DISCRETE-TIME MARTINGALES

STEVEN P. LALLEY

1. Discrete-Time Martingales

1.1. **Definition of a Martingale.** Let $\{\mathscr{F}_n\}_{n\geq 0}$ be an increasing sequence of $\sigma-$algebras in a probability space $(\Omega, \mathscr{F}, P)$. Such a sequence will be called a *filtration*. 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 $E|X_n| < \infty$. The sequence $X_0, X_1, \dots$ is said to be a *martingale* relative to the filtration $\{\mathscr{F}_n\}_{n\geq 0}$ if it is adapted and if for every n,

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

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

$$E(X_{n+1} \,|\, \mathscr{F}_n) \leq (\geq) X_n. \tag{2}$$

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

1.2. **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 such a sequence; then the sequence of partial sums

$$X_n = \sum_{j=1}^{n} Y_j \tag{3}$$

is a martingale 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:

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

The importance of martingales in modern probability theory stems at least in part from the fact that many 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:

$$X_n = X_0 + \sum_{j=1}^{n} \xi_j \quad \text{where} \quad \xi_j = X_j - X_{j-1}. \tag{4}$$

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* $n \geq 0$*,*

$$E(\xi_{n+1} \mid \mathscr{F}_n) = 0. \tag{5}$$

Proof. This is a trivial consequence of 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* $n \geq 0$*,*

$$EX_n = EX_0. \tag{6}$$

Moreover, if $EX_n^2 < \infty$ *for some* $n \geq 1$ *then for* $j \leq n$ *the random variables* ξ_j *are square-integrable and uncorrelated, and so*

$$EX_n^2 = EX_0^2 + \sum_{j=1}^{n} E\xi_j^2. \tag{7}$$

Proof. The first property follows easaily 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 (6).

The second property is only slightly more difficult. For ease of exposition let's assume that $X_0 = 0$. (The general case can then be deduced by re-indexing the random variables.) First, observe that for each $k \leq n$ the random variable X_k is square-integrable, by the Jensen inequality for conditional expectation, since $X_k = E(X_n|\mathscr{F}_k)$. Hence, each of the terms ξ_j has finite variance, because it is the difference of two random variables with finite second moments, and so 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 $j \leq k \leq n$ then

$$\begin{aligned} E\xi_j\xi_{k+1} &= EE(\xi_j\xi_{k+1} \mid \mathscr{F}_k) \\ &= E\xi_j E\xi_{k+1} \mid \mathscr{F}_k) \\ &= E(\xi_j \cdot 0) = 0. \end{aligned}$$

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

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

□

1.3. **Some Examples of Martingales.**

1.3.1. *Paul Lévy's Martingales.* Let X be an integrable $\mathscr{F}-$measurable random variable, and let $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \cdots$ be an increasing sequence of $\sigma-$algebras (i.e., a *filtration*) all contained in $\mathscr{F}$. Then the sequence X_n defined by $X_n = E(X|\mathscr{F}_n)$ is a martingale, by the Tower Property of conditional expectation.

1.3.2. *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 with respect to the standard filtration (i.e., where $\mathscr{F}_n$ is the $\sigma-$algebra generated by the random variables $Y_1, Y_2, \cdots Y_n$).

1.3.3. *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 assume now that $EY_n^2 = \sigma^2 < \infty$. Then the sequence

$$\left(\sum_{j=1}^{n} Y_j\right)^2 - \sigma^2 n \tag{8}$$

is a martingale (again relative to the standard filtration). This is easy to check.

1.3.4. *Likelihood Ratio Martingales.* 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

$$Z_n = Z_n(\theta) = \prod_{j=1}^{n} \frac{e^{\theta X_j}}{\varphi(\theta)} = \frac{e^{\theta S_n}}{\varphi(\theta)^n}. \tag{9}$$

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.)

1.3.5. *Likelihood Ratio Martingales in General.* Let $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \cdots$ be a filtration of a measurable space $(\Omega, \mathscr{F})$, and suppose that P and Q are two probability measure on $\mathscr{F}$ such that for each $n = 0, 1, 2, \cdots$ the probability measure Q is absolutely continuous relative to P *on the* σ*−algebra* $\mathscr{F}_n$. For each n let

$$Z_n := \left(\frac{dQ}{dP}\right)_{\mathscr{F}_n}$$

be the Radon-Nikodym derivative of Q with respect to P on the $\sigma-$algebra $\mathscr{F}_n$. Then the sequence $\{Z_n\}_{n\geq 0}$ is a martingale (under the probability measure P). Proof: homework.

NOTE: It is *not* required here that Q be absolutely continuous with respect to P on the $\sigma-$algebra $\mathscr{F}$. Here is an example: let Ω be the space of all infinite sequences $x_1 x_2 \cdots$ of 0s and 1s, let $\mathscr{F}_n$ be the $\sigma-$algebra generated by the first n coordinate variables, and let $\mathscr{F}$ be the $\sigma-$algebra generated by *all* coordinate variables (equivalently, the smallest $\sigma-$algebra containing the union $\cup_n \mathscr{F}_n$). Let P and Q be the product Bernoulli probability measures on $(\Omega, \mathscr{F})$ with parameters $p = \frac{1}{2}$ and $p = \frac{1}{3}$. Then Q and P are mutually absolutely continuous on each $\mathscr{F}_n$, with likelihood ratio

$$Z_n = 2^n \prod_{i=1}^{n} (1/3)^{x_i} (2/3)^{1-x_i},$$

but P and Q are mutually *singular* on $\mathscr{F}$, by the strong law of large numbers.

1.3.6. *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. Later. □

1.3.7. *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.

1.3.8. *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.[1] 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

$$\mathbb{P}h = h, \tag{10}$$

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

$$h(x) = \sum_{y \in \mathscr{Y}} p(x, y) h(y) = E^x h(Y_1). \tag{11}$$

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 (11) 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 $\mathscr{F}_n$ as a conditional expectation on Y_n. Thus,

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

□

1.3.9. *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

$$Z_n = \varphi(X_n) \tag{12}$$

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) $\varphi(x) = \exp\{\theta x\}$.

[1]We will talk more about Markov chains later in the course; if you don't yet know what a Markov chain is, ignore this example for now.

2. 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.

2.1. **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 $Y_0, Y_1, \dots$ (or to a filtration $\{\mathscr{F}_n\}_{n\geq 0}$). For $n = 1, 2, \dots$, define

$$\xi_n = X_n - X_{n-1}; \tag{13}$$

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

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

$$(Z \cdot X)_n = X_0 + \sum_{k=1}^{n} Z_k \xi_k. \tag{14}$$

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:

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

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

2.2. **Submartingale Transforms.** Submartingales and supermartingales may also be transformed, using equation (14), 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 2. 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

$$(Z \cdot X)_n = Z_0 X_0 + \sum_{k=1}^{n} Z_k \xi_k. \tag{15}$$

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* $n \geq 0$,

$$0 \leq Z_n \leq 1, \tag{16}$$

then

$$E(Z \cdot X)_n \leq EX_n. \tag{17}$$

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 (17). □

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

$$E(Z\cdot X)_n \geq EX_n. \tag{18}$$

3. Optional Stopping

3.1. **Doob's optional sampling theorem.** 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 martingale property, provided the decision about when to stop is based solely on information available up to τ. Such random times are called *stopping times.*[2]

Definition 3. 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. *Let* $\{X_n\}_{n\in\mathbb{Z}_+}$ *be a martingale (respectively submartingale or supermartingale) relative to a filtration* $\{\mathscr{F}_n\}_{n\geq 0}$*, and let* τ *be a stopping time. Then the stopped sequence* $\{X_{\tau\wedge n}\}_{n\geq 0}$ *is a martingale (respectively submartingale or supermartingale). Consequently, 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 transforms are martingales. The crucial fact is that the sequence $\{X_{\tau\wedge n}\}_{n\geq 0}$ may be represented as a transform of the sequence $\{X_n\}_{n\geq 0}$:

$$X_{\tau\wedge n} = (Z\cdot X)_n \tag{19}$$

where

$$Z_n = \begin{cases} 1 & \text{if } \tau \geq n, \qquad \text{and} \\ 0 & \text{if } \tau < n. \end{cases} \tag{20}$$

The equation (19) is easy to verify:

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

since the last sum telescopes. That the sequence Z_n is predictable follows from the hypothesis that τ is a stopping time: in particular, for any integer $n \geq 1$, the event $Z_n = 1$ coincides with the event $\tau \geq n$, which is $\mathscr{F}_{n-1}$–measurable, since this is the complement of the event $\tau \leq n-1$. It now follows immediately from Proposition 4 that if the original sequence X_n is a martingale

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

then so is the sequence $X_{\tau\wedge n}$. Similar arguments apply if the sequence X_n is a submartingale or a supermartingale. (Exercise: Fill in the details.) Finally, if X_n is a martingale then the fact that $X_{\tau\wedge n}$ is also a martingale implies the identity

$$EX_{\tau\wedge n} = EX_0,$$

since for any martingale expectation is a conserved quantity. The corresponding inequalities for submartingales and supermartingales follow from inequalities (15) and (18), respectively. □

3.2. **The Wald Identities.** Important special cases of Doob's Optional Sampling Theorem are the *Wald Identities.* These were formulated and proved by Wald in his development of *sequential statistical testing* in the late 1940s; they pre-date Doob's development of martingale theory by several years, and perhaps partially influenced it. Wald's identities involve partial sums of independent random variables; to simplify things we shall consider only the case where the summands are also *identically distributed.* Thus, assume now (and for the remainder of this section) that $\{\xi\}_{n\geq 1}$ are i.i.d. real random variables, and let $\mathscr{F}_n$ be the natural filtration. Let τ be a stopping time with respect to the filtration $\mathscr{F}_n$, and set

$$S_n = \sum_{k=1}^{n} \xi_k.$$

Theorem 2. *(Wald's First Identity) If* $E|\xi_i| < \infty$ *and* $E\tau < \infty$ *then* $E|S_\tau| < \infty$ *and*

$$ES_\tau = E\tau \cdot E\xi_1. \tag{21}$$

Proof. Assume first that the random variables ξ_i are nonnegative and that $E\xi_1 > 0$. For each $n < \infty$ the stopping time $\tau \wedge n$ is bounded, so Doob's identity implies that $ES_{\tau\wedge n} = E(\tau\wedge n)E\xi_1$. But since $\tau\wedge n \uparrow \tau$ and $S_{\tau\wedge n} \uparrow S_\tau$, the monotone convergence theorem applies on both sides of the identity, and so

$$ES_\tau = E\tau E\xi_1.$$

Since $E\tau < \infty$ by hypothesis, it follows that $ES_\tau < \infty$. This proves the result in the case of nonnegative summands.

Next, consider the general case. By hypothesis, $E|\xi_1| < \infty$, and so by what we have just shown in the case of nonnegative summands,

$$E\sum_{k=1}^{\tau} |\xi_k| < \infty.$$

Now the random variables $|S_{\tau\wedge n}|$ are all dominated by $\sum_{k=1}^{\tau} |\xi_k|$, and clearly $S_{\tau\wedge n} \to S_\tau$, so the dominated convergence theorem implies that

$$ES_\tau = \lim_{n\to\infty} ES_{\tau\wedge n} = \lim_{n\to\infty} E(\tau\wedge n)E\xi_1.$$

But the monotone convergence theorem implies that $E(\tau\wedge n) \uparrow E\tau$, so the theorem follows. □

Exercise 1. It is also possible to deduce Theorem 2 directly from the strong law of large numbers, without using the Doob theorem. Here is an outline. (a) Show that for any stopping time τ, the sequence $\{\xi'_k := \xi_{k+\tau}\}_{k\geq 1}$ consists of independent, identically distributed copies of ξ_1, and show that this sequence is independent of the stopping field $\mathscr{F}_\tau$. (b) Show that there is a stopping time τ' for the natural filtration of the sequence $\{\xi'_k\}_{k\geq 1}$ such that the joint distribution of

τ' and $\sum_{k=1}^{\tau'} \xi'_k$ is identical to that of τ and S_τ. (c) By induction, construct independent copies $(\tau^i, S^i_{\tau^i})$ of (τ, S_τ). (d) Use that strong law of large numbers on both coordinates to deduce that

$$ES_\tau = E\tau \cdot E\xi_1.$$

Theorem 3. *(Wald's Second Identity) If* $E\xi_i = 0$ *and* $\sigma^2 = E\xi_i^2 < \infty$ *then*

$$ES_\tau^2 = \sigma^2 E\tau. \tag{22}$$

Proof. Under the hypotheses of the theorem the sequence $S_n^2 - n\sigma^2$ is a martingale relative to the natural filtration. Hence, by Doob's Optional Sampling Formula,

$$ES_{\tau\wedge n}^2 = E(\tau \wedge n)\sigma^2.$$

The right side converges to $E\tau\sigma^2$, by the monotone converges theorem, and by Fatou's lemma, $ES_\tau^2 \leq \liminf ES_{\tau\wedge n}^2 = E\tau\sigma^2$. Consequently, $S_\tau \in L^2$. It follows, by another application of the dominated convergence theorem, that

$$\lim_{n\to\infty} ES_\tau^2 \mathbf{1}_{\{\tau>n\}} = 0.$$

Thus, to complete the proof it suffices to show that

$$\lim_{n\to\infty} ES_n^2 \mathbf{1}_{\{\tau>n\}} = 0.$$

For this, use the decomposition $S_\tau = S_n + (S_\tau - S_n)$ on the event $\tau > n$, and (exercise!) prove that

$$E((S_\tau - S_n)\mathbf{1}_{\{\tau>n\}} \,|\, \mathscr{F}_n) = 0.$$

HINT: Since τ is possibly unbounded, this is not trivial; you will need the fact that $S_\tau \in L^2$. □

Theorem 4. *(Wald's Third Identity) Assume that* $E\exp\{\theta\xi_1\} = \exp\{-\psi(\theta)\} < \infty$. *Then for every* bounded *stopping time,*

$$E\exp\{\theta S_\tau - \tau\psi(\theta)\} = 1. \tag{23}$$

Proof. Since τ is finite, this follows directly from Doob's theorem, as the sequence $\{\exp\{\theta S_n - n\psi(\theta)\}\}_{n\geq 0}$ is a martingale. □

In applications one would usually want to use the identity (23) for *unbounded* stopping times τ. Unfortunately, it is not easy to find useful sufficient conditions for the validity of (23), and so in most problems one must start with the identity for the truncated stopping times $\tau \wedge n$ and attempt to deduce the desired identity from the dominated convergence theorem or some other considerations. Example 3 below provides a simple example, and also shows that the identity (23) may fail for some unbounded stopping times.

3.3. **Examples.**

Example 1. Let $\{S_n\}_{n\geq 0}$ be simple random walk on the integers $\mathbb{Z}$ starting at $S_0 = 0$. Thus, the increments $\xi_n = S_n - S_{n-1}$ are independent, identically distributed Rademacher random variables, that is $\xi_n = \pm 1$ with probability $\frac{1}{2}$. Let $T = T_{[a,b]}$ be the first time that S_n exits the open interval (a, b), where $a < 0 < b$ are integers. It is easily established that the distribution of T has finite exponential moments, by S. Stein's trick (see below), and so $ET < \infty$. Hence, by Wald I,

$$ES_T = ET \cdot 0 = 0.$$

But S_T can take only two possible values, a or B, so it follows that

$$0 = a(1 - P\{S_T = b\}) + bP\{S_T = b\},$$

and therefore

$$\boxed{P\{S_T = b\} = \frac{-a}{b-a}.}$$

Furthermore, Wald II applies, and since we now have the distribution of the exit point S_T we can explicitly calculate $ES_T^2 = -ab$. THis gives

$$\boxed{ET = ES_T^2 = -ab.}$$

NOTE. Stein's trick is as follows. If there is any successive run of at least $b-a$ consecutive increments of $+1$ then the random walk must exit the interval (a,b) if it hasn't already done so. Now in any consecutive $b-a$ steps, the chance that every step is to the right (i.e., +1) is $2^{-(b-a)} > 0$. Thus, if time is broken into successive blocks of length $(b-a)$ then the number of blocks until the first block of $b-a$ consecutive $+1$ step has a geometric distribution, and therefore exponentially decaying tail. It follows that T must have an exponentially decaying tail.

Example 2. Fix $1 > p > 1/2$, and let $\{S_n\}_{n\geq 0}$ be the random walk with independent, identically distributed increments $\xi_n = S_n - S_{n-1}$ such that $P\{\xi_n = +1\} = p = 1 - P\{\xi_n = -1\}$. Define τ to be the first time that $S_n = -1$, or $+\infty$ if there is no such n. Write $q = 1-p$. Then the sequence $(q/p)^{S_n}$ is a martingale, as is easily checked, and so Doob's Optional Sampling Formula implies that for every $n \geq 1$,

$$E\left(\frac{q}{p}\right)^{S_{\tau\wedge n}} = 1.$$

Since $p > q$, the sequence $(q/p)^{S_{\tau\wedge n}}$ is bounded above by p/q. Moreover, by the strong law of large numbers, $S_n \to \infty$ as $n \to \infty$, so on the event that $\tau = \infty$, the martingale $(q/p)^{S_{\tau\wedge n}}$ will converge to 0, Therefore, by the dominated convergence theorem,

$$E\left(\frac{q}{p}\right)^{S_\tau} \mathbf{1}_{\{\tau<\infty\}} = 1.$$

But $S_\tau = -1$ on the event that $\tau < \infty$, so it follows that

$$\boxed{P\{\tau < \infty\} = \frac{q}{p}.}$$

Example 3. Once again let $\{S_n\}_{n\geq 0}$ be simple random walk on $\mathbb{Z}$ started at $S_0 = 0$, and define τ be the first time that $S_n = 1$, or ∞ if there is no such n. By Example 1, the probability that the random walk reaches $+1$ before $-n$ is $n/(n+1)$, for each integer $n \geq 1$, so $P\{\tau < \infty\} \geq n/(n+1)$. Since n is arbitrary, it follows that $P\{\tau < \infty\} = 1$.

Now let's derive the generating function of the random variable τ. For this we use the third Wald identity. For each θ, the moment generating function of the increments ξ_n is

$$Ee^{\theta\xi_1} = \frac{e^\theta + e^{-\theta}}{2} = \cosh\theta.$$

Therefore, by Wald,

$$E\frac{e^{\theta S_{\tau\wedge n}}}{\cosh\theta^{\tau\wedge n}} = 1.$$

As $n \to \infty$, the integrand converges to $e^{\theta S_\tau}/\cosh\theta^\tau$. To justify passing the limit under the expectation we shall use dominated convergence. Since $S_{\tau\wedge n} \le 1$, if $\theta > 0$ then $\exp\{\theta S_{\tau\wedge n}\} \le e^\theta$. Moreover, $\cosh\theta > 1$, so

$$1/\cosh\theta^{\tau\wedge n} \le 1.$$

Hence, the dominated convergence theorem implies that for $\theta > 0$,

$$E\frac{e^{\theta S_\tau}}{\cosh\theta^\tau} = 1 \quad \Longrightarrow$$
$$E(\cosh\theta)^{-\tau} = e^{-\theta}.$$

To obtain the generating function of τ, make the change of variable $z = 1/\cosh\theta$, which implies that $e^{-\theta} = x$ must satisfy the quadratic equation

$$x + x^{-1} = 2z^{-1} \quad \Rightarrow \quad e^{-\theta} = \frac{1-\sqrt{1-z^2}}{z}.$$

This shows that for $0 < z < 1$,

$$\boxed{Ez^\tau = \frac{1-\sqrt{1-z^2}}{z}.}$$

4. 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 $\{X_n\}_{n\ge0}$ be a sub- or super-martingale relative to $\{Y_n\}_{n\ge0}$, and for each $n \ge 0$ define*

$$M_n = \max_{0\le m\le n} X_m, \qquad \textit{and} \tag{24}$$

$$M_\infty = \sup_{0\le m<\infty} X_m = \lim_{n\to\infty} M_n \tag{25}$$

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

$$P\{M_n \ge \alpha\} \le E(X_n \vee 0)/\alpha \qquad \textit{if } \{X_n\}_{n\ge0} \textit{ is a submartingale, and} \tag{26}$$

$$P\{M_\infty \ge \alpha\} \le EX_0/\alpha \qquad \textit{if } \{X_n\}_{n\ge0} \textit{ is a nonnegative supermartingale.} \tag{27}$$

Proof. Assume first that $\{X_n\}_{n\ge0}$ is a submartingale. Without loss of generality, we may assume that each $X_n \ge 0$, 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 $n \ge 0$ such that $X_n \ge \alpha$, or $+\infty$ is there is no such n. Then for any nonrandom $n \ge 0$, the truncation $\tau \wedge n$ is a stopping time and so, by the Optional Sampling Theorem,

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

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 (26).

The proof of inequality (27) 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 (27) 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 (27) 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?

Proposition 8. *Let $\{X_n\}_{n\geq 0}$ be a nonnegative supermartingale relative to a filtration $\{\mathscr{F}_n\}_{n\geq 0}$. Then for any $\alpha > 0$,,*

$$P\{\sup_{n\geq 0} X_n \geq \alpha\} \leq \frac{EX_0}{\alpha}. \tag{28}$$

Proof. Exercise. Hint: Begin by showing that for any stopping time τ the sequence $\{X_{\tau\wedge n}\}_{n\geq 0}$ is a supermartingale. $\square$

5. Convergence of L^2-Bounded Martingales

To illustrate the usefulness of the Maximal Inequality, we shall prove a special case of the Martingale Convergence Theorem, which will be proved in full generality using different methods in section 7 below.

Theorem 5. *Let $\{X_n\}_{n\geq 0}$ be a martingale relative to a filtration $\{\mathscr{F}_n\}_{n\geq 0}$ such that $\sup_{n\geq 0} E|X_n|^2 < \infty$. Then*

$$X_\infty = \lim_{n\to\infty} X_n \tag{29}$$

exists almost surely. Furthermore, $X_n \to X_\infty$ in L^2, and the martingale is closed, *that is, for every $n \geq 0$*

$$X_n = E(X_\infty \mid \mathscr{F}_n). \tag{30}$$

Proof. Recall (cf. Corollary 1) that the differences $\xi_n := X_n - X_{n-1}$ are uncorrelated, and consequently orthogonal in L^2, and that for each n,

$$EX_n^2 = EX_0^2 + \sum_{k=1}^{n} E\xi_k^2.$$

By hypothesis, the sequence EX_n^2 is bounded, and so $\sum_{k\geq 1} E\xi_k^2 < \infty$. Moreover, the orthogonality of the increments ξ_j implies that for any $n, m \geq 1$,

$$E(X_{n+m} - X_n)^2 = \sum_{k=n+1}^{n+m} E\xi_k^2.$$

Thus, the sequence $\{X_n\}_{n\geq 0}$ is Cauchy in L^2, and so the completeness of L^2 implies that there exists $X_\infty \in L^2$ such that $X_n \to X_\infty$ in the L^2-norm.

The closure property (30) follows easily from L^2-convergence. The martingale property implies that for any two integers $n, m \geq 0$,

$$X_n = E(X_{n+m} \mid \mathscr{F}_n).$$

For random variables in L^2, conditional expectation given $\mathscr{F}_n$ coincides with orthogonal projection onto the subspace $L^2(\Omega, \mathscr{F}_n, P)$. Since this orthogonal projection is a continuous mapping on L^2, the convergence $X_{n+m} \to X_\infty$ (as $m \to \infty$) implies the closure property (30).

It remains to prove that $X_n \to X_\infty$ almost surely. For this it suffices to show that for almost every $\omega \in \Omega$ the sequence $X_n(\omega)$ is a Cauchy sequence of real numbers. To accomplish this, we will exploit the fact that for any m the sequence

$$X_{n+m} - X_m = \sum_{j=n+1}^{n+m} \xi_j \quad \text{for } n = 0, 1, 2, \ldots$$

is a martingale (relative to the filtration $\{\mathscr{F}_{n+m}\}_{n\geq 0}$). For each $k \geq 1$, let n_k be so large that $\sum_{n\geq n_k} E\xi_n^2 < 8^{-k}$. Then by the Maximal Inequality,

$$P\{\sup_{n\geq n_k} |X_n - X_{n_k}| \geq 2^{-k}\} \leq \frac{\sum_{n\geq n_k} E\xi_n^2}{4^{-k}} \leq 2^{-k}.$$

Consequently, by the Borel-Cantelli Lemma, with probability one only finitely many of these events will occur. Clearly, on the event that $\sup_{n\geq n_k} |X_n - X_{n_k}| < 2^{-k}$ for all sufficiently large k the sequence X_n is Cauchy. □

REMARK. This argument does not generalize to martingales that are bounded in L^1 (or L^p, for some $p \in (1,2)$), because for $p \neq 2$ there is no easy way to use L^p−boundedness to show that the sequence X_n has an L^p−limit. Later we will show by more indirect arguments that if $p > 1$ the sequence X_n converges both almost surely and in L^p, but that although every L^1−bounded martingale converges almost surely, not every L^1−bounded martingale converges in L^1.

6. 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

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

with the convention that the min is $+\infty$ if there is no such n. Then

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

Proposition 9. *Let X_n be a submartingale relative to Y_n. Then for any scalars $\alpha < \beta$ and all nonnegative integers m, n,*

$$(32)\qquad (\beta - \alpha) E N_n((\alpha,\beta]) \leq E(X_n \vee 0) + |\alpha|.$$

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:

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

(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,

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

□

For *nonnegative* martingales – or more generally, nonnegative *supermartingales* – there is an even better upcrossings inequality, due to Dubins. Whereas the upcrossings inequality (32) only bounds the expected number of upcrossings, Dubins' inequality shows that the number of upcrossings actually has a geometrically decreasing tail.

Proposition 10. *(Dubins) Let* $\{X_n\}_{n \geq 0}$ *be a nonnegative supermartingale relative to some filtration, and for any* $0 \leq \alpha < \beta < \infty$ *define* $N((\alpha, \beta])$ *to be the number of upcrossings of the interval* $(\alpha, \beta]$ *by the sequence* X_n*. Then*

$$P\{N((\alpha, \beta]) \geq k\} \leq \left(\frac{\alpha}{\beta}\right)^k. \tag{33}$$

Proof. This is by induction on k: we will show that for each $k \geq 0$,

$$P(N((\alpha, \beta]) \geq k + 1) \leq \left(\frac{\alpha}{\beta}\right) P(N((\alpha, \beta]) \geq k). \tag{34}$$

Fix $m \geq 0$ and let τ be the first time $n \geq m$ such that $X_n \geq \beta$, or $+\infty$ if there is no such n. Clearly, τ is a stopping time, so by Doob's Optional Sampling Theorem, the stopped sequence $X_{\tau \wedge n}$ is a nonnegative supermartingale. Consequently, for each $n \geq m$,

$$E(X_{\tau \wedge n} \,|\, \mathscr{F}_m) \leq X_m;$$

since $X_n \geq 0$ and $X_{\tau \wedge n} \geq \beta$ on the event $\tau \leq n$, it follows that $P(\tau \leq n \,|\, \mathscr{F}_m) \leq X_m/\beta$, and since n is arbitrary, the monotone convergence theorem implies that

$$P(\tau < \infty \,|\, \mathscr{F}_m) \leq \frac{X_m}{\beta}. \tag{35}$$

Let σ_j and τ_j be the stopping times defined inductively by (31). Then $N((\alpha,\beta]) \geq k$ if and only if $\tau_k < \infty$, so to establish (34) it suffices to prove that

$$P(\tau_{k+1} < \infty) \leq \left(\frac{\alpha}{\beta}\right) P(\tau_k < \infty).$$

Obviously, $\tau_{k+1} < \infty$ is possible only if $\sigma_k < \infty$, so it will be enough to show that

$$P(\tau_{k+1} < \infty) \leq \left(\frac{\alpha}{\beta}\right) P(\sigma_k < \infty).$$

But for every integer $m \geq 0$ the event $\sigma_k = m$ is in $\mathscr{F}_m$, and on this event it must be the case that $X_m \leq \alpha$, so (35) implies that

$$P(\tau_{k+1} < \infty \,|\, \mathscr{F}_m) \mathbf{1}_{\{\sigma_k = m\}} \leq \alpha/\beta \mathbf{1}_{\{\sigma_k = m\}}.$$

Taking expectations on each side and summing on m does the rest. □

7. Martingale Convergence Theorems

7.1. **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*

$$\lim_{n\to\infty} X_n := X_\infty \tag{36}$$

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

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

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 $X_n \geq 0$,

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

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

7.2. 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 4. 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 $X_\lambda \in A$,

$$E|X_\lambda|\mathbf{1}\{|X_\lambda| \geq C_\delta\} \leq \delta. \tag{37}$$

The next two propositions are standard results in measure theory. Their proofs are not difficult. Proposition 12 is quite useful, as it provides a simple test for uniformly integrability.

Proposition 11. *For any collection $\{X_\lambda\}_{\lambda\in\Lambda}$ of integrable random variables on the probability space $(\Omega, \mathscr{F}, P)$, the following are equivalent:*

(A) The collection $\{X_\lambda\}_{\lambda\in\Lambda}$ is uniformly integrable.
(B) For every $\varepsilon > 0$ there exists $\delta > 0$ such that for every event $B \in \mathscr{F}$ with probability $P(B) < \delta$,

$$\sup_{\lambda\in\Lambda} E|X_\lambda|\mathbf{1}_B < \varepsilon. \tag{38}$$

(C) For every $\varepsilon > 0$ there exists $\delta > 0$ such that for every random variable Y satisfying $0 \leq Y \leq 1$ and $EY < \delta$,

$$\sup_{\lambda\in\Lambda} E|X_\lambda|Y < \varepsilon. \tag{39}$$

Proposition 12. *Any bounded subset of L^p, where $p > 1$, is uniformly integrable.*

It is easily checked that if $|X_\lambda| \leq Y$ for every $\lambda \in \Lambda$, and if $EY < \infty$, then the collection $\{X_\lambda\}_{\lambda\in\Lambda}$ is uniformly integrable. Thus, the following result extends the dominated convergence theorem.

Proposition 13. *Let $\{X_n\}_{n\geq 1}$ 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 $\{X_n\}_{n\geq 1}$ is uniformly integrable.*

Proof. I'll prove the useful direction, that uniform integrability implies L^1-convergence. 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 (37) 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 11) 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 (37)). 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 $EX_n \to EX$. □

Exercise 2. Let Y_n be a uniformly integrable sequence of random variables defined on a probability space $(\Omega, \mathscr{F}, P)$ such that

$$Y(\omega) := \lim_{n\to\infty} Y_n(\omega)$$

exists for every $\omega \in \Omega$. Let $\mathscr{G}$ be a $\sigma-$algebra contained in $\mathscr{F}$.

(A) Prove that $E(Y_n|\mathscr{G})$ converges to $E(Y|\mathscr{G})$ in L^1-norm.

(B) Give an example to show that $E(Y_n|\mathscr{G})$ need not converge to $E(Y|\mathscr{G})$ almost surely.

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:*

$$X_n = E(X\,|\,\mathscr{F}_n). \tag{40}$$

Corollary 4. *Let $(\mathscr{F}_n)_{n\geq 0}$ be a filtration and set $\mathscr{F}_\infty = \sigma(\cup_{n\geq 0}\mathscr{F}_n)$ be the smallest $\sigma-$algebra containing all of the $\sigma-$algebras $\mathscr{F}_n$. If $X \in L^1(\Omega, \mathscr{F}_\infty, P)$ is an integrable random variable that is measurable with respect to $\mathscr{F}_\infty$ then*

$$\lim_{n\to\infty} E(X\,|\,\mathscr{F}_n) = X \text{ almost surely and in } L^1.$$

Proof. The martingale $\{E(X|\mathscr{F}_n)\}_{n\geq 0}$ is uniformly integrable, by Doob's theorem (see note on conditional expectation). By the martingale convergence theorem,

$$Y = \lim_{n\to\infty} E(X|\mathscr{F}_n)$$

exists almost surely, and since the martingale is uniformly integrable, the convergence holds also in L^1, and $E(Y|\mathscr{F}_n) = E(X|\mathscr{F}_n)$ a.s. for every n, by Corollary 3. It follows that for every event $F \in \mathscr{F}_n$,

$$EY\mathbf{1}_F = EX\mathbf{1}_F.$$

But if this identity holds for every $F \in \cup_{n\geq 0}\mathscr{F}_n$ then it must hold for every $F \in \mathscr{F}_\infty$ (why?). It then follows that $X = Y$ a.s. □

The following proposition, although not needed for the theory of discrete-time martingales, naturally belongs here. It is useful in the theory of continuous-time martingales, where conditional expectations against uncountable collections of $\sigma-$algebras arise naturally.

Proposition 14. *Let $\{X_\lambda\}_{\lambda\in\Lambda}$ be a uniformly integrable collection of random variables defined on a probability space $(\Omega,\mathscr{F},P)$, and let $\mathbb{F}$ be the set of all $\sigma-$algebras contained in $\mathscr{F}$. Then the collection*

$$\{E(X_\lambda\,|\,\mathscr{G})\}_{\mathscr{G}\in\mathbb{F}}$$

is uniformly integrable.

Proof. This will use the equivalent characterizations of uniform integrability provided by Proposition 11. Since the collection $\{X_\lambda\}_{\lambda\in\Lambda}$ is, by hypothesis, uniform integrable, for any $\varepsilon>0$ there exists $\delta>0$ such that for all random variables Y satisfying $0\le Y\le 1$ and $EY<\delta$, we have

$$E|X_\lambda|Y<\varepsilon.$$

Now let B be any event such that $P(B)<\delta$; then for any $\mathscr{G}\in\mathbb{F}$,

$$\begin{aligned}E|E(X_\lambda\,|\,\mathscr{G})|\mathbf{1}_B &= E|E(X_\lambda\,|\,\mathscr{G})|E(\mathbf{1}_B\,|\,\mathscr{G})\\ &\le EE(|X_\lambda|\,|\,\mathscr{G})E(\mathbf{1}_B\,|\,\mathscr{G})\\ &= E|X_\lambda|E(\mathbf{1}_B\,|\,\mathscr{G})\\ &<\varepsilon,\end{aligned}$$

the last inequality because the random variable $E(\mathbf{1}_B\,|\,\mathscr{G})$ takes values between 0 and 1 and has expectation $P(B)<\delta$. Therefore, the collection $\{E(X_\lambda\,|\,\mathscr{G})\}_{\mathscr{G}\in\mathbb{F}}$ is uniform integrable. □

7.3. **Reverse Martingales.** The notion of a martingale extends to sequences X_n of random variables indexed by the negative integers $n\le-1$. A *reverse* or *backward filtration* is a sequence $\{\mathscr{F}_n\}_{n\le-1}$ of $\sigma-$algebras indexed by non-positive integers such that $\mathscr{F}_n\subset\mathscr{F}_{n+1}$ for each $n\le-2$. A sequence of random variables or vectors $\{X_n\}_{n\le 0}$ is said to be *adapted* to the filtration $\{\mathscr{F}_n\}_{n\le-1}$ if X_n is measurable relative to $\mathscr{F}_n$ for every n, and an adapted sequence X_n is a *reverse martingale* if for every $n\le-1$,

$$E(X_{n+1}\,|\,\mathscr{F}_n)=X_n. \tag{41}$$

Reverse submartingales and *supermartingales* arise naturally in connection with continuous-time processes, as we will show later.

Reverse Martingale Convergence Theorem . *Let $\{X_n\}_{n\le-1}$ be a reverse martingale relative to a reverse filtration $(\mathscr{F}_n)_{n\le-1}$. Then*

$$\lim_{n\to-\infty}X_n=E(X_{-1}\,|\,\cap_{n\le-1}\mathscr{F}_n) \tag{42}$$

almost surely and in L^1.

Proof. Both the Maximal Inequality and the Upcrossings Inequality apply to any finite stretch $\{X_n\}_{-m\le n\le-1}$ of the martingale. The Upcrossings Inequality (32) implies that for any two rational numbers $\alpha<\beta$ the expected number of upcrossings of $(\alpha,\beta]$ by $\{X_n\}_{-m\le n\le-1}$ is bounded above by $E(X_{-1}\vee 0)+|\alpha|/(\beta-\alpha)$. Since this bound does not depend on $-m$, it follows that the expected number of upcrossings of $(\alpha, \beta]$ by $\{X_n\}_{-\infty<n\leq -1}$ is finite. Therefore, since the rationals are countable, the (reverse) sequence $\{X_n\}_{-\infty<n\leq -1}$ must converge with probability 1.

Denote by $X_{-\infty}$ the almost sure limit of the sequence. By the Maximal Inequality, $X_{-\infty}$ is almost surely finite, and it is clearly measurable with respect to $\mathscr{F}_{-\infty} := \cap_{n\leq -1}\mathscr{F}_n$. Since a reverse martingale is necessarily uniformly integrable, by Doob's theorem, the convergence $X_n \to X_{-\infty}$ holds in L^1 a well as almost surely. Thus,

$$\begin{aligned} X_{-\infty} &= E(X_{-\infty} \,|\, \mathscr{F}_{-\infty}) \\ &= \lim_{m\to-\infty} E(X_m \,|\, \mathscr{F}_{-\infty}) \\ &= \lim_{m\to-\infty} EE(X_{-1} \,|\, \mathscr{F}_m) \,|\, \mathscr{F}_{-\infty}) \\ &= E(X_{-1} \,|\, \mathscr{F}_{-\infty}). \end{aligned}$$

□

8. Exchangeability

8.1. **Strong law of large numbers for exchangeable random variables.**

Definition 5. A sequence $Y_1, Y_2, \ldots$ of (not necessarily real-valued) random variables is said to be *exchangeable* if its joint distribution is invariant under finite permutations of the indices, that is, if for every $N < \infty$ and every permutation σ of the integers $1, 2, \ldots, N$,

$$(Y_{\sigma(1)}, Y_{\sigma(2)}, \ldots, Y_{\sigma(N)}) \stackrel{\mathscr{D}}{=} (Y_1, Y_2, \ldots, Y_N).$$

Observe that if the sequence Y_n is exchangeable then the random variables Y_n are identically distributed. The converse is not true: if X is any non-constant random variable then the sequence $X, X, X, \ldots$ is trivially exchangeable, but its entries are not independent. Here is a more interesting example. Let Θ be a random variable taking values in the unit interval $[0,1]$, and let $\{U_n\}_{n\geq 1}$ be independent, identically distributed Uniform- (0,1) random variables, independent of Θ; set

$$X_n = \mathbf{1}_{\{U_n \leq \Theta\}}. \tag{43}$$

Then the sequence $\{X_n\}_{n\geq 1}$ is exchangeable, as is easily checked, but the entries X_n are not mutually independent unless Θ is constant. The *de Finetti Theorem*, whose proof will be given later in this section, asserts that *every* exchangeable sequence of Bernoulli random variables arises in this manner.

Definition 6. For any sequence $\{Y_n\}_{n\geq 1}$ define the *exchangeable filtration* as follows: for each $n \geq 1$, let $\mathscr{E}_{-n}$ consist of all events B whose indicators $\mathbf{1}_B$ are Borel functions χ_B of the random variables $Y_1, Y_2, \ldots$ such that for every permutation σ of $[n]$,

$$\chi_B(Y_1, Y_2, \ldots) = \chi_B(Y_{\sigma(1)}, Y_{\sigma(2)}, \ldots, Y_{\sigma(n)}, Y_{n+1}, \ldots). \tag{44}$$

Let $\mathscr{E} = \cap_{n\geq 1}\mathscr{E}_{-n}$; this is called the *exchangeable σ−algebra.*

Proposition 15. *Let $Y_1, Y_2, \ldots$ be an exchangeable sequence of integrable random variables, and for each $n \geq 1$ set*

$$\Theta_n = n^{-1} \sum_{j=1}^{n} Y_j.$$

Then the sequence $\{\Theta_{-n}\}_{n\leq -1}$ is a reverse martingale relative to the exchangeable filtration, and consequently

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}Y_j = E(Y_1\,|\,\mathscr{E}) \quad \textit{almost surely.} \tag{45}$$

Proof. It is clear that the sequence $\{\Theta_{-n}\}_{n\leq -1}$ is adapted, because for each n the random variable Θ_n is invariant under permutations of $Y_1, Y_2, \dots, Y_n$. Let $B \in \mathscr{E}_{-n}$; then the indicator function $\mathbf{1}_B = \chi_B(Y_1, Y_2, \dots)$ satisfies (44) for every permutation of $[n]$. Fix $j \in [n]$ and let $\sigma = \sigma_j$ be the permutation that swaps the letters 1 and j and leaves everything else fixed. Then equation (44) and the exchangeability of the sequence $Y_1, Y_2, \dots$ imply that

$$\begin{aligned}
EY_1\chi_B(Y_1, Y_2, \dots) &= EY_1\chi_B(Y_{\sigma(1)}, Y_{\sigma(2)}, \dots, Y_{\sigma(n)}, Y_{n+1}, \dots) \\
&= EY_{\sigma(1)}\chi_B(Y_{\sigma(1)}, Y_{\sigma(2)}, \dots, Y_{\sigma(n)}, Y_{n+1}, \dots) \\
&= EY_j\chi_B(Y_{\sigma(1)}, Y_{\sigma(2)}, \dots, Y_{\sigma(n)}, Y_{n+1}, \dots) \\
&= EY_j\chi_B(Y_1, Y_2, \dots).
\end{aligned}$$

(Note: Exchangeability is used in the second equality.) It follows that for every $j = 1, 2, \dots, n$,

$$E(Y_1\,|\,\mathscr{E}_{-n}) = E(Y_j\,|\,\mathscr{E}_{-n}).$$

Summing over $j \leq n$ and dividing by n yields

$$E(Y_1\,|\,\mathscr{E}_{-n}) = E(\Theta_n\,|\,\mathscr{E}_{-n}) = \Theta_n.$$

This prove that the sequence Θ_{-n} is a reverse martingale. The strong law (45) now follows by the Reverse Martingale Convergence Theorem. □

Essentially the same argument applies to functions of several arguments. This is a useful observation, because many important statistics (for instance, all *U−statistics*) have this form.

Proposition 16. *Let $\{Y_n\}_{n\geq 1}$ be an exchangeable sequence of real random variables, and let $h : \mathbb{R}^d \to \mathbb{R}$ be any measurable function such that $E|h(Y_1, Y_2, \dots, Y_d)| < \infty$. Then*

$$\lim_{n\to\infty}\frac{1}{n^d}\sum_{i_1=1}^{d}\sum_{i_2=1}^{n}\cdots\sum_{i_d=1}^{n}h(Y_{i_1}, Y_{i_2}, \dots, Y_{i_d}) = E(h(Y_1, Y_2, \dots, Y_d)\,|\,\mathscr{E}). \tag{46}$$

Proof. Exercise. □

Since an i.i.d. sequence is exchangeable, Propositions 15 and 16 are valid when the random variables Y_n are i.i.d. It is therefore of interest to understand the exchangeable σ−algebra in this special case.

Proposition 17. *(Hewitt-Savage 0-1 Law) Let $\{Y_n\}_{n\geq 1}$ be independent, identically distributed random variables, and let $\mathscr{E}$ be the exchangeable σ−algebra. Then every event $F \in \mathscr{E}$ has probability* 0 *or* 1*, and consequently, every $\mathscr{E}$−measurable random variable is constant a.s.*

Proof. Let $\{\mathscr{E}_{-n}\}_{n\geq 1}$ be the exchangeable filtration, let $\mathscr{F}_n = \sigma(Y_j)_{j\leq n}$ be the natural filtration for the sequence $\{Y_j\}_{j\geq 1}$, and let $\mathscr{F}_\infty = \sigma(Y_j)_{j\geq 1}$. Then

$$\mathscr{E} \subset \cdots \subset \mathscr{E}_{-n-1} \subset \mathscr{E}_{-n} \subset \cdots \mathscr{E}_{-1} = \mathscr{F}_\infty.$$

Since $\mathscr{F}_\infty$ is the smallest $\sigma-$algebra containing $\cup_{n\geq 1}\mathscr{F}_n$, every event in $\mathscr{E}$ can be arbitrarily well-approximated by events in $\cup_{n\geq 1}\mathscr{F}_n$. In particular, for any event $A \in \mathscr{E}$ there exist events $A_n \in \mathscr{F}_n$ such that

$$\lim_{n\to\infty} E|\mathbf{1}_A - \mathbf{1}_{A_n}| = 0. \tag{47}$$

This, of course, implies that $P(A) = \lim_{n\to\infty} P(A_n)$. Now

$$\mathbf{1}_{A_n} = \chi_{A_n}(Y_1, Y_2, \dots, Y_n) \quad \text{and}$$
$$\mathbf{1}_A = \chi_A(Y_1, Y_2, \dots)$$

where $\chi_{A_n} : \mathbb{R}^n \to \mathbb{R}$ and $\chi_A : \mathbb{R}^\infty \to \mathbb{R}$ are Borel measurable. Furthermore, since $A \in \mathscr{E}$, the function χ_A satisfies (44) for every permutation σ of finitely many coordinates. In particular, (44) holds for the permutation σ_n that swaps the first n integers with the second n integers:

$$\sigma_n(j) = j + n \quad \text{for } 1 \leq j \leq n;$$
$$\sigma_n(j+n) = j \quad \text{for } 1 + n \leq j \leq 2n.$$

Consequently, by (47),since the sequence $\{Y_j\}_{j\geq 1}$ is exchangeable,

$$\lim_{n\to\infty} E|\mathbf{1}_A - \chi_{A_n}(Y_{n+1}, Y_{n+2}, \dots, Y_{2n})| = 0. \tag{48}$$

Combining (47)–(48) and using the trivial observation $\mathbf{1}_A = \mathbf{1}_A \mathbf{1}_A$ we obtain

$$\lim_{n\to\infty} E|\mathbf{1}_A - \chi_{A_n}(Y_1, Y_2, \dots, Y_n)\chi_{A_n}(Y_{n+1}, Y_{n+2}, \dots, Y_{2n})| = 0.$$

Since the random variables Y_j are independent and identically distributed, it follows (by taking the absolute values outside of the expectation) that

$$P(A) = \lim_{n\to\infty} P(A_n)^2.$$

Thus, $P(A) = P(A)^2$, and so $P(A) = 0$ or 1. □

Corollary 5. *(Strong Law of Large Numbers) Let $\{Y_n\}_{n\geq 1}$ be independent, identically distributed random variables with finite first moment $E|Y_1| < \infty$, and let $S_n = \sum_{k=1}^n Y_k$. Then with probability one,*

$$\lim_{n\to\infty} \frac{S_n}{n} = EY_1$$

almost surely and in L^1.

Proof. By Proposition 15, the sequence S_n/n converges almost surely and in L^1 to $E(Y_1|\mathscr{E})$. The Hewitt-Savage 0–1 Law implies that$E(Y_1|\mathscr{E}) = EY_1$. □

8.2. **De Finetti's Theorem.**

Theorem 6. *(de Finetti) For any exchangeable sequence $\{X_n\}_{n\geq 1}$ of Bernoulli random variables there is a unique Borel probability measure μ on the unit interval $[0,1]$ such that for any finite sequence $\{e_j\}_{j\leq m}$ of 0s and 1s,*

$$P\{X_k = e_k \textit{ for every } k \leq m\} = \int_{[0,1]} P_p\{X_k = e_k \textit{ for every } k \leq m\}\mu(dp) \quad \textit{where} \tag{49}$$

$$P_p\{X_k = e_k \textit{ for every } k \leq m\} = p^{\sum_{i=1}^m e_i}(1-p)^{m - \sum_{i=1}^m e_i}. \tag{50}$$

This theorem can be re-formulated as a statement describing the (regular) conditional distribution of the sequence $\{X_n\}_{n\geq 1}$ given the exchangeable $\sigma-$algebra $\mathscr{E}$. By Proposition 15, if $\{X_n\}_{n\geq 1}$ is an exchangeable Bernoulli sequence then

$$\Theta := \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j \tag{51}$$

exists almost surely, and the limit Θ is measurable relative to $\mathscr{E}$.

Theorem 6'. *(de Finetti) The conditional distribution of an exchangeable sequence $\{X_n\}_{n\geq 1}$ of Bernoulli random variables given its exchangeable $\sigma-$algebra $\mathscr{E}$ is the product Bernoulli measure P_Θ, where Θ is given by* (51). *In other words, for any Borel set $B \subset ,1^\infty$,*

$$P((X_1, X_2, \dots) \in B \,|\, \mathscr{E}) = P_\Theta(X_1, X_2, \dots) \in B) \quad a.s. \tag{52}$$

Proof. First, observe that the second formulation (Theorem 6') implies the first, with μ taken to be the distribution of Θ. (Equation (49) follows from equation (52) by taking B to be the cylinder set determined by the finite sequence $\{e_j\}_{j\leq m}$.) Second, note that the uniqueness of μ is an easy consequence of the fact that a probability distribution on $[0,1]$ is uniquely determined by its moments. The identity (49) implies that for any $m \in \mathbb{Z}_+$ the mth moment of μ is $P\{X_k = 1 \;\forall\; k \leq m\}$, so any two probability measures μ such that (49) must have the same moments, and therefore must be equal.

To prove (52), it suffices to consider the case where B is a cylinder event, since these generate the Borel field on $\{0,1\}^\infty$. Thus, fix a finite sequence $\{e_i\}_{i\leq m}$ of 0s and 1s; we must prove that

$$P((X_i = e_i \;\forall\; i \leq m \,|\, \mathscr{E}) = \Theta^{\sum e_i}(1-\Theta)^{m-\sum e_i}. \tag{53}$$

To this end, choose $n > m$ and consider the conditional distribution of the random vector $(X_1, X_2, \dots, X_m)$ given $\mathscr{E}_{-n}$. We will argue that this conditional distribution is the same as the result of sampling *without replacement* from a bin containing S_n balls marked 1 and $n - S_n$ balls marked 0, where $S_n = \sum_{i=1}^n X_i$. The random variable S_n is measurable relative to $\mathscr{E}_{-n}$; in fact,

$$\mathscr{E}_{-n} = \sigma(S_n, X_{n+1}, X_{n+2}, \dots).$$

Since the random variables X_i are exchangeable, for any event $A \in \mathscr{E}_{-n}$, any integer $0 \leq k \leq n$, and any permutation σ of the indices in $[n]$,

$$P((X_i = e_i \;\forall\; i \leq m) \cap A \cap (S_n = k)) = P((X_{\sigma(i)} = e_i \;\forall\; i \leq m) \cap A \cap (S_n = k)).$$

But on the event $S_n = k$ the fraction of permutations $\sigma \in \mathscr{S}_n$ (here $\mathscr{S}_n$ denotes the set of all $n!$ permutations of the set $[n]$]) such that $X_{\sigma(i)} = e_i$ for all $i \leq m$ is

$$\binom{n-k}{m-r}\Big/\binom{n}{k} \quad \text{where} \quad r = \sum_{i=1}^{m} e_i.$$

Hence, averaging over all permutations of $[n]$ gives

$$P((X_i = e_i \;\forall\; i \leq m) \cap A \cap (S_n = k)) = P(A \cap (S_n = k)\binom{n-k}{m-r}\Big/\binom{n}{k}$$

It follows that

$$P((X_i = e_i \ \forall\ i \le m \,|\, \mathscr{E}_{-n}) = \binom{n-S_n}{m-r} \Big/ \binom{n}{S_n}.$$

This is the the distribution of a random sample of size m, drawn without replacement, from a bin with S_n balls marked 1 and $n - S_n$ balls marked 0. As every statistician knows (or ought to know), sampling without replacement is nearly indistinguishable from sampling with replacement when the sample size m is small compared to the number of balls n in the bin. In particular, since $S_n/n \to \Theta$,

$$\lim_{n\to\infty} \binom{n-S_n}{m-\sum e_i} \Big/ \binom{n}{S_n} = \Theta^{\sum e_i}(1-\Theta)^{m-\sum e_i}.$$

Therefore, by the reverse martingale theorem,

$$\begin{aligned} P((X_i = e_i \ \forall\ i \le m \,|\, \mathscr{E}) &= \lim_{n\to\infty} P((X_i = e_i \ \forall\ i \le m \,|\, \mathscr{E}_{-n}) \\ &= \Theta^{\sum e_i}(1-\Theta)^{m-\sum e_i}. \end{aligned}$$

This proves (53). □

The proof of de Finetti's theorem is specific to the case of Bernoulli random variables, as it relies on the relation between sampling without replacement and sampling with replacement. Nevertheless, the theorem itself holds more generally.

Theorem 7. *(Hewitt-Savage-de Finetti) Let $\{X_n\}_{n\ge1}$ be an exchangeable sequence of random variables taking values in a* Borel space*, and let $\mathscr{E}$ be the exchangeable $\sigma-$algebra. Then the (regular) conditional distribution of the sequence $\{X_n\}_{n\ge1}$ given $\mathscr{E}$ is almost surely a* product measure*, that is, conditional on $\mathscr{E}$ the random variables X_n are i.i.d.*

Proof. Recall that a Borel space is a measurable space $(\mathscr{X}, \mathscr{G})$ such that there is a bijective, bimeasurable mapping $T : \mathscr{X} \to [0,1]$. Hence, without loss of generality, we may assume that the random variables X_n are real-valued. Recall further that (i) the sequence space $\mathbb{R}^{\mathbb{N}}$, with the usual Borel sets, is a Borel space, and (ii) if a random variable Z takes values in a Borel space then it has a *regular conditional distribution* given any $\sigma-$algebra. Thus, the sequence $\{X_n\}_{n\ge1}$ has a regular conditional distribution given $\mathscr{E}$.

To show that the conditional distribution is, with probability one, a product measure it suffices to show that for any bounded, Borel measurable functions $\varphi_j : \mathbb{R} \to \mathbb{R}$,

$$E\left(\prod_{j=1}^{m} \varphi_j(X_j) \,\middle|\, \mathscr{E} \right) = \prod_{j=1}^{m} E(\varphi_j(X_1) \,|\, \mathscr{E}). \tag{54}$$

This will follow from Proposition 15 and Proposition 16. First, for each $i \le m$, Proposition 15 implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \varphi_i(X_j) = E(\varphi_i(X_1) \,|\, \mathscr{E}).$$

Next, Proposition 16 implies

$$\lim_{n\to\infty} \frac{1}{n^m} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_m=1}^{n} \prod_{j=1}^{m} \varphi_j(X_{i_j}) = E\left(\prod_{j=1}^{m} \varphi_j(X_j) \,\middle|\, \mathscr{E} \right).$$

Finally, since

$$\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_m=1}^{n} \prod_{j=1}^{m} \varphi_j(X_{i_j}) = \prod_{j=1}^{m} \left(\sum_{i=1}^{m} \varphi_j(X_i) \right),$$

it follows that

$$E\left(\prod_{j=1}^{m} \varphi_j(X_j) \,\middle|\, \mathscr{E} \right) = \prod_{j=1}^{m} E(\varphi_i(X_1) \,|\, \mathscr{E}).$$

□