CONDITIONAL EXPECTATION AND MARTINGALES

1. INTRODUCTION

Martingales play a role in stochastic processes roughly similar to that played by *conserved quantities* in dynamical systems. Unlike a conserved quantity in dynamics, which remains constant in time, a martingale's value can change; however, its *expectation* remains constant in time. More important, the expectation of a martingale is unaffected by *optional sampling*. In fact, this can be used as a provisional definition: A discrete-time*martingale* is a sequence $\{X_n\}_{n\geq 0}$ of integrable real (or complex) random variables with the property that for every bounded stopping time τ, the *Optional Sampling Formula*

$$EX_\tau = EX_0 \tag{1}$$

is valid.

We have seen the Optional Sampling Formula before, in various guises. In particular, the Wald Identities I,II, and III are all instances of (1). Let $\xi_0, \xi_1, \dots$ be independent, identically distributed random variables, and let $S_n = \xi_1 + \xi_2 + \cdots \xi_n$ be the nth partial sum. Denote by μ, σ^2, and $\varphi(\theta)$ the mean, variance, and moment generating function of ξ_1, that is,

$$\begin{aligned} \mu &= E\xi_1, \\ \sigma^2 &= E(\xi_1 - \mu)^2, \qquad \text{and} \\ \varphi(\theta) &= E\exp\{\theta\xi_1\}. \end{aligned}$$

Corresponding to each of these scalar quantities is a *martingale*:

$$M_n := S_n - n\mu, \tag{2}$$

$$V_n := (S_n - n\mu)^2 - n\sigma^2, \qquad \text{and} \tag{3}$$

$$Z_n(\theta) := \exp\{\theta S_n\}/\varphi(\theta)^n. \tag{4}$$

Observe that there is a separate martingale $Z_n(\theta)$ for every real value of θ such that $\varphi(\theta) < \infty$.

The Optional Sampling Formula could be taken as the definition of a martingale, but usually isn't. The standard approach, which we will follow, uses the notion of *conditional expectation*.

2. CONDITIONAL EXPECTATION

2.1. **Definition of Conditional Expectation.** For random variables defined on discrete probability spaces, conditional expectation can be defined in an elementary manner: In particular, the conditional expectation of a discrete random variable X given the value y of another discrete random variable Y may be defined by

$$E(X\,|\,Y = y) = \sum_x xP(X = x\,|\,Y = y), \tag{5}$$

where the sum is over the set of all possible values x of X. Note that this expression depends on the value y. For discrete random variables that take values in finite sets there are no difficulties regarding possible divergence of the sum, nor is there any difficulty regarding the meaning of the conditional probability $P(X = x \mid Y = y)$.

For continuous random variables, or, worse, random variables that are neither discrete nor have probability densities, the definition (5) is problematic. There are two main difficulties: (a) If X is not discrete, then the sum must be replaced by an integral of some sort; and (b) If Y is not discrete then it is no longer clear how to define the condtional probabilities $P(X = x \mid Y = y)$. Fortunately, there is an alternative way of defining conditional expectation that works in both the discrete and the indiscrete cases, and additionally allows for conditioning not only on the value of a single random variable, but for conditioning simultaneously on the values of finitely or even countably many random variables or random vectors:

Definition 1. Let X be a real-valued random variable such that either $E|X| < \infty$ or $X \geq 0$, and let Y be a random variable, random vector, or other random object taking values in a measurable space $(\mathscr{Y}, \mathscr{H})$.The *conditional expectation* $E(X \mid Y)$ is the essentially unique measurable real-valued function of Y such that for every bounded, measurable, real-valued function $g(Y)$,

$$EXg(Y) = E(E(X \mid Y)g(Y)). \tag{6}$$

Definition 2. More generally[1], if X is defined on a probability space $(\Omega, \mathscr{F}, P)$ and $\mathscr{G} \subset \mathscr{F}$ is a $\sigma-$algebra contained in $\mathscr{F}$, then $E(X|\mathscr{G})$ is the essentially unique $\mathscr{G}$-measurable random variable such that

$$E(XZ) = E(E(X|\mathscr{G})Z) \tag{7}$$

for every bounded, $\mathscr{G}-$measurable random variable Z.

It is by no means clear *a priori* that such functions $E(X \mid Y)$ or $E(X|\mathscr{G})$ should always exist, nor is it obvious that they should be unique. In fact, the *existence* of such a function is an important theorem of measure theory, the *Radon-Nikodym* theorem, which I will take as known. The *uniqueness* of the function is not difficult to prove:

Proof of Uniqueness. Suppose that there are distinct functions $h_1(y)$ and $h_2(y)$ such that, for every bounded function $g(y)$,

$$EXg(Y) = Eh_1(Y)g(Y) \qquad \text{and}$$
$$EXg(Y) = Eh_2(Y)g(Y).$$

Then by the linearity of ordinary expectation (taking the difference of the two equations) it must be the case that for every bounded function $g(y)$,

$$0 = E(h_1(Y) - h_2(Y))g(Y);$$

in particular, this equation must hold for the function $g(y)$ that is 1 if $h_1(y) > h_2(y)$ and 0 otherwise. But this implies that $P\{h_1(Y) > h_2(Y)\} = 0$. A similar argument shows that $P\{h_2(Y) = h_1(Y)\} = 0$. It follows that $P\{h_1(Y) = h_2(Y)\} = 1$. □

[1]Nearly all of the $\sigma-$algebras that one encounters in typical situations are generated, either explicitly or implicitly, by random variables Y valued in a Polish space (a complete, separable metric space). In particular, the $\sigma-$algebras that arise naturally in the study of stochastic processes are usually generated by *sample paths*. If Y generates $\mathscr{G}$, that is, if $\mathscr{G}$ consists of all events of the form $\{Y \in B\}$ where B is a Borel subset of the range of Y, then every $\mathscr{G}-$measurable random variable is a measurable function of Y, and conversely (Exercise!). Hence, Definitions 1 and 2 coincide.

2.2. **Equivalence of the Naive and Modern Definitions.** It is not difficult to show that the naive definition (6) coincides with the modern Definition 1 when the random variable X and the random vector $Y = (Y_1, Y_2, \dots, Y_m)$ are discrete and assume only finitely many possible values with positive probability. Define

$$h(y) = \sum_x xP(X = x \,|\, Y = y) = \sum_x x \frac{P\{X = x \text{ and } Y = y\}}{P\{Y = y\}}$$

for those values of y such that $P\{Y = y\} > 0$. To show that $E(X \,|\, Y) = h(Y)$, it suffices, by Definition 1, to show that, for any bounded function $g(y)$,

$$EXg(Y) = Eh(Y)g(Y).$$

But

$$\begin{aligned} EXg(Y) &= \sum_y \sum_x xg(y)P\{X = x \text{ and } Y = y\} \\ &= \sum_y g(y)P\{Y = y\} \sum_x xP(X = x \,|\, Y = y) \\ &= \sum_y g(y)P\{Y = y\}h(y) \\ &= Eh(Y)g(Y). \end{aligned}$$

□

2.3. **Properties of Conditional Expectation.** The raw definition given above can be clumsy to work with directly. In this section we present a short list of important rules for manipulating and calculating conditional expectations. The bottom line will be that, in many important respects, conditional expectations behave like ordinary expectations, with random quantities that are functions of the conditioning random variable being treated as constants.[2]

Let Y be a random variable, vector, or object valued in a measurable space, and let X be an *integrable* random variable (that is, a random variable with $E|X| < \infty$).

Summary: Basic Properties of Conditional Expectation.

(1) **Definition:** $EXg(Y) = EE(X \,|\, Y)g(Y)$ for all bounded measure functions $g(y)$.
(2) **Linearity:** $E(aU + bV|Y) = aE(U|Y) + bE(V|Y)$ for all scalars $a, b \in \mathbb{R}$.
(3) **Positivity:** If $X \geq 0$ then $E(X|Y) \geq 0$.
(4) **Stability:** If X is a (measurable) function of Y, then $E(XZ|Y) = XE(Z|Y)$.
(5) **Independence Law:** If X is independent of Y then $E(X|Y) = EX$ is constant a.s.
(6) **Tower Property:** If Z is a function of Y then $E(E(X \,|\, Y) \,|\, Z) = E(X \,|\, Z)$.
(7) **Expectation Law:** $E(E(X|Y)) = EX$.
(8) **Constants:** For any scalar a, $E(a|Y) = a$.

[2]Later we'll prove a theorem to the effect that conditional expectations *are* ordinary expectations in a certain sense. For those of you who already know the terminology, this theorem is the assertion that every real random variable Y — or more generally, every random variable valued in a Polish space — has a *regular conditional distribution* given any σ–algebra $\mathcal{G}$.

(9) **Jensen Inequalities:** If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $E|X| < \infty$ then

$$E(\varphi(X)) \geq \varphi(EX) \text{ and}$$
$$E(\varphi(X)|Y) \geq \varphi(E(X|Y).$$

In all of these statements, the relations = and ≤ are meant to hold *almost surely*. Similar statements can be formulated for conditional expectation $E(X \mid \mathcal{G})$ on a $\sigma-$algebra. Also, properties (3)–(7) extend to nonnegative random variables X with infinite expectation. All of the properties can be proved easily, using only Definition 1 and elementary properties of *ordinary* expectation. To give an idea of how these arguments go, we shall outline the proofs of the Linearity, Positivity, and Independence properties below. You should try to check the Stability and Tower Properties yourself. The Jensen inequality is of a somewhat different character, but it is not difficult to prove – see below.

Note: The definition (1) requires only that the equation $EXg(Y) = EE(X \mid Y)g(Y)$ be valid for *bounded* functions g. A standard argument from measure theory then implies that it holds for all functions such that the product $Xg(Y)$ has finite first moment. Similarly, Property (4) holds provided the product has finite first moment.

Proof of the Positivity Property. The idea is to exploit the defining property (6) of conditional expectation. First, suppose that $X \geq 0$. Define B to be the set of possible values of Y for which the conditional expectation $E(X \mid Y) < 0$, so that the event $\{E(X \mid Y) < 0\}$ coincides with the event $\{Y \in B\}$. Then by equation (6),

$$EX\mathbf{1}_B(Y) = E(E(X \mid Y)\mathbf{1}_B(y)).$$

Since $X \geq 0$, the left side of this equality is nonnegative; but by definition of B, the right side is negative unless $P\{Y \in B\} = 0$. It follows that $P\{Y \in B\} = 0$, that is, $E(X \mid Y) \geq 0$ with probability one. □

Proof of the Linearity Property. Since each of the conditional expectations $E(U \mid Y)$ and $E(V \mid Y)$ is a function of Y, so is the linear combination $aE(U \mid Y) + bE(V \mid Y)$. Thus, by Definition 1, to show that this linear combination is the conditional expectation $E(aU + bV \mid Y)$, it suffices to show that it satisfies equation (6), that is, that for every bounded nonnegative function $g(Y)$,

(8) $$E(aU + bV)g(Y) = E(aE(U \mid Y) + bE(V \mid Y))g(Y).$$

But equation (6) holds for $X = U$ and for $X = V$:

$$EUg(Y) = EE(U \mid Y)g(Y),$$
$$EVg(Y) = EE(V \mid Y)g(Y).$$

Multiplying these equations by a and b, respectively, and then adding gives (8), because the unconditional expectation operator is linear. □

Proof of the Independence Property. This relies on the fact that if U and V are independent, integrable random variables whose product UV is also integrable, then $E(UV) = EUEV$. Now suppose that X is independent of Y, and let $g(Y)$ be any bounded (measurable) function of Y. Then $EXg(Y) = EXEg(Y) = E(EX)g(Y)$. Since any constant, and in particular EX, is trivially a function of Y, Definition 1 implies that EX must be the conditional expectation $EW(X \mid Y)$. □

Proof of the Jensen Inequalities. One of the basic properties of convex functions is that every point on the graph of a convex function φ has a *support line*: that is, for every argument $x_* \in \mathbb{R}$ there is a linear function $y_{x_*}(x) = ax + b$ such that

$$\varphi(x_*) = y_{x_*}(x_*) \quad \text{and}$$
$$\varphi(x) \geq y_{x_*}(x) \quad \text{for all } x \in \mathbb{R}.$$

Let X be a random variable such that $E|X| < \infty$, so that the expectation EX is well-defined and finite. Let $y_{EX}(x) = ax + b$ be the support line to the convex function at the point $(EX, \varphi(EX))$. Then by definition of a support line, $y_{EX}(EX) = \varphi(EX)$; also, $y_{EX}(X) \leq \varphi(X)$, and so

$$Ey_{EX}(X) \leq E\varphi(X).$$

But because $y_{EX}(x) = ax + b$ is a linear function of x,

$$Ey_{EX}(X) = y_{EX}(EX) = \varphi(EX).$$

This proves the Jensen inequality for ordinary expectation. The proof for conditional expectation is similar. For any value of Y, let $y_{E(X|Y)}(x)$ be the support line at the point $(E(X|Y), \varphi(E(X|Y)))$. Then $y_{E(X|Y)}(E(X|Y)) = \varphi(E(X|Y))$, and for every value of X, $y_{E(X|Y)}(X) \leq \varphi(X)$. Consequently, by the linearity and positivity properties of conditional expectation,

$$\begin{aligned}\varphi(E(X|Y)) &= y_{E(X|Y)}(E(X|Y)) \\ &= E(y_{E(X|Y)}(X)|Y) \\ &\leq E(\varphi(X)|Y).\end{aligned}$$

□

2.4. **Convergence theorems for conditional expectation.** Just as for ordinary expectations, there are versions of Fatou's lemma and the monotone and dominated convergence theorems.

Monotone Convergence Theorem . *Let X_n be a nondecreasing sequence of nonnegative random variables on a probability space $(\Omega, \mathscr{F}, P)$, and let $X = \lim_{n\to\infty} X_n$. Then for any random variable (or vector) Y,*

(9) $$E(X_n \,|\, Y) \uparrow E(X \,|\, Y).$$

Fatou's Lemma . *Let X_n be a sequence of nonnegative random variables on a probability space $(\Omega, \mathscr{F}, P)$, and let $X = \liminf_{n\to\infty} X_n$. Then for any random variable (or vector) Y,*

(10) $$E(X \,|\, Y) \leq \liminf E(X_n \,|\, Y).$$

Dominated Convergence Theorem . *Let X_n be a sequence of real-valued random variables on a probability space $(\Omega, \mathscr{F}, P)$ such that for some integrable random variable Y and all $n \geq 1$,*

(11) $$|X_n| \leq Y.$$

Then for any random variable (or vector) Y,

(12) $$\lim_{n\to\infty} E(X_n \,|\, Y) = E(X \,|\, \mathscr{G}) \quad \textit{and} \quad \lim_{n\to\infty} E(|X_n - X| \,|\, Y) = 0$$

As in Properties 1–9 above, the limiting equalites and inequalities in these statements hold *almost surely.* The proofs are easy, given the corresponding theorems for *ordinary* expectations; I'll give the proof for the Monotone Convergence Theorem and leave the other two, which are easier, as exercises.

Proof of the Monotone Convergence Theorem. By the Positivity and Linearity properties of conditional expectation,

$$E(X_n \mid Y) \le E(X_{n+1} \mid Y) \le E(X \mid Y)$$

for every n. Consequently, the limit $V := \lim_{n\to\infty} \uparrow E(X_n \mid Y)$ exists with probability one, and $V \le E(X \mid Y)$. Moreover, since each conditional expectation is Y-measurable, so is V. Set $B = \{V < E(X \mid Y)\}$; we must show that $P(B) = 0$. Now B is $Y-$measurable, so by definition of conditional expectation,

$$E(X\mathbf{1}_B) = E(E(X \mid Y)\mathbf{1}_B) \quad \text{and} \quad E(X_n\mathbf{1}_B) = E(E(X_n \mid Y)\mathbf{1}_B).$$

But the Monotone Convergence Theorem for ordinary expectation implies that

$$\begin{aligned} EX\mathbf{1}_B &= \lim_{n\to\infty} EX_n\mathbf{1}_B \quad \text{and} \\ EV\mathbf{1}_B &= \lim_{n\to\infty} EE(X_n \mid Y)\mathbf{1}_B, \end{aligned}$$

so $EX\mathbf{1}_B = EV\mathbf{1}_B$. Since $V < X$ on B, this implies that $P(B) = 0$. □

3. Discrete-Time Martingales

3.1. **Definition of a Martingale.** Let $\{\mathscr{F}_n\}_{n\ge 0}$ be an increasing sequence of $\sigma-$algebras in a probability space $(\Omega, \mathscr{F}, P)$. Such sequences will be called *filtrations.* Let $X_0, X_1, \dots$ be an *adapted* sequence of *integrable* real-valued random variables, that is, a sequence with the property that for each n the random variable X_n is measurable relative to $\mathscr{F}_n$ and such that $E|X_n| < \infty$. In a typical application in the study of a Markov chain Y_n, the $\sigma-$algebra $\mathscr{F}_n$ might be the smallest $\sigma-$algebra such that the random vector $(Y_0, Y_1, \dots, Y_n)$ is measurable relative to $\mathscr{F}_n$, and the random variables X_n numerical functions of the states Y_n. The sequence $X_0, X_1, \dots$ is said to be a *martingale* relative to the filtration $\{\mathscr{F}_n\}_{n\ge 0}$ if it is adapted, and if for every n,

$$\boxed{E(X_{n+1} \mid \mathscr{F}_n) = X_n.} \tag{13}$$

Similarly, it is said to be a *supermartingale* (respectively, *submartingale*) if for every n,

$$E(X_{n+1} \mid \mathscr{F}_n) \le (\ge) X_n. \tag{14}$$

Observe that any martingale is automatically both a submartingale and a supermartingale.

3.2. **Martingales and Martingale Difference Sequences.** The most basic examples of martingales are sums of independent, mean zero random variables. Let $Y_0, Y_1, \dots$ be a sequence of independent, identically distributed random variables such that $EY_n = 0$. Then the sequence of partial sums

$$X_n = \sum_{j=1}^{n} Y_j \tag{15}$$

is a martingale relative to the sequence $0, Y_1, Y_2, \dots$ (that is, relative to the natural filtration generated by the variables Y_n). This is easily verified, using the linearity and stability properties and the independence law for conditional expectation:

$$\begin{aligned} E(X_{n+1} \mid \mathscr{F}_n) &= E(X_n + Y_{n+1} \mid \mathscr{F}_n) \\ &= E(X_n \mid \mathscr{F}_n) + E(Y_{n+1} \mid \mathscr{F}_n) \\ &= X_n + EY_{n+1} \\ &= X_n. \end{aligned}$$

The importance of martingales in modern probability stems at least in part from the fact that most of the essential properties of sums of independent, identically distributed random variables are inherited (with minor modification) by martingales: As you will learn, there are versions of the SLLN, the Central Limit Theorem, the Wald indentities, and the Chebyshev, Markov, and Kolmogorov inequalities for martingales. To get some appreciation of why this might be so, consider the decomposition of a martingale $\{X_n\}$ as a partial sum process:

$$X_n = X_0 + \sum_{j=1}^{n} \xi_j \quad \text{where} \quad \xi_j = X_j - X_{j-1}. \tag{16}$$

Proposition 1. *The* martingale difference sequence $\{\xi_n\}$ *has the following properties: (a) the random variable* ξ_n *is a function of* $\mathscr{F}_n$*; and (b) for every* $n \geq 0$*,*

$$E(\xi_{n+1} \mid \mathscr{F}_n) = 0. \tag{17}$$

Proof. Assertion (b) is a three-line calculation using the properties of conditional expectation and the definition of a martingale. □

Corollary 1. *Let* $\{X_n\}$ *be a martingale relative to* $\{Y_n\}$*, with martingale difference sequence* $\{\xi_n\}$*. Then for every* $n \geq 0$*,*

$$EX_n = EX_0. \tag{18}$$

Moreover, if $X_0 = 0$ *and* $EX_n^2 < \infty$ *then the random variables* ξ_j *are uncorrelated, and so*

$$EX_n^2 = \sum_{j=1}^{n} E\xi_j^2. \tag{19}$$

Proof. The first property follows almost trivially from Proposition 1 and the Expectation Law for conditional expectation, as these together imply that $E\xi_n = 0$ for each n. Summing and using the linearity of ordinary expectation, one obtains (18).

The second property is only slightly more difficult: First, observe that each of the terms ξ_j has finite variance, because it is the difference of two random variables with finite second moments. (That $EX_j^2 < \infty$ follows from the hypothesis that $EX_n^2 < \infty$, together with the Tower Property.) Consequently, all of the products $\xi_i \xi_j$ have finite first moments, by the Cauchy-Schwartz inequality. Next, if $j \leq k \leq n$ then ξ_j is measurable relative to $\mathscr{F}_j$; hence, by Properties (1), (4), (6), and (7) of conditional expectation, if $j \le k \le n$ then

$$\begin{aligned} E\xi_j\xi_{k+1} &= EE(\xi_j\xi_{k+1} \mid Y_1, Y_2, \dots, Y_k) \\ &= E\xi_j E\xi_{k+1} \mid Y_1, Y_2, \dots, Y_k) \\ &= E(\xi_j \cdot 0) = 0. \end{aligned}$$

The variance of X_n may now be calculated in exactly the same manner as for sums of independent random variables with mean zero:

$$\begin{aligned} EX_n^2 &= E(\sum_{j=1}^n \xi_j)^2 \\ &= E\sum_{j=1}^n \sum_{k=1}^n \xi_j\xi_k \\ &= \sum_{j=1}^n \sum_{k=1}^n E\xi_j\xi_k \\ &= \sum_{j=1}^n E\xi_j^2 + 2\sum\sum_{j<k} E\xi_j\xi_k \\ &= \sum_{j=1}^n E\xi_j^2 + 0. \end{aligned}$$

□

3.3. **Some Examples of Martingales.**

3.3.1. *Paul Lévy's Martingales.* Let X be any integrable random variable. Then the sequence X_n defined by $X_n = E(X|\mathscr{F}_n)$ is a martingale, by the Tower Property of conditional expectation.

3.3.2. *Random Walk Martingales.* Let $Y_0, Y_1, \dots$ be a sequence of independent, identically distributed random variables such that $EY_n = 0$. Then the sequence $X_n = \sum_{j=1}^n Y_j$ is a martingale, as we have seen.

3.3.3. *Second Moment Martingales.* Once again let $Y_0, Y_1, \dots$ be a sequence of independent, identically distributed random variables such that $EY_n = 0$ and $EY_n^2 = \sigma^2 < \infty$. Then the sequence

$$\left(\sum_{j=1}^n Y_j\right)^2 - \sigma^2 n \tag{20}$$

is a martingale (again relative to the sequence $0, Y_1, Y_2, \dots$). This is also easy to check.

3.3.4. *Likelihood Ratio Martingales: Bernoulli Case.* Let $X_0, X_1, \dots$ be a sequence of independent, identically distributed Bernoulli-p random variables, and let $S_n = \sum_{j=1}^n X_j$. Note that S_n has the binomial-(n,p) distribution. Define

$$Z_n = \left(\frac{q}{p}\right)^{2S_n - n}. \tag{21}$$

Then $Z_0, Z_1, \dots$ is a martingale relative to the usual sequence. Once again, this is easy to check. The martingale $\{Z_n\}_{n\geq 0}$ is quite useful in certain random walk problems, as we have already seen.

3.3.5. *Likelihood Ratio Martingales in General.* Let $X_0, X_1, \dots$ be independent, identically distributed random variables whose moment generating function $\varphi(\theta) = Ee^{\theta X_i}$ is finite for some value $\theta \neq 0$. Define

$$Z_n = Z_n(\theta) = \prod_{j=1}^n \frac{e^{\theta X_j}}{\varphi(\theta)} = \frac{e^{\theta S_n}}{\varphi(\theta)^n}. \tag{22}$$

Then Z_n is a martingale. (It is called a *likelihood ratio* martingale because the random variable Z_n is the likelihood ratio dP_θ / dP_0 based on the sample $X_1, X_2, \dots, X_n$ for probability measures P_θ and P_0 in a certain exponential family.)

3.3.6. *Galton-Watson Martingales.* Let $Z_0 = 1, Z_1, Z_2, \dots$ be a Galton-Watson process whose offspring distribution has mean $\mu > 0$. Denote by $\varphi(s) = Es^{Z_1}$ the probability generating function of the offspring distribution, and by ζ the smallest nonnegative root of the equation $\varphi(\zeta) = \zeta$.

Proposition 2. *Each of the following is a nonnegative martingale:*

$$M_n := Z_n / \mu^n; \qquad \textit{and}$$
$$W_n := \zeta^{Z_n}.$$

Proof. Homework. □

3.3.7. *Polya Urn.* In the traditional Polya urn model, an urn is seeded with $R_0 = r \geq 1$ red balls and $B_0 = b \geq 1$ black balls. At each step $n = 1, 2, \dots$, a ball is drawn at random from the urn and then returned along with a new ball of the same color. Let R_n and B_n be the numbers of red and black balls after n steps, and let $\Theta_n = R_{n/(R_n + B_n)}$ be the fraction of red balls. Then Θ_n is a martingale relatve to the natural filtration.

3.3.8. *Harmonic Functions and Markov Chains.* Yes, surely enough, martingales also arise in connection with Markov chains; in fact, one of Doob's motivations in inventing them was to connect the world of potential theory for Markov processes with the classical theory of sums of independent random variables.[3] Let $Y_0, Y_0, Y_1, \dots$ be a Markov chain on a denumerable state space $\mathscr{Y}$ with transition probability matrix $\mathbb{P}$. A real-valued function $h : \mathscr{Y} \to \mathbb{R}$ is called *harmonic* for the transition probability matrix $\mathbb{P}$ if

$$\mathbb{P}h = h, \tag{23}$$

[3]See his 800-page book *Classical Potential Theory and its Probabilistic Counterpart* for more on this.

equivalently, if for every $x \in \mathscr{Y}$,

$$h(x) = \sum_{y \in \mathscr{Y}} p(x,y)h(y) = E^x h(Y_1). \tag{24}$$

Here E^x denotes the expectation corresponding to the probability measure P^x under which $P^x\{Y_0 = x\} = 1$. Notice the similarity between equation (24) and the equation for the stationary distribution – one is just the *transpose* of the other.

Proposition 3. *If h is harmonic for the transition probability matrix $\mathbb{P}$ then for every starting state $x \in \mathscr{Y}$ the sequence $h(Y_n)$ is a martingale under the probability measure P^x.*

Proof. This is once again nothing more than a routine calculation. The key is the Markov property, which allows us to rewrite any conditional expectation on $Y_0, \mathscr{F}_n$ as a conditional expectation on Y_n. Thus,

$$\begin{aligned} E(h(Y_{n+1}) \mid Y_0, \mathscr{F}_n) &= E(h(Y_{n+1}) \mid Y_n) \\ &= \sum_{y \in \mathscr{Y}} h(y)p(Y_n, y) \\ &= h(Y_n). \end{aligned}$$

□

3.3.9. *Submartingales from Martingales.* Let $\{X_n\}_{n\geq 0}$ be a martingale relative to the sequence $Y_0, Y_1, \dots$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function such that $E\varphi(X_n) < \infty$ for each $n \geq 0$. Then the sequence $\{Z_n\}_{n\geq 0}$ defined by

$$Z_n = \varphi(X_n) \tag{25}$$

is a *sub*martingale. This is a consequence of the Jensen inequality and the martingale property of $\{X_n\}_{n\geq 0}$:

$$\begin{aligned} E(Z_{n+1} | Y_0, Y_1, \dots, Y_n) &= E(\varphi(X_{n+1}) | Y_0, Y_1, \dots, Y_n) \\ &\geq \varphi(E(X_{n+1} | Y_0, Y_1, \dots, Y_n) \\ &= \varphi(X_n) = Z_n \end{aligned}$$

Useful special cases: (a) $\varphi(x) = x^2$, and (b) $\varphi(x) = \exp\{\theta x\}$.

4. Martingale and Submartingale Transforms

According to the Merriam-Webster Collegiate Dictionary, a *martingale* is

> any of several systems of betting in which a player increases the stake usually by doubling each time a bet is lost.

The use of the term in the theory of probability derives from the connection with *fair games* or *fair bets*; and the importance of the theoretical construct in the world of finance also derives from the connection with fair bets. Seen in this light, the notion of a *martingale transform*, which we are about to introduce, becomes most natural. Informally, a martingale transform is nothing more than a system of placing bets on a fair game.

4.1. **Martingale Transforms.** A formal definition of a martingale transform requires two auxiliary notions: *martingale differences* and *predictable sequences*. Let $X_0, X_1, \ldots$ be a martingale relative to another sequence $Y_0, Y_1, \ldots$ (or to a filtration $\{\mathscr{F}_n\}_{n\geq 0}$). For $n = 1, 2, \ldots$, define

$$\xi_n = X_n - X_{n-1}; \tag{26}$$

to be the martingale difference sequence associated with the martingale X_n.

A *predictable sequence* $Z_1, Z_2, \ldots$ relative to the filtration $\mathscr{F}_n$ is a sequence of random variables such that for each $n = 1, 2, \ldots$ the random variable Z_n is measurable relative to $\mathscr{F}_{n-1}$. In gambling (and financial) contexts, Z_n might represent the size (say, in dollars) of a bet paced on the nth play of a game, while ξ_n represents the (random) payoff of the nth play per dollar bet. The requirement that the sequence Z_n be predictable in such contexts is merely an assertion that the gambler not be clairvoyant.

Definition 3. Let $X_0, X_1, \ldots$ be a martingale relative to $\mathscr{F}_n$ and let $\xi_n = X_n - X_{n-1}$ be the associated martingale difference sequence. Let $Z_0, Z_1, \ldots$ be a predictable sequence. Then the *martingale transform* $\{(Z \cdot X)_n\}_{n\geq 0}$ is defined by

$$(Z \cdot X)_n = Z_0 X_0 + \sum_{k=1}^{n} Z_k \xi_k. \tag{27}$$

Example: The St. Petersburg Game. In this game, a referee tosses a fair coin repeatedly, with results $\xi_1, \xi_2, \ldots$, where $\xi_n = +1$ if the nth toss is a Head and $\xi_n = -1$ if the nth toss is a Tail. Before each toss, a gambler is allowed to place a wager of size W_n (in roubles) on the outcome of the next toss. The size of the wager W_n may depend on the observed tosses $\xi_1, \xi_2, \ldots, \xi_{n-1}$, but not on ξ_n (or on any of the future tosses); thus, the sequence $\{W_n\}_{n\geq 1}$ is predictable relative to $\{\xi_n\}_{n\geq 1}$. If the nth toss is a Head, the gambler nets $+W_n$, but if the nth toss is a Tail, the gambler loses W_n. Thus, the net winnings S_n after n tosses is the martingale transform

$$S_n = (W \cdot X)_n = \sum_{k=1}^{n} W_k \xi_k,$$

where $X_n = \xi_1 + \xi_2 + \cdots + \xi_n$. □

The most important fact about martingale transforms is that they are martingales in their own right, as the next proposition asserts:

Proposition 4. *Assume that the predictable sequence $\{Z_n\}_{n\geq 0}$ consists of* bounded *random variables. Then the martingale transform $\{(Z \cdot X)_n\}_{n\geq 0}$ is itself a martingale relative to $\{Y_n\}_{n\geq 0}$.*

Proof. This is a simple exercise in the use of the linearity and stability properties of conditional expectation:

$$\begin{aligned} E((Z \cdot X)_{n+1} \,|\, \mathscr{F}_n) &= (Z \cdot X)_n + E(Z_{n+1}\xi_{n+1} \,|\, \mathscr{F}_n) \\ &= (Z \cdot X)_n + Z_{n+1} E(\xi_{n+1} \,|\, \mathscr{F}_n) \\ &= (Z \cdot X)_n, \end{aligned}$$

the last equation because $\{\xi_n\}_{n\geq 1}$ is a martingale difference sequence relative to $\{Y_n\}_{n\geq 0}$. □

4.2. **Submartingale Transforms.** Submartingales and supermartingales may also be transformed, using equation (27), but the resulting sequences will not necessarily be sub- or super-martingales unless the predictable sequence $\{Z_n\}_{n\geq 0}$ consists of *nonnegative* random variables.

Definition 4. Let $X_0, X_1, \dots$ be a sub- (respectively, super-) martingale relative to $\mathscr{F}_n$ and let $\xi_n = X_n - X_{n-1}$ be the associated sub- (super-) martingale difference sequence. Let $Z_0, Z_1, \dots$ be a predictable sequence consisting of bounded *nonnegative* random variables. Then the *submartingale transform* (respectively, *supermartingale transform*) $\{(Z \cdot X)_n\}_{n\geq 0}$ is defined by

$$(Z \cdot X)_n = Z_0 X_0 + \sum_{k=1}^{n} Z_k \xi_k. \tag{28}$$

Proposition 5. *If the terms Z_n of the predictable sequence are nonnegative and bounded, and if $\{X_n\}_{n\geq 0}$ is a submartingale, then the submartingale transform $(Z \cdot X)_n$ is also a submartingale. Moreover, if, for each $n \geq 0$,*

$$0 \leq Z_n \leq 1, \tag{29}$$

then

$$E(Z \cdot X)_n \leq E X_n. \tag{30}$$

Proof. To show that $(Z \cdot X)_n$ is a submartingale, it suffices to verify that the differences $Z_k \xi_k$ constitute a submartingale difference sequence. Since Z_k is a predictable sequence, the differences $Z_k \xi_k$ are adapted to $\{Y_k\}_{k\geq 0}$, and

$$E(Z_k \xi_k \,|\, \mathscr{F}_{k-1}) = Z_k E(\xi_k \,|\, \mathscr{F}_{k-1}).$$

Since ξ_k is a submartingale difference sequence, $E(\xi_k \,|\, \mathscr{F}_{k-1}) \geq 0$; and therefore, since $0 \leq Z_k \leq 1$,

$$0 \leq E(Z_k \xi_k \,|\, \mathscr{F}_{k-1}) \leq E(\xi_k \,|\, \mathscr{F}_{k-1}).$$

Consequently, $Z_k \xi_k$ is a submartingale difference sequence. Moreover, by taking expectations in the last inequalities, we have

$$E(Z_k \xi_k) \leq E \xi_k,$$

which implies (30). □

There is a similar result for supermartingales:

Proposition 6. *If $\{X_n\}_{n\geq 0}$ is a supermartingale, and if the terms Z_n of the predictable sequence are nonnegative and bounded, then $\{(Z \cdot X)_n\}_{n\geq 0}$ is a supermartingale; and if inequality (29) holds for each $n \geq 0$ then*

$$E(Z \cdot X)_n \geq E X_n. \tag{31}$$

5. Optional Stopping

The cornerstone of martingale theory is Doob's *Optional Sampling Theorem*. This states, roughly, that "stopping" a martingale at a random time τ does not alter the expected "payoff", provided the decision about when to stop is based solely on information available up to τ. Such random times are called *stopping times*.[4]

[4]In some of the older literature, they are called *Markov times* or *optional times*.

Definition 5. A *stopping time* relative to a filtration $\{\mathscr{F}_n\}_{n\geq 0}$ is a nonnegative integer-valued random variable τ such that for each n the event $\{\tau = n\} \in \mathscr{F}_n$.

Theorem 1. *(Optional Sampling Theorem) Let $\{X_n\}_{n\in\mathbb{Z}_+}$ be a martingale, submartingale , or supermartingale relative to a sequence $\{Y_n\}_{n\geq 0}$, and let τ be a stopping time. Then for any $n \in \mathbb{N}$,*

$$\begin{aligned} EX_{\tau\wedge n} &= EX_0 && \textit{(martingales)} \\ EX_{\tau\wedge n} &\leq EX_0 && \textit{(supermartingales)} \\ EX_{\tau\wedge n} &\leq EX_n && \textit{(submartingales)} \end{aligned}$$

Proof. The easiest proof is based on the fact that martingale (respectively, submartingale, supermartingale) transforms are martingales (respectively, submartingales, supermartingales). The connection between transforms and the Optional Sampling Theorem is that the sequence $\{X_{\tau\wedge n}\}_{n\geq 0}$ may be represented as a transform of the sequence $\{X_n\}_{n\geq 0}$:

$$X_{\tau\wedge n} = (Z \cdot X)_n \tag{32}$$

where

$$Z_n = \begin{cases} 1 & \text{if } \tau \geq n, \qquad \text{and} \\ 0 & \text{if } \tau < n. \end{cases} \tag{33}$$

The equation (32) is easy to verify. Note that $Z_0 = 1$, since τ is nonnegative; consequently,

$$\begin{aligned} (Z \cdot X)_n &= X_0 + \sum_{j=1}^{n} Z_j (X_j - X_{j-1}) \\ &= X_0 + \sum_{j=1}^{\tau\wedge n} (X_j - X_{j-1}) \\ &= X_{\tau\wedge n}, \end{aligned}$$

since the last sum telescopes.

In order that the sequence $\{(Z\cdot X)_n\}_{n\geq 0}$ be a martingale transform (respectively, sub- or supermartingale transform) it must be the case that the sequence $\{Z_n\}_{n\geq 0}$ is predictable. This is where the assumption that τ is a stopping time enters: Since τ is a stopping time, for each fixed m the event that $\tau = m$ depends only on $\mathscr{F}_m$. Hence, the event

$$\{\tau \geq n\} = \left(\cup_{m=0}^{n-1}\{\tau = m\}\right)^c$$

depends only on $\mathscr{F}_{n-1}$. But this event is the same as the event that $Z_n = 1$; this proves that Z_n is measurable relative to $\mathscr{F}_{n-1}$, and so the sequence $\{Z_n\}_{n\geq 0}$ is predictable.

The first two assertions of the Optional Sampling Theorem now follow easily from Propositions 4 and 5, in view of the "Conservation of Expectation" properties of martingales and supermartingales. For instance, if $\{X_n\}_{n\geq 0})$ is a martingale, then since martingale transforms are themselves martingales, and since expectation is "preserved" for martingales,

$$EX_{\tau\wedge n} = E(Z \cdot X)_n = E(Z \cdot X)_0 = EX_0.$$

A similar argument establishes the corresponding result for supermartingales. Finally, the last assertion, regarding the case where $\{X_n\}_{n\geq 0}$ is a submartingale, follows from inequality (30), since the terms Z_n of the predictable sequence are between 0 and 1. □

6. Maximal Inequalities

The Optional Sampling Theorem has immediate implications concerning the pathwise behavior of martingales, submartingales, and supermartingales. The most elementary of these concern the maxima of the sample paths, and so are called *maximal inequalities.*

Proposition 7. *Let $\{X_n\}_{n\geq 0}$ be a sub- or super-martingale relative to $\{Y_n\}_{n\geq 0}$, and for each $n \geq 0$ define*

$$M_n = \max_{0\leq m\leq n} X_m, \qquad \text{and} \tag{34}$$

$$M_\infty = \sup_{0\leq m<\infty} X_m = \lim_{n\to\infty} M_n \tag{35}$$

Then for any scalar $\alpha > 0$ and any $n \geq 1$,

$$P\{M_n \geq \alpha\} \leq E(X_n \vee 0)/\alpha \qquad \text{if } \{X_n\}_{n\geq 0} \text{ is a submartingale, and} \tag{36}$$

$$P\{M_\infty \geq \alpha\} \leq EX_0/\alpha \qquad \text{if } \{X_n\}_{n\geq 0} \text{ is a nonnegative supermartingale.} \tag{37}$$

Proof. Assume first that $\{X_n\}_{n\geq 0}$ is a submartingale. Without loss of generality, we may assume that each $X_n \geq 0$, because if not we may replace the original submartingale X_n by the larger submartingale $X_n \vee 0$. Define τ to be the smallest $n \geq 0$ such that $X_n \geq \alpha$, or $+\infty$ is there is no such n. Then for any nonrandom $n \geq 0$, the truncation $\tau \wedge n$ is a stopping time and so, by the Optional Sampling Theorem,

$$EX_{\tau\wedge n} \leq EX_n.$$

But because the random variables X_m are nonnegative, and because $X_{\tau\wedge n} \geq \alpha$ on the event that $\tau \leq n$,

$$\begin{aligned} EX_{\tau\wedge n} &\geq EX_{\tau\wedge n}\mathbf{1}\{\tau \leq n\} \\ &\geq E\alpha\mathbf{1}\{\tau \leq n\} \\ &= \alpha P\{\tau \leq n\}. \end{aligned}$$

This proves the inequality (36).

The proof of inequality (37) is similar, but needs an additional limiting argument. First, for any finite $n \geq 0$, an argument parallel to that of the preceding paragraph shows that

$$P\{M_n \geq \alpha\} \leq EX_0/\alpha.$$

Now the random variables M_n are nondecreasing in n, and converge up to M_∞, so for any $\epsilon > 0$, the event that $M_\infty \geq \alpha$ is contained in the event that $M_n \geq \alpha - \epsilon$ for some n. But by the last displayed inequality and the monotone convergence theorem, the probability of this is no larger than $EX_0/(\alpha - \epsilon)$. Since $\epsilon > 0$ may be taken arbitrarily small, inequality (37) follows. □

Example: The St. Petersburg Game, Revisited. In Dostoevsky's novel *The Gambler*, the hero (?) is faced with the task of winning a certain amount of money at the roulette table, starting with a fixed stake strictly less than the amount he wishes to take home from the casino. What strategy for allocating his stake will maximize his chance of reaching his objective? Here we will consider an analogous problem for the somewhat simpler St. Petersburg game described earlier. Suppose that the gambler starts with 100 roubles, and that he wishes to maximize his chance of leaving with 200 roubles. There is a very simple strategy that gives him a .5 probability of reaching his objective: stake all 100 roubles on the first coin toss, and quit the game after one play. Is there a strategy that will give the gambler more than a .5 probability of reaching the objective?

The answer is *NO*, and we may prove this by appealing to the Maximal Inequality (37) for supermartingales. Let $\{W_n\}_{n\geq 0}$ be any predictable sequence (recall that, for a non-clairvoyant bettor, the sequence of wagers must be predictable). Then the gambler's fortune after n plays equals

$$F_n = 100 + \sum_{k=1}^{n} W_k \xi_k,$$

where ξ_n is the martingale difference sequence of ± 1 valued random variables recording whether the coin tosses are Heads or Tails. By Proposition 4, the sequence F_n is a martingale. Since each $F_n \geq 0$, the Maximal Inequality for nonnegative supermartingales applies, and we conclude that

$$P\{\sup_{n\geq 0} F_n \geq 200\} \leq EX_0/200 = 1/2.$$

Exercise: What is an optimal strategy for maximizing the chance of coming away with at least 300 roubles?

7. Upcrossings Inequalities

The Maximal Inequalities limit the extent to which a submartingale or supermartingale may deviate from it initial value. In particular, if X_n is a submartingale that is bounded in L^1 then the maximal inequality implies that $\sup X_n < \infty$ with probability one. The *Upcrossings Inequalities*, which we shall discuss next, limit the extent to which a submartingale or supermartingale may fluctuate around its initial value.

Fix a sequence X_n of real random variables. For any fixed constants $\alpha < \beta$, define the *upcrossings count* $N_n((\alpha,\beta])$ to be the number of times that the finite sequence $X_0, X_1, X_2, \dots, X_n$ crosses from the interval $(-\infty, \alpha]$ to the interval (β, ∞). Equivalently, define stopping times

$$\begin{aligned}
&\sigma_0 := \min\{n \geq 0 : X_n \leq \alpha\} && \tau_1 := \min\{n \geq \sigma_0 : X_n > \beta\}; \\
&\sigma_1 := \min\{n \geq \tau_1 : X_n \leq \alpha\} && \tau_2 := \min\{n \geq \sigma_1 : X_n > \beta\}; \\
&\quad \cdots \\
&\sigma_m := \min\{n \geq \tau_m : X_n \leq \alpha\} && \tau_{m+1} := \min\{n \geq \sigma_m : X_n > \beta\};
\end{aligned} \tag{38}$$

then

$$N_n((\alpha,\beta]) = \max\{m : \tau_m \leq n\}.$$

Proposition 8. *Let X_n be a submartingale relative to Y_n. Then for any scalars $\alpha < \beta$ and all nonnegative integers m, n,*

$$(\beta - \alpha) EN_n((\alpha,\beta]) \leq E(X_n \vee 0) + |\alpha|. \tag{39}$$

Consequently, if $\sup EX_n < \infty$, then $EN_\infty((\alpha,\beta]) < \infty$, and so the sequence $\{X_n\}_{n\geq 0}$ makes only finitely many upcrossings of any interval $(\alpha,\beta]$.

Proof. The trick is similar to that used in the proof of the Maximal Inequalities: define an appropriate submartingale transform and then use Proposition 5. We begin by making two simplifications: First, it is enough to consider the special case $\alpha = 0$, because the general case may be reduced to this by replacing the original submartingale X_n by the submartingale $X'_n = X_n - \alpha$ (Note that this changes the expectation in the inequality by at most $|\alpha|$.) Second, if $\alpha = 0$, then it is enough to consider the special case where X_n is a *nonnegative* submartingale, because if X_n is not nonnegative, it may replaced by $X''_n = X_n \vee 0$, as this does not change the number of upcrossings of $(0, \beta]$ or the value of $E(X_n \vee 0)$.

Thus, assume that $\alpha = 0$ and that $X_n \geq 0$. Use the stopping times σ_m, τ_m defined above (with $\alpha = 0$) to define a predictable sequence Z_n as follows:

$$
\begin{aligned}
Z_n &= 0 \qquad && \text{if } n \leq \sigma_0; \\
Z_n &= 1 \qquad && \text{if } \sigma_m < n \leq \tau_m; \\
Z_n &= 0 \qquad && \text{if } \tau_m < n \leq \sigma_m.
\end{aligned}
$$

(EXERCISE: Verify that this is a predictable sequence.) This sequence has alternating blocks of 0s and 1s (not necessarily all finite). Over any *complete* finite block of 1s, the increments ξ_k must sum to at least β, because at the beginning of a block (some time σ_m) the value of X is 0, and at the end (the next τ_m), the value is back above β. Furthermore, over any *incomplete* block of 1s (even one which will never terminate!), the sum of the increments ξ_k will be ≥ 0, because at the beginning σ_m of the block the value $X_{\sigma_m} = 0$ and X_n never goes below 0. Hence,

$$
\beta N_n(0, \beta] \leq \sum_{i=1}^{n} Z_i \xi_i = (Z \cdot X)_n.
$$

Therefore, by Proposition 5,

$$
\begin{aligned}
(\beta - \alpha) E N_n(\alpha, \beta] &\leq E(Z \cdot X)_{\tau(M_n)} \\
&\leq E(Z \cdot X)_n \\
&\leq E X_n.
\end{aligned}
$$

□

8. THE MARTINGALE CONVERGENCE THEOREM

8.1. **Pointwise convergence.**

Martingale Convergence Theorem . *Let $\{X_n\}$ be an L^1−bounded submartingale relative to a sequence $\{Y_n\}$, that is, a submartingale such that* $\sup_n E|X_n| < \infty$. *Then with probability one the limit*

$$
\lim_{n \to \infty} X_n := X_\infty \tag{40}
$$

exists, is finite, and has finite first moment.

Proof. By the Upcrossings Inequality, for any interval $(\alpha, \beta]$ with rational endpoints the sequence $\{X_n\}_{n \geq 0}$ can make only finitely many upcrossings of $(\alpha, \beta]$. Equivalently, the probability that $\{X_n\}$ makes infinitely many upcrossings of $(\alpha, \beta]$ is zero. Since there are only countably many intervals $(\alpha, \beta]$ with rational endpoints, and since the union of countably many events of probability zero is an event of probability zero, it follows that with probability one there is no rational interval $(\alpha,\beta]$ such that X_n makes infinitely many upcrossings of $(\alpha,\beta]$.

Now if x_n is a sequence of real numbers that makes only finitely many upcrossings of any rational interval, then x_n must converge to a finite or infinite limit (this is an easy exercise in elementary real analysis). Thus, it follows that with probability one $X_\infty := \lim_{n\to\infty} X_n$ exists (but may be $\pm\infty$). But Fatou's Lemma implies that

$$E|X_\infty| \le \liminf_{n\to\infty} E|X_n| < \infty,$$

and so the limit X_∞ is finite with probability one. □

Corollary 2. *Every nonnegative supermartingale converges almost surely.*

Proof. If X_n is a nonnegative supermartingale, then $-X_n$ is a nonpositive submartingale. Moreover, because $X_n \ge 0$,

$$0 \le E|X_n| = EX_n \le EX_0,$$

the latter because X_n is a supermartingale. Therefore $-X_n$ is an L^1−bounded submartingale, to which the Martingale Convergence Theorem applies. □

8.2. L^1 **convergence and uniform integrability.** The Martingale Convergence Theorem asserts, among other things, that the limit X_∞ has finite first moment. However, it is *not* necessarily the case that $E|X_n - X_\infty| \to 0$. Consider, for example, the martingale X_n that records your fortune at time n when you play the St. Petersburg game with the "double-or-nothing" strategy on every play. At the first time you toss a Tail, you will lose your entire fortune and have 0 forever after. Since this is (almost) certain to happen eventually, $X_n \to 0$ almost surely. But $EX_n = 1 \neq 0$ for every n!

Thus, not every L^1−bounded martingale converges to its pointwise limit in L^1. For which martingales does L^1 convergence occur?

Definition 6. A set of integrable random variables $A = \{X_\lambda\}_{\lambda\in\Lambda}$ is *uniformly integrable* if for every $\delta > 0$ there exists $C_\delta < \infty$ such that for all $X_\lambda \in A$,

$$E|X_\lambda|\mathbf{1}\{|X_\lambda| \ge C_\delta\} \le \delta. \tag{41}$$

Proposition 9. *A set of integrable random variables $A = \{X_\lambda\}_{\lambda\in\Lambda}$ is uniformly integrable if and only if for every $\delta > 0$ there exists $\varepsilon = \varepsilon(\delta) > 0$ such that for any event B satisfying $P(B) < \varepsilon$ and all $X_\lambda \in A$,*

$$E|X_\lambda|\mathbf{1}_B < \delta. \tag{42}$$

Proposition 10. *Any bounded subset of L^p, where $p > 1$, is uniformly integrable.*

These are standard results in measure theory. The proofs are not difficult. Proposition 10 is quite useful, as it provides a simple test for uniformly integrability.

Proposition 11. *Let $\{X_n\}_{n\ge1}$ be a sequence of real random variables such that $\lim X_n = X$ exists almost surely (or in probability). Then $X_n \to X$ in L^1 if and only if the sequence $\{X_n\}_{n\ge1}$ is uniformly integrable. Furthermore, if the collection $\{X_n\}_{n\ge1}$ is uniformly integrable, then for every σ−algebra $\mathscr{G}$,*

$$\lim_{n\to\infty} E(X_n\,|\,\mathscr{G}) = E(X\,|\,\mathscr{G}) \quad \textit{and} \quad \lim_{n\to\infty} E(|X_n - X|\,|\,\mathscr{G}) = 0 \tag{43}$$

Proof. I'll prove the useful direction, that uniform integrability implies L^1-convergence and convergence of conditional expectations. The converse is easier, and is left as an exercise. Assume that $\{X_n\}_{n\geq 1}$ is uniformly integrable; then $\{X_n\}_{n\geq 1}$ is bounded in L^1 (because the inequality (41) implies that the L^1 norms are all bounded by $C_1 + 1$). Hence, by Fatou, the limit $X \in L^1$. It follows (exercise: try using Proposition 9) that the collection $\{|X_n - X|\}_{n\geq 1}$ is uniformly integrable. Let $C_\delta < \infty$ be the uniformity constants for this collection (as in inequality (41)). Fix $\delta > 0$, and set

$$Y_n := |X_n - X| \mathbf{1}\{|X_n - X| \leq C_\delta\}.$$

These random variables are uniformly bounded (by C_δ), and converge to 0 by hypothesis. Consequently, by the dominated convergence theorem, $EY_n \to 0$. Therefore,

$$\limsup E|X_n - X| \leq \delta.$$

Since $\delta > 0$ can be taken arbitrarily small, it follows that the lim sup is actually 0. This proves that $X_n \to X$ in L^1, which in turn implies (by the triangle inequality) that $EX_n \to EX$. The assertions regarding conditional expectations can be proved by similar arguments, using the DCT for conditional expectation. □

Corollary 3. *Let X_n be a uniformly integrable submartingale relative to a filtration $\{\mathscr{F}_n\}_{n\geq 1}$. Then the sequence X_n is bounded in L^1, and therefore has a pointwise limit X; moreover, it converges to its almost sure limit X in L^1. If X_n is a* martingale*, then it is* closed*, in the following sense:*

$$X_n = E(X \,|\, \mathscr{F}_n). \tag{44}$$