CONDITIONAL EXPECTATION AND MARTINGALES

1. Introduction

The stochastic processes known as *martingales* were invented by Joseph Doob a little over fifty years ago at the University of Illinois. Their introduction changed the face of modern probability theory. Martingales now play a central role in probabilistic potential theory, in stochastic calculus, and (as some of you will learn) in mathematical finance. The purpose of these notes is to give an introduction to martingale theory, and to illustrate how martingales arise in the study of Markov chains and random walks.

2. The Wald Identities for Random Walks

A *martingale* is a discrete-time random process $\{X_n\}_{n\geq 0}$ with the property that for every bounded stopping time τ, the *Optional Sampling Formula*

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

is valid. This may be taken as the definition of a martingale, but usually isn't. The traditional definition, together with a proof that this definition implies the Optional Sampling Formula, will be given later.

Some of the simplest and most important martingales occur in connection with sums of independent, identically distributed random variables (random walks). We shall begin by investigating several of these, with the object of showing how the Optional Sampling formula of martingale theory can be used to do concrete calculations of interesting quantities connected with the behavior of the random walk. Later in the notes we will show that the Optional Sampling Formula is valid for the martingales that we introduce here.

2.1. Martingales associated with random walks.

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 process $Z_n(\theta)$ for every real value of θ such that $\varphi(\theta) < \infty$. It is easily checked that the expectations of these random variables are $EM_n = 0$, $EV_n = 0$, and $EZ_n(\theta) = 1$ for all n, and so the Optional Sampling Formula holds at least for *constant* stopping times. The Optional Sampling Formulas for the martingales M_n, V_n, and $Z_n(\theta)$, with arbitrary stopping times, are known as the *Wald Identities.*

2.2. **The Wald Identities.**

Recall that a *stopping time* (relative to the sequence $Y_0, Y_1, \dots$) is a nonnegative-integer-valued random variable τ such that, for each $n = 0, 1, 2, \dots$, the event $\{\tau = n\}$ depends only on the values of $Y_0, Y_1, \dots, Y_n$. If Y_n is one of the three sequences $M_n, V_n, Z_n(\theta)$ defined above, this is equivalent to requiring that the event $\{\tau = n\}$ depend only on the values $\xi_1, \xi_2, \dots, \xi_n$. Thus, the random variables

$$\tau(a) = \min\{n \,:\, S_n \geq a\};$$
$$T(b) = \min\{n \,:\, (S_n - n\mu)^2 \geq 1.96\sigma^2 n\}$$

are stopping times. Note that they are not *bounded* stopping times. It is important in dealing with unbounded stopping times to realize that, in certain circumstances, the Optional Sampling Formula may *fail* to hold. It is also important to keep in mind that any unbounded stopping time may be approximated, in a certain sense, by bounded stopping times: in particular, for any stopping time τ (not necessarily bounded) and any fixed nonrandom integer $n \geq 0$, the truncated random variable $\tau \wedge n$ is a bounded stopping time. (EXERCISE: Prove this!) Moreover, as $n \to \infty$, the truncated times $\tau \wedge n$ converge to τ. (EXERCISE: Verify this!)

First Wald Identity . *Assume that the random variables ξ_j have finite first moment, and let $\mu = E\xi_1$. Then for any stopping time τ with finite expectation,*

$$ES_\tau = \mu E\tau. \tag{5}$$

Second Wald Identity . *Assume that the random variables ξ_j have finite second moment, and let $\mu = E\xi_1$ and $\sigma^2 = E(\xi_1 - \mu)^2$. Then for any stoping time τ with finite expectation,*

$$E(S_\tau - m\tau)^2 = \sigma^2 E\tau. \tag{6}$$

Third Wald Identity . *Assume that the moment generating function $\varphi(\theta) = Ee^{\theta\xi_1}$ of the random variables ξ_j is finite at the argument θ. Then for any bounded stopping time τ,*

$$E\left(\frac{\exp\{\theta S_\tau\}}{\varphi(\theta)^\tau}\right) = 1. \tag{7}$$

Note that the hypothesis on the stopping time τ is stronger in the Third Wald Identity than in the first two. Later we will see an example where equation (7) fails even though $E\tau < \infty$.

All three of these theorems will ultimately be deduced from Theorem 1 below. However, it is not difficult to prove them directly (at least for *bounded* stopping times τ), and this is instructive. We shall limit our attention to the Third Wald Identity.

Proof of the Third Wald Identity. The key to this is that the expectation of a product is the product of the expectations, *provided* that the factors in the product are independent. Fix indices $0 \leq k < m$. The event $\{\tau = k\}$ depends only on the random variables $\xi_1, \xi_2, \dots, \xi_k$, and so is independent of the random variable ξ_m. Similarly, the product $e^{\theta S_k}\mathbf{1}\{\tau = k\}$ is independent of $\sum_{m=k+1}^{n} \xi_m$. Consequently, by the product rule, for any $n \geq k$,

$$\begin{aligned} E\exp\{\theta S_n\}\mathbf{1}\{\tau = k\} &= E\exp\{\theta S_k\}\exp\{\theta(S_n - S_k)\}\mathbf{1}\{\tau = k\} \\ &= E\exp\{\theta(S_n - S_k)\}E\exp\{\theta S_k\}\mathbf{1}\{\tau = k\} \\ &= \varphi(\theta)^{n-k}Ee^{\theta S_k}\mathbf{1}\{\tau = k\}. \end{aligned} \tag{8}$$

Here $\mathbf{1}_F$ denotes the *indicator random variable* for the event F, that is, the random variable that takes the value 1 on F and 0 on F^c.

Suppose now that τ is a *bounded* stopping time, that is, that there is a nonrandom integer $n < \infty$ such that $\tau \leq n$. Then by equation (8),

$$\begin{aligned} E\left(\frac{\exp\{\theta S_\tau\}}{\varphi(\theta)^\tau}\right) &= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_\tau\}}{\varphi(\theta)^\tau}\right)\mathbf{1}\{\tau = k\} \\ &= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_k\}}{\varphi(\theta)^k}\right)\mathbf{1}\{\tau = k\} \\ &= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_k\}}{\varphi(\theta)^k}\right)\left(\frac{\exp\{\theta S_{n-k}\}}{\varphi(\theta)^{n-k}}\right)\mathbf{1}\{\tau = k\} \\ &= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_n\}}{\varphi(\theta)^n}\right)\mathbf{1}\{\tau = k\} \\ &= E\left(\frac{\exp\{\theta S_n\}}{\varphi(\theta)^n}\right) \\ &= 1. \end{aligned}$$

□

2.3. **First Application: The Gambler's Ruin Problem.**

Two gamblers, FATS and SLIM, play the following game: FATS repeatedly tosses a fair coin. After each toss that comes up H, SLIM pays FATS one dollar. After each toss that comes up T, FATS pays SLIM one dollar. The game continues until either one or the other gambler runs out of money. If FATS starts with \$$A$ and SLIM starts with \$$B$,

(A) What is the probability that, when the game ends, FATS has all the cash?
(B) What is the expected duration of the game?

This may be recast as an optional stopping problem. Let $X_0, X_1, \ldots$ be the sequence of play-by-play increments in FATS' fortune: thus, $X_i = \pm 1$ according to whether the ith toss is H or T. The total change in FATS' fortune after n plays is

$$S_n = \sum_{i=1}^{n} X_i.$$

The game continues until time τ, where

$$\tau = \min\{n : S_n = +A \text{ or } -B\}.$$

Solution to Problem A. It is not difficult to see that τ is a stopping time relative to the sequence $X_0, X_1, \dots$ and we shall prove below by an elelemnatary argument that $E\tau < \infty$. Consequently, by the First Wald Identity, for each $n < \infty$,

$$0 = AP\{S_\tau = A\} - BP\{S_\tau = -B\}.$$

Since S_τ must be either A or $-B$, the two probabilities in this equation must sum to 1. Hence, we have two equations in two unknowns, which we may solve to obtain the solution to Problem (A):

$$P\{S_\tau = A\} = \frac{B}{A+B} \qquad \text{and} \tag{9}$$
$$P\{S_\tau = B\} = \frac{A}{A+B}$$

Solution to Problem B. To solve Problem (B), we shall appeal to the Second Wald Identity. Since $\mu = 0$ and $\sigma^2 = 1$, and since $E\tau < \infty$ (see below), the Second Wald Identity implies that

$$ES_\tau^2 = E\tau.$$

But S_τ takes one of only two possible values, $+A$ and $-B$, and by equations (9) above, the probability distribution is known. Thus,

$$ES_\tau^2 = \frac{A^2B}{A+B} + \frac{B^2A}{A+B} = AB$$

This proves that

$$E\tau = AB. \tag{10}$$

To justify the calculations that led to equations (9) and (10), we must show that $E\tau < \infty$. To this end, set $M = A+B$, and consider how the game proceeds in blocks of M consecutive tosses. In each such block, it is possible that there will be M consecutive Heads, in which case the game will end and FATS will take home the money. For each such block of M consecutive tosses, this happens with probability $1/2^M$. Thus, the probability that the game has *not* ended after the first kM tosses is $(1-2^{-M})^k$. Thus,

$$P\{\tau > kM\} \leq \left(1 - \frac{1}{2^M}\right)^k,$$

that is, the distribution of τ has an exponentially decaying tail. This implies that $E\tau < \infty$. (EXERCISE: Explain why.)

2.4. Second Application: Distribution of a First-Passage Time.

Once again let S_n be the simple nearest neighbor random walk on the integers, that is, $S_n = \sum_{j=1}^n X_j$ where $X_0, X_1, \dots$ are independent, identically distributed random variables that take the values ± 1 with probability $1/2$ each. We know that the Markov chain S_n is recurrent, and so it will eventually visit each integer. Hence, the random variable

$$T = \min\{n : S_n = 1\}$$

is finite with probability one, and it is readily apparent that it is a stopping time. It is *not* the case that $ET < \infty$, because if it were then we would have a contradiction to Wald's First Identity:

$$1 = ES_T = EX_1 ET = 0.$$

Earlier in the course we obtained formulas for the distribution of T, in two different ways: first, by using the reflection principle, and second, by analyzing the Renewal Equation for the probability generating function. (Refer to HW set 2.) Now we will show that this distribution can be obtained in a radically different way, using Wald's Third Identity. Fix $\theta > 0$, and let $z = Ee^{\theta X_1} = \cosh\theta$. For each positive integer n, the random variable $T \wedge n$ is a bounded stopping time, and so Wald's Third Identity implies that

$$E\left(\frac{\exp\{\theta S_{T\wedge n}\}}{z^{T\wedge n}}\right) = 1. \tag{11}$$

We will argue that it is permissible to take $n \to \infty$ in this identity. Suppose for the moment that it is; then since $S_T \equiv 1$, the limiting form of the identity will read

$$e^{\theta} E z^{-T} = 1. \tag{12}$$

Set $s = 1/z$. Since $z = \cosh\theta \geq 1$, it follows (by solving a quadratic equation: Exercise!) that

$$e^{-\theta} = \frac{1 - \sqrt{1 - 4s^2}}{2s}.$$

Thus, the limiting form (12) of the Wald Identity may be rewritten so as to give the following exact formula for the probability generating function of the stopping time T:

$$\boxed{Es^T = \frac{1 - \sqrt{1 - 4s^2}}{2s}} \tag{13}$$

You should note that this is exactly the same formula that we obtained by analyzing the Renewal Equation earlier in the course.

To justify taking $n \to \infty$ in the Wald Identity (11) above, we shall use the Dominated Convergence Theorem. First, note that since $T < \infty$ (at least with probability one),

$$\lim_{n\to\infty} \frac{\exp\{\theta S_{T\wedge n}\}}{z^{T\wedge n}} = \frac{\exp\{\theta S_T\}}{z^T}.$$

Hence, by the DCT, it will follow that the limiting equation (12) holds provided the integrands are dominated by an integrable random variable. For this, examine the numerator and the denominator separately. Since $\theta > 0$, the random variable $e^{\theta S_{T\wedge n}}$ cannot be larger than e^{θ}, because on the one hand, $S_T = 1$, and on the other, if $T > n$ then $S_n \leq 0$ and so $e^{S_{T\wedge n}} \leq 1$. The denominator is even easier: since $z = \cosh\theta \geq 1$, it is always the case that $z^{T\wedge n} \geq 1$. Thus,

$$\frac{\exp\{\theta S_{T\wedge n}\}}{z^{T\wedge n}} \leq e^{\theta},$$

and so the integrands are uniformly bounded.

3. CONDITIONAL EXPECTATION

The theory of martingales rests on the notion of *conditional expectation.* For martingales in discrete probability spaces, or for martingales built from independent random variables, such as those considered in section 2, this presents no difficulties, as conditional expectation can be defined in an elementary manner. In general, however, a more sophisticated view of conditional expectation is needed.

3.1. **Definition of Conditional Expectation.**

The conditional expectation of a discrete random variable X given the value y of another discrete random variable Y may be defined in terms of the conditional distribution of X given the event $\{Y = y\}$: in particular,

$$E(X \mid Y = y) = \sum_x x P(X = x \mid Y = y), \tag{14}$$

where the sum is over the set of all possible values x of X. Note that this expression depends on the value y – this will be important in what follows. 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 (14) 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 condtional 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 $Y_1, Y_2, \dots, Y_m$ be random variables taking values in a set $\mathcal{Y}$.*[1] *The* conditional expectation $E(X \mid Y_1, Y_2, \dots, Y_m)$ *is the unique (measurable) real-valued function of $Y_1, Y_2, \dots, Y_m$ such that for every bounded (measurable) function $g(Y_1, Y_2, \dots, Y_m)$,*

$$EXg(Y_1, Y_2, \dots, Y_m) = E(E(X \mid Y_1, Y_2, \dots, Y_m)g(Y_1, Y_2, \dots, Y_m)). \tag{15}$$

It is by no means clear *a priori* that such a function $E(X \mid Y_1, Y_2, \dots, Y_m)$ should always exist, nor is it obvious that it should be unique. In fact, the *existence* of such a function is an important theorem of measure theory, called the *Radon-Nikodym* theorem. You will see its proof next year, in Prof. Wichura's class. The *uniqueness* of the function is not so difficult to prove:

Proof of Uniqueness. For convenience, denote by $Y = (Y_1, Y_2, \dots, Y_m)$ the random vector on which we are conditioning. Suppose that there are distinct functions $h_1(y)$ and $h_2(y)$ such

---

[1] More precisely, $Y_1, Y_2, \dots, Y_m$ take values in a *measurable space* $(\mathcal{Y}, \mathcal{G})$. If the set $\mathcal{Y}$ is finite, there is no need to worry about measurability issues.

that, for every bounded function $g(y)$,

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

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 $P\{h_1(Y) = h_2(Y)\} = 1$. □

3.2. **Equivalence of the Naive and Modern Definitions.** It is not difficult to show that the naive definition (15) 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

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

□

3.3. **Properties of Conditional Expectation.** The raw definition 1 can be rather 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.

In stating the propperties below, it is convenient to use the abbreviation $Y = (Y_1, Y_2, \dots, Y_n)$.

Summary: Basic Properties of Conditional Expectation.

(1) **Definition:** $EXg(Y) = EE(X \,|\, Y)g(Y)$ for all nonnegative 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 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.
(6) **Tower Property:** If Z is a function of Y then $E(E(X \mid Y) \mid Z) = E(X \mid Z)$.
(7) **Expectation Law:** $E(E(X|Y)) = EX$.
(8) **Constants:** For any scalar a, $E(a|Y) = a$.
(9) **Jensen Inequalities:** If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $E|X| < \infty$ then

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

With the exception of the Jensen inequality (7) , all of these properties may be proved easily, using only the 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 Property 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, which you will learn next year, 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 (15) of conditional expectation. First, suppose that $X \geq 0$. Define B to be the set of possible values of Y where 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 (15),

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

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 must be 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 sufficies to show that it satisfies equation (15), that is, that for every bounded nonnegative function $g(Y)$,

$$E(aU + bV)g(Y) = E(aE(U \mid Y) + bE(V \mid Y))g(Y). \tag{16}$$

But equation (15) holds for $X = U$ and for $X = V$:

$$\begin{aligned} EUg(Y) &= EE(U \mid Y)g(Y), \\ EVg(Y) &= EE(V \mid Y)g(Y). \end{aligned}$$

Multiplying these equations by a and b, respectively, and then adding gives (16), because the unconditional expectation operator is linear. $\square$

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 $EW(X \mid Y)$. $\square$

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

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

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,

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

$\square$

4. MARTINGALES

4.1. **Definition of a Martingale.** Let $X_0, X_1, \dots$ and $Y_0, Y_1, \dots$ be sequences of random variables defined on a common probability space, with X_n being real-valued and integrable (that is, $E|X_n| < \infty$), and Y_n valued in an arbitrary (measurable) space. In common applications in stochastic process theory, the random variables Y_n might be the successive states in a Markov chain, and the random variables X_n numerical functions of the states Y_n. The sequence $X_0, X_1, \dots$ is said to be *adapted* to the sequence $Y_0, Y_1, \dots$ if for each n the random variable X_n is a function[2] of $Y_1, Y_2, \dots, Y_n$. The sequence $X_0, X_1, \dots$ is said to be a *martingale* relative to the sequence $Y_0, Y_1, \dots$ if it is adapted, and if for every n,

$$\boxed{E(X_{n+1} \,|\, Y_0, Y_1, Y_2, \dots, Y_n) = X_n.} \tag{17}$$

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

$$E(X_{n+1} \,|\, Y_0, Y_1, Y_2, \dots, Y_n) \leq (\geq) X_n. \tag{18}$$

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

4.2. **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

$$X_n = \sum_{j=1}^{n} Y_j \tag{19}$$

is a martingale relative to the sequence $0, Y_1, Y_2, \dots$. This is easily verified, using the linearity and stability properties and the Independence Law for conditional expectation:

$$\begin{aligned} E(X_{n+1} \,|\, Y_1, Y_2, \dots, Y_n) &= E(X_n + Y_{n+1} \,|\, Y_1, Y_2, \dots, Y_n) \\ &= E(X_n \,|\, Y_1, Y_2, \dots, Y_n) + E(Y_{n+1} \,|\, Y_1, Y_2, \dots, Y_n) \\ &= X_n + EY_{n+1} \\ &= X_n. \end{aligned}$$

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.[3] To get some appreciation of why this might be so, consider the decomposition of a martingale $\{X_n\}$ as a partial sum process:

$$X_n = X_0 + \sum_{j=1}^{n} \xi_j \quad \text{where} \quad \xi_j = X_j - X_{j-1}. \tag{20}$$

Proposition 1. *The* martingale difference sequence $\{\xi_n\}$ *has the following properties: (a) the random variable* ξ_n *is a function of* $Y_1, Y_2, \dots, Y_n$*; and (b) for every* $n \geq 0$*,*

$$E(\xi_{n+1} \,|\, Y_1, Y_2, \dots, Y_n) = 0. \tag{21}$$

Proof. This is another three-line calculation using the properties of conditional expectation and the definition of a martingale. ☐

---

[2] Here and in what follows, the term *function* will always mean *measurable function.* If you do not know yet what this mean, don't worry about it: all reasonable functions are measurable.

[3] We won't have a chance to discuss all of these in this class, but you can rest assured that Prof. Wichura will give a complete tour next year.

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

$$EX_n = EX_0. \tag{22}$$

Moreover, if $X_0 = 0$ *and* $EX_n^2 < \infty$ *then*

$$EX_n^2 = \sum_{j=1}^{n} E\xi_j^2. \tag{23}$$

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 (22).

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 a function of $Y_1, Y_2, \dots, Y_j$, and therefore also a function of $Y_1, Y_2, \dots, Y_k$. Thus, by Properties (1), (4), (6), and (7) of conditional expectation, if $j \leq k \leq n$ then

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

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

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

□

4.3. **Some Examples of Martingales.**

4.3.1. *Paul Lévy's Martingales.* Let X be any integrable random variable. Then the sequence X_n defined by $X_n = E(X|Y_0, Y_1, \dots, Y_n)$ is a martingale, by the Tower Property of conditional expectation.

4.3.2. *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.

4.3.3. *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

$$\left(\sum_{j=1}^{n} Y_j\right)^2 - \sigma^2 n \tag{24}$$

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

4.3.4. *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

$$Z_n = \left(\frac{q}{p}\right)^{2S_n - n}. \tag{25}$$

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.

4.3.5. *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

$$Z_n = Z_n(\theta) = \prod_{j=1}^{n} \frac{e^{\theta X_j}}{\varphi(\theta)} = \frac{e^{\theta S_n}}{\varphi(\theta)^n}. \tag{26}$$

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.)

4.3.6. *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. $\square$

4.3.7. *Harmonic Functions and Markov Chains.* Yes, surely enough, martingales also arise in connection with Markov chains; in fact, Doob's motivation in inventing them was to connect the world of potential theory for Markov processes with the classical theory of sums of independent random variables.[4] Let $Y_0, Y_0, Y_1, \ldots$ be a Markov chain on a denumerable state space $\mathcal{Y}$ with transition probability matrix $\mathbb{P}$. A real-valued function $h : \mathcal{Y} \to \mathbb{R}$ is called *harmonic* for the transition probability matrix $\mathbb{P}$ if

$$\mathbb{P}h = h, \tag{27}$$

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

$$h(x) = \sum_{y \in \mathcal{Y}} p(x,y)h(y) = E^x h(Y_1). \tag{28}$$

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 (28) 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 $x \in \mathcal{Y}$ 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 $Y_0, Y_1, Y_2, \ldots, Y_n$ as a conditional expectation on Y_n. Thus,

$$\begin{aligned} E(h(Y_{n+1}) \mid Y_0, Y_1, Y_2, \ldots, Y_n) &= E(h(Y_{n+1}) \mid Y_n) \\ &= \sum_{y \in \mathcal{Y}} h(y) p(Y_n, y) \\ &= h(Y_n). \end{aligned}$$

□

4.3.8. *Submartingales from Martingales.* Let $\{X_n\}_{n \geq 0}$ be a martingale relative to the sequence $Y_0, Y_1, \ldots$. 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

$$Z_n = \varphi(X_n) \tag{29}$$

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

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

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

---

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

5. 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.

5.1. **Martingale Transforms.** A formal definition of a martingale transform requires two auxiliary notions: *martingale differences* and *predictable sequences*. Let $X_0, X_1, \dots$ be a martingale relative to another sequence $Y_0, Y_1, \dots$. For $n = 1, 2, \dots$, define

$$\xi_n = X_n - X_{n-1}; \tag{30}$$

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

A *predictable sequence* $Z_1, Z_2, \dots$ relative to the sequence $Y_0, Y_1, \dots$ is a sequence of random variables such that for each $n = 1, 2, \dots$ the random variable Z_n is a function of $Y_0, Y_1, \dots, Y_{n-1}$. 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 2. *Let $X_0, X_1, \dots$ be a martingale relative to $Y_0, Y_1, \dots$ and let $\xi_n = X_n - X_{n-1}$ be the associated martingale difference sequence. Let $Z_0, Z_1, \dots$ be a predictable sequence relative to $Y_0, Y_1, \dots$. Then the* martingale transform $\{(Z \cdot X)_n\}_{n\geq 0}$ *is defined by*

$$(Z \cdot X)_n = Z_0 X_0 + \sum_{k=1}^{n} Z_k \xi_k. \tag{31}$$

Example: The St. Petersburg Game. In this game, a referee tosses a fair coin repeatedly, with results $\xi_1, \xi_2, \dots$, 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 $\xi_1, \xi_2, \dots, \xi_{n-1}$, 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 $X_n = \xi_1 + \xi_2 + \dots + \xi_n$. □

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:

$$\begin{aligned} E((Z\cdot X)_{n+1}\,|\,Y_1,Y_2,\dots,Y_n) &= (Z\cdot X)_n + E(Z_{n+1}\xi_{n+1}\,|\,Y_1,Y_2,\dots,Y_n)\\ &= (Z\cdot X)_n + Z_{n+1}E(\xi_{n+1}\,|\,Y_1,Y_2,\dots,Y_n)\\ &= (Z\cdot X)_n, \end{aligned}$$

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

5.2. **Submartingale Transforms.** Submartingales and supermartingales may also be transformed, using equation (31), 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 3. *Let* $X_0, X_1, \dots$ *be a sub- (respectively, super-) martingale relative to* $Y_0, Y_1, \dots$ *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 relative to* $Y_0, Y_1, \dots$ *consisting of bounded* nonnegative *random variables. Then the* submartingale transform *(respectively,* supermartingale transform*)* $\{(Z\cdot X)_n\}_{n\geq 0}$ *is defined by*

$$(Z\cdot X)_n = Z_0X_0 + \sum_{k=1}^{n} Z_k\xi_k. \tag{32}$$

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* $n\geq 0$,

$$0\leq Z_n\leq 1, \tag{33}$$

then

$$E(Z\cdot X)_n \leq EX_n. \tag{34}$$

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

$$E(Z_k\xi_k\,|\,Y_1,Y_2,\dots,Y_{k-1}) = Z_kE(\xi_k\,|\,Y_1,Y_2,\dots,Y_{k-1}).$$

Since ξ_k is a submartingale difference sequence, $E(\xi_k\,|\,Y_1,Y_2,\dots,Y_{k-1}\geq 0)$; and therefore, since $0\leq Z_k\leq 1$,

$$0\leq E(Z_k\xi_k\,|\,Y_1,Y_2,\dots,Y_{k-1})\leq E(\xi_k\,|\,Y_1,Y_2,\dots,Y_{k-1}).$$

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 (34). □

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 (33) holds for each* $n \geq 0$ *then*

$$E(Z\cdot X)_n \geq EX_n. \tag{35}$$

6. 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.*[5]

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 $\{X_n\}_{n\geq 0}$:

$$X_{\tau\wedge n} = (Z\cdot X)_n \tag{36}$$

where

$$Z_n = \begin{cases} 1 & \text{if } \tau \geq n, \qquad \text{and} \\ 0 & \text{if } \tau < n. \end{cases} \tag{37}$$

The equation (36) 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.

---

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

In order that the sequence $\{(Z \cdot X)_n\}_{n \geq 0}$ be a martingale transform (respectively, sub- or super- martingale 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 $Y_0, Y_1, Y_2, \dots, Y_m$. Hence, the event

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

depends only on $Y_0, Y_1, \dots, Y_{n-1}$. But this event is the same as the event that $Z_n = 1$; this proves that Z_n is a function only of $Y_0, Y_1, \dots, Y_{n-1}$, 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 (34), since the terms Z_n of the predictable sequence are between 0 and 1. ☐

7. 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*

$$M_n = \max_{0 \leq m \leq n} X_m, \qquad \text{and} \tag{38}$$

$$M_\infty = \sup_{0 \leq m < \infty} X_m = \lim_{n \to \infty} M_n \tag{39}$$

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

$$P\{M_n \geq \alpha\} \leq E(X_n \vee 0)/\alpha \qquad \text{if } \{X_n\}_{n \geq 0} \text{ is a submartingale, and} \tag{40}$$

$$P\{M_\infty \geq \alpha\} \leq EX_0/\alpha \qquad \text{if } \{X_n\}_{n \geq 0} \text{ is a nonnegative supermartingale.} \tag{41}$$

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 (40).

The proof of inequality (41) 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 (41) follows. $\square$

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 (41) 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

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

8. 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

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

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,*

$$(\beta - \alpha) E N_n((\alpha, \beta]) \leq E(X_n \vee 0) + |\alpha|. \tag{43}$$

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:

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

(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,

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

□

9. The Martingale Convergence Theorem

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*

$$\lim_{n \to \infty} X_n := X_\infty \tag{44}$$

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 $X_\infty := \lim_{nxg\infty} X_n$ exists (but may be $\pm\infty$). But Fatou's Lemma implies that

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

and so in particular 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 $X_n \geq 0$,

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

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

Note: 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!

(To be continued.)