Ma140a Probability
Conditional Probability and Expectation

Juan Pablo Vigneaux

Caltech

Department of Mathematics
California Institute of Technology
Pasadena, CA, USA
vigneaux@caltech.edu

August 3, 2024

Elementary definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and $A \in \mathcal{F}$ such that $\mathbb{P}(A) > 0$. The expression

$$\mathbb{P}_A(B) := \mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

is the conditional probability of B given A. The map $B \mapsto \mathbb{P}_A(B)$ is a probability measure on $\mathcal{F}$.

The problem: If $Z : \Omega \to \mathbb{R}$ is a continuous probability measure, the events $\{Z = z\}$ have probability zero, hence $\mathbb{P}_{\{Z=z\}}(\cdot)$ is not well defined.

An example illustrating a change of perspective

Let $\mathcal{A}$ be a finite partition of Ω into sets $\{A_1, ..., A_n\}$ with positive probability (for instance, the events $\{X = x_i\}$ for a variable X taking finitely many different values.)

One can introduce $\mathbb{P}_{\mathcal{A}}(B)$, the conditional probability of $B \in \mathcal{F}$ after the "experiment" $\mathcal{A}$ is performed, which is a random variable that maps $\omega \in \Omega$ to $\mathbb{P}_{A(\omega)}(B)$, where $A(\omega)$ is the set that contains ω.

Characterization of the conditional probability $\mathbb{P}_{\mathcal{A}}(B)$

The function $\mathbb{P}_{\mathcal{A}}(B) : \Omega \to \mathbb{R}$ is uniquely characterized by the requirement of being $\sigma(\mathcal{A})$-measurable and the condition:

$$\forall A \in \mathcal{A} \,\text{(or, equivalently, in } \sigma(A)), \quad \mathbb{P}(A \cap B) = \int_A P_{\mathcal{A}}(B) \, \mathrm{d}\mathbb{P}. \tag{1}$$

This characterization also makes sense when we replace $\sigma(\mathcal{A})$ by any sub-σ-algebra of $\mathcal{F}$.

Conditional expectation

More generally, one can introduce conditional expectations (remark that $\mathbb{P}_{\mathcal{A}}(B) = \mathbb{E}(I_B|\mathcal{A})$).

Definition

Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{A}$ be a sub-σ-algebra of $\mathbb{F}$. A version of the conditional expectation is a function $Y \equiv \mathbb{E}(X|\mathcal{A})$ in $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ such that,

$$\forall A \in \mathcal{A}, \quad \int_A X \,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{A}) \,\mathrm{d}\mathbb{P}. \tag{2}$$

Example (illustrating the intuitions coming from “elementary” probability)

Suppose X and Z discrete (taking finitely many values), then $\sigma(Z)$ has atoms $\{Z = z\}$, $Y = \mathbb{E}(X|\sigma(Z))$ is constant on each atom, and for each $\omega \in \Omega$ such that $Z(\omega) = z$, one has $Y(\omega) = \mathbb{E}(X|Z = z) := \sum_{j=1}^{N} x_j \mathbb{P}(X = x_j|Z = z)$.

Frame Title

In the general case, there might be infinitely many versions of $\mathbb{E}(X|\mathcal{A})$, but any two version agree almost surely.

Notation: if Z random variable, we set $\mathbb{E}(X|Z) = \mathbb{E}(X|\sigma(Z))$. Because it is Z measurable, it is a measurable function of Z. *However, in general, the function f cannot be determined explicitly and one has to limit oneself to use the properties of $\mathbb{E}(\cdot|\mathcal{G})$ listed below.*

Example (Again, an "elementary" case where everything is very explicit)

Let $N \sim \mathrm{Poisson}(\lambda)$ i.e. $p(k) = \lambda^k e^{-\lambda}/k!$ for $k \in \mathbb{N}$ where $\lambda > 0$.
Let $K \sim \mathrm{Bin}(N, p)$ i.e. $f_{K|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}$.
Then $\psi(n) = \mathbb{E}(K|N = n) = pn$ and $\mathbb{E}(K|N) = \psi(N) = pN$.

Conditional expectation as L^2-projection

Suppose that $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Then $Y = \mathbb{E}(X|\mathcal{G})$ is a version of the orthogonal projection of $[X] \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ on $L^2(\Omega, \mathcal{A}, \mathbb{P})$ (which is a vector subspace of the former).

Remember that L^2 is an inner product space. The orthogonal projection $\tilde{Y}$ satisfies the condition

$$\forall Z \in L^2(\Omega, \mathcal{A}, \mathbb{P}), \quad \langle X - \tilde{Y}, Z \rangle = 0. \tag{3}$$

In particular, if $Z = I_A$ for $A \in \mathcal{A}$, we get $\int_A X \,\mathrm{d}\mathbb{P} = \int_A \tilde{Y} \,\mathrm{d}\mathbb{P}$. But this seemingly weaker condition also implies (3) via the standard machine. So $Y = \mathbb{E}(X|\mathcal{A})$ is a version (representative) of the projection.

(The existence of an orthogonal projection is proved in Sec. 6.11 of Williams' book and depends on the sequential completeness of L^2; $\tilde{Y}$ can be equivalently characterized by the equality $\left\| X - \tilde{Y} \right\|_2 = \inf_{W \in L^2(\mathcal{A})} \| X - W \|_2$.)

General construction of the conditional expectation for $X \in \mathcal{L}^1$

For general $X \in \mathcal{L}^1$. We suppose that X is positive. (For general X, decompose $X = X^+ - X^-$ first, then use linearity.)

Introduce $X_n = X \wedge n$, which are bounded variables such that $X_n \uparrow X$.

For each X_n, there exists $Y_n \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ versions of $\mathbb{E}(X_n|\mathcal{A})$ introduced via projections.

Claim: $(Y_n)_n$ is a sequence of positive functions that is increasing.

Then $Y = \limsup Y_n$ is the desired function: it follows from (MON) that $\int_A Y \,\mathrm{d}\mathbb{P} = \int_A X \,\mathrm{d}\mathbb{P}$ for all $A \in \mathcal{A}$.

(Prove the claim by contradiction.)

Another proof of the existence of $\mathbb{E}(X|\mathcal{A})$

We assume again that X is positive. Then $Q : \mathcal{A} \to [0, \infty)$, $A \mapsto \int_A X \,\mathrm{d}\mathbb{P}$ is a finite measure on $\mathcal{A}$, such that $Q(\Omega) = \mathbb{E}(X)$.

Moreover, it is clear that if $\mathbb{P}(A) = 0$ then $Q(A) = 0$; in other words, Q is absolutely continuous with respect to $\mathbb{P}$. Hence, by the Radon-Nikodym theorem, there exists a function $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ such that for all $A \in \mathcal{A}$, $Q(A) = \int_A Y \,\mathrm{d}\mathbb{P}$.

The general case follows by linearity.

This was Kolmogorov's proof. However, following Williams, we'll take a different path here, and prove the Radon-nikodym theorem using conditional expectations and martingales.

Properties of the conditional expectation, see Williams 9.7

Let X be integrable and $\mathcal{G}$ and $\mathcal{H}$ denote sub-σ-algebras of $\mathcal{F}$.

1. If Y is any version of $\mathbb{E}(X|\mathcal{G})$, then $\mathbb{E}(Y) = \mathbb{E}(X)$.
2. If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$ a.s.
3. Linearity: $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ a.s.
4. Positivity: If $X \geq 0$, then $\mathbb{E}(X|\mathcal{G}) \geq 0$ a.s.
5. cMON: If $0 \leq X_n \uparrow X$, then $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$ a.s.
   Similarly there's a cFATOU and cDOM (**exercise**).
6. cJensen: If $c : \mathbb{R} \to \mathbb{R}$ is convex, and $\mathbb{E}|c(X)| < \infty$, then
   $\mathbb{E}(c(X)|\mathcal{G}) \geq c(\mathbb{E}(X|\mathcal{G}))$ a.s.
   Consequence: contractivity, $\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p$ for $p \geq 1$.
7. If Z is $\mathcal{G}$-measurable and bounded, $\mathbb{E}(ZX|\mathcal{G}) = Z\mathbb{E}(X|\mathcal{G})$ a.s.
   Also if $X \in \mathcal{L}^p$, $Z \in \mathcal{L}^q$ and p, q conjugates i.e. $p^{-1} + q^{-1} = 1$.
   Also if $X \in (m\mathcal{F})^+$, $Z \in (m\mathcal{G})^+$, X integrable, and $\mathbb{E}(XZ) < \infty$.
8. If $\mathcal{H}$ independent of $\sigma(\sigma(X), \mathcal{G})$, then $\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X|\mathcal{G})$ a.s.

Regular versions

Given a sequence $A = (A_n)_n \subset F$ of pairwise disjoint sets, one can use additivity and cMON to show that

$$\mathbb{E}\left(\sum_n I_{A_n}|\mathcal{G}\right) = \sum_n \mathbb{E}(I_{A_n}|\mathcal{G}) \quad \text{a.s.} \tag{4}$$

which holds outside a set N_A of $\mathbb{P}$-measure zero. Since there are uncountable many sequences A, it is not true in general that one can find a unique set N of measure zero such that the σ-additivity (4) holds on $\Omega \setminus N$ for any sequence A of pairwise disjoint sets.

Definition

A **regular conditional probability on $\mathcal{F}$ given $\mathcal{G}$** is a map $P_{\mathcal{G}} : \Omega \times \mathcal{F} \to [0, 1]$ such that

1. for all $F \in \mathcal{F}$, $\omega \mapsto P_{\mathcal{G}}(\omega, F)$ is a version of $\mathbb{P}(F|\mathcal{G})$,
2. for $\mathbb{P}$-almost every ω, the map $F \mapsto P_{\mathcal{G}}(\omega, F)$ is a probability measure on $\mathcal{F}$.