Math 6810
(Probability and Fractals)

Spring 2016

Lecture notes

Pieter Allaart
University of North Texas

February 15, 2016

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

1. *Real Analysis* by H. L. Royden (4th edition), Prentice Hall.
2. *Probability and Measure* by P. Billingsley (3rd edition), Wiley.
3. *Probability with Martingales* by D. Williams, Cambridge University Press.
4. *Fractal Geometry: Foundations and Applications* by K. Falconer (2nd edition), Wiley.

Chapter 2

Conditional expectation and martingales

2.1 Conditional expectation

From here on we write I_A for the characteristic function χ_A.

Theorem 2.1. *Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. Then there is a $\mathcal{G}$-measurable and integrable random variable Y such that*

$$\mathrm{E}[X\,\mathrm{I}_G] = \mathrm{E}[Y\,\mathrm{I}_G] \qquad \text{for every } G \in \mathcal{G}, \tag{2.1}$$

or in terms of integrals,

$$\int_G X\,d\,\mathrm{P} = \int_G Y\,d\,\mathrm{P} \qquad \text{for every } G \in \mathcal{G}. \tag{2.2}$$

Moreover, Y is unique up to sets of measure zero; that is, if Y' is another r.v. satisfying (2.2), then $\mathrm{P}(Y' = Y) = 1$.

Proof. Assume first that X is nonnegative. Let P' denote the restriction of P to $\mathcal{G}$, and define a measure Q on $\mathcal{G}$ by $\mathrm{Q}(G) = \int_G X\,d\,\mathrm{P}'$, $G \in \mathcal{G}$. The measure Q is finite because X is integrable, and $\mathrm{Q} \ll \mathrm{P}'$. So by the Radon-Nikodym theorem there exists a (nonnegative) $\mathcal{G}$-measurable function (random variable) Y such that $\mathrm{Q}(G) = \int_G Y\,d\,\mathrm{P}'$ for $G \in \mathcal{G}$, and this Y is unique up to sets of measure zero. Since P and P' agree on $\mathcal{G}$, we have (2.2).

For arbitrary integrable X, apply the above to X^+ and X^- and let Y be the difference of the resulting functions. □

Definition 2.2. The random variable Y in the above theorem is called (a version of) the *conditional expectation of X given $\mathcal{G}$*, and denoted $\mathrm{E}(X|\mathcal{G})$. Thus,

$$\mathrm{E}(\mathrm{E}[X|\mathcal{G}]\,\mathrm{I}_G) = \mathrm{E}(X\,\mathrm{I}_G) \qquad \text{for every } G \in \mathcal{G}.$$

Note that whereas the expectation of a random variable is a (non-random) number, its conditional expectation given $\mathcal{G}$ is again a random variable. By the last statement of Theorem 2.1, any two versions of the conditional expectation $\mathrm{E}(X|\mathcal{G})$ are equal with probability 1.

The most familiar special case is the following.

Definition 2.3. Let X be a random variable. The *σ-algebra generated by X*, denoted $\sigma(X)$, is the collection of all sets of the form $\{\omega : X(\omega) \in B\}$, where $B \in \mathcal{B}(\mathbb{R})$. (Check that this really is a σ-algebra!)

Similarly, if $X_1, \ldots, X_n$ are random variables on the same probability space, then the σ-algebra generated by $X_1, \ldots, X_n$ is the collection of all sets of the form

$$\{\omega : (X_1(\omega), \ldots, X_n(\omega)) \in B\}, \qquad B \in \mathcal{B}(\mathbb{R}^n),$$

and is denoted by $\sigma(X_1, \ldots, X_n)$.

Instead of $\mathrm{E}[X|\sigma(Y)]$ we simply write $\mathrm{E}[X|Y]$, and instead of $\mathrm{E}[X|\sigma(Y_1, \ldots, Y_n)]$ we write $\mathrm{E}[X|Y_1, \ldots, Y_n]$. One interpretation of $\mathrm{E}[X|Y]$ is that it is the "best possible prediction" of X when you know the value of Y.

The following proposition shows that for discrete random variables X and Y, the conditional expectation $\mathrm{E}[X|Y]$ corresponds with our earlier notion of conditional expectation of X given $Y = y$.

Proposition 2.4. *Let X and Y be discrete random variables, where Y takes the values $y_1, y_2, \ldots$. There is a version of $\mathrm{E}[X|Y]$ such that for each i, if $\omega \in \{Y = y_i\}$, then $\mathrm{E}[X|Y](\omega) = \mathrm{E}[X|Y = y_i]$.*

Proof. Note that each set in $\sigma(Y)$ is a union of (finitely or countably many) of the sets $\{Y = y_i\}$. So it suffices to show that

$$\int_{\{Y=y_i\}} X \, d\mathrm{P} = \int_{\{Y=y_i\}} \mathrm{E}[X|Y = y_i] \, d\mathrm{P}\,. \tag{2.3}$$

Let $x_1, x_2, \ldots$ be the possible values of X. Since

$$\mathrm{E}[X|Y = y_i] = \sum_j x_j \, \mathrm{P}(X = x_j|Y = y_i) = \sum_j x_j \frac{\mathrm{P}(X = x_j, Y = y_i)}{\mathrm{P}(Y = y_i)},$$

it follows that

$$\int_{\{Y=y_i\}} \mathrm{E}[X|Y = y_i] \, d\mathrm{P} = \mathrm{E}[X|Y = y_i] \, \mathrm{P}(Y = y_i) = \sum_j x_j \, \mathrm{P}(X = x_j, Y = y_i).$$

On the other hand,

$$\int_{\{Y=y_i\}} X \, d\mathrm{P} = \int X \, \mathrm{I}_{\{Y=y_i\}} \, d\mathrm{P} = \sum_j x_j \, \mathrm{P}(X = x_j, Y = y_i).$$

Hence, we have (2.3). □

Conditional expectation has the same properties as expectation (monotonicity, linearity). There is a conditional version of Jensen's inequality as well as conditional versions of the Monotone and Dominated Convergence Theorems and Fatou's lemma. Below we collect some properties which are specific to conditional expectation. First we need a definition:

Definition 2.5. Two σ-algebras $\mathcal{G}_1$ and $\mathcal{G}_2$ in a common space Ω are *independent* if for each $A \in \mathcal{G}_1$ and each $B \in \mathcal{G}_2$, A and B are independent. A random variable X and a σ-algebra $\mathcal{G}$ are independent if $\sigma(X)$ and $\mathcal{G}$ are independent.

Note that from the definition follows immediately that random variables X and Y are independent if and only if $\sigma(X)$ and $\sigma(Y)$ are independent.

Theorem 2.6 (Properties of conditional expectation)**.** *Let X be an integrable r.v. on $(\Omega, \mathcal{F}, \mathrm{P})$ and let $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2$ be sub-σ-algebras of $\mathcal{F}$.*

(i) ("Taking out what is known") If X is $\mathcal{G}$-measurable, then

$$\mathrm{E}[XY|\mathcal{G}] = X\,\mathrm{E}[Y|\mathcal{G}] \tag{2.4}$$

for every random variable Y for which Y and XY are integrable. In particular, if X is $\mathcal{G}$-measurable then

$$\mathrm{E}[X|\mathcal{G}] = X.$$

(This property does not require that X be integrable!)

(ii) If X and $\mathcal{G}$ are independent, then

$$\mathrm{E}[X|\mathcal{G}] = \mathrm{E}(X).$$

This holds in particular if $\mathcal{G}$ is the trivial σ-algebra, $\mathcal{G} = \{\emptyset, \Omega\}$.

(iii) (Tower law) *If $\mathcal{G}_1 \subset \mathcal{G}_2$, then*

$$\mathrm{E}[E[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathrm{E}[X|\mathcal{G}_1].$$

(iv) The law of double expectation:

$$\mathrm{E}[\mathrm{E}[X|\mathcal{G}]] = \mathrm{E}(X). \tag{2.5}$$

Proof. Property (i) is the most involved; we prove it first for indicator random variables. Let $X = \mathrm{I}_{G_0}$, where $G_0 \in \mathcal{G}$. We must show that

$$\mathrm{E}[\mathrm{I}_{G_0}\,Y|\mathcal{G}] = \mathrm{I}_{G_0}\,\mathrm{E}[Y|\mathcal{G}].$$

This follows since

$$\begin{aligned}
\int_G \mathrm{I}_{G_0}\,Y\,d\,\mathrm{P} &= \int_{G\cap G_0} Y\,d\,\mathrm{P} \\
&= \int_{G\cap G_0} \mathrm{E}[Y|\mathcal{G}]\,d\,\mathrm{P} \qquad \text{(by definition of } \mathrm{E}[Y|\mathcal{G}] \text{ and } G\cap G_0 \in \mathcal{G}) \\
&= \int_G \mathrm{I}_{G_0}\,\mathrm{E}[Y|\mathcal{G}]\,d\,\mathrm{P}\,.
\end{aligned}$$

Thus, (2.4) holds for indicator random variables X. By linearity, it then holds for all simple random variables X. Now let X be an arbitrary random variable. Then there are simple r.v.'s $\{X_n\}$ such that $|X_n| \leq |X|$ and $X_n \to X$ a.s. Since (2.4) holds for each X_n, the conditional form of the DCT implies that it holds for X as well. (Check the details!)

(ii) If X and $\mathcal{G}$ are independent, then X and I_G are independent for every $G \in \mathcal{G}$, and so

$$\int_G X \, d\mathrm{P} = \mathrm{E}(X \, \mathrm{I}_G) = \mathrm{E}(X) \, \mathrm{E}(I_G) = \mathrm{E}(X) \, \mathrm{P}(G) = \int_G \mathrm{E}(X) \, d\mathrm{P} \, .$$

Hence, $\mathrm{E}[X|\mathcal{G}] = \mathrm{E}(X)$.

(iii) Let $\mathcal{G}_1 \subset \mathcal{G}_2$. Then for $G \in \mathcal{G}_1$,

$$\begin{aligned} \int_G \mathrm{E}[X|\mathcal{G}_2] \, d\mathrm{P} &= \int_G X \, d\mathrm{P} \qquad \text{since } G \in \mathcal{G}_1 \text{ implies } G \in \mathcal{G}_2 \\ &= \int_G \mathrm{E}[X|\mathcal{G}_1] \, d\mathrm{P} \qquad \text{by definition of } \mathrm{E}[X|\mathcal{G}_1]. \end{aligned}$$

Hence, $\mathrm{E}[E[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathrm{E}[X|\mathcal{G}_1]$.

Finally, (iv) follows from (ii) and (iii) by taking $\mathcal{G}_1 = \{\emptyset, \Omega\}$. □

Remark 2.7. If $B_1, \ldots, B_n$ is a partition of Ω and $\mathcal{G}$ is the smallest σ-algebra containing each B_i, then (2.5) is just a more compact statement of (1.2).

Definition 2.8. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. For $A \in \mathcal{F}$, the *conditional probability of A given $\mathcal{G}$* is

$$\mathrm{P}(A|\mathcal{G}) := \mathrm{E}[\mathrm{I}_A \, |\mathcal{G}].$$

Exercise 2.9. Show that this definition is consistent with our earlier definition of $\mathrm{P}(A|B)$. (Hint: take $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$.)

2.2 Martingales

Definition 2.10. 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 2.11. 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 2.12. 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 2.13. 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 2.14. Let $\{X_n\}$ be a supermartingale. Show that $\mathrm{E}(X_n) \leq \mathrm{E}(X_0)$ 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 2.15 (Simple random walk)**.** 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

$$S_0 \equiv 0, \qquad \text{and} \quad S_n = X_1 + \cdots + X_n, \quad n \geq 1. \tag{2.6}$$

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 the outcome of a fair coin toss.

Theorem 2.16. *Simple random walk has the following properties:*

(i) (Independent increments) *For all $n_1 < n_2 < \cdots < n_k$, the random variables $S_{n_1}, S_{n_2} - S_{n_1}, \dots, S_{n_k} - S_{n_{k-1}}$ are independent.*

(ii) (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) $S_{n_1}, S_{n_2} - S_{n_1}, \dots, S_{n_k} - S_{n_{k-1}}$ 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, $S_{n-m} \stackrel{d}{=} S_n - S_m$. □

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 2.17 (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 $S_n = X_1 + \cdots + X_n$. Then $\{S_n\}$ is a martingale.

Example 2.18 (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 2.19. *(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:

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

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

2.2.1 Doob's submartingale inequality

Theorem 2.20 (Submartingale inequality)**.** *Let $\{X_0, X_1, \dots, X_n\}$ be a submartingale. Then for any $c > 0$,*

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

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

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

and

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

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 2.19, $\{Y_k\}$ 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 2.21 (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. □

2.2.2 Martingale transforms

Definition 2.22. A process $\{C_n\}_{n\geq 1}$ is *previsible* if C_n is $\mathcal{F}_{n-1}$-measurable for each $n \geq 1$.

Definition 2.23. 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 2.24. *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 2.6(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. □

2.2.3 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 2.25. 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 2.26 (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

$$Y_N \ge (b-a)U_N(a,b) - (X_N - a)^-. \tag{2.7}$$

By Theorem 2.24, Y is a supermartingale, and hence, since $Y_0 = 0$, $\mathrm{E}(Y_N) \le 0$. Taking expectations on both sides of (2.7) now gives the result. □

Corollary 2.27. *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*

$$\mathrm{P}(U_\infty(a,b) = \infty) = 0. \tag{2.8}$$

Proof. Note that $(X_N - a)^- \le |X_N| + |a|$, so Lemma 2.26 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 (2.8). □

Theorem 2.28 (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 (2.8), 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 2.29. *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 2.30. A collection $\mathcal{C}$ of random variables is *uniformly integrable* (UI) if for each $\varepsilon > 0$ there is $K > 0$ such that

$$\mathrm{E}\left(|X|\,\mathrm{I}_{\{|X|>K\}}\right) < \varepsilon \qquad \forall\, X \in \mathcal{C}. \tag{2.9}$$

Exercise 2.31. If $\{X_n\}$ is UI, then $\{X_n\}$ is bounded in L^1. Give an example to show that the converse if false.

Proposition 2.32. *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 2.33. *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 2.34. *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,*

$$\mathrm{E}[M_\infty|\mathcal{F}_n] = M_n \quad a.s. \tag{2.10}$$

(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 (2.10).

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 2.33, $M_n \to M_\infty$ in L^1. Now for $k > n$, we have $\mathrm{E}[M_k|\mathcal{F}_n] = M_k$, and hence, for $F \in \mathcal{F}_n$,

$$\mathrm{E}(M_k \, \mathrm{I}_F) = \mathrm{E}(M_n \, \mathrm{I}_F). \tag{2.11}$$

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 (2.11) 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 (2.10). □

2.2.4 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 necessarily independent!); that is, if $s \leq t \leq u \leq v$, then

$$\mathrm{E}\left[(M_v - M_u)(M_t - M_s)\right] = 0. \tag{2.12}$$

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 (2.12), any L^2-martingale satisfies

$$\mathrm{E}(M_n^2) = \mathrm{E}(M_0^2) + \sum_{i=1}^{n} \mathrm{E}[(M_i - M_{i-1})^2]. \tag{2.13}$$

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

Theorem 2.35. *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 (2.13). 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 (2.12) 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 2.36. 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 2.37. Since the orthogonal increments play a crucial role in the above proof, Theorem 2.35 has no analog for sub- or supermartingales in L^2.