Conditional Expectation

By Prof. Jim Fill

April 12, 2010

Throughout, $(\Omega, \mathcal{A}, P)$ is a fixed probability space.

Proposition 1. *Let X be a P-quasi-integrable r.v. on Ω, and let $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$. Then there exists a P-quasi-integrable r.v. $E^{\mathcal{B}}X$ such that*

(i) $E^{\mathcal{B}}X$ is $\mathcal{B}$-measurable.

(ii) $\int_B E^{\mathcal{B}}X \mathrm{d}P = \int_B X \mathrm{d}P$ for all $B \in \mathcal{B}$.

Moreover, any two such $E^{\mathcal{B}}X$'s are equal a.s. P.

Definition 2. Let X and $\mathcal{B}$ be as in the proposition. Any of the P-equivalent $E^{\mathcal{B}}X$'s is called a (or "the") *conditional expectation of X given $\mathcal{B}$.*

Four proofs of proposition :

$1^{\underline{\circ}}$. *Motivational.* Valid only in the case $\Omega = \sum_{i \in I} B_i$[1], $\mathcal{B} = \{\sum_{j \in J} B_j : J \subset I\}$, with I countable.

$2^{\underline{\circ}}$. *"Build-'em-up" from case $X \in L^2(\Omega, \mathcal{A}, P)$*, for which the projection of X onto $L^2(\Omega, \mathcal{B}, P)$ serves as $E^{\mathcal{B}}X$.

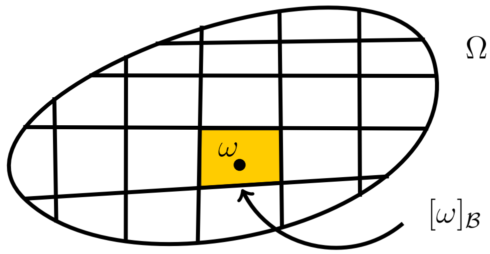
$3^{\underline{\circ}}$. *Use Radon-Nikodym theorem.* See Billingsley, Section 34.

$4^{\underline{\circ}}$. *Use martingale theory* (coming soon!).

Proof. $1^{\underline{\circ}}$ of basic proposition.

This will be an honest proof, but only subject to a very restrictive assumption; namely, that there exists a partition of Ω into countably many sets $B_i \in \mathcal{B} (i \in I)$, and that $\mathcal{B} = \{\sum_{j \in J} B_j : J \subset I\}$ (i.e., that any set in $\mathcal{B}$ is the union of some B_j's). (In essence, the measurable space $(\Omega, \mathcal{B})$ is discrete.) Note that a real function on Ω is $\mathcal{B}$-measurable if and only if it is constant on each cell B_i of the partition. For each $\omega \in \Omega$, *we define $[\omega]_{\mathcal{B}}$ to be the smallest element of $\mathcal{B}$ containing ω, i.e., the B_i that contains ω.*

[1] "$\sum$" denotes disjoint union.



Uniqueness. (i) implies that $E^{\mathcal{B}}X$ must be constant on $[\omega]_{\mathcal{B}}$. Using this in (ii) we get

$$\int_{[\omega]_{\mathcal{B}}} X \mathrm{d}P = \int_{[\omega]_{\mathcal{B}}} E^{\mathcal{B}} X \mathrm{d}P = (E^{\mathcal{B}}X)(\omega) P([\omega]_{\mathcal{B}}).$$

Thus if $P([\omega]_{\mathcal{B}}) > 0$, we must have

$$(E^{\mathcal{B}}X)(\omega) = \frac{\int_{[\omega]_{\mathcal{B}}} X \mathrm{d}P}{P([\omega]_{\mathcal{B}})} = E(X \mid [\omega]_{\mathcal{B}})$$

i.e.,

$$\boxed{(E^{\mathcal{B}}X)(\omega) = \text{the average value of } X \text{ over } [\omega]_{\mathcal{B}}}.$$

It follows that any two $E^{\mathcal{B}}X$'s can differ only at ω's such that $P([\omega]_{\mathcal{B}}) = 0$. But the set of such ω's, namely, $\sum_{j:P(B_j)=0} B_j$, is a set of probability zero, so we have the almost sure uniqueness.

Existence. It is clear how to proceed. We set

$$(E^{\mathcal{B}}X)(\omega) := \begin{cases} E(X \mid [\omega]_{\mathcal{B}}) & \text{if } P([\omega]_{\mathcal{B}}) > 0 \\ 0 (\text{say, or } EX) & \text{if } P([\omega]_{\mathcal{B}}) = 0. \end{cases}$$

Our $E^{\mathcal{B}}X$ is constant over each cell B_i of the partition and therefore $\mathcal{B}$-measurable. Moreover, for any $B = \sum_{j \in J} B_j \in \mathcal{B}$, we have, with $\tilde{J} := \{j \in J : P(B_j) > 0\}$,

$$
\begin{aligned}
\int_B X\mathrm{d}P &= \sum_{j\in J}\int_{B_j} X\mathrm{d}P = \sum_{j\in\tilde J}\int_{B_j} X\mathrm{d}P = \sum_{j\in\tilde J} E(X\,|\,B_j)P(B_j) \\
&= \sum_{j\in\tilde J}\int_{B_j} E(X\,|\,[\omega]_{\mathcal{B}})P(\mathrm{d}\omega) \quad (\text{since } B_j=[\omega]_{\mathcal{B}} \text{ for } \omega\in B_j) \\
&= \sum_{j\in J}\int_{B_j} (E^{\mathcal{B}}X)(\omega)P(\mathrm{d}\omega) \\
&= \int_B E^{\mathcal{B}}X\mathrm{d}P
\end{aligned} \tag{1}
$$

provided $E^{\mathcal{B}}X$ is quasi-integrable over B. But $E^{\mathcal{B}}X$ *is* quasi-integrable over $\{E^{\mathcal{B}}X\geq 0\}\in\mathcal{B}$, so (1) for this event implies

$$
\begin{aligned}
E\big[(E^{\mathcal{B}}X)^+\big] &= \int_{\{E^{\mathcal{B}}X\geq 0\}} E^{\mathcal{B}}X\mathrm{d}P = \int_{\{E^{\mathcal{B}}X\geq 0\}} X\mathrm{d}P \\
&\leq \int_{\{E^{\mathcal{B}}X\geq 0\}} X^+\mathrm{d}P \leq EX^+.
\end{aligned}
$$

Similarly, (1) implies $E[(E^{\mathcal{B}}X)^-]\leq EX^-$. Therefore $E^{\mathcal{B}}X$ is in fact quasi-integrable (in the same sense(s) as X) over the entire sample space Ω, and, in particular, quasi-integrable over each $B\in\mathcal{B}$, as required. Thus (1), i.e., (*ii*), holds in general. □

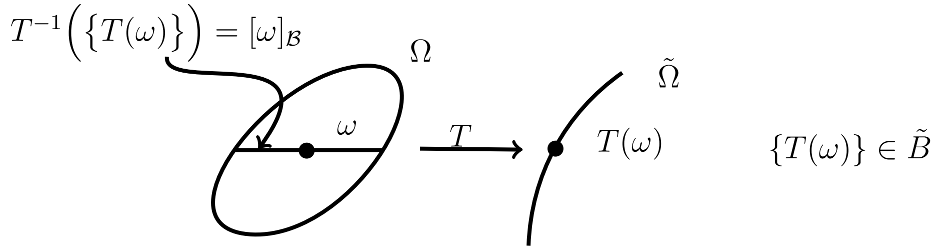
Remarks 3. The general $\mathcal{B}$ cannot be so neatly associated with a partition of Ω. To see what's involved, use $\mathcal{B}$ to define an equivalence relation on Ω: say $\omega_1\sim_{\mathcal{B}}\omega_2$ exactly when every $B\in\mathcal{B}$ containing ω_1 also contains ω_2, and vice versa. Call the resulting equivalence classes $\mathcal{B}$-*cosets*, and denote by $[\omega]_{\mathcal{B}}$ the $\mathcal{B}$-coset containing $\omega\in\Omega$:

$$
[\omega]_{\mathcal{B}} = \bigcap_{B\in\mathcal{B}\,:\,\omega\in B} B.
$$

Example 4. (a) $\mathcal{B}$ *as in "proof"* $1^{\underline{\circ}}$. The $\mathcal{B}$-cosets are precisely the B_i's, and $[\omega]_{\mathcal{B}}$ here is the same thing as before – the B_i containing ω.

(b) $\mathcal{B}=\sigma\langle T;\tilde{\mathcal{B}}\rangle$, where $T:\sigma\to\tilde\sigma$ for some measurable space $(\tilde\Omega,\tilde{\mathcal{B}})$ such that $\tilde{\mathcal{B}}$ contains all singletons. Since $\mathcal{B}=\{T^{-1}(\tilde B):\tilde B\in\tilde{\mathcal{B}}\}$ (by definition) and since $\{T(\omega)\}\in\tilde{\mathcal{B}}$, it follows that $[\omega]_{\mathcal{B}}=T^{-1}(\{T(\omega)\})$.

The $\mathcal{B}$-cosets do always partition Ω, and every $B\in\mathcal{B}$ is the sum of those $\mathcal{B}$-cosets it contains. But the sum is in general uncountable (cf. Example(b), with $(\tilde\Omega,\tilde{\mathcal{B}})=(\mathbb{R},\mathcal{R})$); moreover, not every $[\omega]_{\mathcal{B}}$, much less every sum of $\mathcal{B}$-cosets, need belong to $\mathcal{B}$. (In Example (b), every $[\omega]_{\mathcal{B}}$, but not necessarily every sum of $\mathcal{B}$-cosets, belongs to $\mathcal{B}$.) So we cannot say $\mathcal{B}=\{\sum_{j\in J}B_j : \text{each } B_j \text{ is a } \mathcal{B}\text{-coset}\}$, and we cannot identify constancy over $\mathcal{B}$-cosets with $\mathcal{B}$-measurability.



Also, even if $[\omega]_{\mathcal{B}}$ belongs to $\mathcal{B}$, it may (and in interesting cases will) have probability zero; typically all the expressive $E([\omega]_{\mathcal{B}}) \equiv \dfrac{\int_{[\omega]_{\mathcal{B}}} X \mathrm{d}P}{P([\omega]_{\mathcal{B}})}$ are either meaningless ($[\omega]_{\mathcal{B}} \notin \mathcal{B}$) or indeterminate ($P([\omega]_{\mathcal{B}}) = 0$). However, if we ignore these unpleasantnesses, and go through the previous proof informally, we see that

(a) one can intuitively think of $(E^{\mathcal{B}}X)(\omega)$ as $E(X \,|\, [\omega]_{\mathcal{B}})$, the "average value of X over the smallest event in $\mathcal{B}$ containing ω", and

(b) conditions (i) and (ii) in the proposition express the global behavior of the locally (*un*-) defined function $E(X \,|\, [\,\centerdot\,]_{\mathcal{B}})$.

As an aid to the intuition, we may sometimes denote the value of an $E^{\mathcal{B}}X$ at ω by $E(X \,|\, [\omega]_{\mathcal{B}})$. But keep in mind that this quantity is not defined unless $[\omega]_{\mathcal{B}} \in \mathcal{B}$ and $P([\omega]_{\mathcal{B}}) > 0$, in which case it is in fact that the value of *any* $E^{\mathcal{B}}X$ at ω.

Example 5. (1) $\mathcal{B} = \{\phi, \Omega\}$, $E^{\mathcal{B}}X = (EX)I_{\Omega}$ (uniquely). $E^{\{\phi,\Omega\}}X$ is the ultimate *smoothing* of X – to a constant function.

(2) $\mathcal{B} = \{\,\sum_{j \in J} B_j : J \subset \{1,2,3\}\}$ with $P(B_j) > 0$ for all j, $E^{\mathcal{B}}X = \sum_{j=1}^{3} E(X \,|\, B_j)I_{B_j}$ (uniquely). $E^{\mathcal{B}}X$ is a partial *smoothing* of X.

(3) Same $\mathcal{B}$ as (2), except $P(B_3) = 0$. General $E^{\mathcal{B}}X = \sum_{j=1}^{2} E(X \,|\, B_j)I_{B_j} + cI_{B_3}$, c arbitrary.

(4) $\Omega = [-1, +1]$, $\mathcal{A} = \{\text{Borels}\}$, $\mathrm{d}P = f \mathrm{d}\lambda$, $\lambda =$ Lebesgue measure, $0 < f < \infty$ (for convenience). $\mathcal{B} = \sigma\langle T \rangle$, where $T : \Omega \to [0,1]$ is defined by $T(\omega) = |\omega|$. $(E^{\mathcal{B}}X)(\omega) = X(\omega)\frac{f(\omega)}{f(\omega)+f(-\omega)} + X(-\omega)\frac{f(-\omega)}{f(\omega)+f(-\omega)}$ (at least for $X \geq 0$). $E^{\mathcal{B}}X$ is the *smoothing* of X to an even function.

One easily checks that $\mathcal{B} = \{A \in \mathcal{A} : A = -A\}$, where $-A \coloneqq \{-a : a \in A\}$. Let $X \geq 0$. We have $[\omega]_{\mathcal{B}} = \{-\omega, \omega\}$ (recall Example (b) above), which has probability zero under P. But given that $T(\omega) = t$, one intuitively feels that $\omega = t$ with probability $\frac{f(t)}{f(t)+f(-t)}$ and $\omega = -t$ with probability $\frac{f(-t)}{f(t)+f(-t)}$. Hence $E(X \,|\, [\omega]_{\mathcal{B}})$ "ought" to be $X(\omega)\frac{f(\omega)}{f(\omega)+f(-\omega)} + X(-\omega)\frac{f(-\omega)}{f(\omega)+f(-\omega)} =: Z(\omega)$. So we propose Z as a candidate for $E^{\mathcal{B}}X$.

Z is $\mathcal{B}$-measurable because it is $\mathcal{A}$-measurable and even ($Z(\omega) = Z(-\omega)\ \forall\omega \in \Omega$). Moreover, if $0 \leq U$ is $\mathcal{B}$-measurable, i.e., $\mathcal{A}$-measurable and even, then we have

$$\begin{aligned}\int_\Omega U(\omega)Z(\omega)P(\mathrm{d}\omega) &= \int U(\omega)\Big[X(\omega)\frac{f(\omega)}{f(\omega)+f(-\omega)} + X(-\omega)\frac{f(-\omega)}{f(\omega)+f(-\omega)}\Big]f(\omega)\mathrm{d}\omega \\ &= \int U(\omega)X(\omega)\frac{f(\omega)f(\omega)}{f(\omega)+f(-\omega)}\mathrm{d}\omega + \int U(\omega)X(-\omega)\frac{f(-\omega)f(\omega)}{f(\omega)+f(-\omega)}\mathrm{d}\omega \\ &= \int U(\omega)X(\omega)\frac{f(\omega)f(\omega)}{f(\omega)+f(-\omega)}\mathrm{d}\omega + \int U(-\omega)X(\omega)\frac{f(\omega)f(-\omega)}{f(-\omega)+f(\omega)}\mathrm{d}\omega \\ &= \int U(\omega)X(\omega)\frac{f(\omega)}{f(\omega)+f(-\omega)}\big[f(\omega)+f(-\omega)\big]\mathrm{d}\omega \\ &= \int U(\omega)X(\omega)f(\omega)\mathrm{d}\omega = \int U(\omega)X(\omega)P(\mathrm{d}\omega)\end{aligned}$$

It follows (take $U = I_B$, $B \in \mathcal{B}$) that our Z is indeed a conditional expectation of X given $\mathcal{B}$. Later on we shall develop a machine which will crank out Z for $E^{\mathcal{B}}X$ automatically.

Basic properties of $E^{\mathcal{B}}X$:

Note. Whenever we write $E^{\mathcal{B}}X$, we automatically assure X P-quasi-integrable (unless a special fuss is made), and we always view $E^{\mathcal{B}}X$ as an arbitrary but fixed version of the conditional expectation of X given $\mathcal{B}$.

Smoothing-type properties :

(S1) $E^{\mathcal{B}}X \in \begin{Bmatrix} Q_+ \\ Q_- \\ L^1 \end{Bmatrix}$ iff $X \in \begin{Bmatrix} Q_+ \\ Q_- \\ L^1 \end{Bmatrix}$. $EE^{\mathcal{B}}X = EX$.

(S2) If X is $\mathcal{B}$-measurable, then $E^{\mathcal{B}}(XY) \overset{\text{a.s.}}{=} XE^{\mathcal{B}}Y$ and, in particular, $E^{\mathcal{B}}X \overset{\text{a.s.}}{=} X$.

(S3) If $\mathcal{B}_1 \subset \mathcal{B}_2$, then $E^{\mathcal{B}_1}(E^{\mathcal{B}_2}X) \overset{\text{a.s.}}{=} E^{\mathcal{B}_1}X \overset{\text{a.s.}}{=} E^{\mathcal{B}_2}(E^{\mathcal{B}_1}X)$.

Expectation operator type properties :

(E1) $E^{\mathcal{B}}1 \overset{\text{a.s.}}{=} 1$.

(E2) $E^{\mathcal{B}}(cX) = cE^{\mathcal{B}}X$; $E^{\mathcal{B}}(X+Y) \overset{\text{a.s.}}{=} E^{\mathcal{B}}X + E^{\mathcal{B}}Y$ if X and Y are both in $\begin{Bmatrix} Q_+ \\ Q_- \end{Bmatrix}$.

(E3) (a) If $X_1 \leq X_2$, then $E^{\mathcal{B}}X_1 \overset{\text{a.s.}}{\leq} E^{\mathcal{B}}X_2$. In particular, if $0 \leq X$ then $0 \overset{\text{a.s.}}{\leq} E^{\mathcal{B}}$.

(b) $(E^{\mathcal{B}}X)^+ \overset{\text{a.s.}}{\leq} E^{\mathcal{B}}(X^+)$; $(E^{\mathcal{B}}X)^- \overset{\text{a.s.}}{\leq} E^{\mathcal{B}}(X^-)$; $|E^{\mathcal{B}}X| \overset{\text{a.s.}}{\leq} E^{\mathcal{B}}|X|$.

(c) If $X_1 \overset{\text{a.s.}}{<} X_2$ and at least one of X_1 and X_2 is integrable, then $E^{\mathcal{B}}X_1 \overset{\text{a.s.}}{<} E^{\mathcal{B}}X_2$.

(E4) (*MCT*) If $X_n \uparrow X$, then $E^{\mathcal{B}} X_n \overset{\text{a.s.}}{\uparrow} E^{\mathcal{B}} X$ over $\cup_{m\geq 1}\{E^{\mathcal{B}} X_m > -\infty\}$.
If $X_n \downarrow X$, then $E^{\mathcal{B}} X_n \overset{\text{a.s.}}{\downarrow} E^{\mathcal{B}} X$ over $\cup_{m\geq 1}\{E^{\mathcal{B}} X_m < \infty\}$.

(E5) (*Fatou*) If $E(\inf_n X_n) > -\infty$, then $E^{\mathcal{B}}(\liminf_n X_n) \overset{\text{a.s.}}{\leq} \liminf_n E^{\mathcal{B}} X_n$.
If $E(\sup_n X_n) < \infty$, then $\limsup_n E^{\mathcal{B}} \overset{\text{a.s.}}{\leq} X_n E^{\mathcal{B}}(\limsup_n X_n)$.

(E6) (*DCT*) If $E(sup_n|X_n|) < \infty$ and $\lim_n X_n$ exists a.s., then $E^{\mathcal{B}}(\lim_n X_n) \overset{\text{a.s.}}{=} \lim_n E^{\mathcal{B}} X_n$.

(E7) If X and $\mathcal{B}$ are independent, then $E^{\mathcal{B}} X \overset{\text{a.s.}}{=} (EX) I_\Omega$.

Conditional expectation given a measurable function :

Motivation. So far we have considered conditioning relative to sub-σ-algebras $\mathcal{B}$. There is a closely related notion, involving conditioning by a measurable function. Let $T : (\Omega, \mathcal{A}) \to (\mathcal{T}, \mathcal{C})$ be measurable, and set $\mathcal{B} \equiv \mathcal{B}_T := T^{-1}(\mathcal{C})$. Look at $E^{\mathcal{B}_T} X$. This is a $\mathcal{B}_T$-measurable function and therefore (by the Factorization Theorem) can be written as a measurable function of T, say, as

$$E(X \mid T) \circ T,$$

where $E(X \mid T) : \mathcal{T} \to \overline{\mathbb{R}}$ is $\mathcal{C}$-measurable. (Intuitively, $E(X|T)(t) =$ "the constant value of $E^{\mathcal{B}_T} X$ over all the $\mathcal{B}_T$-coset $\{T = t\}$".) Moreover, we have

$$\begin{aligned} \int_{\{T\in C\}} X \,\mathrm{d}P = \int_{\{T\in C\}} E^{\mathcal{B}_T} X \,\mathrm{d}P &= \int_{\{T\in C\}} E(X|T) \circ T \,\mathrm{d}P \\ &= \int_C E(X|T)\,\mathrm{d}(PT^{-1}) \quad \text{(change of variable)} \end{aligned} \tag{2}$$

for all $C \in \mathcal{C}$. It's easy to check that $E(X|T)$ is PT^{-1}-essentially uniquely determined by (2).

For the sake of intuition, we sometimes write $E(X|T)(t)$ as $E(X|T = t)$.

Definition 6. Let X be P-quasi-integrable, and let $T : \Omega \to \mathcal{T}$ be measurable between the σ-fields $\mathcal{A}$ and $\mathcal{C}$. Anyone of the PT^{-1}-equivalent $\mathcal{C}$-measurable functions $E(X|T) : \mathcal{T} \to \overline{\mathbb{R}}$ such that

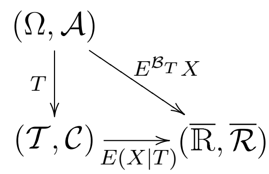
$$\int_{\{T\in C\}} X \,\mathrm{d}P = \int_C E(X|T)\mathrm{d}(PT^{-1}) \quad \forall C \in \mathcal{C}$$

is called a *conditional expectation of* X *given* T.

Correspondence between $E^{\lfloor} X$ and $E(X|T)$:

(a) If $\mathcal{B}_T := \sigma\langle T\rangle = T^{-1}(\mathcal{C})$, then

$$E^{\mathcal{B}_T} = E(X|T) \circ T.$$



(b) If $T :=$ (identity mapping on Ω) and $(\mathcal{T}, \mathcal{C}) := (\Omega, \mathcal{B})$, then

$$E^{\mathcal{B}} T = E(X|T).$$

Warning. Often $E(X|T)$ is used as the notation for $E^{\mathcal{B}_T} X$.

Proofs :

Proof. $2^{\underline{o}}$ of basic proposition.

Uniqueness. It suffices to show that if Y and Z are $\mathcal{B}$-measurable and P-quasi-integrable and $\int_B Y \, \mathrm{d}P \leq \int_B Z \, \mathrm{d}P \ \forall B \in \mathcal{B}$, then $Y \leq Z$ a.s. For this, consider the event $B := \{Z < t \leq Y\} \in \mathcal{B}$. We have

$$tP(\mathcal{B}_t) \leq \int_{B_t} Y \, \mathrm{d}P \leq \int_{B_t} Z \, \mathrm{d}P < tP(B_t),$$

a contradiction, unless $P(B_t) = 0$. But $\{Z < Y\} = \cup_{t \text{ rational}} B_t$, so $P\{Z < Y\} = 0$, i.e., $Y \leq Z$ a.s.

Existence. If we had $E^{\mathcal{B}} X$, we would expect (using the "build-'em-up" technique) that $\int U(X - E^{\mathcal{B}} X) \, \mathrm{d}P = 0$, i.e., $(X - E^{\mathcal{B}} X) \perp U$, for $\mathcal{B}$-functions U. [Indeed, with $U = I_B$, $B \in \mathcal{B}$, this is required property (ii).] This suggests what to do:

Step 1. $X \in L^2(\Omega, \mathcal{A}, P)$. Check that $L^2(\Omega, \mathcal{B}, P)$ is a Hilbert subspace of $L^2(\Omega, \mathcal{A}, P)$. By the basic projection theorem for Hilbert spaces, there is a vector, say, Z, in $L^2(\mathcal{B})$ such that $\|X - Z\|_2 = \inf_{Y \in L^2(\mathcal{B})} \|X - Y\|_2$. This Z satisfies

$$(X - Z) \perp U \quad \forall U \in L^2(\mathcal{B}). \tag{3}$$

We claim that Z serves as $E^{\mathcal{B}} X$: Z is $\mathcal{B}$-measurable, $Z \in L^2 \subset L^1 \subset Q$, and by (3) $\int_B Z \, \mathrm{d}P = \int_B X \, \mathrm{d}P$ for each $B \in \mathcal{B}$.

Step 2. $X \geq 0$. Define $X_n = X_{\wedge n} \in L^2(\Omega, \mathcal{A}, P)$. Use Step 1 to determine $E^{\mathcal{B}} X_n$ — $\mathcal{B}$-measurable and satisfying

$$\int_B E^{\mathcal{B}} X_n \, \mathrm{d}P = \int_B X_n \, \mathrm{d}P. \tag{4}$$

Since $0 \leq X_n \leq X_{n+1}$, it follows (cf. pf. of uniqueness) that $0 \leq E^{\mathcal{B}} X_n \leq E^{\mathcal{B}} X_{n+1}$ a.s. Thus the $(E^{\mathcal{B}} X_n)$'s increases on a set of probability one; in fact (why?), they may be taken to increase everywhere. We claim $\lim_n \uparrow E^{\mathcal{B}} X_n$ serves as $E^{\mathcal{B}} X$. Clearly, $\lim_n \uparrow E^{\mathcal{B}} X_n$ is $\mathcal{B}$-measurable and P-quasi-integrable (in fact, ≥ 0). Apply the MCT to (4) to get

$$\begin{aligned} \int_B (\lim_n \uparrow E^{\mathcal{B}} X_n)\,\mathrm{d}P &= \lim_n \uparrow \int_B E^{\mathcal{B}} X_n\,\mathrm{d}P = \lim_n \uparrow \int_B X_n\,\mathrm{d}P \\ &= \int_B (\lim_n \uparrow X_n)\,\mathrm{d}P = \int_B X\,\mathrm{d}P \quad \text{for each } B \in \mathcal{B}. \end{aligned}$$

Step 3. $X \in Q^-$. Use Step 2 to determine $E^{\mathcal{B}}(X^+)$ and $E^{\mathcal{B}}(X^-)$. Observe $E\big(E^{\mathcal{B}}(X^-)\big) = E(X^-) < \infty$ (consider $B = \Omega$). In particular, $E^{\mathcal{B}}(X^-) < \infty$ a.s.; without loss of generality $E^{\mathcal{B}}(X^-) < \infty$ everywhere. We claim that $E^{\mathcal{B}}(X^+) - E^{\mathcal{B}}(X^-)$ serves as $E^{\mathcal{B}} X$. Clearly, $E^{\mathcal{B}}(X^+) - E^{\mathcal{B}}(X^-)$ is well defined, $\mathcal{B}$-measurable, and quasi-integrable (since $E^{\mathcal{B}}(X^-) \in L^1$), and for each $B \in \mathcal{B}$

$$\begin{aligned} \int_B \big(E^{\mathcal{B}}(X^+) - E^{\mathcal{B}}(X^-)\big)\,\mathrm{d}P &= \int_B E^{\mathcal{B}}(X^+)\,\mathrm{d}P - \int_B E^{\mathcal{B}}(X^-)\,\mathrm{d}P \quad (\text{since } E^{\mathcal{B}}(X^-) \in L^1) \\ &= \int_B X^+\,\mathrm{d}P - \int_B X^-\,\mathrm{d}P = \int_B X\,\mathrm{d}P. \end{aligned}$$

□

Notes.

(a) The constructions in Steps 2 and 3 are entirely analogous to what we did in developing (unconditional) expectation.

(b) If $X \in L^2(\mathcal{A})$, we have seen that $E^{\mathcal{B}} X$, the $\mathcal{B}$-smoothing of X, is obtained by moving X as little as possible.

Proofs of properties of conditional expectation :

(E1) Simple.

(S1) Simple.

(E2) First part simple. If both X and Y are in Q_-, then by (S1) so are $E^{\mathcal{B}} X$ and $E^{\mathcal{B}} Y$, and so are $X + Y$ and $E^{\mathcal{B}} X + E^{\mathcal{B}} Y$. The last is $\mathcal{B}$-measurable, and for each $B \in \mathcal{B}$

$$\int_B (E^{\mathcal{B}} X + E^{\mathcal{B}} Y)\,\mathrm{d}P = \int_B E^{\mathcal{B}} X\,\mathrm{d}P + \int_B E^{\mathcal{B}} Y\,\mathrm{d}P = \int_B X\,\mathrm{d}P + \int_B Y\,\mathrm{d}P = \int_B (X+Y)\,\mathrm{d}P.$$

(E3) (a) Proved in uniqueness part of proof $2^{\underline{\circ}}$.

(b) For example, $X \leq X^+$ and so by (a) $E^{\mathcal{B}}X \overset{\text{a.s.}}{\leq} E^{\mathcal{B}}(X^+)$. But $E^{\mathcal{B}}(X^+) \overset{\text{a.s.}}{\geq} 0$, so $(E^{\mathcal{B}}X)^+ \overset{\text{a.s.}}{\leq} E^{\mathcal{B}}(X^+)$.

(c) Taking $X = X_2 - X_1$ and using (E2), it suffices to show that if $0 \overset{\text{a.s.}}{<} X$, then $0 \overset{\text{a.s.}}{<} E^{\mathcal{B}}X$. Actually, we do a bit more: If $x \geq 0$, then $\{E^{\mathcal{B}}X = 0\} \overset{\text{a.s.}}{\subset} \{X = 0\}$, because $0 = \int_{\{E^{\mathcal{B}}X=0\}} E^{\mathcal{B}}X = \int_{\{E^{\mathcal{B}}X=0\}} X$.

(S2) *Case* 1: $X \geq 0, Y \geq 0$. Since $\int UE^{\mathcal{B}}Y = \int UY$ holds for $\mathcal{B}$-indicators U, it holds (build-'em-up !) for non-negative $\mathcal{B}$-functions U. Hence for $B \in \mathcal{B}$

$$\int_B XE^{\mathcal{B}}Y = \int (I_B X)E^{\mathcal{B}}Y = \int (I_B X)Y = \int_B XY.$$

Case 2: XY and Y are both quasi-integrable. Note $XY = (X^+ - X^-)(Y^+ - Y^-) = X^+Y^+ + X^-Y^- + (-X^+Y^-) + (-X^-Y^+)$, where the four r.v.'s here have disjoint supports. This, e.g., if $XY \in Q_-$, the all four r.v.'s are in Q_-, and (E2) gives (a.s. throughout)

$$\begin{aligned} E^{\mathcal{B}}(XY) &= E^{\mathcal{B}}(X^+Y^+) + E^{\mathcal{B}}(X^-Y^-) - E^{\mathcal{B}}(X^+Y^-) - E^{\mathcal{B}}(X^-Y^+) \\ &= X^+E^{\mathcal{B}}Y^+ + X^-E^{\mathcal{B}}Y^- - X^+E^{\mathcal{B}}Y^- - X^-E^{\mathcal{B}}Y^+ \\ &= (X^+ - Y^-)(E^{\mathcal{B}}Y^+ - E^{\mathcal{B}}Y^-) \\ &= XE^{\mathcal{B}}Y. \end{aligned}$$

(S3) Simple.

(E4) Put $B_{c,m} = \{E^{\mathcal{B}}X_m \geq c\} \in \mathcal{B} \quad (0 > c > -\infty)$. Now $X_n \uparrow X$

$$\begin{aligned} &\Rightarrow X_n I_{B_{c,m}} \uparrow X I_{B_{c,m}} \Rightarrow \text{ for } B \in \mathcal{B}, \int_B \overset{\text{(a.s.)}}{\lim_n \uparrow} \underbrace{E^{\mathcal{B}}(X_n I_{B_{c,m}})}_{\geq c \text{ for } n \geq m \text{ (a.s.) [use (E3), (S2)]}} \\ &= \lim_n \uparrow \int_B E^{\mathcal{B}}(X_n I_{B_{c,m}}) \quad \text{by MCT} \\ &= \lim_n \uparrow \int_B X_n I_{B_{c,m}} = \int_B \lim_n \uparrow (X_n I_{B_{c,m}}) = \int_B X I_{B_{c,m}} \quad \text{by MCT} \\ &\Rightarrow E^{\mathcal{B}}(X; B_{c,m}) \uparrow E^{\mathcal{B}}(X; B_{c,m}) \text{ a.s. by uniqueness of conditional expectation} \\ &\Rightarrow I_{B_{c,m}} E^{\mathcal{B}}X_n \uparrow I_{B_{c,m}} E^{\mathcal{B}}X \text{ a.s. by (S2), i.e. } E^{\mathcal{B}}X_n \overset{\text{a.s.}}{\uparrow} E^{\mathcal{B}}X \text{ over } B_{c,m}. \end{aligned}$$

But $\cup_{m\geq 1}\{E^{\mathcal{B}}X_m > -\infty\} = \cup_{m \geq 1} \cup_{c \text{ rational}} B_{c,m}$.

(E5)-(E6) Repeat unconditional proofs.

(E7) Use $E(I_B X) = P(B)EX$ for $B \in \mathcal{B}$.

□