Regular Conditional Probability

This document reviews, drawing from the literature, the topic of regular conditional probability, which arises in the attempt to formalize the notion of a conditional probability $\mathbb{P}[\,\cdot \mid A]$ given an event A with zero probability. A classical example is given by a pair of random variables (X, Y), where we would like to define and make sense of the distribution of Y when the event $\{ X = x \}$ has been observed, i.e. of $\mathbb{P}[\{ Y \in B \} \mid X = x]$.

1 *Motivation & Limitations*

Conditional Densities

Let X, Y real valued random variables with a joint density f. Assume we have observed X taking a concrete value $x \in \mathbb{R}$ and want to predict likely values of Y, i.e. we seek a *conditional distribution* of Y given $X = x$. Unfortunately, since $\{ X = x \}$ is an event of probability zero, $\mathbb{P}[Y \in \cdot \mid \{ X = x \}]$ is not well defined from the point of view of elementary probability. However, if $\lim_{\varepsilon \to 0} \mathbb{P}[Y \in \cdot \mid \{ |X - x| \leq \varepsilon \}]$ exists, it is tempting to define a conditional distribution as this limit. Under some assumptions on the joint density f the limit does indeed exist, is a probability distribution and even has a density that has a simple relation to f.

Lemma 1.1. *If f is a "nice enough" density, then*

$$\mathbb{P}[Y \in B | X = x] := \lim_{\varepsilon \to 0} \mathbb{P}[\{ Y \in B \} \mid \{ |X - x| \leq \varepsilon \}]$$

exists for all $x \in \mathbb{R}$ and all $B \in \mathcal{B}_{\mathbb{R}}$ and defines a probability measure with density

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \tag{1}$$

where $f_X(x) = \int_{\mathbb{R}} f(x,y)\,\mathrm{d}y$ denotes the marginal density of X.

Proof. Fix $a \in \mathbb{R}$ and $B \in \mathcal{B}_{\mathbb{R}}$ and let $\delta(x,y) = f(x,y) - f(a,y)$. We make the following assumptions about f:

- f_X is strictly positive, and
- $\sup_{x \in [a-\varepsilon, a+\varepsilon]} \int_{\mathbb{R}} |\delta(x,y)|\,\mathrm{d}y \to 0$ as $\varepsilon \to 0$.

The second assumption is certainly met when $M_\varepsilon(y) := \sup_{|x-a| \leq \varepsilon} |\delta(x,y)| \in \mathcal{L}^1(\mathbb{R})$ for small enough ε and $M_\varepsilon(y) \to 0$ $\forall y \in \mathbb{R}$ as $\varepsilon \to 0$.

In this case,

$$\begin{aligned}
\mathbb{P}[\{Y \in B\} \cap \{|X - a| \leq \varepsilon\}] &= \int_B \int_{a-\varepsilon}^{a+\varepsilon} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_B \int_{a-\varepsilon}^{a+\varepsilon} f(a,y) + \delta(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= 2\varepsilon \int_B f(a,y)\,\mathrm{d}y + \int_{a-\varepsilon}^{a+\varepsilon} \int_B \delta(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= 2\varepsilon \int_B f(a,y)\,\mathrm{d}y + o(\varepsilon).
\end{aligned}$$

In particular for $B = \mathbb{R}$, we find

$$\mathbb{P}[|X - a| \leq \varepsilon] = 2\varepsilon \int_{\mathbb{R}} f(a,y)\,\mathrm{d}y + o(\varepsilon) = 2\varepsilon f_X(a) + o(\varepsilon).$$

Consequently,

$$\begin{aligned}
\lim_{\varepsilon \to 0} \mathbb{P}[\{Y \in B\} \mid \{|X - a| \leq \varepsilon\}] &= \lim_{\varepsilon \to 0} \frac{2\varepsilon \int_B f(a,y)\,\mathrm{d}y + o(\varepsilon)}{2\varepsilon f_X(a) + o(\varepsilon)} \\
&= \lim_{\varepsilon \to 0} \frac{\int_B f(a,y)\,\mathrm{d}y}{f_X(a)} = \int_B f_{Y|X}(y|a)\,\mathrm{d}y\,,
\end{aligned}$$

An alternative way of evaluating this limit would be L'Hôpital's rule. In this case, different "niceness" assumptions on f might be needed.

proving the claim. ∎

Using more abstract machinery, we can show that also in higher dimensions and for arbitrary joint densities, the limit $\mathbb{P}[Y \in B | X \in B(x, \varepsilon)]$ exists for almost every x, and can be computed with the conditional density.

Proposition 1.2. *Let (X, Y) an $\mathbb{R}^m \times \mathbb{R}^n$-valued random vector with joint density f w.r.t. the $m{+}n$-dimensional Lebesgue measure λ^{m+n}. Denote f_X the marginal density of X, let $S = \{ f_X > 0 \}$ and define $f_{Y|X}(y|x)$ as in (1) for $x \in S$ with arbitrary extension to a density for $x \in S^c$. Then for every $B \in \mathcal{B}_{\mathbb{R}^n}$, for λ^m-a.e. $x \in S$,*

$$\lim_{\varepsilon \to 0} \mathbb{P}[Y \in B \mid X \in B(x, \varepsilon)] = \int_B f_{Y|X}(y|x)\, \mathrm{d}y\,. \tag{2}$$

The exceptional set here possibly depends on B.

Proof. For $B \in \mathcal{B}_{\mathbb{R}^n}$, let $\nu_B = \mathbb{P}[X \in \cdot\,, Y \in B]$, and let $\nu = \mathbb{P}_X$. Then for $A \in \mathcal{B}_{\mathbb{R}^m}$, $\nu_B(A) = \int_A \int_B f(x, y)\, \mathrm{d}y\, \mathrm{d}x$, so $\nu_B = f_B \odot \lambda^m$ where $f_B(x) = \int_B f(x, y)\, \mathrm{d}y$. Similarly, $\nu = f_X \odot \lambda^m$. From ν_B, $\nu \ll \lambda^m$, it can be shown (see [1, Thm. 3.22]) that $\lim_{\varepsilon \to 0} \nu_B(B(x, \varepsilon))/\lambda^m(B(x, \varepsilon)) = f_B(x)$ for λ^m-a.e. x, and similarly for ν. Now

$$\begin{aligned} \mathbb{P}[Y \in B | X \in B(x, \varepsilon)] &= \frac{\mathbb{P}[X \in B(x, \varepsilon), Y \in B]}{\mathbb{P}[X \in B(x, \varepsilon)]} \\ &= \frac{\nu_B(B(x, \varepsilon))}{\lambda^m(B(x, \varepsilon))} \frac{\lambda^m(B(x, \varepsilon))}{\nu(B(x, \varepsilon))} \xrightarrow{\varepsilon \to 0} \frac{f_B(x)}{f_X(x)} = \int_B f_{Y|X}(y|x)\, \mathrm{d}y\,, \end{aligned}$$

proving the claim. □

In fact, this proof works for all families of sets "shrinking nicely to x" (rather than merely the ε-balls) in a sense precisely defined in [1, §3].

Remark 1.3. The form of the conditional density (1) obtained in lemma 1.1 and proposition 1.2 is reminiscent of Bayes' rule for the conditional probability of events and can be shown to have many useful computational properties. There are some drawbacks to this approach however:

(i) In proposition 1.2, equation (2) holds for any fixed $B \in \mathcal{B}_{\mathbb{R}^n}$ and for all $x \in \mathbb{R}^m$ outside of a λ^m-null set. Unfortunately the exceptional sets for different B's can differ, and we cannot rule out that their union could have positive measure. Thus establishing a probability measure as the limit $\mathbb{P}[Y \in \cdot \mid X \in B(x, \varepsilon)]$ as $\varepsilon \to 0$ for λ^m-a.e. x might fail.
(ii) While lemma 1.1 is not plagued by the issue described in the previous bullet, this comes at the price of much stronger assumptions. Moreover, the generalization to higher dimensions and random variables taking values in more general measurable spaces is, where feasible at all, laborious.
(iii) The assumptions on the density in lemma 1.1 are too strong and cumbersome to check.
(iv) Random variables with densities are not sufficiently general.
(v) The limiting procedure chosen to obtain (1) and (2) is arbitrary and it is unclear at this point whether all such limiting procedures yield a sensible limit, and if so, whether this limit is unique.

The following section presents a famous paradox due to Kolmogorov, illustrating that different limiting procedures can indeed yield different conditional distributions.

The Borel-Kolmogorov Paradox[1]

[1] The classical Borel-Kolmogorov Paradox is based on great circles on the sphere S^2 in $\mathbb{R}^3$. The approach followed here, inspired by [2], is slightly different.

We consider the unit half disc $H \subseteq \mathbb{R}^2$, as pictured in fig. 1, together with the uniform distribution $\mathcal{U}_H$, which has density $f = \frac{2}{\pi}$ w.r.t. the Lebesgue-measure λ^2 on H. Let $(X, Y) \sim \mathcal{U}_H$, $L = \{ X = 0 \} = [(0, 0) : (0, 1)]$ the line segment connecting the origin with the point $(0, 1)$, and denote $A = \{ X = 0, Y \geq 1/2 \} = [(0, 1/2) : (0, 1)]$ the upper half of L.

We compute the conditional distribution of Y given $X = 0$ in two different ways: once in Cartesian and once in polar coordinates. With respect to these conditional distributions, we compute the conditional probabilities $\mathbb{P}_{\text{Cart}}[A|L]$ and $\mathbb{P}_{\text{Pol}}[A|L]$ of A given L.

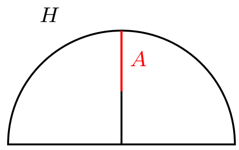

Figure 1: The unit half-disc H.

(i) By lemma 1.1, the conditional density of Y given $X = x$ is $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$. For $x = 0$, we obtain $f_{Y|X}(y|0) = \mathbb{1}_{[0,1]}(y)$, so $Y \sim \mathcal{U}_{[0,1]}$ given $X = 0$. From this,

$$\mathbb{P}_{\text{Cart}}[A|L] = \mathbb{P}[Y \geq 1/2 | X = 0] = \frac{1}{2}$$

follows.

(ii) Denote (R, Φ) the polar coordinates of (X, Y), which have density $g(r, \varphi) = \frac{2}{\pi} r$. It holds that $L = \{ X = 0 \} = \{ \Phi = \pi/2 \}$ and $A = \{ X = 0, Y \geq 1/2 \} = \{ \Phi = \pi/2, R \geq \frac{1}{2} \}$. The marginal density of R given $\Phi = \pi/2$ is computed using lemma 1.1 as $f_{R|\Phi}(r|\pi/2) = 2r$. Thus

$$\mathbb{P}_{\text{Pol}}[A|L] = \mathbb{P}[R \geq 1/2 | \Phi = \pi/2] = \int_{1/2}^{1} 2r \, \mathrm{d}r = 3/4.$$

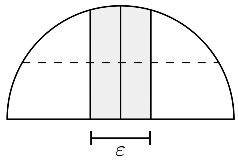

Figure 2: The limiting procedure in Cartesian coordinates. Given $\{ |X| \leq \varepsilon \}$, roughly half of the probability is concentrated on $\{ Y \geq 1/2 \}$.

The reason for the discrepancy becomes clearer from a geometric point of view. We have seen in the proof of lemma 1.1 that the conditional density (1) can be obtained from a limiting procedure. In Cartesian coordinates,

$$\mathbb{P}_{\text{Cart}}[A|L] = \lim_{\varepsilon \to 0} \mathbb{P}[Y \geq 1/2 | |X| \leq \varepsilon] = \frac{1}{2},$$

as seen in fig. 2. On the other hand, in polar coordinates,

$$\mathbb{P}_{\text{Pol}}[A|L] = \lim_{\varepsilon \to 0} \mathbb{P}[R \geq 1/2 | |\Phi - \pi/2| \leq \varepsilon] = \frac{3}{4},$$

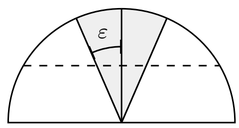

Figure 3: The limiting procedure in polar coordinates. Given $\{ |\Phi - \pi/2| \leq \varepsilon \}$, roughly $3/4^{\text{ths}}$ of the probability are concentrated on $\{ R \geq 1/2 \}$.

see also fig. 3. The conclusion is that

(i) conditional probability given an event of probability zero can in general not be defined in a coordinate independent manner, and

(ii) defining a conditional probability given an event L as a limit of conditional probabilities given events L_ε "approximating L" can be valid only when the approximating events L_ε are precisely specified.

E.g. the event $L = \{ X = 0 \}$ has $L_\varepsilon := \{|X| \leq \varepsilon\}$ and $\widetilde{L}_\varepsilon := \{|\Phi - \pi/2| \leq \varepsilon\}$ approximating sequences, and the corresponding limits disagree.

The next section presents an abstract, measure-theoretic approach that starts from conditional expectation and leads to the concept of regular conditional probability. The abstract construction is sufficiently general and overcomes some of the drawbacks of the more direct limit definition outlined in remark 1.3. It also yields identical results to the limit approach in many cases, in particular it reproduces the conditional density formula (1) (see proposition 3.22). While the Borel-Kolmogorov paradox shows that care is needed when working with conditional probabilities, regular conditional probability has properties that are often both intuitive and computationally useful, making it an interesting object of study.

2 *Definitions, Existence*

We work on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with $\mathcal{F} \subseteq \mathcal{A}$ a sub-σ-algebra. Let $(E', \mathcal{E}')$ and $(E, \mathcal{E})$ measurable spaces, X a random variable taking values in E' and Y a random variable taking values in E .

Lemma 2.1 (Factorization lemma). *Let $(E', \mathcal{E}')$ a measurable space, Ω a nonempty set and $f \colon \Omega \to E'$ a map. A map $g \colon \Omega \to \overline{\mathbb{R}}$ is $\sigma(f)$-$\mathcal{B}_{\overline{\mathbb{R}}}$-measurable iff there is a measurable map $\varphi \colon (E', \mathcal{E}') \to (\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$ such that $g = \varphi \circ f$.*

$$\begin{tikzcd}
(\Omega, \sigma(f)) \arrow[r, "g"] \arrow[d, "f"'] & (\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}}) \\
(E', \mathcal{E}') \arrow[ur, "\varphi"'] &
\end{tikzcd}$$

Definition 2.2. The map φ from the factorization lemma is called a **factorizing map** and denoted $g \circ f^{-1}$, even when the inverse map of f does not exist.

If f is surjective, the factorizing map φ is unique. Otherwise, the values of φ are determined on $f(\Omega)$, but any measurable $\widetilde{\varphi}$ with $\widetilde{\varphi}|_{f(\Omega)} = \varphi|_{f(\Omega)}$ is also a factorizing map. Even if f is not surjective, we will write $g \circ f^{-1}$ to denote some (choice of) factorizing map, understanding the indeterminacy of the values of this map on $E' \setminus f(\Omega)$.

Definition 2.3. Let $A \in \mathcal{A}$ an event. We call $\mathbb{P}[A|\mathcal{F}] \coloneqq \mathbb{E}[\mathbb{1}_A \,|\mathcal{F}]$ the **conditional probability** of A given the σ-algebra $\mathcal{F}$.

Remark 2.4. The conditional expectation and the conditional probability are random variables which are uniquely defined only up to equality almost surely. Thus concrete computations sometimes involve choice of a *version*. When no mention is made of a choice of version, we follow the convention that equalities involving conditional expectations and probabilities are to be read as equalities almost surely.

Definition 2.5. Let Y a random variable taking values in $(E, \mathcal{E})$. For every $B \in \mathcal{E}$, fix a version $\mathbb{P}[\{ Y \in B \} \,|\mathcal{F}]$ of the conditional probability of $\{ Y \in B \}$ given $\mathcal{F}$. Then

$$(\omega, A) \mapsto \mathbb{P}[\{ Y \in B \} \,|\mathcal{F}](\omega)$$

is called a (version of the) **conditional distribution** of Y given $\mathcal{F}$.

Remark 2.6. Let $(A_n) \subset \mathcal{A}$ disjoint and $A = \bigsqcup_{n \geq 1} A_n$. Fixing versions of $\mathbb{P}[A|\mathcal{F}]$ and $\mathbb{P}[A_n|\mathcal{F}]$, $n \in \mathbb{N}$, the conditional monotone convergence theorem implies that almost surely,

$$\mathbb{P}[A|\mathcal{F}] = \sum_{n \geq 1} \mathbb{P}[A_n|\mathcal{F}]. \tag{3}$$

Conditional probability can therefore be thought of as a countably additive map $A \mapsto \mathbb{P}[A|\mathcal{F}]$ from events to equivalence classes of random variables. Unfortunately, for fixed $\omega \in \Omega$, $A \mapsto \mathbb{P}[A|\mathcal{F}](\omega)$ is in general no measure. Even worse, it is possible that the measure property fails with positive probability. The reason for this is that (3) holds for ω outside of a $\mathbb{P}$-negligible set depending on the partition $(A_n)_{n \in \mathbb{N}}$ of the event A. Different partitions lead to different exceptional sets, and these exceptional sets can pile up, potentially causing countable additivity to fail with positive probability.

However, when viewing conditional probabilities as elements of $L^1(\mathbb{P})$ rather than (after choice of version) $\mathcal{L}^1(\mathbb{P})$, the map $A \mapsto \mathbb{P}[A|\mathcal{F}]$ is a vector measure, cf. [3].

It turns out that in many cases, there is a "regular enough" version of the conditional probability that is a measure for every $\omega \in \Omega$, which leads to some nice analytical properties. The following definition is key.

Definition 2.7. Let $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ measurable spaces. A map $\kappa \colon \Omega_1 \times \mathcal{A}_2 \to \overline{\mathbb{R}}_+$ is called a (σ-)finite **transition kernel** from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ if

(i) $\omega_1 \mapsto \kappa(\omega_1, A_2)$ is $\mathcal{A}_1$-measurable for any $A_2 \in \mathcal{A}_2$,
(ii) $A_2 \mapsto \kappa(\omega_1, A_2)$ is a (σ-)finite measure on $(\Omega_2, \mathcal{A}_2)$ for any $\omega_1 \in \Omega_1$.

If in (ii) the measure is a probability measure for all $\omega_1 \in \Omega_1$, then κ is called a **stochastic kernel** or Markov kernel. If $\kappa(\omega_1, \Omega_2) \leq 1$ for any ω_1, the kernel is called sub-Markov or sub-stochastic.

Remark 2.8. For a finite kernel, a Dynkin- or monotone class argument shows that it is sufficient to check the measurability property (i) for sets A_2 in a π-system $\mathcal{E}$ that generates $\mathcal{A}_2$ and either contains Ω_2 or a sequence $E_n \nearrow \Omega_2$.

A conditional distribution is called regular, if it can be represented by a stochastic kernel.

Definition 2.9. Let Y be a random variable taking values in $(E,\mathcal{E})$ and $\mathcal{F} \subset \mathcal{A}$ a sub-σ-algebra. A stochastic kernel $\kappa_{Y|\mathcal{F}}$ from $(\Omega,\mathcal{F})$ to $(E,\mathcal{E})$ is called a **regular conditional distribution of Y given $\mathcal{F}$**, if

$$\kappa_{Y|\mathcal{F}}(\omega, B) = \mathbb{P}[\{\, Y \in B \,\}\,|\mathcal{F}](\omega)$$

for $\mathbb{P}$-a.e. $\omega \in \Omega$ and all $B \in \mathcal{E}$.

Remark 2.10. (a) By definition, for every $\omega \in \Omega$, $\kappa_{Y|\mathcal{F}}(\omega,\,\cdot\,)$ is a measure such that in addition $\kappa_{Y|\mathcal{F}}(\,\cdot\,, B)$ is a version of $\mathbb{P}[\{\, Y \in B \,\}\,|\mathcal{F}]$. In particular for every $A \in \mathcal{F}$ it holds that

$$\mathbb{P}[A \cap \{\, Y \in B \,\}] = \mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{\, Y \in B \,\}}] = \int \mathbb{1}_A \, \kappa_{Y|\mathcal{F}}(\,\cdot\,, B) \, \mathrm{d}\mathbb{P}\,.$$

(b) A regular conditional distribution (r.c.d.) may not exist if $(E,\mathcal{E})$ is "too large", but it does exist when $(E,\mathcal{E})$ is a Borel space[2], e.g. a Polish space with its Borel-σ-algebra.

[2] See definition 2.18.

Definition 2.11. In the special case where $(E,\mathcal{E}) = (\Omega,\mathcal{A})$ the r.c.d. of the identity map $\mathrm{id}\colon \omega \mapsto \omega$ given $\mathcal{F}$ (if it exists) is called the **regular conditional probability given $\mathcal{F}$**. In this case, for all $B \in \mathcal{A}$,

$$\kappa_{\mathrm{id}|\mathcal{F}}(\,\cdot\,, B) = \mathbb{P}[\{\, Y \in B \,\}\,|\mathcal{F}] = \mathbb{P}[B|\mathcal{F}].$$

The following lemma shows that existence of a regular conditional probability is much stronger than existence of regular conditional distributions of specific random variables.

Lemma 2.12. *If there exists a regular conditional probability given $\mathcal{F}$ on $(\Omega,\mathcal{A},\mathbb{P})$, then any random variable defined on $(\Omega,\mathcal{A})$ has a regular conditional distribution given $\mathcal{F}$.*

Proof. Let Y a random variable defined on Ω taking values in $(E,\mathcal{E})$. Define $\kappa_{Y|\mathcal{F}}(\omega,\,\cdot\,) = \kappa_{\mathrm{id}|\mathcal{F}}(\omega, \{\, Y \in \,\cdot\, \}) = \kappa_{\mathrm{id}|\mathcal{F}}(\omega,\,\cdot\,) \circ Y^{-1}$, which is clearly a stochastic kernel from $(\Omega,\mathcal{F})$ to $(E,\mathcal{E})$. Furthermore,

$$\kappa_{Y|\mathcal{F}}(\omega, B) = \kappa_{\mathrm{id}|\mathcal{F}}(\omega, \{\, Y \in B \,\}) = \mathbb{P}[\{\, Y \in B \,\}\,|\mathcal{F}],$$

so $\kappa_{Y|\mathcal{F}}$ is a regular conditional distribution of Y given $\mathcal{F}$. □

Let X a random variable taking values in the measurable space $(E',\mathcal{E}')$. In analogy to how it is sometimes advantageous to work with the law $\mathbb{P}_X$ of X instead of the underlying probability measure $\mathbb{P}$, it is of interest to consider conditional distributions "given $X = x$" rather than $\omega \in \Omega$. Key to this is the factorization of X-measurable random variables according to lemma 2.1, which will recur in similar spirit several times.

The fact that factorizing maps are uniquely determined $X(\Omega)$ but can take arbitrary (up to measurability) values on $E' \setminus X(\Omega)$ causes some technical headaches.

Definition 2.13. For X a random variable taking values in $(E',\mathcal{E}')$, we call a map $E' \times \mathcal{E} \to \overline{\mathbb{R}}_+$,

$$(x, B) \mapsto \mathbb{P}[\{\, Y \in B \,\}\,|X = x] := \mathbb{P}[\{\, Y \in B \,\}\,|\sigma(X)] \circ X^{-1}$$

a (version of the) **conditional distribution of Y given X**.

The notation $\ldots|X = x$ will reappear in the sequel, always resulting from an application of the factorization lemma.

Remark 2.14. (i) Two versions of the conditional distribution of Y given X can disagree due to (a) different versions of the conditional probabilities of $\{\, Y \in B \,\}$ given $\sigma(X)$ being factorized and (b) the factorizing maps being indetermined on $E' \setminus X(\Omega)$.

(ii) Modifying a conditional distribution of Y given X on a $\mathbb{P}_X$-negligible set (in a way that perseveres measurability for fixed $B \in \mathcal{E}$) yields another conditional distribution.

As with the r.c.d. given a σ-algebra, an additional regularity assumption is required in order to obtain strong results.

Definition 2.15. Let X, Y be random variable taking values in measurable spaces $(E', \mathcal{E}')$, $(E, \mathcal{E})$ respectively.

(i) A stochastic kernel $\kappa_{Y|X}$ from $(E', \mathcal{E}')$ to $(E, \mathcal{E})$ is called a (version of the) **regular conditional distribution of** Y **given** X, if it holds that for every $x \in E'$ and $B \in \mathcal{E}$,
$$\kappa_{Y|X}(x, B) = \mathbb{P}[\{ Y \in B \} | X = x].$$

(ii) A version $\kappa_{Y|X}$ of the conditional distribution of Y given X is called **semi-regular**, if it has all the properties of a kernel with the exception that $\kappa_{Y|X}(x, \cdot\,)$ can fail to be a measure on $E' \setminus X(\Omega)$.

Remark 2.16. (i) A regular conditional distribution of Y given X is a kernel $\kappa_{Y|X}$ from $(E', \mathcal{E}')$ to $(E, \mathcal{E})$ for which in addition $\kappa_{Y|X}(\,\cdot\,, B)$ is a version of the conditional probability $\mathbb{P}[\{ Y \in B \} | X = \,\cdot\,]$.

(ii) A stochastic kernel $\kappa_{Y|X}$ from $(E', \mathcal{E}')$ to $(E, \mathcal{E})$ is a r.c.d. of Y given X iff for every $B \in \mathcal{E}$, $\kappa_{Y|X}(\,\cdot\,, B) \circ X = \mathbb{P}[\{ Y \in B \} | \sigma(X)]$.

(iii) On the terminology: conditional objects given a sub-σ-algebra $\mathcal{F}$ are random variables depending on $\omega \in \Omega$, while conditional objects "given X" depend on $x \in E'$. This clashes with the terminology for the conditional expectation where $\mathbb{E}[Y|\sigma(X)]$ can be written interchangeably as $\mathbb{E}[Y|X]$ (which we will avoid in this document). For additional clarity, we could speak (in slight abuse of notation) e.g. of $\mathbb{P}[\{ Y \in B \} | X = \,\cdot\,]$ as a conditional probability of $\{ Y \in B \}$ given $X = x$.

The reason for introducing semi-regularity is the indeterminacy of factorizing maps on $E' \setminus X(\Omega)$. If there exists a regular conditional distribution $\kappa_{Y|\sigma(X)}$ of Y given $\sigma(X)$, we can obtain a semi-regular conditional distribution $\kappa_{Y|X}$ of Y given X by factorizing $\kappa_{Y|\sigma(X)}$. Conversely, from any semi-regular conditional distribution $\kappa_{Y|X}$ we obtain a r.c.d. of Y given $\sigma(X)$ by composing with X in the first argument.

Lemma 2.17. *(a) If $\kappa_{Y|\sigma(X)}$ is a regular conditional distribution of Y given $\sigma(X)$, then any factorizing map $\kappa_{Y|X}(\,\cdot\,, B) := \kappa_{Y|\sigma(X)}(\,\cdot\,, B) \circ X^{-1}$ is a semi-regular distribution of Y given X. If in addition $X(\Omega)$ is measurable, there exists a r.c.d. of Y given X.*

(b) For any semi-regular conditional distribution $\kappa_{Y|X}$ of Y given X, the map $(\omega, B) \mapsto \kappa_{Y|X}(X(\omega), B)$ is a r.c.d of Y given $\sigma(X)$

Proof. (a) For $B \in \mathcal{E}$, let $\varphi_B = \kappa_{Y|\sigma(X)}(\,\cdot\,, B) \circ X^{-1} \colon E' \to \overline{\mathbb{R}}_+$ a factorizing map. By definition, $\kappa_{Y|\sigma(X)}(\,\cdot\,, B) = \mathbb{P}[\{ Y \in B \} | \sigma(X)]$ and φ_B is measurable. Let $\kappa_{Y|X}(x, B) := \varphi_B(x)$, then for $x = X(\omega_x) \in X(\Omega)$, it holds that
$$\kappa_{Y|X}(x, B) = \varphi_B(X(\omega_x)) = \kappa_{Y|\sigma(X)}(\omega_x, B).$$
It follows that $\kappa_{Y|X}(x, \cdot\,)$ is a measure since $\kappa_{Y|\sigma(X)}(\omega_x, \cdot\,)$ is. Consequently $\kappa_{Y|X}$ is a semi-regular conditional distribution of Y given X. If in addition $X(\Omega)$ is measurable, we define $\kappa_{Y|X}$ by
$$\kappa_{Y|X}(x, B) = \begin{cases} \delta_x(B) & x \in E' \setminus X(\Omega), \\ \varphi_B(x) & x \in X(\Omega). \end{cases}$$

Note that instead of defining $\kappa_{Y|X}(x, \cdot\,)$ as a Dirac measure for $x \notin X(\Omega)$, we could have chosen any other measure on E. In particular for $E = \mathbb{R}^n$, a measure defined by a density is admissible, which is useful when defining regular conditional distributions using densities.

(b) Let $\kappa_{Y|X}$ a semi-regular distribution of Y given X and let $\kappa(\omega, B) := \kappa_{Y|X}(X(\omega), B)$. For any $B \in \mathcal{E}$, $\kappa_{Y|X}(\,\cdot\,, B) = \mathbb{P}[\{ Y \in B \} | X = \,\cdot\,]$ by definition, which is $\mathcal{E}'$-$\mathcal{B}_{\overline{\mathbb{R}}}$-measurable as a factorizing map. Hence $\kappa(\,\cdot\,, B)$ is $\sigma(X)$-$\mathcal{B}_{\overline{\mathbb{R}}}$-measurable and a version of the conditional probability of $\{ Y \in B \}$:

$$\begin{aligned}\kappa(\,\cdot\,, B) &= \kappa_{Y|X}(\,\cdot\,, B) \circ X = \big(\mathbb{P}[\{ Y \in B \} \,|\sigma(X)] \circ X^{-1}\big) \circ X \\ &= \mathbb{P}[\{ Y \in B \} \,|\sigma(X)].\end{aligned}$$

Finally, $\kappa_{Y|X}(x, \cdot\,)$ is a measure for $x \in X(\Omega)$, so for any $\omega \in \Omega$, $\kappa(\omega, \cdot\,) = \kappa_{Y|X}(X(\omega), \cdot\,)$ is a measure, proving the claim.

□

Lemma 2.17 offers a way of computing a r.c.d. $\kappa_{Y|X}$ explicitly by first computing for fixed $B \in \mathcal{E}$ the (ordinary) conditional expectation $\mathbb{E}[\mathbb{1}_B(Y)|\sigma(X)]$. If this results in an expression $\Phi(X, B)$ with an explicit dependence on X, we can take (after checking the kernel properties) $\kappa_{Y|X}(x, B) = \Phi(x, B)$. See also the examples at the beginning of section 3.

We now come to the question of existence of regular conditional distributions, which is guaranteed when Y takes values in a "nice enough" space.

Definition 2.18. A measurable space $(E, \mathcal{E})$ is called a **Borel space**, if there exists a Borel set $B \in \mathcal{B}_\mathbb{R}$ such that $(E, \mathcal{E})$ and $(B, \mathcal{B}_B)$ are isomorphic.[3]

[3] Two measure spaces are said to be isomorphic, if there exists a measurable bijection between them, whose inverse is also measurable.

Theorem 2.19. *Let E a Polish space with Borel-σ-algebra $\mathcal{E}$. Then $(E, \mathcal{E})$ is a Borel space.*

Theorem 2.20. *Let $\mathcal{F} \subset \mathcal{A}$ a sub-σ-algebra. Let Y a random variable taking values in a Borel space $(E, \mathcal{E})$. Then there exists a regular conditional distribution $\kappa_{Y|\mathcal{F}}$ of Y given $\mathcal{F}$*

Proof sketch. In the case $E = \mathbb{R}$, construct a regular version of the distribution function $F(r, \omega) = \mathbb{P}[Y \in (-\infty, r]|\mathcal{F}](\omega)$ of the conditional distribution of Y by first defining it for rational r (and $\omega \in \Omega$ up to a null set) and then extending it to $r \in \mathbb{R}$. For $(E, \mathcal{E})$ a Borel space with an isomorphism $\varphi \colon E \to B \in \mathcal{B}_\mathbb{R}$, obtain a r.c.d. of $Y' = \varphi \circ Y$ and let $\kappa_{Y|\mathcal{F}}(\omega, B) \coloneqq \kappa_{Y'|\mathcal{F}}(\omega, \varphi(B))$ for $B \in \mathcal{E}$. For details, see [3, 4] □

3 Computation with regular conditional distributions

3.1 Independence

Let X, Y be random variable taking values in measurable spaces $(E', \mathcal{E}')$, $(E, \mathcal{E})$ respectively.

Lemma 3.1. *Y is independent of X iff the regular conditional distribution of Y given $\sigma(X)$ (or X) equals its unconditional distribution.*

Proof. since $\kappa_{Y|\sigma(X)}(\,\cdot\,, B) = \kappa_{Y|X}(\,\cdot\,, B) \circ X$ and $\kappa_{Y|X}(\,\cdot\,, B) = \kappa_{Y|\sigma(X)}(\,\cdot\,, B) \circ X^{-1}$, it suffices to show the claim "given $\sigma(X)$".

"$\Rightarrow$" This follows easily from the definition of the regular conditional distribution, since for $A \in \mathcal{F}$, $\mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{ Y \in B \}}] = \mathbb{P}[A]\mathbb{P}[Y \in B]$ and thus $\kappa_{Y|\mathcal{F}}(\omega, B) = \mathbb{P}[Y \in B]$.
"$\Leftarrow$" In this case, for any $A \in \mathcal{E}'$,

$$\begin{aligned}\mathbb{P}[X \in A, Y \in B] &= \mathbb{E}[\mathbb{1}_A(X)\kappa_{Y|\sigma(X)}(\,\cdot\,, B)] \\ &= \mathbb{E}[\mathbb{1}_A(X)\mathbb{P}[Y \in B]] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B].\end{aligned}$$

□

The regular conditional distribution behaves intuitively in the sense that when X is independent of other random quantities appearing in an expression, we can determine the r.c.d. of that expression given $X = x$ by replacing X with x, as exemplified in the following.

Example 3.2. Let X, Y independent $\mathbb{R}^n$-valued random vectors. Then a regular conditional distribution of $X + Y$ given X is given by the convolution

$$\mathbb{P}[X + Y \in \cdot\,|X = x] = \delta_x * \mathbb{P}_Y = \mathbb{P}_Y(\cdot - x).$$

Indeed, $\kappa(x, B) := (\delta_x * \mathbb{P}_Y)(B)$ is a kernel: (i) for any $B \in \mathcal{B}_{\mathbb{R}^n}$, $x \mapsto \kappa(x, B)$ is measurable (this holds for cubes by left-continuity and extends by remark 2.8), (ii) for any $x \in \mathbb{R}^n$, $\kappa(x, \cdot)$ is clearly a measure. To conclude, we check that $\kappa(\cdot, B) \circ X = \mathbb{P}[\{ X + Y \in B \}\,|\sigma(X)]$. For any $A \in \mathcal{B}_{\mathbb{R}^n}$,

$$\begin{aligned}\mathbb{E}[\mathbb{1}_A(X)\kappa(X, B)] &= \int_A \kappa(x, B)\,\mathrm{d}\mathbb{P}_X = \int_A \int_{\mathbb{R}^n} \mathbb{1}_{B-x}(y)\,\mathrm{d}\mathbb{P}_Y(y)\,\mathrm{d}\mathbb{P}_X(x) \\ &= \int_A \int_{\mathbb{R}^n} \mathbb{1}_B(x + y)\,\mathrm{d}\mathbb{P}_Y(y)\,\mathrm{d}\mathbb{P}_X(x) = \mathbb{E}[\mathbb{1}_A(X)\,\mathbb{1}_B(X + Y)],\end{aligned}$$

proving the claim.

Example 3.3. More generally, if X, Y are independent random variables taking values in the measurable spaces $(E', \mathcal{E}')$ and $(E, \mathcal{E})$ respectively and $g\colon E' \times E \to \mathbb{R}$ is measurable s.t. $g(X, Y) \in \mathcal{L}^1(\mathbb{P})$, then a regular conditional distribution of $g(X, Y)$ given X is given by

$$\mathbb{P}[g(X, Y) \in \cdot\,|X = x] = \mathbb{P}_{g(x,Y)}$$

whenever $E' \ni x \mapsto \mathbb{P}_{g(x,Y)}(B)$ is measurable for all $B \in \mathcal{B}_{\mathbb{R}}$. This follows, since

$$\begin{aligned}\mathbb{P}[\{ g(X, Y) \in B \}\,|\sigma(X)] &= \mathbb{E}[\mathbb{1}_B(g(X, Y))|\sigma(X)] \\ &= \mathbb{E}[\mathbb{1}_B(g(x, Y))]|_{x=X} \\ &= \mathbb{P}[g(x, Y) \in B]|_{x=X} = \mathbb{P}_{g(x,Y)}(B) \circ X,\end{aligned}$$

where in the second equality we have used a result about the conditional expectation of a function of independent random variables.

Example 3.4. Let $X, Z_1, \ldots, Z_n \overset{\perp\!\!\!\perp}{\sim} \mathcal{U}_{[0,1]}$ be independent uniform random variables on $[0, 1]$, and let $Y_k = \mathbb{1}_{\{ Z_k \le X \}}$, $1 \le k \le n$. Intuitively, given $X = x$, $Y = (Y_1, \ldots, Y_n)$ has distribution $\mathrm{Ber}_x^{\otimes n}$. Indeed, for $w \in \{ 0, 1 \}^n$, we can check directly or infer from the previous example that

$$\mathbb{E}[\mathbb{1}_{\{ Y=w \}}\,|X] = X^{|w|}(1 - X)^{n-|w|} = \mathrm{Ber}_X^{\otimes n}(\{ w \}),$$

where $|w| := \sum_{i=1}^n w_i$. This proves that a r.c.d. of Y given X is given by $\kappa_{Y|X}(x, B) = \mathrm{Ber}_x(B)$ for $x \in [0, 1]$, $B \subset \{ 0, 1 \}^n$.

An alternative approach to independence is provided by conditional characteristic functions.

Definition 3.5. The **conditional characteristic function** of Y given $\mathcal{F}$ is the map

$$\varphi_{Y|\mathcal{F}}(\omega, t) := \int_{\mathbb{R}^n} e^{i\langle t,y\rangle}\,\mathrm{d}\kappa_{Y|\mathcal{F}}(\omega, \mathrm{d}y) \overset{\text{a.s.}}{=} \mathbb{E}[e^{i\langle t,Y\rangle}|\mathcal{F}](\omega)$$

A random variable Y is independent of $\mathcal{F}$ iff its conditional characteristic function $\varphi_{Y|\mathcal{F}}$ given $\mathcal{F}$ does not depend on ω:

Proposition 3.6. *Let Y a random vector taking values in $\mathbb{R}^n$ and $\mathcal{F} \subset \mathcal{A}$ a sub-σ-algebra. Then the following are equivalent:*

(i) Y is independent of $\mathcal{F}$
(ii) There exists a (deterministic) function $\varphi\colon \mathbb{R}^n \to \mathbb{C}$ such that for $\mathbb{P}$-a.e. $\omega \in \Omega$,

$$\varphi_{Y|\mathcal{F}}(\omega, t) = \varphi(t), \text{ for all } t \in \mathbb{R}^n$$

Proof. (i) $\Rightarrow$ (ii). If Y is independent of $\mathcal{F}$, we have

$$\varphi_{Y|\mathcal{F}}(\omega, t) = \mathbb{E}[e^{i\langle t,Y\rangle}|\mathcal{F}](\omega) \overset{\perp\!\!\!\perp}{=} \mathbb{E}[e^{i\langle t,Y\rangle}] = \varphi_Y(t)$$

almost surely, by irrelevance of independent information.
(ii) $\Rightarrow$ (i). For any X bounded and $\mathcal{F}$-measurable, we have

$$\mathbb{E}[Xe^{i\langle t,Y\rangle}] = \mathbb{E}[X \underbrace{\mathbb{E}[e^{i\langle t,Y\rangle}|\mathcal{F}]}_{=\varphi_{Y|\mathcal{F}}(\cdot,t)}] = \varphi(t)\,\mathbb{E}[X].$$

Taking $X = e^{isZ}$ with Z an $\mathcal{F}$-measurable random variable, we obtain that $Y \perp\!\!\!\perp Z$ by factorization of characteristic functions (Kác's theorem). Since Z was arbitrary, $Y \perp\!\!\!\perp \mathcal{F}$ follows. □

3.2 Conditional expectations from regular conditional distributions

With a regular conditional distribution of Y, we can obtain (a version of) the conditional expectation of $g(Y)$ given $\mathcal{F}$ as the expectation of g w.r.t. the conditional distribution $\kappa_{Y|\mathcal{F}}(\omega, \cdot\,)$.

Theorem 3.7 (Conditional expectation from r.c.d.). *Let $\kappa_{Y|\mathcal{F}}$ a regular conditional distribution of Y given $\mathcal{F}$. Further, let $g\colon E \to \mathbb{R}$ be measurable and either nonnegative or such that $g(Y) \in \mathcal{L}^1(\mathbb{P})$. Then*

$$\mathbb{E}[g(Y)|\mathcal{F}](\omega) = \int_E g(y)\kappa_{Y|\mathcal{F}}(\omega, \mathrm{d}y) \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega \tag{4}$$

Proof. We check that the right-hand side $\int g(y)\kappa_{Y|\mathcal{F}}(\,\cdot\,, \mathrm{d}y)$ has the properties of the conditional expectation. For $g = \mathbb{1}_B$, $B \in \mathcal{E}$, this holds by definition of the regular conditional distribution and extends to the general case using the standard machinery. □

Remark 3.8. The "standard machinery" involves extending the equality (4) from indicators and simple functions to general measurable g using the (conditional) monotone and dominated convergence theorems, which on the right-hand side of (4) hinge on $\kappa_{Y|\mathcal{F}}(\omega, \cdot\,)$ being a measure. That is, in proposition 3.7, regularity of the conditional distribution enters in an essential manner.

A well known result about the conditional expectation states that if X is $\mathcal{F}$-measurable and Y independent of $\mathcal{F}$, then for measurable g with $g(X,Y) \in \mathcal{L}^1(\mathbb{P})$, it holds that $\mathbb{E}[g(X,Y)|\mathcal{F}] = \mathbb{E}[g(x,Y)]_{x=X}$. The next result generalizes this fact. It states that the conditional expectation w.r.t. $\mathcal{F}$ is obtained by "holding all $\mathcal{F}$-measurable quantities fixed and averaging the rest w.r.t. their regular conditional distribution given $\mathcal{F}$". In particular it shows that $\mathbb{E}[g(X,Y)|\mathcal{F}] = \mathbb{E}[g(x,Y)|\mathcal{F}]|_{x=X}$.

Theorem 3.9. *Let X measurable w.r.t. $\mathcal{F}$ and $\kappa_{Y|\mathcal{F}}$ a regular conditional distribution of Y given $\mathcal{F}$. Furthermore, let $g\colon E' \times E \to \overline{\mathbb{R}}$ measurable and either nonnegative or s.t. $g(X,Y) \in \mathcal{L}^1(\mathbb{P})$. Then*

$$\begin{aligned}\mathbb{E}[g(X,Y)|\mathcal{F}](\omega) &= \int_E g(X(\omega),y)\kappa_{Y|\mathcal{F}}(\omega,\mathrm{d}y)\\ &= \mathbb{E}[g(x,Y)|\mathcal{F}](\omega)|_{x=X(\omega)}\end{aligned}$$

for $\mathbb{P}$-a.e. $\omega \in \Omega$.

Proof. Let $\mathcal{M} \subseteq B(E' \times E; \overline{\mathbb{R}})$ the set of bounded measurable maps such that (3.9) holds, and note that $\mathcal{M}$ is a vector space containing the constants. Let $B \in \mathcal{E}$, $B' \in \mathcal{E}'$ and $g = \mathbb{1}_{B' \times B}$. Then

$$\begin{aligned}\int_E g(X(\omega),y)\kappa_{Y|\mathcal{F}}(\omega,\mathrm{d}y) &= \int_E \mathbb{1}_{B'}(X(\omega))\,\mathbb{1}_B(y)\kappa_{Y|\mathcal{F}}(\omega,\mathrm{d}y)\\ &= \mathbb{1}_{B'}(X(\omega))\kappa_{Y|\mathcal{F}}(\omega,B) = \mathbb{1}_{B'}(X(\omega))\,\mathbb{E}[\mathbb{1}_{\{\,Y\in B\,\}}\,|\mathcal{F}]\\ &= \mathbb{E}[\mathbb{1}_{B'}(X)\,\mathbb{1}_B(Y)|\mathcal{F}](\omega) = \mathbb{E}[g(X,Y)|\mathcal{F}](\omega),\end{aligned}$$

so $\mathcal{M}$ contains the indicators of sets in $\mathcal{E}' \times \mathcal{E}$, which is a π-system generating $\mathcal{E}' \otimes \mathcal{E}$. Furthermore, $\mathcal{M}$ is closed under bounded nonnegative limits, and thus contains all bounded $(\mathcal{E}' \otimes \mathcal{E})$-measurable functions by the functional monotone class theorem.
The extension to measurable g with $g(X,Y) \in \mathcal{L}^1(\mathbb{P})$ is a routine argument, complicated only by the fact that the r.h.s. of (3.9) may be ill-defined a priori. Consider first $g \geq 0$. Both sides of (3.9) satisfy a monotone convergence theorem, so the equality holds (almost surely) in the non-negative case. With the l.h.s., the r.h.s. is integrable, so the (by the MCT measurable) map $\int_E g(X(\,\cdot\,),y)\kappa_{Y|\mathcal{F}}(\,\cdot\,,\mathrm{d}y)$ defined by the r.h.s. is finite almost surely. The result then extends a last time by splitting a general measurable g with $g(X,Y) \in \mathcal{L}^1(\mathbb{P})$ into positive and negative part and possibly redefining on a null set. ☐

Definition 3.10. Let X a random variable taking values in $(E', \mathcal{E}')$ and $g\colon E \to \overline{\mathbb{R}}$ measurable such that $g(Y) \in \mathcal{L}^1(\mathbb{P})$. Then

$$\mathbb{E}[g(Y)|X=x] \coloneqq \mathbb{E}[g(Y)|\sigma(X)] \circ X^{-1} \tag{5}$$

is called a (version of the) **conditional expectation** of $g(Y)$ given $X = x$.

Remark 3.11. (a) By definition, a measurable map $\Phi\colon E \to \overline{\mathbb{R}}$ is a version of the conditional expectation of $g(Y)$ given X iff $\Phi \circ X = \mathbb{E}[g(Y)|\sigma(X)]$ a.s.. In this case, if $\widetilde{\Phi} = \Phi$ $\mathbb{P}_X$-a.s., $\widetilde{\Phi}$ is also a version of the conditional expectation of $g(Y)$ given $X = x$.

(b) There are two sources of indeterminacy in defining $\mathbb{E}[g(Y)|X=x]$.

1. The factorizing map $\mathbb{E}[g(Y)|\sigma(X)] \circ X^{-1}\colon E' \to \overline{\mathbb{R}}$ is unique only for $x \in X(\Omega)$, and with arbitrary (up to measurability) values on $E' \setminus X(\Omega)$.
2. Two versions Z_1 and Z_2 of $\mathbb{E}[g(Y)|\sigma(X)]$ agree only almost surely. However, if $A = \{\,Z_1 \neq Z_2\,\}$, then $Z_1 \circ X^{-1}$ and $Z_2 \circ X^{-1}$ agree on $X(\Omega \setminus A)$.

Leaving issues of measurability aside, both of these indeterminacies are confined to (subsets of) $\mathbb{P}_X$-negligible sets, hence different versions of $\mathbb{E}[g(Y)|X=\,\cdot\,]$ agree $\mathbb{P}_X$-a.s.

(c) If $\mathbb{P}[X=x] > 0$, it holds that $\mathbb{E}[g(Y)|\{\,X=x\,\}] = \mathbb{E}[g(Y)|X=x]$. Indeed, in this case all versions of $\mathbb{E}[g(Y)|X]$ agree on $\{\,X=x\,\}$ as a consequence of the factorization lemma, so there is some constant c such that $\mathbb{E}[g(Y)|X]|_{\{X=x\}} = \mathbb{E}[g(Y)|X = x] = c$. Then $\mathbb{E}[g(Y)|\{X = x\}]\mathbb{P}[X = x] = \mathbb{E}[g(Y)\mathbb{1}_{\{X=x\}}] = \mathbb{E}[\mathbb{E}[g(Y)|X]\mathbb{1}_{\{X=x\}}] = c\,\mathbb{P}[X = x]$ and the claim follows.

There is a tower rule: $\mathbb{E}[g(Y)]$ is the expectation of the conditional expectation $\mathbb{E}[g(Y)|X = \cdot]$ w.r.t. $\mathbb{P}_X$.

Lemma 3.12. *Let $g\colon E \to \overline{\mathbb{R}}$ measurable such that $g(Y) \in \mathcal{L}^1(\mathbb{P})$. Then*

$$\mathbb{E}[g(Y)] = \int_{E'} \mathbb{E}[g(Y)|X = x]\,\mathrm{d}\mathbb{P}_X(x)$$

This generalizes the "classical" formula $\mathbb{E}[g(Y)] = \sum_x \mathbb{E}[g(Y)|X = x]\mathbb{P}[X = x]$ for X taking values in a discrete space.

Proof. By factorizing $\mathbb{E}[g(Y)|X] = (\mathbb{E}[g(Y)|X] \circ X^{-1}) \circ X = \mathbb{E}[g(Y)|X = \cdot\,] \circ X$, we can rewrite the expectation as an integral against $\mathbb{P}_X$:

$$\begin{aligned}\mathbb{E}[g(Y)] &= \mathbb{E}[\mathbb{E}[g(Y)|X]] = \mathbb{E}[\mathbb{E}[g(Y)|X = \cdot\,] \circ X] \\ &= \int_{E'} \mathbb{E}[g(Y)|X = x]\,\mathrm{d}\mathbb{P}_X(x)\,.\end{aligned}$$

□

In analogy to theorem 3.7, a version of $\mathbb{E}[g(Y)|X = \cdot\,]$ can be computed from a (semi-)regular conditional distribution $\kappa_{Y|X}$ of Y given X.

Proposition 3.13. *Let $g\colon E \to \overline{\mathbb{R}}$ measurable such that $g(Y) \in \mathcal{L}^1(\mathbb{P})$ and $\kappa_{Y|X}$ a regular conditional distribution of Y given X. Then a version of $\mathbb{E}[g(Y)|X = \cdot\,]$ is given by*

$$\mathbb{E}[g(Y)|X = \cdot\,] = \int_E g(y)\kappa_{Y|X}(\,\cdot\,, \mathrm{d}y).$$

for $\mathbb{P}_X$-a.e. $x \in E'$.

If $X(\Omega)$ is measurable, the above integral defines (for $x \in X(\Omega)$, with arbitrary measurable extension to E') a version of $\mathbb{E}[g(Y)|X = \cdot\,]$ even when $\kappa_{Y|X}$ is semi-regular.

Proof. Consider first the regular case. Let $\Phi(x) = \int_E g(y)\kappa_{Y|X}(x, \mathrm{d}y)$, which is measurable by $\kappa_{Y|X}$ being a kernel together with a monotone class argument. Note that integrability is no issue by arguments analogous to those in the proof of theorem 3.9. It then suffices to show that $\Phi \circ X = \mathbb{E}[g(Y)|\sigma(X)]$. By lemma 2.17 $(\omega, B) \mapsto \kappa_{Y|X}(X(\omega), B)$ defines a regular conditional distribution of Y given $\sigma(X)$, hence

$$\begin{aligned}(\Phi \circ X)(\omega) &= \int_E g(y)\kappa_{Y|X}(X(\omega), \mathrm{d}y) = \int_E g(y)\kappa_{Y|\sigma(X)}(\omega, \mathrm{d}y) \\ &= \mathbb{E}[g(Y)|\sigma(X)](\omega)\end{aligned}$$

That is, a version of $\mathbb{E}[g(Y)|X = x]$ is given by the mean of $g(Y)$ w.r.t. $\kappa_{Y|X}(x,\,\cdot\,)$ which we can think of as a conditional mean.

where the last equality follows from theorem 3.7. The main obstruction in semi-regular case is that Φ can only be defined as above for $x \in X(\Omega)$, where $\kappa_{Y|X}(x,\,\cdot\,)$ is a measure. But $X(\Omega)$ is measurable and $\Phi|_{X(\Omega)}$ is $X(\Omega) \cap \mathcal{E}'$-measurable. Extending $\Phi|_{X(\Omega)}$ to a measurable map $\widehat{\Phi}\colon E' \to \overline{\mathbb{R}}$, yields a version of $\mathbb{E}[g(Y)|X = \cdot\,]$. □

In analogy to theorem theorem 3.9, the conditional expectation of $g(X, Y)$ given $X = x$ may be computed by replacing X by x.

Proposition 3.14. *Let $g\colon E' \times E \to \overline{\mathbb{R}}$ measurable such that $g(X,Y) \in \mathcal{L}^1(\mathbb{P})$ and let $\kappa_{Y|X}$ a regular conditional distribution of Y given X. Then a version of $\mathbb{E}[g(X,Y)|X = \cdot\,]$ is given by*

$$\begin{aligned}\mathbb{E}[g(X,Y)|X = x] &= \int_E g(x,y)\kappa_{Y|X}(x, \mathrm{d}y) \\ &= \mathbb{E}[g(x,Y)|X = x]\end{aligned} \tag{6}$$

for $\mathbb{P}_X$-a.e. $x \in E'$.

If $X(\Omega)$ is measurable, the above integral defines (for $x \in X(\Omega)$, with arbitrary measurable extension to E') a version of $\mathbb{E}[g(X,Y)|X = \cdot\,]$ even when $\kappa_{Y|X}$ is semi-regular.

3.3 Conditional densities

We now introduce conditional analogues of the density and cdf of distributions on $\mathbb{R}^n$.

Definition 3.15. Let $\mathcal{F} \subset \mathcal{A}$ a sub-σ-algebra and $Y\colon \Omega \to \mathbb{R}^n$ a random vector with regular conditional distribution $\kappa_{Y|\mathcal{F}}\colon \Omega \times \mathcal{B}_{\mathbb{R}^n} \to [0,1]$.

(i) The **(regular) conditional cdf of Y given $\mathcal{F}$** is the map $F\colon \Omega \times \mathbb{R}^n \to [0,1]$ given by

$$F(\omega, y) \coloneqq \kappa_{Y|\mathcal{F}}(\omega, \{\, x \in \mathbb{R}^n : x \leq_n y \,\})$$

(ii) A map $f_{Y|\mathcal{F}}\colon \Omega \times \mathbb{R}^n \to \overline{\mathbb{R}}_+$ is called the **(regular) conditional density of Y given $\mathcal{F}$** if

  (a) $f_{Y|\mathcal{F}}(\omega, \cdot\,)$ is Borel-measurable for all $\omega \in \Omega$,
  (b) $f_{Y|\mathcal{F}}(\,\cdot\,, y)$ is $\mathcal{F}$-measurable for each $y \in \mathbb{R}^n$, and
  (c) for every $\omega \in \Omega$, $\kappa_{Y|\mathcal{F}}(\omega, \cdot\,)$ has density $f_{Y|\mathcal{F}}(\omega, \cdot\,)$, i.e. $\kappa_{Y|\mathcal{F}}(\omega, B) = \int_B f_{Y|\mathcal{F}}(\omega, y)\,\mathrm{d}y$ for all $B \in \mathcal{B}_{\mathbb{R}^n}$

From now on, let X a random variable taking values in some measurable space $(E', \mathcal{E}')$ such that Y has a conditional density $f_{Y|\sigma(X)}$ given $\sigma(X)$.

Definition 3.16. Let $\kappa_{Y|X}$ a regular conditional distribution of Y given X.

(i) A map $f_{Y|X}\colon \mathbb{R}^n \times E' \to \overline{\mathbb{R}}_+$ is called (a version of) the **regular conditional density** of Y given X (for the particular version of $\kappa_{Y|X}$), if

  (a) $f_{Y|X}(\,\cdot\,|x)$ is Borel-measurable for all $x \in E'$,
  (b) $f_{Y|X}(y|\,\cdot\,)$ is $\mathcal{E}'$-measurable for each $y \in \mathbb{R}^n$, and
  (c) for every $x \in E'$, $\kappa_{Y|X}(x, \cdot\,)$ has density $f_{Y|X}(x, \cdot\,)$, i.e. $\kappa_{Y|X}(x, B) = \int_B f_{Y|X}(x, y)\,\mathrm{d}y$ for all $B \in \mathcal{B}_{\mathbb{R}^n}$

(ii) If $\kappa_{Y|X}$ is semi-regular and $f_{Y|X}$ satisfies (b) and violates (a) and (c) above at most for $x \notin X(\Omega)$, it is called a **semi-regular conditional density** of Y given X

Remark 3.17. Since $f_{Y|\sigma(X)}(\,\cdot\,, y)$ is $\sigma(X)$-measurable, by lemma 2.1 there exists a factorizing map $\eta_y\colon E' \to \overline{\mathbb{R}}_+$, also denoted as $\eta_y = f_{Y|\sigma(X)}(\,\cdot\,, y) \circ X^{-1}$, such that $f_{Y|\sigma(X)}(\,\cdot\,, y) = \eta_y \circ X$.

Lemma 3.18. *(i) If $f_{Y|\sigma(X)}$ is a regular conditional density of Y given $\sigma(X)$, then any factorizing map $f_{Y|X}(\,\cdot\,, y) \coloneqq f_{X|\sigma(X)}(\,\cdot\,, y) \circ X^{-1}$ is a semi-regular conditional density of Y given X. In addition, if $X(\Omega)$ is measurable, there exists a regular conditional density for some r.c.d. $\kappa_{Y|X}$ of Y given X.*

(ii) For any semi-regular conditional density $f_{Y|X}$ of Y given X, the map $(\omega, y) \mapsto f_{Y|X}(y|X(\omega))$ is a regular conditional density of Y given $\mathcal{F}$.

Proof. (i) We first check (a)-(c) from definition 3.16. Let η_y the factorizing map from remark 3.17.

(b) Measurability of $f_{Y|X}(y|\,\cdot\,)$ follows from the factorization lemma.
(a,c) For $x = X(\omega_x) \in X(\Omega)$, $f_{Y|X}(\,\cdot\,|x) = \eta_\cdot(X(\omega_x)) = f_{Y|\sigma(X)}(\omega_x, \,\cdot\,)$, which is Borel-measurable, and it holds that

$$\begin{aligned}\kappa_{Y|X}(x, B) &= \kappa_{Y|\sigma(X)}(\omega_x, B)\\ &= \int_B f_{Y|\sigma(X)}(\omega_x, y)\,\mathrm{d}y = \int_B f_{Y|X}(y|x)\,\mathrm{d}y\,,\end{aligned}$$

so $f_{Y|X}$ is a semi-regular conditional density as claimed.

If $X(\Omega)$ is measurable, we can, by (the proof of) lemma 2.17, modify a r.c.d. $\kappa_{Y|X}$ of Y given X to equal some arbitrary measure with density f_0 w.r.t. the Lebesgue measure for fixed $x \in X(\Omega)^c$ and thereby obtain a different version of the r.c.d. of Y given X. Similarly, we can modify a semi-regular conditional density $f_{Y|X}$ to equal f_0 on $X(\Omega)^c$. These modifications preserve the required measurability properties, such that the modified density is a regular conditional density for the modified regular conditional distribution.

(ii) Let $f_{Y|X}$ a semi-regular conditional density and $f(\omega, y) := f_{Y|X}(y|X(\omega))$. Clearly $f(\omega, \,\cdot\,)$ is Borel-measurable, and similarly $f(\,\cdot\,, y)$ is $\sigma(X)$-measurable. For $\omega \in \Omega$ and $B \in \mathcal{B}_{\mathbb{R}^n}$,

$$\int_B f_{Y|X}(y|X(\omega))\,\mathrm{d}y = \kappa_{Y|X}(X(\omega), B),$$

which is a r.c.d. of Y given $\sigma(X)$ by lemma 2.17, proving the claim.

□

Corollary 3.19. *Let $g\colon E \to \overline{\mathbb{R}}$ measurable such that $g(Y) \in \mathcal{L}^1(\mathbb{P})$, and let $f_{Y|X}$ a regular conditional density of Y given X. Then a version of the conditional expectation of $g(Y)$ given $X = x$ is given by*

$$\mathbb{E}[g(Y)|X = x] = \int_{\mathbb{R}^n} g(y) f_{Y|X}(y|x)\,\mathrm{d}y\,.$$

If $X(\Omega)$ is measurable, the above integral (for $x \in X(\Omega)$ with measurable extension to E') defines a conditional distribution of Y given $X = x$ even for a semi-regular density $f_{Y|X}$.

Proof. In the regular case, by definition, (a version of) $\kappa_{Y|X}(x, \,\cdot\,)$ has density $f_{Y|X}(x, \,\cdot\,)$ for all $x \in E'$. Thus $\int_{\mathbb{R}^n} g(y) f_{Y|X}(y|x)\,\mathrm{d}y = \int_{\mathbb{R}^n} g(y)\kappa_{Y|X}(x, \mathrm{d}y)$ and the claim follows from proposition 3.13.
Now consider the semi-regular case and assume $X(\Omega)$ is measurable. The map $\Phi\colon X(\Omega) \ni x \mapsto \int_{\mathbb{R}^n} g(y) f_{Y|X}(y|x)\,\mathrm{d}y$ is $(\mathcal{E}' \cap X(\Omega))$-$\mathcal{B}_{\overline{\mathbb{R}}}$-measurable since this holds for $g = \mathbb{1}_B$, $B \in \mathcal{B}_{\mathbb{R}^n}$ and extends using a monotone class argument. Then by measurability of $X(\Omega)$, Φ is $\mathcal{E}'$-measurable. Hence Φ can be extended to a measurable map $\widehat{\Phi}\colon (E', \mathcal{E}') \mapsto (\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$. For $x \in X(\Omega)$, $\widehat{\Phi}(x) = \int_{\mathbb{R}^n} g(y)\kappa_{Y|X}(x, \mathrm{d}y)$, from which the result follows using proposition 3.13. □

Corollary 3.20. *It holds that*

$$\mathbb{E}[g(Y)] = \int_{E'} \int_{\mathbb{R}^n} g(y) f_{Y|X}(y|x)\,\mathrm{d}y\,\mathrm{d}\mathbb{P}_X(x)\,.$$

Proof. This follows from lemma 3.12 together with corollary 3.19. □

We now consider the important case where $X = (X_1, \dots, X_m)$ is a random vector in $\mathbb{R}^m$, $Y = (Y_1, \dots, Y_n)$ is a random vector in $\mathbb{R}^n$ and (X, Y) has a joint density f. In this case, there exists a regular conditional density of Y given $\sigma(X)$, which can be explicitly computed from the joint density f.

Theorem 3.21. *Let X, Y random vector taking values in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively and suppose that (X, Y) has a density $f\colon \mathbb{R}^{m+n} \to \overline{\mathbb{R}}_+$. Then there exists a regular conditional density $f_{Y|\sigma(X)}\colon \Omega \times \mathbb{R}^n \to \overline{\mathbb{R}}_+$ of Y given $\sigma(X)$ and (a version of it) is given by*

$$f_{Y|\sigma(X)}(\omega, y) = \begin{cases} \dfrac{f(X(\omega), y)}{f_X(X(\omega))} & f_X(X(\omega)) > 0, \\ f_0(y) & \text{otherwise}, \end{cases} \tag{7}$$

where f_0 is an arbitrary density and $f_X = \int_{\mathbb{R}^n} f(\,\cdot\,, y)\,\mathrm{d}y$ denotes the marginal pdf of X.

Proof. By joint measurability of f, f_X is (jointly) measurable. Since $f_{Y|\sigma(X)}$ is constructed from f, f_X in an elementary way, it is $(\sigma(X) \otimes \mathcal{B}_{\mathbb{R}^n})$-measurable. Hence $\kappa(\omega, B) := \int_B f_{Y|\sigma(X)}(\omega, y)\,\mathrm{d}y$ defines a stochastic kernel from Ω to $\mathbb{R}^n$. To verify that $f_{Y|\sigma(X)}$ is a regular conditional density, it suffices to check that κ is a version of the conditional distribution, i.e. that for all $B \in \mathcal{B}_{\mathbb{R}^n}$,

$$\kappa(\omega, B) = \mathbb{P}[\{\, Y \in B \,\}\,|\sigma(X)](\omega).$$

Let $N = \{\, x \in \mathbb{R}^m : f_X(x) = 0 \,\}$, which is $\mathbb{P}_X$-negligible. For all $A \in \mathcal{B}_{\mathbb{R}^m}$,

$$\begin{aligned} \mathbb{E}\left[\kappa(\,\cdot\,, B)\,\mathbb{1}_{\{\, X\in A \,\}}\right] &= \mathbb{E}\left[\mathbb{1}_{\{\, X\in A\cap N^c \,\}} \int_B f_{Y|\sigma(X)}(\,\cdot\,, y)\,\mathrm{d}y\right] \\ &= \int_{A\cap N^c} f_X(x) \int_B f(x,y)/f_X(x)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{A\cap N^c} \int_B f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \mathbb{P}[X \in A, Y \in B], \end{aligned}$$

proving the claim. □

From this, we obtain a regular conditional density of Y given X.

Proposition 3.22. *In the setting of theorem 3.21, a regular conditional density of Y given X is given by*

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{f(x, y)}{f_X(x)} & \text{if } f_X(x) > 0 \\ f_0(y) & \text{otherwise}, \end{cases} \tag{8}$$

where f_0 is some arbitrary density.

Proof. The density f is jointly measurable and f_X is measurable, hence $N = \{\, f_X = 0 \,\}$ is measurable. As a consequence, $f_{Y|X}$ is jointly measurable. For $x \in \mathbb{R}^m$ and $B \in \mathcal{B}_{\mathbb{R}^n}$, let $\kappa(x, B) = \int_B f_{Y|X}(y|x)\,\mathrm{d}y$. By joint measurability of $f_{Y|X}$, κ is a stochastic kernel, and it holds that $\kappa(X(\omega), B)$ equals the r.c.d. of Y given $\sigma(X)$ constructed in (the proof of) theorem 3.21. Thus κ is a r.c.d. of Y given X, proving the claim. □

Remark 3.23. (i) Note how the formula (8) for the conditional density formally resembles Bayes' rule:

$$\underbrace{f_{Y|X}(y|x)}_{\mathbb{P}[Y=y|X=x]}\ \underbrace{f_X(x)}_{\mathbb{P}[X=x]} = \underbrace{f(x,y)}_{\mathbb{P}[X=x\wedge Y=y]}$$

for $x \in \{\, f_X > 0 \,\}$, i.e. $\mathbb{P}_X$-almost surely.

(ii) The conditional density (8) agrees with the one found in lemma 1.1, motivating the following "empirical" interpretation: If, for $\varepsilon > 0$, we were to sample (X, Y) repeatedly in order to obtain an estimate of the conditional probabilities $p_\varepsilon(B) := \mathbb{P}[Y \in B \,|\, \|X - x\| \le \varepsilon]$, then $\lim_{\varepsilon\to 0} p_\varepsilon(B)$ exists (for any "reasonably nice" joint density f) and agrees with $\int_B f_{Y|X}(y|x)\,\mathrm{d}y = \kappa_{Y|X}(x, B)$.

TODO: update with proposition 1.2

Example 3.24 (Conditioning in the multivariate Gaussian setting). Let (X, Y) a centered Gaussian vector taking values in $\mathbb{R}^m \times \mathbb{R}^n$. It is known (cf. [5, §1.3]) that

$$\mathbb{E}[Y|\sigma(X)] = p_X(Y),$$

where p_X denotes the orthogonal projection of Y onto the (closed) subspace $\langle X\rangle := \operatorname{span}\{X_1, \ldots, X_m\}$ of the Gaussian space generated by $X_1, \ldots, Y_n$. For computing $p_X(Y)$, an orthonormal basis of $\langle X\rangle$ is helpful. Assuming that the covariance Σ_X of X is invertible (otherwise we can consider a subset of components of X generating the same σ-algebra), we can take $Z := \sqrt{\Sigma_X^{-1}}X$, which is a Gaussian vector with independent components spanning $\langle X\rangle$ and generating $\sigma(X)$. Then for $i \in [n]$,

The matrix square root can be computed from a diagonalization of the covariance.

$$p_X(Y_i) = \sum_{j=1}^m \langle Y_i, Z_j\rangle_{L^2} Z_j = \sum_{j=1}^m \operatorname{Cov}[Y_i, Z_j] Z_j,$$

i.e. $p_X(Y) = \Sigma_{Y,Z} Z$ where $\Sigma_{Y,Z} = \operatorname{Cov}[Y, Z] = \mathbb{E}[YZ^T] = \mathbb{E}[YX^T]\sqrt{\Sigma_X^{-1}} = \Sigma_{Y,X}\sqrt{\Sigma_X^{-1}}$. Combining, we find

$$p_X(Y) = \Sigma_{Y,X}\Sigma_X^{-1}X = AX, \qquad A = \Sigma_{Y,X}\Sigma_X^{-1}.$$

If (X, Y) is Gaussian but not necessarily centered with mean $\mathbb{E}[(X, Y)] = (\mu_X, \mu_Y)$, applying the above to the centered Gaussian vector $(\widetilde{X}, \widetilde{Y}) = (X, Y) - (\mu_X, \mu_Y)$ implies

$$p_X(Y) = \mu_Y + A(X - \mu_X) = \mathbb{E}[Y|\sigma(X)]. \tag{9}$$

We now compute the r.c.d. of Y given $\sigma(X)$ and X. Let $Z = Y - p_X(Y)$, which is independent from $\sigma(X)$. Then for any nonnegative measurable g, it holds that

$$\mathbb{E}[g(Y)|\sigma(X)] = \mathbb{E}[g(Z + p_X(Y))|\sigma(X)] = \int g(p_X(Y) + z)\,\mathrm{d}\mathbb{P}_Z(z)\,.$$

Since Z is centered Gaussian, $\mathbb{P}_Z$ has a density

$$f_{0,\Sigma}(z) = |2\pi\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\langle z, \Sigma^{-1}z\rangle\right),$$

with the "unexplained variance"

$$\Sigma := \operatorname{Cov}[Z] = \operatorname{Cov}[Y - AX] = \Sigma_Y - \Sigma_{Y,X}\Sigma_X^{-1}\Sigma_{X,Y},$$

where the last equality follows from a simple calculation. Taking $g = \mathbb{1}_A$ for $A \in \mathcal{B}_{\mathbb{R}^n}$, we find

$$\mathbb{P}[\{Y \in A\}|\sigma(X)] = \int_A f_{p_X(Y),\Sigma}(z)\,\mathrm{d}z\,,$$

that is, a regular conditional density of Y given $\sigma(X)$ is given by $f_{Y|\sigma(X)}(\omega, y) = f_{p_X(Y)(\omega),\Sigma}(y)$. With (9), we have explicitly

$$\begin{aligned} f_{Y|\sigma(X)}(\omega, y) &= |2\pi\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\langle y - \mu_{Y|X}, \Sigma^{-1}(y - \mu_{Y|X})\rangle\right) \\ &= f_{\mu_{Y|X},\Sigma}(y) \end{aligned} \tag{10}$$

with the conditional mean

$$\mu_{Y|X} = p_X(Y) = \mu_Y + A(X - \mu_X).$$

Observe that the mean μ_Y gets corrected by a multiple of the difference between X and its mean μ_X. Similarly the covariance Σ_Y gets corrected by $\Sigma_{Y,X}\Sigma_X^{-1}\Sigma_{X,Y}$, although this does not depend on X. From (10), we obtain the conditional density of Y given X,

$$f_{Y|X}(\,\cdot\,|x) = f_{\mu_{Y|X=x},\Sigma}(\,\cdot\,)$$

with the conditional mean given $X = x$

$$\mu_{Y|X=x} = \mu_Y + A(x - \mu_X).$$

Example 3.25. Let $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$ independent Gaussians. We compute the r.c.d. of $Y \coloneqq X_1$ given $X \coloneqq X_1 + X_2$. Let $\sigma_X^2 = \mathbb{V}(X) = \sigma_1^2 + \sigma_2^2$ and $\sigma_{X,Y} = \mathrm{Cov}[X,Y] = \mathrm{Cov}[X_1 + X_2, X_1] = \sigma_1^2$. Then in the terminology of example 3.24, $A = \frac{\sigma_{X,Y}}{\sigma_X^2} = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$, $\mu_{Y|X=x} = \mu_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}(x - \mu_1 - \mu_2)$ and $\Sigma = \mathbb{V}(Y) - \mathrm{Cov}[X,Y]^2/\,\mathbb{V}(X) = \sigma_1^2 - \frac{\sigma_1^4}{\sigma_1^2+\sigma_2^2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$. Writing $\mu_x = \mu_{Y|X=x}$ and $\sigma_x^2 = \Sigma$, we have

$$f_{Y|X}(y|x) = f_{\mu_x,\sigma_x^2}(y)$$

and thus

$$\mathbb{P}[\{\, Z_1 \in \cdot\,\}\,|Z_1 + Z_2 = x] = \mathcal{N}(\mu_x, \sigma_x^2).$$

4 Defining distributions with conditionality

In the examples so far, we considered the case of a pair (X, Y) of random variables of known distribution and were interested in finding a regular conditional distribution of Y given X. In this section, we consider the case where we know the marginal distribution of of X and the conditional distribution of Y given X and seek a joint distribution.

Definition 4.1. Let $(\Omega_1, \mathcal{A}_1, \mu)$ a finite measure space, $(\Omega_2, \mathcal{A}_2)$ a measurable space and let κ a finite transition kernel from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$. Then

$$(\mu \otimes \kappa)(A) = \int_{\Omega_1} \mu(\mathrm{d}\omega_1) \int_{\Omega_2} \kappa(\omega_1, \mathrm{d}\omega_2)\, \mathbb{1}_A(\omega_1, \omega_2), \qquad A \in \mathcal{A}_1 \otimes \mathcal{A}_2$$

defines a σ-finite measure on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ called the product of μ and κ. If μ is a probability measure and κ is a stochastic kernel, $\mu \otimes \kappa$ is a probability measure.

Notation. If a random variable Y has a regular conditional distribution κ given X, we write $Y \mid X = x \sim \kappa(x, \,\cdot\,)$.

The following lemma guarantees that the pair of marginal and (regular) conditional distributions uniquely specifies the joint distribution, which is given by their product in the above sense.

Lemma 4.2. *Let $(E', \mathcal{E}', \mu)$ a probability space, $(E, \mathcal{E})$ a measurable space and κ a stochastic kernel from E' to E. Denote X and Y the canonical projections from $E' \times E$ onto the first and second coordinates, respectively. Then the product $\mu \otimes \kappa$ is the unique probability measure on $(E' \times E, \mathcal{E}' \otimes \mathcal{E})$ w.r.t. which $X \sim \mu$ and $Y|X = x \sim \kappa(x, \,\cdot\,)$*

Proof. Clearly $\mathbb{P} \coloneqq \mu \otimes \kappa$ is a probability measure, and for any $A \in \mathcal{E}'$, it holds that $\mathbb{P}[X \in A] = \mu(A)$. For $B \in \mathcal{E}$,

$$\begin{aligned}
\mathbb{P}[X \in A, Y \in B] &= \int_{E'} \mu(\mathrm{d}x)\, \mathbb{1}_A(x) \int_E \kappa(x, \mathrm{d}y)\, \mathbb{1}_B(y) \\
&= \int_{E'} \mu(\mathrm{d}x)\, \mathbb{1}_A(x)\kappa(x, B) \\
&= \mathbb{E}[\mathbb{1}_A(X)\kappa(X, B)],
\end{aligned}$$

i.e. $\mathbb{P}[\{\, Y \in B \,\}\,|\sigma(X)](\,\cdot\,) = \kappa(X(\,\cdot\,), B)$. Thus, w.r.t. $\mathbb{P}$, κ is indeed a regular conditional distribution of Y given X, as claimed. Uniqueness follows from a probability measure on $\mathcal{E}' \otimes \mathcal{E}$ being uniquely defined by its values on $\mathcal{E}' \times \mathcal{E}$. ☐

Corollary 4.3. *If Y has a regular conditional distribution $\kappa_{Y|X}$ given X, then the law of (X, Y) is given by $\mathbb{P}_X \otimes \kappa_{Y|X}$.*

Corollary 4.4. *If X has a density f_X w.r.t. λ^d and Y has a regular conditional density $(x, y) \mapsto f_{Y|X}(y|x)$ given X, then (X, Y) has a density given by $f(x, y) = f_{Y|X}(y|x)f(x)$.*

Proof. This follows, since in this case $\mu \otimes \kappa$ has f as its density. □

Remark 4.5. It is also of interest to consider (families of) random variables with multiple levels of conditionality, as is the case with Markov chains. The main difference here is that the joint distribution is given by a product of one probability measure (for the initial state) and multiple stochastic kernels (one for each level of conditional dependence). For details, see [4, § 14].

Example 4.6. Let $X \sim \mathcal{U}_{[0,1]}$ and $Y|X = x \sim \mathcal{N}(x, 1)$. Since X has a density $f_X(x) = \mathbb{1}_{[0,1]}(x)$ and $Y|X = x$ has the conditional density $f_{Y|X}(y|x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}(y-x)^2)$, the pair (X, Y) has the joint density

$$f(x, y) = \frac{\mathbb{1}_{[0,1]}(x)}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - x)^2\right)$$

Figure 4: The joint density f from example 4.6

5 Appendix

Further reading: see [6] and [7, 8].

References

[1] Gerald B Folland. *Real analysis: modern techniques and their applications.* Vol. 40. John Wiley & Sons, 1999.

[2] *Borel's Paradox - Conditional Probability is not Well-Defined.* URL: `https://youtu.be/PL_L-Sv9p5M`.

[3] Gordan Zitkovic. *Lecture notes: Theory of Probability I.* 2013. URL: `https://web.ma.utexas.edu/users/gordanz/notes/conditional_expectation.pdf`.

[4] Achim Klenke. *Wahrscheinlichkeitstheorie.* Vol. 1. Springer, 2006.

[5] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus.* Springer, 2016.

[6] Joseph T Chang and David Pollard. "Conditioning as disintegration". In: *Statistica Neerlandica* 51.3 (1997), pp. 287–317. URL: `http://www.stat.yale.edu/~jtc5/papers/ConditioningAsDisintegration.pdf`.

[7] *Regular conditional distribution vs conditional distribution.* URL: `https://stats.stackexchange.com/questions/531705/regular-conditional-distribution-vs-conditional-distribution`.

[8] *Formal definition of conditional probability.* URL: `https://math.stackexchange.com/questions/496608/formal-definition-of-conditional-probability/498338#498338`.