QUICK REMAINDERS

1. A quick remainder on conditional expectations

Definitions and first properties. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space, namely Ω is a non-empty set, $\mathfrak{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on $(\Omega, \mathfrak{F})$. In this framework, specifying the σ-algebra $\mathfrak{F}$ is a way to precisely asses how much information is available in our model space Ω, in other words what are the measurable sets or equivalently the measurable functions. As such, it is nor surprising that measurable sets are called *events* in this framework, and measurable functions are called *random variables* or *observables*.

Now, let $\mathfrak{G}$ be a σ-algebra coarser than $\mathfrak{F}$, namely $\mathfrak{G} \subset \mathfrak{F}$. One may for instance think that $\mathfrak{F}$ and $\mathfrak{G}$ corresponds to the σ-algebra of events regarding some phenomenon, relative to two different persons, the first one having more information than the second (so that they can see more *events* or *measure* more things than the second one). If we know the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, there is a natural way to equip the space $(\Omega, \mathfrak{G})$ with a probability measure, which is just the restriction of $\mathbb{P}$ to $\mathfrak{G}$. In other words $(\Omega, \mathfrak{G}, \mathbb{P}_{|\mathfrak{G}})$ is the probability space modeling the same random phenomenon under a smaller amount of information.

So, given $\mathfrak{G} \subset F$, we can create, somehow trivially, a corresponding probability space. Can we also associate to a random variables on $(\Omega, \mathfrak{F}, \mathbb{P})$ some corresponding random variables on $(\Omega, \mathfrak{G}, \mathbb{P}_{|\mathfrak{G}})$? Say that X is an $\mathfrak{F}$-measurable function $X : \Omega \to \mathbb{R}$, can we define a $\mathfrak{G}$-measurable function $X' : \Omega \to \mathbb{R}$ that somehow correspond to X? If we get back to our idea that $\mathfrak{F}$ and $\mathfrak{G}$ correspond to the different amount of information of two people observing the same phenomenon, and X is a function depending on the phenomenon, a natural way to do it, is to take X' so that it somehow coincides in some weak sense with X when only events in $\mathfrak{G}$ are considered (see below).

Example 1. *We roll two cubic dice, and we receive an amount of roubles X given by the sum of the squares of their results. We can model this game in several ways (as probability is concerned with quantities that are defined regardless of the space we use to represent them), for instance we can take $\Omega := \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$ corresponding to the results of the two dice, $\mathfrak{F}$ to be the σ-algebra of the power set of Ω (any subset is measurable), $\mathbb{P}$ to be uniform and*

$$X : \Omega \to \mathbb{R}$$
$$X(i, j) = i^2 + j^2$$

corresponding to the payoff.

Now suppose that some observer has less information than us, they can only see the result of the first die. It is still useful to consider the same space Ω but with a small σ-algebra $\mathfrak{G}$, corresponding to the observations of the first die. Namely a set $E \subset \Omega$ is $\mathfrak{G}$-measurable iff $E = A \times \{1, 2, 3, 4, 5, 6\}$ for some $A \subset \{1, 2, 3, 4, 5, 6\}$. Now, this observer cannot know the value of the payoff, since they do not see the second die. The best guess they can have, is some X' that should be $\mathfrak{G}$-measurable (namely only depend on the result of the first die) and somehow corresponding to the real payoff X averaged on all the possible results of the second die. Namely we look for X' such that for all the results i of the first die

$$\mathbb{E}[X' \mathbf{1}_{\textit{first die}=i}] = \mathbb{E}[X \mathbf{1}_{\textit{first die}=i}] \tag{1}$$

Since the payoff associated to the second die is $\sum_{j=1}^{6} j^2/6 = 91/6$, we have that $X'(i, j) = i^2 + 91/6$ satisfies (1) and is $\mathfrak{G}$-measurable (depending only on the result i of the first die and not on j).

We can easily generalize the above idea to any probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and σ-algebra $\mathfrak{G} \subset \mathfrak{F}$ and $\mathfrak{F}$-measurable $X : \Omega \to \mathbb{E}$ such that $\mathbb{E}[|X|] := \int |X|(\omega) d\mathbb{P}(\omega) < \infty$. The resulting random variable X' is called the expected value of X conditioned to $\mathfrak{G}$, and it is denoted $\mathbb{E}[X|\mathfrak{G}]$. To proceed formally, consider the following remark.

Remark 2. *Let, as above, $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and $\mathfrak{G}$ a σ-algebra with $\mathfrak{G} \subset \mathfrak{F}$. Let X be a real integrable random variable on $(\Omega, \mathfrak{F}, \mathbb{P})$. Then there exists a unique, up to a.e. equivalence, random variable $X' : \Omega \to \mathbb{R}$ that is $\mathfrak{G}$-measurable and integrable and such that*

$$\mathbb{E}[X \mathbf{1}_A] = \mathbb{E}[X' \mathbf{1}_A] \qquad \textit{for all } A \in \mathfrak{G}. \tag{2}$$

Moreover if $X_1 = X_2$ a.e., then $X'_1 = X'_2$ a.e. and $\mathbb{E}[|X'|] < \infty$.

Proof. Denote by μ the finite measure on $(\Omega, \mathfrak{G})$ given by $\mu(A) = \mathbb{E}[X \mathbf{1}_A]$. If $\mathbb{P}(A) = 0$ then $\mu(A) = 0$ so that μ is absolutely continuous w.r.t. $\mathbb{P}_{|\mathfrak{G}}$. Then $X' = \frac{d\mu}{d\mathbb{P}_{|\mathfrak{G}}}$ has the required properties due to Radon-Nikodym theorem. A.e. equivalences are easily checked. □

The above remark defines X' up to a.e. equivalence, given X up to a.e. equivalence. Thus the following definition is well-posed.

Definition 3. *Given $X \in L_1(\Omega, \mathfrak{F}, \mathbb{P})$, the conditional expectation $\mathbb{E}[X|\mathfrak{G}] \in L_1(\Omega, \mathfrak{F}, \mathbb{P})$ is defined as the unique $\mathfrak{G}$-measurable X' such that (2) holds.*

Thus the conditional expectation defines a (measurable) map on $L_1(\Omega, \mathfrak{F}, \mathbb{P})$, whose image is actually in $L_1(\Omega, \mathfrak{G}, \mathbb{P})$, and thus it can also be regarded as a map $L_1(\Omega, \mathfrak{F}, \mathbb{P}) \to L_1(\Omega, \mathfrak{G}, \mathbb{P})$.

Let's check some properties, that can be easily proved.

Remark 4. *It holds*

(1) If X is $\mathfrak{G}$-measurable, then $\mathbb{E}[X|\mathfrak{G}] = X$. In particular $\mathbb{E}[X|\mathfrak{F}] = X$.
(2) If X is independent of $\mathfrak{G}$, then $\mathbb{E}[X|\mathfrak{G}] = \mathbb{E}[X]$. In particular if $\mathfrak{G} = \{\emptyset, \Omega\}$, the $\mathbb{E}[X|\mathfrak{G}] = \mathbb{E}[X]$.
(3) If $\mathfrak{H} \subset \mathfrak{G}$, then $\mathbb{E}[[X|\mathfrak{G}]|\mathfrak{H}] = \mathbb{E}[X|\mathfrak{H}]$.
(4) If $f\colon \mathbb{R} \to \mathbb{R}$ is a convex function $f(X) \in L_1$, then $\mathbb{E}[f(\mathbb{E}[X|\mathfrak{G}]) \le \mathbb{E}[f(X)]$. In particular if $X \in L_p$ for $p \ge 1$, then $\mathbb{E}[X|\mathfrak{G}] \in L_p$.
(5) If $X \in L_2$, then

$$\mathbb{E}[(X - \mathbb{E}[X|\mathfrak{G}])^2] \le \mathbb{E}[(X - Y)^2] \tag{3}$$

for every $Y \in L_2$ that is $\mathfrak{G}$-measurable, with equality holding off $Y = \mathbb{E}[X|\mathfrak{G}]$. In particular, for square-integrable variables, the conditional expectation can be regarded as the orthogonal projection from $L_2(\Omega, \mathfrak{F}, \mathbb{P})$ to $L_2(\Omega, \mathfrak{G}, \mathbb{P})$.
(6) The conditional expectation $X' = \mathbb{E}[X|\mathfrak{G}]$ can also be characterized requiring that

$$\mathbb{E}[X\,Y] = \mathbb{E}[X'Y]$$

for all $Y \in L_\infty(\Omega, \mathfrak{G}, \mathbb{P})$, which is a slight generalization of (2).

Remark 5. *If $X \in L_1$ and Y is another random variables, then $\mathbb{E}[X|Y]$ is a short-hand notation for $\mathbb{E}[X|\mathfrak{G}]$ where $\mathfrak{G}$ is the weakest σ-algebra such that Y is $\mathfrak{G}$-measurable. Since any $\mathfrak{G}$-measurable variable is a function of Y, we have that $\mathbb{E}[X|Y] = f(Y)$ for some measurable $f\colon \mathbb{R} \to \mathbb{R}$. f is defined up to a.e. equivalence of $f(Y)$.*

Examples. Let X and Y such that the law $\theta \in \mathcal{P}(E \times F)$ of (X, Y) has a density w.r.t. Lebesgue

$$d\theta(x, y) = \varrho(x, y)dxdy$$

Given a measurable function $f\colon \mathbb{R}^2 \to \mathbb{R}$, such that $f(X, Y) \in L_1$, one has that $\mathbb{E}[f(X, Y)|Y] = g(Y)$ where g is given by

$$g(y) = \frac{\int f(x, y)\varrho(x, y)dx}{\int \varrho(x, y)dx} \tag{4}$$

Indeed, it is enough to check that (5), namely that $\mathbb{E}[f(X, y)h(Y)] = \mathbb{E}[g(Y)h(Y)]$ for each bounded measurable h, which easily holds.

At the seminar, we discussed the following examples.

Example 6. *Let $X = (X_1, X_2)$ be a Gaussian $\mathbb{R}^2$ valued random variable with $(X_1, X_2) \sim \mathcal{N}(m, S)$ where $m \in \mathbb{R}^2$ and S is an positive definite $2x2$ symmetric tensor.*

Then, by (1), $\mathbb{E}[X_1|X_2] = g(X_2)$ where

$$\mathbb{E}[X_1|X_2] = \frac{\int x_1 \exp(-\frac{1}{2}\langle S^{-1}(x_1 - m_1, x_2 - m_2), (x_1 - m_1, x_2 - m_2)dx_1\rangle}{\int \exp(-\frac{1}{2}\langle S^{-1}(x_1 - m_1, x_2 - m_2), (x_1 - m_1, x_2 - m_2)dx_1} = m_1 + \frac{S_{12}}{S_{22}}(x_2 - m_2)$$

The case where $S_{22} = 0$ can be treated separately. In this case $X_2 = m_2$ a.s., so that $\mathbb{E}[X_1|X_2] = \mathbb{E}[X_1] = m_1$.

Example 7. *Let X and Y be discrete random variables, with $\mathbb{P}(X = x_i, Y = y_j) = p_{i,j}$ for some countable values $(x_i)_{i\in\mathbb{N}}$ and $(y_i)_{i\in\mathbb{N}}$ with $y_i \ne y_j$ for $i \ne j$. Then $\mathbb{E}[X|Y] = g(Y)$ where $g(Y)$ is defined up to a.e.-equivalence, and thus it is enough to define g on the y_j as*

$$g(y_j) = \frac{\sum_i x_i p_{i,j}}{\sum_i p_{i,j}}$$

This is sometimes informally written $\mathbb{E}[X|Y = y_j] = \frac{\sum_i x_i p_{i,j}}{\sum_i p_{i,j}}$.

Example 8. *Let X and Y be discrete random variables, with $\mathbb{P}(X = x_i, Y = y_j) = p_{i,j}$ for some countable values $(x_i)_{i\in\mathbb{N}}$ and $(y_i)_{i\in\mathbb{Z}}$ with $y_i \ne y_j$ for $i \ne j$. Then $\mathbb{E}[X|Y^2] = g(Y^2)$. Again it is enough to define*

$$g(y_j) = g(y_{-j}) = \frac{\sum_i x_i(p_{i,j} + p_{i,-j})}{\sum_i p_{i,j} + p_{i,-j}}$$

Generalizations. If X takes values in a topological vector space V, one can still say that $\mathbb{E}[X] = \ell$ for some $\ell \in V^{**}$ provided $\mathbb{E}[\langle X, h\rangle] = \langle \ell, h\rangle$ for all $h \in V^*$. This idea goes through conditional expectation as $\mathbb{E}[X|\mathfrak{G}] = X'$ if X' is $\mathfrak{G}$-measurable and $\mathbb{E}[\langle X, h\rangle|\mathfrak{G}] = \langle X', h\rangle$ for all $h \in V^*$.

Another possible generalization holds for variables with values in metric spaces E, provided the $d(X, x) \in L_2$ for some fixed (any fixed) point $x \in E$. One may say in this case that $X' = \mathbb{E}[X|\mathfrak{G}]$ if

$$\mathbb{E}[d(X, X')^2] \leq \mathbb{E}[d(X, Y)^2]$$

for all $\mathfrak{G}$-measurable Y such that $d(Y, x) \in L_2$.

On $\mathbb{R}^n$ or Hilbert spaces these definitions coincide, and on separable Hilbert spaces they can be constructed by standard integration tools. All the elementary properties are then easily verified. In complete generality however, such conditional expectation may fail to exist or to be uniquely defined.

In any case, even when variable takes value in rather wild spaces, one may often be interested in conditional expectations of some functions of the variable.

2. A quick remainder on change of variables

Definitions and basic properties. If $(E, \mathfrak{E})$ and $(F, \mathfrak{F})$ are two measurable spaces, $\pi \colon E \to F$ is a measurable function and μ is a measure on E, we can build a new measure μ on F simply composing $\nu := m \circ \pi^{-1}$. This means that for a set A in $\mathfrak{F}$ we define

$$\nu(A) := \mu(\pi^{-1}(A)) = \mu(\{e \in E\, : \pi(e) \in A\}$$

Notice that

- If μ is additive, ν is additive.
- If μ is σ-additive, ν is σ-additive.
- If μ is σ-finite, ν is σ-finite.
- If μ is positive, ν is positive.
- If μ is a probability, ν is a probability.
- If μ is discrete, ν is discrete.
- If μ is absolutely continuous w.r.t. μ', then ν is absolutely continuous w.r.t. $\nu' = \mu' \circ \pi^{-1}$.

This is the unique canonical way to lift π to a map $\pi_\sharp$ from measures on E to measures on F. $\pi_\sharp$, defined by $\pi_\sharp\mu := \mu \circ \pi^{-1}$ is called the *pushforward* of μ via π. Notice that $\pi_\sharp\delta_x = \delta_{\pi(x)}$ and $\pi_\sharp(\alpha\mu_1 + (1-\alpha)\mu_2 = \alpha\pi_\sharp\mu_1 + (1-\alpha)\mu_2$.

If X is random variables taking values in E and μ is its law, then the random variable $Y = \pi(X)$ has law $\mu \circ \pi^{-1}$, as indeed $\mathbb{P}(Y \in B) = \mathbb{P}(X \in \pi^{-1}(B)) = \mu(\pi^{-1}(B))$.

Change of variables. In the same framework as above, the change of variables formula (2) below holds in full generality. It is indeed enough to (trivially) check it on simple functions, and then to extend it to any integrable function. If $f \colon F \to \mathbb{R}$ is measurable, then $f \circ \pi \in L_1(\mu)$ iff $f \in L_1(\mu \circ \pi^{-1})$ and in such a case

$$\int_E f(\pi(x))\, d\mu(x) = \int_F f(y) d(\mu \circ \pi^{-1})(y) \tag{5}$$

One can actually calculate $\mu \circ \pi^{-1}$ explicitly (явно) in some cases. The first two examples follow from direct computations.

Example 9. *If π is constant, $\pi(x) = c \in F$ for every x, then for every probability measure μ on E it holds $\mu \circ \pi^{-1} = \delta_c$.*

More generally, if π takes countable many values, namely there is a countable partition of $E = E_1 \cup E_2 \cup \dots$ such that $\pi(x) = y_i$ for all $x \in E_i$, then $mu \circ \pi^{-1} = \sum_i \beta_i \delta_{y_i}$ where $\beta_i = \mu(E_i)$.

Example 10. *If μ is discrete, namely it is concentrated on a countable set say $\mu = \sum_i \alpha_i \delta_{x_i}$, then $\mu \circ \pi^{-1} = \sum_i \alpha_i \delta_{\pi(x_i)}$.*

Example 11. *Another class of functions for which it is possible to calculate the pushforward explicitely are differentiable functions between manifold. Suppose that $\pi \colon \mathbb{R}^n \to \mathbb{R}^n$ is injective, with smooth left inverse π^{-1}. Assume that μ has a density w.r.t. the Lebesgue measure on $\mathbb{R}^n$, $d\mu(x) = \varrho(x)dx$. Recall that this just means*

$$\mu(A) = \int_A \varrho(x) dx \qquad \textit{for every measurable } A \subset \mathbb{R}^n.$$

Then

$$\mu \circ \pi^{-1}(B) = \int_{\pi^{-1}(B)} \varrho(x) dx = \int_B \varrho(\pi^{-1}(y))|J_{\pi^{-1}}|(y)\, dy$$

where $|J_{\pi^{-1}}|$ is the absolute value of the Jacobian determinant of π^{-1}. Thus, $\mu \circ \pi^{-1}$ has density $\varrho(\pi^{-1})|J_{\pi^{-1}}|$.

General change of variables. Actually (2) has a more general version to represent $\int f(x)d\mu(x)$, even if f in a generic function of x (and not of $\pi(x)$). We assume here that μ is σ-finite and E and F are Polish spaces (completely metrizable and separable), and the the σ-algebras are Borel. Then there exists a measurable map (kernel) $p\colon F \to \mathcal{P}(E)$ (that associates to an $y \in F$, a probability measure on E) such that $p(y,\cdot)$ is concentrated on $\pi^{-1}(y)$ and for each $f \in L_1(\mu)$

$$\int f(x)d\mu(x) = \int_F d(\mu \circ \pi^{-1})(y) \int_E p(y,dx)\, f(x)$$

If $E = F$, p is nothing but a Markov kernel. This is sometimes called the transfer operator associated to the transformation (or some more general algebraic action) π. But indeed, is nothing but the usual formal of change of variables!

If $\pi_1 \colon E \to G$ and $\pi_2 \colon G \to F$, and p_1 and p_2 are the respective kernels, then it is easily seem that $\pi_2 \circ pi_1$ is associated with the kernel $p(y,dx) := \int_G p_2(y,dz)\, p_1(z,dx)$. If $E = F = G$, this is the usual formula for iteration of Markov kernels or transfer operators for actions of semigroups.

Example 12. *Suppose that the law of real random variables X and Y has density $\varrho(x,y)$. Let $R = X^2 + Y^2$. Let us find the law of the random variable R. We can for instance consider the change of variable related to polar coordinates, $x = r\cos(\theta)$, $y = r\sin(\theta)$. The Jacobian is r. We gather that the density of the corresponding random variables (R,Θ) w.r.t. $dr\, d\theta$ on $\mathbb{R}^+ \times S^1$ (up to measure zero) is $r\varrho(R\cos(\theta), r\sin(\Theta))$, so that for R we obtain the density*

$$h(r) := r\,\mathbf{1}_{r\geq 0} \int_0^{2\pi} \varrho(r\cos(\theta), r\sin(\theta))\, d\theta$$

Notice that the steps we described in full generality before read in this example as follows. For every measurable $A \subset \mathbb{R}$

$$\mathbb{P}(R \in A) = \mathbb{P}(X^2 + Y^2 \in A) = \int_{\{(x,y)\,:\,x^2+y^2 \in A} \varrho(x,y)\, dx\, dy = \int_{\{(r,\theta)\,:\,r\in A,\, r\geq 0\}} r\,\varrho(r\cos(\theta), r\sin(\theta))\, dr\, d\theta = \int_A h(r)\, dr$$

Since $\mathbb{P}(R \in A)$ writes as the integral of A of some h for every measurable A, h is indeed the density of the law of $\mathbb{R}$.