MARTINGALES AND SOME APPLICATIONS

By

Anselmo Ramón Sánchez Titla

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
SPECIALTY IN PROBABILITY AND STATISTICS
AT
CENTRO DE INVESTIGACION EN MATEMÁTICAS
CALLEJON DE JALISCO S/N
JANUARY 2004

CENTRO DE INVESTIGACION EN MATEMÁTICAS
DEPARTMENT OF
PROBABILITY AND STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "**Martingales and Some Applications**" by **Anselmo Ramón Sánchez Titla** in partial fulfillment of the requirements for the degree of **Specialty in Probability and Statistics.**

Dated: February 2004

Supervisor: Dr. Daniel Hernndez Hernndez

Readers: Joaqun Ortega

CENTRO DE INVESTIGACION EN MATEMÁTICAS

Date: **January 2004**

Author: **Anselmo Ramón Sánchez Titla**

Title: **Martingales and Some Applications**

Department: **Probability and Statistics**

Convocation: **January** Year: **2004**

Signature of Author

Table of Contents

Abstract

The aim of this work is to give a summary of what martingales are and how they are used in the convergence of random variables. We start by giving the concept of stopping time, then the concept of martingale is introduced and applied to obtain important results such as Doob's optional stopping theorem and Wald's formula on the sum of a random number of random variables and the supermartingale convergence theorem.

Introduction

The study of dependent random variables is one of the main subjects in the theory of probability and statistics. The stochastic processes that are martingales deal with a type of dependence expressed through the concept of conditional expectation and its importance falls in that many processes are studied in a convenient way within this context.

Probability theory has its roots in games of chance and it is often profitable to interpret results in terms of gambling situations. For example, if $X_1, X_2, \ldots$ is a sequence of random variables, we may think of X_n as our total winnings after n trials in a sequence of games. Having survived the first n trails, our expected fortune after trial $n+1$ is $E(X_{n+1} \mid X_1, \ldots X_n)$. If this equals X_n, the game is "fair" since the expected gain on trial $n+1$ is $E(X_{n+1} - X_n \mid X_1, \ldots X_n) = X_n - X_n = 0$. If $E(X_{n+1} - X_n \mid X_1, \ldots X_n) \geq 0$, the game is "favorable," and if $E(X_{n+1} - X_n \mid X_1, \ldots X_n) \leq 0$, the game is "unfavorable". This is what motivates the concept of martingale, submartingale and supermartingale. As we will see, there are two basic facts about martingales. The first is that you are betting on a fair game, and in particular if you choose to stop playing at some bounded time N then your expected winnings $E(X_N)$ are equal to your initial fortune X_0. Our second fact concerns submartingales; they are the stochastic analogues of non decreasing sequences and so if they are bounded above i.e. $\sup E(X_n^+) < \infty$, they converge almost everywhere.

Chapter 1

Martingale Theory

In this chapter we give the concepts of a stopping time and martingale, and establish important results such as the Doob sampling theorem and Wald's formula on the sum of a random number of random variables, as well as the supermartingale convergence theorem.

1.1 Stopping times

Definition 1.1.1. Let $\{\mathcal{F}_n, n = 0, 1, \ldots\}$ be an increasing sequence of sub σ-fields of $\mathcal{F}$. A stopping time for $\mathcal{F}_n$ is a map $T\colon \Omega \longrightarrow \{0, 1, \ldots, \infty\}$ such that $\{T = n\} \in \mathcal{F}_n$ for each nonnegative integer n. Since $\{T \leq n\} = \bigcup_{k=0}^{n} \{T = k\}$ and $\{T = n\} = \{T \leq n\} - \{T \leq n-1\}$, the definition is equivalent to the requirement that $\{T \leq n\} \in \mathcal{F}_n$ for all $n = 0, 1, \ldots$. If $\{X_n, n = 0, 1, \ldots\}$ is a sequence of random variables, a stopping time for $\{X_n\}$ is, by definition, a stopping time relative to the σ-fields $\mathcal{F}_n = \mathcal{F}(X_0, \ldots, X_n)$.

One of the most important examples of stopping time is the *hitting time* of a set. If $\{X_n\}$ is a sequence of random variables and $B \in \mathcal{B}(\mathbb{R})$, let $T(w) = \inf\{n \;:\; X_n \in B\}$

if $X_n(w) \in B$ for some n; $T(w) = \infty$ if $X_n(w)$ never hits B. T is a stopping time since $\{T \leq n\} = \bigcup_{k \leq n} \{X_k \in B\} \in \mathcal{F}(X_k, k \leq n)$ or $\{T = n\} = \{X_1 \notin B\} \cap \cdots \cap \{X_{n-1} \notin B\} \cap \{X_n \in B\} \in \mathcal{F}(X_k, k \leq n)$.

Interpretation. $n = 1, 2, \ldots$ represents the time, $\mathcal{F}_n$ the information available at time n (the events that have occurred at time n), and X_n denotes some random quantity whose value $X_n(w)$ is revealed at time n.

Example 1.1.1. Consider the problem of coin tossing, that is, let $X_1, X_2, \ldots$ be independent random variables, each taking on values ± 1 according to if the coin comes up head or tail with probabilities p and q, respectively. Let $Y_n = X_1 + \cdots + X_n$, the amount accumulated at time n. For instance, if $T = \inf\{n : \; Y_n = 5\}$, the hitting time for $\{5\}$, we decide to finish the game when we have 5 pesos.

One important question in this example is: what is the sample space Ω and the σ- field $\mathcal{F}$? Of course, Ω consists of the infinite sequences $(w_1, w_2, \ldots)$ where each component is head or tail. Consider a typical sample point $w_{i_1,\ldots,i_k,i_{k+1},\ldots,i_N}$ with heads in positions $i_1, \ldots, i_k$ and tails in positions $i_{k+1}, \ldots, i_N$. If A_i is the event of obtaining a head on trial i, so that $A_i = \{w :$ the ith coordinate of w is head$\}$, we have

$$\{w_{i_1,\ldots,i_k,i_{k+1},\ldots,i_N}\} = A_{i_1} \cap \cdots \cap A_{i_k} \cap A^c_{i_{k+1}} \cap \cdots \cap A^c_{i_N}.$$

If $\mathcal{F} = 2^\Omega$ we have that $\{w_{i_1,\ldots,i_k,i_{k+1},\ldots,i_N}\} \in \mathcal{F}$. Since the trials are independent and $P(A_i) = p$ for all i, the probability assigned to $w_{i_1,\ldots,i_k,i_{k+1},\ldots,i_N}$ is determined; it must be

$$P(w_{i_1,\ldots,i_k,i_{k+1},\ldots,i_N}) = P(A_{i_1}) \ldots P(A_{i_k}) P(A^c_{i_{k+1}}) \ldots P(A^c_{i_N}) = p^k q^{N-k}.$$

But this natural definition of probability measure has problems when N is infinity, therefore we have to give another σ-field and probability measure.

We may now define the sub σ- fields $\mathcal{F}_i$ in order to give an alternative probability measure. To do so, we realize we would like that $\mathcal{F}_1$, for example, has all the information about the first toss, $\mathcal{F}_2$ about the second toss and so on.

Let $HHTHTT...THTT$ be a permutation of the letters H and T in the positions $i_1, \ldots, i_j$ and $i_{j+1}, \ldots, i_m$ respectively. If we denote with this permutation the collection of all sample points $w_{i_1,\ldots,i_j,i_{j+1},\ldots,i_m}$, with heads in positions $i_1, \ldots, i_k$ and tails in positions $i_{k+1}, \ldots, i_N$, we have

$$\mathcal{F}_1 = \{\emptyset, \Omega, T, H\}$$

is a σ-field.

$$\mathcal{F}_2 = \{\emptyset, \Omega, TT, TH, HH, HT, TT \cup TH = T, \ldots\}$$

We may assign a measure to each event in $\mathcal{F}_2$. We define

$$P(TT) = p^2, P(TH) = pq, P(HH) = q^2, P(HT) = qp, P(T) = p, \ldots.$$

Analogously

$$\mathcal{F}_3 = \{\emptyset, \Omega, TTT, TTH, THT, HTT, THH, HTH, HHT, HHH, \ldots\}.$$

Proceeding inductively we define the sub σ-fields $\mathcal{F}_n$ for $n > 3$.

Setting $\mathcal{F} = \sigma(\bigcup_{n\geq 1} \mathcal{F}_n)$, we get the desired construction. Here we also have that the σ-field $\mathcal{F}$, can no longer be the power set. We will now show that $T = \inf\{n \mid Y_n = 5\}$ is a stopping time. We have that for $k = 1, 2, 3, 4$, $\{T = k\} = \emptyset \in \mathcal{F}_k$, while $\{T = 5\} = HHHHH \in \mathcal{F}_5$, $\{T = 6\} = \emptyset \in \mathcal{F}_6$, $\{T = 7\} = THHHHHH \cup HTHHHHH \cup \cdots \cup HHHHTHH \in \mathcal{F}_7$, and so on.

Example 1.1.2. Let $T(w) = n$, that is, we stop when we have played n times. Since $\{T = k\} = \emptyset$ for $k \neq n$ and $\{T = k\} = \Omega$ for $k = n$, T is a stopping time.

If $\mathcal{F}_i$ corresponds to the information available up to time i, how should we define a σ-field $\mathcal{F}_T$ in such a way that it contains the information available up to a random time T? Intuitively on the part of Ω where $T = i$ the sets in the σ-field $\mathcal{F}_T$ should be the same as the sets in the σ-field $\mathcal{F}_i$. That is, we could hope that

$$\{A \cap \{T = i\} \ : \ A \in \mathcal{F}_T\} = \{A \cap \{T = i\} \ : \ A \in \mathcal{F}_i\} \qquad \text{for each } i.$$

These relations would be suitable as a definition of $\mathcal{F}_T$ in discrete time.

Definition 1.1.2. Let T be a stopping time for the σ-fields $\mathcal{F}_n, n = 0, 1, \ldots$. The σ-field $\mathcal{F}_T$ consists of all the events $A \in \mathcal{F}$ for which $A \cap \{T = n\} \in \mathcal{F}_n$ for all $n = 0, 1, \ldots$. It follows that $\mathcal{F}_T$ is a σ-field.

The σ-field $\mathcal{F}_T$ arises very often in the following way. Let T be a finite stopping time for $\{X_n\}$, and define X_T in the natural way: if $T(w) = n$, let $X_T(w) = X_n(w)$. If $B \in \mathcal{B}(\mathbb{R})$, then $\{X_T \in B\} \in \mathcal{F}_T$, in other words, X_T is $\mathcal{F}_T$-measurable. To see this, write

$$\{X_T \in B\} \cap \{T \leq n\} = \bigcup_{k=0}^{n} [\{X_k \in B\} \cap \{T = k\}].$$

Since $\{X_k \in B\} \cap \{T = k\} \in \mathcal{F}(X_0, \ldots, X_n)$ for $k \leq n$, we have

$$\{X_T \in B\} \cap \{T \leq n\} \in \mathcal{F}(X_0, \ldots, X_n).$$

Hence $\{X_T \in B\} \in \mathcal{F}_T$, i.e. X_T is $\mathcal{F}_T$-measurable.

Example 1.1.3. If T_1 and T_2 are both stopping times, so are $T_1 \wedge T_2$ and $T_1 \vee T_2$ because $\{T_1 \wedge T_2 \leq n\} = \{T_1 \leq n\} \cup \{T_2 \leq n\} \in \mathcal{F}_n$ and $\{T_1 \vee T_2 \leq n\} = \{T_1 \leq n\} \cap \{T_2 \leq n\} \in \mathcal{F}_n$.

Example 1.1.4. Suppose T_1 and T_2 are both stopping times, such that $T_1 \leq T_2$. Then, $\mathcal{F}_{T_1} \subseteq \mathcal{F}_{T_2}$, because

$$A \cap \{T_2 \leq n\} = (A \cap \{T_1 \leq n\}) \cap \{T_2 \leq n\} \qquad \text{for all } n = 0, 1, \ldots,$$

and both sets on the right-hand side are $\mathcal{F}_n$-measurable if $A \in \mathcal{F}_{T_1}$.

Example 1.1.5. The stopping time T is measurable with respect to $\mathcal{F}_T$, because for each $\alpha \in \mathbb{R}^+$ and $n = 0, 1, \ldots,$

$$\{T \leq \alpha\} \cap \{T \leq n\} = \{T \leq m \wedge n\} \in \mathcal{F}_{m \wedge n} \subseteq \mathcal{F}_n,$$

where m is the greatest integer not greater than α. That is, $\{T \leq \alpha\} \in \mathcal{F}_T$ for all $\alpha \in \mathbb{R}^+$, and hence T is $\mathcal{F}_T$-measurable.

1.2 Martingales

A little notation goes a long way in martingale theory. A fixed probability space $(\Omega, \mathcal{F}, P)$ sits in the background throughout. The key new ingredients are:

(i) a set T partially ordered by a relation "$\leq$";

(ii) a filtration $\{\mathcal{F}_t : t \in T\}$, that is a collection of sub σ-fields of $\mathcal{F}$ for which $\mathcal{F}_s \subseteq \mathcal{F}_t$ if $s \leq t$;

(iii) a family of integrable random variables $\{X_t : t \in T\}$ **adapted** to the filtration, that is, X_t is $\mathcal{F}_t$-measurable for each $t \in T$.

Definition 1.2.1. A family of integrable random variables $\{X_t : t \in T\}$ adapted to a filtration $\{\mathcal{F}_t : t \in T\}$ is said to be a *martingale* relative to the $\mathcal{F}_t$ (alternatively, we say that $\{X_t, \mathcal{F}_t\}$ is a martingale) iff for all $s \leq t$, $E(X_t \mid \mathcal{F}_s) = X_s$ a.e., a *submartingale* if $E(X_t \mid \mathcal{F}_s) \geq X_s$ a.e., and a *supermartingale* if $E(X_t \mid \mathcal{F}_s) \leq X_s$ a.e.

Remarks 1.2.1. (a) If $\{X_t, \mathcal{F}_t\}$ is a martingale, then $E(X_t)$ is constant, increases in a submartingale, and decreases in a supermartingale.

(b) $\{X_t, \mathcal{F}_t\}$ is a martingale iff

$$\int_A X_t \, dP = \int_A X_s \, dP \qquad \text{for all } s \leq t \text{ and } A \in \mathcal{F}_s.$$

Similarly, $\{X_t, \mathcal{F}_t\}$ is a submartingale iff

$$\int_A X_t \, dP \leq \int_A X_s \, dP \qquad \text{for all } s \leq t \text{ and } A \in \mathcal{F}_s,$$

and a supermartingale iff

$$\int_A X_t \, dP \geq \int_A X_s \, dP \qquad \text{for all } s \leq t \text{ and } A \in \mathcal{F}_s.$$

(c) The set T has the interpretation of time. For discrete time we have that if $\{X_n, \mathcal{F}_n\}$ is a martingale, then

$$E(X_{n+k} \mid \mathcal{F}_n) = X_n, \qquad n, k = 1, 2, \ldots,$$

(with corresponding statements for sub- and supermartingales). For

$$\begin{aligned} E(X_{n+2} \mid \mathcal{F}_n) &= E[E(X_{n+2} \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n] \\ &= E(X_{n+1} \mid \mathcal{F}_n) \\ &= X_n. \end{aligned}$$

The general statement follows by induction.

Example 1.2.1. Let $Y_1, Y_2, \ldots$ be independent random variables and set $X_n = \sum_{k=1}^{n} Y_k$ and $\mathcal{F}_n = \mathcal{F}(Y_1, \ldots, Y_n)$. Then,

$$\begin{aligned} E(X_{n+1} \mid \mathcal{F}_n) =& E(X_n + Y_{n+1} \mid Y_1, \ldots, Y_n) \\ =& X_n + E(Y_{n+1} \mid Y_1, \ldots, Y_n) \\ & \text{since } X_n \text{ is } \mathcal{F}_n - \text{measurable} \\ =& X_n + E(Y_{n+1}) \\ & \text{by independence.} \end{aligned}$$

Thus, $\{X_n, \mathcal{F}_n\}$ is a martingale if $E(Y_n) = 0$, a submartingale if $E(Y_n) \geq 0$ and a supermartingale if $E(Y_n) \leq 0$, for all $n = 1, 2 \ldots$.

Example 1.2.2. Let $\{X_n : n = 0, 1, \ldots\}$ be a martingale and let Ψ be a convex function for which each $\Psi(X_n)$ is integrable. Then $\{\Psi(X_n) : n = 0, 1, \ldots\}$ is a submartingale: the require almost everywhere inequality $E(\Psi(X_{n+1}) \mid \mathcal{F}_n) \geq \Psi(X_n)$, is a direct application of the conditional expectation form of Jensen's inequality.

The companion result for submartingales is: if the convex function Ψ is increasing, X_n is a submartingale and each $\Psi(X_n)$ is integrable, then $\{\Psi(X_n) : n = 0, 1, \ldots\}$ is a submartingale, because

$$E(\Psi(X_{n+1}) \mid \mathcal{F}_n) \geq \Psi(E(X_{n+1} \mid \mathcal{F}_n)) \geq \Psi(X_n).$$

Thus, if $r \geq 1$, $\{X_n\}$ is a martingale and $|X_n|^r$ is integrable for all n, then $\{|X_n|^r\}$ is a submartingale.

Two good examples to remember: if $\{X_n\}$ is a martingale and each X_n is square integrable, then $\{X_n^2\}$ is a submartingale; if $\{X_n\}$ is a submartingale then $\{X_n^+\}$ is also a submartingale.

We now introduce a concept that has important applications to martingale theory.

Definition 1.2.2. Let $\{X_t\}_{t\in T}$ be a family of random variables on the probability space $(\Omega, \mathcal{F}, P)$. This family is said to be *uniformly integrable* iff

$$\int_{\{|X_t|\geq c\}} |X_t|\, dP \longrightarrow 0 \qquad \text{as} \quad c \longrightarrow \infty,$$

uniformly in $t \in T$.

One basic application of uniform integrability is the following extension of Fatou's lemma and the dominated convergence theorem.

Theorem 1.2.1. *Let $f_1, f_2, \dots$ be a real-valued and uniformly integrable family of random variables.*

(a)

$$\begin{aligned}\int_\Omega \liminf_n f_n\, dP &\leq \liminf_n \int_\Omega f_n\, dP \\ &\leq \limsup_n \int_\Omega f_n\, dP \leq \int_\Omega \limsup_n f_n\, dP.\end{aligned}$$

(b) *If $f_n \longrightarrow f$ a.e., then f is integrable and*

$$\int_\Omega f_n\, dP \longrightarrow \int_\Omega f\, dP.$$

Proof. (a) We have

$$\int_\Omega f_n\, dP = \int_{\{f_n<-c\}} f_n\, dP + \int_{\{f_n\geq -c\}} f_n\, dP, \quad c > 0.$$

By uniform integrability, c may be chosen so large that $\left|\int_{\{f_n<-c\}} f_n\, dP\right| < \epsilon$ for all n, where $\epsilon > 0$ is preassigned. Since $f_n I_{f_n \geq -c} \geq -c$, which is integrable since P is finite, Fatou's lemma yields

$$\liminf_n \int_{\{f_n\geq -c\}} f_n\, dP \geq \int_\Omega \liminf_n (f_n I_{\{f_n\geq -c\}})\, dP.$$

Since $f_n I_{\{f_n \geq -c\}} \geq f_n$, this integral is in turn greater than or equal to $\int \liminf_n f_n \, dP$. Thus

$$\liminf_n \int_\Omega f_n \, dP \geq \int_\Omega \liminf_n f_n \, dP - \epsilon,$$

proving the "lim inf" part. The "lim sup" part is done by a symmetric argument.

(b) This is immediate from (a) if $f_n \longrightarrow f$ a.e.

□

Theorem 1.2.2. *The family of random variables $\{X_t\}$ is uniformly integrable iff the integrals $\int_\Omega |X_t| \, dP$ are uniformly bounded i.e. $\sup_t \int_\Omega |X_t| \, dP < \infty$ and also uniformly continuous, that is, $\int_A |X_t| \, dP \longrightarrow 0$ as $P(A) \longrightarrow 0$, uniformly in $t \in T$.*

Proof. Assume the integrals are uniformly bounded and uniformly continuous. Then

$$P\{|X_t| \geq c\} \leq \frac{1}{c} \int_\Omega |X_t| \, dP$$

by Chebyshev's inequality, and this approaches 0 as $c \longrightarrow \infty$, uniformly in n, by the uniform boundedness. Thus $\int_{\{|X_t| \geq c\}} |X_t| \, dP \longrightarrow 0$ as $c \longrightarrow \infty$, uniformly in t, by uniform continuity.

Conversely, assume uniform integrability. We have

$$\begin{aligned}\int_A |X_t| \, dP &= \int_{A \cap \{|X_t| \geq c\}} |X_t| \, dP + \int_{A \cap \{|X_t| < c\}} |X_t| \, dP \\ &= \int_{\{|X_t| \geq c\}} |X_t| \, dP + cP(A). \qquad (1.2.1)\end{aligned}$$

Choose c so that $\int_{\{|X_t| \geq c\}} |X_t| \, dP < \epsilon/2$ for all t; if $P(A) < \epsilon/2c$, then by (1.2.1), $\int_A |X_t| \, dP < (\epsilon/2) + (\epsilon/2) = \epsilon$ for all t, proving uniform continuity. To verify uniform boundedness, if $\epsilon > 0$, we have

$$\int_\Omega |X_t| \, dP = \int_{\{|X_t| \geq c\}} |X_t| \, dP + \int_{\{|X_t| < c\}} |X_t| \, dP \leq \epsilon + cP(\Omega),$$

for large t. Then, $\sup_t \int_\Omega |X_t| \, dP < \infty$. □

Theorem 1.2.3. *Let $\{X_n\} \subseteq L^p(\Omega, \mathcal{F}, P)$, $0 < p < \infty$, and $X_n \longrightarrow X$ in probability. Then, the following statements are equivalent.*

(a) $\{|X_n|^p\}$ *is uniformly integrable.*

(b) $X_n \longrightarrow X$ *in* L^p.

(c) $E\,|X_n|^p \longrightarrow E\,|X|^p$.

Also, if $|X_n| \leq Y$, where Y is integrable, in particular, if $\{X_n\}$ is uniformly bounded, then $\{X_n\}$ is uniformly integrable.

Proof. (a) $\Rightarrow$ (b) First assume that $\{|X_n - X|^p\}$ is uniformly integrable. Since $X_n \xrightarrow{P} X$, there is a subsequence X_{n_k} converging to X a.e., which in turn implies that $|X_{n_k} - X|^p \longrightarrow 0$ a.e. Then, by Theorem 1.2.1(b) $\int |X_{n_k} - X|^p \longrightarrow 0$ as $k \longrightarrow \infty$. The same argument shows that any subsequence of $\{X_n\}$ has a subsequence converging to X in L^p. Hence, $X_n \xrightarrow{L^p} X$, because if this would not be the case, there would be an $\epsilon > 0$ and a subsequence $\{X_{n_i}\}$ such that $\int |X_{n_i} - X|^p \geq \epsilon$ for all i.

Now assume that $\{|X_n|^p\}$ is uniformly integrable. We have $|X_n - X|^p \leq |X_n|^p + |X|^p$ if $p \leq 1$, and $|X_n - X|^p \leq 2^{p-1}(|X_n|^p + |X|^p)$ if $p \geq 1$. As above, we have a subsequence $X_{n_k} \rightarrow X$ a.e. By Theorem 1.2.1(b) $|X|^p$ is integrable, and it follows that the integrals $\int_\Omega |X_n - X|^p \, dp$ and $\int (|X_n|^p + |X|^p) \, dp$ are uniformly bounded and uniformly continuous, by uniform integrability. Therefore, $\{|X_n - X|\}^p$ is uniformly integrable, and the previous argument applies. Here we have used the above Theorem 1.2.2.

(b) $\Rightarrow$ (c) We have by Minkowski's inequality, since $X_n = X + (X_n - X)$ and $X = X_n + (X - X_n)$ that

$$E(|X_n|^r)^{1/r} \leq E(|X|^r)^{1/r} + E(|X_n - X|^r)^{1/r}$$

and

$$E(|X|^r)^{1/r} \leq E(|X_n|^r)^{1/r} + E(|X_n - X|^r)^{1/r}$$

letting $n \longrightarrow \infty$ we obtain the desired result.

(c) $\Rightarrow$ (a) Let $Y_n = |X_n|^p$ and $Y = |X|^p$, thus $E(Y_n) \longrightarrow E(Y)$ by hypothesis. Then $|Y_n - Y| = (Y_n \vee Y) - (Y_n \wedge Y)$ and $Y_n + Y = (Y_n \vee Y) + (Y_n \wedge Y)$. By hypothesis $E(Y_n + Y) \longrightarrow 2E(Y)$, and by the *extension of the dominated convergence theorem* (just changing the hypothesis $Y_n \xrightarrow{\text{a.e.}} Y$ by $Y_n \xrightarrow{P} Y$), $E(Y_n \wedge Y) \longrightarrow E(Y)$. Hence $E(Y_n \vee Y) \longrightarrow E(Y)$, so $E\,|Y_n - Y| \longrightarrow E(Y) - E(Y) = 0$. Thus $Y_n \longrightarrow Y$ in L^1, then, for $\epsilon > 0$ and some n_0 we have

$$\int |Y_n|\, dP \leq \int |Y_n - Y|\, dP + \int |Y|\, dP < \epsilon/2 + \int |Y|\, dP \qquad \forall n > n_0.$$

Hence $\int |Y_n|\, dP \leq \max\left\{\epsilon/2 + \int |Y|\, dP, \int |Y_1 - Y|\, dP, \ldots, \int |Y_{n_0} - Y|\, dP\right\}$, that is the $\{Y_n\}$ are uniformly bounded.

Analogously, there exist n_0 and $\delta > 0$ such that

$$\begin{aligned} \int_A |Y_n|\, dP &\leq \int_A |Y_n - Y|\, dP + \int_A |Y|\, dP \\ &\leq \epsilon/2 + \epsilon/2 \qquad \text{for } n > n_0 \text{ and } P(A) < \delta. \end{aligned}$$

Then, the random variables $\{Y_n\}_{n>n_0}$ are uniformly continuous. For the other set of random variables, $\{Y_k\}_{k\leq n_0}$, we have that there exist $\delta_1, \ldots, \delta_{n_0}$ such that

$$\int_A |Y_k|\, dP < \epsilon/2 \qquad \text{if} \quad P(A) < \delta_k \qquad 1 \leq k \leq n_0.$$

If $\delta = \min\{\delta_1, \ldots, \delta_{n_0}\}$, then $\int_A |Y_k|\, dP < \epsilon/2$ for $1 \leq k \leq n_0$. Therefore, both $\{Y_k\}_{k>n_0}$ and $\{Y_k\}_{k\leq n_0}$ are uniformly continuous. Hence the $\{Y_n\}_{n>1}$ are uniformly continuous. Accordingly, by theorem 1.2.2 the $\{Y_n\}_{n>1}$ are uniformly integrable. Therefore the $\{|X_n|^p\}$ are uniformly integrable. □

Theorem 1.2.4. ***Doob's Optional Stopping Theorem.*** *Let $\{X_n, \mathcal{F}_n\}$ be a supermartingale and T_1 and T_2 be stopping times for $\{\mathcal{F}_n\}$ such that $T_1 \leq T_2$. If T_2 is bounded or if $\{X_n\}$ is uniformly integrable, then*

$$E(X_{T_2} \mid \mathcal{F}_{T_1}) \leq X_{T_1}.$$

If X_n is martingale, we have equality.

Remarks 1.2.2. The theorem does not hold for arbitrary stopping times. For instance, consider the problem of fair coin tossing, that is, let $Y_1, Y_2, \ldots$ be independent random variables taking values ± 1 with equal probability. If $X_n = Y_1 + \cdots + Y_n$, X_n is a martingale because $E(X_n) = 0$ for all n (by example 1.2.1). Let a, b be integers with $0 < a < b$, $T_1 = \inf\{n : X_n = a\}$ and $T_2 = \inf\{n : X_n = b\}$. Then $T_1 \leq T_2$ (the basic idea is: in order to have 10 pesos first we have to have 7 pesos). $X_{T_1} = a$ and $X_{T_2} = b$, therefore $E(X_{T_2} \mid \mathcal{F}_{T_1}) = b \neq X_{T_1}$. The problem is that T_1 and T_2 are not bounded stopping times and $\{X_n\}$ is not uniformly integrable.

Theorem 1.2.5. ***Wald's formula on the sum of a random number of random variables.*** *Let $Y_1, Y_2, \ldots$ be independent, identically distributed random variables with finite mean m, and let $X_n = \sum_{k=1}^{n} Y_k$. If T is a finite stopping time for $\{X_n\}$, the following hold:*

(a) *If all $Y_j \geq 0$, then $E(X_T) = mE(T)$.*

(b) *If $E(T) < \infty$, then $E|X_T| < \infty$ and $E(X_T) = mE(T)$.*

(c) *If T is a positive integer-valued random variable independent of $(Y_1, Y_2, \ldots)$, but not necessarily a stopping time, then (a) and (b) still hold.*

Proof. (a) Since $X_n - nm = \sum_{k=1}^{n}(Y_k - m)$, and $E(Y_k - m) = 0$, $\{X_n - nm\}$ is a martingale. Applying Doob's Optional Stopping Theorem to the stopping times $T_n = T \wedge n$ and $T_1 = 1$, we have $E(X_{T_n} - T_n m \mid \mathcal{F}_1) = X_1 - m$, that is, $E(X_{T_n} - T_n m) = E(X_1 - m) = 0$; hence $E(X_{T_n}) = mE(T_n)$. Since $Y_j \geq 0, X_{T_n} = \sum_{k=1}^{T_n} Y_k \uparrow \sum_{k=1}^{T} Y_k = X_T$, as $n \longrightarrow \infty$, by the monotone convergence theorem, $E(X_T) = mE(T)$.

(b) We write $X_n = \sum_{k=1}^{T_n} Y_k^+ - \sum_{k=1}^{T_n} Y_k^- = X_n' - X_n''$. By (a), $E(X_T') = E(Y_1^+)E(T)$, $E(X_T'') = E(Y_1^-)E(T)$. Since $E(T)$ is finite, so are $E(X_T')$ and $E(X_T'')$; hence $E\,|X_T| < \infty$ and

$$E(X_T) = E(X_T') - E(X_T'') = [E(Y_1^+) - E(Y_1^-)]E(T) = mE(T).$$

(c) To prove (a), observe that if all $Y_j \geq 0$, then

$$\begin{aligned} E(X_T) &= \sum_{n=1}^{\infty} E(X_n I_{\{T=n\}}) \\ &= \sum_{n=1}^{\infty} E(X_n) P\{T = n\} \\ &= m \sum_{n=1}^{\infty} nP\{T = n\} \\ &= mE(T). \end{aligned}$$

Part (b) is proved just as above. □

1.3 Martingale Convergence Theorems

Let $\{X_i, \mathcal{F}_i : i = 0, 1, \ldots\}$ be a martingale. For fixed real numbers α and β, with $\alpha < \beta$, define the increasing sequences of random times at which the process might drop below α or arise above β:

$$T_0 := \inf\{i \geq 0 : X_i \leq \alpha\}, \qquad T_1 := \inf\{i \geq T_0 : X_i \geq \beta\},$$

$$T_2 := \inf\{i \geq T_1 : X_i \leq \alpha\}, \qquad T_3 := \inf\{i \geq T_2 : X_i \geq \beta\},$$

$$\vdots$$

$$T_{2k} := \inf\{i \geq T_{2k-1} : X_i \leq \alpha\}, \qquad T_{2k+1} := \inf\{i \geq T_{2k} : X_i \geq \beta\},$$

with the convention that the infimum of an empty set is taken as $+\infty$. Because $\{X_i\}$ is adapted to $\{\mathcal{F}_i\}$, each T_r is a stopping time for the filtration. For example,

$$\{T_1 \leq k\} = \{X_i \leq \alpha, X_j \geq \beta \text{ for some } i \leq j \leq k\},$$

which could be written out explicitly as a finite union of events involving only $X_0, \ldots, X_k$.

T_{2k-1} is the time of the kth upcrossing of the interval $[\alpha, \beta]$ by the process $\{X_n\}$. Set

$$U_\alpha^\beta = \begin{cases} \sup\{k : T_{2k-1} < \infty\} & \text{if } T_1 < \infty, \\ 0 & \text{if } T_1 = \infty. \end{cases}$$

U_α^β is the total number of upcrossings of $[\alpha, \beta]$ made by $\{X_n\}$.

Lemma 1.3.1. *A sequence of real numbers $\{x_n\}$ converges (the limit can be $+\infty$ or $-\infty$) iff $U_\alpha^\beta < \infty$, $\forall\, \alpha, \beta \in \mathbb{Q}$ with $\alpha < \beta$.*

Proof. If for some $\alpha < \beta$, $U_\alpha^\beta = \infty$, then there exist sequences $n_k \longrightarrow \infty$ and $m_k \longrightarrow \infty$ such that $x_{n_k} \leq \alpha$ and $x_{n_k} \geq \beta\ \forall\, k$. Hence, $\limsup_n x_n \geq \limsup_k x_{n_k} \geq \beta > \alpha \geq \liminf_k x_{m_k} \geq \liminf_n x_n$. That is, x_n does not converge to a finite or infinite limit.

Conversely, if the x_n does not converge i.e. $\limsup_n x_n > \liminf_n x_n$, then for some $\alpha, \beta \in \mathbb{Q}$ with $\alpha < \beta$, $\limsup_n x_n \geq \alpha > \beta \geq \liminf_n x_n$, thus x_n has an infinite number of upcrossings of $[\alpha, \beta]$, i.e. $U_\alpha^\beta = \infty$. □

Lemma 1.3.2. ***Doob's Upcrossing Lemma.*** *Let $\{X_n\}$ be a supermartingale. Let $U_\alpha^\beta(w)$ be the number of upcrossings of $[\alpha, \beta]$ made by $n \mapsto X_n(w)$. Then U_α^β is a random variable that can take the value $+\infty$ with*

$$E(U_\alpha^\beta) \leq \frac{1}{\beta - \alpha}(\sup_n E\,|X_n| + |\alpha|).$$

Theorem 1.3.3. ***Doob's Convergence Theorem.*** *Let $\{X_n, \mathcal{F}_n\}$ be a supermartingale bounded in L^1, i.e. $\sup_n E\,|X_n| < \infty$. Then almost everywhere, $X_\infty := \lim X_n$ exists and is finite.*

Proof. From the above lemma we have for each pair $\alpha, \beta \in \mathbb{Q}$ with $\alpha < \beta$, $U_\alpha^\beta < \infty$ a.e. since its expectation is finite. Hence $U_\alpha^\beta < \infty$ a.e. $\forall\, \alpha, \beta \in \mathbb{Q}, \alpha < \beta$. Now $U_\alpha^\beta < \infty$ a.e. implies there exists $\Omega(\alpha, \beta)$ such that $U_\alpha^\beta(w) < \infty \;\; \forall\, w \in \Omega(\alpha, \beta)$ and $P(\Omega(\alpha, \beta)) = 1$. Thus $\Lambda := \bigcap_{\substack{\alpha, \beta \in \mathbb{Q} \\ \alpha < \beta}} \Omega(\alpha, \beta)$ has probability 1. To see this we compute

$$\begin{aligned} P(\Lambda^c) &= P(\bigcup_{\substack{\alpha, \beta \in \mathbb{Q} \\ \alpha < \beta}} \Omega^c(\alpha, \beta)) \\ &= \sum_{\substack{\alpha, \beta \in \mathbb{Q} \\ \alpha < \beta}} P(\Omega^c(\alpha, \beta)) \\ &= 0. \end{aligned}$$

This shows that the $X_n(w)$ converges a.e. On the other hand,

$$\begin{aligned} E\,|\lim X_n| &\leq \liminf E\,|X_n| \quad \text{by Fatou's lemma} \\ &\leq \sup E\,|X_n| \\ &< \infty \qquad\qquad\quad \text{by hypothesis.} \end{aligned}$$

Therefore, $\lim X_n \in L^1$. □

Theorem 1.3.4. *Let $\xi_1, \xi_2 \ldots$ be independent random variables with finite expectation. If $\sum_{n=1}^{\infty} Var\,(\xi_n) < \infty$, then $\sum_{n=1}^{\infty}[\xi_n - E(\xi_n)]$ converges a.e. to an integrable random variable.*

Proof. We may assume that $E(\xi_n) \equiv 0$. Let $X_n = \xi_1 + \cdots + \xi_n$ and $\mathcal{F}_n = \mathcal{F}(\xi_1, \ldots, \xi_n)$. Then, $\{X_n, \mathcal{F}_n\}$ is a martingale bounded in L^1, which can be verified by using the *Cauchy-Schwarz* inequality,

$$\begin{aligned} \sup_n \; E\,|X_n| &\leq \sup_n \; [E\,|X_n|^2]^{1/2} \\ &= \sup_n \; [\sum_{k=1}^{n} E(\xi_k^2)]^{1/2} \\ &\leq \sup_n \; [\sum_{k=1}^{\infty} Var\,(\xi_k)]^{1/2} \\ &< \infty. \end{aligned}$$

Thus by Doob's Convergence Theorem we arrive at the desired conclusion. □

Example 1.3.1. Consider again the problem of fair coin tossing known also as the gamblers ruin problem. We have $Y_1, Y_2, \ldots$ independent random variables taking values ± 1 with equal probability. Define $X_0 = 0$, and let $X_n = Y_1 + \cdots Y_n$ (random walk). Let a and b be positive integers and set $T = \inf\{n : X_n = -a \text{ or } X_n = b\}$ and $T_a = \inf\{n : X_n = -a\}$. Taking for granted the fact that $T_a < \infty$ a.e., we can compute $P\{X_T = a\}$ and $P\{X_T = b\}$.

The process $\{X_n, \mathcal{F}_n\}$ is a martingale, with $\mathcal{F}_n = \mathcal{F}(X_1, \ldots, X_n)$, and then by the Doob's optional stopping theorem, $E(X_{T\wedge n}) = E(X_0) = 0$. On the other hand $T < \infty$ a.e. since $T < T_a < \infty$ a.e., therefore $|X_T - X_{T\wedge n}| = \left|\sum_{k=T\wedge n+1}^{T} Y_k\right| \Longrightarrow 0$ as $n \longrightarrow \infty$, since $T \wedge n \longrightarrow T$. Hence $X_{T\wedge n} \longrightarrow X_T$. Next, by definition of T, $|X_{T\wedge n}| \leq a \vee b$ and, by the dominated convergence theorem, $E(X_{T\wedge n}) \longrightarrow E(X_T)$. Therefore $0 = E(X_T) = -aP(X_T = -a) + bP(X_T = b)$ and $P(X_T = -a) + P(X_T = b) = 1$. From these equations we get

$$P(X_T = b) = \frac{a}{a+b},$$
$$P(X_T = -a) = \frac{b}{a+b}.$$

Now, by Wald's formula we might have that $mE(T) = E(X_T)$, however it does not work here since $m = E(X_T) = 0$ and we don't know whether $E(T)$ is finite. So we proceed in another fashion. We have that $\{X_n^2 - n, \mathcal{F}_n\}$ is a martingale, since

$$\begin{aligned}
E(X_{n+1}^2 - (n+1) \mid \mathcal{F}_n) &= E[(X_n + Y_{n+1})^2 - (n+1) \mid \mathcal{F}_n] \\
&= E[(X_n^2 + 2X_nY_{n+1} + Y_{n+1}^2 - (n+1) \mid \mathcal{F}_n] \\
&= X_n^2 + 2X_nE[Y_{n+1} \mid \mathcal{F}_n] + E[Y_{n+1}^2 \mid \mathcal{F}_n] - (n+1) \\
&= X_n^2 + 2X_nE[Y_{n+1}] + E[Y_{n+1}^2] - (n+1) \\
&= X_n^2 - n,
\end{aligned}$$

because $E[Y_{n+1}] = 0$ and $E[Y_{n+1}^2] = 1$.

Now, by the Doob's optional stopping theorem, $E(X_{T\wedge n}^2 - T\wedge n) = E(X_0^2 - 0) = 0$. Therefore $E(X_{T\wedge n}^2) = E(T\wedge n)$. Next, $X_{T\wedge n}^2 \longrightarrow X_T^2$ and $X_{T\wedge n}^2 \leq a^2 \vee b^2$, thus by the dominated convergence theorem $E(X_T^2) = E(T)$. But $E(X_T^2) = a^2 \frac{b}{a+b} + b^2 \frac{b}{a+b} = ab$. Hence, $E(T) = ab$.

Remarks 1.3.1. (a) The same results hold for the Brownian motion for $a, b \in \mathbb{R}$ with $a > b > 0$. (b) X_n and $X_n^2 - n$ are martingales, and for the Brownian motion W_t and $W_t^2 - t$ are also martingales.

Example 1.3.2. Let $\{\mathcal{F}_n : n = 0, 1, \ldots\}$ be a filtration on $(\Omega, \mathcal{F}, P)$, X a random variable in L^1 and define $X_n := E(X \mid \mathcal{F}_n)$. Then X_n is a martingale. For

$$\begin{aligned} E(X_{n+1} \mid \mathcal{F}_n) &= E[E(X \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n] \\ &= E[X \mid \mathcal{F}_n] \\ &= X_n. \end{aligned}$$

One interesting question is: would it be true that all the martingales are of this form? That is, given $\{X_n, \mathcal{F}_n\}$ a martingale, does there exist X a random variable in L^1 such that $X_n = E(X \mid \mathcal{F}_n)$?

Lemma 1.3.5. *Let Y be an integrable random variable on $(\Omega, \mathcal{F}, P)$ and let $\mathcal{F}_i, i \in I$, be arbitrary sub σ-fields of $\mathcal{F}$. Then the family $\{X_i = E(Y \mid \mathcal{F}_i), i \in I\}$, is uniformly integrable, that is,*

$$\int_{\{|X_i| \geq c\}} |X_i| \, dP \longrightarrow 0 \qquad as \qquad c \longrightarrow \infty,$$

uniformly in $i \in I$.

Proof. Since $|E(Y \mid \mathcal{F}_i)| \leq E(|Y| \mid \mathcal{F}_i)$,

$$\int_{\{|X_i| \geq c\}} |X_i| \, dP \leq \int_{\{|X_i| \geq c\}} E(|Y| \mid \mathcal{F}_i) \, dP = \int_{\{|X_i| \geq c\}} |Y| \, dP$$

since $\{|X_i| \geq c\} \in \mathcal{F}_i$ (remember that X_i is $\mathcal{F}_i$-measurable). Now, by Chebyshev's inequality,

$$P\{|X_i| \geq c\} \leq c^{-1}E(|X_i|) \leq c^{-1}E[E(|Y| \mid \mathcal{F}_i)] = c^{-1}E(|Y|) \longrightarrow 0 \quad \text{as } c \longrightarrow \infty$$

uniformly in i. □

Theorem 1.3.6. *Let $\{X_n, \mathcal{F}_n\}$ be a martingale. The following statements are equivalent.*

a) *$\{X_n, \mathcal{F}_n\}$ is uniformly integrable.*

b) *$\{X_n\}$ converges in L^1.*

c) *There exists a random variable $\xi \in L^1$ such that $X_n = E(\xi \mid \mathcal{F}_n)$.*

In each one of these cases $X_n \longrightarrow X_\infty = E(\xi \mid \mathcal{F}_\infty)$ a.e., where $\mathcal{F}_\infty$ is the σ-field generated by $\bigcup_{n=1}^{\infty} \mathcal{F}_n$ and X_∞ is the unique random variable $\mathcal{F}_\infty$-measurable such that $X_n = E(X_\infty \mid \mathcal{F}_n)$ for all n.

Proof. (a) $\Rightarrow$ (b). By Theorem 1.2.2, $\sup_n E\,|X_n| < \infty$. Then by Doob's Convergence Theorem, $X_n \longrightarrow X_\infty$ a.e. Since P is a finite measure, $X_n \longrightarrow X_\infty$ in probability and the $|X_n|$ are uniformly integrable by hypothesis, thus by theorem 1.2.3, $X_n \xrightarrow{L^1} X_\infty$.

(b) $\Rightarrow$ (c). Let $X_n \xrightarrow{L^1} X_\infty$. If $A \in \mathcal{F}_n$ and $k \geq n$, then $\int_A X_n\,dP = \int_A X_k\,dP$ by Remarks 1.2.1(b) and (c). Letting $k \longrightarrow \infty$, the L^1-convergence yields $\int_A X_n\,dP \longrightarrow \int_A X_\infty\,dP$. Thus, $X_n = E(X_\infty \mid \mathcal{F}_n)$ and the result follows with $\xi := X_\infty$.

(c) $\Rightarrow$ (a). It follows from the above lemma. □

Finally, we give a martingale proof of the Kolmogorov Zero-One Law, but first we give the following

Definition 1.3.1. Let $X_1, X_2, \ldots$ be a sequence of random variables, and let $\mathcal{F}_n = \mathcal{F}(X_n, X_{n+1}, \ldots)$, $n = 1, 2, \ldots$. The σ-field $\mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \mathcal{F}_n$ is called the tail σ-field of X_n, and elements of $\mathcal{F}_\infty$ are called *tail events.*

Theorem 1.3.7. *All tail events relative to a sequence of independent random variables have probability 0 or 1.*

Proof. Let $X_1, X_2, \ldots$ be a sequence of independent random variables, and let $\mathcal{H}_n = \sigma(X_1, \ldots, X_n)$, $n = 1, 2, \ldots$, and $A \in \mathcal{F}_\infty$. I_A is $\sigma(\bigcup_{n=1}^{\infty} \mathcal{H}_n)$-measurable. Then $E(I_A \mid \mathcal{H}_n) \xrightarrow{L^1} I_A$ because $E(I_A \mid \mathcal{H}_n) \xrightarrow{L^1} E(I_A \mid \sigma(\bigcup_{n=1}^{\infty} \mathcal{H}_n)) = I_A$ by the comment following Theorem 1.3.6. Now, since $A \in \mathcal{F}_{n+1}$, A can be written in the form $\{(X_{n+1}, X_{n+2}, \ldots) \in A_{n+1}\}$, $A_{n+1} \in \mathcal{B}(R)^\infty$, it follows that A and $\mathcal{H}_n$ are independent, hence $E(I_A \mid \mathcal{H}_n) = P(I_A \mid \mathcal{H}_n) = P(A)$. Consequently $P(A) = I_A$, that is, $P(A) = 0$ or 1. □

Bibliography

[1] Robert B. Ash, *Real analysis and probability*, Academic Press, Inc., 1972.

[2] Patrick Billingsley, *Probability and measure*, John Wiley Sons, New York, 1986.

[3] Kai Lai Chung, *A course in probability theory*, Academic Press, Inc., New York, 1974.

[4] David Pollard, *A user's guide to measure theoretic probability*, Cambridge University Press, 2002.

[5] David Williams, *Probability with martingales*, Cambridge University Press, 1991.