Math 6810 (Probability)

Fall 2012

Lecture notes

Pieter Allaart
University of North Texas

September 23, 2012

Text: *Introduction to Stochastic Calculus with Applications*, by Fima C. Klebaner (3rd edition), Imperial College Press.

Other recommended reading: (Do not purchase these books before consulting with your instructor!)

1. *Real Analysis* by H. L. Royden (3rd edition), Prentice Hall.

2. *Probability and Measure* by Patrick Billingsley (3rd edition), Wiley.

3. *Probability with Martingales* by David Williams, Cambridge University Press.

4. *Stochastic Calculus for Finance I and II* by Steven E. Shreve, Springer.

5. *Brownian Motion and Stochastic Calculus* by Ioannis Karatzas and Steven E. Shreve, Springer. (Warning: this requires stamina, but is one of the few texts that is complete and mathematically rigorous)

Chapter 2

Stochastic Processes in Discrete Time

By a *stochastic process in discrete time* we mean simply a sequence $\{X_n\}_{n=0}^{\infty} = (X_0, X_1, \dots)$ of random variables defined on a common probability space, with some joint distribution. We think of the index n in X_n as a "time parameter", of time $n = 0$ as the "start" of the process, and of X_{n+1} as being observed "after" X_n. These intuitive concepts are made precise using the mathematically rigorous notions of filtration and stopping time, which will be introduced in the present chapter. We begin with the most important example of a discrete time stochastic process.

2.1 Random walk

Setup: Let $X_1, X_2, \dots$ be independent, identically distributed (i.i.d.) $\{-1, 1\}$-valued r.v.'s with $\mathrm{P}(X_i = 1) = p$ and $\mathrm{P}(X_i = -1) = q := 1 - p$ for all i, where $p \in (0, 1)$ is a constant parameter. Define

$$S_0 \equiv 0, \qquad \text{and} \quad S_n = X_1 + \cdots + X_n, \quad n \geq 1. \tag{2.1}$$

The process $\{S_n\}$ is called a *simple random walk* or *Bernoulli random walk* or a *nearest-neighbor random walk* on $\mathbb{Z}$. When $p = 1/2$, we speak of a *symmetric* simple random walk. This random walk can be thought of as the evolving fortune of a gambler who repeatedly bets \$1 on red at roulette, or (appropriately scaled) as the approximate path of a particle suspended in liquid.

The first basic question to ask is: what is the distribution of S_n for any fixed n? Note that when n is even, S_n must be even, and when n is odd, S_n must be odd. Thus, $n + S_n$ is always even. Furthermore, $|S_n| \leq n$. Fix n, and let k be of the same parity as n (i.e. $n + k$ is even), with $|k| \leq n$. Consider ordered pairs (i, S_i). Any path from $(0, 0)$ to (n, k) must include exactly $(n + k)/2$ up-steps and exactly $(n - k)/2$ down-steps. Thus (compare the binomial distribution!),

$$\mathrm{P}(S_n = k) = \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2}, \qquad k \in \mathbb{Z}, \tag{2.2}$$

where we interpret $\binom{n}{x}$ as 0 when x is not integer or when $x < 0$ or $x > n$. The mean and variance of S_n follow directly from the definition (2.1): since $\mathrm{E}(X_1) = p - q$ and $\mathrm{Var}(X_1) = 4pq$ (check!), we have

$$\mathrm{E}(S_n) = n\,\mathrm{E}(X_1) = (p-q)n, \qquad \mathrm{Var}(S_n) = n\,\mathrm{Var}(X_1) = 4pqn.$$

Whereas (2.2) gives the *marginal distribution* of S_n, the joint distribution of the process $\{S_n\}$ follows from the following property.

Theorem 2.1. *Simple random walk has the following properties:*

1. (Independent increments) *For all $n_1 < n_2 < \cdots < n_k$, the random variables $S_{n_1}, S_{n_2} - S_{n_1}, \ldots, S_{n_k} - S_{n_{k-1}}$ are independent.*
2. (Stationary increments) *For all n and m with $m < n$, $S_n - S_m \stackrel{d}{=} S_{n-m}$.*

Proof. (i) Since $S_{n_1} = \sum_{i=1}^{n_1} X_i$, $S_{n_2} - S_{n_1} = \sum_{i=n_1+1}^{n_2} X_i$, etc., the random variables (increments) $S_{n_1}, S_{n_2} - S_{n_1}, \ldots, S_{n_k} - S_{n_{k-1}}$ are functions of disjoint subcollections of the X_i. Since the X_i are independent, that makes the increments independent.

(ii) We have $S_{n-m} = \sum_{i=1}^{n-m} X_i$, and $S_n - S_m = \sum_{i=m+1}^{n} X_i$. So each of S_{n-m} and $S_n - S_m$ is a sum of the same number $(n-m)$ of the X_i, which are independent and have the same distribution. Hence, $S_{n-m} \stackrel{d}{=} S_n - S_m$. □

Exercise 2.2. How would you use the independent and stationary increments of the random walk to compute, say, $\mathrm{P}(S_1 = -1, S_4 = 0, S_{10} = 4)$?

2.1.1 Hitting times and recurrence

Definition 2.3. The *hitting time* or *first-passage time* of a point $r \in \mathbb{Z}\backslash\{0\}$ is the r.v.

$$T_r := \inf\{n \geq 1 : S_n = r\},$$

with the convention that $\inf \emptyset = \infty$. We can define T_r by this formula also for $r = 0$. The r.v. T_0 is the *first return time* to the origin.

In this subsection we focus on the event $\{T_r < \infty\}$. In the next subsection we specialize to the case $p = 1/2$ and calculate the full distribution of T_r.

We begin with T_1. To start, note that T_1 must be odd. We compute $\mathrm{P}(T_1 = 2n+1)$ for $n = 0, 1, 2, \ldots$. Note that

$$\begin{aligned}
\mathrm{P}(T_1 = 2n+1) &= \mathrm{P}(S_1 \leq 0, \ldots, S_{2n} \leq 0, S_{2n+1} = 1) \\
&= \mathrm{P}(S_1 \leq 0, \ldots, S_{2n-1} \leq 0, S_{2n} = 0)\,\mathrm{P}(X_{2n+1} = 1) \\
&= p\,\mathrm{P}(S_1 \leq 0, \ldots, S_{2n-1} \leq 0, S_{2n} = 0).
\end{aligned}$$

Now each path of $2n$ steps that begins and ends at 0 has probability $p^n q^n$, and we must count the number of such paths that do not go above 0. This number is

$$C_n := \frac{1}{n+1}\binom{2n}{n}.$$

The number C_n is called the nth *Catalan number*. For a proof that C_n is the answer to our counting problem, see the Wikipedia entry on Catalan numbers! (There are in fact four different proofs given.) We thus have:

$$\mathrm{P}(T_1 = 2n+1) = \frac{1}{n+1}\binom{2n}{n} p^{n+1} q^n, \qquad n = 0, 1, 2, \dots. \tag{2.3}$$

Calculus exercise: Show (by summing the above over n) that

$$\mathrm{P}(T_1 < \infty) = \frac{1 - \sqrt{1 - 4pq}}{2q} = \min\{1, p/q\}.$$

Hint: compare with the series $\sum_{n=0}^{\infty} \frac{1}{n+1}\binom{2n}{n} x^{n+1}$. The derivative of this is $\sum_{n=0}^{\infty} \binom{2n}{n} x^n$. Verify the identities

$$\binom{2n}{n} = (-4)^n \binom{-1/2}{n}$$

and the generalization of the binomial theorem:

$$\sum_{n=0}^{\infty} \binom{\alpha}{n} x^n = (1+x)^{\alpha}.$$

We see that $\mathrm{P}(T_1 < \infty) = 1$ when $p \geq 1/2$, but $\mathrm{P}(T_1 < \infty) < 1$ when $p < 1/2$. For $r \in \mathbb{N}$, we have

$$\mathrm{P}(T_r < \infty) = (\min\{1, p/q\})^r. \tag{2.4}$$

This can be seen as follows: write

$$T_r = T_1 + (T_2 - T_1) + \cdots + (T_r - T_{r-1}).$$

Thus T_r is expressed as the sum of r r.v.'s which are independent and all have the same distribution as T_1. (This is intuitively obvious, though making it precise is a somewhat tedious exercise using the independent and stationary increments of the walk.) Of course, T_r is finite if and only if all the summands in the above equation are finite. Thus, we have (2.4). It follows that when $p \geq 1/2$, the walk will eventually reach any level $r \geq 1$ with certainty, but when $p < 1/2$, there is a positive probability that it will not.

Analogous results hold for finiteness of T_r with $r < 0$; just switch p and q. But the case $r = 0$, when T_r is the first return time to the origin, is special. Here we calculate

$$\begin{aligned}
\mathrm{P}(T_0 < \infty) &= \mathrm{P}(T_0 < \infty | X_1 = 1)\,\mathrm{P}(X_1 = 1) + \mathrm{P}(T_0 < \infty | X_1 = -1)\,\mathrm{P}(X_1 = -1) \\
&= p\,\mathrm{P}(T_{-1} < \infty) + q\,\mathrm{P}(T_1 < \infty) \\
&= p \min\{1, q/p\} + q \min\{1, p/q\} \\
&= \min\{2p, 2q\}.
\end{aligned}$$

The second equality follows since, once the walk moves up to level 1, the probability of it ever returning to 0 is the same as the probability that a walk starting at 0 would ever reach -1 (and similarly for the second term).

Thus, $\mathrm{P}(T_0 < \infty) = 1$ if and only if $p = 1/2$. When the walk $\{S_n\}$ returns to its initial position eventually with probability one we say the walk is *recurrent* (otherwise: *transient*). Thus, simple random walk is recurrent if $p = 1/2$, and transient otherwise.

2.1.2 The reflection principle and the record process

Assume in this subsection that $p = 1/2$, i.e. the walk is symmetric. Define the *record* at time n by

$$M_n := \max\{S_0, S_1, \dots, S_n\}, \qquad n = 0, 1, \dots.$$

The *reflection principle* says this: for $r \geq 1$ and $k \leq r$, the number of paths (starting at 0) with $M_n \geq r$ and $S_n = k$ is equal to the number of paths with $S_n = 2r - k$. To see this, note that $\{M_n \geq r\} = \{T_r \leq n\}$. Consider a path with $M_n \geq r$ and $S_n = k$, and reflect the portion of the path from time T_r until time n in the horizontal line $y = r$. This gives a new path for which $S_n = 2r - k \geq r$. It is easy to see that this operation is bijective, and hence the reflection principle follows. Since all paths of length n have the same probability, we conclude that

$$\mathrm{P}(M_n \geq r, S_n = k) = \mathrm{P}(S_n = 2r - k), \qquad r \geq 1, k < r. \tag{2.5}$$

From this, it is easy to derive the joint distribution of S_n and M_n, and with a little more effort, the marginal distributions of M_n and T_r. First, we have simply

$$\begin{aligned} \mathrm{P}(M_n = r, S_n = k) &= \mathrm{P}(M_n \geq r, S_n = k) - \mathrm{P}(M_n \geq r+1, S_n = k) \\ &= \mathrm{P}(S_n = 2r - k) - \mathrm{P}(S_n = 2r + 2 - k). \end{aligned} \tag{2.6}$$

Next, observe that for $k > r$, $S_n = k$ implies $M_n \geq r$ and so

$$\begin{aligned} \mathrm{P}(M_n \geq r) &= \sum_{k \leq r} \mathrm{P}(M_n \geq r, S_n = k) + \sum_{k > r} \mathrm{P}(M_n \geq r, S_n = k) \\ &= \sum_{k \leq r} \mathrm{P}(S_n = 2r - k) + \sum_{k > r} \mathrm{P}(S_n = k) \\ &= \mathrm{P}(S_n \geq r) + \mathrm{P}(S_n > r) \\ &= 2\,\mathrm{P}(S_n > r) + \mathrm{P}(S_n = r). \end{aligned}$$

From this we obtain, after some algebra,

$$\mathrm{P}(M_n = r) = \mathrm{P}(M_n \geq r) - \mathrm{P}(M_n \geq r+1) = \mathrm{P}(S_n = r) + \mathrm{P}(S_n = r+1), \qquad r \geq 0.$$

And for $r \geq 1$ we have, by (2.6),

$$\begin{aligned}
\mathrm{P}(T_r = n) &= \mathrm{P}(S_n = r, M_{n-1} = r-1) \\
&= \mathrm{P}(M_{n-1} = r-1, S_{n-1} = r-1, X_n = 1) \\
&= \frac{1}{2}\,\mathrm{P}(M_{n-1} = r-1, S_{n-1} = r-1) \\
&= \frac{1}{2}\left[\mathrm{P}(S_{n-1} = r-1) - \mathrm{P}(S_{n-1} = r+1)\right].
\end{aligned}$$

The calculations can then be completed using (2.2), which simplifies substantially in the case $p = 1/2$.

(In fact, the distributions of M_n and T_r can be calculated fully for arbitrary p using the method of generating functions; we will not go into that here.)

2.2 Martingales

Definition 2.4. Let $(\Omega, \mathcal{F})$ be a measurable space. A *filtration* is an increasing sequence $\{\mathcal{F}_n\}_{n=0}^{\infty}$ of sub-σ-algebras of $\mathcal{F}$; that is,

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_n \subset \cdots \subset \mathcal{F}.$$

A probability space $(\Omega, \mathcal{F}, \mathrm{P})$ together with a filtration $\{\mathcal{F}_n\}$ on it is called a *filtered probability space*, denoted $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathrm{P})$.

The most common example of a filtration is that generated by a stochastic process:

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n).$$

We call $\{\mathcal{F}_n\}$ the *natural filtration* of the process $\{X_n\}$. We think of $\mathcal{F}_n$ as containing all information (in this case about the process $\{X_n\}$) "up to time n".

Definition 2.5. A stochastic process $\{X_n\}$ is *adapted* to a filtration $\{\mathcal{F}_n\}$ if X_n is $\mathcal{F}_n$-measurable for each n.

Definition 2.6. A process $\{X_n\}$ is called a *submartingale* relative to the filtration $\{\mathcal{F}_n\}$ if:

(i) $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$;

(ii) $\mathrm{E}|X_n| < \infty$ for all n; and

(iii) $\mathrm{E}[X_n|\mathcal{F}_{n-1}] \geq X_{n-1}$ a.s. for all n.

A process $\{X_n\}$ is a *supermartingale* if $\{-X_n\}$ is a submartingale. A process that is both a submartingale and a supermartingale is called a *martingale*.

When the filtration $\{\mathcal{F}_n\}$ is not mentioned explicitly, $\{\mathcal{F}_n\}$ is normally clear from the context, or else is understood to be the natural filtration of the process $\{X_n\}$.

Example 2.7. Let X be an integrable random variable and $\{\mathcal{F}_n\}$ a filtration. Then the process

$$X_n := \mathrm{E}[X|\mathcal{F}_n], \qquad n = 0, 1, 2, \dots$$

is a martingale relative to $\{\mathcal{F}_n\}$. To see this, note that X_n is clearly $\mathcal{F}_n$-measurable and use the tower property:

$$\mathrm{E}[X_n|\mathcal{F}_{n-1}] = \mathrm{E}[\mathrm{E}[X|\mathcal{F}_n]|\mathcal{F}_{n-1}] = \mathrm{E}[X|\mathcal{F}_{n-1}] = X_{n-1} \quad \text{a.s.}$$

Exercise 2.8. Let $\{X_n\}$ be a supermartingale. Show that $\mathrm{E}(X_n) \le \mathrm{E}(X_0)$ for all n. If $\{X_n\}$ is a martingale, then $\mathrm{E}(X_n) = \mathrm{E}(X_0)$ for all n. (Use induction and the law of double expectation.)

Example 2.9. Let $\{S_n\}$ be symmetric simple random walk ($p = 1/2$). Then $\{S_n\}$ and $\{S_n^2 - n\}$ are martingales. For simple random walk with arbitrary p, the following are martingales:

(i) $S_n - \mu n$, where $\mu = \mathrm{E}(X_1) = p - q$;

(ii) $(S_n - \mu n)^2 - \sigma^2 n$, where $\sigma^2 = \mathrm{Var}(X_1) = 4pq$;

(iii) $(p/q)^{S_n}$.

Example 2.10 (Sums of independent, zero-mean r.v.'s)**.** More generally, let $X_1, X_2, \dots$ be independent r.v.'s with mean 0, and put $S_n = X_1 + \dots + X_n$. Then $\{S_n\}$ is a martingale.

Example 2.11 (Products of independent, mean 1 r.v.'s)**.** Let $Z_1, Z_2, \dots$ be independent r.v.'s with $\mathrm{E}(Z_n) = 1$ for each n, and put $M_n = Z_1 \cdots Z_n$. Then $\{M_n\}$ is a martingale.

Proposition 2.12. *(i) Let $\{X_n\}$ be a martingale and φ a convex real function. Put $Y_n = \varphi(X_n)$. If $\mathrm{E}|Y_n| < \infty$ for all n, then $\{Y_n\}$ is a submartingale.*

(ii) Let $\{X_n\}$ be a submartingale and φ a nondecreasing, convex real function. Put $Y_n = \varphi(X_n)$. If $\mathrm{E}|Y_n| < \infty$ for all n, then $\{Y_n\}$ is a submartingale.

Proof. (i) This follows from the conditional version of Jensen's inequality:

$$\mathrm{E}[Y_n|\mathcal{F}_{n-1}] = \mathrm{E}[\varphi(X_n)|\mathcal{F}_{n-1}] \ge \varphi(\mathrm{E}[X_n|\mathcal{F}_{n-1}]) = \varphi(X_{n-1}) = Y_{n-1} \qquad \text{a.s.}$$

(ii) In this case we can replace the second equality above by "$\ge$" and obtain the desired result. □

2.2.1 Stopping times and the optional stopping theorem

Definition 2.13. A *stopping time* with respect to a filtration $\{\mathcal{F}_n\}$ is a $\mathbb{Z}_+ \cup \{\infty\}$-valued r.v. τ such that

$$\{\tau \le n\} \in \mathcal{F}_n, \qquad n \ge 0. \tag{2.7}$$

Equivalently (check!), τ is a stopping time if and only if

$$\{\tau = n\} \in \mathcal{F}_n \qquad \text{for all } n \geq 0.$$

Example 2.14. The *hitting time* of a Borel set B by an adapted stochastic process $\{X_n\}$, that is,

$$\tau = \inf\{n : X_n \in B\},$$

is a stopping time:

$$\{\tau = n\} = \{X_0 \notin B, \dots, X_{n-1} \notin B, X_n \in B\},$$

an intersection of events that all lie in $\mathcal{F}_n$ since $\mathcal{F}_n$ contains $\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_{n-1}$. Intuitively, by time n it is "known" whether the process has entered B on or before that time. In particular, the hitting times T_r (including the first return time T_0) of Section 2.1.1 are stopping times.

On the other hand, a r.v. like $\tau = \sup\{n \leq 20 : X_n \in B\}$ is, in general, *not* a stopping time.

Proposition 2.15. *Let τ_1 and τ_2 be stopping times. Then* $\max\{\tau_1, \tau_2\}$ *and* $\min\{\tau_1, \tau_2\}$ *are also stopping times. This extends to maxima and minima of any finite number of stopping times.*

Proof. Exercise. $\square$

Recall that if $\{X_n\}$ is a martingale, then $\mathrm{E}(X_n) = \mathrm{E}(X_0)$ for every (non-random!) n. This does *not* imply, however, that $\mathrm{E}(X_\tau) = \mathrm{E}(X_0)$ for random τ, even if τ is a stopping time. For one thing, X_τ is not defined on the set $\{\tau = \infty\}$. But even if $\tau < \infty$ almost surely, things can go wrong.

Example 2.16. As a case in point, consider again the symmetric simple random walk $\{S_n\}$ (starting at 0), and let

$$\tau = \inf\{n : S_n = 1\} = T_1.$$

We have seen in Subsection 2.1.1 that $\tau < \infty$ with probability 1, but clearly, $S_\tau = 1$ almost surely and hence $\mathrm{E}(S_\tau) = 1 \neq 0 = \mathrm{E}(S_0)$.

An important question is then, under what conditions on a martingale $\{X_n\}$ and a stopping time τ can we expect that $\mathrm{E}(X_\tau) = \mathrm{E}(X_0)$?

Definition 2.17. For a stochastic process $\{X_n\}$ and a stopping time τ, the *stopped process* determined by τ is the process $\{X_n^\tau\}$ defined by

$$X_n^\tau := X_{\tau \wedge n}, \qquad n = 0, 1, 2, \dots,$$

where $x \wedge y = \min\{x, y\}$.

Proposition 2.18. *A stopped (sub-, super-) martingale is a (sub-, super-) martingale: Let $\{X_n\}$ be a (sub)martingale and T a stopping time. Then $\{X_n^\tau\}$ is a (sub)martingale.*

Proof. Let $\{X_n\}$ be a submartingale. Since $\{\tau > n\} \in \mathcal{F}_n$, we have

$$\begin{aligned}
\mathrm{E}[X_{n+1}^\tau - X_n^\tau | \mathcal{F}_n] &= \mathrm{E}[X_{\tau\wedge(n+1)} - X_{\tau\wedge n} | \mathcal{F}_n] \\
&= \mathrm{E}\left[\mathrm{I}_{\{\tau\le n\}}\left(X_{\tau\wedge(n+1)} - X_{\tau\wedge n}\right)|\mathcal{F}_n\right] + \mathrm{E}\left[\mathrm{I}_{\{\tau> n\}}\left(X_{\tau\wedge(n+1)} - X_{\tau\wedge n}\right)|\mathcal{F}_n\right] \\
&= \mathrm{E}\left[\mathrm{I}_{\{\tau\le n\}}(X_\tau - X_\tau)|\mathcal{F}_n\right] + \mathrm{E}\left[\mathrm{I}_{\{\tau> n\}}(X_{n+1} - X_n)|\mathcal{F}_n\right] \\
&= \mathrm{I}_{\{\tau> n\}}\,\mathrm{E}[X_{n+1} - X_n|\mathcal{F}_n] \ge 0.
\end{aligned}$$

Hence, $\{X_n^\tau\}$ is a submartingale. □

It follows immediately that if $\{X_n\}$ is a submartingale and τ is bounded (i.e. $\mathrm{P}(\tau \le n) = 1$ for some n), then $\mathrm{E}(X_\tau) \ge \mathrm{E}(X_0)$. The following theorem, due to J. L. Doob, gives two other sufficient conditions.

Theorem 2.19 (Optional stopping theorem)**.** *Let $\{X_n\}$ be a submartingale and τ an a.s. finite stopping time. If either*

(i) $\{X_n\}$ is uniformly bounded (i.e. there is $K > 0$ such that $|X_n(\omega)| \le K$ for all n and all ω); or

(ii) the increments $X_{n+1} - X_n$ are uniformly bounded and $\mathrm{E}(\tau) < \infty$,

then X_τ is integrable and $\mathrm{E}(X_\tau) \ge \mathrm{E}(X_0)$. If $\{X_n\}$ is in fact a martingale and (i) or (ii) holds, then $\mathrm{E}(X_\tau) = \mathrm{E}(X_0)$.

Proof. Assume (i). By Proposition 2.18,

$$\mathrm{E}(X_{\tau\wedge n} - X_0) \ge 0. \tag{2.8}$$

Since $\tau < \infty$ a.s., $\tau \wedge n \to \tau$ a.s. Let $n \to \infty$ and use the BCT to conclude

$$\mathrm{E}(X_\tau - X_0) \ge 0,$$

i.e. $\mathrm{E}(X_\tau) \ge \mathrm{E}(X_0)$.

Next, assume (ii). Let K be such that $|X_n(\omega) - X_{n-1}(\omega)| \le K$ for all n and all ω. Then

$$|X_{\tau\wedge n} - X_0| = \left|\sum_{i=1}^{\tau\wedge n}(X_i - X_{i-1})\right| \le K(\tau\wedge n) \le K\tau.$$

Since $\mathrm{E}(\tau) < \infty$, the DCT gives the desired result. □

Let us return now to Example 2.16. Since the increments of $\{S_n\}$ are clearly uniformly bounded, it must be the case that the stopping time T_1 considered there has infinite expectation. This can indeed be verified using the distribution of T_1 given in (2.3) and Stirling's formula.

Exercise 2.20. Let $\{X_n\}$ be a nonnegative supermartingale and τ a finite stopping time. Show that $\mathrm{E}(X_\tau) \le \mathrm{E}(X_0)$.

A related result is the *optional sampling theorem*. For a stopping time τ, define

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \le n\} \in \mathcal{F}_n \ \forall\, n \ge 0\}.$$

Then $\mathcal{F}_\tau$ is a σ-algebra and τ is $\mathcal{F}_\tau$-measurable (check!). Furthermore, if τ_1 and τ_2 are stopping times and $\tau_1 \le \tau_2$ (pointwise on Ω), then $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$. (Check!)

Theorem 2.21 (Optional sampling theorem). *Let $\{X_0, X_1, \dots, X_n\}$ be a submartingale and let τ_1, τ_2 be stopping times with $\tau_1 \le \tau_2 \le n$. Then*

$$\mathrm{E}\left[X_{\tau_2} | \mathcal{F}_{\tau_1}\right] \ge X_{\tau_1} \quad a.s.$$

Proof. Let $A \in \mathcal{F}_{\tau_1}$. We must show that

$$\int_A (X_{\tau_2} - X_{\tau_1})\, d\,\mathrm{P} \ge 0.$$

Put $\Delta_k = X_k - X_{k-1}$, and write

$$\int_A (X_{\tau_2} - X_{\tau_1})\, d\,\mathrm{P} = \int_A \sum_{k=1}^{n} \mathrm{I}_{\{\tau_1 < k \le \tau_2\}}\, \Delta_k\, d\,\mathrm{P} = \sum_{i=1}^{n} \int_{A \cap \{\tau_1 < k \le \tau_2\}} \Delta_k\, d\,\mathrm{P}\,.$$

Now verify that $A \in \mathcal{F}_{\tau_1}$ implies $A \cap \{\tau_1 < k \le \tau_2\} \in \mathcal{F}_{k-1}$, and conclude by the submartingale property of $\{X_k\}$ that the integral in the last expression is nonnegative. $\square$

Of course, "$\ge$" may be replaced with "$=$" in the optional sampling theorem when $\{X_k\}$ is a martingale.

2.2.2 Doob's submartingale inequality

Theorem 2.22 (Submartingale inequality). *Let $\{X_0, X_1, \dots, X_n\}$ be a submartingale. Then for any $c > 0$,*

$$c\,\mathrm{P}\left(\max_{k \le n} X_k \ge c\right) \le \mathrm{E}(X_n^+).$$

Proof. Assume first that X_k is nonnegative. Let $A = \{\max_{k \le n} X_k \ge c\}$. Then $A = A_0 \cup A_1 \cup \cdots \cup A_n$ with the union disjoint, where

$$A_0 = \{X_0 \ge c\},$$

and

$$A_k = \{X_0 < c, \dots, X_{k-1} < c, X_k \ge c\} \quad \text{for } k = 1, \dots, n.$$

Since $A_k \in \mathcal{F}_k$ and $X_k \geq c$ on A_k, we have

$$\int_{A_k} X_n \, d\mathrm{P} \geq \int_{A_k} X_k \, d\mathrm{P} \geq c\,\mathrm{P}(A_k).$$

Summing over k gives $c\,\mathrm{P}(A) \leq \mathrm{E}(X_n)$, as required.

If X_k is not necessarily nonnegative, put $Y_k = X_k^+$. Then Y_k is a nondecreasing convex function of X_k and hence, by Proposition 2.12, $\{Y_k\}$ is a nonnegative submartingale. Now apply the submartingale inequality to $\{Y_k\}$. □

An application of the submartingale inequality is the following, which strengthens Chebyshev's inequality for partial sums of independent mean-zero random variables.

Theorem 2.23 (Kolmogorov's inequality)**.** *Let $X_1, X_2, \ldots$ be independent r.v.'s with mean 0 and finite variance. Put $S_n = X_1 + \cdots + X_n$. Then for any $c > 0$,*

$$c^2 \mathrm{P}\left(\max_{k \leq n} S_k \geq c\right) \leq \mathrm{Var}(S_n).$$

Proof. Since $\{S_n\}$ is a martingale, $\{S_n^2\}$ is a submartingale and it is nonnegative, with $\mathrm{E}(S_n^2) = \mathrm{Var}(S_n)$ because $\mathrm{E}(S_n) = 0$. The result now follows directly from the submartingale inequality. □

2.2.3 Martingale transforms

Definition 2.24. A process $\{C_n\}_{n\geq 1}$ is *previsible* if C_n is $\mathcal{F}_{n-1}$-measurable for each $n \geq 1$.

Definition 2.25. Let $X = \{X_n\}$ be an adapted stochastic process and $C = \{C_n\}$ a previsible process. The *martingale transform* of X by C is the process $Y = \{Y_n\}$ defined by

$$Y_n = \sum_{i=1}^{n} C_i (X_i - X_{i-1}).$$

We denote $Y = C \bullet X$.

Note that if $C_i \equiv 1$ for all i, we have simply $Y_n = X_n$. The martingale transform $C \bullet X$ has a gambling interpretation: Let $X_i - X_{i-1}$ be your net winnings per unit stake at the ith game in a sequence of games. Your stake C_i in the ith game should depend only on the outcomes of the first $i-1$ games, hence C_i should be $\mathcal{F}_{i-1}$-measurable, i.e. the stake process C is previsible. The r.v. $Y_n = (C \bullet X)_n$ represents your total fortune immediately after the nth game. Note that, by definition, $Y_0 \equiv 0$.

Theorem 2.26. *Let $X = \{X_n\}$ be an adapted process and $C = \{C_n\}$ a previsible process.*

(i) If C is nonnegative and uniformly bounded and X is a supermartingale, then $C \bullet X$ is a supermartingale.

(ii) If C is uniformly bounded and X is a martingale, then $C \bullet X$ is a martingale.

(iii) If $\mathrm{E}(C_n^2) < \infty$ and $\mathrm{E}(X_n^2) < \infty$ for all n and X is a martingale, then $C \bullet X$ is a martingale.

Proof. Write $Y = C \bullet X$. Since

$$Y_n - Y_{n-1} = C_n(X_n - X_{n-1})$$

and C_n is $\mathcal{F}_{n-1}$-measurable, it follows from Theorem 1.100(i) that

$$\mathrm{E}[Y_n - Y_{n-1}|\mathcal{F}_{n-1}] = C_n\,\mathrm{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] \le 0$$

if X is a supermartingale, or $= 0$ if X is a martingale. In each of (i)-(iii), the hypothesis implies that $C_n(X_n - X_{n-1})$ is integrable, in the last case because of the Hölder (or Schwartz) inequality. □

2.2.4 Convergence theorems

An important question in martingale theory is, when and in what sense we can expect a (sub-, super-)martingale $\{X_n\}$ to converge as $n \to \infty$. The first main result is known as the *martingale convergence theorem.* We follow Williams, chap. 11.

Definition 2.27. Let $X = \{X_n\}$ be a stochastic process. Fix $N \in \mathbb{N}$, and fix real numbers $a < b$. The *number of upcrossings* $U_N(a, b)$ of the interval $[a, b]$ by X in the time interval $[0, N]$ is the largest integer m for which there exist indices

$$0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_m < t_m \le N$$

such that

$$X_{s_i} < a < b < X_{t_i}, \qquad i = 1, \ldots, m.$$

Lemma 2.28 (Doob's Upcrossing Lemma)**.** *Let X be a supermartingale. Then*

$$(b - a)\,\mathrm{E}\left(U_N(a, b)\right) \le \mathrm{E}\left[(X_N - a)^-\right].$$

Proof. Define a process $C = \{C_n\}$ by

$$C_1 = \mathrm{I}_{\{X_0 < a\}}$$

and for $n \ge 2$,

$$C_n = \mathrm{I}_{\{C_{n-1}=1, X_{n-1} \le b\}} + \mathrm{I}_{\{C_{n-1}=0, X_{n-1} < a\}}\,.$$

Gambling interpretation: wait until the process falls below a. Then play unit stakes until the process gets above b. Then stop playing until the process gets back below a, etc.

Note that C is previsible. Define $Y = C \bullet X$, and verify the inequality

$$Y_N \ge (b - a)U_N(a, b) - (X_N - a)^-. \tag{2.9}$$

By Theorem 2.26, Y is a supermartingale, and hence, since $Y_0 = 0$, $\mathrm{E}(Y_N) \le 0$. Taking expectations on both sides of (2.9) now gives the result. □

Corollary 2.29. *Let X be a supermartingale bounded in L^1; that is,* $\sup_n \mathrm{E}\,|X_n| < \infty$. *Define* $U_\infty(a,b) = \lim_{N\to\infty} U_N(a,b)$ *(which exists in* $\mathbb{Z}_+ \cup \{\infty\}$ *since* $U_N(a,b)$ *is increasing in* N*). Then*

$$\mathrm{P}(U_\infty(a,b) = \infty) = 0. \tag{2.10}$$

Proof. Note that $(X_N - a)^- \le |X_N| + |a|$, so Lemma 2.28 implies

$$(b-a)\,\mathrm{E}\,(U_N(a,b)) \le |a| + \mathrm{E}\,|X_N| \le |a| + \sup_n \mathrm{E}\,|X_n|.$$

Letting $N \to \infty$ gives, by MCT,

$$(b-a)\,\mathrm{E}\,(U_\infty(a,b)) \le |a| + \sup_n \mathrm{E}\,|X_n| < \infty.$$

Any r.v. with finite expectation is finite a.s., hence (2.10). □

Theorem 2.30 (Martingale Convergence Theorem)**.** *Let X be a supermartingale bounded in L^1. Then almost surely,* $X_\infty := \lim_{n\to\infty} X_n$ *exists and is finite.*

Proof. Let $X_* = \liminf X_n$ and $X^* = \limsup X_n$. Define

$$\Lambda := \{X_n \text{ does not converge to a limit in } [-\infty,\infty]\}.$$

Then

$$\Lambda = \{X_* < X^*\} = \bigcup_{a,b\in\mathbb{Q},a<b} \{X_* < a < b < X^*\} =: \bigcup_{a,b\in\mathbb{Q},a<b} \Lambda_{a,b}.$$

Now $\mathrm{P}(\Lambda_{a,b}) = 0$ by (2.10), since $\Lambda_{a,b} \subset \{U_\infty(a,b) = \infty\}$. Therefore, $\mathrm{P}(\Lambda) = 0$ so that $X_\infty := \lim_{n\to\infty} X_n$ exists a.s. in $[-\infty,\infty]$. By Fatous's lemma,

$$\mathrm{E}|X_\infty| = \mathrm{E}(\liminf |X_n|) \le \liminf \mathrm{E}\,|X_n| \le \sup_n \mathrm{E}\,|X_n| < \infty,$$

and hence, X_∞ is finite almost surely. □

Corollary 2.31. *Let X be a nonnegative supermartingale. Then almost surely,* $X_\infty := \lim_{n\to\infty} X_n$ *exists and is finite.*

Proof. Check that X is bounded in L^1. □

The martingale convergence theorem is a good start, but we want more. For instance, we would like to also be able to conclude that $X_n \to X_\infty$ in L^1 (i.e. $\mathrm{E}|X_n - X_\infty| \to 0$) and that X_∞ is itself "part of" the supermartingale, i.e. $\mathrm{E}(X_\infty|\mathcal{F}_n) \le X_n$ a.s. To obtain this stronger conclusion we need a stronger hypothesis. This is where uniform integrability comes in.

Definition 2.32. A collection $\mathcal{C}$ of random variables is *uniformly integrable* (UI) if for each $\varepsilon > 0$ there is $K > 0$ such that

$$\mathrm{E}\left(|X|\,\mathrm{I}_{\{|X|>K\}}\right) < \varepsilon \qquad \forall\, X \in \mathcal{C}. \tag{2.11}$$

Exercise 2.33. If $\{X_n\}$ is UI, then $\{X_n\}$ is bounded in L^1. Give an example to show that the converse if false.

Proposition 2.34. *Let $\mathcal{C}$ be a collection of r.v.'s and suppose that either:*

(i) There is $p > 1$ and $A > 0$ such that $\mathrm{E}(|X|^p) \le A$ for all $X \in \mathcal{C}$ (i.e. $\mathcal{C}$ is bounded in L^p); or

(ii) There is an integrable nonnegative r.v. Y such that $|X| \le Y$ for all $X \in \mathcal{C}$.

Then $\mathcal{C}$ is UI.

Proof. Assume (i). If $x \ge K > 0$, then $x \le K^{1-p}x^p$. Hence for $X \in \mathcal{C}$,

$$\mathrm{E}[|X|\,\mathrm{I}_{\{|X|>K\}}] \le K^{1-p}\,\mathrm{E}[|X|^p\,\mathrm{I}_{\{|X|>K\}}] \le K^{1-p}A,$$

and since the last expression tends to 0 as $K \to \infty$, it follows that $\mathcal{C}$ is UI.

Next, assume (ii). Then for all $X \in \mathcal{C}$ and $K > 0$,

$$\mathrm{E}[|X|\,\mathrm{I}_{\{|X|>K\}}] \le \mathrm{E}[Y\,\mathrm{I}_{\{Y>K\}}] \to 0 \qquad (K \to \infty).$$

Hence, $\mathcal{C}$ is UI. □

The following theorem is what makes uniform integrability a useful concept.

Theorem 2.35. *Let $\{X_n\}$ be a sequence of r.v.'s such that $X_n \to X$ a.s. If $\{X_n\}$ is UI, then $X_n \to X$ in L^1; that is, $\mathrm{E}|X_n - X| \to 0$.*

Proof. Let $\varepsilon > 0$ be given. Define

$$\varphi_K(x) = \begin{cases} -K, & x < K \\ x, & |x| \le K \\ K, & x > K. \end{cases}$$

Note that for all x, $|\varphi(x) - x| \le |x|$. Hence we have for each n,

$$\mathrm{E}(|\varphi_K(X_n) - X_n|) = \mathrm{E}[|\varphi_K(X_n) - X_n|\,\mathrm{I}_{\{|X_n|>K\}}] \le \mathrm{E}[|X_n|\,\mathrm{I}_{\{|X_n|>K\}}]$$

and likewise,

$$\mathrm{E}(|\varphi_K(X) - X|) = \mathrm{E}[|\varphi_K(X) - X|\,\mathrm{I}_{\{|X|>K\}}] \le \mathrm{E}[|X|\,\mathrm{I}_{\{|X|>K\}}].$$

Hence, since $\{X_n\}$ is UI we can find K so large that

$$\mathrm{E}|\varphi_K(X_n) - X_n| < \varepsilon/3 \qquad (n \in \mathbb{N})$$

and

$$\mathrm{E}|\varphi_K(X) - X| < \varepsilon/3.$$

Since φ_K is continuous and $X_n \to X$ a.s., we also have $\varphi_K(X_n) \to \varphi_K(X)$ a.s. And since φ_K is bounded, the BCT implies the existence of $N \in \mathbb{N}$ such that, for $n \geq N$,

$$\mathrm{E}|\varphi_K(X_n) - \varphi_K(X)| < \varepsilon/3.$$

Hence, by the triangle inequality, we have for $n \geq N$,

$$\mathrm{E}|X_n - X| < \varepsilon,$$

and the proof is complete. □

Theorem 2.36. *Let $\{M_n\}$ be a UI martingale. Then $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. and $M_n \to M_\infty$ in L^1. Moreover, for each n,*

$$\mathrm{E}[M_\infty|\mathcal{F}_n] = M_n \quad a.s. \tag{2.12}$$

(Of course, the analogous statements hold for UI sub- or supermartingales.) The important second part of the theorem can be interpreted as saying that M_∞ is "part of" the martingale and is in fact its "last element". From here on, when considering UI (sub-, super-)martingales, we will routinely use the fact that M_∞ exists and satisfies (2.12).

Proof. Since $\{M_n\}$ is UI it is bounded in L^1, and hence, by the Martingale Convergence Theorem, $M_\infty := \lim_{n\to\infty} M_n$ exists and is finite a.s. By Theorem 2.35, $M_n \to M_\infty$ in L^1. Now for $k > n$, we have $\mathrm{E}[M_k|\mathcal{F}_n] = M_k$, and hence, for $F \in \mathcal{F}_n$,

$$\mathrm{E}(M_k\, \mathrm{I}_F) = \mathrm{E}(M_n\, \mathrm{I}_F). \tag{2.13}$$

Now $\mathrm{E}(M_k\, \mathrm{I}_F) \to \mathrm{E}(M_\infty\, \mathrm{I}_F)$ because

$$|\,\mathrm{E}(M_k\, \mathrm{I}_F) - \mathrm{E}(M_\infty\, \mathrm{I}_F)| \leq \mathrm{E}(|M_k - M_\infty|\, \mathrm{I}_F) \leq \mathrm{E}\,|M_k - M_\infty| \to 0.$$

Hence, letting $k \to \infty$ in (2.13) gives

$$\mathrm{E}(M_\infty\, \mathrm{I}_F) = \mathrm{E}(M_n\, \mathrm{I}_F)$$

for all $F \in \mathcal{F}_n$, and this is equivalent to (2.12). □

The last theorem in this section is about convergence of martingales bounded in L^2. A martingale M is an L^2-*martingale* if $\mathrm{E}(M_n^2) < \infty$ for each n. L^2-martingales have the special property that their increments are orthogonal (but not independent!); that is, if $s \leq t \leq u \leq v$, then

$$\mathrm{E}\left[(M_v - M_u)(M_t - M_s)\right] = 0. \tag{2.14}$$

To see this, note that $\mathrm{E}[M_v|\mathcal{F}_u] = M_u$ a.s., or equivalently,

$$\mathrm{E}[M_v - M_u|\mathcal{F}_u] = 0 \qquad \text{a.s.}$$

Thus (since $M_t - M_s$ is $\mathcal{F}_u$-measurable),

$$\begin{aligned}\mathrm{E}\left[(M_v - M_u)(M_t - M_s)\right] &= \mathrm{E}\left[\mathrm{E}\left[(M_v - M_u)(M_t - M_s)|\mathcal{F}_u\right]\right] \\ &= \mathrm{E}\left[(M_t - M_s)\,\mathrm{E}\left[M_v - M_u|\mathcal{F}_u\right]\right] \\ &= 0.\end{aligned}$$

In view of (2.14), any L^2-martingale satisfies

$$\mathrm{E}(M_n^2) = \mathrm{E}(M_0^2) + \sum_{i=1}^{n} \mathrm{E}[(M_i - M_{i-1})^2]. \tag{2.15}$$

Say a martingale M is *bounded in* L^2 if $\sup_n \mathrm{E}(M_n^2) < \infty$.

Theorem 2.37. *Let M be an L^2-martingale.*

(i) M is bounded in L^2 if and only if

$$\sum_{n=1}^{\infty} \mathrm{E}[(M_n - M_{n-1})^2] < \infty.$$

(ii) If M is bounded in L^2, then $M_\infty = \lim_{n\to\infty} M_n$ exists and is finite almost surely, and $M_n \to M_\infty$ in L^2.

Proof. Statement (i) is obvious from (2.15). For (ii), note first that if M is bounded in L^2, then M is bounded in L^1 (why?), so the martingale convergence theorem implies the existence of M_∞. Now for $r \in \mathbb{N}$, the orthogonal increment property (2.14) gives

$$\mathrm{E}[(M_{n+r} - M_n)^2] = \sum_{i=n+1}^{n+r} \mathrm{E}[(M_i - M_{i-1})^2].$$

Hence, by Fatou's lemma,

$$\mathrm{E}[(M_\infty - M_n)^2] \le \sum_{i=n+1}^{\infty} \mathrm{E}[(M_i - M_{i-1})^2].$$

Since the right hand side is the tail of a convergent series, we conclude

$$\lim_{n\to\infty} \mathrm{E}[(M_\infty - M_n)^2] = 0,$$

in other words, $M_n \to M_\infty$ in L^2. □

Exercise 2.38. Show that if M is a martingale bounded in L^2, then

$$\mathrm{E}[(M_\infty - M_n)^2] = \sum_{i=n+1}^{\infty} \mathrm{E}[(M_i - M_{i-1})^2].$$

(Hint: Write $M_\infty - M_n = (M_\infty - M_{n+r}) + (M_{n+r} - M_n)$. Expand the square, and consider what happens upon lettig $r \to \infty$. The Schwartz (or Hölder) inequality could be helpful.)

Remark 2.39. Since the orthogonal increments play a crucial role in the above proof, Theorem 2.37 has no analog for sub- or supermartingales in L^2.

2.2.5 Doob decomposition and the angle-bracket process

Theorem 2.40. *Let $X = \{X_n\}$ be an adapted process. Then there exist a martingale $M = \{M_n\}$ and a previsible process $A = \{A_n\}$ with $M_0 = A_0 = 0$ such that*

$$X_n = X_0 + M_n + A_n, \qquad n \geq 0. \tag{2.16}$$

If $X = X_0 + \tilde{M} + \tilde{A}$ is another such decomposition, then

$$\mathrm{P}(M_n = \tilde{M}_n \text{ and } A_n = \tilde{A}_n \ \forall\, n) = 1.$$

We call (2.16) *the* Doob decomposition *of X. If in addition X is a submartingale, then A is almost surely increasing; that is,*

$$\mathrm{P}(A_{n+1} \geq A_n \ \forall\, n) = 1.$$

Proof. Suppose martingale M and previsible A satisfy (2.16) and are null at zero. Then

$$\mathrm{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] = \mathrm{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}] + \mathrm{E}[A_n - A_{n-1}|\mathcal{F}_{n-1}] = A_n - A_{n-1} \quad \text{a.s.}$$

Thus (since $A_0 = 0$) with probability one,

$$A_n = \sum_{i=1}^{n} \mathrm{E}[X_i - X_{i-1}|\mathcal{F}_{i-1}], \qquad n = 1, 2, \ldots,$$

and then, of course,

$$M_n = X_n - X_0 - A_n = (X_n - X_0) - \sum_{i=1}^{n} \mathrm{E}[X_i - X_{i-1}|\mathcal{F}_{i-1}].$$

This proves the a.s. uniqueness of M and A. Conversely, if we define A_n and M_n as in the above equations, it is easy to see that A is previsible and M is a martingale. □

Now let X be a martingale with $X_0 = 0$ and $\mathrm{E}(X_n^2) < \infty$. Then X^2 is a martingale, and has (a.s. unique) Doob decomposition

$$X_n^2 = M_n + A_n,$$

where M is a martingale and A an increasing, previsible process, both null at zero. Since A is increasing, the limit $A_\infty := \lim_{n\to\infty} A_n$ exists. Finiteness of A_∞ is associated with convergence of the martingale X; see Williams, sec. 12.13. The process A is often written $\langle X \rangle$ and is called the *angle-bracket process* associated with X.

2.2.6 Improved OSTs

The Optional Stopping Theorem and Optional Sampling Theorem in Subsection 2.2.1 impose strong conditions on the stopping time τ, but relatively weak conditions on the (sub-, super-)martingale X. In this subsection we aim to do the opposite. It turns out that the "right" condition on X is uniform integrability. Here we follow Williams, chap. A14.

Lemma 2.41. *Let X be a submartingale and τ a stopping time with $\tau \leq N$, where $N \in \mathbb{N}$. Then $\mathrm{E}|X_\tau| < \infty$ and*

$$\mathrm{E}[X_N|\mathcal{F}_\tau] \geq X_\tau \qquad a.s.$$

Proof. That $\mathrm{E}|X_\tau| < \infty$ follows since

$$|X_\tau| \leq \max\{|X_1|, \dots, |X_N|\} \leq |X_1| + \cdots + |X_n|.$$

The displayed inequality is a special case of Theorem 2.21. (Take $\tau_2 \equiv N$ and $\tau_1 = \tau$ there.) □

Theorem 2.42 (Doob)**.**

(i) Let M be a UI martingale (so M_∞ exists!). Then for any stopping time τ,

$$\mathrm{E}[M_\infty|\mathcal{F}_\tau] = M_\tau \qquad a.s. \tag{2.17}$$

(ii) If X is a UI submartingale and τ is any stopping time, then

$$\mathrm{E}[X_\infty|\mathcal{F}_\tau] \geq X_\tau \qquad a.s.$$

(iii) If X is a UI supermartingale and τ is any stopping time, then

$$\mathrm{E}[X_\infty|\mathcal{F}_\tau] \leq X_\tau \qquad a.s.$$

Proof. By Theorem 2.36 and Lemma 2.41, we have for $k \in \mathbb{N}$,

$$\mathrm{E}[M_\infty|\mathcal{F}_k] = M_k, \qquad \mathrm{E}[M_k|\mathcal{F}_{\tau\wedge k}] = M_{\tau\wedge k}$$

(since $\tau \wedge k$ is a stopping time bounded by k). Thus, the tower law gives

$$\mathrm{E}[M_\infty|\mathcal{F}_{\tau\wedge k}] = M_{\tau\wedge k}. \tag{2.18}$$

Check that if $A \in \mathcal{F}_\tau$, then $A \cap \{\tau \leq k\} \in \mathcal{F}_{\tau\wedge k}$, so (2.18) gives

$$\int_{A\cap\{\tau\leq k\}} M_\infty \, d\mathrm{P} = \int_{A\cap\{\tau\leq k\}} M_{\tau\wedge k} \, d\mathrm{P} = \int_{A\cap\{\tau\leq k\}} M_\tau \, d\mathrm{P}\,. \tag{2.19}$$

We would like to let $k \to \infty$ in this equation and conclude that

$$\int_{A\cap\{\tau<\infty\}} M_\infty \, d\,\mathrm{P} = \int_{A\cap\{\tau<\infty\}} M_\tau \, d\,\mathrm{P}\,. \tag{2.20}$$

The problem is, that we do not yet know if M_τ is integrable, so the DCT can not be applied directly. But for given $F \in \mathcal{F}_\tau$, we can take $A = F \cap \{\tau \le k\} \cap \{M_\tau \ge 0\}$ in (2.19). This set lies in $\mathcal{F}_k$ because M_τ is $\mathcal{F}_\tau$-measurable, so $F \cap \{M_\tau \ge 0\} \in \mathcal{F}_\tau$. We now obtain

$$\int_{F\cap\{\tau\le k\}\cap\{M_\tau\ge 0\}} M_\infty \, d\,\mathrm{P} = \int_{F\cap\{\tau\le k\}} M_\tau^+ \, d\,\mathrm{P}\,. \tag{2.21}$$

The right hand side of this tends by MCT to

$$\int_{F\cap\{\tau<\infty\}} M_\tau^+ \, d\,\mathrm{P},$$

and the left hand side tends by DCT to

$$\int_{F\cap\{\tau<\infty\}\cap\{M_\tau\ge 0\}} M_\infty \, d\,\mathrm{P},$$

since we already know that M_∞ is integrable. Hence

$$\int_{F\cap\{\tau<\infty\}} M_\tau^+ \, d\,\mathrm{P} = \int_{F\cap\{\tau<\infty\}\cap\{M_\tau\ge 0\}} M_\infty \, d\,\mathrm{P}, \tag{2.22}$$

and this shows at the same time that the integral on the left is finite. In the same way, we can show

$$\int_{F\cap\{\tau<\infty\}} M_\tau^- \, d\,\mathrm{P} = -\int_{F\cap\{\tau<\infty\}\cap\{M_\tau< 0\}} M_\infty \, d\,\mathrm{P}, \tag{2.23}$$

so that $\int_{F\cap\{\tau<\infty\}} M_\tau^- \, d\,\mathrm{P}$ is finite as well. But then, we can subtract (2.22) and (2.23) from each other and conclude

$$\int_{F\cap\{\tau<\infty\}} M_\tau \, d\,\mathrm{P} = \int_{F\cap\{\tau<\infty\}} M_\infty \, d\,\mathrm{P}\,.$$

Of course, we also have

$$\int_{F\cap\{\tau=\infty\}} M_\tau \, d\,\mathrm{P} = \int_{F\cap\{\tau=\infty\}} M_\infty \, d\,\mathrm{P},$$

and hence,

$$\int_F M_\tau \, d\,\mathrm{P} = \int_F M_\infty \, d\,\mathrm{P},$$

as desired. This proves (i).

Next, let X be a UI submartingale. Then X has Doob decomposition $X = X_0 + M + A$, where M is a martingale and A is increasing. Since X is bounded in L^1 and M is a martingale, it follows that A is bounded in L^1 and therefore, $\mathrm{E}(A_\infty) < \infty$. Thus, A is UI and as a result, M is UI as well. We now obtain

$$\mathrm{E}[X_\infty|\mathcal{F}_\tau] = X_0 + \mathrm{E}[M_\infty|\mathcal{F}_\tau] + \mathrm{E}[A_\infty|\mathcal{F}_\tau] \geq X_0 + M_\tau + \mathrm{E}[A_\tau|\mathcal{F}_\tau] = X_\tau.$$

This gives (ii). Finally, (iii) follows from (ii) in the usual way, by considering the process $-X$. □

Corollary 2.43 (Optional Sampling Theorem)**.** *If M is a UI martingale and σ and τ are stopping times with $\sigma \leq \tau$, then*

$$\mathrm{E}[M_\tau|\mathcal{F}_\sigma] = M_\sigma \qquad a.s.$$

The analogous statements hold for UI sub- and supermartingales.

Proof. From (2.17) and the tower law, we obtain

$$\mathrm{E}[M_\tau|\mathcal{F}_\sigma] = \mathrm{E}[\mathrm{E}[M_\infty|\mathcal{F}_\tau]|\mathcal{F}_\sigma] = \mathrm{E}[M_\infty|\mathcal{F}_\sigma] = M_\sigma.$$

□

Corollary 2.44 (Optional Stopping Theorem)**.** *Let M be a UI martingale and τ a stopping time. Then M_τ is integrable, and*

$$\mathrm{E}(M_\tau) = \mathrm{E}(M_0).$$

The analogous statements hold for UI sub- and supermartingales.

Proof. That M_τ is integrable follows from (2.17) and the conditional form of Jensen's inequality. Apply the previous corollary with $\sigma = 0$ to get $\mathrm{E}(M_\tau) = \mathrm{E}(M_0)$. If X is a UI sub- or supermartingale, integrability of X_τ follows from the Doob decomposition of X and the argument in the proof of Theorem 2.42. □

Note that Corollary 2.44 does not render Theorem 2.19 redundant: part (iii) of that theorem, which applies to random walks, does not follow from Corollary 2.44!

2.3 Gambler's ruin

See Klebaner, p. 194.

2.4 Markov chains

Definition 2.45. Let $\mathcal{S}$ be a countable (perhaps finite) set. A *Markov chain* with *state space* $\mathcal{S}$ is a stochastic process $\{X_n\}_{n\in\mathbb{Z}_+}$ such that $X_n \in \mathcal{S}$ for each n, and

$$\mathrm{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathrm{P}(X_{n+1} = j | X_n = i_n), \tag{2.24}$$

for all $i_0, i_1, \dots, i_n, j \in \mathcal{S}$. The Markov chain is called *time-homogeneous* if $\mathrm{P}(X_{n+1} = j | X_n = i)$ does not depend on n. In that case, we write

$$p_{ij} = \mathrm{P}(X_{n+1} = j | X_n = i), \qquad i, j \in \mathcal{S}.$$

We call the distribution of X_0 the *initial distribution* of the chain and p_{ij} the *transition probabilities.*

For a time-homogenous Markov chain, the initial distribution and the transition probabilities p_{ij} together determine the finite-dimensional distributions of the process by the usual multiplication rule. We often put the p_{ij} in a matrix called the *transition matrix* of the Markov chain, with p_{ij} being the entry in the ith row and jth column. Note that the row sums of the matrix are equal to 1.

The defining condition of a Markov chain (2.24) can be stated more compactly as

$$\mathrm{P}(X_{n+1} = j | X_0, \dots, X_n) = \mathrm{P}(X_{n+1} = j | X_n).$$

More generally, a real-valued stochastic process $\{X_n\}$ is called a *Markov process* if for each Borel set B in $\mathbb{R}$,

$$\mathrm{P}(X_{n+1} \in B | X_0, \dots, X_n) = \mathrm{P}(X_{n+1} \in B | X_n),$$

and the process is *time-homogeneous* if this probability is independent of n.

Example 2.46. Simple random walk is a time-homogeneous Markov chain with state space $\mathbb{Z}$. More generally, if $X_1, X_2, \dots$ are i.i.d. random variables and $S_n = X_1 + \dots + X_n$, then $\{S_n\}$ is in general not a Markov chain (because it doesn't have a countable state space, unless X_i is discrete), but it is a Markov process.

Example 2.47 (Random walk with absorbing boundaries)**.** Let $\{X_n\}$ be a Markov chain with state space $\{0, 1, \dots, N\}$ and transition matrix

$$\mathcal{P} = \begin{bmatrix} 1 & 0 & 0 & & \dots & 0 \\ q & 0 & p & & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & & \dots & q & 0 & p \\ 0 & & \dots & 0 & 0 & 1 \end{bmatrix}$$

(where $q = 1 - p$). Since $p_{00} = p_{NN} = 1$, we call 0 and N *absorbing states.* The chain behaves just like a random walk until it reaches one of these states, then it stays there forever.

Other examples of Markov chains include higher-dimensional random walks as well as branching processes, birth-death processes, the Ehrenfest model, etc.

The theory of Markov chains revolves mostly around classification of states (recurrent or transient) and existence and uniqueness of limiting distributions. We will not really use Markov chains in this course, but will occasionally refer to the "Markov property" of a process. A particularly important aspect of Markov chains for our purposes is that they possess the so-called *strong Markov property*: If $X = \{X_n\}$ is a Markov chain and τ is a stopping time relative to the natural filtration of X, then

$$\mathrm{P}(X_{\tau+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_{\tau-1} = i_{\tau-1}, X_\tau = i) = \mathrm{P}(X_{\tau+1} = j | X_\tau = i) = p_{ij}, \tag{2.25}$$

for all $i_0, i_1, \dots$ and j in $\mathcal{S}$.

Exercise 2.48. Prove the strong Markov property (2.25). (Hint: partition over the values of τ, and use the fact that $\{\tau = n\} \in \sigma(\{X_0, \dots, X_n\})$.)