{P}(A|\mathcal{G})(\omega) = \mathbb{P}(A|\Omega_i) \text{ if } \omega \in \Omega_i. \tag{9}$$

Review cont. (cont.)

Hence

$$\mathbb{P}(A|\Omega_i) \cdot \mathbb{P}(\Omega_i) = \mathbb{P}(A \cap \Omega_i)$$

This implies that the following assertions hold:

(i) $\mathbb{P}(A|\mathcal{G}) \in \mathcal{G}$.

(ii) $\mathbb{P}(A|\mathcal{G}) \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ and

(iii)

$$\int_G \mathbb{P}(A|\mathcal{G})\, d\mathbb{P} = \mathbb{P}(A \cap G) \text{ for all } G \in \mathcal{G}. \tag{10}$$

Review cont. (cont.)

(i)-(iii) above could serve as an alternative definition of the conditional probability $\mathbb{P}(A|\mathcal{G})$.
The proof of the following theorem is immediate from the properties of the conditional expectation.

Review cont. (cont.)

Theorem 2.4 (Basic properties)

Given $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.

(a) $\mathbb{P}(\emptyset|\mathcal{G}) = 0$ *and* $\mathbb{P}(\Omega|\mathcal{G}) = 1$ *a.s.*

(b) $0 \le \mathbb{P}(A|\mathcal{G}) \le 1$ *a.s. for all* $A \in \mathcal{F}$.

(c) *Let* $A = \bigsqcup_{n=1}^{\infty} A_n$ *(recall:* $\sqcup$ *means disjoint union) and* $A_n \in \mathcal{F}$ *then*

$$\mathbb{P}(A|\mathcal{G}) = \sum_{n=1}^{\infty} \mathbb{P}(A_n|\mathcal{G}) \quad a.s. \tag{11}$$

Review cont. (cont.)

Remark 2.5 (We have a problem)

$\mathbb{E}[\mathbb{1}_A|\mathcal{G}] = \mathbb{P}(A|\mathcal{G})$ is defined on $\Omega_A \subset \Omega$, $\mathbb{P}(\Omega_A) = 1$.
So, for every $A \in \mathcal{F}$ there is a set Z_A of zero measure where $\mathbb{P}(A|\mathcal{G})$ is not defined. In order to satisfy (11) we need to insure that for all countable collections simultaneously the exceptional set is still a set of zero measure.

That is why we need to study this problem with more detail. But at the end everything will be alright, at least in the "nice" cases.

R.C.D.

Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the measurable space $(S, \mathcal{S})$. Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ be a measurable map. In other words: X is an S-valued r.v.. Further, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.

R.C.D. (cont.)

Definition 3.1 (Regular conditional Distribution)

We say that $\mu : \Omega \times \mathcal{S} \to [0, 1]$ is a Regular conditional Distribution for X given $\mathcal{G}$ if

(a) For any fixed $A \in \mathcal{S}$ the map $\omega \mapsto \mu(\omega, A)$ is a version of $\mathbb{P}(X \in A|\mathcal{G})$.

(b) For almost every fixed ω, $\mu(\omega, \bullet)$ is a probability measure on $(S, \mathcal{S})$.

If $S = \Omega$ and X is the identity map $X(\omega) = \omega$ then we say that μ is a regular conditional probability.

R.C.D. (cont.)

Example 3.2

Assume that (X, Y) has density $f(x, y) > 0$. Let

$$\mu(y, A) := \int_A f(x, y)dx \Big/ \int_{-\infty}^{\infty} f(x, y)dx.$$

Then $\mu(Y(\omega), A)$ is an r.c.d. for X given $\sigma(Y)$.

R.C.D. (cont.)

Theorem 3.3

Let $\mu(\omega, A)$ be a r.c.d. for X given $\mathcal{F}$ and let $f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{R})$ be measurable. (This means that $f : S \to \mathbb{R}$ and for every Borel set $B \in \mathcal{R}$ we have $f^{-1}(B) \in \mathcal{S}$.) Further, we assume that $\mathbb{E}[|f(X)|] < \infty$. Then

$$\mathbb{E}[f(X)|\mathcal{F}] = \int f(x) \cdot \mu(\omega, dx). \quad (12)$$

We say that a space is nice if there is an injective map $\varphi : S \to \mathbb{R}$ such that both φ and φ^{-1} are measurable.

R.C.D. (cont.)

Theorem 3.4

Let S be a complete separable metric space and $\mathcal{S}$ be the Borel sets on S. Then $(S, \mathcal{S})$ is nice.

Theorem 3.5

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ Further, let $(S, \mathcal{S})$ be a measurable space which is nice. Then any S-valued r.v. X admits a regular conditional distribution given $\mathcal{G}$.

The proof follows [13, Proposition 7.14].

Proof of Theorem. 3.5 for $S = \mathbb{R}$

First we assume that $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$.
For a rational number $q \in \mathbb{Q}$ we define the r.v.

$$P^q(\omega) := \mathbb{P}(X \leq q|\mathcal{G})(\omega).$$

By throwing away countably many null sets we may suppose that

$$P^q(\omega) \leq P^r(\omega), \quad \forall q \leq r, \ q, r \in \mathbb{Q} \text{ and } \forall \omega \quad (13)$$

and

$$0 = \lim_{q \to -\infty} P^q(\omega), \quad \lim_{q \to \infty} P^q(\omega) = 1, \quad \forall \omega.$$

Proof of Theorem. 3.5 for $S = \mathbb{R}$ cont.

For an $x \in \mathbb{R}$ let

$$F(\omega, x) := \lim_{q \in \mathbb{Q}, q > x} P^q(\omega). \quad (14)$$

Fix an arbitrary ω. Then $\forall \omega$ the function $x \mapsto F(\omega, x)$:

- is right continuous,
- non-decreasing,
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Fix an arbitrary x. Then $\omega \mapsto F(\omega, x)$ is r.v. that is measurable (as an infimum of measurable functions).

Proof of Theorem. 3.5 for $S = \mathbb{R}$ cont.

Combining the first comment on the previous slide with Theorem **??** from File "Some basic facts form probability" we obtain that for

$$\forall \omega, \exists \text{ a probability measure } \mu_{X|\mathcal{G}}(\omega, \bullet). \quad (15)$$

satisfying

$$\mu_{X|\mathcal{G}}(\omega, (-\infty, x]) = F(\omega, x), \quad \forall \omega, \forall x. \quad (16)$$

Proof of Theorem. 3.5 for $S = \mathbb{R}$ cont.

Now we write $\mathcal{L}$ for the family of all Borel sets $B \in \mathcal{R}$ satisfying the following two conditions:

(i) $\omega \mapsto \mu_{X|\mathcal{G}}(\omega, B)$ is a r.v..

(ii) $\mu_{X|\mathcal{G}}(\omega, B)$ is a version of $\mathbb{P}(X \in B|\mathcal{G})(\omega)$.

Now the strategy is as follows:

(α) It is not hard to prove that $\mathcal{L}$ is λ-system (we omit this proof).

(β) We prove that $\mathcal{L}$ contains a π-system $\mathcal{P}$ such that $\mathcal{R} = \sigma(\mathcal{P})$.

Proof of Theorem. 3.5 cont.

Assuming that (α) and (β) hold the proof of Theorem 3.5 is complete in the case of $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$. Namely, using ($\alpha$) and ($\beta$), it follows from Theorem **??** from File "Some basic facts form probability" that $\mathcal{R} \subset \mathcal{L}$. Part (a) of Definition 3.1 will be verified by this. Part (b) is immediate from (15).

Proof of Theorem. 3.5 for $S = \mathbb{R}$ cont.

To complete part (α) on the previous slide: a convenient π-system is

$$\mathcal{P} := \{(-\infty, x] : x \in \mathbb{R}\}.$$

(i) holds: $mu_{X|\mathcal{G}}(\omega, B) = F(\omega, x)$ by definition and then we use the second comment on slide #28.
(ii) holds: We need to verify that

$$F(\omega, x) = \mathbb{P}(X \leq x|\mathcal{G})(\omega). \quad (17)$$

Recall that F was defined in (14).

Proof of Theorem. 3.5 for $S = \mathbb{R}$ cont.

(17) follows from:

$$F(\omega, x) = \inf_{q>x} P^q(\omega) = \lim_{q \downarrow x} P^q(\omega)$$
$$= \lim_{q \downarrow x} \mathbb{P}(X \leq q | \mathcal{G})(\omega) = \mathbb{P}(X \leq x | \mathcal{G})(\omega), \text{ for a.a. } \omega.$$

In the last step we used the Dominated Convergence Theorem. This verifies part (β) on # slide 30. Which completes the proof of the Theorem in the case when $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$.

Proof of Theorem. 3.5 in the general case

Now we turn to the general case when X is an S-valued random variable:

$$X : \Omega \to S.$$

That is $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ is measurable. Using that $(S, \mathcal{S})$ is a nice space, there exists an injective map $\rho : S \to \mathbb{R}$ such that both ρ and ρ^{-1} are r.v.. Then the composition

$$Y := \rho \circ X : \Omega \to \mathbb{R}$$

is also a r.v. for which we consider the corresponding r.c.d.:

$$\mu_{Y|\mathcal{G}}(\omega, A) := \mathbb{P}(Y \in A | \mathcal{G}), \quad A \in \mathcal{R}.$$

Proof of Theorem. 3.5 in the general case cont.

Now we can define the r.c.d for X:

$$\mu_{X|\mathcal{G}}(\omega, B) := \mu_{Y|\mathcal{G}}(\omega, \rho(B)).$$

Then it is not hard to prove that $\mu_{X|\mathcal{G}}(\omega, B)$ satisfies the conditions (a) and (b) of Definition 3.1.

Conditional Characteristic Function

Notation for the next slides:

- $(\Omega, \mathcal{F}, \mathbb{P})$ is the given probabiliyt space,
- $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$,
- $\mathbf{X} : \Omega \to \mathbb{R}^n$ is a given vector-valued r.v.,
- $\mu_{X|\mathcal{G}} : \Omega \times \mathcal{R}^n \to [0, 1]$ be the regular conditional distribution of X given $\mathcal{G}$.

Definition 3.6 (Regular conditional cdf)

$$F(\omega, x) := \mu_{\mathbf{X}|\mathcal{G}}(\omega, \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \leq_n \mathbf{x}\}) \quad \mathbf{x} \in \mathbb{R}^n$$

Conditional Characteristic Function cont.

Definition 3.7

$f_{\mathbf{X}|\mathcal{G}} : \Omega \times \mathbb{R}^n \to [0, \infty)$ is the conditional density function of X given $\mathcal{G}$ if

- $\mathbf{x} \mapsto f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x})$ is Borel measurable,
- $\omega \mapsto f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x})$ is $\mathcal{G}$-measurable for every $\mathbf{x} \in \mathbb{R}^n$,
- $\int_B f_{\mathbf{X}|\mathcal{G}(\omega,\mathbf{x})} dx = \mu_{\mathbf{X}|\mathcal{G}}(\omega, B)$.

Conditional Characteristic Function cont.

Definition 3.8 (Conditional characteristic function)

The conditional characteristic function of X given $\mathcal{G}$, $\varphi_{X|\mathcal{G}} : \Omega \times \mathbb{R}^n \to \mathbb{C}$ is

$$\varphi_{X|\mathcal{G}}(\omega, \mathbf{t}) := \int_{\mathbb{R}^n} e^{i\mathbf{t}\cdot\mathbf{x}} d\mu_{\mathbf{X}|\mathcal{G}}(\omega, dx), \quad \mathbf{t} \in \mathbb{R}^n,$$

where $\mathbf{t} \cdot \mathbf{x}$ above means the scalar product of $\mathbf{t}$ and $\mathbf{x}$.

Conditional Characteristic Function cont.

Theorem 3.9

The following two assertions are equivalent

(a) *The function*

$$\omega \mapsto \varphi_{X|\mathcal{G}}(\omega, \mathbf{t})$$

is constant for $\mathbb{P}$-almost all ω. This constant is denoted by $\varphi(t)$.

(b) *$\sigma(\mathbf{X})$ is independent of $\mathcal{G}$.*

Proof of Theorem 3.9 (a)⇒ (b)

By Theorem 3.3

$$\varphi_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{t}) = \mathbb{E}\left[e^{i\mathbf{t}\cdot\mathbf{X}} | \mathcal{G}\right](\omega) \quad a.s..$$

here we assume that this is a constant (in ω) denoted by $\varphi(t)$. Multiply both sides with a r.v. Y which is bounded (real-valued) and $\mathcal{G}$-measurable, we get

$$\varphi(t) \cdot \mathbb{E}[Y] = \mathbb{E}\left[Y e^{i\mathbf{t}\cdot\mathbf{X}}.\right]$$

For $Y = 1$ we get $\varphi(t) = \mathbb{E}\left[e^{i\mathbf{t}\cdot\mathbf{X}}\right]$. Substitute this to the previous equality to get

(18) $$\mathbb{E}\left[Y e^{i\mathbf{t}\cdot\mathbf{X}}\right] = \mathbb{E}[Y] \cdot \mathbb{E}\left[e^{i\mathbf{t}\mathbf{X}}\right]$$

Proof of Theorem 3.9 (a)⇒ (b) cont

holds for all $\mathcal{G}$-measurable bounded Y and $\mathbf{t} \in \mathbb{R}^n$. So, (18) holds for all r.v.

$$Y = e^{i\mathbf{s} \cdot Z},$$

where Z is any $\mathcal{G}$-measurable $\mathbb{R}^n$-valued r.v. and $\mathbf{s} \in \mathbb{R}^n$. So from (18)

$$\mathbb{E}\left[e^{i\mathbf{t} \cdot \mathbf{X} + i\mathbf{s} \cdot \mathbf{Z}}\right] = \mathbb{E}\left[e^{i\mathbf{t}\mathbf{X}}\right] \cdot \mathbb{E}\left[e^{i\mathbf{s}\mathbf{Z}}\right], \quad \forall \mathbf{s}, \mathbf{t} \in \mathbb{R}^n.$$

Using this and the assertion of Homework # **??** we get that X and Z are independent.

Proof of Theorem 3.9 (a)⇒ (b) cont

On the other hand, this implies that X and $\mathcal{G}$ are independent since Z was an arbitrary $\mathcal{G}$-measurable r.v..

Proof of Theorem 3.9 cont (b)⇒ (a)

This is immediate from the "irrelevance of independent information" property of the conditional expectation.

The continuous case

Theorem 3.10

On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *we are give a random vector*

$$\mathbf{Z} = (\underbrace{X_1, \dots, X_k}_{\mathbf{X}}, \underbrace{Y_1, \dots, Y_\ell}_{\mathbf{Y}}) = (\mathbf{X}, \mathbf{Y}).$$

We assume that $\mathbf{Z}$ *admits a density* $f_{\mathbf{Z}} : \mathbb{R}^{k+\ell} \to [0, \infty)$.
Let $\mathcal{G} := \sigma(\mathbf{Y})$.
Then there exists a conditional density $f_{\mathbf{X}|\mathcal{G}} : \mathbb{R}^k \to [0, \infty)$ *of* $\mathbf{X}$ *given* $\mathcal{G}$ *by the formula:*

The continuous case cont.

Theorem 3.10 cont.

(19)

$$f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x}) = \begin{cases} \frac{f_{\mathbf{Z}}(\mathbf{x}, \mathbf{Y}(\omega))}{\int_{\mathbb{R}^k} f_{\mathbf{Z}}(\mathbf{x}, \mathbf{Y}(\omega)) d\mathbf{x}}, & \text{if } \int_{\mathbb{R}^\ell} f(\mathbf{x}, \mathbf{Y}(\omega)) d\mathbf{x} > 0; \\ f_0(\mathbf{x}), & \text{otherwise,} \end{cases}$$

where $f_0 : \mathbb{R}^k \to [0, \infty)$ is an arbitrary density function.

The continuous case cont.

proof

We have to check that for all $A \in \mathcal{R}^k$,

$$\int_A f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x}) d\mu_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x})$$

is a version of $\mathbb{P}(\mathbf{X} \in A|\mathcal{G})(\omega)$. This follows if (20)

$$\mathbb{E}\left[\mathbb{1}_{\mathbf{Y} \in B}(\omega) \cdot \int_A f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x}) d\mathbf{x}\right] = \mathbb{E}\left[\mathbb{1}_{\mathbf{Y} \in B}(\omega) \cdot \mathbb{1}_{\mathbf{X} \in A}(\omega)\right],$$

holds for $\forall A \in \mathcal{R}^k$ and $B \in \mathcal{R}^\ell$. We verify this:

The continuous case cont.

proof cont.

$$\mathbb{E}\left[\mathbb{1}_{\mathbf{Y} \in B}(\omega) \cdot \int_A f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x}) d\mathbf{x}\right] = \int_A \mathbb{E}\left[\mathbb{1}_{\mathbf{Y} \in B}(\omega) \cdot f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x})\right] d\mathbf{x}$$

Observe that by definition of $f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x})$ and change of variables formula:

$$\mathbb{E}\left[\mathbb{1}_{\mathbf{Y} \in B}(\omega) \cdot f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x})\right] = \int_B f_{\mathbf{Z}}(x, y) d\mathbf{y}.$$

The continuous case cont.

proof cont.

So,

$$\mathbb{E}\left[\mathbb{1}_{\mathbf{Y} \in B}(\omega) \cdot \int_A f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x}) d\mathbf{x}\right] = \int_A \int_B f_{\mathbf{Z}}(x, y) d\mathbf{y} d\mathbf{x} = \mathbb{P}(X \in A; Y \in B.) \blacksquare$$

Definition 4.1 (Normal distribution (on $\mathbb{R}$))

Let $\mu \in \mathbb{R}$ and $\sigma > 0$. Random variable The r.v. X has normal (or Gaussian) distribution with parameters (μ, σ^2), if its density function:

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

Then we write $X \sim \mathcal{N}(\mu, \sigma^2)$. If $\mu = 0$ and $\sigma = 1$, then we get the standard normal distribution $\mathcal{N}(0,1)$. Let us use the following notation:

$$(21) \quad \varphi(x) := \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}, \quad \Phi(x) := \int_{-\infty}^{x} \varphi(y) dy.$$

Some properties

$X \sim \mathcal{N}(\mu, \sigma^2)$ and $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$. Then

(a) $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$.

(b) $F_X(x) = \mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

(c) $X_1 + X_2 = \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

(d) $X \sim \mathcal{N}(0,1)$, then

(22)

$$\frac{1}{\sqrt{2\pi}} \cdot \left(x^{-1} - x^{-3}\right) \cdot e^{-x^2/2} \leq \mathbb{P}\left(X \geq x\right) \leq \frac{1}{\sqrt{2\pi}} \cdot x^{-1} \cdot e^{-x^2/2}$$

(e) Fix a $p \in (0,1)$. Let $Y_n \sim \mathrm{Bin}(n,p)$, $a < b$, then

$$(23) \quad \lim_{n\to\infty} \mathbb{P}\left(a < \frac{Y_n - np}{\sqrt{np(1-p)}} < b\right) = \Phi(b) - \Phi(a).$$

Multivariate normal distribution

Definition 4.2

A random vector $\mathbf{X} \in \mathbb{R}^d$ is non-degenerate multivariate normal or jointly Gaussian, if the density function $f(\mathbf{x})$ of $\mathbf{X}$

$$(24) \quad f(\mathbf{x}) = \frac{\sqrt{\det(A)}}{(2\pi)^{d/2}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \cdot A \cdot (\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{R}^d,$$

or

(25)

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \cdot \det(\Sigma)}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \cdot \Sigma^{-1} \cdot (\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{R}^d,$$

Multivariate normal distribution cont.

where A and $\boldsymbol{\mu}$ and Σ satisfy:

- A is a $d \times d$ matrix which is
  1. symmetric and
  2. positive definit. Further,
- $\boldsymbol{\mu} \in \mathbb{R}^d$ is a fixed vector

The meaning of matrix A is as follows:

$$\left(A^{-1}\right)_{ij} = Cov(X_i, X_j) = \mathbb{E}\left[(X_i - \mathbb{E}[X_i]) \cdot (X_j - \mathbb{E}[X_j]\right],$$

where $\mathbf{X} = (X_1, \dots, X_d)$. The $d \times d$ matrix $\Sigma = A^{-1}$ with

$$\Sigma_{ij} := Cov(X_i, X_j)$$

is called covariance matrix. We write $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$

Multivariate normal distribution cont.

Definition 4.3

Let $\mathbf{X}$ be as above. Let $\lambda_1, \dots, \lambda_d$ be the eigenvalues of A, and $\mathbf{v}_1, \dots, \mathbf{v}_d$ be the ortonormal basis of $\mathbb{R}^d$ with the appropriate eigenvectors. Let us define diagonal matrix

$$D := \mathrm{diag}(\lambda_1, \dots, \lambda_d).$$

We define the orthogonal $d \times d$ matrix $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_d \end{bmatrix}$ from the eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_d$ as column vectors.

Multivariate normal distribution cont.

Lemma 4.4

Let $\mathbf{X}$ be as above. Then

$$(26) \quad \mathbf{X} = P \cdot D^{-1/2} \cdot (Y_1, \dots, Y_d) + \boldsymbol{\mu},$$

where $Y_i \sim \mathcal{N}(0,1), i = 1, \dots, d$ and they are independent. In this case we call $\mathbf{Y}$ standard multivariate normal vector.

That is the random vector $\mathbf{Y}$ is presented as the affine transform of independent standard normal r.v.. See [1, chapters 6 and 7].

Converse of the previous lemma

Lemma 4.5

Let $\mathbf{Y}$ be a standard multivariate normal vector in $\mathbb{R}^n$. Let B be a non-singular $d \times d$ matrix and $\boldsymbol{\mu} \in \mathbb{R}^n$. Let

$$\mathbf{X} := B \cdot \mathbf{Y} + \boldsymbol{\mu}$$

Then $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, A \cdot A^T)$.

An equivalent definition

Lemma 4.6

The random vector $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$ has a multivariate normal distribution if for all $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ the following holds:

$a_1 X_1 + \dots + a_n X_n$ has univariate normal distribution.

The proof are available in [3]

The bivariate Case

Assume that $\mathbf{Z} = (X, Y)$ has a bivariate normal distribution. Let

$$\mu_X,\ \mu_Y,\ \sigma_X,\ \sigma_Y$$

be the expectation and standard deviation of X and Y respectively. Further, recall the definitions of covariance and correlation:

$$\mathrm{cov}(X,Y) := \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]$$

The bivariate Case cont.

The correlation of (X, Y) is:

$$(27)\quad \rho := \rho_{X,Y} := \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)} = \frac{\mathbb{E}\left[(X-\mu_X)][(Y-\mu_Y)\right]}{\sigma(X)\sigma(Y)}$$

The mean vector and the variance-covariance matrix is:

$$\boldsymbol{\mu} := \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

The bivariate Case cont.

Let

$$Q(x,y) := \frac{1}{1-\rho^2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)$$

So, the density is

$$f_{\mathbf{Z}}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}Q(x,y)\right).$$

The bivariate Case cont.

Consider the marginal densities:

$$f_X := \frac{1}{\sigma_X\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-\mu_X)^2}{2\sigma^2}} \text{ and } f_Y := \frac{1}{\sigma_Y\cdot\sqrt{2\pi}}\cdot e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}.$$

Observe that whenever X and Y are uncorrelated, that is $\rho = 0$ then

$$f_{\mathbf{Z}} = f_X \cdot f_Y.$$

This means that X and Y are independent. In a similar way one can prove the same in higher dimension:

Uncorrelated ⇒ independent for Gussian

Theorem 4.7

Let $\mathbf{X} = (X_1, \dots, X_n)$ be multivariate normal vector. Assume that $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$. Then $X_1, \dots, X_n$ are independent.

Multivariate normal distribution cont.

A more general theorem in this direction is:

Theorem 4.8

Let $\mathbf{X} = (X_1, \dots, X_n)$ be random vector such that the marginal distributions (the distributions of the component vectors X_i) are

- *normal and*
- *independent*

Then X has a multivariate normal distribution.

CF and MGF

Theorem 4.9

Let $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Then The characteristic function is

$$(28)\quad \varphi_{\mathbf{X}}(\mathbf{t}) := \mathbb{E}\left[\exp(i\mathbf{t}^T\cdot\mathbf{X})\right] = \exp\left(i\boldsymbol{\mu}^T\mathbf{t} - \tfrac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\right)$$

The moment generating function is

$$(29)\quad M_{\mathbf{X}}(\mathbf{t}) := \mathbb{E}\left[\exp(\mathbf{t}^T\cdot\mathbf{X})\right] = \exp\left(i\boldsymbol{\mu}^T\cdot\mathbf{t} - \tfrac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\right).$$

Conditioning normals

Given the multivariate normal vector

$$\mathbf{Z} = (\underbrace{X_1, \dots, X_k}_{\mathbf{X}}, \underbrace{Y_1, \dots, Y_\ell}_{\mathbf{Y}}) = (\mathbf{X}, \mathbf{Y}).$$

with mean $\boldsymbol{\mu}$ and variance-covariance matrix Σ:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \mathbb{E}\left[\widetilde{\mathbf{Z}}\cdot\widetilde{\mathbf{Z}}^T\right] = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix},$$

where $\widetilde{\mathbf{Z}} := \mathbf{Z} - \boldsymbol{\mu}$ and for $\widetilde{\mathbf{X}} := \mathbf{X} - \boldsymbol{\mu}_X$, $\widetilde{\mathbf{Y}} := \mathbf{Y} - \boldsymbol{\mu}_Y$

$$\Sigma_{XX} = \mathbb{E}\left[\widetilde{\mathbf{X}}\cdot\widetilde{\mathbf{X}}^T\right] \qquad \Sigma_{XY} = \mathbb{E}\left[\widetilde{\mathbf{X}}\cdot\widetilde{\mathbf{Y}}^T\right]$$
$$\Sigma_{YX} = \mathbb{E}\left[\widetilde{\mathbf{Y}}\cdot\widetilde{\mathbf{X}}^T\right] \qquad \Sigma_{YY} = \mathbb{E}\left[\widetilde{\mathbf{Y}}\cdot\widetilde{\mathbf{Y}}^T\right]$$

Conditioning normals cont.

We may assume that Σ_{YY} is invertible. (Why?) Then for $A := \Sigma_{XY} \cdot \Sigma_{YY}^{-1}$ we have (simply by definitions) that

$$(30) \qquad \mathbb{E}\left[\left(\widetilde{\mathbf{X}} - A\widetilde{\mathbf{Y}}\right) \cdot \widetilde{\mathbf{Y}}^T\right] = 0.$$

By Theorem 4.7 this implies that $\widetilde{\mathbf{X}} - A\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}$ are independent. By Theorem 3.9 we have that the characteristic function of $\widetilde{\mathbf{X}} - A\widetilde{\mathbf{Y}}$ given $\mathcal{G} = \sigma(Y)$ is **deterministic** and is equal to (for every ω):

$$\varphi_{\widetilde{\mathbf{X}}-A\widetilde{\mathbf{Y}}}(\mathbf{t}) = \mathbb{E}\left[e^{i\mathbf{t}(\widetilde{\mathbf{X}}-A\widetilde{\mathbf{Y}})}|\mathcal{G}\right], \quad \forall \mathbf{t} \in \mathbb{R}^k.$$

Since $A\widetilde{\mathbf{Y}}$ is $\mathcal{G}$-measurable, we can pull out what is known and use (4.9):

Conditioning normals cont.

$$\mathbb{E}\left[e^{i\mathbf{t}\cdot\mathbf{X}}|\mathcal{G}\right] = e^{i\mathbf{t}\mu_X} e^{i\mathbf{t}A\widetilde{\mathbf{Y}}} e^{-\frac{1}{2}\mathbf{t}^T\widehat{\Sigma}\mathbf{t}} \text{ for } \mathbf{t} \in \mathbb{R}^k,$$

where

$$\widehat{\Sigma} = \mathbb{E}\left[(\mathbf{X} - A\widetilde{\mathbf{Y}})(\mathbf{X} - A\widetilde{\mathbf{Y}})^T\right].$$

Then an easy calculation shows that conditionally, $\mathbf{X}$ given $\mathcal{G}$ is multivariate normal $\mathcal{N}\left(\mu_{\mathbf{X}|\mathcal{G}}, \Sigma_{\mathbf{X}|\mathcal{G}}\right)$ with mean and variance-covariance matrix:

$\mu_{\mathbf{X}|\mathcal{G}} = \mu_Y + A(\mathbf{Y} - \mu_{\mathbf{Y}})$ and $\Sigma_{\mathbf{X}|\mathcal{G}} = \Sigma_{\mathbf{XX}} - \Sigma_{\mathbf{XY}}\Sigma_{YY}^{-1}\Sigma_{\mathbf{YX}}$.

[1] Mrton Balázs, Bálint Tóth
Lecture notes: Introductory probability (in Hungarian)
Click here.

[2] P. Billingsley
Probability and measure
Wiley, 1995

[3] M. Bolla, A. Kámli
Statisztikai kővetkeztetések elmélete
Typotex, 2005

[4] R. Durrett
Essentials of Stochastic Processes, Second edition
Springer, 2012. Click here

[5] R. Durrett
Probability: Theory with examples, 4th edition
Cambridge University Press, 2010.

[6] R. Durrett
Probability: Theory and Examples
Click here

[7] S. Karlin, H.M. Taylor
A first course in stochastic processes
Academic Press, New York, 1975

[8] S. Karlin, H.M. Taylor
Sztochasztikus Folyamatok
Gondolat, Budapest, 1985

[9] S. Karlin, H.M. Taylor
A second course in stochastic processes
, Academic Press, 1981

[10] P. Mattila Geometry of sets and measure in Euclidean spaces. Cambridge, 1995.

[11] S. I. Resnik
A probability Path
Birkhäuser 2005

[12] D. Williams
Probability with Martingales
Cambridge 2005

[13] G. Zitkovic *Theory of Probabilty I, Lecture 7*
Click here