# 23. Conditional Expectation

Let $X$ and $Y$ be two random variables with $Y$ taking values in $\mathbf{R}$ with $X$ taking on *only countably many values.* It often arises that we know already the value of $X$ and want to calculate the expected value of $Y$ taking into account the knowledge of $X$. That is, suppose we know that the event $\{X = j\}$ for some value $j$ has occurred. The expectation of $Y$ may change given this knowledge. Indeed, if $Q(\Lambda) = P(\Lambda | X = j)$, it makes more sense to calculate $E_Q\{Y\}$ than it does to calculate $E_P\{Y\}$ ($E_R\{\cdot\}$ denotes expectation with respect to the Probability measure $R$.)

**Definition 23.1.** *Let $X$ have values $\{x_1, x_2, \dots, x_n, \dots\}$ and $Y$ be a random variable. Then if $P(X = x_j) > 0$ the conditional expectation of $Y$ given $\{X = x_j\}$ is defined to be*

$$E\{Y | X = x_j\} = E_Q\{Y\},$$

*where $Q$ is the probability given by $Q(\Lambda) = P(\Lambda | X = x_j)$, provided $E_Q\{|Y|\} < \infty$.*

**Theorem 23.1.** *In the previous setting, and if further $Y$ is countably valued with values $\{y_1, y_2, \dots, y_n, \dots\}$ and if $P(X = x_j) > 0$, then*

$$E\{Y | X = x_j\} = \sum_{k=1}^{\infty} y_k P(Y = y_k | X = x_j),$$

*provided the series is absolutely convergent.*

*Proof.*

$$E\{Y | X = x_j\} = E_Q\{Y\} = \sum_{k=1}^{\infty} y_k Q(Y = y_k) = \sum_{k=1}^{\infty} y_k P(Y = y_k | X = x_j).$$

□

Next, still with $X$ having at most a countable number of values, we wish to define the conditional expectation of any real valued r.v. $Y$ given knowledge of the *random variable* $X$, rather than given only the event $\{X = x_j\}$. To this effect we consider the function

$$f(x) = \begin{cases} E\{Y|X = x_j\} & \text{if } P(X = x_j) > 0 \\ \text{any arbitrary value} & \text{if } P(X = x_j) = 0. \end{cases} \tag{23.1}$$

**Definition 23.2.** *Let $X$ be countably valued and let $Y$ be a real valued random variable. The conditional expectation of $Y$ given $X$ is defined to be*

$$E\{Y|X\} = f(X),$$

*where $f$ is given by (23.1) provided $f$ is well defined* (that is, $Y$ is integrable with respect to the probability measure $Q_j$ defined by by $Q_j(\Lambda) = P(\Lambda|X = x_j)$, for all $j$ such that $P(X = x_j) > 0$).

**Remark 23.1.** The above definition does not really define $E\{Y|X\}$ everywhere, but only almost everywhere since it is arbitrary on each set $\{X = x\}$ such that $P(X = x) = 0$: this will be a distinctive feature of the conditional expectation for more general r.v. $X$'s as defined below.

**Example:** Let $X$ be a Poisson random variable with parameter $\lambda$. When $X = n$, we have that each one of the $n$ outcomes has a probability of success $p$, independently of the others. Let $S$ denote the total number of successes. Let us find $E\{S|X\}$ and $E\{X|S\}$.

We first compute $E\{S|X = n\}$. If $X = n$, then $S$ is binomial with parameters $n$ and $p$, and $E\{S|X = n\} = pn$. Thus $E\{S|X\} = pX$.

To compute $E\{X|S\}$, we need to compute $E\{X|S = k\}$; to do this we first compute $P(X = n|S = k)$:

$$\begin{aligned} P(X = n|S = k) &= \frac{P(S = k|X = n)P(X = n)}{P(S = k)} \\ &= \frac{\binom{n}{k}p^k(1-p)^{n-k}\left(\frac{\lambda^n}{n!}\right)e^{-\lambda}}{\sum_{m \geq k}\binom{m}{k}p^k(1-p)^{m-k}\left(\frac{\lambda^m}{m!}\right)e^{-\lambda}} \\ &= \frac{((1-p)\lambda)^{n-k}}{(n-k)!}e^{-(1-p)\lambda} \end{aligned}$$

for $n \geq k$. Thus,

$$E\{X|S = k\} = \sum_{n \geq k} n\frac{((1-p)\lambda)^{n-k}}{(n-k)!}e^{-(1-p)\lambda} = k + (1-p)\lambda,$$

hence,

$$E\{X|S\} = S + (1-p)\lambda.$$

Finally, one can check directly that $E\{S\} = E\{E\{S|X\}\}$; also this follows from Theorem 23.3 below. Therefore, we also have that

$$E\{S\} = pE\{X\} = p\lambda.$$



Next we wish to consider the general case: that is, we wish to treat $E\{Y|X\}$ where $X$ is no longer assumed to take only countably many values. The preceding approach does not work, because the events $\{X = x\}$ in general have probability zero. Nevertheless we found in the countable case that $E\{Y|X\} = f(X)$ for a function $f$, and it is this idea that extends to the general case, with the aid of the next theorem. Let us recall a definition already given in Chapter 10:

**Definition 23.3.** *Let $X: (\Omega, \mathcal{A}) \to (\mathbf{R}^n, \mathcal{B}^n)$ be measurable. The* $\sigma$-algebra generated by $X$ *is* $\sigma(X) = X^{-1}(\mathcal{B}^n)$ (it is a $\sigma$-algebra: see the proof of Theorem 8.1), *which is also given by*

$$\sigma(X) = \left\{A \subset \Omega : X^{-1}(B) = A, \textit{ for some } B \in \mathcal{B}^n\right\}.$$

**Theorem 23.2.** *Let $X$ be an $\mathbf{R}^n$ valued random variable and let $Y$ be an $\mathbf{R}$-valued random variable. $Y$ is measurable with respect to $\sigma(X)$ if and only if there exists a Borel measurable function $f$ on $\mathbf{R}^n$ such that $Y = f(X)$.*

*Proof.* Suppose such a function $f$ exists. Let $B \in \mathcal{B}$. Then $Y^{-1}(B) = X^{-1}(f^{-1}(B))$. But $\Lambda = f^{-1}(B) \in \mathcal{B}^n$, whence $X^{-1}(\Lambda) \in \sigma(X)$ (alternatively, see Theorem 8.2).

Next suppose $Y^{-1}(B) \in \sigma(X)$, for each $B \in \mathcal{B}$. Suppose first $Y = \sum_{i=1}^k a_i 1_{A_i}$ for some $k < \infty$, with the $a_i$'s all distinct and the $A_i$'s pairwise disjoint. Then $A_i \in \sigma(X)$, hence there exists $B_i \in \mathcal{B}^n$ such that $A_i = X^{-1}(B_i)$. Let $f(x) = \sum_{i=1}^k a_i 1_{B_i}(x)$, and we have $Y = f(X)$, with $f$ Borel measurable: so the result is proved for every simple r.v. $Y$ which is $\sigma(X)$-measurable. If $Y$ is next assumed only positive, it can be written $Y = \lim_{n\to\infty} Y_n$, where $Y_n$ are simple and non-decreasing in $n$. (See for example such a construction in Chapter 9.) Each $Y_n$ is $\sigma(X)$ measurable and also $Y_n = f_n(X)$ as we have just seen. Set $f(x) = \limsup_{n\to\infty} f_n(x)$. Then

$$Y = \lim_{n\to\infty} Y_n = \lim_n f_n(X).$$

But

$$(\limsup_{n\to\infty} f_n)(X) = \limsup_n (f_n(X)).$$

and since $\limsup_{n\to\infty} f_n(x)$ is Borel measurable, we are done.

For general $Y$, we can write $Y = Y^+ - Y^-$, and we are reduced to the preceding case. □

In what follows, let $(\Omega, \mathcal{A}, P)$ be a fixed and given probability space, and let $X : \Omega \to \mathbf{R}^n$. The space $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ is the space of all random variables $Y$ such that $E\{Y^2\} < \infty$. If we identify all random variables that are equal a.s., we get the space $L^2(\Omega, \mathcal{A}, P)$. We can define an inner product (or "scalar product") by

$$\langle Y, Z\rangle = E\{YZ\}.$$

Then $L^2(\Omega, \mathcal{A}, P)$ is a Hilbert space, as we saw in Chapter 22. Since $\sigma(X) \subset \mathcal{A}$, the set $L^2(\Omega, \sigma(X), P)$ is also a Hilbert space, and it is a (closed) Hilbert subspace of $L^2(\Omega, \mathcal{A}, P)$. (Note that $L^2(\Omega, \sigma(X), P)$ has the same inner product as does $L^2(\Omega, \mathcal{A}, P)$.)

**Definition 23.4.** *Let* $Y \in L^2(\Omega, \mathcal{A}, P)$. *Then* the conditional expectation of $Y$ given $X$ *is the unique element* $\hat{Y}$ *in* $L^2(\Omega, \sigma(X), P)$ *such that*

$$E\{\hat{Y}Z\} = E\{YZ\} \text{ for all } Z \in L^2(\Omega, \sigma(X), P). \tag{23.2}$$

*We write*

$$E\{Y|X\}$$

*for the conditional expectation of* $Y$ *given* $X$, *namely* $\hat{Y}$.

Note that $\hat{Y}$ is simply the Hilbert space projection of $Y$ on the closed linear subspace $L^2(\Omega, \sigma(X), P)$ of $L^2(\Omega, \mathcal{A}, P)$: this is a consequence of Corollary 22.1 (or Exercise 23.4), and thus the conditional expectation does exist.

Observe that since $E\{Y|X\}$ is $\sigma(X)$ measurable, by Theorem 23.2 there exists a Borel measurable $f$ such that $E\{Y|X\} = f(X)$. Therefore (23.2) is equivalent to

$$E\{f(X)g(X)\} = E\{Yg(X)\} \tag{23.3}$$

for each Borel $g$ such that $g(X) \in \mathcal{L}^2$.

Next let us replace $\sigma(X)$ with simply a $\sigma$-algebra $\mathcal{G}$ with $\mathcal{G} \in \mathcal{A}$. Then $L^2(\Omega, \mathcal{G}, P)$ is a sub-Hilbert space of $L^2(\Omega, \mathcal{A}, P)$, and we can make an analogous definition:

**Definition 23.5.** *Let* $Y \in L^2(\Omega, \mathcal{A}, P)$ *and let* $\mathcal{G}$ *be a sub* $\sigma$*-algebra of* $\mathcal{A}$. *Then the conditional expectation of* $Y$ *given* $\mathcal{G}$ *is the unique element* $E\{Y|\mathcal{G}\}$ *of* $L^2(\Omega, \mathcal{G}, P)$ *such that*

$$E\{YZ\} = E\{E\{Y|\mathcal{G}\}Z\} \tag{23.4}$$

*for all* $Z \in L^2(\Omega, \mathcal{G}, P)$.

*Important Note:* The conditional expectation is an element of $L^2$, that is an "equivalence class" of random variables. Thus any statement like $E\{Y|\mathcal{G}\} \geq 0$ or $E\{Y|\mathcal{G}\} = Z$, etc... should be understood with an implicit "almost surely" qualifier, or equivalently as such: there is a "version" of $E\{Y|\mathcal{G}\}$ that is positive, or equal to $Z$, etc...

**Theorem 23.3.** *Let* $Y \in L^2(\Omega, \mathcal{A}, P)$ *and* $\mathcal{G}$ *be a sub* $\sigma$*-algebra of* $\mathcal{A}$.

a) *If* $Y \geq 0$ *then* $E\{Y|\mathcal{G}\} \geq 0$;
b) *If* $\mathcal{G} = \sigma(X)$ *for some random variable* $X$, *there exists a Borel measurable function* $f$ *such that* $E\{Y|\mathcal{G}\} = f(X)$;
c) $E\{E\{Y|\mathcal{G}\}\} = E\{Y\}$;
d) *The map* $Y \to E\{Y|\mathcal{G}\}$ *is linear.*



*Proof.* Property (b) we proved immediately preceding the theorem. For (c) we need only to apply (23.4) with $Z = 1$. Property (d) follows from (23.4) as well: if $U, V$ are in $L^2$, then

$$\begin{aligned} E\{(U + \alpha V)Z\} &= E\{UZ\} + \alpha E\{VZ\} \\ &= E\{E\{U|\mathcal{G}\}Z\} + \alpha E\{E\{V|\mathcal{G}\}Z\} \\ &= E\{(E\{U|\mathcal{G}\} + \alpha E\{V|\mathcal{G}\})Z\}, \end{aligned}$$

and thus $E\{U + \alpha V|\mathcal{G}\} = E\{U|\mathcal{G}\} + \alpha E\{V|\mathcal{G}\}$ by uniqueness (alternatively, as said before, $E\{Y|\mathcal{G}\}$ is the projection of $Y$ on the subspace $L^2(\Omega, \mathcal{G}, P)$, and projections have been shown to be linear in Corollary 22.2).

Finally for (a) we again use (23.4) and take $Z$ to be $1_{\{E\{Y|\mathcal{G}\}<0\}}$, assuming $Y \geq 0$ a.s. Then $E\{YZ\} \geq 0$ since both $Y$ and $Z$ are nonnegative, but

$$E\{E\{Y|\mathcal{G}\}Z\} = E\{E\{Y|\mathcal{G}\}1_{\{E\{Y|\mathcal{G}\}<0\}}\} < 0 \quad \text{if} \quad P(\{E\{Y|\mathcal{G}\} < 0\}) > 0.$$

This violates (23.3), so we conclude $P(\{E\{Y|\mathcal{G}\} < 0\}) = 0$. □

**Remark 23.2.** As one can see from Theorem 23.3, the key property of conditional expectation is the property *(23.4)*; our only use of Hilbert space projection was to show that the conditional expectation exists.

We now wish to extend the conditional expectation of Definition 23.4 to random variables in $L^1$, not just random variables in $L^2$. Here the technique of Hilbert space projection is no longer available to us.

Once again let $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ be the space of all $L^1$ random variables; we identify all random variables that are equal a.s. and we get the (Banach) space $L^1(\Omega, \mathcal{A}, P)$. Analogously, let $L^+(\Omega, \mathcal{A}, P)$ be all nonnegative random variables, again identifying all a.s. equal random variables. We allow random variables to assume the value $+\infty$.

**Lemma 23.1.** *Let* $Y \in L^+(\Omega, \mathcal{A}, P)$ *and let* $\mathcal{G}$ *be a sub* $\sigma$*-algebra of* $\mathcal{A}$. *There exists a unique element* $E\{Y|\mathcal{G}\}$ *of* $L^+(\Omega, \mathcal{G}, P)$ *such that*

$$E\{YX\} = E\{E\{Y|\mathcal{G}\}X\} \tag{23.5}$$

*for all* $X$ *in* $L^+(\Omega, \mathcal{G}, P)$ *and this conditional expectation agrees with the one in Definition 23.5 if further* $Y \in L^2(\Omega, \mathcal{A}, P)$. *Moreover, if* $0 \leq Y \leq Y'$, *then*

$$E\{Y|\mathcal{G}\} \leq E\{Y'|\mathcal{G}\}. \tag{23.6}$$

*Proof.* If $Y$ is in $L^2(\Omega, \mathcal{A}, P)$ and positive, we define $E\{Y|\mathcal{G}\}$ as in Definition 23.5. If $X$ in $L^+(\Omega, \mathcal{G}, P)$ then $X_n = X \wedge n$ is square-integrable. Hence the Monotone Convergence Theorem (applied twice) and (23.5) yield

$$\begin{aligned} E\{YX\} &= \lim_n E\{YX_n\} \\ &= \lim_n E\{E\{Y|\mathcal{G}\}X_n\} \\ &= E\{E\{Y|\mathcal{G}\}X\} \end{aligned} \tag{23.7}$$

and (23.5) holds for all positive $X$.

Let now $Y$ be in $L^+(\Omega, \mathcal{A}, P)$. Each $Y_m = Y \wedge m$ is bounded and hence in $L^2$, and by Theorem 23.3, conditional expectation on $L^2$ is a positive operator, so $E\{Y \wedge m|\mathcal{G}\}$ is increasing; therefore the following limit exists and we can set

$$E\{Y|\mathcal{G}\} = \lim_{m\to\infty} E\{Y_m|\mathcal{G}\}. \tag{23.8}$$

If $X \in L^+(\Omega, \mathcal{G}, P)$, we apply the Monotone Convergence Theorem several times as well as (23.8)to deduce that:

$$\begin{aligned} E\{YX\} &= \lim_m E\{Y_m X\} \\ &= E\left\{\lim_m E\{Y_m|\mathcal{G}\}X\right\} \\ &= E\left\{E\{Y|\mathcal{G}\}X\right\}. \end{aligned}$$

Furthermore if $Y \le Y'$ we have $Y\wedge m \le Y'\wedge m$ for all $m$, hence $E\{Y\wedge m|\mathcal{G}\} \le E\{Y' \wedge m|\mathcal{G}\}$ as well by Theorem 23.3(a). Therefore (23.6) holds.

It remains to establish the uniqueness of $E\{Y|\mathcal{G}\}$ as defined above. Let $U$ and $V$ be two versions of $E\{Y|\mathcal{G}\}$ and let $\Lambda_n = \{U < V \le n\}$ and suppose $P(\Lambda_n) > 0$. Note that $\Lambda_n \in \mathcal{G}$. We then have

$$E\{Y1_{\Lambda_n}\} = E\{U1_{\Lambda_n}\} = E\{V1_{\Lambda_n}\},$$

since $E\{Y1_\Lambda\} = E\{E\{Y|\mathcal{G}\}1_\Lambda\}$ for all $\Lambda \in \mathcal{G}$ by (23.7). Further, $0 \le U1_{\Lambda_n} \le V1_{\Lambda_n} \le n$, and $P(\Lambda_n) > 0$ implies that the r.v. $V1_{\Lambda_n}$ and $U1_{\Lambda_n}$ are not a.s. equal: we deduce that $E\{U1_\Lambda\} < E\{V1_\Lambda\}$, whence a contradiction. Therefore $P(\Lambda_n) = 0$ for all $n$, and since $\{U > V\} = \cup_{n\ge1}\Lambda_n$ we get $P\{U < V\}) = 0$; analogously $P(\{V > U\}) = 0$, and we have uniqueness. □

**Theorem 23.4.** *Let $Y \in L^1(\Omega, \mathcal{A}, P)$ and let $\mathcal{G}$ be a sub $\sigma$-algebra of $\mathcal{A}$. There exists a unique element $E\{Y|\mathcal{G}\}$ of $L^1(\Omega, \mathcal{G}, P)$ such that*

$$E\{YX\} = E\{E\{Y|\mathcal{G}\}X\} \tag{23.9}$$

*for all bounded $\mathcal{G}$-measurable $X$ and this conditional expectation agrees with the one in Definition 23.5 (resp. Lemma 23.1) when further $Y \in L^2(\Omega, \mathcal{A}, P)$ (resp. $Y \ge 0$), and satisfies*

a) *If $Y \ge 0$ then $E\{Y|\mathcal{G}\} \ge 0$;*
b) *The map $Y \to E\{Y|\mathcal{G}\}$ is linear.*

*Proof.* Since $Y$ is in $L^1$, we can write

$$Y = Y^+ - Y^-$$

where $Y^+ = \max(Y, 0)$ and $Y^- = -\min(Y, 0)$: moreover $Y^+$ and $Y^-$ are also in $L^1(\Omega, \mathcal{G}, P)$. Next set

$$E\{Y|\mathcal{G}\} = E\{Y^+|\mathcal{G}\} - E\{Y^-|\mathcal{G}\}.$$



This formula makes sense: indeed the r.v. $Y^+$ and $Y^-$, hence $E\{Y^+|\mathcal{G}\}$ and $E\{Y^-|\mathcal{G}\}$ as well by Theorem 23.3(c), are integrable, hence a.s. finite. That $E\{Y|\mathcal{G}\}$ satisfies (23.9) follows from Lemma 23.1. For uniqueness, let $U, V$ be two versions of $E\{Y|\mathcal{G}\}$, and let $\Lambda = \{U < V\}$. Then $\Lambda \in \mathcal{G}$, so $1_\Lambda$ is bounded and $\mathcal{G}$-measurable. Then $E\{Y1_\Lambda\} = E\{E\{Y|\mathcal{G}\}1_\Lambda\} = E\{U1_\Lambda\} = E\{V1_\Lambda\}$. But if $P(\Lambda) > 0$, then $E\{U1_\Lambda\} < E\{V1_\Lambda\}$, which is a contradiction. So $P(\Lambda) = 0$ and analogously $P(\{V < U\}) = 0$ as well.

The final statements are trivial consequences of the previous definition of $E\{Y|\mathcal{G}\}$ and of Lemma 23.1 and Theorem 23.3. □

**Example:** Let $(X, Z)$ be real-valued random variables having a joint density $f(x, z)$. Let $g$ be a bounded function and let

$$Y = g(Z).$$

We wish to compute $E\{Y|X\} = E\{g(Z)|X\}$. Recall that $X$ has density $f_X$ given by

$$f_X(x) = \int f(x, z)dz$$

and we defined in Chapter 12 (see Theorem 12.2) a *conditional density for $Z$ given $X = x$* by:

$$f_{X=x}(z) = \frac{f(x, z)}{f_X(x)},$$

whenever $f_X(x) \ne 0$. Next consider

$$h(x) = \int g(z)f_{X=x}(z)dz.$$

We then have, for any bounded Borel function $k(x)$:

$$\begin{aligned} E\{h(X)k(X)\} &= \int h(x)k(x)f_X(x)dx \\ &= \iint g(z)f_{X=x}(z)dz\, k(x)f_X(x)dx \\ &= \iint g(z)\frac{f(x, z)}{f_X(x)}k(x)f_X(x)dz\, dx \\ &= \iint g(z)k(x)f(x, z)dz\, dx \\ &= E\{g(Z)k(X)\} = E\{Yk(X)\}. \end{aligned}$$

Therefore by (23.9) we have that

$$E\{Y|X\} = h(X).$$

This gives us an explicit way to calculate conditional expectations in the case when we have densities.

**Theorem 23.5.** *Let $Y$ be a positive or integrable r.v. on $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G}$ be a sub $\sigma$-algebra. Then $E\{Y|\mathcal{G}\} = Y$ if and only if $Y$ is $\mathcal{G}$-measurable.*

*Proof.* This is trivial from the definition of conditional expectation. □

**Theorem 23.6.** *Let $Y \in L^1(\Omega, \mathcal{A}, P)$ and suppose $X$ and $Y$ are independent. Then*

$$E\{Y|X\} = E\{Y\}.$$

*Proof.* Let $g$ be bounded Borel. Then $E\{Yg(X)\} = E\{Y\}E\{g(X)\}$ by independence. Thus taking $f(x) = E\{Y\}$ for all $x$ (the constant function) in Theorem 23.2, we have the result by (23.9). □

**Theorem 23.7.** *Let $X, Y$ be random variables on $(\Omega, \mathcal{A}, P)$, let $\mathcal{G}$ be a sub $\sigma$-algebra of $\mathcal{A}$, and suppose that $X$ is $\mathcal{G}$-measurable. In the two following cases:*

a) *the variables $X$, $Y$ and $XY$ are integrable,*
b) *the variables $X$ and $Y$ are positive,*

*we have*

$$E\{XY|\mathcal{G}\} = XE\{Y|\mathcal{G}\}.$$

*Proof.* Assume first (b). For any $\mathcal{G}$-measurable positive r.v. $Z$ we have

$$E\{XYZ\} = E\{XZE\{Y|\mathcal{G}\}\}$$

by (23.5). Since $XE\{Y|\mathcal{G}\}$ is also $\mathcal{G}$-measurable, we deduce the result by another application of the characterization (23.5).

In case (a), we observe that $X^+Y^+$, $X^-Y^+$, $X^+Y^-$ and $X^-Y^-$ are all integrable and positive. Then $E\{X^+Y^+|\mathcal{G}\} = X^+E\{Y^+|\mathcal{G}\}$ by what precedes, and similarly for the other three products, and all these quantities are finite. It remains to apply the linearity of the conditional expectation and the property $XY = X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+$. □

Let us note the important observation that the principal convergence theorems also hold for conditional expectations (we choose to emphasize below the fact that all statements about conditional expectations are "almost sure"):

**Theorem 23.8.** *Let $(Y_n)_{n\geq 1}$ be a sequence of r.v.'s on $(\Omega, \mathcal{A}, P)$ and let $\mathcal{G}$ be a sub $\sigma$-algebra of $\mathcal{A}$.*

a) *(Monotone Convergence.) If $Y_n \geq 0$, $n \geq 1$, and $Y_n$ increases to $Y$ a.s., then*

$$\lim_{n\to\infty} E\{Y_n|\mathcal{G}\} = E\{Y|\mathcal{G}\} \quad a.s.;$$



b) *(Fatou's Lemma.) If $Y_n \geq 0$, $n \geq 1$, then*

$$E\{\liminf_{n\to\infty} Y_n|\mathcal{G}\} \leq \liminf_{n\to\infty} E\{Y_n|\mathcal{G}\} \quad a.s.;$$

c) *(Lebesgue's dominated convergence theorem.) If $\lim_{n\to\infty} Y_n = Y$ a.s. and $|Y_n| \leq Z$ $(n \geq 1)$ for some $Z \in L^1(\Omega, \mathcal{A}, P)$, then*

$$\lim_{n\to\infty} E\{Y_n|\mathcal{G}\} = E\{Y|\mathcal{G}\} \quad a.s..$$

*Proof.* a) By (23.6) we have $E\{Y_{n+1}|\mathcal{G}\} \geq E\{Y_n|\mathcal{G}\}$ a.s., each $n$; hence $U = \lim_{n\to\infty} E\{Y_n|\mathcal{G}\}$ exists a.s. Then for all positive and $\mathcal{G}$-measurable r.v. $X$ we have:

$$E\{UX\} = \lim_{n\to\infty} E\{E\{Y_n|\mathcal{G}\}X\}$$
$$= \lim_{n\to\infty} E\{Y_nX\}$$

by (23.5); and

$$= \lim_{n\to\infty} E\{YX\}$$

by the usual monotone convergence theorem. Thus $U = E\{Y|\mathcal{G}\}$, again by (23.5).

The proofs of (b) and (c) are analogous in a similar vein to the proofs of Fatou's lemma and the Dominated Convergence Theorem without conditioning. □

We end with three useful inequalities.

**Theorem 23.9 (Jensen's Inequality).** *Let $\varphi: \mathbf{R} \to \mathbf{R}$ be convex, and let $X$ and $\varphi(X)$ be integrable random variables. For any $\sigma$-algebra $\mathcal{G}$,*

$$\varphi \circ E\{X|\mathcal{G}\} \leq E\{\varphi(X)|\mathcal{G}\}.$$

*Proof.* A result in real analysis is that if $\varphi : \mathbf{R} \to \mathbf{R}$ is convex, then $\varphi(x) = \sup_n(a_nx + b_n)$ for a countable collection of real numbers $(a_n, b_n)$. Then

$$E\{a_nX + b_n|\mathcal{G}\} = a_nE\{X|\mathcal{G}\} + b_n.$$

But $E\{a_nX + b_n|\mathcal{G}\} \leq E\{\varphi(X)|\mathcal{G}\}$, hence $a_nE\{X|\mathcal{G}\} + b_n \leq E\{\varphi(X)|\mathcal{G}\}$, all $n$. Taking the supremum in $n$, we get the result. □

Note that $\varphi(x) = x^2$ is of course convex, and thus as a consequence of Jensen's inequality we have

$$(E\{X|\mathcal{G}\})^2 \leq E\{X^2|\mathcal{G}\}.$$

An important consequence of Jensen's inequality is Hölder's inequality for random variables.

**Theorem 23.10 (Hölder's Inequality).** *Let $X, Y$ be random variables with $E\{|X|^p\} < \infty$, $E\{|Y|^q\} < \infty$, where $p > 1$, and $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$|E\{XY\}| \leq E\{|XY|\} \leq E\{|X|^p\}^{\frac{1}{p}} E\{|X|^q\}^{\frac{1}{q}}.$$

*(Hence if $X \in L^p$ and $Y \in L^q$ with $p, q$ as above, then the product $XY$ belongs to $L^1$).*

*Proof.* Without loss of generality we can assume $X \geq 0$, $Y \geq 0$ and $E\{X^p\} > 0$, since $E\{X^p\} = 0$ implies $X^p = 0$ a.s., thus $X = 0$ a.s. and there is nothing to prove. Let $C = E\{X^p\} < \infty$. Define a new probability measure $Q$ by

$$Q(\Lambda) = \frac{1}{C} E\{1_\Lambda X^p\}.$$

Next define $Z = \frac{Y}{X^{p-1}} 1_{\{X>0\}}$. Since $\varphi(x) = |x|^p$ is convex, Jensen's inequality (Theorem 23.9) yields

$$(E_Q\{Z\})^q \leq E\{Z^q\}.$$

Thus,

$$\begin{aligned}
\frac{1}{C^q} E\{XY\}^q &= \frac{1}{C^q} E\left\{\frac{Y}{X^{p-1}} X^p\right\}^q \\
&= \left(E_Q\left\{\frac{Y}{X^{p-1}}\right\}\right)^q \\
&\leq E_Q\left\{\left(\frac{Y}{X^{p-1}}\right)^q\right\} \\
&= \frac{1}{C} E\left\{\left(\frac{Y}{X^{p-1}}\right)^q X^p\right\} \\
&= \frac{1}{C} E\left\{Y^q \frac{1}{X^{(p-1)q}} X^p\right\},
\end{aligned}$$

and $q = \frac{p}{p-1}$ while $(p-1)q = p$, hence

$$\begin{aligned}
&= \frac{1}{C} E\left\{Y^q \frac{1}{X^p} X^p\right\} \\
&= \frac{1}{C} E\{Y^q\}.
\end{aligned}$$

Thus

$$E\{XY\}^q \leq C^{q-1} E\{Y^q\},$$

and taking $q^{th}$ roots yields

$$E\{XY\} \leq C^{\frac{q-1}{q}} E\{Y^q\}^{\frac{1}{q}}.$$

Since $\frac{q-1}{q} = \frac{1}{p}$ and $C = E\{X^p\}$, we have the result. □



**Corollary 23.1 (Minkowski's Inequality).** *Let $X, Y$ be random variables and $1 \leq p < \infty$ with $E\{|X|^p\} < \infty$ and $E\{|Y|^p\} < \infty$. Then*

$$E\{|X+Y|^p\}^{\frac{1}{p}} \leq E\{X^p\}^{\frac{1}{p}} + E\{Y^p\}^{\frac{1}{p}}.$$

*Proof.* If $p = 1$ the result is trivial. We therefore asume that $p > 1$. We use Hölder's inequality (Theorem 23.10). We have

$$\begin{aligned}
E\{|X+Y|^p\} &= E\left\{|X|\,|X+Y|^{p-1}\right\} + E\left\{|Y|\,|X+Y|^{p-1}\right\} \\
&\leq E\{|X|^p\}^{\frac{1}{p}} E\{|X+Y|^{(p-1)q}\}^{\frac{1}{q}} + E\{|Y|^p\}^{\frac{1}{p}} E\{|X+Y|^{(p-1)q}\}^{\frac{1}{q}}.
\end{aligned}$$

But $(p-1)q = p$, and $\frac{1}{q} = 1 - \frac{1}{p}$, hence

$$= \left(E\{|X|^p\}^{\frac{1}{p}} + E\{|Y|^p\}^{\frac{1}{p}}\right) E\{|X+Y|^p\}^{1-\frac{1}{p}}$$

and we have the result. □

Minkowski's inequality allows one to define a norm (satisfying a triangle inequality) on the space $L^p$ of equivalence classes (for the relation "equality a.s.") of random variables with $E\{|X|^p\} < \infty$.

**Definition 23.6.** *For $X$ in $L^p$, define a norm by*

$$\|X\|_p = E\{|X^p|\}^{\frac{1}{p}}.$$

Note that Minkowski's inequality shows that $L^p$ is a bonafide normed linear space. In fact it is even a *complete* normed linear space (called a "Banach space"). But for $p \neq 2$ it is not a Hilbert space: the norm is not associated with an inner product.

## Exercises for Chapter 23

For Exercises 23.1–23.6, let $Y$ be a positive or integrable random variable on the space $(\Omega, \mathcal{A}, P)$ and $\mathcal{G}$ be a sub $\sigma$-algebra of $\mathcal{A}$.

**23.1** Show $|E\{Y|\mathcal{G}\}| \leq E\{|Y||\mathcal{G}\}$.

**23.2** Suppose $\mathcal{H} \subset \mathcal{G}$ where $\mathcal{H}$ is a sub $\sigma$-algebra of $\mathcal{G}$. Show that

$$E\{E\{Y|\mathcal{G}\}|\mathcal{H}\} = E\{Y|\mathcal{H}\}.$$

**23.3** Show that $E\{Y|Y\} = Y$ a.s.

**23.4** Show that if $|Y| \leq c$ a.s. then $|E\{Y|\mathcal{G}\}| \leq c$ a.s. also.

**23.5** If $Y = \alpha$ a.s., with $\alpha$ a constant, show that $E\{Y|\mathcal{G}\} = \alpha$ a.s.

**23.6** If $Y$ is positive, show that $\{E\{Y|\mathcal{G}\} = 0\} \subset \{Y = 0\}$ and $\{Y = +\infty\} \subset \{E\{Y|\mathcal{G}\} = +\infty\}$ almost surely.

**23.7** * Let $X, Y$ be independent and let $f$ be Borel such that $f(X,Y) \in L^1(\Omega, \mathcal{A}, P)$. Let

$$g(x) = \begin{cases} E\{f(x,Y)\} & \text{if } |E\{f(x,Y)\}| < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $g$ is Borel on $\mathbf{R}$ and that

$$E\{f(X,Y)|X\} = g(X).$$

**23.8** Let $Y$ be in $L^2(\Omega, \mathcal{A}, P)$ and suppose $E\{Y^2 \mid X\} = X^2$ and $E\{Y \mid X\} = X$. Show $Y = X$ a.s.

**23.9** * Let $Y$ be an exponential r.v. such that $P(\{Y > t\}) = e^{-t}$ for $t > 0$. Calculate $E\{Y \mid Y \wedge t\}$, where $Y \wedge t = \min(t, Y)$.

**23.10** (Chebyshev's inequality). Prove that for $X \in L^2$ and $a > 0$, $P(|X| \geq a|\mathcal{G}) \leq \frac{E\{X^2|\mathcal{G}\}}{a^2}$.

**23.11** (Cauchy-Schwarz). For $X, Y$ in $L^2$ show

$$(E\{XY|\mathcal{G}\})^2 \leq E\{X^2|\mathcal{G}\}E\{Y^2|\mathcal{G}\}.$$

**23.12** Let $X \in L^2$. Show that

$$E\{(X - E\{X|\mathcal{G}\})^2\} \leq E\{(X - E\{X\})^2\}.$$

**23.13** Let $p \geq 1$ and $r \geq p$. Show that $L^p \supset L^r$, for expectation with respect to a probability measure.



**23.14** * Let $Z$ be defined on $(\Omega, \mathcal{F}, P)$ with $Z \geq 0$ and $E\{Z\} = 1$. Define a new probability $Q$ by $Q(\Lambda) = E\{1_\Lambda Z\}$. Let $\mathcal{G}$ be a sub $\sigma$-algebra of $\mathcal{F}$, and let $U = E\{Z|\mathcal{G}\}$. Show that $E_Q\{X|\mathcal{G}\} = \frac{E\{XZ|\mathcal{G}\}}{U}$, for any bounded $\mathcal{F}$-measurable random variable $X$. (Here $E_Q\{X|\mathcal{G}\}$ denotes the conditional expectation of $X$ relative to the probability measure $Q$.)

**23.15** Show that the normed linear space $L^p$ is complete for each $p$, $1 \leq p < \infty$. (*Hint:* See the proof of Theorem 22.2.)

**23.16** Let $X \in L^1(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}, \mathcal{H}$ be sub $\sigma$-algebras of $\mathcal{F}$. Moreover let $\mathcal{H}$ be independent of $\sigma(\sigma(X), \mathcal{G})$. Show that $E\{X|\sigma(\mathcal{G}, \mathcal{H})\} = E\{X|\mathcal{G}\}$.

**23.17** Let $(X_n)_{n\geq 1}$ be independent and in $L^1$ and let $S_n = \sum_{i=1}^n X_i$ and $\mathcal{G}_n = \sigma(S_n, S_{n+1}, \ldots)$. Show that $E\{X_1|\mathcal{G}_n\} = E\{X_1 \mid S_n\}$ and also $E\{X_j|\mathcal{G}_n\} = E\{X_j \mid S_n\}$ for $1 \leq j \leq n$. Also show that $E\{X_j|\mathcal{G}_n\} = E\{X_1|S_n\}$ for $1 \leq j \leq n$ (*Hint:* Use Exercise 23.16.)