NOTES AND EXERCISES, MARTINGALES, MATH 642:592, Spring 2008

The material covered in lectures 2, 3, and 4 on conditional expectation and discrete time martingales is standard and may be found in RW, Volume I. Therefore, we mainly summarize here and give intuition, so that the student may get an overview of the main results. Similar ground is covered, but less efficiently because of many applications to Markov chains and large number laws, in chapter 6 of O.

1. *Conditional Expectation*

(a) *A Special Case.* We begin with a special case which is easy to understand and contains all the intuition. Remember that events typically come in σ-algebra bundles in probability theory. The simplest class of examples are σ-algebras generated by a disjoint partition, $\{A_1, A_2, \ldots\}$, of Ω into events $A_i \in \mathcal{F}$. This σ-algebra is easily described; it is the collection of all unions, necessarily countable, of events of the partition, plus the empty set. For this discussion, suppose the partition $\{A_1, A_2, \ldots\}$ is fixed and denote the associate σ-algebra by $\mathcal{A}$. Any $\mathcal{A}$ measurable random variable takes the form

$$X(\omega) = \sum_1^\infty c_i \mathbf{1}_{A_i}(\omega).$$

Conversely, if the c_i's are all distinct, $\sigma\{X\}$ (the salgebra generated by X) is equal to $\mathcal{A}$.

Let us suppose now that $\mathbb{P}(A_i) > 0$ for each i. Let X be an integrable random variable. In elementary probability theory the conditional expecation of X given A_i is defined by

$$\mathbb{E}\left[X \mid A_i\right] \triangleq \frac{\mathbb{E}\left[X \mathbf{1}_{A_i}\right]}{I\!\!P(A_i)},$$

and is interpreted as the expectation of X given that we know event A_i has occured, but nothing more. If you take a frequentist approach to probability, the justification for this interpretation comes from the law of large numbers. Imagine a sequence of successive trials of the experiment modeled by the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each trial being independent of the others. If we keep a running average of the values of X observed each time A_i occurs, this average will converge almost surely to $\mathbb{E}\left[X \mid A_i\right]$ as the number of trials increases towards infinity.

In advanced probability, we define the conditional expectation of X given the σ-algebra $\mathcal{A}$ as the *random variable*

$$\mathbb{E}\left[X \mid \mathcal{A}\right](\omega) \triangleq \sum_1^\infty \mathbb{E}\left[X \mid A_i\right] \mathbf{1}_{A_i}(\omega). \tag{1}$$

The idea is that we are analyzing an experiment producing a point ω, we are told only which A_i the outcome ω falls in, and we want to calculate the conditional expectation of X given that outcome. This will be a random variable depending on the outcome. Viewing conditional expectations as random variables allows us to discuss the conditional relationships between different random occurences.

The following lemma is easily verified and gives us an alternative way to think about what a conditional expectation does.

Lemma 1 *Let $\mathcal{A}$ be the σ-algebra generated by a countable disjoint partition into events of positive probability. Let $\mathbb{E}[|X|] < \infty$. Then $\mathbb{E}[X \mid \mathcal{A}]$ is the unique random variable Z satisfying*

(i) Z is $\mathcal{A}$-measurable.

(ii) $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Z]$ for every $A \in \mathcal{A}$.

Exercise 1. Prove this lemma.

This Lemma is extremely important. Property (ii) is one of the principal uses of conditional expection, and it is the basis for generalizing the definition of conditional expectation to any σ-algebra.

(b) *The general definition.* We will extend the definition of conditional expectation to general σ-algebras. The restriction of the probability measure $\mathbb{P}$ to $\mathcal{G}$ will be denoted by $\mathbb{P}\mid_{\mathcal{G}}$. Given a random variable X, let $X^+ = X\mathbf{1}_{\{X\geq 0\}}$ and $X^- = -X\mathbf{1}_{\{X<0\}}$ be its positive and negative parts. We say that X is extended integrable if at most one of $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ is infinite. In this case, $\lambda_X(A) \stackrel{\triangle}{=} \mathbb{E}[\mathbf{1}_A X]$ defines a σ-finite signed measure on $(\Omega, \mathcal{F}, \mathbb{P})$. Note that this measure assign infinite values to a set.

Theorem 1 (and Definition) *Let $\mathcal{G}$ be a σ-algebra contained in $\mathcal{F}$. Let X be a random variable. Assume that $\lambda_X(A) = \mathbb{E}[\mathbf{1}_A X]$ defines a σ-finite signed measure on $(\Omega, \mathcal{G}, \mathbb{P}\mid_{\mathcal{G}})$. (Signed meaures are defined in the lecture 1 notes.) There exists an extended integrable random variable Z such that*

(i) Z is $\mathcal{G}$-measurable.

(ii) $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Z]$ for every $A \in \mathcal{A}$.

This Z is unique up to sets of probability 0; that is, if Z' is a second random variable that satisfies (i) and (ii), then $\mathbb{P}(Z' = Z) = 0$. $\mathbb{E}[X \mid \mathcal{G}]$ is used to denote a version of Z and is called the conditional expectation of X given $\mathcal{G}$. A sufficient condition for $\lambda_X(A)$ to define a signed measure on $(\Omega, \mathcal{G}, \mathbb{P}\mid_{\mathcal{G}})$ is that X be extended integrable.

Remarks: 1. In class, we defined conditional expectation under the assumption that X be integrable, but this is not necessary. For example, the conditional expectation is defined for any positive random variable. Remember that we allow random variables to take values in the extended reals.

2. The proof is just to observe that λ_X is absolutely continuous with respect to $\mathbb{P}\mid_{\mathcal{G}}$ on $(\Omega, \mathcal{G}, \mathbb{P}\mid_{\mathcal{G}})$, $\mathbb{P}\mid_{\mathcal{G}}$ being the restriction of $\mathbb{P}$ to $\mathcal{G}$. Then apply the Radon-Nikodym theorem; we stated this only for bounded signed measures in the lecture notes, but it is true for any signed measure (see the exercise 2 below). It provides a $\mathcal{G}$ measurable random variable Z satisfying $\lambda_X(A) = \mathbb{E}[\mathbf{1}_A Z]$ for $A \in \mathcal{G}$ and so Z automatically satisfies (ii). Note that by this construction of Z, if X is extended integrable

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}\left[X^+ \mid \mathcal{G}\right] - \mathbb{E}\left[X^- \mid \mathcal{G}\right].$$

(c) Let X be a random variable. Recall that $\sigma\{X\}$ is the smallest σ-algebra with respect to which X is measurable; it is the collection of all events of the form $\{\omega; X(\omega) \in U\}$, where U is a Borel subset of $\mathbb{R}$. Similarly $\sigma\{X_1, \dots, X_n\}$ is the σ-algebra of all events of the form $\{\omega; (X_1(\omega), \dots, X_n(\omega)) \in U\}$, where U is a Borel subset of $\mathbb{R}^n$. We define

$$\mathbb{E}[Y \mid X_1, \dots, X_n] \triangleq \mathbb{E}[Y \mid \sigma\{X_1, \dots, X_n\}].$$

(d) The basic properties of conditional expectation are all listed in RW. We restate here just three of them that are especially important.

The tower property. Suppose that $\mathcal{G} \subset \mathcal{H}$, then if X is extended integrable

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{H}] \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}].$$

Factoring property. This says that if Y is $\mathcal{G}$ measurable, then

$$\mathbb{E}[YX \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}] \quad \text{a.s.}$$

To be rigorous this needs assumptions. Thus usual statement assumes that both X and YX are integrable. This can be relaxed.

Conditional Jensen inequality Let ϕ be a convex function and suppose that X and $\phi(X)$ are both integrable, then

$$\phi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\phi(X) \mid \mathcal{G}].$$

- (e) Exercises

  Exercise 2. Prove the Tower property. To simplify, you can restrict to the case that X is integrable. This is a good exercise in applying the definition of conditional expectation and you should certainly do it if you are new to this definition.

  Exercise 3. This exercise help connect our general definition of conditional expectation to that we learn in elementary probability. Let (X, Y) have the joint probability density function $f(x, y)$ (thus $\mathbb{P}((X,Y) \in U) = \int\int_U f(x,y)\,\mathrm{d}x\mathrm{d}y$ for any Borel $U \subset \mathbb{R}^2$.) Let $f_X(x)$ be the density of X. In elementary probability theory we define

  $$\mathbb{E}[Y \mid X = x] \triangleq \int_{\mathbb{R}} y \frac{f(x,y)}{F_X(x)}\,\mathrm{d}y \cdot \mathbf{1}_{\{f_X(x)>0\}}$$

  This is a function of x and for temporary convenience, call it $\psi(x)$. Show that $\psi(X)$ is in fact $\mathbb{E}[Y \mid X]$.

  Exercise 4. Do at least (a)-(d) if you are not yet that familiar with conditioning)

  (a) Prove the factor property when Y is a simple $\mathcal{G}$-measurable random variable, that is, $Y = \sum_1^M c_i \mathbf{1}_{A_i}$, where each A_i belongs to $\mathcal{G}$.

  (b) Prove that when X is non-negative, so is $\mathbb{E}[X \mid \mathcal{G}]$ (this actually follows from the Radon-Nikodym theorem construction, but can be proved directly). Prove also that $|\mathbb{E}[X \mid \mathcal{G}]| \le \mathbb{E}[|X| \mid \mathcal{G}]$.

  (c) Prove that factor property if X and Y are non-negative random variables. Hint: use a sequence of simple $\mathcal{G}$-measurable random variables Y_n such that $Y_n \uparrow Y$ and the monotone convergence theorem.

  (d) Prove the factor property when X and XY are integrable. (Note that using (b) and (c) $|Y||\mathbb{E}[X \mid \mathcal{G}]|$ is integrable. Use simple functions converging to Y such that $|Y_n| \le Y$ for all n.)

  (e) Show that the factor property is true if Y is a non-negative or nonpositive random variable and X and YX are extended integrable. Show that it is true if X is positive or negative and YX is extended integrable. Is it true if both X and YX are assumed to be extended integrable?

  (f) Let ϕ be convex and assume that X is extended integrable and $\mathbb{E}[X] > -\infty$, then $\phi(X)$ is extended integrable and $\mathbb{E}[\phi(X)] > -\infty$. Is conditional Jensen still true?

  Exercise 5. (This is optional. A full discussion/solution may be found in Folland, *Real Analysis: Modern Techniques and their Application*, John-Wiley.) Extending the Radon-Nikodym theorem on a probability space to arbitrary signed measures. (The theorem extend also to the case when the probability measure $\mathbb{P}$ is replaced by a σ-finite measure, as defined in this exercise.)

A signed measure or measure λ on $(\Omega, \mathcal{F})$ is σ-finite if there is a countable disjoint partition of Ω into events A_n such that $|\lambda|(A_n) < \infty$ for every n. (a) First extend the Radon-Nikodym theorem stated in Lecture 1 to σ-finite λ. It suffices to treat the case that λ is positive, because the Jordan decomposition implies that λ can be written as $\lambda^+ - \lambda^-$, where λ^+ and λ^- are mutually singular, postive measures, one of which is finite and the other σ-finite. (remember that by definition a signed measure cannot take on both $-\infty$ and ∞ as values.) Then let $\{A_n\}$ be a partition such that $\lambda(A_n) < \infty$ for each n and apply the Radon-Nikodym theorem for the bounded case on each A_n. (b) Now extend to the arbitrary, signed measure case. Again it suffices to prove the result when λ is positive. In this case, show there is an $A \in \mathcal{F}$ on which λ is σ-finite and such that $\mathbb{P}(B) \leq \mathbb{P}(A)$ for all events B on which λ is σ-finite. Apply the Radon-Nikodym theorem for the σ-finite case on A. If C is an event disjoint from A, either $\lambda(A) = \mathbb{P}(A) = 0$ or $\mathbb{P}(C) > 0$ and $\lambda(F) = \infty$. Use this to complete the proof of existence. Prove almost-sure uniqueness.

2. *Martingales in discrete time; definition*

(a) *Filtrations:* A filtration $\mathbb{F}$ is an increasing family of σ-algebras, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \cdots$. The members of the family are always assumed to be contained in the σ-algebra $\mathcal{F}$ of the base probability space.

We shall think of $\mathcal{F}_n$ as the σ-algebra of events that can happen up to time n; specifically, at time n, we "know" $\mathcal{F}_n$ if for evey $A \in \mathcal{F}_n$, we can say whether ω, the outcome of the experiment modeled by $(\Omega, \mathcal{F}, \mathbb{P})$, belongs to A_n or not.

A nice and simple example to keep in mind as we go along is to let each $\mathcal{F}_n$ is the σ-algebra associated to a countable partition of Ω. This will be a filtration if each partition is a refinement of the preceding partition. If $\mathcal{F}_n$ represents what is known at time n, then we know at each time n which event of the partition ω is in, and so we are thus getting a more and more refined idea of the location of ω as time passes. For a concrete example, let $\Omega = (0, 1]$ and let $\mathcal{F}_n$ be the partition defined by the intervals $((k-1)/2^n, k/2^n]$, $1 \leq k \leq 2^n$.

Important definition: A stochastics process $\{X_n\}_{n \geq 0}$ is **adapted** to the filtration $\mathbb{F}$ (or is $\mathbb{F}$-adapted) if X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$. Intuitively, this means that the value of X_n is recorded in the information represented by $\mathcal{F}_n$.

(b) *Martingales.* Martingales, submartingales, and supermartingales are models, imposing the minimal assumptions that correspond to fair, favorable, and unfavorable games, respectively.

Definition. A (discrete-time) $\mathbb{F}$-martingale (or martingale with respect to $\mathbb{F}$) is a random process $\{X_n\}_{n\geq 0}$ satisfying:

(i) $\mathbb{E}[|X_n|] < \infty$ for all $n \geq 0$;

(ii) $\{X_n\}_{n\geq 0}$ is $\mathbb{F}$-adapted;

(iii) For each $n \geq 0$, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$, almost-surely, for all n.

$\{X_n\}_{n\geq 0}$ is an $\mathbb{F}$-submartingale if instead of (iii), $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$, a.s, for all n; it is a supermartingale if instead $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n$, a.s., for all n. Of course, a martingale is both a submartingale and a martingale.

When the filtration is clear, we usually omit refering to the filtration in specifying that a process is a martingale.

It is helpful for intuition to think of X_n as ones current fortune in a game of chance in which a dollar is bet on each play; the increment $X_n - X_{n-1}$ is the amount won or lost on play n; X_0 is the initial amount of money we start with. The conditional increment $\mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] - X_{n-1}$ is what we expect to win or lose given everything we know before play n.

Exercise 6. This is a simple exercise using the tower property of conditional expectation. Let $\{X_n\}$ be an $\mathbb{F}$-martingale (sub- or supermartingale). Let $\mathcal{G}_n = \sigma\{X_0, X_1, \ldots, X_n\}$. Then $\{X_n\}$ is a $\{\mathcal{G}_n\}$-martingale (sub- or supermartingale).

Examples. (a) Mean zero random walks. Let $\xi_1, \xi_2, \ldots$ be independent, integrable random variables with 0 mean. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and, for $n \geq 1$, let $\mathcal{F}_n = \sigma\{\xi_1, \xi_2, \ldots, \xi_n\}$. Then

$$X_n \stackrel{\triangle}{=\!=} \begin{cases} 0, & \text{if } n = 0; \\ \sum_1^n \xi_i, & \text{if } n \geq 1. \end{cases}$$

defines an $\mathbb{F}$-martingale.

(b) "Geometric" random walks. Let $\xi_1, \xi_2, \ldots$ be independent, identically distributed random variables, and let $\mathbb{F}$ be the filtration they generate, as defined in the previous example. Assume that there is an $s \neq 0$, such that the moment generating function $M(s) = \mathbb{E}\left[e^{s\xi_i}\right]$ is finite at s. Define $X_0 = 1$ and for $n \geq 1$,

$$X_n = \frac{\exp\{\sum_1^n \xi_i\}}{M^n(s)}$$

This is a martingale. If the ξ_i's are normal with mean μ and variance σ^2, $M(s) = e^{\mu s + \sigma^2 s/2}$ and

$$X_n = \exp\{\sum_1^n (\xi_i - \mu) - n\sigma^2/2\}.$$

(c) Let ϕ be a convex function and let $\{X_n\}$ be an $\mathbb{F}$-martingale. Then $\{\phi(X_n)\}$ is a $\mathbb{F}$-submartingale. If $\{X_n\}$ is a submartingale and ϕ is increasing and convex, then $\{\phi(X_n)\}$ is a submartingale. These facts follow from application of the conditional Jensen's inequality.

(d) Let Z be an integrable random variable and let $\mathbb{F}$ be a filtration. Define $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$, for each $n \geq 0$. Then $\{X_n\}$ is a martingale with respect to $\mathbb{F}$. If we append ∞ to the index set $\{0, 1, \ldots\}$ as a last index, bigger than all the rest, and define $\mathcal{F}_\infty = \sigma(\cup_i \mathcal{F}_i)$ and $X_\infty \stackrel{\triangle}{=} \mathbb{E}[Z \mid \mathcal{F}_\infty]$, we can think of $\{X_n\}_{0 \leq 1 \leq \infty}$ as a martingale on the extended index set.

(e) Let $\mathbb{F}$ be a filtration and let $\{Y_n\}$ be adapted to $\mathbb{F}$ and assume all its terms are integrable. Then, $X_0 = Y_0$ and

$$X_n = Y_0 + \sum_1^n [Y_k - \mathbb{E}[Y_k \mid \mathcal{F}_{k-1}]]$$

defines a martingale. In this very important construction, we are just subtracting the conditional mean of each next value given the past. Therefore the conditional expectation of each forward increment given the past, namely

$$\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = \mathbb{E}[Y_{n+1} - \mathbb{E}[Y_n \mid \mathcal{F}_n] \mid \mathcal{F}_n] = 0.$$

3. *Martingale integrals in discrete time.* Given two processes $\{X_n\}_{n\geq 1}$ and $h := \{h_n\}_{n\geq 1}$, define

$$(h.X)_n = \sum_1^n h_k [X_k - X_{k-1}]$$

Let $(h.X)_0 = 0$. This is called the discrete time integral of h with respect to $\{X_n\}$. If $\{X_n\}$ is interpreted as the fortune after n plays of a game on which a dollar is bet on each play, then $(h.X)_n$ is the total gain in n plays if, instead, h_k dollars is bet on play k for each k.

Let $\mathbb{F}$ be a filtration. A process $\{h_n\}_{n\geq 1}$ is called $\mathbb{F}$-predictable if h_n is $\mathbb{F}_{n-1}$-measurable for each n. If $\{h_n\}_{n\geq 1}$ is thought of as a betting strategy and $\mathbb{F}_n$ the information available by the end of play n, predictability means that the money we bet on play n can depend only on the information we have up to play $n-1$ for all n. It is a condition that forbids clairvoyance.

If we are betting predictably on a martingale, we should not expect that we can create a betting strategy that makes the game favorable to us. This idea is captured in the following fundamental result, which is summarized in RW as "you can't beat the system."

Theorem 2 *(a) If $\{X_n\}$ is an $\mathbb{F}$-martingale, if $\{h_n\}$ is $\mathbb{F}$-predictable, and if h_n is bounded for each n (there exists $K_n < \infty$ such that $|h_n| \leq K_n$, a.s.), then $\{(h.X)_n\}$ is also a martingale with respect to $\mathbb{F}$.*

(b) If $\{X_n\}$ is an $\mathbb{F}$-submartingale, if $\{h_n\}$ is $\mathbb{F}$-predictable, non-negative, and if h_n is bounded for each n, then $\{(h.X)_n\}$ is also a submartingale. Moreover, if $0 \le H_k \le K$ for all $k \le n$, $\mathbb{E}\left[(h.X)_n\right] \le K\mathbb{E}\left[X_n - X_0\right]$.

The predictability is key in this theorem. Consider the proof of (a). It is easy to show that $\{(h.X)_n\}$ is adapted to $\mathbb{F}$ and that $(h.X)_n$ is integrable for each n. To show the martingale property, we use the factoring property of conditional expectation. Since h_{n+1} is bounded and $\mathcal{F}_n$-measurable,

$$\mathbb{E}\left[(h.X)_{n+1} - (h.X)_n \mid \mathcal{F}_n\right] = \mathbb{E}\left[h_{n+1}\left(X_{n+1} - X_n\right) \mid \mathcal{F}_n\right] = h_{n+1}\mathbb{E}\left[X_{n+1} - X_n \mid \mathcal{F}_n\right] = 0.$$

The proof of (b) is similar. To prove the last statement of (b), one must use that $\mathbb{E}\left[(X_{k+1} - X_k) \mid \mathcal{F}_k\right] \ge 0$, and, if $0 \le h_k \le K$,

$$\begin{aligned}\mathbb{E}\left[h_{k+1}\left(X_{k+1} - X_k\right)\right] &= \mathbb{E}\left[h_{k+1}\mathbb{E}\left[(X_{k+1} - X_k) \mid \mathcal{F}_k\right]\right] \\ &\le K\mathbb{E}\left[\mathbb{E}\left[(X_{k+1} - X_k) \mid \mathcal{F}_k\right]\right] = K\mathbb{E}\left[X_{k+1} - X_k\right]\end{aligned}$$

Also the results of martingale theory spring ultimately from this rather simple theorem.

4. *Some martingale inequalities.*

Let $\{X_n\}$ be process. For $a < b$, let $U_n([a,b],\{X_k\})$ denote the number of upcrossings of $[a,b]$ by $X_0, X_1, \ldots, X_n\}$, an upcrossing being a piece of the path of $\{X_n\}$ from the first time (after the previous upcrossing) it reaches level a or less until the next time it rises to level b or above. The notation y^+ is used to denote $y\mathbf{1}_{\{y \ge 0\}}$.

Theorem 3 *Doob's Upcrossing Inequality. Let $\{X_n\}$ be a submartingale. For any $a < b$,*

$$\mathbb{E}\left[U_n([a,b],\{X_k\})\right] \le \frac{\mathbb{E}\left[X_n^+\right] + |a|}{b-a}.$$

The proof (sketch): Let $Y_n = (X_n - a)^+$. This is a positive submartingale because $y \to (y-a)^+$ is an increasing convex function. (See, section 2, example (c).) Let $h_n = 1$ if at time $n-1$ the process is in a segment of the path of X_n corresponding to an upcrossing of $[a,b]$; otherwise, set $h_n = 0$. Then $\{h_n\}$ is predictable and $(h.Y)_n \ge (b-a)U_n([a,b],\{X_k\})$. Using Theorem 2 (b), we obtain that

$$\mathbb{E}\left[U_n([a,b],\{X_k\})\right] \le \frac{\mathbb{E}\left[(Y_n - Y_0\right]}{b-a} \le \frac{\mathbb{E}\left[X_n+\right] + |a|}{b-a}.$$

Let $X_n^* = \max\{X_k; 0 \le k \le n\}$. For $p \ge 1$, the L^p norm of a random variable X is denoted by

$$\|X\|_p \triangleq \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p}.$$

We remind the student of Hölder's inequality; if $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$,

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Theorem 4 *Doob's inequalities. (a) Let X_n be a submartingale and let $\lambda > 0$. Then*

$$\mathbb{P}\left(X_n^* \geq \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}\left[X_n \mathbf{1}_{\{X_n^* \geq \lambda\}}\right] \leq \frac{\mathbb{E}\left[X_n^+\right]}{\lambda} \leq \frac{\mathbb{E}\left[X_n\right]}{\lambda}. \tag{2}$$

(b) If $\{X_n\}$ is a martingale, then

$$\mathbb{P}\left(\mid X_n \mid^* \geq \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}\left[\mid X_n \mid \mathbf{1}_{\{\mid X_n \mid^* \geq \lambda\}}\right] \leq \frac{\mathbb{E}\left[\mid X_n \mid\right]}{\lambda}. \tag{3}$$

(c) If $\{X_n\}$ is a martingale or a positive submartingale and $p > 1$,

$$\||X_n|^*\|_p \leq \frac{p}{p-1}\|X_n\|_p. \tag{4}$$

In this theorem, (b) is an immediate consequence of (a) applied to the submartingale $\{|X_n|\}$. (c) is a consequence of a lemma that if X and Y are positive random variables and

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}\left[Y \mathbf{1}_{\{X \geq \lambda\}}\right],$$

then

$$\|X\|_p \leq \frac{p}{p-1}\|Y\|_p.$$

The is proved nicely in RW. The proof uses Hölder's inequality and the following very useful identity, which, if you have not seen it you should derive as an exercise (use Fubini's theorem): if Z is a non-negative random variable.

$$\mathbb{E}\left[Z\right] = \int_0^\infty \mathbb{P}(Z \geq \lambda)\, \mathrm{d}\lambda.$$

We sketch here the standard proof of (a)–see also RW. Let $T \triangleq \min\{k; X_n \geq \lambda\}$ ($T = \infty$ if $\{X_n\}$ never reaches level λ). For $k \geq 1$, let $h_k = \mathbf{1}_{\{T \geq k\}}$. Since

$$\{T \geq k\} = \{X_0 < \lambda, X_1 < \lambda, \ldots, X_{k-1} < \lambda\},$$

it is an event in $\mathcal{F}_{k-1}$ since each X_j, $j \leq k-1$ is $\mathcal{F}_{k-1}$-measurable. This implies that h_k is $\mathcal{F}_{k-1}$-measurable for every k and hence that $\{h_k\}$ is $\mathbb{F}$-predictable. It follows by Theorem 2 that if $\{X_n\}$ is submartingale, then so also is $\{X_0 + (h.X)_n\}$ and

$$\mathbb{E}\left[X_0 + (h.X)_n\right] \leq \mathbb{E}\left[X_n\right]. \tag{5}$$

However $\quad (h.X)_n = \sum_{k=1}^{n} \mathbf{1}_{\{T \geq k\}}(X_k - X_{k-1}) = \sum_{1}^{T\wedge n}(X_k - X_{k-1}) = X_{T\wedge n} - X_0$. Here $T \wedge n$ denotes the minimum of T and n; you can check that the formula works even when $T = 0$. Now observe that the events $\{T \leq n\}$ and $\{X_n^* \geq \lambda\}$. It follows that

$$(h.X)_n = X_{T\wedge n} - X_0 = X_T \mathbf{1}_{\{T\leq n\}} + X_n \mathbf{1}_{\{T>n\}} \geq \lambda \mathbf{1}_{\{X_n^* \geq \lambda\}} + X_n \mathbf{1}_{\{X_n^* < \lambda\}}.$$

By taking expectations on both sides and applying (5), we obtain

$$\lambda \mathbb{P}\left(X_n^* \geq \lambda\right) \leq \mathbb{E}\left[X_n\right] - \mathbb{E}\left[X_n \mathbf{1}_{\{X_n^* < \lambda\}}\right]$$

and the inequalities of (a) then follow easily. $\diamond$

The upcrossing inequality is the basis for an important class of theorems called martingale convergence theorems.

Theorem 5 *(a) Let $\{X_n\}$ be a submartingale. If $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$, then $X_\infty = \lim_{n\to\infty} X_n$ exists and is finite almost-surely and $\mathbb{E}\left[|X_\infty|\right] < \infty$ and*

(b) Let $\{X_n\}$ be a uniformly integrable martingale (submartingale). Then $X_\infty = \lim_{n\to\infty} X_n$ exists and is finite almost-surely. Moreover, $\mathbb{E}\left[|X_n - X_\infty|\right] \to 0$ as $n \to \infty$ and $X_n = \mathbb{E}\left[X_\infty \mid \mathcal{F}_n\right]$ almost-surely ($X_n \leq \mathbb{E}\left[X_\infty \mid \mathcal{F}_n\right]$ almost-surely).

(c) If $\{X_n\}$ is a positive supermartingale, then $X_\infty = \lim_{n\to\infty} X_n$ exists almost-surely.

Note that RW prefer to state the convergence theorems for supermartingales. One can translate between sub- and supermartingale cases, since X is a supermartingale if and only if $-X$ is a submartingale.

The essential element in the proof of (a) is using the upcrossing inequality. Letting $n \to \infty$ in the upcrossing inequality and applying the assumption $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$

$$\mathbb{E}\left[U_\infty([a,b];\{X_k\})\right] < \frac{\sup_n \mathbb{E}\left[X_n^+\right] + |a|}{b-a} < \infty$$

and it follows that $\mathbb{P}\left(U_\infty([a,b];\{X_k\}) = \infty\right) = 0$ whenever $a < b$. Through the identity,

$$\{\lim_{n\to\infty} X_n \ \text{ does not exist}\} = \cup_{a<b;a,b \text{ rational}} U_\infty([a,b];\{X_k\}) \tag{6}$$

it follows that $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely. That $\mathbb{E}\left[|X_\infty|\right] < \infty$ requires an application of Fatou's lemma, and it then follows that X_∞ is finite a.s.

(b) follows from (a) using the fact that a.s. convergence implies convergence in probability and then using Theorem 3 of the notes on lectures 1 and 2 and properties of conditional expectation.

(c) is a consequence of (a) and the fact that if X_n is a positive supermartingale, then $-X_n$ is a negative submartingale and so $X_n^+ = 0$ for all n.

Stopping times and optional stopping theorems.

All what follows is very important in the generalization to continuous time and to the application of martingale theory.

Exercise 7. Prove that the set where X_n does not converge to a finite limit is the set $\cup_{a<b;a,b \text{ rational}} U_\infty([a,b];\{X_k\})$.

We note as corollaries:

Theorem 6 *(a) Lévy's upward convergence. Let* $\mathbb{E}[|Z|] < \infty$. *If* $\mathcal{F}_\infty \triangleq \sigma(\cup_1^\infty \mathcal{F}_i)$, $\mathbb{E}[Z \mid \mathcal{F}_n] \to \mathbb{E}[Z \mid \mathcal{F}_\infty]$ *a.s. and in* L^1.

(b) Lévy-Doob downward convergence. Let $\mathbb{E}[|Z|] < \infty$. *If* $\mathcal{F}_0 \supset \mathcal{F}_{-1} \supset \mathcal{F}_{-2} \supset \cdots$ *and* $\mathcal{F}_{-\infty} = \cap_0^\infty \mathcal{F}_{-n}$, *then* $\mathbb{E}[Z \mid \mathcal{F}_{-n}]$ *converges a.s. and in* L^1 *to* $\mathbb{E}[Z \mid \mathcal{F}_{-\infty}]$.

Notes: The upward convergence theorem is a direct Corollary of Theorem 5 (b) and the fact that the martingale $\{\mathbb{E}[Z \mid \mathcal{F}_n]\}$ is uniformly integrable. This last fact is a consequence of:

Exercise 8. Let $\mathcal{S}$ be a family of σ-algebras. Let $\mathbb{E}[|Z|] < \infty$. Then $\{\mathbb{E}[Z \mid \mathcal{G}]\,;\, \mathcal{G} \in \mathcal{S}\}$ is uniformly integrable.

The downward convergence theorem is not a direct consequence of Theorem 5. Rather, one follows the proof of Theorem 5, using the upcrossing inequality and the fact that for every n, the sequence $(\mathbb{E}[Z \mid \mathcal{F}_{-n}], \ldots, \mathbb{E}[Z \mid \mathcal{F}_0])$ is a martingale.

Definition of stopping time. Given a filtration $\mathbb{F}$, a random variable T taking values in $\{0,1,2,\ldots,\} \cup \{\infty\}$ is called an $\mathbb{F}$-stopping time if

$$\{T \le n\} \in \mathcal{F}_n, \text{ for all } 0 \le n. \tag{7}$$

One should check that this is equivalent to requiring that $\{T = n\}$ be an event in $\mathcal{F}_n$ for each finite n. However the condition (7) is the one that generalizes best to continuous time. The condition says that a decision to stop at time n can be based only on the information available up to time n.

We have already encountered a stopping time in the sketch of the proof of Doob's martingale inequality in Theorem 4. Most stopping times in practice are first entrance times like the one in this proof. Generally, if $\{Y_n\}$ is an $\mathbb{F}$-adapted process and if U is a Borel set,

$$T_U = \min\{n;\, Y_n \in U\}$$

defines a stopping time with respect to $\mathbb{F}$; the student should verify this as an easy exercise. Also an easy exercise is (we essentially did it above): if $h_n = \mathbf{1}_{\{T \ge n\}}$, $n \ge 1$ and if $n \wedge T \triangleq \min(n,T)$, then

$$X_0 + (h.X)_n = X_0 + X_{T\wedge n}.$$

This is true of any random time. But if T is an $\mathbb{F}$-stopping time, then $\{\mathbf{1}_{\{T \ge n\}}\}$ is $\mathbb{F}$-predictable. As an immediate consequence we derive,

Theorem 7 *Let $\{X_n\}$ be an $\mathbb{F}$-martingale (sub- or supermartingale). Let T be an $\mathbb{F}$-stopping time. Then $\{X_{T\wedge n}\}$ is a martingale (sub- or supermartingale.)*

This theorem is essentially what is used in the proof of Theorem 4. Heuristically, it says that we cannot beat the system using a stopping strategy.

$\{X_{T\wedge n}\}$ is refered to as the process $\{X_n\}$ stopped at T.

When $\{X_n\}$ is a martingale and T a stopping time, we find that because $\{X_{T\wedge n}\}$ is a martingale, $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]$ for all n—our expected fortune at T or n, whichever comes first, is exactly what we have today at time 0. If we could exchange limits with respect to n and expectation, and if $T < \infty$ almost surely, we would get $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. This will not always be true; think of the symmetric Bernoulli random walk starting at 0, so that $\mathbb{E}[X_o] = 0$. This random walk is a martingale. But it will reach any level eventually. So, if T is the first time under which this is valid are called *optional stopping theorems.* They include theorems extending the martingale property to stopping times. To state such a theorem, we need another very important definition:

Definition of stopped σ-algebra. Let T be an $\mathbb{F} = \{\mathcal{F}_n\}_{n\geq 0}$-stopping time. Define

$$\mathcal{F}_T \triangleq \{A;\, S \in \mathcal{F}, A \cap \{T \leq n\} \in \mathcal{F}_n \ \ \forall\, n \geq 0\}\,.$$

Exercise 9. Verify that $\mathcal{F}_T$ is a σ-algebra.

Heuristically, A is the set of events of the filtration which can occur up to the random time. It is very useful to note that if T is a stopping time and if $A \in \mathcal{F}_T$, then the random time

$$T_A(\omega) \triangleq \begin{cases} T(\omega) & \text{if } \omega \in A; \\ \infty, & \quad\text{if not.} \end{cases}$$

is also a stopping time and $\mathbf{1}_{\{T_A<\infty\}} = A \cap \{T < \infty\}$.

Exercise 9. If S and T are $\mathbb{F}$ stopping times and $S \leq T$, show that $\mathcal{F}_S \subset \mathcal{F}_T$. In fact, verify all the properties listed in the exercise on page 159 of RW, Volume I.

Theorem 8 *(Optional stopping) Let $\{X_n\}$ be an $\mathbb{F}$-martingale (respectively, submartingale), and let T be an $\mathcal{F}$-stopping time. Suppose*

(i) $E\left[|X_T|\mathbf{1}_{\{T<\infty\}}\right] < \infty$*; and*

(ii) $\liminf_{n\to\infty} E\left[|X_n|\mathbf{1}_{\{T>n\}}\right] = 0.$

Then

$$E\left[X_T\mathbf{1}_{\{T<\infty\}}\right] = E[X_0] \quad \Big(\textit{respectively,}\ \ E\left[X_T\mathbf{1}_{\{T<\infty\}}\right] \geq E[X_0]\,.\Big) \tag{8}$$

Under the same conditions, if S and T are stopping times and if $S \leq T < \infty$ a.s., then

$$\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S, \quad a.s., (\textit{respectively,} \mathbb{E}[X_T \mid \mathcal{F}_S] \geq X_S, a.s. \tag{9}$$

Conditions (i) and (ii) supply what is needed to interchange limit and integration in the identity $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ in the martingale case.

The the proof of the second statement includes the first (with a slight modification) in the case $S = 0$. So we stick to the second statement. For the second statement, take $A \in \mathcal{F}_S$. It is enough (why?) to show that

$$\mathbb{E}[\mathbf{1}_A X_T] = \mathbb{E}[\mathbf{1}_A X_S] \quad \text{for every } A \in \mathcal{F}_S.$$

Fix $A \in \mathcal{F}_S$. Let $m \leq n$. Then, since $A \cap \{S = m\} \in \mathcal{F}_m$ and $\{X_{T\wedge n}\}$ is a martingale,

$$\mathbb{E}\left[\mathbf{1}_{A\cap\{S=m\}} X_{T\wedge n}\right] = \mathbb{E}\left[\mathbf{1}_{A\cap\{S=m\}} X_{T\wedge m}\right] = \mathbb{E}\left[\mathbf{1}_{A\cap\{S=m\}} X_{T\wedge S}\right] = \mathbb{E}\left[\mathbf{1}_{A\cap\{S=m\}} X_S\right].$$

On the other hand, since $S > n$ implies $T > n$, $\mathbb{E}\left[\mathbf{1}_{A\cap\{S>n\}} X_{T\wedge n}\}\right] = \mathbb{E}\left[\mathbf{1}_{A\cap\{S>n\}} X_n\}\right]$. Therefore

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_A X_{T\wedge n}] &= \sum_0^n \mathbb{E}\left[\mathbf{1}_{A\cap\{S=m\}} X_{T\wedge n}\right] + \mathbb{E}\left[\mathbf{1}_{A\cap\{S>n\}} X_{T\wedge n}\}\right] \\
&= \mathbb{E}\left[\mathbf{1}_{A\cap\{S\leq n\}} X_S\right] + \mathbb{E}\left[\mathbf{1}_{A\cap\{S>n\}} X_n\}\right].
\end{aligned}$$

By taking $n \to \infty$ along a subsequence for which the second term goes to zero, which we can do by (ii), and by using (i) and dominated convergence, the right hand side will have the limit $\mathbb{E}[\mathbf{1}_A X_S]$. On the other hand

$$\mathbb{E}[\mathbf{1}_A X_{T\wedge n}] = \mathbb{E}\left[\mathbf{1}_A X_T \mathbf{1}_{\{T\leq n\}}\right] + \mathbb{E}\left[\mathbf{1}_A X_n \mathbf{1}_{\{T>n\}}\right],$$

and by taking $n \to \infty$ along a subsequence for which, by (ii), the second term goes to 0, and applying dominated convergence for the first term, by virtue of (i), we can conclude that this tends to $\mathbb{E}[\mathbf{1}_A X_T]$. We therefore conclude that $\mathbb{E}[\mathbf{1}_A X_T] = \mathbb{E}[\mathbf{1}_A X_S]$, as required. $\diamond$

There are a host of different, more easily checked conditions implying the assumption (i) and (ii) of the theorem. Some of these are stated in RW. Perhaps the simplest set of conditions is simply that $\mathbb{P}(T < \infty) = 1$ and that $\{X_n\}$ be uniformly integrable; see RW page 159, Theorem 59.1. In this case X_∞ is defined an in L^1 and (9) is true even for stopping times with positive probability of being infinite. See RW for a direct proof.

For many applications of optional stopping see Chapter 6 of O.