.

Notes on Elementary Martingale Theory

by John B. Walsh

1 Conditional Expectations

1.1 Motivation

Probability is a measure of ignorance. When new information decreases that ignorance, it changes our probabilities. Suppose we roll a pair of dice, but don't look immediately at the outcome. The result is there for anyone to see, but if we haven't yet looked, as far as we are concerned, the probability that a two ("snake eyes") is showing is the same as it was before we rolled the dice, 1/36. Now suppose that we happen to see that one of the two dice shows a one, but we do not see the second. We reason that we have rolled a two if—and only if—the unseen die is a one. This has probability 1/6. The extra information has changed the probability that our roll is a two from 1/36 to 1/6. It has given us a new probability, which we call a *conditional probability.*

In general, if A and B are events, we say the conditional probability that B occurs given that A occurs is the *conditional probability of* B *given* A. This is given by the well-known formula

$$P\{B \mid A\} = \frac{P\{A \cap B\}}{P\{A\}}, \tag{1}$$

providing $P\{A\} > 0$. (Just to keep ourselves out of trouble if we need to apply this to a set of probability zero, we make the convention that $P\{B \mid A\} = 0$ if $P\{A\} = 0$.) Conditional probabilities are familiar, but that doesn't stop them from giving rise to many of the most puzzling paradoxes in probability. We want to study a far-reaching generalization of this, called a conditional expectation. The final definition is going to look rather abstract if we give it without preparation, so we will try to sneak up on it.

Note that (1) defines a new probability measure on the sets B of $(\Omega, \mathcal{F})$, and we can define an expectation with respect to it. If X is an integrable random variable, then it will also be integrable with respect to the conditional probability of (1). (See the exercises.) Thus we can define the *conditional expectation of* X *given* A by

$$E\{X \mid A\} = \sum_x x P\{X = x \mid A\},$$

for discrete X, and we can extend it to the general integrable X by taking limits, as in Section **??**.

First, why do we insist on conditional expectations rather than conditional probabilities? That is simple: conditional probabilities are just a special case of conditional expectations.

If B is an event, then, I_B is a random variable with $P\{I_B = 1 \mid A\} = P\{B \mid A\}$, and $P\{I_B = 0 \mid A\} = 1 - P\{B \mid A\}$ so that

$$P\{B \mid A\} = E\{I_B \mid A\},$$

i.e. we can get conditional probabilities by taking conditional expectations of indicator functions. Thus we will concentrate on conditional expectations rather than conditional probabilities.

If X has possible values (x_i),

$$\begin{aligned} E\{X \mid A\} &= \sum_i x_i P\left\{X = x_i \mid A\right\} = \frac{1}{P\{A\}} \sum_i x_i P\{\{X = x_i\} \cap A\} \\ &= \frac{1}{P\{A\}} \sum_i x_i E\{I_{\{X=x_i\}} I_A\} = \frac{1}{P\{A\}} E\left\{\left(\sum_i x_i I_{\{X=x_i\}}\right) I_A\right\} \\ &= \frac{1}{P\{A\}} E\{X I_A\} \equiv \frac{1}{P\{\,A\,\}} \int_A X\, dP\,. \end{aligned}$$

Thus, in general, $E\{X \mid A\} = (\int_A X\, dP)/P\{\,A\,\}$: the conditional expectation is the average of X over A.

We will need much more general conditioning than this. For instance, we could condition on the value of a random variable. Let X and Y be random variables; suppose Y is discrete. Then we can define $E\{X \mid Y = y\}$ for every value of y such that $P\{Y = y\} > 0$. If $P\{Y = y\} = 0$, we arbitrarily set $E\{X \mid Y = y\} = 0$. Now of course the conditional expectation depends on Y, so it is actually a function of Y, which we can write:

$$E\{X \mid Y\}(\omega) \stackrel{\text{def}}{=} \sum_y E\{X \mid Y = y\} I_{\{Y=y\}}(\omega),\ \omega \in \Omega\,. \tag{2}$$

By our remarks above, $E\{X \mid Y\}$ is a random variable, and

$$E\{X \mid Y\} = \frac{1}{P\{Y = y_i\}} \int_{\{Y=y_i\}} X\, dP \tag{3}$$

on the set $\{Y = y_i\}$.

Remark 1.1 *Note that* $E\{X \mid Y\}$ *is constant on each set* $\{Y = y_i\}$*, and its value is equal to the* average *value of* X *on* $\{Y = y_i\}$*. This means that first, conditional expectations are random variables, and second, conditional expectations are averages.*

Let us summarize the properties of $E\{X \mid Y\}$ for a discrete random variable Y.

(A) It is a random variable.

(B) It is a function of Y.

(C) It has the same integral as X over sets of the form $\{\omega : Y(\omega) = y\}$

$$\int_{\{Y=y\}} E\{X \mid Y\}\, dP = \int_{\{Y=y\}} \left[\frac{1}{P\{Y = y\}} \int_{\{Y=y\}} X\, dP\right] dP = \int_{\{Y=y\}} X\, dP/,.$$

since the term in square brackets is a constant, hence the $P\{Y = y\}$ cancel.

If B is a Borel set and (y_i) are the possible values of Y, $\{Y \in B\} = \cup_{i:y_i \in B}\{Y = y_i\}$, and (3) tells us that

$$\int_{\{Y\in B\}} E\{X \mid Y\}\, dP = \sum_{i:y_i\in B} \int_{\{Y=y_i\}} E\{X \mid Y\}\, dP = \sum_{i:y_i\in B} \int_{\{Y=y_i\}} X\, dP = \int_{\{Y\in B\}} X\, dP\,,$$

as long as the sum converges absolutely. But $|\int_{\{Y=y_i\}} X\, dP| \le \sum_{i:y_i\in B} \int_{\{Y=y_i\}} |X|\, dP$ and $\sum_i \int_{\{Y=y_i\}} |X|\, dP = E\{|X|\} < \infty$ since X is integrable. In short, $E\{X \mid Y\}$ satisfies the following properties:

(a) It is a function of Y.

(b) $\displaystyle\int_{\{Y\in B\}} E\{X \mid Y\}\, dP = \int_{\{Y\in B\}} X\, dP\,.$

Let's translate this into the language of sigma fields. According to Proposition **??** a random variable is a function of Y iff it is $\mathcal{F}_Y$-measurable, and by Propsition **??**, sets of the form $\{Y \in B\}$ comprise $\mathcal{F}_Y$, so we can express (a) purely in terms of measurability and (b) in terms of $\mathcal{F}_Y$, so that we have

(4) $$E\{X \mid Y\} \text{ is } \mathcal{F}_Y\text{-measurable.}$$

(5) $$\int_\Lambda E\{X \mid Y\}\, dP = \int_\Lambda X\, dP\,, \quad \forall \Lambda \in \mathcal{F}_Y\,.$$

1.2 Conditional Expectations Defined

Rather than trying to give a constructive definition of conditional expectations, we will give a set of properties which they satisfy. As long as these properties characterize them uniquely, this is a perfectly legal definition. At the same time, we will generalize the idea. Notice that (4) and (5) characterize $E\{ X \mid Y \}$ entirely in terms of the sigma field $\mathcal{F}_Y$. Sigma fields contain information—in fact, to all intents and purposes, sigma fields *are* information, so we will define conditional expectations relative to argitrary sigma fields. If $\mathcal{G}$ is a sigma field, then $E\{ X \mid \mathcal{G} \}$ will be the conditional expectation of X, given all the information in $\mathcal{G}$. This will seem strange at first, but it is in fact the most natural way to define it. So, we just replace $\mathcal{F}_Y$ by a general sigma field $\mathcal{G}$ in (4) and (5) to get:

Definition 1.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}$ be a sub-sigma field of $\mathcal{F}$. If X is an integrable random variable, then the* **conditional expectation of X given $\mathcal{G}$** *is any random variable Z which satisfies the following two properties:*

(CE1) Z is $\mathcal{G}$-measurable;

(CE2) if $\Lambda \in \mathcal{G}$, then

$$\int_\Lambda Z \, dP = \int_\Lambda X \, dP \,. \tag{6}$$

We denote Z by $E\{X \mid \mathcal{G}\}$.

Remark 1.2 *It is implicit in (CE2) that Z must be integrable.*

There are some immediate questions which arise; most of them can be answered by looking at the example of $E\{X \mid Y\}$. Let's take them in order. Perhaps the most pressing is, "What is the role of the sigma field in all this?" The answer is that a sigma field represents information. For instance, if $\mathcal{G}$ is generated by a random variable Y, $\mathcal{G}$ will contain all of the sets $\{Y \in B\}$ for Borel B. Thus we can reconstruct Y from the sets in $\mathcal{G}$. In other words, if we know $\mathcal{G}$, we also know Y.

The second question is, "Why should the conditional expectation be a random variable?" Look again at the example where $\mathcal{G}$ is generated by Y. We expect that $E\{X \mid \mathcal{G}\} = E\{X \mid Y\}$, and the latter is a function of Y; as Y is a random variable, so is a function of Y, and the conditional expectation should indeed be a random variable.

The third question is, "How does measurability come into this?" In the case $\mathcal{G} = \mathcal{F}_Y$, the conditional expectation should be a function of Y, as we saw just above. But by Proposition **??**, if Z is $\mathcal{F}_Y$-measurable, it is indeed a function of Y. The next—but not last—question we might ask is "What does (CE2) mean?" Look at (C). This equation says that the conditional expectation is an average of X over certain sets. Since (CE2) is a generalization of (C), we can interpret it to mean that $E\{X \mid \mathcal{G}\}$ is an average in a certain sense over sets of $\mathcal{G}$.

It remains to be seen how we can actually use (CE1) and (CE2). This is best seen by simply proving some facts about conditional expectations and observing how they are used. It happens, rather surprisingly, that even in cases where we have an explicit formula for the conditional expectation, it is much quicker to use the definition, rather than the formula, to prove things.

Proposition 1.3 *If Z and Z' are two random variables satisfying (CE1) and (CE2), then $Z = Z'$ a.e.*

Proof. Since Z and Z' are $\mathcal{G}$-measurable by (CE1), $\{Z - Z' > 0\} \in \mathcal{G}$. (This takes a little proof: $\{Z - Z' > 0\} = \{Z > Z'\} = \cup_{r \in \mathcal{Q}} \{Z' < r\} \cap \{Z \geq r\}$ which is in $\mathcal{G}$ since both $\{Z' < r\}$ and $\{Z \geq r\}$ are.) Apply (CE2) to both Z and Z':

$$\int_{\{Z>Z'\}} (Z - Z')\, dP = \int_{\{Z>Z'\}} Z\, dP - \int_{\{Z>Z'\}} Z'\, dP = \int_{\{Z>Z'\}} X\, dP - \int_{\{Z>Z'\}} X\, dP = 0.$$

But since $Z - Z' > 0$ on $\{Z > Z'\}$, we must have $P\{Z > Z'\} = 0$. Thus $Z \leq Z'$ a.e. Now reverse Z and Z' to see $Z' \leq Z$ a.e. as well, so $Z = Z'$ a.e. ♣

Unfortunately, the conditional expectation is not uniquely defined. It is only defined up to sets of measure zero—changing Z on a set of measure zero doesn't change the integrals in (CE2)—so two candidates are only equal a.e., but not necessarily identical. If one wanted to be pedantic, one could insist that the conditional expectation was an equivalence class of r.v.s which are a.e. equal and satisfy (CE1) and (CE2). But we won't do this, since in fact, ambiguities on a set of probability zero seldom difficulties. (We had better quickly qualify that: while null sets seldom cause difficulties, when they do, they are likely to be serious ones! As we are just about to see.)

This definition may appear to be overly abstract, but it is not: it is needed in order to handle the general case—indeed, one of the triumphs of the measure-theoretic approach is that it handles conditional expectations correctly. The reason that conditional expectations are so tricky to handle is a problem of null sets–and this is a case where the null-set difficulty is serious: if $P\{Y = y\} = 0$ for all y, how is one to define $E\{X \mid Y\}$? The formula (3) simply does not work. Yet this exactly what happens for any r.v. with a continuous distribution function. One can extend the formula to random variables having probability densities, but this still leaves many important cases uncovered.

This leaves us with an important question: "Does the conditional expectation always exist?" That is, we have a definition: a conditional expectation is any r.v. satisfying the two properties. But does *any* r.v. actually satisfy them? The answer turns out to be "yes," but—and this is almost embarrassing—there is no way we can prove this with the tools we have at hand. Still, there is one case in which we can prove there is.

Consider the case where the sigma field $\mathcal{G}$ is generated by a partition of disjoint sets $\Lambda_1, \Lambda_2, \ldots$, with $\Lambda_i \in \mathcal{F}$, $\cup_i \Lambda_i = \Omega$.

Proposition 1.4 *Let X be an integrable random variable, and $\mathcal{G} \subset \mathcal{F}$ the sigma-field generated by the partition Λ_i, $i = 1, 2, \ldots$. Then with probability one,*

$$E\{X \mid \mathcal{G}\} = \sum_i \frac{E\{X I_{\Lambda_i}\}}{P\{\Lambda_i\}} I_{\Lambda_i}\,. \tag{7}$$

In particular, with probability one

$$E\{X \mid Y\} = E\{X \mid \mathcal{F}_Y\}. \tag{8}$$

PROOF. In fact, we have already proved (8), since (4) and (5) are equivalent to (CE1) and (CE2) in this case. But the seemingly more general case (7) where $\mathcal{G}$ is generated by the Λ_i follows from (8). We need only define the random variable Y by $Y = i$ on the set Λ_i. Then $\mathcal{G} = \mathcal{F}_Y$ and $E\{X \mid \mathcal{G}\} = E\{X \mid \mathcal{F}_Y\}$. ♣

Remark 1.5 *Warning Number One In what follows, we can prove the existence of conditional expectations when the sigma fields are generated by partitions, but not otherwise. So, in effect, we will be assuming that all sigma fields are generated by partitions in what follows. In the end, we will show the existence of conditional expectations in general—using only sigma fields generated by partitions—so that all we do below will be correct in full generality.*

Remark 1.6 *Warning Number Two Everything will work out neatly in the end, so don't worry about Warning Number One.*

1.3 Elementary Properties

Theorem 1.7 *Let X and Y be integrable random variables, a and b real numbers. Then*
(i) $E\{E\{X \mid \mathcal{G}\}\} = E\{X\}$.
(ii) $\mathcal{G} = \{\phi, \Omega\} \Longrightarrow E\{X \mid \mathcal{G}\} = E\{X\}$ *a.e.*
(iii) If X is $\mathcal{G}$-measurable, $E\{X \mid \mathcal{G}\} = X$ *a.e.*
(iv) $E\{aX + bY \mid \mathcal{G}\} = aE\{X \mid \mathcal{G}\} + bE\{Y \mid \mathcal{G}\}$ *a.e.*
(v) If $X \geq 0$ a.e., $E\{X \mid \mathcal{G}\} \geq 0$ *a.e.*
(vi) If $X \leq Y$ a.e., $E\{X \mid \mathcal{G}\} \leq E\{Y \mid \mathcal{G}\}$ *a.e.*
(vii) $|E\{X \mid \mathcal{G}\}| \leq E\{|X| \mid \mathcal{G}\}$ *a.e.*
(viii) Suppose Y is $\mathcal{G}$-measurable and XY is integrable. Then

$$E\{XY \mid \mathcal{G}\} = YE\{X \mid \mathcal{G}\} \ \text{a.e.} \tag{9}$$

(ix) If X_n and X are integrable, and if either $X_n \uparrow X$, or $X_n \downarrow X$, then

$$E\{X_n \mid \mathcal{G}\} \longrightarrow E\{X \mid \mathcal{G}\} \ \text{a.e.}$$

PROOF. (i) Just take $\Lambda = \Omega$ in (CE2).

(ii) E
$lbr X$ }, considered as a r.v., is constant, hence $\mathcal{G}$-measurable. The only sets on which to check (CE2) are ϕ, where it is trivial, and Ω, where it is obvious. This means $Z \stackrel{\text{def}}{=} E\{X\}$ satisfies the properties of the conditional expectation, and hence it *is* the conditional

expectation. Note that we only have equality a.e. since the conditional expectation is only defined up to sets of measure zero. Another way of putting this is to say that $E\{X\}$ is a *version* of the conditional expectation.

(iii) Set $Z \equiv X$. Then (CE1) and (CE2) are immediate, and X must be a version of the conditional expectation.

(iv) Once more we verify that the right-hand side satisfies (CE1) and (CE2). It is clearly $\mathcal{G}$-measurable, and if $\Lambda \in \mathcal{G}$, apply (CE2) to X and Y to see that

$$\int_\Lambda aE\{X \mid \mathcal{G}\} + bE\{Y \mid \mathcal{G}\}\, dP = a\int_\Lambda X\, dP + b\int_\Lambda Y\, dP = \int_\Lambda aX + bY\, dP\,.$$

(v) Take $\Lambda = \{E\{X \mid \mathcal{G}\} < 0\} \in \mathcal{G}$. Then by (CE2)

$$0 \geq \int_\Lambda E\{X \mid \mathcal{G}\} = \int_\Lambda X \geq 0 \Longrightarrow P\{\Lambda\} = 0.$$

(vi) Let $Z = Y - X$ and apply (v).

(vii) This follows from (vi) and (iv) since $X \leq |X|$ and $-X \leq |X|$.

$(viii)$ The right-hand side of (9) is $\mathcal{G}$-measurable, so (CE1) holds. Assume first that both X and Y are positive and let $\Lambda \in \mathcal{G}$. First suppose $Y = I_\Gamma$ for some $\Gamma \in \mathcal{G}$. Then

$$\int_\Lambda Y E\{X \mid \mathcal{G}\}\, dP = \int_{\Lambda\cap\Gamma} E\{X \mid \mathcal{G}\}\, dP = \int_{\Lambda\cap\Gamma} X\, dP = \int_\Lambda XY\, dP$$

Thus (CE2) holds in this case. Now we pull ourselves up by our bootstraps. It follows from (iv) that (CE2) holds if Y is a finite linear combination of indicator functions, i.e. if Y is simple. Now let $\underline{Y}_n = k2^{-n}$ on the set $\{k2^{-n} \leq Y < (k+1)2^{-n}\}$, $k = 0, 1, \ldots$ and set $\underline{Y}'_n = \min\{\underline{Y}_n, n\}$. Then $\underline{Y}'_n \leq Y$ and $\underline{Y}'_n$ increases to Y as $n \to \infty$. $\underline{Y}'_n$ is simple, so

$$\int_\Lambda \underline{Y}'_n E\{X \mid \mathcal{G}\}\, dP = \int_\Lambda X\underline{Y}'_n\, dP\,. \tag{10}$$

Let $n \to \infty$. Since $X \geq 0$, $E\{X|\mathcal{G}\} \geq 0$ by (v) and we can apply the Monotone Convergence Theorem to both sides of (10) to see that

$$\int_\Lambda Y E\{X \mid \mathcal{G}\}\, dP = \int_\Lambda XY\, dP\,. \tag{11}$$

Now in the general case, write $X^+ = \max\{X, 0\}$ and $X^- = X^+ - X = \max\{-X, 0\}$. Then both X^+ and X^- are positive, $X = X^+ - X^-$ and $|X| = X^+ + X^-$. X^+ is the *positive part of* X and X^- is the *negative part of* X. Define the positive and negative parts of Y similarly, and note that both are $\mathcal{G}$-measurable. If X is integrable, so is $|X|$, and hence so are X^+ and X^-. Moreover, $XY = X^+Y^+ + X^-Y^- - X^+Y^- - X^-Y^+$. Since (11) holds for all four products, it holds for XY. Thus $YE\{X \mid \mathcal{G}\}$ satisfies (CE1) and (CE2), proving $(viii)$.

(ix) Assume without loss of generality that (X_n) increases. (Otherwise consider the sequence $(-X_n)$.) Let $Z_n = E\{X_n \mid \mathcal{G}\}$. Then $Z_n \leq \{X \mid \mathcal{G}\}$ and by (vi), Z_n increases to a limit Z. By the Monotone Convergence Theorem **??**, $E\{Z\} \leq E\{E\{X \mid \mathcal{G}\}\} = E\{X\}$ by (i). X is integrable, so this is finite and Z must be integrable. Each Z_n is $\mathcal{G}$-measurable, hence so is Z. Moreover, again by the Monotone Convergence Theorem, for $\Lambda \in \mathcal{G}$

$$\int_\Lambda Z\, dP = \lim \int_\Lambda Z_n\, dP = \int_\Lambda Z\, dp\,.$$

Thus Z satisfies both (CE1) and (CE2), and therefore equals $E\{X \mid \mathcal{G}\}$.

♣

1.4 Changing the Sigma Field

What happens if we take successive conditional expectations with respect to different sigma fields? If one sigma field is contained in the other, the answer is that we end up with the conditional expectation with respect to the coarsest sigma field of the two.

Theorem 1.8 *If X is an integrable r.v. and if $\mathcal{G}_1 \subset \mathcal{G}_2$, then*

(12) $$E\{E\{X \mid \mathcal{G}_1\} \mid \mathcal{G}_2\} = E\{E\{X \mid \mathcal{G}_2\} \mid \mathcal{G}_1\} = E\{X \mid \mathcal{G}_1\}$$

PROOF. We will show that each of the first two terms of (12) equals the third. First, $E\{X \mid \mathcal{G}_1\}$ is $\mathcal{G}_1$-measurable, and therefore $\mathcal{G}_2$ measurable as well, since $\mathcal{G}_1 \subset \mathcal{G}_2$. By ($iii$), then, $E\{E\{X \mid \mathcal{G}_1\} \mid \mathcal{G}_2\} = E\{X \mid \mathcal{G}_1\}$.

Consider the second term, $E\{E\{X \mid \mathcal{G}_2\} \mid \mathcal{G}_1\}$. To show this equals $E\{X \mid \mathcal{G}_1\}$, note that it is $\mathcal{G}_1$-measurable, as required. If $\Lambda \in \mathcal{G}_1$, apply (CE2) to $E\{X \mid \mathcal{G}_2\}$ and X successively to see that

$$\int_\Lambda E\{E\{X \mid \mathcal{G}_2\} \mid \mathcal{G}_1\}\, dP = \int_\Lambda E\{X \mid \mathcal{G}_2\}\, dP = \int_\Lambda X\, dP\,.$$

Thus it satisfies both (CE1) and (CE2). This completes the proof. ♣.

Remark 1.9 *Note that (i) of Theorem 1.7 is a special case of Theorem 1.8. If we think of the conditional expectation as an average, this result is an instance of the principle that the average of sub-averages is the grand average.*

Corollary 1.10 *Let X be an integrable r.v. and let $\mathcal{G}_1 \subset \mathcal{G}_2$ be sigma fields. Then a necessary and sufficient condition that $E\{X \mid \mathcal{G}_2\} = E\{X \mid \mathcal{G}_1\}$ is that $E\{X \mid \mathcal{G}_2\}$ be $\mathcal{G}_1$-measurable.*

PROOF. Suppose $E\{X \mid \mathcal{G}_2\}$ is $\mathcal{G}_1$-measurable. By Theorem 1.7 (iii), $E\{X \mid \mathcal{G}_2\} = E\{E\{X \mid \mathcal{G}_2\} \mid \mathcal{G}_1\}$ and by Theorem 1.8 this equals $E\{X \mid \mathcal{G}_1\}$. The converse is clear. ♣

1.5 Jensen's Inequality

A function ϕ on $I\!R$ is *convex* if for any $a < b \in I\!R$ and $0 \le \lambda \le 1$, we have $\phi(\lambda a+(1-\lambda)b) \le \lambda\phi(a) + (1-\lambda)\phi(b)$. A well-known and useful inequality concerning these is the following, known as *Jensen's inequality.*

Theorem 1.11 *Let X be a r.v. and ϕ a convex function. If both X and $\phi(X)$ are integrable, then*

$$\phi(E\{X\}) \le E\{\phi(X)\}\,. \tag{13}$$

We will assume this, along with some of the elementary properties of convex functions. We want to extend it to conditional expectations. This turns out to be not quite trivial—the usual proof, elegant though it is, doesn't extend easily to conditional expectations.

Theorem 1.12 *Let X be a r.v. and let ϕ be a convex function on $I\!R$. Suppose both X and $\phi(X)$ are integrable, and $\mathcal{G} \subset \mathcal{F}$ is a sigma field. Then*

$$\phi\left(E\{X \mid \mathcal{G}\}\right) \le E\{\phi(X) \mid \mathcal{G}\} \ \textit{a.e.} \tag{14}$$

Proof. Case 1: suppose X is discrete, say $X = \sum_i x_i I_{\Gamma_i}$ where the Γ_i are disjoint and $\cup_i \Gamma_i = \Omega$.

$$\phi\left(E\{X \mid \mathcal{G}\}\right) = \phi\left(\sum_i x_i P\{\Gamma_i \mid \mathcal{G}\}\right).$$

For a.e. fixed ω, $\sum_i P\{\Gamma_i \mid \mathcal{G}\} = P\{\cup_i \Gamma_i \mid \mathcal{G}\} = 1$ a.e. (see the exercises.) Thus by Theorem 1.11, this is

$$\le \sum_i \phi(x_i) P\{\Gamma_i \mid \mathcal{G}\} = E\{\phi(X) \mid \mathcal{G}\}\,.$$

In the general case, there exists a sequence of integrable X_n such that $X_n \uparrow X$ and $|X - X_n| \le 2^{-n}$. Then $E\{X_n \mid \mathcal{G}\} \uparrow E\{X \mid \mathcal{G}\}$ a.e. by the Monotone Convergence Theorem for conditional expectations, 1.7 (ix). ϕ is convex, therefore continuous, so $\phi(E\{X_n \mid \mathcal{G}\}) \to \phi(E\{X \mid \mathcal{G}\})$. Now either ϕ is monotone or else it has a minimum a, and ϕ is decreasing on $(-\infty, a]$ and increasing on $[a, \infty)$. If ϕ is monotone, we can apply the monotone convergence to $\phi(X_n)$ to finish the proof.

If not, for a positive integer m, and $n > m$ write

$$\phi(X_n) = I_{\{X < a - 2^{-m}\}}\,\phi(X_n) + I_{\{a-2^{-m} \le X < a+2^{-m}\}}\,\phi(X_n) + I_{\{X > a + 2^{-m}\}}\,\phi(X_n)$$

so

$$\begin{aligned} E\{\phi(X_n) \mid \mathcal{G}\} &= E\{I_{\{X<a-2^{-m}\}}\,\phi(X_n) \mid \mathcal{G}\} + E\{I_{\{a-2^{-m} \le X \le a+2^{-m}\}}\,\phi(X_n) \mid \mathcal{G}\} \\ &\quad + E\{I_{\{X > a+2^{-m}\}}\,\phi(X_n) \mid \mathcal{G}\}\,. \end{aligned}$$

In the first expectation, $X_n < X \leq a$, and ϕ decreases on $(-\infty, a]$, so that $\phi(X_n)$ decreases to $\phi(X)$, and we can go to the limit by Theorem 1.7 (*ix*). In the third expectation, $a \leq X_n$; since ϕ is increasing there, $\phi(X_n)$ increases to $\phi(X)$ and we can again go to the limit. In the middle expectation, $\phi(X)$ and $\phi(X_n)$ are caught between $\phi(a)$ (the minimum value of ϕ) and $\phi(a) + M_m$, where $M_m = \max\{\phi(a - 2^{-m}), \phi(a + 2^{-m})\} - \phi(a)$. Thus $E\{I_{\{a-2^{-m} \leq X_n \leq a+2^{-m}\}}\, \phi(X_n) \mid \mathcal{G}\}$ and $E\{I_{\{a-2^{-m} \leq X \leq a+2^{-m}\}}\, \phi(X_n) \mid \mathcal{G}\}$ are both caught between $\phi(a)P\{a - 2^{-m} \leq X \leq a + 2^{-m}\}$ and $(\phi(a) + M_n)P\{a - 2^{-m} \leq X \leq a + 2^{-m}\}$. Since ϕ is continuous, these can be made as close together as we wish by making m large enough. This proves convergence. ♣

Remark 1.13 *Suppose X is square-integrable. Then $E\{X \mid \mathcal{G}\}$ is also square-integrable by Jensen's inequality. Here is another interesting characterizaton of $E\{X \mid \mathcal{G}\}$: it is the best mean-square approximation of X among* all *$\mathcal{G}$-measurable random variables.*

Indeed, suppose Y is square-integrable and $\mathcal{G}$-measurable. Then

$$\begin{aligned} E\{(Y - X)^2\} &= E\left\{(Y - E\{X \mid \mathcal{G}\} + (E\{X \mid \mathcal{G}\} - X)^2\right\} \\ &= E\left\{(Y - E\{X \mid \mathcal{G}\})^2\right\} + \left(E\left\{E\{X \mid \mathcal{G}\} - X\right)^2\right\} \\ &\qquad +2E\{(Y - E\{X \mid \mathcal{G}\})(E\{X \mid \mathcal{G}\} - X)\}\,. \end{aligned}$$

Now $Y - E\{X \mid \mathcal{G}\}$ is $\mathcal{G}$-measurable, so that by Theorem 1.7 *(iii) and (i)*

$$E\{(Y - E\{X \mid \mathcal{G}\})\,(E\{X \mid \mathcal{G}\} - X)\} = (Y - E\{X \mid \mathcal{G}\})\,E\{E\{X \mid \mathcal{G}\} - X\} = 0\,,$$

so the cross terms above drop out. Thus we can minimize $E\{(Y - X)^2\}$ by setting $Y = E\{X \mid \mathcal{G}\}$.

Remark 1.14 *In the language of Hilbert spaces, $E\{X \mid \mathcal{G}\}$ is the* projection *of X on the space of $\mathcal{G}$-measurable random variables. If we had already studied Hilbert spaces, we could settle the problem of the existence of the general conditional expectation right now: projections exist—this is a consequence of the Riesz-Fischer Theorem—and the existence of general conditional expectations follows easily. However, we will wait a bit, and show this existence later as a result of some martingale convergence theorems.*

1.6 Independence

We will quickly translate some well known facts about independence to our setting. Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition 1.2 *(i) Two sigma fields $\mathcal{G}$ and $\mathcal{H}$ are* **independent** *if for $\Lambda \in \mathcal{G}$ and $\Gamma \in \mathcal{H}$,*

$$P\{\Lambda \cap \Gamma\} = P\{\Lambda\}P\{\Gamma\}\,.$$

(ii) A finite family $\mathcal{G}_1, \ldots, \mathcal{G}_n$ of sigma fields is **independent** *if for any $\Lambda_i \in \mathcal{G}_i,\ i = 1, \ldots, n$,*

$$P\{\cap_i \Lambda_i\} = \prod_i P\{\Lambda_i\}.$$

(iii) An infinite family $\{\mathcal{G}_\alpha,\ \alpha \in I\}$ is **independent** *if any finite sub-family is.*

We define independence of random variables in terms of this.

Definition 1.3 *(i) A finite family $X_1, \ldots, X_n$ of r.v. is* **independent** *if the sigma fields $\mathcal{F}_{X_1}, \ldots, \mathcal{F}_{X_n}$ are independent.*

(ii) an infinite family $\{X_\alpha, \alpha \in I\}$ of r.v. is **independent** *if any finite subfamily is.*

Proposition 1.15 *Let X be a r.v. and $\mathcal{G}$ a sigma field. Suppose X and $\mathcal{G}$ are independent (i.e. $\mathcal{F}_X$ and $\mathcal{G}$ are independent.) Then $E\{X \mid \mathcal{G}\} = E\{X\}$ a.e.*

1.7 Exercises

2° Let S be an integrable random variable and B a set of positive probability. Show that X will be integrable with respect to the conditional probability measure $P\{A \mid B\}$.

2° Let Λ_i be a countable partition of Ω and $\mathcal{G}$ a sigma field. Show that with probability one, $\sum_i P\{\Lambda_i \mid \mathcal{G}\} = 1$.

3° Show that the definition of independence of given above is equivalent to the usual definition: $P\{X_1 \in B_1, \ldots, X_n \in B_n\} = \Pi_{i=1}^n P\{X_i \in B_n\}$ for Borel sets $B_1, \ldots, B_n$.

4° Prove Proposition 1.15.

2 Martingales

A martingale is a mathematical model for a fair wager. It takes its name from "la grande martingale," the strategy for even-odds bets in which one doubles the bet after each loss. (If you double the bet with each loss, the first win will recoup all previous losses, with a slight profit left over. Since one is bound to win a bet sooner or later, this system guarantees a profit. Its drawbacks have been thoroughly explored.) The term also refers to the back belt of a dress or jacket, a part of a sailing ship's rigging, and a tie-down used on horses' bridles, which might explain some bizarre birthday gifts to prominent probabilists.

What do we mean by "fair"? One reasonable criterion is that on the average, the wagerer should come out even. In mathematical language, the expected winnings should be zero. But we must go a little deeper than that. It is not enough to have the total expectation zero—the expectation should be zero *at the time of the bet*. This is not an arbitrary distinction. Consider, for example, the three-coin game: we are given a fair coin, an unbalanced coin which comes up heads with probability $2/3$, and another unbalanced coin which comes up tails with probability $2/3$. We first toss the fair coin. If it comes up heads, we toss the coin with $P\{\text{heads}\} = 2/3$. If it comes up tails, we toss the other coin. We bet on the final outcome—heads or tails—at even odds. In this case, it is important to know exactly when we bet. If we bet before the first coin is tossed, by symmetry we will have a fifty-fifty chance of winning our bet, and we will gain a dollar if we guess correctly, and lose it if we are wrong. Thus the wager is fair. But if we bet *after* we see the result of the first toss, we have a much better chance of winning our bet. The one-for-one payoff is no longer fair, but is greatly in our favor. Thus, it is really the *conditional* expectation—given all our knowledge at the time of the bet—which should be zero.

We will model a succession of bets, keeping track of the total amount of money we have—our fortune—at each bet. Let X_0 be our initial fortune, X_1 our fortune after the first bet, X_2 our fortune after the second, and so on. At the time we place each bet, we will know certain things, including our fortune at the time, but we will *not* know the result of the bet. Our winnings at the nth bet are $X_n - X_{n-1}$. The requirement of fairness says that that the conditional expectation of $X_n - X_{n-1}$ *given our knowledge at the time we bet*, is zero. As we have seen, we can represent knowledge by sigma fields. So let us represent the knowledge we have at the time we make the nth bet by a sigma field $\mathcal{F}_n$. Then we should have $E\{X_n - X_{n-1} \mid \mathcal{F}_{n-1}\} = 0$, or $E\{X_n \mid \mathcal{F}_{n-1}\} = X_n$. This leads us to the following definitions.

Definition 2.1 *A* **filtration** *on the probability space* $(\Omega, \mathcal{F}, P)$ *is a sequence* $\{\mathcal{F}_n : n = 0, 1, 2, \ldots\}$ *of sub-sigma fields of* $\mathcal{F}$ *such that for all* n, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

The filtration represents our knowledge at the successive betting times. This increases with time—in this model we don't forget things—so that the sigma fields increase.

Definition 2.2 *A* **stochastic process** *is a collection of random variables defined on the same probability space.*

That is a fairly general definition—it is almost hard to think of something numerical which is *not* a stochastic process. However, we have something more specific in mind.

Definition 2.3 *A stochastic process* $X = \{X_n,\ n = 0, 1, 2, \ldots\}$*, is* **adapted** *to the filtration* $(\mathcal{F}_n)$ *if for all* n*,* X_n *is* $\mathcal{F}_n$*-measurable.*

Definition 2.4 *A process* $X = \{X_n, \mathcal{F}_n,\ n = 0, 1, 2, \ldots\}$*, is a* **martingale** *if for each* $n = 0, 1, 2, \ldots$,

(i) $\{\mathcal{F}_n,\ n = 0, 1, 2 \ldots\}$ *is a filtration and* X *is adapted to* $(\mathcal{F}_n)$*;*
(ii) for each n*,* X_n *is integrable;*
(iii) for each n*,* $E\{X_{n+1} \mid \mathcal{F}_n\} = X_n$.

At the same time, we can define the related notions of submartingales and supermartingales.

Definition 2.5 *A process* $X = \{X_n, \mathcal{F}_n,\ n = 0, 1, 2, \ldots\}$*, is a* **submartingale** *(resp.* **supermartingale***) if for each* $n = 0, 1, 2, \ldots$,

(i) $\{\mathcal{F}_n,\ n = 0, 1, 2 \ldots\}$ *is a filtration and* X *is adapted to* $(\mathcal{F}_n)$*;*
(ii) for each n*,* X_n *is integrable;*
(iii) for each n*,* $X_n \leq E\{X_{n+1} \mid \mathcal{F}_n\}$ *(resp.* $X_n \geq E\{X_{n+1} \mid \mathcal{F}_n\}$ *.)*

Remark 2.1 *Think of the parameter* n *as time, and* $\mathcal{F}_n$ *as the history of the world up to time* n*. If a r.v. is measurable with respect to* $\mathcal{F}_n$*, it depends only on the past before* n*. If a process* (X_n) *is adapted to* $(\mathcal{F}_n)$*, then each* X_n *depends only on what has already happened before time* n*. One sometimes calls such processes "non-anticipating" because, quite simply, they can't look into the future. This is, needless to say, a rather practical hypothesis, one which is satisfied by all living beings, with the possible exception of a few (doubtless rich) seers.*

Remark 2.2 *We can define the notion of a martingale relative to any subset of the real line. If* $I \subset I\!R$ *is any set, and if* $\{\mathcal{F}_t,\ t \in I\}$ *is a filtration* $(s, t \in I$ *and* $s \leq t \Longrightarrow \mathcal{F}_s \subset \mathcal{F}_t)$*, then a stochastic process* $\{X_t,\ t \in I\}$ *is a* submartingale *if* $s \leq t \Longrightarrow X_s \leq E\{X_t \mid \mathcal{F}_s\}$ *a.e. However, for the minute we will restrict ourselves to discrete parameter sets. Continuous parameter martingales, though useful and interesting, will have to await their turn.*

We could consider an arbitrary discrete parameter set $t_1 < t_2 < \ldots$*, but the main property of a discrete parameter set is simply its order, so without any loss of generality, we may map* t_n *to* n*, and take it to be a subset of the integers. If it is finite, we take it to be* $0, 1, 2, \ldots, n$*. If it is infinite, we can take either* $I\!N^+ \equiv 0, 1, 2 \ldots$*,* $I\!N^- \equiv \ldots - 2, -1$*, or* $I\!N \equiv \ldots 0, \pm 1, \pm 2, \ldots$*. Let us start by considering processes of the form* $\{X_n,\ n = 0, 1, 2, \ldots\}$*, with the understanding that this includes the case where the parameter set is finite.*

If a martingale represents a fair game, then a submartingale represents a favorable game (the expected winnings are positive) and a supermartingale represents an unfavorable game. As a practical matter, the patron of a gambling casino plays games which are supermartingales (with the exception of blackjack where there is a known winning strategy) while the casino plays submartingales. In the business world, financiers go to great length to make sure they are playing submartingales, not supermartingales.

First, here are some elementary properties.

Proposition 2.3 *(i) A stochastic process* $X \equiv \{X_n, \mathcal{F}_n,\ n = 0, 1, 2 \ldots\}$ *is a submartingale if and only if* $-X$ *is a supermartingale. It is a martingale if and only if it is both a sub- and supermartingale.*

(ii) Suppose (X_n) *is a submartingale relative to the filtration* $(\mathcal{F}_n)$*. Then for each* $m < n$*,* $X_m \le E\{X_n \mid \mathcal{F}_m\}$*.*

(iii) If X *is a martingale,* $E\{X_n\} = E\{X_0\}$ *for all* n*. If* $m < n$ *and if* X *is a submartingale, then* $E\{X_m\} \le E\{X_n\}$*; if* X *is a supermartingale, then* $E\{X_m\} \ge E\{X_n\}$*.*

(iv) If (X_n) *is a submartingale relative to some filtration* $(\mathcal{F}_n)$*, then it is also a submartingale with respect to its* natural filtration $\mathcal{G}_n \equiv \sigma\{X_0, \ldots, X_n\}$.

Remark 2.4 *Because of (i), we will state most of our results for submartingales. The corresponding results for martingales and submartingales follow immediately.*

By (iii), martingales have constant expectations, while the expectations of submartingales increase with time, and the expectations of supermartingales decrease. (iv) shows us that if necessary, we can always use the natural filtration of the processes. However, it is useful to have the flexibility to choose larger filtrations.

PROOF. (i) is clear. Note (ii) is true for $n = m + 1$ from the definition of submartingale. Suppose it is true for $n = m + k$. Then $X_m \le E\{X_{m+k} \mid \mathcal{F}_m\}$. But by the submartingale inequality, $X_{m+k} \le E\{X_{m+k+1} \mid \mathcal{F}_{m+k}\}$, so that $X_m \le E\{E\{X_{m+k+1} \mid \mathcal{F}_{m+k}\} \mid \mathcal{F}_m\}$. Since $\mathcal{F}_m \subset \mathcal{F}_{m+k}$, Theorem 1.8 implies that this equals $E\{X_{m+k+1} \mid \mathcal{F}_m\}$, hence (ii) is true for $n = k + 1$, hence for all $n > m$ by induction.

(iii) If X_n is a submartingale, $X_n \le E\{X_{n+1} \mid \mathcal{F}_n\}$. Take the expectation of both sides and note that the expectation of the right-hand side is $E\{X_{n+1}\}$, so $E\{\, X_n \,\} \le E\{\, X_{n+1} \,\}$ for all n.

(iv) Note that $\mathcal{G}_n \subset \mathcal{F}_n$ (why?) so X is adapted to its natural filtrations, and if $m < n$, $X_m \le E\{X_n \mid F_n\}$ so

$$E\{X_n \mid \mathcal{G}_m\} = E\{E\{X_n \mid \mathcal{F}_m\} \mid \mathcal{G}_m\} \ge E\{X_m \mid \mathcal{G}_m\} = X_m\,.$$

♣

Proposition 2.5 *(i) Suppose* $\{X_n, \mathcal{F}_n,\ n = 0, 1, 2, \dots\}$ *is a martingale and* ϕ *is a convex function on* $\mathbb{R}$*. Then, if* $\phi(X_n)$ *is integrable for all* n*,* $\{\phi(X_n), \mathcal{F}_n,\ n = 0, 1, 2, \dots\}$ *is a submartingale.*

(ii) Suppose that $\{X_n, \mathcal{F}_n,\ n = 0, 1, 2, \dots\}$ *is a submartingale and* ϕ *is an increasing convex function on* $\mathbb{R}$*. Then, if* $\phi(X_n)$ *is integrable for all* n*,* $\{\phi(X_n), \mathcal{F}_n\ n = 0, 1, 2, \dots\}$ *is a submartingale.*

PROOF. (i) By Jensen's inequality for conditional expectations,

$$\phi(X_m) = \phi(E\{X_n \mid \mathcal{F}_m\}) \leq E\{\phi(X_n) \mid \mathcal{F}_m\}\,.$$

The proof for (ii) is almost the same: $X_m \leq E\{X_m \mid \mathcal{F}_m\}$ so $\phi(X_m) \leq \phi(E\{X_n \mid \mathcal{F}_m\}) \leq E\{(\phi(X_n) \mid \mathcal{F}_m\}$, where the last inequality follows by Jensen. ♣

Thus if X_n is a martingale, then (subject to integrability) $|X_n|$, X_n^2, e^{X_n}, and e^{-X_n} are all submartingales, while if $X_n > 0$, $\sqrt{X_n}$ and $\log(X_n)$ are supermartingales. If X_n is a submartingale and K a constant, then $\max\{X_n, K\}$ is a submartingale, while if X_n is a supermartingale, so is $\min\{X_n, K\}$. (The last two follow because $x \mapsto \max\{x, K\}$ is convex and increasing.)

2.1 Examples

1° Let $x_0, x_1, x_2, \dots$ be real numbers. If the sequence is increasing, it is a submartingale, if it is constant, it is a martingale, and if it is decreasing, it is a supermartingale.

Indeed, we can define constant random variables $X_n \equiv x_n$ on any probability space whatsoever, and take $\mathcal{F}_n$ to be the trivial filtration $\mathcal{F}_n = \{\phi, \Omega\}$ for all n to make this into a stochastic process. Trivial tho it may be, this example is sometimes useful to curb over-enthusiastic conjectures.

2° For a more interesting example, let $Y_1, Y_2, \dots$ be a sequence of independent random variables with $E\{Y_n\} \geq 0$ for all n. Set $X_0 = 0$, $X_n = \sum_{i=1}^n Y_i$ for $n \geq 1$. Let $\mathcal{F}_n = \sigma\{Y_1, \dots, Y_n\}$ be the sigma field generated by the first n Y_i's. Then $\{X_n,\ n \geq 0\}$ is a submartingale relative to the filtration $(\mathcal{F}_n)$. If the Y_n all have expectation zero, it is a martingale.

Indeed, $E\{X_{n+1} \mid \mathcal{F}_n\} = E\{X_n + Y_{n+1} \mid \mathcal{F}_n\}$. Now X_n is easily seen to be $\mathcal{F}_n$-measurable, so this equals $X_n + E\{Y_{n+1} \mid \mathcal{F}_n\}$. But Y_{n+1} is independent of $Y_1, \dots, Y_n$, and hence of $\mathcal{F}_n$, so this equals $X_n + E\{Y_{n+1}\} \geq X_n$. Thus X is a submartingale.

3° Example 2° could describe a sequence of gambles: the gambler bets one dollar each time and Y_j is the gain or loss at the jth play. We would say the game is fair if $E\{Y_j\} = 0$ for all j, that it is favorable if $E\{Y_j\} \geq 0$ for all j, and that it is unfair (or realistic) if $E\{Y_j\} \leq 0$. A more worldly model would allow the gambler to vary the bets. Suppose at the time of the nth bet the gambler decides to wager a stake of K_n dollars; K_n must be positive, and it may well be random—the gambler might vary the bets according to the situation (this is the principal strategy in the system for blackjack) or on a random whim. At the time of the wager, the bettor does not know the result Y_n of the nth bet, but knows all the previous results $Y_1, \dots, Y_{n-1}$, so that K should be measurable with respect to $\mathcal{F}_{n-1}$. Call the new process $\hat{X}$. Thus the gambler's fortune after the nth wager is $\hat{X}_n = \hat{X}_{n-1} + K_n Y_n$, where K_n is an $\mathcal{F}_{n-1}$ measurable positive random variable. (We will also assume it is bounded to preserve integrability.) Then $E\{\hat{X}_n \mid \mathcal{F}_{n-1}\} = \hat{X}_{n-1} + K_n E\{Y_n \mid \mathcal{F}_{n-1}\} = \hat{X}_{n-1} + K_n E\{Y_n\}$, since both $\hat{X}_{n-1}$ and K_n are $\mathcal{F}_{n-1}$-measurable. Since K_n is positive, $\hat{X}$ will be a martingale (resp. submartingale, supermartingale) if the fixed-bet process X is a martingale (resp. submartingale, supermartingale)

This tells us that we can't change the basic nature of the game by changing the size of the bets. (And, incidentally, it shows the superiority of the stock market to casinos. In casinos, one can't bet negative dollars. In the stock market one can, by selling short.)

4° Here is an important example. Let $(\mathcal{F}_n)$ be a filtration and let X be any integrable random variable. Define $X_n \equiv E\{X \mid \mathcal{F}_n\}$. Then $\{X_n,\ n = 0, 1, \dots\}$ is a martingale relative to the filtration $(\mathcal{F}_n)$.

Indeed, $E\{X_{n+1} \mid \mathcal{F}_n\} = E\{E\{X \mid \mathcal{F}_{n+1}\} \mid \mathcal{F}_n\} \equiv E\{X \mid \mathcal{F}_n\} = X_n$ by Theorem 1.8.

5° Let $Y_1, Y_2, \dots$ be a sequence of i.i.d. r.v. with $P\{Y_j = 1\} = p$, and $P\{Y_j = 0\} = 1 - p$. Then

$$X_n = \prod_{j=1}^{n} \frac{Y_j}{p}$$

is a martingale. Indeed,

$$E\{X_{n+1} \mid Y_1, \dots Y_n\} = E\{X_n Y_{n+1}/p \mid Y_1, \dots, Y_n\} = X_n E\{Y_{n+1}/p\} = X_n \,.$$

This is an example of a positive martingale with common expectation one, which converges to zero with probability one. (Since $p < 1$, one of the Y_n will eventually vanish, and $X_n = 0$ from then on.)

6° There are numerous examples of martingales in statistics. Here is one which arises in sequential analysis.

Let $X_1, X_2, \dots$ be a sequence of random variables, not necessarily independent. Assume that there are two candidates for their joint densities, which we call H1 and H2. The statistical problem is to observe some of the random variables, and then decide whether H1 or H2 is correct.

(H1) The joint density of the first n random variables $X_1, \dots, X_n$ is $q_n(x_1, \dots, x_n)$ for $n = 1, 2, \dots$

(H2) The joint density of the first n random variables $X_1, \dots, X_n$ is $p_n(x_1, \dots, x_n)$ for $n = 1, 2, \dots$

In either case, we have a sequence of joint densities; in (H1) they are given by $q_1(x_1), q_2(x_1, x_2), \ldots$ and in (H2) they are a different sequence, $p_1(x_1), p_2(x_1, x_2), \ldots$. Let $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$ be the natural sigma fields of the X_j, and define

$$Z_n = \begin{cases} \dfrac{q_n(X_1, \ldots, X_n)}{p_n(X_1, \ldots, X_n)} & \text{if } p_n(X_1, \ldots, X_n) \neq 0 \\ 0 & \text{if } p_n(X_1, \ldots, X_n) = 0\,. \end{cases} \tag{15}$$

Then if (H2) is actually the correct series of densities, $\{Z_n, \mathcal{F}_n,\ n = 1, 2, \ldots\}$ is a positive super-martingale, and it is a martingale if $p_n(x_1, \ldots, x_n)$ never vanishes.

To see this, note that Z_n is certainly $\mathcal{F}_n$-measurable, and let $\Lambda \in \mathcal{F}_n$, so that $\Lambda = \{(X_1, \ldots, X_n) \in B\}$ for some Borel set B in $\mathbb{R}^n$. We are supposing that the the p_n are the correct densities, so

$$\begin{aligned}
\int_\Lambda Z_{n+1}\, dP &= \int_{B\times\mathbb{R}} \cdots \int \frac{q_{n+1}(x_1, \ldots, x_{n+1})}{p_{n+1}(x_1, \ldots, x_{n+1})} I_{\{p_{n+1}(x_1,\ldots,x_{n+1})\neq 0\}}\, dx_1 \ldots dx_{n+1} \\
&= \int_{B\times\mathbb{R}} \cdots \int q_{n+1}(x_1, \ldots, x_{n+1}) I_{\{p_{n+1}(x_1,\ldots,x_{n+1})\neq 0\}}\, dx_1 \ldots dx_{n+1} \\
&\leq \int_{B\times\mathbb{R}} \cdots \int q_{n+1}(x_1, \ldots, x_{n+1}) I_{\{p_n(x_1,\ldots,x_n)\neq 0\}}\, dx_1 \ldots dx_{n+1}
\end{aligned}$$

since $p_n(x_1, \ldots, x_n) = 0 \Longrightarrow p_{n+1}(x_1, \ldots, x_n, \cdot) = 0$ a.e., so $I_{\{p_{n+1}>0\}} \leq I_{\{p_n>0\}}$. Now integrate out x_{n+1} and use the properties of joint densities to see that this is

$$\begin{aligned}
&= \int_B \cdots \int q_n(x_1, \ldots x_n) I_{\{p_n(x_1,\ldots,x_n)\neq 0\}}\, dx_1 \ldots dx_n \\
&= \int_B \cdots \int \frac{q_n(x_1, \ldots, x_n)}{p_n(x_1, \ldots, x_n)} p_n(x_1, \ldots, x_n) I_{\{p_n(x_1,\ldots,x_n)\neq 0\}}\, dx_1 \ldots dx_n \\
&= \int_\Lambda X_n\, dP\,.
\end{aligned}$$

There is equality if the p_n never vanish.

2.2 Exercises

1° Find the probabilities for the three-coin game which make it fair if we bet after the first coin is tossed.

2° Show that the maximum of two submartingales (relative to the same filtration) is a submartingale.

3° Let (Y_n) be a sequence of positive independent r.v. with $E\{Y_j\} = 1$ for all j. Set $X_0 = 1$, and $X_n = \prod_1^n Y_j$ for $n \geq 1$. Show that $X_0, X_1, X_2, \ldots$ is a martingale relative to its natural filtration.

2.3 The Doob Decomposition

Submartingales turn out to satisfy many of the same theorems as martingales. This is initially surprising, since a martingale is a rather special case of a submartingale, but the following result indicates why this might be true. In fact, any submartingale is the sum of a martingale plus an increasing process.

Theorem 2.6 *(Doob Decomposition) Let $X = \{X_n, n \geq 0\}$ be a submartingale relative to the filtration $(\mathcal{F}_n)$. Then there exists a martingale $M = \{M_n,\ n \geq 0\}$ and a process $A = \{A_n,\ n \geq 0\}$ such that*

(i) M is a martingale relative to $(\mathcal{F}_n)$;

(ii) A is an increasing process: $A_n \leq A_{n+1}$ a.e.;

(iii) A_n is $\mathcal{F}_{n-1}$-measurable $\forall\ n$;

(iv) $X_n = M_n + A_n$.

Proof. Let $d_n \equiv E\{X_{n+1} - X_n \mid \mathcal{F}_n\}$. By the submartingale inequality, d_n is positive and $\mathcal{F}_n$-measurable. Set $A_0 = 0$, $A_n = \sum_{j=1}^{n-1} d_j$ and $M_n = X_n - A_n$. Then (ii), (iii), and (iv) hold. To see (i), write

$$\begin{aligned} E\{M_{n+1} \mid \mathcal{F}_n\} &= E\{X_{n+1} - A_{n+1} \mid \mathcal{F}_n\} \\ &= E\{X_{n+1} \mid \mathcal{F}_n\} - A_{n+1} \\ &= X_n + d_n - \sum_{j=1}^{n} d_j \\ &= X_n - \sum_{j=1}^{n-1} d_j \\ &= M_n\,. \end{aligned}$$

This finishes the proof. ♣

2.4 Stopping Times

Consider a random instant of time, such as the first time heads comes up in a series of coin tosses, the first time the Dow-Jones takes a one-day fall of more than five hundred points, or the moment we stopped believing in Santa Claus. There are two fundamental types of random times: the first class consists of those we can determine in “real time”—that is, times which we can recognize when they arrive—times which can be determined without reference to the future. The second class consists of those we can’t. The three examples above are in the first class, but something like the time when the Dow Jones reaches its maximum in the month of February is not: in general, we must wait until the end of the month to determine the maximum, and by then the time has passed. The first class of times can be used for strategies of investing and other forms of gambling. We can recognize them when they arrive, and make decisions based on them.

We will give a general characterization of such times. Let I be a subset of the integers and let $\{\mathcal{F}_n,\ n \in I\}$ be a filtration.

Definition 2.6 *A random variable T taking values in $I \cup \{\infty\}$ is a* **stopping time** *(or* **optional time***) if for each $n \in I$, $\{\omega : T(\omega) = n\} \in \mathcal{F}_n$.*

Definition 2.7 *Let T be a stopping time. The sigma field $\mathcal{F}_T$, sometimes called "the past before T" is*

$$\mathcal{F}_T = \{\Lambda \in \mathcal{F} : \Lambda \cap \{T = n\} \in \mathcal{F}_n, \quad \forall n\}.$$

Remark 2.7 *$\mathcal{F}_n$ represents the information available at time n, so $\{T = n\} \in \mathcal{F}_n,\ \forall n$ is exactly the condition which says we can recognize T when it arrives.*

T may take on the value ∞. This is interpreted to mean that "T never happens." This is all too often the case with times such as "the time I broke the bank at Monte Carlo."

Remark 2.8 *The the definition of $\mathcal{F}_T$ may require reflection to understand. If the $\mathcal{F}_n$ are the natural sigma fields, then $\mathcal{F}_T$ should just be $\sigma\{X_1, \dots X_T\}$. The problem with this is that the length of the sequence $X_1, \dots X_T$ is random—that is, the sequence itself can contain differing numbers of elements, so we must be careful to say exactly what we mean by $\sigma\{X_1, \dots X_T\}$. To handle it, we have to consider the sets on which T takes a specific value. An example may help to clarify this. Suppose T is the first time that there is a multicar chain collision on a new interstate highway. One event of interest would be whether there was a fog which caused the collision. Let Λ be the event "there was fog just before the collision." This event is certainly in the past of T. To determine whether Λ happened, one could go to the newspaper every day, and see if there is an article about such an accident—if it is the newspaper of day n, this tells us if $T = n$. If the accident did happen, read further in the article to find if there was fog—this is the event $\{T = n\} \cap \Lambda$. For "newspaper of day n," read $\mathcal{F}_n$.*

It is easy to see that a constant time is a stopping time. A more interesting example is the first hitting time:

Proposition 2.9 *Let $\{X_n,\ n = 0, 1, 2, \dots\}$ be a stochastic process adapted to the filtration $(\mathcal{F}_n)$. Let $B \in \mathbb{R}$ be a Borel set and define*

$$T_B(\omega) = \begin{cases} \inf\{n \geq 0 : X_n(\omega) \in B\} & \\ \infty & \text{if there is no such } n. \end{cases}$$

Then T_B is a stopping time, called the first hitting time of B.

Proof. $\{T_B = 0\}\{X_0 \in B\} \in \mathcal{F}_0$ and for $n \geq 1$, $\{T_B = n\} = \{X_0 \in B^C, \dots X_{n-1} \in B^c, X_n \in B\}$. Since $X_0, \dots, X_n$ are all $\mathcal{F}_n$-measurable (why?) this set is in $\mathcal{F}_n$, showing that T_B is a stopping time. ♣

One can extend this: if S is a stopping time, then $T = \inf\{n \geq S : X_n \in B\}$ is also a stopping time, which shows that second hitting times, etc. are stopping times.

Here are some elementary properties of stopping times.

Proposition 2.10 *(i) If T is a stopping time, the sets $\{T < n\}$, $\{T \leq n\}$, $\{T = n\}$ $\{T \geq n\}$ and $\{T > n\}$ are in $\mathcal{F}_n$.*

(ii) A random variable T with values in $I \cup \{\infty\}$ is a stopping time if and only if $\{T \leq n\} \in \mathcal{F}_n$ for all $n \in I$.

(iii) A constant random variable T taking its value in $I \cup \{\infty\}$ is a stopping time.

(iv) If T_1 and T_2 are stopping times, so are $T_1 \wedge T_2$ and $T_1 \vee T_2$.

Proof. (i) $\{T \leq n\} = \cup_{\{j \in I, j \leq n\}}\{T = j\}$. But if $j \leq n$, $\{T = j\} \in \mathcal{F}_j \subset \mathcal{F}_n$, so the union is in $\mathcal{F}_n$. $\{T < n\} = \{T \leq n\} - \{T = n\}$, which is also in $\mathcal{F}_n$, and the remaining two sets are complements of these two.

We leave (ii) as an exercise. To see (iii), note that $\{T = n\}$ is either the empty set or the whole space, and both are in $\mathcal{F}_n$.

(iv) $\{T_1 \wedge T_2 = n\} = (\{T_1 = n\} \cap \{T_2 \geq n\}) \cup (\{T_2 = n\} \cap \{T_1 \geq n\}) \in \mathcal{F}_n$ by (i). Similarly, $\{T_1 \vee T_2 = n\} = (\{T_1 = n\} \cap \{T_2 \geq n\}) \cup (\{T_2 = n\} \cap \{T_1 \leq n\}) \in \mathcal{F}_n$ by (i). ♣

Proposition 2.11 *(i) Let T be a stopping time. Then $\mathcal{F}_T$ is a sigma field, and T is $\mathcal{F}_T$-measurable.*

(ii) If $T_1 \leq T_2$ are stopping times, then $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$, and $\{T_1 = T_2\} \in \mathcal{F}_{T_1}$.

(iii) If $T_1, T_2, T_3, \dots$ are stopping times and if $T \equiv \lim_n T_n$ exists, then T is a stopping time.

Proof. (i) To show $\mathcal{F}_T$ is a sigma field, we verify the properties of the definition. Clearly $\phi \in \mathcal{F}_T$. Suppose $\Lambda \in \mathcal{F}_T$. Then $\Lambda^c \cap \{T = n\} = \{T = n\} - \Lambda \cap \{T = n\}$. But $\{T = n\} \in \mathcal{F}_n$ by the definition of a stopping time, and $\Lambda \cap \{T = n\} \in \mathcal{F}_n$ since $\Lambda \in \mathcal{F}_T$. Thus their difference is in $\mathcal{F}_n$, so $\Lambda^c \in \mathcal{F}_T$. If Λ_j, $j = 1, 2, \dots$ is a sequence of sets in $\mathcal{F}_T$, then $(\cup_j \Lambda_j) \cap \{T = n\} = \cup_j(\Lambda_j \cap \{T = n\}) \in \mathcal{F}_n$, so $\cup_j \Lambda_j \in \mathcal{F}_T$. Thus $\mathcal{F}_T$ is a sigma field.

To see T is $\mathcal{F}_T$-measurable, it is enough to show that $\{T = j\} \in \mathcal{F}_T$ for each j. But $\{T = j\} \cap \{T = n\}$ is either empty or equal to $\{T = n\}$, and in either case is in $\mathcal{F}_n$.

To see (ii), let Λ in $\mathcal{F}_{T_1}$. We claim $\Lambda \in \mathcal{F}_{T_2}$. Indeed, $\Lambda \cap \{T_2 = n\} = \cup_{j \leq n} \Lambda \cap \{T_1 = j\} \cap \{T_2 = n\}$. But $\Lambda \cap \{T_1 = j\} \in \mathcal{F}_j \subset \mathcal{F}_n$ and $\{T_2 = n\} \in \mathcal{F}_n$, hence $\Lambda \cap \{T_2 = n\} \in \mathcal{F}_n \Longrightarrow \Lambda \in \mathcal{F}_{T_2} \Longrightarrow \mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$.

Consider $\{T_1 = T_2\} \cap \{T_1 = n\} = \{T_1 = n\} \cap \{T_2 = n\}$. This is in $\mathcal{F}_n$ since both sets in the intersection are, proving (ii).

(iii) Because the T_n take values in a discrete parameter set, $\lim_n T_n(\omega) = T(\omega) \Longrightarrow T_n(\omega) = T(\omega)$ for all large enough n. Thus, $\{T = n\} = \liminf_j \{T_j = n\} \in \mathcal{F}_n$, hence T is a stopping time.

2.5 Exercises

1° Let S and T be stopping times, and $a > 0$ an integer. Show that $S + T$ and $S + a$ are stopping times, providing they take values in the parameter set.

2° Let $\{X_n, \mathcal{F}_n, n = 1, 2, \dots\}$ be a submartingale, and let $(\mathcal{G}_n)$ be a filtration with the property that X_n is $\mathcal{G}_n$-measurable for all n, and $\mathcal{G}_n \subset \mathcal{F}_n$ for all n. Then $\{X_n, \mathcal{G}_n, n = 1, 2, \dots\}$ is a submartingale.

3° Let S be a stopping time and $B \subset I\!\!R$ a Borel set. Show that $T \equiv \inf\{n > S : X_n \in B\}$ is a stopping time. (By convention, the inf of the empty set is ∞.)

4° Prove Proposition 2.10 (ii).

5° Let T be a stopping time. Prove that

$$\mathcal{F}_T = \{\, \lambda \in \mathcal{F} : \Lambda \cap \{\, T \le n \,\} \in \mathcal{F}_n, \forall\, n \}\ .$$

6° Let T be a stopping time, $\Lambda \in \mathcal{F}_T$ and $m \le n$ Show that $\Lambda \cap \{T = m\} \in \mathcal{F}_{T \wedge n}$.

7° Let $T_1 \ge T_2 \ge T_3 \ge \dots$ be a sequence of stopping times with limit T. Show that $\mathcal{F}_T = \cap_n \mathcal{F}_{T_n}$. Show that if the sequence of stopping times is increasing, i.e. if $T_1 \le T_2 \le \dots$ and if the limit T is finite, that $\mathcal{F}_T = \sigma\{\cup_n \mathcal{F}_{T_n}$. (Hint: if $T_n \downarrow T$ are all stopping times, they are integer-valued (or infinite), so that for large enough n, $T_n(\omega) = T(\omega)$.)

8° Find an example of a filtration and a sequence of stopping times $T_1 \le T_2 \le T_3 \dots$ with limit T for which the sigma-field generated by $\cup_n \mathcal{F}_{T_n}$ is strictly smaller than $\mathcal{F}_T$. (Hint: this is a rather technical point, having to do with the way the field $\mathcal{F}_T$ is defined when T has infinite values. Show that if $T \equiv \infty$, then $\mathcal{F}_T = \mathcal{F}$.)

2.6 System Theorems

Martingale system theorems get their name from the fact they are often invoked to show that gambling systems don't work. They are also known as martingale stopping theorems for the good reason that they involve martingales at stopping times. Many gambling systems involve the use of random times, and martingale system theorems show that if one starts with a (super) martingale—i.e. a fair or unfavorable game—then applying one of these systems to it won't change it into a favorable game. Consider for example the system whereby one observes the roulette wheel until there is a run of ten reds or ten blacks, and then one bets on the other color the next spin. This is often justified by the law of averages, but the justification is spurious: if the results of the spins are independent, then the probabilities of the wheel at this random time are exactly the same as they are at a fixed time. The system neither increases nor decreases the odds. This is true, but surprisingly subtle to prove. Try it! (In fact, it might be more reasonable to turn things around and bet on the *same* color after the run, on the grounds that such a long run might be evidence that something was fishy with the roulette wheel. But that's a fundamentally different system.)

The system theorems we will prove have the form: "a submartingale remains a submartingale under optional stopping" and "a submartingale remains a submartingale under optional sampling."

Definition 2.8 *Let $\{X_n,\ n = 0, 1, 2, \ldots\}$ be a stochastic process and T a positive-integer-valued random variable. Then*

$$X_T(\omega) = \sum_{n=0}^{\infty} X_n(\omega) I_{\{\omega : T(\omega) = n\}}\,.$$

Remark 2.12 *In other words, $X_T = X_n$ on the set $\{T = n\}$. Note that if X is adapted and T is a stopping time, X_T is $\mathcal{F}_T$-measurable. Indeed, if B is Borel, $\{X_T \in B\} \cap \{T = n\} = \{X_m = B\} \cap \{T = m\} \in \mathcal{F}_m$.*

Definition 2.9 *Let $\{X_n,\ n = 0, 1, 2, \ldots\}$ be a stochastic process and let T be a random variable taking values in $\mathbb{N} \cup \{\infty\}$. Then the* **process** X **stopped at** T *is the process $\{X_{n\wedge T},\ n = 0, 1, 2 \ldots\}$.*

If $T \equiv 3$, the stopped process is $X_0, X_1, X_2, X_3, X_3, X_3 \ldots$. The gambling interpretation of the process stopped at T is that the gambler stops playing the game at time T. Thereafter, his fortune remains unchanged.

Theorem 2.13 *Let $\{X_n, \mathcal{F}_n, n = 0, 1, 2, \ldots\}$ be a submartingale, and let T be a stopping time, finite or infinite. Then $\{X_{n\wedge T}, \mathcal{F}_n,\ n = 0, 1, 2, \ldots\}$ is also a submartingale, as is $\{X_{n\wedge T}, \mathcal{F}_{n\wedge T},\ n = 0, 1, 2, \ldots\}$.*

Proof. $X_{n\wedge T}$ is $\mathcal{F}_{n\wedge T}$-measurable and, as $n \wedge T \le n$, $\mathcal{F}_{n\wedge T} \le \mathcal{F}_n$, so $X_{n\wedge T}$ is also $\mathcal{F}_n$-measurable. Let $\Lambda \in \mathcal{F}_n$.

$$\int_\Lambda E\{X_{(n+1)\wedge T} \mid \mathcal{F}_n\}\, dP = \int_\Lambda X_{(n+1)\wedge T}\, dP = \int_{\Lambda\cap\{T\le n\}} X_{(n+1)\wedge T}\, dP + \int_{\Lambda\cap\{T>n\}} X_{n+1}\, dP.$$

$\Lambda \cap \{T > n\} \in \mathcal{F}_n$, so by the submartingale inequality this is

$$\ge \int_{\Lambda\cap\{T\le n\}} X_{n\wedge T}\, dP + \int_{\Lambda\cap\{T>n\}} X_n\, dP = \int_\Lambda X_{n\wedge T}\, dP.$$

Thus $\int_\Lambda X_{n\wedge T}\, dP \le \int_\Lambda E\{X_{(n+1)\wedge T} \mid \mathcal{F}_n\}\, dP$ for each $\Lambda \in \mathcal{F}_n$. As $X_{n\wedge T}$ is $\mathcal{F}_n$-measurable, this implies that $X_{n\wedge T} \le E\{X_{(n+1)\wedge T} \mid \mathcal{F}_n\}$ a.e., showing the first process is a submartingale. The fact that the second is also a submartingale follows from Exercise 2.5 2°. ♣

Remark 2.14 *The usefulness of this theorem lies in the fact that it has so few hypotheses on X and T. Stronger theorems require extra hypotheses. In effect there is a trade-off: we may either assume the process is bounded in some way, in which case we the stopping times can be arbitrary, or we may assume the stopping times are bounded, in which case no further hypotheses on the process are needed.*

Theorem 2.15 *Let $\{X_n, \mathcal{F}_n,\ n = 0, 1, 2, \dots\}$ be a sub-martingale and let $S \le T$ be bounded stopping times. Then X_S, X_T is a submartingale relative to the filtration $\mathcal{F}_S, \mathcal{F}_T$.*

PROOF. T is bounded, so let N be an integer such that $T \le N$. $S \le T \Longrightarrow \mathcal{F}_S \subset \mathcal{F}_T$, so the pair constitutes a filtration, and we know X_S is $\mathcal{F}_S$-measurable and X_T is $\mathcal{F}_T$-measurable.

$$E\{|X_T|\} = \sum_{j=0}^{N} E\{|X_j|; T = j\} \le \sum_{j=0}^{N} E\{|X_j|\} < \infty\,,$$

so that X_T and X_S are integrable. Suppose $\Lambda \in \mathcal{F}_S$.

$$\begin{aligned} \int_\Lambda X_T - X_S\, dP &= \sum_{j=0}^{N} \int_{\Lambda\cap\{S=j\}} X_T - X_j\, dP \\ &= \sum_{j=0}^{N}\sum_{k=j}^{N} \int_{\Lambda\cap\{S=j\}} X_{T\wedge(k+1)} - X_{T\wedge k}\, dP \\ &= \sum_{j=0}^{N}\sum_{k=j}^{N} \int_{\Lambda\cap\{S=j\}} E\{X_{T\wedge(k+1)} - X_{T\wedge k} \mid \mathcal{F}_j\}\,. \end{aligned}$$

But $X_{T\wedge k}$ is a submartingale relative to the $\mathcal{F}_k$ by Theorem 2.13, so $E\{X_{T\wedge(k+1)} - X_{T\wedge k} \mid \mathcal{F}_j\} = E\{E\{X_{T\wedge(k+1)} - X_{T\wedge k} \mid \mathcal{F}_k\} \mid \mathcal{F}_j\} \ge 0$, so the above is positive. Thus $\int_\Lambda X_T - X_S\, dP \ge 0$ for all $\Lambda \in \mathcal{F}_S$, and X_S, X_T is a martingale, as claimed. ♣

Corollary 2.16 *Let $\{X_n, \mathcal{F}_n,\ n = 0, 1, 2, \dots\}$ be a bounded sub-martingale and let $S \le T$ be finite stopping times. Then X_S, X_T is a submartingale relative to the filtration $\mathcal{F}_S, \mathcal{F}_T$.*

PROOF. X_S and X_T are bounded, hence integrable. To show that the pair X_S, X_T is a submartingale, it is enough to show that $X_S \le E\{X_T \mid \mathcal{F}_S\}$. Fix n for the moment. Then, as $S \wedge n \le T \wedge n$ are bounded stopping times, $X_{S\wedge n}, X_{T\wedge n}$ is a submartingale. Fix m. If $n \ge m$, and $\Lambda \in \mathcal{F}_S$, $\Lambda \cap \{\, S = m\,\} \in \mathcal{F}_{S\wedge n}$ (see the exercises) so that

$$\int_{\Lambda\cap\{\,S=m\,\}} X_{S\wedge n}\, dP \le \int_{\Lambda\cap\{\,S=m\,\}} X_{T\wedge n}\, dP \tag{16}$$

for all $n \ge m$. Let $n \to \infty$. $X_{S\wedge n} \to X_S$ and $X_{T\wedge n} \to X_T$ boundedly, so we can go to both the limit on both sides of (16) by the bounded convergence theorem to get

$$\int_{\Lambda\cap\{\,S=m\,\}} X_S\, dP \le \int_{\Lambda\cap\{\,S=m\,\}} X_T\, dP$$

Add over m to see

$$\int_\Lambda X_S\,dP \le \int_\Lambda X_T\,dP$$

which shows that $X_S \le E\{X_T \mid \mathcal{F}_S\}$ a.s. ♣

Notice that the above results are symmetric: they also apply, with the obvious changes, if the initial processes are supermartingales or martingales. We can extend Corollary 2.16 to processes which are bounded on one side only.

Corollary 2.17 *Let $\{X_n, \mathcal{F}_n\ n = 0, 1, 2, \ldots\}$ be a negative sub-martingale (resp. positive supermartingale) and let $S \le T$ be finite stopping times. Then X_S, X_T is a submartingale (resp. supermartingale) relative to the filtration $\mathcal{F}_S, \mathcal{F}_T$.*

PROOF. Let $\phi_N(x) = \max(x, -N)$. Then $\phi_N(X_n)$ is a submartingale by Proposition 2.5 (ii), and, since X is negative, it is bounded. Note that as $N \to \infty$, $\phi_N(x)$ decreases to x. Let us first show that X_S and X_T are integrable. (This is not obvious, since X is not bounded below.) However, the X_n are all integrable, so that

$$-\infty < E\{X_0\} \le E\{\phi_N(X_0)\} \le E\{\phi_N(X_S)\}\,,$$

where we have used the submartingale inequality. Now let $N \uparrow \infty$. $\phi_N(X_S) \downarrow X_S$, so by the Monotone Convergence Theorem, the last term decreases to $E\{X_S\} > -\infty$, implying X_S is integrable. The same argument works for X_T.

Apply Corollary 2.16 to the bounded submartingale $\phi_N(X_n)$: $\phi_N(X_S), \phi_N(X_T)$ is a submartingale relative to $\mathcal{F}_S, \mathcal{F}_T$, so for $\Lambda \in \mathcal{F}_S$,

$$\int_\Lambda \phi_N(X_S)\,dP \le \int_\Lambda \phi_N(X_T)\,dP\,. \tag{17}$$

Now let $N \uparrow \infty$. $\phi_N(x) \downarrow x$, so we can go to the limit on both sides of (17) to get

$$\int_\Lambda X_S\,dP \le \int_\Lambda X_T\,dP\,.$$

proving that X_S, X_T is a submartingale as claimed. ♣

Remark 2.18 *Corollary 2.17 says that a negative submartingale remains a submartingale under optional sampling and that a positive supermartingale remains a supermartingale under optional sampling. But it does* not *imply that a positive martingale remains a martingale under optional sampling. In fact, this is not necessarily true. Take, for instance, the martingale X_n of Example 5°. This is a positive martingale of expectation one, such that the stopping time $T \equiv \inf\{n : X_n = 0\}$ is almost surely finite. Take $S \equiv 1$. Then $X_S \equiv 1$ while $X_T \equiv 0$, and X_S, X_T is* not *a martingale. In fact, all one can conclude from the corollary is that a positive martingale becomes a supermartingale under optional sampling, and a negative martingale becomes a submartingale.*

2.7 Applications

Let $Y_1, Y_2, \dots$ be a sequence of independent random variables, with $P\{Y_j = 1\} = P\{Y_j = -1\} = 1/2$ for all j. Let $X_0 = x_0$, for some integer x_0, and $X_n = x_0 + \sum_{j=1}^{n} Y_j$. Then X is a symmetric random walk. One can use it to model one's fortune at gambling (the initial fortune is x_0 and the gambler bets one dollar at even odds each time) or the price of a stock (the stock starts at price x_0, and at each trade its value goes up or down by one unit.) In either case, one might be interested in the following problem: let $a < x_0 < b$ be integers. What is the probability p that X_n reaches b before a? This problem is called the gamblers ruin: a gambler starts with x_0 dollars, and decides to gamble until he makes b dollars or goes broke. if one takes $a = 0$, then we are asking "What is the probability that the gamblers fortune reaches b before he is ruined?" The probability of his ruin is $1 - p$.

Let us assume that X_n eventually hits either a or b—it does, as we will see later as an easy consequence of the martingale convergence theorem.

Let $\mathcal{F}_n$ be the natural filtration generated by the X_n (or, equivalently, by the Y_n). Note that $\{X_n, \mathcal{F}_n, n \geq 0\}$ is a martingale—it was one of the examples we gave. Now define a stopping time

$$T = \inf\{n \geq 0 : X_n = b \text{ or } a\}\,.$$

Then T is a finite stopping time. Now (X_n) is a martingale, so the stopped process $X_{n\wedge T}$ is also a martingale. Moreover, it never leaves the interval $[a, b]$, so it is bounded, and we can apply the stopping theorem to the stopping times $S \equiv 0$ and T: X_0, X_T is a martingale. Thus $E\{X_T\} = E\{X_0\} = x_0$. But if p is the probability of hitting b before a—which is exactly the probability that $X_T = b$—then $x_0 = E\{X_T\} = pb + (1-p)a$ from which we see

$$p = \frac{x_0 - a}{b - a}\,.$$

We might ask how long it takes to hit a or b. Let us compute the expectation of T. First, we claim that $Z_n \equiv X_n^2 - n$ is a martingale. The measurability properties and integrability properties are clear, so we need only check

$$\begin{aligned} E\{Z_{n+1} - Z_n \mid \mathcal{F}_n\} &= E\{X_{n+1}^2 - n - 1 - X_n^2 + n \mid \mathcal{F}_n\} \\ &= E\{(X_{n+1} - X_n)^2 - 2X_n(X_{n+1} - X_n) - 1 \mid \mathcal{F}_n\} \end{aligned}$$

Now $(X_{n+1} - X_n)^2 \equiv 1$, so this is

$$= E\{X_n(X_{n+1} - X_n) \mid \mathcal{F}_n\} = X_n E\{X_{n+1} - X_n \mid \mathcal{F}_n\} = 0$$

since X is a martingale. This verifies the claim.

Once again we apply the system theorem to the bounded stopping times $S \equiv 0$ and $T \wedge N$ for a (large) integer N. Thus

$$x_0^2 = E\{Z_{T\wedge N}\} = E\{X_{T\wedge N}^2 - T \wedge N\}$$

so that for each N

(18) $$E\{T \wedge N\} = E\{X_{T\wedge N}^2\} - x_0^2.$$

As $N \to \infty$, $T \wedge N \uparrow T$ and $X_{T\wedge N} \to X_T$ boundedly, since $a \leq X_{T\wedge N} \leq b$, so we can go to the limit on both sides of (18), using monotone convergence on the left and bounded convergence on the right. Thus $E\{T\} = E\{X_T^2\} - x_0^2$. We found the distribution of X_T above, so we see this is

$$= \frac{x_0 - a}{b - a} b^2 + \frac{b - x_0}{b - a} a^2 - x_0^2 = (b - x_0)(x_0 - a).$$

Thus $E\{T\} = (b - x_0)(x_0 - a)$.

Let's go a little further. Suppose the gambler keeps on gambling until he goes broke–that is, he does not stop if his fortune reaches b.

This is almost immediate: let $b \to \infty$ in the above. Note that the stopping time must increase with b, and so as $b \uparrow \infty$, $T \uparrow T_a \equiv \inf\{n : X_t = a\}$, so by monotone convergence

$$E\{T_a\} = \lim_{b\to\infty} (b - x_0)(x_0 - a) = \infty.$$

There is hope for the gambler! Yes, he will go broke in the end, but it will take him an (expected) infinite time! (Of course, by symmetry, the expected time to reach his goal of b is also infinite.)

2.8 The Maximal Inequality

According to Chebyshev's inequality, if X is a random variable and $\lambda > 0$, $\lambda P\{|X| \geq \lambda\} \leq E\{|X|\}$. It turns out that martingales satisfy a similar, but much more powerful inequality, which bounds the maximum of the process.

Theorem 2.19 *Let $\{X_n, n = 0, 1, 2, \ldots, N\}$ be a positive submartingale. Then*

(19) $$\lambda P\{\max_{n\leq N} X_n \geq \lambda\} \leq E\{X_N\}.$$

Remark 2.20 *If (X_n) is a martingale, $(|X_n|)$ is a positive submartingale, and we have $\lambda P\{\max_{n\leq N} |X_n| \geq \lambda\} \leq E\{|X_N|\}$, which is the extension of Chebyshev inequality we mentioned above.*

PROOF. Define a bounded stopping time T by

$$T = \begin{cases} \inf\{\, n \le N : X_n \ge \lambda \,\} & \\ N & \text{if there is no such } n\,. \end{cases}$$

Note that $\max X_n \ge \lambda$ if and only if $X_T \ge \lambda$. The system theorem implies that X_T, X_N is a submartingale. Since X is positive, $X_T \ge 0$, so

$$\lambda P\{\, \max_n X_n \ge \lambda \,\} = \lambda P\{\, X_T \ge \lambda \,\} \le E\{\, X_T \,\} \le E\{\, X_N \,\}\,.$$

♣

2.9 The Upcrossing Inequality

The inequality of this section is one of the most satisfying kind of results: it is elegant, unexpected, powerful, and its proof may even teach us more about something we thought we understood quite well.

Let us define the number of *upcrossings of an interval* $[a, b]$ by the finite sequence of real variables $x_0, x_1, \ldots, x_N$. This is the number of times the sequence goes from below a to above b.

Set $\alpha_0 = 0$,

$$\alpha_1 = \begin{cases} \inf\{\, n \le N : x_n \le a \,\} & \\ N+1 & \text{if there is no such } n\,. \end{cases}$$

and for $k \ge 1$

$$\beta_k = \begin{cases} \inf\{\, n \ge \alpha_k : x_n \ge b \,\} & \\ N+1 & \text{if there is no such } n\,. \end{cases}$$

$$\alpha_{k+1} = \begin{cases} \inf\{\, n \ge \beta_k : x_n \ge b \,\} & \\ N+1 & \text{if there is no such } n\,. \end{cases}$$

If $\beta_k \le N$, $x_{\beta_k} \ge b$ and $x_{\alpha_k} \le a$, so the sequence makes an upcrossing of $[a, b]$ between α_k and β_k. It makes its downcrossings of the interval during the complementary intervals $[\beta_k, \alpha_{k+1}]$.

Definition 2.10 The number of upcrossings of the interval $[a, b]$ **by the sequence** $x_0, x_1, \ldots, x_N$ *is* $\nu_N(a, b) \equiv \sup\{\, k : \beta_k \le N \,\}$.

Now let $\{\, \mathcal{F}_n,\ n = 0, 1, \ldots N \,\}$ be a filtration and $\{\, X_n,\ n = 0, 1, \ldots \}$ a stochastic process adapted to $(\mathcal{F}_n)$. Replace the sequence $x_0, \ldots, x_n$ by $X_0(\omega), \ldots X_N(\omega)$. Then the α_k and β_k defined above are stopping times (see Exercise 2.5 3°) and $\nu_N(a, b)$ is the number of upcrossings of $[a, b]$ by the process $X_0, \ldots X_N$.

Theorem 2.21 *(Upcrossing Inequality) Let* $\{ X_n, \mathcal{F}_n,\ n = 0, 1, \dots, N \}$ *be a submartingale and let* $a < b$ *be real numbers. Then the number of upcrossings* $\nu_N(a, b)$ *of* $[a, b]$ *by* (X_n) *satisfies*

$$E\{ \nu_N(a,b) \} \leq \frac{E\{ (X_N - a)^+ \}}{b-a} \tag{20}$$

Proof. Let $\hat{X}_n = X_n \vee a$ for $n \geq 0$. Then $(\hat{X}_n)$ is again a submartingale, and it has the same number of upcrossings of $[a, b]$ as does the original process X. Define the process for $n = N+1$ by setting $\hat{X}_{N+1} = \hat{X}_N$. The process remains a submartingale. Now write

$$\hat{X}_N - \hat{X}_0 = \hat{X}_{\alpha_1} - \hat{X}_0 + \sum_{n=1}^{N} (\hat{X}_{\beta_n} - \hat{X}_{\alpha_n}) + \sum_{n=1}^{N} (\hat{X}_{\alpha_{n+1}} - \hat{X}_{\beta_n})$$

(Notice that if $\alpha_n \leq N$, then $\hat{X}_n \leq a$, hence $\beta_n \geq 1 + \alpha_n$, so that $\alpha_n = n$, and in particular, $\alpha_{N+1} = N+1$ in the above sum.) Take the expectation of this:

$$E\{ \hat{X}_N - \hat{X}_0 \} \;=\; E\{ \hat{X}_{\alpha_1} - \hat{X}_0 \} + E\left\{ \sum_{n=1}^{N} (\hat{X}_{\beta_n} - \hat{X}_{\alpha_n}) \right\} + E\left\{ \sum_{n=1}^{N} (\hat{X}_{\alpha_{n+1}} - \hat{X}_{\beta_n}) \right\}.$$

But now, $\hat{X}$ is a submartingale and the α_n and β_n are stopping times. By the system theorem, $\hat{X}_N - \hat{X}_0$, $\hat{X}_{\alpha_1} - \hat{X}_0$, and $\hat{X}_{\alpha_{n+1}} - \hat{X}_{\beta_n}$ all have positive expectations, so that the above is

$$\geq E\left\{ \sum_{n=1}^{N} (\hat{X}_{\beta_n} - \hat{X}_{\alpha_n}) \right\}$$

If $\beta_n \leq N$, it marks the end of an upcrossing of $[a, b]$, so $\hat{X}_{\beta_n} - \hat{X}_{\alpha_n} \geq b - a$. Thus, there are at least $\nu_N(a, b)$ terms of the first sum which exceed $b - a$, and the expectation of any remaining terms is positive, so that this is

$$\geq (b-a) E\{ \nu_N(a,b) \}.$$

Thus $E\{ \nu_N(a,b) \} \leq E\{ \hat{X}_N - \hat{X}_0 \}/(b-a)$. But now, in terms of the original process, $\hat{X}_N - \hat{X}_0 \leq X_N \vee a - a = (X_N - a)^+$, which implies (20). ♣

Remark 2.22 *In the proof, we threw away* $E\left\{ \sum_{n=1}^{N} (\hat{X}_{\alpha_{n+1}} - \hat{X}_{\beta_n}) \right\}$ *because it was positive. However, the downcrossings of* $[a, b]$ *occur between the* β_n *and* α_{n+1}*: we expect* $X_{\alpha_{n+1}} \leq a$ *and* $X_{\beta_n} \geq b$ *so that each of these terms would seem to be negative ...? Why is this not a contradiction?*

The above result extends to an infinite parameter set. If $X = \{ X_n,\ n = 0, 1, 2, \dots \}$ is a process, define the number of upcrossings $\nu_\infty(a, b)$ by X as $\nu_\infty(a, b) = \lim_{N \to \infty} \nu_N(a, b)$. Then the upcrossing inequality extends to the following.

Corollary 2.23 *Let* $\{ X_n, \mathcal{F}_n, \ n = 0, 1, \dots \}$ *be a submartingale and let* $a < b$ *be real numbers. Then the number of upcrossings* $\nu_\infty(a, b)$ *of* $[a, b]$ *by* (X_n) *satisfies*

$$E\{ \nu_\infty(a, b) \} \le \frac{\sup_N E\{ (X_N - a)^+ \}}{b - a} \tag{21}$$

PROOF. The proof is almost immediate. As $N \to \infty$, the number $\nu_N(a, b)$ of upcrossings of $[a, b]$ by $X_0, \dots, X_N$ increases to $\nu_\infty(a, b)$, so $E\{ \nu_\infty(a, b) \} = \lim_{N \to \infty} E\{ \nu_N(a, b) \}$. By the theorem, this is bounded by $\sup E\{ (X_N - a)^+ \}/(b - a)$. ♣

2.10 Martingale Convergence

One of the basic facts about martingales is this: if you give it half a chance, a martingale will converge. There are several nuances, but here is the most basic form of the convergence theorem.

Theorem 2.24 *(Martingale Convergence Theorem) Let* $\{ X_n, \mathcal{F}_n, \ n = 0, 1, \dots \}$ *be a submartingale and suppose that* $E\{ |X_n| \}$ *is bounded. Then with probability one, there exists a finite integrable r.v.* X_∞ *such that*

$$\lim_{n \to \infty} X_n = X_\infty \ a.e.$$

PROOF. First, let us show that $\liminf X_n(\omega) = \limsup X_n(\omega)$. This is where the upcrossing inequality enters.

Suppose the sequence does not converge. Then there exist rational numbers a and b such that $\liminf X_n(\omega) < a < b < \limsup X_n(\omega)$. It follows that there exists a subsequence n_k such that $X_{n_k}(\omega) \to \liminf X_n(\omega)$ and another subsequence n_j such that $X_{n_j}(\omega) \to \limsup X_n(\omega)$, and in particular, $X_{n_k}(\omega) < a$ for infinitely many k, and $X_{n_j}(\omega) > b$ for infinitely many j. This implies that there are infinitely many upcrossings of $[a, b]$. Thus, if $\limsup X_n(\omega) > \liminf X_n(\omega)$, there exist rational $a < b$ such that the number of upcrossings of $[a, b]$ is infinite.

Now for every pair $r_1 < r_2$ of rationals, $E\{ \nu_\infty(r_1, r_2) \} \le \sup_n E\{ (X_N - a)^+ \}/(b - a) \le \sup_N (E\{ |X_N| \} + a)/(b - a) < \infty$. It follows that $P\{ \nu_\infty(r_1, r_2) = \infty \} = 0$. This is true for each of the countable number of pairs (r_1, r_2) of rationals, hence $P\{ \exists \ r_1 < r_2 \in \mathbb{Q} : \nu_\infty(r_1, r_2) = \infty \} = 0$. Thus for a.e. ω, $\limsup X_n(\omega) = \liminf X_n(\omega)$.

Thus the limit exists a.e. but, a priori, it might be infinite. To see it is finite, note that $\lim |X_n|$ exists a.s. and by Fatou's Lemma, $E\{ \lim |X_n| \} \le \liminf E\{ |X_n| \} < \infty$. In particular, $\lim |X_n| < \infty$ a.e. ♣

Thus, to show convergence of a martingale, submartingale, or supermartingale, it is enough to show that its absolute expectation is bounded.

Corollary 2.25 *A positive supermartingale and a negative submartingale each converge with probability one.*

PROOF. If $\{ X_n, \mathcal{F}_n,\ n = 0, 1, \ldots \}$ is a positive supermartingale, then $E\{ |X_n| \} = E\{ X_n \} \leq E\{ X_0 \}$, so the process converges a.e. by Theorem 2.24.

Remark 2.26 *For a very quick application of this, recall the random walk (X_n) introduced in the Gambler's Ruin. We stated, but did not show, that (X_n) eventually reached the complement of any finite interval. Let us show that for any a, there exists n such that $X_n \leq a$.*

Let $T = \inf\{ n : X_n < a \}$. Now X_n is a martingale, $X_0 = x_0$, and $|X_{n+1} - X_n| = 1$ for all n. Then $\hat{X}_n \equiv X_{n \wedge T}$ is a martingale which is bounded below by the minimum of x_0 and $a - 1$, so it converges a.e. by the Martingale Convergence Theorem. But the only way such a process can converge is to have $T < \infty$. Indeed, $n < T \Longrightarrow |\hat{X}_{n+1} - \hat{X}_n| = 1$, so convergence is impossible on the set $\{ T = \infty \}$. But if $T < \infty$, $X_T < a$, so the process does eventually go below a.

It is a rather easy consequence of this to see that X_n must visit all integers, positive or negative. In fact, it must visit each integer infinitely often, and with probability one, both $\liminf X_n = -\infty$ and $\limsup X_n = \infty$. We leave the details to the reader as an exercise.

2.11 Exercises

1° Suppose $\{ X_n, \mathcal{F}_n,\ n = 0, 1, 2, \ldots \}$ is a submartingale and $T_1 \leq T_2, \leq \ldots$ is an increasing sequence of stopping times. Suppose that either X is bounded, or that each of the stopping times is bounded. Let $Z_n = X_{T_n}$ and $\mathcal{G}_n = \mathcal{F}_{T_n}$. Show that $\{ Z_n, \mathcal{G}_n,\ n = 0, 1, 2, \ldots \}$ is a submartingale.

2° Find the probability of the gambler's ruin if the random walk is *not* symmetric, i.e. if $P\{ Y_j = 1 \} = p$ and $P\{ Y_j = -1 \} = 1 - p$ for $0 < p < 1$, $p \neq 1/2$. (Hint: look for a martingale of the form $Z_n = r^{X_n}$.)

3° Suppose that X_1, X_2, ... is a sequence of i.i.d. random variables and that T is a finite stopping time relative to the natural filtration $\mathcal{F}_n = \sigma\{ X_1, \ldots, X_n \}$. Show that X_{T+1} has the same distribution as X_1, and is independent of $\mathcal{F}_T$. Apply this to roulette: if T is the first time that a string of ten reds in a row comes up, what is the probability of red on the $T + 1$st spin?

3° Let $x = x_1, \ldots, x_n$ be a sequence of real numbers and let $x' \stackrel{\text{def}}{=} x_{n_1}, \ldots, x_{n_k}$ be a subsequence. Show that if $a < b$, the number of upcrossings of $[a, b]$ by x' is less than or equal to the number of upcrossings of $[a, b]$ by x.

5° Show that with probability one, the simple symmetric random walk visits all integers infinitely often.

2.12 Uniform Integrability

If X is an integrable random variable, and if Λ_n is a sequence of sets with $P\{\Lambda_n\} \to 0$, then $\int_{\Lambda_n} X\, dP \to 0$. (This is a consequence of the dominated convergence theorem: $|X|$ is integrable, $|XI_{\Lambda_n}| \le |X|$, and XI_{Λ_n} converges to zero in probability.) This is often called *uniform integrability*. We will extend it to arbitrary sets of random variables. In what follows, I is an arbitrary index set.

Definition 2.11 *A family $\{X_\alpha, \alpha \in I\}$ of random variables is* **uniformly integrable** *if*

$$\lim_{N\to\infty} \int_{\{|X_\alpha| \ge N\}} |X_\alpha|\, dP = 0 \tag{22}$$

uniformly in α.

In other words, the family is uniformly integrable if the supremum over α of the integral in (22) tends to zero as $N \to \infty$. There are two reasons why this property is important:

(i) uniform integrability is a necessary and sufficient condition for going to the limit under an expectation; and

(ii) it is often easy to verify in the context of martingale theory.

Property (i) would seem to be enough to guarantee that uniform integrability is interesting, but it isn't: the property is not often used in analysis; when one wants to justify the interchange of integration and limits, one usually uses the dominated or monotone convergence theorems. It is only in probability where one takes this property seriously, and that is because of (ii).

Example 2.1 *1° Any finite family of integrable r.v. is uniformly integrable. More generally, if each of a finite number of families is uniformly integrable, so is their union.*

2° If there exists a r.v. Y such that $|X_\alpha| \le Y$ a.e. for all $\alpha \in I$, and if $E\{Y\} < \infty$, then $\{X_\alpha,\ \alpha \in I\}$ is uniformly integrable.

3° If there exists $K < \infty$ for which $E\{|X_\alpha|^2\} \le K$ for all α, then the family $\{X_\alpha,\ \alpha \in I\}$ is uniformly integrable.

To see 1°, if $(X^i_{\alpha_i})$ are uniformly integrable families for $i = 1, \dots, n$, then for each i, $\sup_{\alpha_i} \int_{\{|X^i_{\alpha_i}| \ge N\}} X^i_{\alpha_i}\, dP$ tends to zero as $N \to \infty$, hence so does the maximum over i. But this is the supremum of the integrals over the union of the families for $i = 1, \dots, n$.

To see 2°, note that $P\{|X_\alpha| \ge N\} \le E\{|X_\alpha|\}/N \le E\{Y\}/N \to 0$ as $N \to \infty$. Thus $\int_{\{|X_\alpha| > N\}} |X_\alpha|\, dP \le \int_{\{|X_\alpha| > N\}} Y\, dP$, which tends to zero as $N \to \infty$ since the probability of the set over which we integrate goes to zero. We leave 3° as an exercise for the reader.

Note that the second example shows that the hypotheses of the Dominated Convergence Theorem imply uniform integrability. The definition of uniform integrability looks different from the property of uniform integrability of a single random variable we gave above. Here is the connection.

Proposition 2.27 *Suppose $\{ X_\alpha, \alpha \in I \}$ is uniformly integrable. Then*

(i) $\sup_\alpha E\{ |X_\alpha| \} < \infty$.

(ii) $\lim_{N\to\infty} P\{ |X_\alpha| > N \} = 0$ *uniformly in* α.

(iii) $\lim_{P\{\Lambda\}\to 0} \int_\Lambda |X_\alpha| \, dP = 0$ *uniformly in* α.

Conversely, either (i) and (iii) or (ii) and (iii) imply uniform integrability.

PROOF. (i): There exists N_0 such that for all α, $\int_{\{ |X_\alpha| \geq N_0 \}} |X_\alpha| \, dP \leq 1$. Then for all α, $E\{ |X_\alpha| \} = \int_{\{ |X_\alpha| < N_0 \}} |X_\alpha| \, dP + \int_{\{ |X_\alpha| \geq N_0 \}} |X_\alpha| \, dP \leq N_0 + 1$.

Note that $(i) \Longrightarrow (ii)$, for $P\{ |X_\alpha| > N \} \leq (1/N) \sup_\alpha E\{ |X_\alpha| \} \longrightarrow 0$.

To see (iii), let $\epsilon > 0$ and choose N_ε such that $\int_{\{ |X_\alpha| > N_\varepsilon \}} |X_\alpha| \, dP < \varepsilon/2$ for all α, which we can do since the X_α are uniformly integrable. Then

$$\begin{aligned}
\int_\Lambda |X_\alpha| \, dP &= \int_{\Lambda \cap \{ |X_\alpha| \leq N_\varepsilon \}} |X_\alpha| \, dP + \int_{\Lambda \cap \{ |X_\alpha| > N_\varepsilon \}} |X_\alpha| \, dP \\
&\leq N_\varepsilon P\{ \Lambda \} + \int_{\{ |X_\alpha| > N_\varepsilon \}} |X_\alpha| \, dP \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2}
\end{aligned}$$

if $P\{ \Lambda \}$ is small enough.

Conversely, $(i) \Longrightarrow (ii)$, so suppose (ii) and (iii) hold. Choose $\varepsilon > 0$. By (iii) there exists $\delta > 0$ such that $P\{ \Lambda \} < \delta \Longrightarrow \int_\Lambda |X_\alpha| \, dP < \varepsilon$ for all α. Then by (ii) there exists N_0 such that $N \geq N_0 \Longrightarrow P\{ |X_\alpha| > N \} < \delta$ for all α. This implies in turn that

$$\int_{\{ |X_\alpha| > N \}} |X_\alpha| \, dP < \varepsilon \, .$$

This is true for all α, which implies uniform integrability. ♣

As we said before, uniform integrability is a necessary and sufficient condition for going to the limit under the integral. Another way of saying this is this. Recall that a process $\{ X_n, \ n = 0, 1, 2 \dots \}$ is said to *converge in L^1* or *converge in the mean* to an integrable r.v. X_∞ if $\lim_{n\to\infty} E\{ |X_n - X_\infty| \} = 0$.

If X_n converges to X_∞ in L^1, it is easy to see that $E\{ X_n \} \to E\{ X_\infty \}$ and that for any set Λ, $\int_\Lambda X_n \, dP \longrightarrow \int_\Lambda X_\infty \, dP$, so that L^1 convergence implies that we can go to the limit under the integral sign. So in fact, we are really talking about L^1 convergence here, and the point is that while L^1 convergence implies convergence in probability, the converse is not necessarily true, and conditions implying L^1 convergence are important.

Theorem 2.28 *Let $\{ X_n, \ n = 0, 1, \dots \}$ be a sequence of integrable r.v. which converge in probability to a r.v. X. Then the following are equivalent.*

(i) $\{ X_n,\ n = 0, 1, \dots \}$ *is uniformly integrable.*
(ii) $E\{ |X_n - X| \} \longrightarrow 0$.
(iii) $E\{ |X_n| \} \longrightarrow E\{ |X| \}$.

PROOF. *(i)* $\Longrightarrow$ *(ii)*: $|X_n|$ converges to $|X|$ in probability, so that some subsequence (X_{n_k}) converges a.e.. By Fatou's Lemma, $E\{ |X| \} \le \liminf E\{ |X_{n_k}| \}$, which is finite by Proposition 2.27 *(i)*. Therefore X is integrable. Let $\varepsilon > 0$.

$$\begin{aligned} E\{ |X - X_n| \} &= \int_{\{ |X-X_n| \le \varepsilon/3 \}} |X - X_n| \, dP + \int_{\{ |X-X_n| > \varepsilon/3 \}} |X - X_n| \, dP \\ &\le \frac{\varepsilon}{3} + \int_{\{ |X-X_n| > \varepsilon/3 \}} |X| \, dP + \int_{\{ |X-X_n| > \varepsilon/3 \}} |X_n| \, dP \,. \end{aligned}$$

Now $P\{ |X - X_n| > \varepsilon/3 \} \to 0$ as $n \to \infty$ by convergence in probability, so each of the last two integrals tend to zero as $n \to \infty$, the first because X is integrable, and the second by Proposition 2.27 (iii). This implies *(ii)*.

It is clear that *(ii)* $\Longrightarrow$ *(iii)*. To see that *(iii)* $\Longrightarrow$ *(i)*, suppose $E\{ |X_n| \} \to E\{ |X| \}$. Choose a real number M such that $P\{ |X| = M \} = 0$, and truncate X at M:

$$X^{(M)} = \begin{cases} X & \text{if } |X| \le M \\ 0 & \text{if } |X| > M \,. \end{cases}$$

Now, using the same notation for $X_n^{(M)}$,

$$\int_{\{ |X_n| > M \}} |X_n| \, dP = E\{ |X_n| \} - E\{ |X_n^{(M)}| \} \longrightarrow E\{ |X| \} - E\{ |X^{(M)}| \}$$

since $E\{ |X_n| \}$ converges by hypothesis and $E\{ |X_n^{(M)}| \}$ converges by the Bounded Convergence Theorem. Let $\varepsilon > 0$ and choose M_ε large enough so that $E\{ |X| \} - E\{ |X^{(M)}| \} < \varepsilon/3$. There exists N_0 such that for $n \ge N_0$, we have both

$$\begin{aligned} |E\{ |X_n| \} - E\{ |X| \}| &< \frac{\varepsilon}{3} \\ |E\{ |X_n^{(M_\varepsilon)}| \} - E\{ |X^{(M_\varepsilon)}| \}| &< \frac{\varepsilon}{3} \,. \end{aligned}$$

Thus if $n \ge N_0$, we have

$$\int_{\{ |X_n| > M_\varepsilon \}} |X_n| \, dP < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \,.$$

But there are only finitely many $n < N_0$, so there exist $M'_\varepsilon > M_\varepsilon$ such that

$$\int_{\{ |X_n| > M'_\varepsilon \}} |X_n| \, dP < \varepsilon$$

for *all* n. This implies (i). ♣

The second reason that uniform integrability is important is that it is readily verified for martingales and submartingales.

Theorem 2.29 *Let $\{\mathcal{F}_\alpha,\ \alpha \in I\}$ be a family of sub sigma fields of $\mathcal{F}$ and let X be an integrable random variable. Define $X_\alpha = E\{X \mid \mathcal{F}_\alpha\}$. Then $\{X_\alpha,\ \alpha \in I\}$ is uniformly integrable.*

PROOF. $P\{|X_\alpha| > N\} \leq \frac{1}{N}E\{|X_\alpha|\} \leq \frac{1}{N}E\{|X|\}$, which clearly goes to 0 as $N \to \infty$, uniformly α. Now X_α, X is a martingale, so $|X_\alpha|, |X|$ is a submartingale, and

$$\int_{\{|X_\alpha|>N\}} |X_\alpha|\, dP \leq \int_{\{|X_\alpha|>N\}} |X|\, dP\,.$$

We have just seen that $P\{|X_\alpha| > N\}$ tends to zero uniformly in α, so the last integral tends to zero, uniformly in α by the uniform integrability of X itself. ♣

Corollary 2.30 *Let $I \subset \bar{\mathbb{R}}$ be a set with a largest element, and suppose $\{X_t, \mathcal{F}_t,\ t \in I\}$ is a positive submartingale. Then $\{X_t,\ t \in I\}$ is uniformly integrable.*

PROOF. Let t_0 be the largest element of I. Then $0 \leq X_t \leq E\{X_{t_0} \mid \mathcal{F}_t\}$. But the conditional expectations of X_{t_0} are uniformly integrable by Theorem 2.29, hence so are the X_t. ♣

Corollary 2.31 *A martingale with a last element is uniformly integrable.*

PROOF. If (X_t) is a martingale, $(|X_t|)$ is a positive submartingale, and the result follows from Corollary 2.30. ♣

Let us now use this idea to refine the martingale system theorems.

Theorem 2.32 *Let $\{X_n, \mathcal{F}_n,\ n = 0, 1, 2, \dots\}$ be a uniformly integrable submartingale, and let $S \leq T$ be finite stopping times. Then X_S, X_T is a submartingale relative to $\mathcal{F}_S, \mathcal{F}_T$.*

PROOF. First, X_S and X_T are integrable. Indeed, by Fatou's Lemma,

$$\begin{aligned} E\{|X_T|\} &\leq \liminf E\{|X_{T\wedge n}|\} \\ &= \liminf \left(2E\{X^+_{T\wedge n}\} - E\{X_{T\wedge n}\}\right) \\ &\leq 2\liminf E\{|X_n|\} - E\{X_0\}\,. \end{aligned}$$

where we have used the fact that $X_0, X_{T\wedge n}$ and $X^+_{T\wedge n}, X^+_n$ are submartingales. But $E\{\,|X_n|\,\}$ is bounded in n since (X_n) is uniformly integrable, so this is finite. The same calculation holds for X_S.

$S \wedge n \leq T \wedge n$ are bounded stopping times, so $X_{S\wedge n}, X_{T\wedge n}$ is a submartingale relative to $\mathcal{F}_{S\wedge n}, \mathcal{F}_{T\wedge n}$. Let $\Lambda \in \mathcal{F}_S$. If $m < n$, $\Lambda \cap \{\, S = m \,\} \in \mathcal{F}_{S\wedge n}$ (why?) so

$$\int_{\Lambda\cap\{\, S=m\,\}} X_{S\wedge n}\, dP \leq \int_{\Lambda\cap\{\, S=m\,\}} X_{T\wedge n}\, dP$$

The processes $(|X_{S\wedge n}|)$ and $(|X_{T\wedge n}|)$ are uniformly integrable: for example the latter is bounded above by $|X_T| + |X_n|$, which is uniformly integrable by hypothesis. Thus we can let $n \to \infty$ above: $X_{S\wedge n} \to X_S$ and $X_{T\wedge n} \to X_T$, so by Theorem 2.28 we can go to the limit to see that

$$\int_{\Lambda\cap\{\, S=m\,\}} X_S\, dP \leq \int_{\Lambda\cap\{\, S=m\,\}} X_T\, dP\,.$$

Now add over m to get

$$\int_\Lambda X_S\, dP \leq \int_\Lambda X_T\, dP, \quad \forall\, \Lambda \in \mathcal{F}_S\,,$$

which shows that $X_S \leq E\{\, X_T \mid \mathcal{F}_S\,\}$. ♣

2.13 Martingale Convergence, Part Two

A submartingale with a bounded absolute expectation converges a.e., but there is no guarantee that it converges in L^1. Indeed, it may not. Example 5° of Section 2.1 concerned a martingale (X_n) with $E\{\, X_n\,\} = 1$ for all n, whose limit $X_\infty \equiv 0$. Thus $\lim E\{\, X_n\,\} = 1 \neq E\{\, X_\infty\,\} = 0$. We need an extra condition on the martingale or submartingale to ensure that it converges in L^1. This extra condition is uniform integrability.

We will look at two different kinds of convergence. First, when the parameter set is $0, 1, 2, \ldots$, we ask whether the sequence X_n converges to a limit X_∞ as $n \to \infty$. In the second case, when the parameter set is $\ldots, -2, -1$, we ask whether X_n converges to a limit $X_{-\infty}$ as $n \to -\infty$. These two are equivalent if we are just talking about the convergence of a sequence of r.v.—just replace the parameter n by $-n$—but if we are talking about martingales, a filtration $(\mathcal{F}_n)$ is involved. The filtration increases with n, which fixes the direction of time. So as $n \to \infty$, we are looking at at the behavior as the filtrations get larger and larger, but when $n \to -\infty$, the filtrations are getting smaller and smaller. For n represents time, and the far, far future may not be the same as the far, far past. As it happens, the results are similar but not identical in the two cases. A martingale of the form $\{\, X_n,\ n = \ldots - 2, -1\,\}$ is sometimes called a *backwards martingale.*

A closely related question is this: suppose a martingale X_n converges to X_∞ as $n \to \infty$. Can we add X_∞ to the martingale as a final element? That is, is $\{\, X_n, \mathcal{F}_n,\ n =$

$1, 2, \ldots, \infty \}$ a martingale? This is equivalent to asking if $X_n = E\{ X_\infty \mid \mathcal{F}_n \}$ for each n. We speak of *closing* the martingale with X_∞, to get a martingale with parameter set $0, 1 \ldots, \infty$ in the extended reals. Similarly, if $n \to -\infty$ in a backward martingale, can we add $X_{-\infty}$ as an *initial* element of the martingale? This is equivalent to setting $\mathcal{F}_{-\infty} = \cap_n \mathcal{F}_n$ and asking if $X_{-\infty} = E\{ X_n \mid \mathcal{F}_{-\infty} \}$. This time the closed process has the parameter set $-\infty, \ldots, -1$.

Theorem 2.33 *Let $\{ X_n, \mathcal{F}_n,\ n = 0, 1, 2, \ldots \}$ be a uniformly integrable submartingale. Then $X_\infty \equiv \lim_{n\to\infty} X_n$ exists a.e. and in L^1, and $\{ X_n, n = 0, 1, \ldots, \infty \}$ is a submartingale, where $\mathcal{F}_\infty = \sigma\{ \cup_n \mathcal{F}_n \}$.*

Proof. If (X_n) is uniformly integrable, $E\{ |X_n| \}$ is bounded (Proposition 2.27), so X_n converges a.e. by the Martingale Convergence Theorem to an integrable r.v. X_∞. It follows by uniform integrability that the sequence also converges in L^1. X_∞ is clearly $\mathcal{F}_\infty$-measurable, so we need only check the submartingale inequality. Let $\Lambda \in \mathcal{F}_m$. Then for all $n \geq m$ $\int_\Lambda X_m \, dP \leq \int_\Lambda X_n \, dP$. Let $n \to \infty$. Since $X_n \to X_\infty$ in L^1, we can go to the limit: $\int_\Lambda X_n \, dP \longrightarrow \int_\Lambda X_\infty \, dP$, implying that $\int_\Lambda X_m \, dP \leq \int_\Lambda X_\infty \, dP$, which is the desired inequality. ♣

The theorem for backward submartingales is somewhat easier: the limits always exist, and uniform integrability, for martingales, at least, comes for free.

Theorem 2.34 *Let $\{ X_n, \mathcal{F}_n,\ n = \ldots, -2, -1 \}$ be a submartingale. Then*

(i) $\lim\limits_{n\to-\infty} E\{ X_n \} \equiv L$ exists, and $X_{-\infty} \equiv \lim\limits_{n\to-\infty} X_n$ exists a.e., where $-\infty \leq L < \infty$ and $-\infty \leq X_{-\infty} < \infty$ a.e.;

(ii) If $L \neq -\infty$, then the submartingale (X_n) is uniformly integrable, and if we put $\mathcal{F}_{-\infty} = \cap_n \mathcal{F}_n$, then X_n converges a.e. and in L^1 to $X_{-\infty}$, and $\{ X_n, \mathcal{F}_n,\ n =- \infty, \ldots, -2, -1 \}$ is a submartingale.

(iii) If $\{ X_n, \mathcal{F}_n,\ n = \ldots, -2, -1 \}$ is a martingale, it is uniformly integrable and convergent a.e. and in L^1; moreover, its limit satisfies $X_{-\infty} = E\{ X_{-1} \mid \mathcal{F}_{-\infty} \}$.

Proof. Let $\nu_{-N}[a, b]$ be the number of upcrossings of $[a, b]$ by the $X_{-N}, \ldots, X_{-2}, X_{-1}$. Note that the upcrossing inequality only depends on the expectation of the last element of the submartingale; the first elements don't enter it. Thus we have

$$E\{ \nu_{-N}[a, b] \} \leq \frac{E\{ |X_{-1}| \} + |a|}{b - a} < \infty,$$

independent of N. Next, notice that adding a term to the sequence won't decrease the number of upcrossings, so $\nu_{-N-1}[a, b] \geq \nu_{-N}[a, b]$. Thus by the Monotone Convergence Theorem, $\nu_{-\infty}[a, b] \equiv \lim_{N\to\infty} \nu_{-N}[a, b]$ exists and satisfies

$$E\{\, \nu_{-\infty}[a,b] \,\} \le \frac{E\{\, |X_{-1}| \,\} + |a|}{b-a} < \infty$$

Then the number of upcrossings of $[a,b]$ by the entire sequence (X_n) is a.e. finite, and consequently, with probability one, the number of upcrossings of each interval $[a,b]$ is finite simultaneously for all $a<b$. This implies that $\liminf_{n\to-\infty} X_n = \limsup_{n\to-\infty} X_n \equiv X_{-\infty}$ as in the Martingale Convergence Theorem.

Let M be a real number and consider $X_n \vee M$. As $X_n \longrightarrow X_{-\infty}$, $X_n \vee M \longrightarrow X_{-\infty} \vee M$ a.e. Moreover, $(X_n \vee M)$ is a submartingale which is bounded below by M and has a last element, namely $X_{-1} \vee M$. Thus it is uniformly integrable by Corollary 2.30, so it converges in L^1 and its limit $X_{-\infty} \vee M$ is integrable and consequently finite. Thus $X_{-\infty} < \infty$ a.e., regardless of whether or not L is finite.

Now suppose $L > -\infty$. First note that (X_n) is bounded in L^1. Indeed, $L \le E\{\, X_n^+ \,\} - E\{\, X_n^- \,\}$, so $E\{\, |X_n| \,\} = E\{\, X_n^+ \,\} + E\{\, X_n^- \,\} \le 2E\{\, X_n^+ \,\} - L$. As (X_n^+) is a submartingale, this is bounded by $2E\{\, X_{-1}^+ \,\} - L$, which shows that (X_n) is L^1-bounded. Thus for $N > 0$

$$P\{\, |X_n| > N \,\} \le \frac{1}{N}(2E\{\, X_{-1}^+ \,\} - L)\,. \tag{23}$$

Choose $\varepsilon > 0$ and a negative integer k such that $E\{\, X_k \,\} - L < \varepsilon/2$. Then choose $\delta > 0$ so that $P\{\, \Lambda \,\} < \delta \Longrightarrow \int_\Lambda |X_k|\, dP < \varepsilon/2$. Consider

$$\int_{\{\, |X_n| \ge N \,\}} |X_n|\, dP\,.$$

Any finite family is uniformly integrable, so we can choose N large enough so that this is less that ε for $n = k, k+1, \ldots, -1$. On the other hand, for $n \le k$, this is

$$\le \int_{\{\, |X_n| \ge N \,\}} X_n\, dP - \int_{\{\, X_n \le -N \,\}} X_n\, dP\,.$$

Apply the submartingale inequality to the first term and rewrite the second. This is then

$$\begin{aligned} &\le \int_{\{\, X_n \ge N \,\}} X_k\, dP - E\{\, X_n \,\} + \int_{\{\, X_n > -N \,\}} X_n\, dP \\ &\le \int_{\{\, X_n \ge N \,\}} X_k\, dP - E\{\, X_n \,\} + E\{\, X_k \,\} - \int_{\{\, X_n \le -N \,\}} X_k\, dP\,. \end{aligned}$$

But $E\{\, X_k \,\} - E\{\, X_n \,\} \le E\{\, X_k \,\} - L \le \varepsilon/2$, so this is

$$\le \frac{\varepsilon}{2} + \int_{\{\, |X_n| \ge N \,\}} |X_k|\, dP \le \varepsilon\,,$$

which proves that $\{\, X_n,\ n \le 0 \,\}$ is uniformly integrable.

Now the submartingale converges a.e. (by (i)) and in L^1 (by uniform integrability.) Clearly, $n < m \Longrightarrow X_n \in \mathcal{F}_m$, so $\lim_{n\to-\infty} X_n$ is $\mathcal{F}_m$-measurable for all m, and hence is measurable with respect to $\mathcal{F}_{-\infty} = \cap_m \mathcal{F}_m$.

Suppose $m < n$ and $\Lambda \subset \mathcal{F}_{-m}$. Then $\int_\Lambda X_m\, dP \le \int_\Lambda X_n\, dP$. Let $m \to -\infty$. We can go to the limit under the integral to get

$$\int_\Lambda X_{-\infty}\, dP \le \int_\Lambda X_n\, dP\,.$$

This is true for all $\Lambda \in \mathcal{F}_m \supset \mathcal{F}_{-\infty}$, and in particular, it is true for all $\Lambda \in \mathcal{F}_{-\infty}$. Thus $X_{-\infty} \le E\{\, X_n \mid \mathcal{F}_{-\infty} \,\}$, as claimed.

Finally, if $\{\, X_n, \mathcal{F}_n,\ n = \ldots, -2, -1 \,\}$ is a martingale, it has a last element and is therefore uniformly integrable. Since it is both a sub- and supermartingale, by (ii), $X_{-\infty} = E\{\, X_n \mid \mathcal{F}_{-\infty} \,\}$, proving (iii). ♣

It follows immediately that:

Corollary 2.35 *A martingale with a last element is uniformly integrable. A submartingale or supermartingale with both a first and last element is uniformly integrable.*

Remark 2.36 *Notice that a backward submartingale always has a (possibly infinite) limit. One might ask if the same is true for a forward martingale or submartingale. The answer is "no". Indeed, the gamblers fortune in the (unrestricted) gambler's ruin problem is a martingale, and it has its limsup equal to infinity, its liminf equal to negative infinity. Thus there really is an asymmetry between backward and forward martingale limits.*

A useful consequence of these theorems is Paul Lévy's limit theorem on conditional expectations. Conditional expectations with respect to sequences of sigma fields which either increase or decrease monotonically, converge to the best possible limits. Before proving this we will need a result from measure theory.

Definition 2.12 *A class $\mathcal{G}$ of sets is a* **field** *(or* **countably additive class** *or* **algebra***) if it contains the empty set, and is closed under complementation and finite unions.*

The difference between a field and a sigma field is that the sigma field is closed under *countable* unions, not just finite unions. If $(\mathcal{F}_n)$ is a filtration, then $\cup_n \mathcal{F}_n$ is a field, but not necessarily a sigma-field. The following extension lemma is a well-known result in measure theory. We shall accept it without proof.

Lemma 2.37 *Suppose P and Q are finite measures on $(\Omega, \mathcal{F})$ and that $\mathcal{G} \subset \mathcal{F}$ is a field. If $P\{\,\Lambda\,\} = Q\{\,\Lambda\,\}$ for all $\Lambda \in \mathcal{G}$, then $P\{\,\Lambda\,\} = Q\{\,\Lambda\,\}$ for all $\Lambda \in \sigma\{\,\mathcal{G}\,\}$.*

Corollary 2.38 *Let X be an integrable random variable. Let $\mathcal{G}_n$ be a sequence of sub-sigma fields of $\mathcal{F}$.*

(i) If $\mathcal{G}_n \subset \mathcal{G}_{n+1}$ for all n, then

$$\lim_{n\to\infty} E\{ X \mid \mathcal{G}_n \} = E\{ X \mid \sigma(\cup_n \mathcal{G}_n) \} . \tag{24}$$

(ii) If $\mathcal{G}_n \supset \mathcal{G}_{n+1}$ for all n, then

$$\lim_{n\to\infty} E\{ X \mid \mathcal{G}_n \} = E\{ X \mid \cap_n \mathcal{G}_n \} . \tag{25}$$

Proof. (i). $X_n \equiv E\{ X \mid \mathcal{G}_n \}$ is a martingale. It is uniformly integrable by Theorem 2.29, so it converges to a limit X_∞, and $X_n = E\{ X_\infty \mid \mathcal{G}_n \}$ by Theorem 2.33. We must identify X_∞ as the conditional expectation. First, X_∞ is measurable with respect to $\sigma\{ \cup_n \mathcal{G}_n \}$. If $\Lambda \in \mathcal{G}_n$, $\int_\Lambda X_n \, dP = \int_\Lambda X_\infty \, dP$ and, as $X_n = E\{ X \mid \mathcal{G}_n \}$, we also have $\int_\Lambda X_n \, dP = \int_\Lambda X \, dP$. Thus

$$\int_\Lambda X_\infty \, dP = \int_\Lambda X \, dP \tag{26}$$

for all $\Lambda \in \mathcal{G}_n$, and hence for all $\Lambda \in \cup_n \mathcal{G}_n$. Now both sides of (26) define finite measures, so by Lemma 2.37, there is equality for all $\Lambda \in \sigma\{ \cup_n \mathcal{G}_n \}$, proving (i).

The proof of (ii) is direct from the convergence theorem. Set $X_{-n} = E\{ X \mid \mathcal{G}_n \}$, $\mathcal{F}_{-n} = \mathcal{G}_n$. Then $\{ X_n, \mathcal{F}_n, \ n = \ldots, -2, -1 \}$ is a backwards martingale, so by Theorem 2.34 (iii), it has the limit $X_{-\infty} = E\{ X_{-1} \mid \mathcal{F}_{-\infty} \}$. But $\mathcal{F}_{-\infty} = \cap_n \mathcal{G}_n$, so $X_{-\infty} = E\{ E\{ X \mid \mathcal{G}_1 \} \mid \cap_n \mathcal{G}_n \} = E\{ X \mid \cap_n \mathcal{G}_n \}$. ♣

2.14 Conditional Expectations and the Radon-Nikodym Theorem

Martingales have numerous applications. We will try to limit ourselves to those we will actually need in the sequel. One, the Radon-Nikodym Theorem, is quite important, not only because it is important in its own right (which it is), but because we have been working without a good existence theorem for conditional expectations. We know they exist when the conditioning field is finite, or more generally, generated by a partition, but we don't know about general sigma fields. And (now it can be revealed) we have been playing a little game. If the basic filtrations are generated by partitions, everything we have done is rigorous. In the general case, what we have proved is equally rigorous, *providing the conditional expectations exist.* Effectively, we have been operating under the assumption that the sigma fields $\mathcal{F}_n$ of the filtrations are generated by partitions. We need a theorem guaranteeing the existence of conditional expectations to close the circle. Once we have this, we can figuratively lift ourself by our own bootstraps: our previous results and proofs will then go through *without* any assumptions on the filtrations. (One might object that we have dealt with large filtrations in the limit: even if all the $\mathcal{F}_n$ are finite, the filtration $\mathcal{F}_\infty \equiv \sigma\{\cup_n \mathcal{F}_n\}$ need not be generated by a partition. This is true, but in fact we have never needed conditional expectations with respect to this sigma field.) We will get the general existence from the Radon-Nikodym theorem: a conditional expectation is really a Radon-Nikodym derivative. We will prove it using martingale convergence theorems, but of course—and this is important if we are to play our game fairly—we only will use finite filtrations in the proof.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let Q be another measure on $(\Omega, \mathcal{F})$. We say Q is **absolutely continuous** with respect to P, and we write $Q \ll P$, if $\Lambda \in \mathcal{F}$ and $P\{\Lambda\} = 0 \Longrightarrow Q\{\Lambda\} = 0$. We say the sigma field $\mathcal{F}$ is **separable** if there exists a sequence Λ_n of sets such that $\mathcal{F} = \sigma\{\Lambda_n,\ n = 1, 2, \dots\}$.

Lemma 2.39 *Suppose $Q \ll P$ is a finite measure. Then, given $\varepsilon > 0$ there exists $\delta > 0$ such that $\Lambda \in \mathcal{F}$ and $P\{\Lambda\} \leq \delta \Longrightarrow Q\{\Lambda\} < \varepsilon$.*

Proof. If not, there exists a sequence Λ_n of events and $\varepsilon > 0$ such that $P\{\Lambda\} \leq 2^{-n}$ while $Q\{\Lambda_n\} > \varepsilon$. Set $\Gamma_n = \cup_{j>n}\Lambda_j$. Then $P\{\Gamma_n\} \leq 2^{-n}$ so $P\{\cap_n \Gamma_n\} = 0$ by the countable additivity of P. But as Q is finite and Γ_n decreasing, $Q\{\cap_n \Gamma_n\} = \lim Q\{\cap_n \Gamma_n\} \geq \varepsilon$, which contradicts the absolute continuity of Q ♣

Theorem 2.40 *(Radon-Nikodym Theorem) Let $(\Omega, \mathcal{F}, P)$ be a probability space and let Q be a finite measure such that $Q \ll P$. Then there exists an integrable random variable X such that for all $\Lambda \in \mathcal{F}$, $Q\{\Lambda\} = \int_\Lambda X\, dP$.*

Remark 2.41 *We call X the* **Radon-Nikodym derivative** *of Q with respect to P, and write*

$$X = \frac{dQ}{dP}\,.$$

The restriction to finite Q is not necessary: Q can be sigma-finite.

Proof. We will only prove this for the case where $\mathcal{F}$ is separable, generated by a sequence of sets $\Lambda_n,\ n = 1, 2, \dots$. Once this is proved, it can be extended to the general case by a measure-theoretical argument which is more standard than it is interesting, so we will skip it.

Define sigma fields $\mathcal{F}_n$ by $\mathcal{F}_n = \sigma\{\Lambda_1, \dots, \Lambda_n\}$. Then $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, so the $(\mathcal{F}_n)$ form a filtration. Moreover, each $\mathcal{F}_n$ is generated by a finite partition: the elements of the partition are sets of the form $G_1 \cap \ldots \cap G_n$, where each G_i is either Λ_i or Λ_i^c; all sets in $\mathcal{F}_n$ are finite unions of these sets. Let $\Gamma_j^n,\ j = 1, 2, \dots, N_n$ be the partition generating $\mathcal{F}_n$. Define X_n by

$$X_n(\omega) = \begin{cases} \frac{Q\{\Gamma_j^n\}}{P\{\Gamma_j^n\}} & \text{if } \omega \in \Gamma_j^n \text{ and } P\{\Gamma_j^n\} > 0 \\ 0 & \text{if } \omega \in \Gamma_j^n \text{ and } P\{\Gamma_j^n\} = 0\,. \end{cases} \tag{27}$$

X_n is clearly positive and $\mathcal{F}_n$-measurable. Notice that $\int_{\Gamma_j^n} X_n\, dP = Q\{\, \Gamma_j^n \,\}$. This is immediate from (27) if $P\{\, \Gamma_j^n \,\} \neq 0$, and it also holds if $P\{\, \Gamma_j^n \,\} = 0$, since then the fact that $Q \ll P$ implies that $Q\{\, \Gamma_j^n \,\} = 0$ too, while $X \equiv 0$ on Γ_j^n so the integral also vanishes. It follows that $\int_\Lambda X_n\, dP = Q\{\, \Lambda \,\}$ for all $\Lambda \in \mathcal{F}_n$, since all such sets are finite unions of the Γ_j^n.

We claim that $\{\, X_n, \mathcal{F}_n,\ n = 0, 1, 2, \dots \}$ is a martingale. Indeed, if $m < n$, and $\Lambda \in \mathcal{F}_m \subset \mathcal{F}_n$, then $\int_\Lambda X_m\, dP = Q\{\, \Lambda \,\} = \int_\Lambda X_n\, dP$ which shows that $X_m = E\{\, X_n \mid \mathcal{F}_m \,\}$, so (X_n) is a martingale as claimed. Furthermore, it is uniformly integrable. Indeed, $\int_{\{\, X_n \geq N \,\}} X_n\, dP = Q\{\, X_n \geq N \,\}$, while $P\{\, X_n \geq N \,\} \leq E\{\, X_n \,\}/N = Q\{\, \Omega \,\}/N$. As Q is finite, this tends to zero, independent of n. Let $\varepsilon > 0$ and choose $\delta > 0$ by the lemma so that $P\{\, \Lambda \,\} < \delta \Longrightarrow Q\{\, \Lambda \,\} < \varepsilon$. Choose N large enough so that $Q\{\, \Omega \,\}/N < \delta$. It follows that $\int_{\{\, X_n \geq N \,\}} X_n\, dP < \varepsilon$ for all n, and (X_n) is uniformly integrable.

Thus the X_n converge a.e. and in L^1 to an integrable limit X_∞, and for all n, $X_n = E\{\, X_\infty \mid \mathcal{F}_n \,\}$. For any $\Lambda \in \mathcal{F}_n$, $\int_\Lambda X_\infty\, dP = \int_\Lambda X_n\, dP = Q\{\, \Lambda \,\}$, which implies

$$\int_\Lambda X_\infty\, dP = Q\{\, \Lambda \,\} \tag{28}$$

for all $\Lambda \in \cup_n \mathcal{F}_n$. But both sides of (28) define finite measures on $\mathcal{F}$, so by Lemma 2.37, (28) holds for all $\Lambda \in \sigma\{\, \cup_n \mathcal{F}_n \,\} = \mathcal{F}$, which completes the proof. ♣

Now we can show the existence of the conditional expectation of an arbitrary integrable random variable with respect to an arbitrary sub-sigma field of $\mathcal{F}$.

Corollary 2.42 *(Existence of Conditional Expectations) Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X be an integrable random variable and let $\mathcal{G} \subset \mathcal{F}$ be a sigma field. Then the conditional expectation $E\{\, X \mid \mathcal{G} \,\}$ exists.*

Proof. It is enough to prove this for the case where $X \geq 0$, since otherwise we can consider the positive and negative parts X^+ and X^- separately. Define a measure Q on $(\Omega, \mathcal{F}, P)$ by $Q\{\, \Lambda \,\} = \int_\Lambda X\, dP$, for $\Lambda \in \mathcal{F}$. Note that, as $\mathcal{G} \subset \mathcal{F}$, Q is also a measure on the probability space $(\Omega, \mathcal{G}, P)$. If $P\{\, \Lambda \,\} = 0$, then $Q\{\, \Lambda \,\} = \int_\Lambda X\, dP = 0$, so $Q \ll P$. Thus there exists a Radon-Nikodym density Z of Q with respect to P *on the probability space* $(\Omega, \mathcal{G}, P)$. (We can apply the theorem on *any* probability space, so we choose $(\Omega, \mathcal{G}, P)$ rather than $(\Omega, \mathcal{F}, P)$.) Then $Z = E\{\, X \mid \mathcal{G} \,\}$. Indeed, Z is integrable and, being a r.v. on $(\Omega, \mathcal{G}, P)$, it is $\mathcal{G}$-measurable. Moreover, if $\Lambda \in \mathcal{G}$, then $\int_\Lambda Z\, dP = Q\{\, \Lambda \,\} = \int_\Lambda X\, dP$, which is exactly what we needed to show. ♣

2.15 Two Other Applications

Martingales offer a number of striking applications to various parts of probability. We will concentrate on applications to mathematical finance below, so we will give just two here: the Borel Zero-One Law, and Kolmogorov's Strong Law of Large Numbers. Both of these have other proofs. Both fall easily to martingale methods.

Let $X_1, X_2, \dots$ be a sequence of independent random variables. We want to look at the limiting behavior of this sequence in a general way. Let $\mathcal{F}_n^\star = \sigma\{X_n, X_{n+1}, \dots\}$ be the sigma field generated by the sequence after n. Note that $\mathcal{F}_n^\star \supset \mathcal{F}_{n+1}^\star$, so the $\mathcal{F}_n^\star$ decrease with n. Let $\mathcal{F}^\star = \cap_n \mathcal{F}_n^\star$ be the limiting field. This is called the *tail field* or *remote field* and events in it are called *tail* or *remote* events. It contains information about the limiting behavior of the sequence. For instance, quantities like $\limsup X_n$, $\liminf X_n$, and $\limsup (X_1 + \ldots + X_n)/n$ are all $\mathcal{F}^\star$-measurable, and events like $\{$ *the sequence* $(X_1 + \ldots + X_n)/n$ *converges* $\}$ and $\{ X_n = 0$ *for infinitely many* $n \}$ are in $\mathcal{F}^\star$. It turns out that the structure of $\mathcal{F}^\star$ is remarkable simple—trivial, even—according to the Borel Zero-One Law.

Theorem 2.43 *(Borel Zero-One Law) If* $\Lambda \in \mathcal{F}^\star$ *then* $P\{\Lambda\} = 0$ *or* 1.

Thus, $\mathcal{F}^\star$ is trivial in the sense that any $\mathcal{F}^\star$-measurable random variable is a.e. constant, and any event in $\mathcal{F}^\star$ is either sure to happen, or sure not to happen; no tail event has probability one half. Thus for example, a series such as $\sum_1^\infty X_n$ will either converge with probability one, or diverge with probability one. Which of the two occurs depends on the distributions of the X_n, but we know beforehand that one of the two alternatives happens.

PROOF. (Warning: this proof is short, but it turns in rather tight logical circles. Don't get twisted up!) Let $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$. Note that $\mathcal{F}_n$ and $\mathcal{F}_{n+1}^\star$ are independent, hence $\mathcal{F}_n$ and $\mathcal{F}^\star$ are independent for all n (for $\mathcal{F}^\star \subset \mathcal{F}_{n+1}^\star$.) Let $\Lambda \in \mathcal{F}^\star$, and let $Y_n = E\{ I_\Lambda \mid \mathcal{F}_n \}$. Then Λ is independent of $\mathcal{F}_n$, so that $Y_n = E\{I_\Lambda\} = P\{\Lambda\}$ a.e. On the other hand, (Y_n) is a uniformly integrable martingale, which converges to $E\{ I_\Lambda \mid \sigma\{\cup_n \mathcal{F}_n\}\}$ by Theorem 2.33. But $\sigma\{\cup_n \mathcal{F}_n\} = \sigma\{X_1, X_2, \dots\}$, which in particular contains Λ, so that $E\{ I_\Lambda \mid \sigma\{\cup_n \mathcal{F}_n\}\} = I_\Lambda$. This gives us two expressions for $\lim Y_n$. Putting them together, we see

$$P\{\Lambda\} = I_\Lambda \text{ a.e.}$$

The left-hand side is a real number; the right-hand side is a r.v. which, being an indicator function, can take on only the values zero and one. Thus $P\{\Lambda\}$ has to equal either zero or one! ♣

The Law of Large numbers states that if $X_1, X_2, \dots$ is a sequence of iid random variables with common expectation m, then $\lim_{n\to\infty}(1/n)(X_1 + \ldots + X_n) = m$. The X_j have to be integrable in order to even state the theorem, but most versions of the Law of Large Numbers require extra integrability hypotheses. In fact, though, if the random variables are i.i.s., only first moments are needed. Kolmogorov's theorem is decidedly non-trivial, but we will be able to prove it fairly easily once we have made one observation.

Theorem 2.44 *(Kolmogorov) Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, with $E\{X_1\} = m$. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j = m \ a.e.$$

Proof. Let

$$S_{-n} = \frac{1}{n}\sum_{j=1}^{n} X_j\,,$$

and let $\mathcal{F}_{-n} = \sigma\{S_{-n}, S_{-n-1}, \ldots\}$. Notice that $\mathcal{F}_{-n} \supset \mathcal{F}_{-n-1}$, so that $(\mathcal{F}_n)$ is a filtration. Moreover, we claim that $\{S_n, \mathcal{F}_n,\ n = \ldots, -2, -1\}$ is a martingale. Indeed, notice that $\mathcal{F}_{-n} = \sigma\{S_{-n}, X_{n+1}, X_{n+2}, \ldots\} = \sigma\{(X_1 + \ldots X_n), X_{n+1}, X_{n+2} \ldots\}$. If $j \le n$, let $Z_j \equiv E\{X_j \mid \mathcal{F}_{-n}\} = E\{X_j \mid (X_1 + \ldots X_n), X_{n+1}, X_{n+2} \ldots\}$. Since $X_1, \ldots X_n$ is independent of $X_{n+1}, X_{n+2} \ldots$, this is equal to $E\{X_j \mid X_1 + \ldots X_n\}$. But now, $Z_1 = \ldots = Z_n$ by symmetry—the X_i are identically distributed—and $X_1 + \ldots + X_n = E\{X_1 + \ldots + X_n \mid X_1 + \ldots + X_n\} = Z_1 + \ldots + Z_n = nZ_1$. Thus $j \le n \Longrightarrow E\{X_j \mid \mathcal{F}_{-n}\} = S_{-n}$.

Thus

$$\begin{aligned} E\{S_{-n+1} \mid \mathcal{F}_{-n}\} &= \frac{1}{n-1} E\{X_1 + \ldots + X_{n-1} \mid X_1 + \ldots + X_n\} \\ &= \frac{1}{n-1}(n-1)S_{-n} \\ &= S_{-n}. \end{aligned}$$

Thus (S_n) is a *backward* martingale. (This is surprising in itself!) By Theorem 2.34 it is uniformly integrable and converges a.e. and in L^1. Moreover, its limit $S_{-\infty}$ can be added on as the initial element of the martingale, so $E\{S_{-\infty}\} = E\{S_{-1}\} = E\{X_1\} = m$. It is not hard to see that in fact $S_{-\infty}$ is measurable with respect to the tail field. (Prove it!) Thus it is a.e. constant by the Borel Zero-One Law, so $S_{-\infty} = m$ a.e. ♣

2.16 Exercises

1° Prove that if $X_1, X_2, \ldots$ are iid and integrable, that $\lim_{n\to\infty} 1/n) \sum_{j=1}^{n} X_j$ is measurable with respect to the tail field.

2° Show that every sequence of integrable random variables is the sum of a submartingale and a supermartingale.

3° Give an example of a martingale X_n with the property that $X_n \to -\infty$ a.s. as $n \to \infty$. (Hint: consider sums of independent but not identically-distributed random variables.)

4° Let (X_n) be a positive supermartingale. Show that for a.e. ω, if $X_k(\omega) = 0$, then $X_n(\omega) = 0$ for all $n > k$.

5° (Krickeberg decomposition) Let $\{ X_n, \mathcal{F}_n,\ n \geq 0 \}$ be a sub martingale with $E\{ |X_n| \}$ bounded. Put $X_n^+ = \max(X_n, 0)$. Show there exists a positive martingale $\{ Y_n, \mathcal{F}_n,\ n \geq 0 \}$ such that $X_n^+ \leq Y_n$ and $E\{ Y_n \} = \lim E\{ X_n+ \}$. Deduce from this that every L^1-bounded martingale is the difference of two positive martingales. (Hint: try $Y_n = \lim_{k\to\infty} E\{ X_k^+ \mid \mathcal{F}_n \}$, and show the limit exists a.e.)

6° Suppose, in the set-up of the proof of the Radon-Nikodym theorem, that instead of being absolutely continuous with respect to P, Q is *singular* with respect to P, i.e. there exist a set $\Lambda \in \mathcal{F}$ such that $P\{ \Lambda \} = 0$ and $Q\{ \Lambda^c \} = 0$. Show that in that case the process X_n defined in the proof converges a.e. to zero. (Note that X_n may not be a martingale, but only a supermartingale. The calculation in Example 6 of Section 2.1 may help.)

7° Consider the martingale Z_n of Example 6 of Section 2.1. (Assume that the densities p and q never vanish, so that it is in fact a martingale, not a supermartingale.) Suppose that the statistician decides to use the following criterion to decide between the two hypotheses (H1) and (H2): choose two numbers, $0 < a < 1 < b$. Continue sampling as long as $a < Z_n < b$. Let $T = \inf\{ n : Z_n \leq a \text{ or } Z_n \geq b \}$. Assume that $T < \infty$ a.e. and that, at time T, $Z_t = a$ or $Z_T = b$. (That is, we ignore the possible overshoot.) If $Z_T = a$, the statistician will decide that (H2) is correct, and if $Z_T = b$, that (H1) is correct. How can one choose a and b to make both $P\{ \text{statistician makes error} \mid \text{(H1) is correct} \}$ and $P\{ \text{statistician makes error} \mid \text{(H2) is correct} \}$ equal to a fixed number $0 < \alpha < 1/2$?

8° A family $\{ X_\alpha,\ \alpha \in I \}$ of random variables is uniformly integrable if and only if there exists a function ϕ on $[0, \infty)$ which is increasing, satisfies $\lim_{x\to\infty} \phi(x)/x = \infty$, and $\sup_\alpha E\{ \phi(X_\alpha) \} < \infty$.