On densities and conditional expectation

SML Introduction to Stochastic Calculus

Rasmus Peter Skadborg

The aim of this report is to outline the proof of existence and uniqueness of conditional expectation. Such a proof is given in Theorem 4.1 of [Bal] and in Appendix A6 of [Mik], but some notions from measure-theory are assumed by both. We shall therefore first recall these notions, and then outline in slightly more detail how they are used to prove the existence of conditional expectation.

First, we recall some definitions regarding how two measures defined on the same measurable space may relate to each other

Definition 1. *Let $(\Omega, \mathcal{F})$ be a measurable space, and let μ, ν be two measures on this space.*

(a) If there exists an $\mathcal{F}$-measurable function $f : \Omega \to [0, +\infty]$ such that

$$\nu(A) = \int_A f \mathrm{d}\mu \qquad \text{for all } A \in \mathcal{F},$$

then we say ν has density f with respect to μ. In that case, f is also known as the Radon-Nikodym derivative of ν with respect to μ, and we write $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

(b) If every μ-null set is also a ν-null set, we say that ν is absolutely continuous with respect to μ, and we write $\nu \ll \mu$.

An important result about measures with densities is that the integrals work in the way one would expect:

Proposition 2. *Let $(\Omega, \mathcal{F})$ be a measurable space, and let μ, ν be two measures on this space. Suppose that ν has density f with respect to μ. Then the ν-integral is given by*

$$\int g \mathrm{d}\nu = \int g f \mathrm{d}\mu,$$

where g can be any positive $\mathcal{F}$-measurable function $\Omega \to [0, +\infty]$, or any ν-integrable function $\Omega \to \mathbb{R}$.

Proof. The proof is by a standard approximation argument: When g is an indicator-function $\mathbb{1}_A$ where $A \in \mathcal{F}$, the result is clear, since

$$\int \mathbb{1}_A \mathrm{d}\nu = \nu(A) = \int_A f \mathrm{d}\mu = \int \mathbb{1}_A f \mathrm{d}\mu.$$

The result then follows for simple functions by linearity, for positive measurable functions by approximating from below by simple functions and applying the monotone convergence theorem, and for general integrable functions by considering the difference of two positive integrable functions. □

Let us remark that if ν has a density f with respect to μ, then it is easy to see that $\nu \ll \mu$; for if $A \in \mathcal{F}$ is a μ-null set, then we have

$$\nu(A) = \int_A f \mathrm{d}\mu = \int f \mathbb{1}_A \mathrm{d}\mu = 0,$$

since $f\mathbb{1}_A = 0$ μ-a.e.. In fact, in the case where μ and ν are both finite measures (e.g. probability measures), the converse also holds:

Theorem 3 (Radon-Nikodym Theorem)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space, and let μ, ν be two finite measures on this space. If $\nu \ll \mu$, then there exists an $\mathcal{F}$-measurable function $f : \Omega \to [0, +\infty]$ such that*

$$\nu(A) = \int_A f \mathrm{d}\mu \qquad \textit{for all } A \in \mathcal{F}.$$

Proving this theorem in full is quite a bit outside our scope, but let us briefly sketch out the main ideas (for a full proof, see any textbook on measure-theory. [Tho] was consulted for this outline, but beware that it is in Danish). The key idea is to consider the continuous functional φ on the Hilbert-space $L^2(\Omega, \mathcal{F}, \mu + \nu)$ given by

$$g \mapsto \int g \mathrm{d}\nu$$

(of course, one has to actually show that this is in fact continuous). The Riesz representation theorem then says that there exists a function $f \in L^2(\Omega, \mathcal{F}, \mu + \nu)$ such that $\varphi(g) = \langle g, f \rangle$ for any $g \in L^2(\Omega, \mathcal{F}, \mu + \nu)$, in other words

$$\int g \mathrm{d}\nu = \langle g, f \rangle = \int g\overline{f} \mathrm{d}(\mu + \nu) = \int g\overline{f} \mathrm{d}\mu + \int g\overline{f} \mathrm{d}\nu. \tag{1}$$

One can then show that $f(\omega)$ in fact takes real values in the interval $[0, 1)$ for μ-a.a ω (and therefore also for ν-a.a. ω, since $\nu \ll \mu$). Now, define two measures on $(\Omega, \mathcal{F})$ by

$$m_1(A) = \int_A (1 - f) \mathrm{d}\nu, \qquad m_2(A) = \int_A f \mathrm{d}\mu,$$

for any $A \in \mathcal{F}$. Then applying (1) with $g = \mathbb{1}_A$ (which is an element of $L^2(\Omega, \mathcal{F}, \mu + \nu)$ since $\mu + \nu$ is a finite measure), we see that $m_1 = m_2$. From this we can conclude that $\frac{f}{1-f}$ gives the required density (after modifying it on a null-set to be sure it only takes on real, non-negative values), since for any $A \in \mathcal{F}$, Proposition 2 implies

$$\begin{aligned} \int_A \frac{f}{1-f} \mathrm{d}\mu &= \int_A \frac{1}{1-f} \mathrm{d}m_2 \\ &= \int_A \frac{1}{1-f} \mathrm{d}m_1 \\ &= \int_A \frac{1-f}{1-f} \mathrm{d}\nu \\ &= \int_A 1 \mathrm{d}\nu = \nu(A). \end{aligned}$$

Another remark on Definition 1 is that we called $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ "the" Radon-Nikodym-derivative. This language use is justified by the folllowing uniqueness-result:

Proposition 4. *Let $(\Omega, \mathcal{F})$ be a measurable space, and let μ and ν be two finite measures on this space. If ν has two densities f and g with respect to μ, then $f = g$ μ-a.e.*

Proof. We have

$$\int f \mathrm{d}\mu = \int g \mathrm{d}\mu = \nu(\Omega) < \infty,$$

which implies that the functions f, g and $f - g$ are μ-integrable, and in particular that

$$\int_A (f-g)\mathrm{d}\mu = \int_A f\mathrm{d}\mu - \int_A g\mathrm{d}\mu = \nu(A) - \nu(A) = 0$$

for any $A \in \mathcal{F}$. In particular, taking $A = \{\omega \in \Omega \mid f(\omega) > g(\omega)\}$, we get

$$\int (f-g)\mathbb{1}_A \mathrm{d}\mu = \int_A (f-g)\mathrm{d}\mu = 0,$$

and therefore $\mathbb{1}_A(f-g) = 0$, μ-a.e., since $\mathbb{1}_A(f-g) \geq 0$. But then

$$0 = \mu(\{\omega \in \Omega \mid (f(\omega) - g(\omega))\mathbb{1}_A(\omega) > 0\}) = \mu(\{\omega \in A \mid f(\omega) > g(\omega)\}) = \mu(A),$$

so $f \leq g$ μ-a.e.. A similar argument shows that $f \geq g$ μ-a.e., so $f = g$ μ-a.e. □

As an aside, let us note that Theorem 3 and Proposition 4 in fact remain valid if μ and ν are only σ-finite (e.g. the Lebesgue-measure). We will not need this greater generality here, but note that our definition of absolutely continuous random variables now has a different interpretation: We defined X to be absolutely continuous if and only if the induced measure μ_X had a density with respect to the Lebesgue-measure λ, which we now know holds if and only if $\mu_X \ll \lambda$. In other words, X is absolutely continuous if and only if $\mathbb{P}(X \in A) = 0$ for every Lebesgue-null set A.

We are now ready to prove the existence and uniqueness (up to "almost surely") of conditional expectation. The following theorem is essentially "Definition and Theorem 4.1" of [Bal], though we restrict ourselves only to integrable, rather than left semi-integrable, random variables (i.e. we require that $\int |X| \mathrm{d}\mathbb{P} = \mathbb{E}[|X|] < \infty$). This is also how [Mik] states the theorem in Appendix A6. Note also that unlike Baldi (but like Mikosch), we make a clear notational distinction between the probability measure $\mathbb{P}$ on $\mathcal{F}$ and its restriction $\mathbb{P}|_{\mathcal{G}}$ to the σ-subalgebra $\mathcal{G}$. This distinction is essential to understanding both the content and proof of the theorem.

Theorem 5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability-space, and let $X : \Omega \to \mathbb{R}$ be a $\mathbb{P}$-integrable univariate random variable. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-subalgebra. Then there exists a $\mathbb{P}|_{\mathcal{G}}$-integrable univariate random variable Y on $(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$, such that for every $G \in \mathcal{G}$*

$$\int_G Y \mathrm{d}\,\mathbb{P}|_{\mathcal{G}} = \int_G X \mathrm{d}\mathbb{P}.$$

Further, if Z is another random variable on $(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$ with the same property, then $Y = Z$ almost surely (with respect to $\mathbb{P}|_{\mathcal{G}}$). It therefore makes sense to define the conditional expectation of X given $\mathcal{G}$ as $\mathbb{E}[X|\, \mathcal{G}] := Y$.

Proof. Let us first prove the theorem in the case $X \geq 0$. Define a measure μ on $(\Omega, \mathcal{G})$ by

$$\mu(G) = \int_G X \mathrm{d}\mathbb{P},$$

for $G \in \mathcal{G}$. Clearly $\mu \ll \mathbb{P}|_{\mathcal{G}}$, for if $G \in \mathcal{G}$ is a $\mathbb{P}|_{\mathcal{G}}$-null set, then clearly $\mathbb{P}(G) = \mathbb{P}|_{\mathcal{G}}(G) = 0$, so $\mu(G) = \int_G X \mathrm{d}\mathbb{P} = 0$. The fact that X is $\mathbb{P}$-integrable implies that μ is a finite measure, since

$$\mu(\Omega) = \int X \mathrm{d}\mathbb{P} < \infty.$$

$\mathbb{P}|_{\mathcal{G}}$ is also clearly a finite measure ($\mathbb{P}|_{\mathcal{G}}(\Omega) = \mathbb{P}(\Omega) = 1$), so the hypotheses for the Radon-Nikodym theorem are satisfied. We can therefore let Y be the Radon-Nikodym derivative:

$$Y := \frac{\mathrm{d}\mu}{\mathrm{d}\left(\mathbb{P}|_{\mathcal{G}}\right)},$$

which by definition is a density of μ with respect to $\mathbb{P}|_{\mathcal{G}}$. This means in particular that Y is $\mathbb{P}|_{\mathcal{G}}$-integrable, since $\int_\Omega Y \,\mathrm{d}\,\mathbb{P}|_{\mathcal{G}} = \mu(\Omega) < \infty$, and more to the point it means that for any $G \in \mathcal{G}$:

$$\int_G Y \,\mathrm{d}\,\mathbb{P}|_{\mathcal{G}} = \mu(G) = \int_G X \,\mathrm{d}\mathbb{P},$$

so Y is our required conditional expectation. Conversely, any Y satisfying the above equation would be a density for μ with respect to $\mathbb{P}|_{\mathcal{G}}$, so the uniqueness follows from Proposition 4.

For the general case, we can decompose X as $X = X^+ - X^-$, where

$$X^+ = X\mathbb{1}_{\{X>0\}}, \qquad X^- = -X\mathbb{1}_{\{X<0\}}.$$

Then X^+ and X^- are non-negative and $\mathbb{P}$-integrable, so letting

$$\mu^+(G) = \int_G X^+ \mathrm{d}\mathbb{P}, \qquad \mu^-(G) = \int_G X^- \mathrm{d}\mathbb{P}, \qquad\qquad (G \in \mathcal{G}),$$

we can apply the previous argument to see that

$$Y := \frac{\mathrm{d}\mu^+}{\mathrm{d}\left(\mathbb{P}|_{\mathcal{G}}\right)} - \frac{\mathrm{d}\mu^-}{\mathrm{d}\left(\mathbb{P}|_{\mathcal{G}}\right)}$$

is a $\mathbb{P}|_{\mathcal{G}}$-integrable random variable satisfying

$$\int_G Y \,\mathrm{d}\,\mathbb{P}|_{\mathcal{G}} = \mu^+(G) - \mu^-(G) = \int_G X \,\mathrm{d}\mathbb{P},$$

for any $G \in \mathcal{G}$. Y thus gives us our required conditional expectation.

For the uniqueness, we note that $Y^+ = Y\mathbb{1}_{\{Y>0\}}$ and $Y^- = -Y\mathbb{1}_{\{Y<0\}}$ are conditional expectations for the positive integrable random variables X^+ and X^- respectively. Therefore, by the uniqueness in the positive case, any other conditional expectation Z would have $Z^+ = Y^+$ and $Z^- = Y^-$ $\mathbb{P}|_{\mathcal{G}}$-almost surely, and thus

$$Z = Z^+ - Z^- = Y^+ - Y^- = Y \qquad \mathbb{P}|_{\mathcal{G}}\text{-almost surely.}$$

□

As a final remark, note that a statement analogous to Theorem 5 also holds for multivariate random variables, as one can simply apply the theorem to each component.

References

[Bal] Paolo Baldi. *Stochastic calculus, an introduction through theory and exercises.* Springer, 2017.

[Mik] Thomas Mikosch. *Elementary stochastic calculus with finance in mind.* World Scientific, 1998.

[Tho] Steen Thorbjørnsen. *Grundlæggende mål- og integralteori.* Aarhus Universitetsforlag, 2014.