1 Conditional expectation

1.1 Discrete case

Let $\{G_i : i \in I\}$ denote a countable family of disjoint events, whose union is the whole probability space. Set $\mathcal{G} = \sigma(G_i : i \in I)$. For any integrable random variable X, we can define

$$Y \overset{\mathsf{def}}{=} \sum_i \mathsf{E}\,(X \mid G_i)\,\mathbb{1}_{G_i}$$

where we set $\mathsf{E}\,(X \mid G_i) \overset{\mathsf{def}}{=} \mathsf{E}(X\mathbb{1}_{G_i})/\mathsf{P}(G_i)$ when $\mathsf{P}(G_i) > 0$ and define $\mathsf{E}\,(X \mid G_i)$ in some arbitrary way when $\mathsf{P}(G_i) = 0$. Then Y has the following properties:

(a) Y is $\mathcal{G}$-measurable,

(b) Y is integrable and $\mathsf{E}(X\mathbb{1}_A) = \mathsf{E}(Y\mathbb{1}_A)$ for all $A \in \mathcal{G}$.

1.2 Gaussian case

Let (W, X) be a Gaussian random variable in $\mathbb{R}^2$. Set $\mathcal{G} = \sigma(W)$ and $Y \overset{\mathsf{def}}{=} aW + b$, where a, $b \in \mathbb{R}$ are chosen to satisfy

$$a\mathsf{E}(W) + b = \mathsf{E}(X), \qquad a\mathsf{Var}\,W = \mathsf{Cov}\,(W, X).$$

Then $\mathsf{E}(X - Y) = 0$ and

$$\mathsf{Cov}\,\big(W, X - Y\big) = \mathsf{Cov}\,(W, X) - \mathsf{Cov}\,(W, Y) = 0$$

so W and $X - Y$ are independent. Hence Y satisfies:

(a) Y is $\mathcal{G}$-measurable,

(b) Y is integrable and $\mathsf{E}(X\mathbb{1}_A) = \mathsf{E}(Y\mathbb{1}_A)$ for all $A \in \mathcal{G}$.

1.3 Conditional density functions

Suppose that U and V are random variables having a joint density function $f_{U,V}(u, v)$ in $\mathbb{R}^2$. Then U has a density function f_U, given by

$$f_U(u) = \int_{\mathbb{R}} f_{U,V}(u, v)\,dv.$$

The *conditional density function* $f_{V|U}(v|u)$ of V given U is defined by

$$f_{V|U}(v|u) \overset{\mathsf{def}}{=} f_{U,V}(u, v)/f_U(u)$$

where we agree that $0/0 = 0$. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function and suppose that $X = h(V)$ is integrable. Let

$$g(u) = \int_{\mathbb{R}} h(v) f_{V|U}(v|u)\, dv.$$

Set $\mathcal{G} = \sigma(U)$ and $Y = g(U)$. Then Y satisfies:

(a) Y is $\mathcal{G}$-measurable,

(b) Y is integrable and $\mathsf{E}(X\mathbb{1}_A) = \mathsf{E}(Y\mathbb{1}_A)$ for all $A \in \mathcal{G}$.

To see (b), note that every $A \in \mathcal{G}$ takes the form $A = \{U \in B\}$, for some Borel set B. Then, by Fubini's theorem,

$$\begin{aligned}\mathsf{E}(X\mathbb{1}_A) &= \int_{\mathbb{R}^2} h(v)\mathbb{1}_B(u) f_{U,V}(u,v)\, dudv \\ &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} h(v) f_{V|U}(v|u)\, dv \Big) f_U(u)\mathbb{1}_B(u)\, du = \mathsf{E}\big(Y\mathbb{1}_A\big).\end{aligned}$$

1.4 Existence and uniqueness

Theorem 1.4.1 *Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Then there exists a random variable Y such that:*

(a) Y is $\mathcal{G}$-measurable,

(b) Y is integrable and $\mathsf{E}(X\mathbb{1}_A) = \mathsf{E}(Y\mathbb{1}_A)$ for all $A \in \mathcal{G}$.

Moreover, if Y' also satisfies (a) and (b), then $Y = Y'$ a.s..

We call Y (*a version of*) *the conditional expectation of X given $\mathcal{G}$* and write $Y = \mathsf{E}\,(X \mid \mathcal{G})$ a.s.. In the case $\mathcal{G} = \sigma(G)$ for some random variable G, we also write $Y = \mathsf{E}\,(X \mid G)$ a.s.. The preceding three examples show how to construct explicit versions of the conditional expectation in certain simple cases. In general, we have to live with the indirect approach provided by the theorem.

Proof. (*Uniqueness.*) Suppose that Y satisfies (a) and (b) and that Y' satisfies (a) and (b) for another integrable random variable X' with $X \le X'$ a.s.. Consider the non-negative random variable $Z = (Y - Y')\mathbb{1}_A$, where $A \overset{\mathsf{def}}{=} \{Y \ge Y'\} \in \mathcal{G}$. Then

$$\mathsf{E}(Z) = \mathsf{E}(Y\mathbb{1}_A) - \mathsf{E}(Y'\mathbb{1}_A) = \mathsf{E}(X\mathbb{1}_A) - \mathsf{E}(X'\mathbb{1}_A) \le 0$$

so $Z = 0$ a.s., which implies $Y \le Y'$ a.s.. In the case $X = X'$, we deduce that $Y = Y'$ a.s..

(*Existence.*) Assume to begin that $X \in L^2(\mathcal{F})$. Since $V = L^2(\mathcal{G})$ is a closed subspace of $L^2(\mathcal{F})$, we have $X = Y + W$ for some $Y \in V$ and $W \in V^{\perp}$. Then, for any $A \in \mathcal{G}$, we have $\mathbb{1}_A \in V$, so

$$\mathsf{E}(X\mathbb{1}_A) - \mathsf{E}(Y\mathbb{1}_A) = \mathsf{E}(W\mathbb{1}_A) = 0.$$

Hence Y satisfies (a) and (b).

Assume now that X is any non-negative random variable. Then $X_n \stackrel{\mathsf{def}}{=} X \wedge n \in L^2(\mathcal{F})$ and $0 \leq X_n \uparrow X$ as $n \to \infty$. We have shown, for each n, that there exists $Y_n \in L^2(\mathcal{G})$ such that, for all $A \in \mathcal{G}$,

$$\mathsf{E}(X_n 1\!\!1_A) = \mathsf{E}(Y_n 1\!\!1_A)$$

and moreover that $0 \leq Y_n \leq Y_{n+1}$ a.s.. Set $Y \stackrel{\mathsf{def}}{=} \lim_{n\to\infty} Y_n$, then Y is $\mathcal{G}$-measurable and, by monotone convergence, for all $A \in \mathcal{G}$,

$$\mathsf{E}(X 1\!\!1_A) = \mathsf{E}(Y 1\!\!1_A).$$

In particular, if $\mathsf{E}(X)$ is finite then so is $\mathsf{E}(Y)$.

Finally, for a general integrable random variable X, we can apply the preceding construction to X^- and X^+ to obtain Y^- and Y^+. Then $Y = Y^+ - Y^-$ satisfies (a) and (b). □

1.5 Properties of conditional expectation

Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. The following properties follow directly from Theorem 1.4.1:

(i) $\mathsf{E}\big(\mathsf{E}\,(X \mid \mathcal{G})\big) = \mathsf{E}(X)$,

(ii) *if X is $\mathcal{G}$-measurable, then $\mathsf{E}\,(X \mid \mathcal{G}) = X$ a.s.,*

(iii) *if X is independent of $\mathcal{G}$, then $\mathsf{E}\,(X \mid \mathcal{G}) = \mathsf{E}(X)$ a.s..*

In the proof of Theorem 1.4.1, we showed also

(iv) *if $X \geq 0$ a.s., then $\mathsf{E}\,(X \mid \mathcal{G}) \geq 0$ a.s..*

Next, for α, $\beta \in \mathbb{R}$ and any integrable random variable Y, we have

(v) $\mathsf{E}\,(\alpha X + \beta Y \mid \mathcal{G}) = \alpha \mathsf{E}\,(X \mid \mathcal{G}) + \beta \mathsf{E}\,(Y \mid \mathcal{G})$ *a.s..*

To see this, one checks that the right hand side has the defining properties (a) and (b) of the left hand side.

The basic convergence theorems for expectation have counterparts for conditional expectation. Let us consider a sequence of random variables X_n in the limit $n \to \infty$. If $0 \leq X_n \uparrow X$ a.s., then $\mathsf{E}\,(X_n \mid \mathcal{G}) \uparrow Y$ a.s., for some $\mathcal{G}$-measurable random variable Y; so, by monotone convergence, for all $A \in \mathcal{G}$,

$$\mathsf{E}(X 1\!\!1_A) = \lim_{n\to\infty} \mathsf{E}(X_n 1\!\!1_A) = \lim_{n\to\infty} \mathsf{E}\big(\mathsf{E}\,(X_n \mid \mathcal{G})\, 1\!\!1_A\big) = \mathsf{E}(Y 1\!\!1_A),$$

which implies $Y = \mathsf{E}\,(X \mid \mathcal{G})$ a.s.. We have proved the conditional monotone convergence theorem:

(vi) *if $0 \leq X_n \uparrow X$ a.s., then $\mathsf{E}\,(X_n \mid \mathcal{G}) \uparrow \mathsf{E}\,(X \mid \mathcal{G})$ a.s..*

Next, by essentially the same arguments used for the original results, we can deduce conditional forms of Fatou's lemma and the dominated convergence theorem:

(vii) *if* $X_n \geq 0$ *for all* n, *then* $\mathsf{E}\,(\liminf X_n \mid \mathcal{G}) \leq \liminf \mathsf{E}\,(X_n \mid \mathcal{G})$ *a.s.*,

(viii) *if* $X_n \to X$ *and* $|X_n| \leq Y$ *for all* n, *a.s., for some integrable random variable* Y, *then* $\mathsf{E}\,(X_n \mid \mathcal{G}) \to \mathsf{E}\,(X \mid \mathcal{G})$ *a.s.*.

There is a conditional form of Jensen's inequality. Let $c : \mathbb{R} \to (-\infty, \infty]$ be a convex function. Then c is the supremum of countably many affine functions:

$$c(x) = \sup_i \big(a_i x + b_i\big), \quad x \in \mathbb{R}.$$

Hence, $\mathsf{E}\,(c(X) \mid \mathcal{G})$ is well defined and, almost surely, for all i,

$$\mathsf{E}\,(c(X) \mid \mathcal{G}) \geq a_i \mathsf{E}\,(X \mid \mathcal{G}) + b_i.$$

So we obtain

(ix) *if* $c : \mathbb{R} \to (-\infty, \infty]$ *is convex, then* $\mathsf{E}\,(c(X) \mid \mathcal{G}) \geq c\big(\mathsf{E}\,(X \mid \mathcal{G})\big)$ *a.s.*.

In particular, for $1 \leq p < \infty$,

$$\big\|\mathsf{E}\,(X \mid \mathcal{G})\big\|_p^p = \mathsf{E}\big(|\mathsf{E}\,(X \mid \mathcal{G})\,|^p\big) \leq \mathsf{E}\big(\mathsf{E}\,(|X|^p \mid \mathcal{G})\big) = \mathsf{E}\big(|X|^p\big) = \|X\|_p^p.$$

So we have

(x) $\big\|\mathsf{E}\,(X \mid \mathcal{G})\big\|_p \leq \|X\|_p$ *for all* $1 \leq p < \infty$.

For any σ-algebra $\mathcal{H} \subseteq \mathcal{G}$, the random variable $Y \overset{\mathsf{def}}{=} \mathsf{E}\,(\mathsf{E}\,(X \mid \mathcal{G}) \mid \mathcal{H})$ is $\mathcal{H}$-measurable and satisfies, for all $A \in \mathcal{H}$

$$\mathsf{E}\big(Y 1\!\!1_A\big) = \mathsf{E}\big(\mathsf{E}\,(X \mid \mathcal{G})\, 1\!\!1_A\big) = \mathsf{E}\big(X 1\!\!1_A\big)$$

so we have the *tower property*:

(xi) *if* $\mathcal{H} \subseteq \mathcal{G}$, *then* $\mathsf{E}\,(\mathsf{E}\,(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathsf{E}\,(X \mid \mathcal{H})$ *a.s.*.

We can always *take out what is known*:

(xii) *if* Y *is bounded and* $\mathcal{G}$*-measurable, then* $\mathsf{E}\,(YX \mid \mathcal{G}) = Y\mathsf{E}\,(X \mid \mathcal{G})$ *a.s.*.

To see this, consider first the case when $Y = 1\!\!1_B$ for some $B \in \mathcal{G}$. Then, for $A \in \mathcal{G}$,

$$\mathsf{E}\big(Y\mathsf{E}\,(X \mid \mathcal{G})\, 1\!\!1_A\big) = \mathsf{E}\big(\mathsf{E}\,(X \mid \mathcal{G})\, 1\!\!1_{A\cap B}\big) = \mathsf{E}\big(X 1\!\!1_{A\cap B}\big) = \mathsf{E}\big(YX 1\!\!1_A\big),$$

which implies $\mathsf{E}\,(YX \mid \mathcal{G}) = Y\mathsf{E}\,(X \mid \mathcal{G})$ a.s.. The result extends to simple $\mathcal{G}$-measurable random variables Y by linearity, then to the case $X \geq 0$ and any non-negative $\mathcal{G}$-measurable random variable Y by monotone convergence. The general case follows by writing $X = X^+ - X^-$ and $Y = Y^+ - Y^-$.

Finally,

(xiii) *if $\sigma(X, \mathcal{G})$ is independent of $\mathcal{H}$, then* $\mathsf{E}\,(X \mid \sigma(\mathcal{G}, \mathcal{H})) = \mathsf{E}\,(X \mid \mathcal{G})$ a.s..

For, suppose $A \in \mathcal{G}$ and $B \in \mathcal{H}$, then

$$\begin{aligned} \mathsf{E}\big(\mathsf{E}\,(X \mid \sigma(\mathcal{G}, \mathcal{H}))\, \mathbb{1}_{A\cap B}\big) &= \mathsf{E}\big(X\,\mathbb{1}_{A\cap B}\big) = \mathsf{E}\big(X\,\mathbb{1}_A\big)\mathsf{P}(B) \\ &= \mathsf{E}\big(\mathsf{E}\,(X \mid \mathcal{G})\,\mathbb{1}_A\big)\mathsf{P}(B) = \mathsf{E}\big(\mathsf{E}\,(X \mid \mathcal{G})\,\mathbb{1}_{A\cap B}\big). \end{aligned}$$

It remains to observe that the set of such intersections $A \cap B$ is a π-system generating $\sigma(\mathcal{G}, \mathcal{H})$.

Lemma 1.5.2 *Let $X \in L^1$. Then the set of random variables Y of the form $Y = \mathsf{E}\,(X \mid \mathcal{G})$, where $\mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra, is uniformly integrable.*

Proof. Since $X \in L^1$, given $\varepsilon > 0$, we can find $\delta > 0$ so that $\mathsf{E}\big(|X|\mathbb{1}_A\big) \le \varepsilon$ whenever $\mathsf{P}(A) \le \delta$. Then choose $\lambda < \infty$ so that $\mathsf{E}(|X|) \le \lambda\delta$. Suppose $Y = \mathsf{E}\,(X \mid \mathcal{G})$, then $|Y| \le \mathsf{E}\,(|X| \mid \mathcal{G})$. In particular, $\mathsf{E}(|Y|) \le \mathsf{E}(|X|)$ so

$$\mathsf{P}(|Y| \ge \lambda) \le \lambda^{-1}\mathsf{E}(|Y|) \le \delta.$$

Then

$$\mathsf{E}\big(|Y|\mathbb{1}_{\{|Y|\ge\lambda\}}\big) \le \mathsf{E}\big(\mathsf{E}\,(|X| \mid \mathcal{G})\,\mathbb{1}_{\{|Y|\ge\lambda\}}\big) = \mathsf{E}\big(|X|\mathbb{1}_{\{|Y|\ge\lambda\}}\big) \le \varepsilon.$$

Since λ was chosen independently of $\mathcal{G}$, we are done. □