Markov Processes and Martingales

Károly Simon

Department of Stochastics
Institute of Mathematics
Budapest University of Technology and Economics
www.math.bme.hu/~simonk

File B 2025

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

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

Lemma 1.1 (Independence Lemma)

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

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

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

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

Example. Let $X \in \mathcal{G}$, and let Y be independent of $\mathcal{G}$. Define

$$\varphi(X, Y) = XY.$$

Then,

$$f_\varphi(X) := \mathbb{E}[\varphi(X, Y)] = \mathbb{E}[XY] = X\mathbb{E}[Y].$$

$$\mathbb{E}[\varphi(X, Y) \mid \mathcal{G}] = \mathbb{E}[XY \mid \mathcal{G}] = X\mathbb{E}[Y \mid \mathcal{G}] = X\mathbb{E}[Y].$$

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.

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

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

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

Then $\mathrm{RECTS} \subset \mathcal{C}$. Now we verify that $\mathcal{C}$ is a λ-system. 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

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

(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 we get

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

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

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

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 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])

Definition (π-system). A collection of sets $\mathcal{A}$ is called a π-system if:

$$A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}.$$

Example:

$$\mathcal{A} = \{(-\infty, x] \subseteq \mathbb{R} \mid x \in \mathbb{R}\}.$$

Monotone Class Theorem cont.

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) $f, g \in \mathcal{H} \implies f + g \in \mathcal{H}$ further, $\forall c \in \mathbb{R} : \quad c \cdot f \in \mathcal{H}$

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

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

Application of Monotone Class Theorem

The Monotone Class Theorem plays a crucial role in proving that conditional expectation satisfies key properties, such as:

(i) $\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$.

(ii) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{A}] = \mathbb{E}[X \mid \mathcal{A}], \quad$ if $\mathcal{A} \subseteq \mathcal{G}$.

(iii) If Y is $\mathcal{G}$-measurable, then:

$$\mathbb{E}[YX \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}].$$

Idea of proof:

$$\mathcal{H} := \{X : \mathbb{E}[YX \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}] \text{ for all } \mathcal{G}\text{-measurable } Y\}.$$

Review

Lemma 2.1

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

$$\mathbb{E}[X|\mathcal{G}](\omega) = \frac{\mathbb{E}[X;\Omega_i](\omega)}{\mathbb{P}(\Omega_i)} \quad \textit{for a.s. } \omega \in \Omega.$$

That is,

$$\mathbb{E}[X|\mathcal{G}](\omega) = \sum_i \mathbb{1}_{\Omega_i}(\omega) \frac{\int_{\Omega_i} X(\omega) d\mathbb{P}}{\mathbb{P}(\Omega_i)} \quad \text{for a.s. } \omega \in \Omega. \tag{5}$$

Review cont

Example 2.2

Let $\Omega = \mathbb{R}$ and $X \sim \mathcal{N}(0,1)$. Let $\Omega_1 = \{\omega : X < 0\}$ and $\Omega_1 = \{\omega : X \geq 0\}$. Let $\mathcal{G} = \sigma(\{\Omega_1, \Omega_2\})$. Then

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \begin{cases} 2\frac{\int_{-\infty}^{0} xe^{-x^2/2}\,dx}{\sqrt{2\pi}}, & \text{if } \omega \in \Omega_1, \\ 2\frac{\int_{0}^{\infty} xe^{-x^2/2}\,dx}{\sqrt{2\pi}}, & \text{if } \omega \in \Omega_2. \end{cases}$$

By Lemma 2.1,

$$\mathbb{E}\left[X|\mathcal{G}\right](\omega) = \sum_i \mathbb{1}_{\Omega_i}(\omega)\frac{\int_{\Omega_i} X(\omega)d\mathbb{P}}{\mathbb{P}(\Omega_i)}.$$

If we apply Lemma 2.1 with $X = \mathbb{1}_A$:

$$\mathbb{E}\left[\mathbb{1}_A|\mathcal{G}\right](\omega) = \frac{\mathbb{P}(A \cap \Omega_i)}{\mathbb{P}(\Omega_i)} = \mathbb{P}(A|\Omega_i), \text{ if } \omega \in \Omega_i. \tag{6}$$

We define the conditional probability w.r.t. sub-σ-algebra:

$$\mathbb{P}(A|\mathcal{G})(\omega) := \mathbb{E}\left[\mathbb{1}_A|\mathcal{G}\right](\omega). \tag{7}$$

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

Proof of (iii):

$$\begin{aligned}
\mathbb{P}(A \cap G) &= \sum_{i:\Omega_i \subseteq G} \mathbb{P}(A \cap \Omega_i) \\
&= \sum_{i:\Omega_i \subseteq G} \mathbb{P}(\Omega_i)\, \mathbb{P}(A|\Omega_i) \\
&= \sum_{i:\Omega_i \subseteq G} \mathbb{P}(\Omega_i)\, \mathbb{P}(A|\mathcal{G}) \quad \text{by (6) and (7)} \\
&= \int_G \mathbb{P}(A|\mathcal{G})\, d\mathbb{P}.
\end{aligned}$$

Review cont

Remark 2.3

For $A \in \mathcal{F}$, $\mathbb{E}\left[\mathbb{1}_A | \mathcal{G}\right] = \mathbb{P}\left(A|\mathcal{G}\right)$ is defined on $\Omega_A \subset \Omega$, $\mathbb{P}\left(\Omega_A\right) = 1$. So, $\exists Z_A \in \mathcal{G}$ s.t. $\mathbb{P}(Z_A) = 0$ and $\mathbb{P}\left(A|\mathcal{G}\right)$ is not defined on Z_A.

Review 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) $\mathbb{P}(\emptyset|\mathcal{G})(\omega) = 0$ *and* $\mathbb{P}(\Omega|\mathcal{G})(\omega) = 1$ *for* $\omega \in \Omega \setminus (Z_1 \cup Z_2)$.

(b) *For* $A \in \mathcal{F}$,

$0 \leq \mathbb{P}(A|\mathcal{G}) \leq 1$, *for* $\omega \in \Omega \setminus Z_A$.

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

$$\mathbb{P}(A|\mathcal{G}) = \sum_{n=1}^{\infty} \mathbb{P}(A_n|\mathcal{G}), \quad \text{for } \omega \in \Omega \setminus \bigcup_n Z_{A_n}.$$

Review cont

We have a problem: For each $\alpha \in [0,1]$, let $\{B_{\alpha,n}\}_n \in \mathcal{F}$. Then there exists $\cup_n Z_{\alpha,n}$ with $\mathbb{P}(\cup_n Z_{\alpha,n}) = 0$. Do we have

$$\mathbb{P}(\bigcup_{\alpha\in[0,1]} \bigcup_n Z_{\alpha,n}) = 0?$$

We wish that there exists $\widetilde{Z} \in \mathcal{G}$ with $\mathbb{P}(\widetilde{Z}) = 0$ such that for any fixed $\omega \in \Omega \setminus \widetilde{Z}$,

$$\mathbb{P}\left(\bigsqcup_{n=1}^{\infty} A_n|\mathcal{G}\right)(\omega) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n|\mathcal{G}\right)(\omega), \quad \forall\{A_n\} \in \mathcal{F}.$$

which implies that $\mathbb{P}\left(\cdot|\mathcal{G}\right)$ is a conditional probability measure.

Review cont

Goal: Find a sufficient condition on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ such that for a.s. $\omega \in \Omega$, $\mathbb{P}(\cdot|\mathcal{G})$ is a conditional probability measure.

Before we state the sufficient condition, let's start with a description of an abstract object that corresponds to conditional probability.

R.C.D.

- Probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- Sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$.
- Measurable space $(S, \mathcal{S})$.
- Measurable map $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$.

Definition 3.1 (Regular conditional Distribution)

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

(a) Fix $A \in \mathcal{S}$,
$\omega \mapsto \mu_{X|\mathcal{G}}(\omega, A)$ is $\mathcal{G}$ measurable.

(b) Fix $\omega \in \Omega \setminus \widetilde{Z}$ with $\mathbb{P}(\widetilde{Z}) = 0$,
$B \mapsto \mu_{X|\mathcal{G}}(\omega, B)$ is a probability measure on $(S, \mathcal{S})$.
Moreover, $\mu_{X|\mathcal{G}}(\omega, B) = \mathbb{P}\left(X \in B|\mathcal{G}\right), \forall B \in \mathcal{S}$.

If $S = \Omega$ and X is the identity map $X(\omega) = \omega$ then we say that $\mu_{X|\mathcal{G}}$ is a regular conditional probability.

Existence of R.C.D.

Theorem 3.2 (Existence of R.C.D.)

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 Borel space. Then any S-valued r.v. X admits a regular conditional distribution given $\mathcal{G}$.

The proof follows [13, Proposition 7.14].

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

Existence of R.C.P.

Corollary of Theorem 3.2:
If $(\Omega, \mathcal{F}, \mathbb{P})$ is a Borel space, then $\mu_{X|\mathcal{G}}$ is a regular conditional probability.

Example 3.3 (Example of R.C.D.)

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

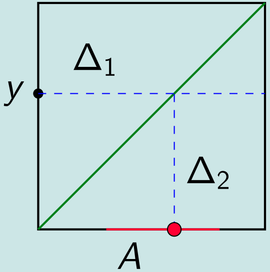
$$\mu(y, A) := \int_A f(x, y)dx \Big/ \int_{-\infty}^{\infty} f(x, y)dx.$$

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

Concrete example:

Let

$$\Delta_1 = \{(x, y) \in [0, 1]^2 : y > x\},$$
$$\Delta_2 = \{(x, y) \in [0, 1]^2 : y \leq x\}.$$



$$f(x,y) = \begin{cases} \frac{1}{2}, & (x,y) \in \Delta_1 \\ \frac{3}{2}, & (x,y) \in \Delta_2 \end{cases}$$

- f is a density function.

The conditional measure $\mu(y, A)$ is given by:

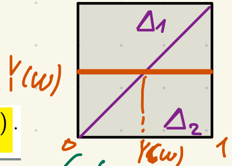
$$\mu(y,A) = \frac{\int_A f(x,y)dx}{f_Y(y)} = \frac{\frac{1}{2}\mathcal{L}(A \cap [0,y]) + \frac{3}{2}\mathcal{L}([y,1] \cap A)}{\frac{1}{2}y + \frac{3}{2}(1-y)}.$$

$$\mu_{X|\mathcal{G}}(\omega, A) = \mathbb{P}(X \in A | \mathcal{G}) = \frac{\mathcal{L}(A \cap [0, Y(\omega)] + 3\mathcal{L}([Y(\omega), 1] \cap A)}{3 - 2Y(\omega)}$$

The conditional density function $f_{X|\mathcal{G}} : \Omega\times\mathbb{R}\to[0,\infty)$ satisfies (by def.)

$x \mapsto f_{X|\mathcal{G}}(\omega,x)$ is Borel measurable, $\quad \mathcal{G}=\sigma(Y)$

$\omega \mapsto f_{X|\mathcal{G}}(\omega,x)$ is $\mathcal{G}$ measurable $\forall x\in\mathbb{R}$.



$$\int_B f_{\mathbf{X}|\mathcal{G}(\omega,\mathbf{x})}\,dx = \mu_{\mathbf{X}|\mathcal{G}}(\omega,B).$$

SUBSTITUTE equation (15)

$$f_{X|\mathcal{G}}(\omega,x)=\begin{cases}\dfrac{\frac{1}{2}}{\frac{1}{2}Y(\omega)+\frac{3}{2}(1-Y(\omega))} & \text{if } (x,Y(\omega))\in\Delta_1\\[2ex] \dfrac{3}{\frac{1}{2}Y(\omega)+\frac{3}{2}(1-Y(\omega))} & \text{if } (x,Y(\omega))\in\Delta_2\end{cases}$$

That is $f_{X|\mathcal{G}}(\omega,x)=\begin{cases}\dfrac{1}{3-2Y(\omega)} & \text{if } (x,Y(\omega))\in\Delta_1\\[2ex] \dfrac{3}{3-2Y(\omega)} & \text{if } (x,Y(\omega))\in\Delta_2\end{cases}$

For every fixed $\omega\in\Omega$, $A\subset\mathbb{R}$ Borel set

$$\int_A f_{X|\mathcal{G}}(\omega,x)\,dx = \frac{\ell[A\cap[0,Y(\omega)]]+3\ell(A\cap[Y(\omega),1])}{3-2Y(\omega)} = \mu_{X|\mathcal{G}}(\omega,A).$$

Proof of Theorem. 3.2 for $S = \mathbb{R}$

First we assume that $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$.
We first consider the collection of sets $\mathcal{A} = \{(-\infty, x) : x \in \mathbb{R}\}$. We claim that for a.s. $\omega \in \Omega$, there exists a probability measure $\mu_{X|\mathcal{G}}(\omega, \cdot)$ on $\mathbb{R}$ such that

(*i) $\mu_{X|\mathcal{G}}(\omega, (-\infty, x])$ is $\mathcal{G}$- measurable function, $\forall x \in \mathbb{R}$.

(*ii) $\mu_{X|\mathcal{G}}(\omega, (-\infty, x]) = \mathbb{P}(X \leq x|\mathcal{G})(\omega)$.

For a rational number $q \in \mathbb{Q}$ we define the r.v.

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

$P^q(\cdot)$ is $\mathcal{G}$-measurable.

By throwing away countably many null sets we may suppose that

$$P^q(\omega) \le P^r(\omega), \quad \forall q \le r, \; q, r \in \mathbb{Q} \text{ and } \forall \omega \tag{8}$$

and

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

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

$$F(\omega, x) := \lim_{q \in \mathbb{Q}, q > x} P^q(\omega). \tag{9}$$

For each $x \in \mathbb{R}$, $F(\cdot, x)$ is $\mathcal{G}$-measurable.

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 $\lim_{x\to\infty} F(x) = 1$.

Hence there exists a probability measure $\mu_{X|\mathcal{G}}(\omega, \bullet)$ satisfying

$$\mu_{X|\mathcal{G}}(\omega, (-\infty, x]) = F(\omega, x), \quad \forall\omega, \forall x. \tag{10}$$

For each $x \in \mathbb{R}$, $\mu_{X|\mathcal{G}}(\cdot, (-\infty, x])$ is $\mathcal{G}$-measurable, which proves (*i).

Moreover, since for a.s. ω,

$$F(\omega, x) = \inf_{q>x, q\in\mathbb{Q}} P^q(\omega) = \lim_{q\downarrow x, q\in\mathbb{Q}} P^q(\omega)$$
$$= \lim_{q\downarrow x} \mathbb{P}\left(X \le q|\mathcal{G}\right)(\omega) = \mathbb{P}\left(X \le x|\mathcal{G}\right)(\omega), \quad \forall x \in \mathbb{R}.$$

By this and (10) we have for a.s. ω,

$$\mu_{X|\mathcal{G}}(\omega, (-\infty, x]) = \mathbb{P}\left(X \le x|\mathcal{G}\right)(\omega), \quad \forall x \in \mathbb{R}.$$

This proves (*ii).

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 $\mathbb{P}\left(X \in B|\mathcal{G}\right)(\omega)$.

Cleary,

$$\mathcal{L} \supseteq \mathcal{A}(:= \{(-\infty, x) : x \in \mathbb{R}\}).$$

Check that

- $\mathcal{L}$ is λ-system (we omit this proof).
- $\mathcal{A}$ is a π-system such that $\mathcal{R} = \sigma(\mathcal{A})$.

Then $\mathcal{L} \supseteq \mathcal{R}$. The proof of Theorem 3.2 is completed in the case of $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$.

Proof of Theorem. 3.2 in the general case

Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ 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.:

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

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.

Corollary of Theorem 3.2:

Theorem 3.4 (Expectation w.r.t. the R.C.D.)

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 $\mathbb{E}\left[|f(X)|\right] < \infty$. Then

$$\mathbb{E}\left[f(X)|\mathcal{F}\right] = \int f(x) \cdot \mu(\omega, dx). \tag{11}$$

E.g. If $f = \mathbb{1}_A$, then

$$\mathbb{E}\left[\mathbb{1}_A|\mathcal{F}\right](\omega) = \mu(\omega, A).$$

Conditional Characteristic Function

Notation for the next slides:

- $(\Omega, \mathcal{F}, \mathbb{P})$ is the given probability 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 3.5 (Regular conditional cdf)

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

Conditional Characteristic Function cont.

Definition 3.6

$f_{\mathbf{X}|\mathcal{G}} : \Omega \times \mathbb{R}^n \to [0, \infty)$ 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 3.7 (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

$$\varphi_{X|\mathcal{G}}(\omega, \mathbf{t}) := \int_{\mathbb{R}^n} e^{i\mathbf{t}\cdot\mathbf{x}} d\mu_{\mathbf{X}|\mathcal{G}}(\omega, d\mathbf{x}) \overset{\text{By Theorem 3.4}}{=} \mathbb{E}\left[e^{i\mathbf{t}\cdot\mathbf{X}}|\mathcal{G}\right](\omega), \quad \mathbf{t} \in \mathbb{R}^n, \tag{12}$$

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

Conditional Characteristic Function cont.

Theorem 3.8

The following two assertions are equivalent

(a) *There exists a function* $\varphi : \mathbb{R}^n \to \mathbb{C}$ *such that for* $\mathbb{P}$*-almost all* $\omega \in \Omega$,

$$\varphi_{X|\mathcal{G}}(\omega, \mathbf{t}) = \varphi(t), \quad \forall t \in \mathbb{R}^n.$$

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

Proof of Theorem 3.8 (**a**)⇒ (**b**):

By (12),

$$\mathbb{E}\left[e^{i\mathbf{t}\cdot\mathbf{X}}|\mathcal{G}\right](\omega) = \varphi_{X|\mathcal{G}}(\omega,\mathbf{t}). \tag{13}$$

Multiply both sides with a r.v. Y which is bounded (real-valued) and $\mathcal{G}$-measurable, we get

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

Taking expectations,

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

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

$$\mathbb{E}\left[Ye^{i\mathbf{t}\cdot\mathbf{X}}\right] = \mathbb{E}\left[e^{i\mathbf{t}\cdot\mathbf{X}}\right]\cdot\mathbb{E}\left[Y\right] \tag{14}$$

Proof of Theorem 3.8 (**a**)$\Rightarrow$ (**b**)

holds for all $\mathcal{G}$-measurable bounded Y and $\mathbf{t}\in\mathbb{R}^n$. So, (14) 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 (14)

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

This implies that X and Z are independent, and thus, X and $\mathcal{G}$ are independent.

Proof of Theorem 3.8 cont (**b**)$\Rightarrow$ (**a**)

By (13),

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

The continuous case

$\mathbb{P}(Z \in H) = \int_H f_Z(t)\,dt$, $H \subset \mathbb{R}^{k+\ell}$

Theorem 3.9

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:

(defined on slide 38)

The continuous case cont.

$\mathcal{G} = \sigma(Y)$

Theorem 3.9 cont.

$$f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x}) = \begin{cases} \dfrac{f_{\mathbf{Z}}(\mathbf{x}, \mathbf{Y}(\omega))}{\int_{\mathbb{R}^k} f_{\mathbf{Z}}(\mathbf{x}, \mathbf{Y}(\omega)) d\mathbf{x}}, & \text{if } \int_{\mathbb{R}^\ell} f(\mathbf{x}, \mathbf{Y}(\omega)) d\mathbf{x} > 0; \\ f_0(\mathbf{x}), & \text{otherwise,} \end{cases} \tag{15}$$

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

$$\mathbb{E}\left[\mathbb{1}_{\mathbf{Y}\in B}(\omega) \cdot \int_A f_{\mathbf{X}|\mathcal{G}}(\omega, \mathbf{x}) d\mathbf{x}\right] = \mathbb{E}\left[\mathbb{1}_{\mathbf{Y}\in B}(\omega) \cdot \mathbb{1}_{\mathbf{X}\in A}(\omega)\right], \tag{16}$$

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

$\mathcal{G} = \sigma(Y)$ Clearly $\int_A f_{X|\mathcal{G}}(\omega, x)\,dx$ is $\mathcal{G}$ measurable by (15).

If $\int_{Y^{-1}(B)} \int_A f_{X|\mathcal{G}}(\omega,x)\,dx\,d\mathbb{P}(\omega) = \int_{Y^{-1}(B)} \mathbb{P}(X\in A|\mathcal{G})\,d\mathbb{P}(\omega)$ then $\int_A f_{X|\mathcal{G}}(\omega,x)\,dx$ is a version of $\mathbb{P}(X\in A|\mathcal{G})$.

$\{\omega : Y \in B\} = Y^{-1}(B)$

$$\int_{Y^{-1}(B)} \int_A f_{X|\mathcal{G}}(\omega,x)\,dx\,d\mathbb{P}(\omega) = \mathbb{E}\left[\mathbb{1}_{Y\in B}(\omega) \cdot \int_A f_{X|\mathcal{G}}(\omega,x)\,dx\right]$$ the l.h.s. in (16)

$$\int_{Y^{-1}(B)} \mathbb{P}(X\in A|\mathcal{G})\,d\mathbb{P}(\omega) = \int_{Y^{-1}(B)} \mathbb{P}(X^{-1}(A)|\mathcal{G})\,d\mathbb{P}(\omega) = \int_{Y^{-1}(B)} \mathbb{E}[\mathbb{1}_{X^{-1}(A)}(\omega)|\mathcal{G}]\,d\mathbb{P}(\omega) = \mathbb{E}\left[\underbrace{\mathbb{1}_{\{Y\in B\}}}_{\in\mathcal{G}} \cdot \mathbb{E}[\mathbb{1}_{\{X\in A\}}|\mathcal{G}]\right]$$

$$= \mathbb{E}\left[\mathbb{E}[\mathbb{1}_{Y\in B} \cdot \mathbb{1}_{X\in A}|\mathcal{G}]\right] = \mathbb{E}[\mathbb{1}_{Y\in B} \cdot \mathbb{1}_{X\in A}]$$ the r.h.s. of (16).

The continuous case cont.

proof cont.

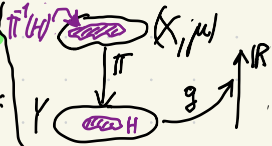
$$\mathbb{E}\left[\mathbb{1}_{\mathbf{Y}\in B}(\omega)\cdot\int_A f_{\mathbf{X}|\mathcal{G}}(\omega,\mathbf{x})d\mathbf{x}\right]=\int_A \mathbb{E}\left[\mathbb{1}_{\mathbf{Y}\in B}(\omega)\cdot f_{\mathbf{X}|\mathcal{G}}(\omega,\mathbf{x})\right]d\mathbf{x}$$

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

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

Namely,

Change of variables formula in general

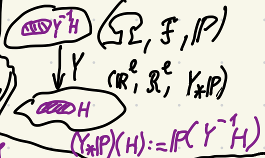


① Define the push forward measure $\pi_* \mu$ by $\pi_* \mu(H) := \mu(\pi^{-1} H)$

② The change of variables formula states:

$$\int_{y \in Y} g(y)\, d(\pi_* \mu)(y) = \int_{x \in X} g(\pi(x))\, d\mu(x).$$

A special case:



What is $Y_* \mathbb{P} = ?$ This is the distribution of Y (by definition)

Let $H < \mathbb{R}^\ell$. Then $(Y_* \mathbb{P})(H) = \mathbb{P}(Y^{-1} H) = \mathbb{P}(X \in \mathbb{R}^k, Y \in H) = \int_{y \in H} \int_{x \in \mathbb{R}^k} f_Z(x, y)\, dx\, dy$

← this is denoted by $f_Y(y)$ the marginal density function of (X, Y) corresponding to Y.

This means that

$$d(\mathbb{P}_* Y)(y) = f_Y(y)\, dy = \left(\int_{t \in \mathbb{R}^k} f_Z(t, y)\, dt \right) dy$$

$$\int_{y \in \mathbb{R}^\ell} g(y)\, d(Y_* \mathbb{P})(y) = \int g(Y(\omega)) \cdot d\mathbb{P}(\omega) \quad (*)$$

$g \geq 0$ measurable.

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

$$\mathbb{E}\left[\mathbb{1}_{Y\in B}(\omega)\cdot f_{X|\mathcal{G}}(\omega,x)\right] = \int_{Y^{-1}(B)} \frac{f_Z(x,Y(\omega))}{\int_{\mathbb{R}^k} f_Z(t,Y(\omega))\,dt}\,dP(\omega)$$

$$= \int_{Y^{-1}(B)} \overbrace{\frac{f_Z(x,Y(\omega))}{f_Y(Y(\omega))}}^{g(Y(\omega))} d\mathbb{P}(\omega) \overset{(*)}{=} \int_B \overbrace{\frac{f_Z(x,y)}{f_Y(y)}}^{g(y)} \cdot \overbrace{f_Y(y)\,dy}^{d(Y_*\mathbb{P})(y)} = \int_B f_Z(x,y)\,dy$$

The continuous case cont.

proof cont.

So,

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

$$=\int_A\int_B f_{\mathbf{Z}}(x,y)d\mathbf{y}d\mathbf{x}$$

$$=\mathbb{P}\left(X\in A;Y\in B.\right)\blacksquare$$

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:

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}, \quad \Phi(x) := \int_{-\infty}^{x} \varphi(y)dy. \tag{17}$$

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

$$\frac{1}{\sqrt{2\pi}} \cdot \left(x^{-1} - x^{-3}\right) \cdot e^{-x^2/2} \leq \mathbb{P}\left(X \geq x\right) \leq \frac{1}{\sqrt{2\pi}} \cdot x^{-1} \cdot e^{-x^2/2} \tag{18}$$

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

$$\lim_{n \to \infty} \mathbb{P}\left(a < \frac{Y_n - np}{\sqrt{np(1-p)}} < b\right) = \Phi(b) - \Phi(a). \tag{19}$$

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 $\mathbf{X}$

$$f(\mathbf{x}) = \frac{\sqrt{\det(A)}}{(2\pi)^{d/2}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \cdot A \cdot (\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{R}^d, \tag{20}$$

or

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \cdot \det(\Sigma)}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \cdot \Sigma^{-1} \cdot (\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{R}^d, \tag{21}$$

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:

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

where $\mathbf{X} = (X_1, \ldots, X_d)$. 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 $\lambda_1, \ldots, \lambda_d$ be the eigenvalues of A, and $\mathbf{v}_1, \ldots, \mathbf{v}_d$ be the ortonormal basis of $\mathbb{R}^d$ with the appropriate eigenvectors. Let us define diagonal matrix

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

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

Multivariate normal distribution cont.

Lemma 4.4

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

$$\mathbf{X} = P \cdot D^{-1/2} \cdot (Y_1, \ldots, Y_d) + \boldsymbol{\mu}, \tag{22}$$

where $Y_i \sim \mathcal{N}(0,1), i = 1, \ldots, d$ *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 $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ *has a multivariate normal distribution if for all* $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ *the following holds:*

$$a_1X_1 + \cdots + a_nX_n \text{ has univariate normal distribution.}$$

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:

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

The bivariate Case cont.

The correlation of (X, Y) is:

$$(23)\quad \rho := \rho_{X,Y} := \mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{\mathbb{E}\left[(X - \mu_X)[(Y - \mu_Y)\right]}{\sigma(X)\sigma(Y)}$$

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 $\Rightarrow$ independent for Gussian

Theorem 4.7

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

Multivariate normal distribution cont.

A more general theorem in this direction is:

Theorem 4.8

Let $\mathbf{X} = (X_1, \ldots, X_n)$ 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

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

The moment generating function is

typos

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

Conditioning normals

Given the multivariate normal vector

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

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

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

where $\tilde{\mathbf{Z}} := \mathbf{Z} - \boldsymbol{\mu}$ 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. Then for $A := \Sigma_{XY} \cdot \Sigma_{YY}^{-1}$ we have (simply by definitions) that

$$\mathbb{E}\left[\left(\widetilde{\mathbf{X}} - A\widetilde{\mathbf{Y}}\right) \cdot \widetilde{\mathbf{Y}}^T\right] = 0. \tag{24}$$

By Theorem 4.7 this implies that $\widetilde{\mathbf{X}} - A\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}$ are independent. By Theorem 3.8 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

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

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:

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

Assume that $V=(X,Y)$ is a multivariate normal r.v. and $\mathbb{E}[Y]=0$.

Question $\mathbb{E}[X|Y]=?$ $A=\Sigma_{XY}\cdot\Sigma_{YY}^{-1}$, where $\Sigma_{XY}=\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]=Cov(X,Y)$

$\Sigma_{YY}=\left(\mathbb{E}[(Y-\mu_Y)^2]\right)=Var(Y)$. $A=\Sigma_{XY}\cdot\Sigma_{YY}^{-1}=\frac{Cov(X,Y)}{Var(Y)}$. So, by the formula at the bottom of slide 67

(*)

$$\mathbb{E}[X|Y]=\mu_X+\frac{Cov(X,Y)}{Var(Y)}(Y-\mu_Y).$$

Using that $\mu_Y=0$ we get

(*') $$\mathbb{E}[X|Y]=\mu_X+\frac{\mathbb{E}[(X-\mu_X)\cdot Y]}{\mathbb{E}[Y^2]}\cdot Y.$$

Example for using this formula. Suppose that the weights X in (lbs) and the heights (in inches) Y of undergraduate collage men have a multivariate normal distribution with $\mu=\begin{pmatrix}175\\71\end{pmatrix}$ & $\Sigma=\begin{pmatrix}550 & 40\\40 & 8\end{pmatrix}$. Then by formula (*) the conditional distribution of X given that $Y=y$, is a normal distribution

Mean: $\mu_1+\frac{\sigma_{12}}{\sigma_{22}}(y-\mu_2)=175+\frac{40}{8}(y-71)=5y-180.$

Variance $\Sigma_{XX}-\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}=\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}=550-\frac{40^2}{8}=350$ ← By the last formula on the previous slide.

Assume that $Z=(X, Y_1, Y_2)$ is a multivariate normal vector. Let $\mathcal{G} := \sigma(Y_1, Y_2)$.

Question What is $\mathbb{E}[X \mid \mathcal{G}] = ?$ This is $\mu_{X|\mathcal{G}} = \mu_X + A(\underline{Y} - \mu_{\underline{Y}})$. What is A?

$A = \Sigma_{XY} \cdot \Sigma_{YY}^{-1}$. $\Sigma_{XY} = \mathbb{E}[\tilde{X} \cdot \tilde{Y}^T] = \mathbb{E}[(X-\mu_X)\cdot(Y_1-\mu_{Y_1}, Y_2-\mu_{Y_2})] = (\operatorname{cov}(X, Y_1), \operatorname{cov}(X, Y_2))$.

$$\Sigma_{YY} = \mathbb{E}\left[\begin{bmatrix} Y_1-\mu_{Y_1} \\ Y_2-\mu_{Y_2}\end{bmatrix} \cdot [Y_1-\mu_{Y_1}, Y_2-\mu_{Y_2}]\right] = \begin{bmatrix} \mathbb{E}[(Y_1-\mu_{Y_1})^2] & \mathbb{E}[(Y_1-\mu_{Y_1})(Y_2-\mu_{Y_2})] \\ \mathbb{E}[(Y_1-\mu_{Y_1})\cdot(Y_2-\mu_{Y_2})] & \mathbb{E}[(Y_2-\mu_{Y_2})^2] \end{bmatrix} = \begin{bmatrix} \operatorname{var}(Y_1) & \operatorname{cov}(Y_1, Y_2) \\ \operatorname{cov}(Y_1, Y_2) & \operatorname{var}(Y_2) \end{bmatrix}$$

$$A = [\operatorname{cov}(X, Y_1), \operatorname{cov}(X, Y_2)] \cdot \begin{bmatrix} \operatorname{var}(Y_1) & \operatorname{cov}(Y_1, Y_2) \\ \operatorname{cov}(Y_1, Y_2) & \operatorname{var}(Y_2) \end{bmatrix}^{-1}.$$

So,

$$\mathbb{E}[X \mid \mathcal{G}] = \mu_{X|\mathcal{G}} = \mu_X + [\operatorname{cov}(X, Y_1), \operatorname{cov}(X, Y_2)] \cdot \begin{bmatrix} \operatorname{var}(Y_1) & \operatorname{cov}(Y_1, Y_2) \\ \operatorname{cov}(Y_1, Y_2) & \operatorname{var}(Y_2) \end{bmatrix}^{-1} \cdot \begin{pmatrix} Y_1 - \mu_{Y_1} \\ Y_2 - \mu_{Y_2} \end{pmatrix}.$$

Assume that $\mathbb{E}[Y_1]=\mathbb{E}[Y_2]=0$ & Y_1, Y_2 are independent.

$$\mathbb{E}[X|\mathcal{G}]=\mu_X+\left[\mathbb{E}[(X-\mu_X)Y_1],\mathbb{E}[(X-\mu_X)\cdot Y_2]\right]\cdot\begin{bmatrix}\mathbb{E}[Y_1^2] & 0\\ 0 & \mathbb{E}[Y_2^2]\end{bmatrix}^{-1}\cdot\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}$$

$$=\mu_X+\left[\mathbb{E}[(X-\mu_X)\cdot Y_1],\mathbb{E}[(X-\mu_X)\cdot Y_2]\right]\cdot\begin{bmatrix}(\mathbb{E}[Y_1^2])^{-1} & 0\\ 0 & (\mathbb{E}[Y_2^2])^{-1}\end{bmatrix}\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}$$

$$\begin{bmatrix}(\mathbb{E}[Y_1^2])^{-1}\cdot Y_1\\ (\mathbb{E}[Y_2^2])^{-1}\cdot Y_2\end{bmatrix}$$

$$=\mu_X+\mathbb{E}[(X-\mu_X)\cdot\left[(\mathbb{E}[Y_1^2])^{-1}\cdot Y_1+(\mathbb{E}[Y_2^2])^{-1}\cdot Y_2\right].$$

So,

(**) $$\boxed{\mathbb{E}[X|\mathcal{G}]=\mu_X+\mathbb{E}[(X-\mu_X)\cdot\left[(\mathbb{E}[Y_1^2])^{-1}\cdot Y_1+(\mathbb{E}[Y_2^2])^{-1}\cdot Y_2\right]}.$$

We put together (*) & (**) to get $\boxed{\mathbb{E}[X|Y_1,Y_2]=\mathbb{E}[X|Y_1]+\mathbb{E}[X|Y_2]-\mathbb{E}[X]}$

[1] MŔTON 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 kv̈etkezttetések 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