MATH 230B Notes

Logan Bell

January 27, 2022

Contents

1 January 4

1.1 Introduction and Motivation

In this class, we will primarily discuss conditional probability, expectation, and martingales.

We know from undergraduate studies that conditional probability looks like $\mathbb{P}[A \mid X = x]$, and similarly for conditional expectation $\mathbb{E}[Y \mid X = x]$. The meanings are clear for discrete random variables:

$$\mathbb{P}[A \mid X = x] = \frac{\mathbb{P}[A \cap \{X = x\}]}{\mathbb{P}[X = x]}, \quad \text{provided } \mathbb{P}[X = x] \neq 0.$$

Things become unclear when $\mathbb{P}[X = x] = 0$. But this is important when X is a continuous random variable. Usually this is interpreted as

$$\lim_{\varepsilon \to 0} \mathbb{P}[A \mid X \in (x - \varepsilon, x + \varepsilon)]$$

for continuous random variables X. However, there are many problems with this interpretation, such as existence of the limit. The bigger problem is the **Borel-Kolmogorov Paradox**.

Example 1.1: Borel-Kolmogorov Paradox

Let (X, Y) be uniformly distributed on the unit disk in $\mathbb{R}^2$.

Question: What is the distribution of Y given $X = 0$, i.e., given that (X, Y) is on the y-axis?

Answer: Obviously uniform on $[-1, 1]$. Take the limit of the conditional distribution given that $X \in (-\varepsilon, \varepsilon)$.

Objection: The event $\{X = 0\}$ can also be written in a different way! Let (R, Θ) be (X, Y) in polar coordinates, chosen so that $\Theta \in [0, 2\pi)$. Then $\{X = 0\}$ is the same as the event $\big\{\Theta \in \{\pi/2, 3\pi/2\}\big\}$. So we can think alternatively think of the conditional distribution of Y given $X = 0$ as a limit of the conditional distribution of Y given

$$\Theta \in (\pi/2 - \varepsilon, \pi/2 + \varepsilon) \cup (3\pi/2 - \varepsilon, 3\pi/2 + \varepsilon)$$

as $\varepsilon \to 0$. We can check that this limiting conditional distribution is *not* uniform on $[-1, 1]$. It has density proportional to $|y|$.

Why does this happen? Though the limiting events are the same, the approximate events for a given ε are slightly different and give a different conditional distribution on Y. The limiting distributions thus maintain a discrepancy.

This example shows that probability conditioned on an event of probability 0 that is obtained as a limit of a sequence of events with non-zero probability generally depends on the choice of approximating events. This is not acceptable for a rigorous treatment of conditional probability. How can we resolve the paradox?

In the above example, we can begin to resolve the paradox by abandoning the view that the conditional distribution of Y given $X = 0$ is the same object as the conditional distribution of Y given $\Theta \in \{\pi/2, 3\pi/2\}$. The idea is that, fixing A, we can define a function $f(x) = \mathbb{P}[Y \in A \mid X = x]$ and interpret $\mathbb{P}[Y \in A \mid X = 0]$ as $f(0)$. The function $g(\theta) = \mathbb{P}[Y \in A \mid \Theta = \theta]$ is a different function. Given these functions, we can plug in X or Θ to get a random variable: $f(X)$ is written as $\mathbb{P}[Y \in A \mid X]$.

1.2 Rigorously Defining Conditional Expectation

Let us now discuss conditional expectation. In rigorous proability theory, we define conditional expectation of a random variable X given a σ-algebra $\mathcal{G}$ instead of given an event. In Example 1.1, we condition Y on $\sigma(X)$. Conditional expectation is denoted by $\mathbb{E}[X \mid \mathcal{G}]$, and it is a random variable. In the case that $\mathcal{G} = \sigma(Y)$ for another random variable Y, we generally write $\mathbb{E}[X \mid Y]$.

Definition 1.1: Conditional Expectation

*Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let X be a real-valued integrable random variable defined on Ω. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. An integrable random variable Y is called **the conditional expectation of X given $\mathcal{G}$**, i.e., $Y = \mathbb{E}[X \mid \mathcal{G}]$, if*

1. *Y is $\mathcal{G}$-measurable, and*
2. *For any $B \in \mathcal{G}$, we have $\mathbb{E}[X\delta_B] = \mathbb{E}[Y\delta_B]$.*

We will later see that this definition is logical in the sense that conditional expectation exists. For now, we will check that conditional expectation is essentially unique if it exists.

Lemma 1.1: Uniqueness of Conditional Expectation

If $\mathbb{E}[X \mid \mathcal{G}]$ exists, then it is unique almost everywhere.

Before proving the lemma, we give an example.

Example 1.2: Uniqueness of Conditional Expectation

Consider again the random variable (X, Y) that is uniform on the unit disk in $\mathbb{R}^2$. Let $\mathcal{G} = \sigma(X)$. What is $\mathbb{E}\big[Y^2 \mid \mathcal{G}\big]$?

Guess: Given $X = x$, we know $Y \sim \text{Unif}\,[-\sqrt{1-x^2}, \sqrt{1-x^2}]$. So

$$\begin{aligned}
\mathbb{E}\big[Y^2 \mid \mathcal{G}\big] &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{2\sqrt{1-x^2}} y^2\, dy \\
&= \frac{1}{2\sqrt{1-x^2}} \frac{1}{3}\Big(2(1-x^2)^{3/2}\Big) \\
&= \frac{\sqrt{2}}{3}(1-x^2)
\end{aligned}$$

Verification: If $\mathcal{G} = \sigma(X)$ and $B \in \mathcal{G}$, then B must be of the form $\{X \in A\}$ for some (Borel) set A. We need to show

$$\mathbb{E}\left[\frac{\sqrt{2}}{3}(1-X^2)\delta_{\{X\in A\}}\right] = \mathbb{E}\big[Y^2\delta_{\{X\in A\}}\big]$$

for any A. Indeed, we have

$$\begin{aligned}
\mathbb{E}\big[Y^2\delta_{\{X\in A\}}\big] &= \frac{1}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2\delta_{\{x\in A\}}\, dy\, dx \\
&= \frac{1}{\pi}\int_A \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2\, dy\, dx \\
&= \frac{1}{\pi}\int_A 2\sqrt{1-x^2}\frac{\sqrt{2}}{3}(1-x^2)\, dx \\
&= \mathbb{E}\left[\frac{\sqrt{2}}{3}(1-X^2)\delta_{\{X\in A\}}\right]
\end{aligned}$$

where the last equality follows since $\frac{2}{\pi}\sqrt{1-x^2}$ is the pdf of X. We have thus verified the guess.

We now prove Lemma 1.1.

Proof of Lemma 1.1. Let Y_1 and Y_2 be two candidates for $\mathbb{E}[X \mid \mathcal{G}]$. Let $B = \{\omega : Y_1(\omega) > Y_2(\omega)\}$. Since Y_1 and Y_2 are both $\mathcal{G}$-measurable by Definition 1.1, we know $B \in \mathcal{G}$, as it is the pre-image $(Y_1 - Y_2)^{-1}((0,\infty))$. Thus $\mathbb{E}[Y_1\delta_B] = \mathbb{E}[X\delta_B] = \mathbb{E}[Y_2\delta_B]$. Hence, $\mathbb{E}[(Y_1 - Y_2)\delta_B] = 0$. The random variable $(Y_1 - Y_2)\delta_B$ is non-negative, since we force it to be 0 whenever $Y_1 \leq Y_2$. Thus, since its expectation is 0, we know $(Y_1 - Y_2)\delta_B = 0$ almost surely. Therefore $Y_1 \leq Y_2$ almost surely. By symmetry, $Y_2 \leq Y_1$ almost surely. □

Going back to Example 1.2, we see that $\mathbb{E}\big[Y^2 \mid X = x\big]$ is not well-defined. The random variable

$$Z = \begin{cases} \frac{\sqrt{2}}{3}(1 - X^2) & \text{if } X \neq x \\ 100 & \text{if } X = x \end{cases}$$

is also a viable candidate for $\mathbb{E}\big[Y^2 \mid X\big]$. So we can change our conditional expectation at a "small" number of points and still have a valid conditional expectation, making conditioning X on a single point to be ill-defined.

We now turn our focus to existence.

Theorem 1.1: Existence and Uniqueness of Conditional Expectation

For any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, any integrable random variable X defined on Ω, and any sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, we have $\mathbb{E}[X \mid \mathcal{G}]$ exists and is unique almost everywhere.

To prove Theorem 1.1, we need some preliminary results.

Lemma 1.1: Existence of Conditional Expectation for L^2 Random Variables

$\mathbb{E}[X \mid \mathcal{G}]$ *exists if* $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Proof. Let $X \in L^2$. Let $a = \inf_{Y \in L^2(\mathcal{G})} \mathbb{E}\big[(X - Y)^2\big]$. Then there exists $Y_1, Y_2, \ldots \in L^2(\mathcal{G})$ such that $\mathbb{E}\big[(X - Y_n)^2\big] \to a$ as $n \to \infty$. Take any integers n, m. Now recall the **parallelogram identity**:

$$(u - v)^2 + (u + v)^2 = 2u^2 + 2v^2.$$

Take $u = X - Y_n$ and $v = X - Y_m$. This gives

$$(Y_n - Y_m)^2 + (2X - Y_n - Y_m)^2 = 2(X - Y_n)^2 + 2(X - Y_m)^2.$$

Taking expectation on both sides gives

$$\mathbb{E}\big[(Y_n - Y_m)^2\big] = 2\,\mathbb{E}\big[(X - Y_n)^2\big] + 2\,\mathbb{E}\big[(X - Y_m)^2\big] - 4\,\mathbb{E}\Big[(X - \frac{Y_n + Y_m}{2})^2\Big]. \tag{1}$$

We know $\frac{Y_n + Y_m}{2} \in L^2(\mathcal{G})$. Thus, $\mathbb{E}\big[(X - \frac{Y_n+Y_m}{2})^2\big] \geq a$, and so (1) implies

$$\mathbb{E}\big[(Y_n - Y_m)^2\big] \leq 2\,\mathbb{E}\big[(X - Y_n)^2\big] + 2\,\mathbb{E}\big[(X - Y_m)^2\big] - 4a.$$

As $n, m \to a$, we know $\mathbb{E}\big[(X - Y_n)^2\big] \to a$ and $\mathbb{E}\big[(X - Y_m)^2\big] \to a$ by assumption, and so

$$\limsup_{m,n\to a} \mathbb{E}\big[(Y_n - Y_m)^2\big] \le 0$$

which implies equality since $(Y_n - Y_m)^2$ is non-negative. Thus, $\{Y_n\}_{n\ge 1}$ is Cauchy in $L^2(\mathcal{G})$. But $L^2(\mathcal{G})$ is complete, so there exists $Y \in L^2(\mathcal{G})$ such that $Y_n \to Y$ in L^2 as $n \to \infty$ (i.e., $\lim_{n\to\infty} \mathbb{E}\big[|Y_n - Y|^2\big] \to 0$). Hence,

$$\mathbb{E}\big[(X - Y)^2\big] = \lim_{n\to\infty} \mathbb{E}\big[(X - Y_n)^2\big] = a.$$

We now need to show that $Y = \mathbb{E}[X \mid \mathcal{G}]$. Take any $B \in \mathcal{G}$. We need to show that $\mathbb{E}[X\delta_B] = \mathbb{E}[Y\delta_B]$. For any $\lambda \in \mathbb{R}$, we know $Y + \lambda\delta_B$ is in $L^2(\mathcal{G})$. Thus, $\mathbb{E}\big[(X - Y - \lambda\delta_B)^2\big] \ge a$. But we know

$$\mathbb{E}\big[(X - Y - \lambda\delta_B)^2\big] = \mathbb{E}\big[(X - Y)^2\big] - 2\lambda\,\mathbb{E}[(X - Y)\delta_B] + \lambda^2\mathbb{P}[B]. \tag{2}$$

This is a quadratic polynomial in λ, which is minimized at $\lambda = 0$ (at this point, $\mathbb{E}\big[(X - Y - \lambda\delta_B)^2\big]$ achieves the infimum a). Thus the derivative of (2) with respect to λ at the point $\lambda = 0$ is 0, so $\mathbb{E}[(X - Y)\delta_B] = 0$. $\square$

Recall now that a random variable is called **simple** if it takes value in a finite set (of finite values). Clearly simple random variables are in L^2. The idea for how to use Lemma 1.1 to prove Theorem 1.1 (which we will prove tomorrow) is to approximate the random variable X by simple random variables and take the limit.

2 January 6

2.1 Lemmas Involving Simple Random Variables

We first prove some lemmas about the properties of conditional expectation for simple random variables, which we will combine to prove later theorems.

Lemma 2.1: Linearity for Simple Random Variables

Let X_1, X_2 be simple random variables. Then for any $a, b \in \mathbb{R}$,

$$\mathbb{E}[aX_1 + bX_2 \mid \mathcal{G}] = a\,\mathbb{E}[X_1 \mid \mathcal{G}] + b\,\mathbb{E}[X_2 \mid \mathcal{G}] \qquad \text{a.e.}$$

Proof. Let $Y_1 = \mathbb{E}[X_1 \mid \mathcal{G}]$ and $Y_2 = \mathbb{E}[X_2 \mid \mathcal{G}]$ (here, we mean Y_1 and Y_2 are *versions* of conditional expectation, unique almost everywhere by Lemma 1.1). Then for any $B \in \mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}[(aY_1 + bY_2)\delta_B] &= a\,\mathbb{E}[Y\delta_B] + b\,\mathbb{E}[Y_2\delta_B]\\ &= a\,\mathbb{E}[X_1\delta_B] + b\,\mathbb{E}[X_2\delta_B]\\ &= \mathbb{E}[(aX_1 + bX_2)\delta_B].\end{aligned}$$

Also, $aY_1 + bY_2$ is $\mathcal{G}$-measurable. Thus $aY_1 + bY_2 = \mathbb{E}[aX_1 + bX_2 \mid \mathcal{G}]$. $\square$

Lemma 2.2: Non-negativity for Simple Random Variables

If X is a non-negative simple random variable, then $\mathbb{E}[X \mid \mathcal{G}]$ is also a non-negative random variable.

Proof. Let $Y = \mathbb{E}[X \mid \mathcal{G}]$, and let $B = \{\omega : Y(\omega) < 0\}$. Then $B \in \mathcal{G}$, so $\mathbb{E}[Y\delta_B] = \mathbb{E}[X\delta_B]$. But X is a non-negative random variable, so $\mathbb{E}[X\delta_B] \ge 0$. Thus $\mathbb{E}[Y\delta_B] \ge 0$. But $Y\delta_B$ is a non-positive random variable. Thus $Y\delta_B = 0$ almost surely, and so $Y \ge 0$ almost surely. $\square$

2.2 Existence of Conditional Expectation

Recall that we proved the existence of conditional expectation for L^2 random variables, which we accomplished in a manner similar to projection onto Hilbert spaces. We will now prove that conditional expectation exists for general integrable random variables by building them up from non-negative random variables. Note that we will now generally omit qualifiers such as "almost surely" when equating random variables.

In the next lemma, we do not require integrability because we are dealing with non-negative random variables. Conditional expectation will still be used in an identical sense: $\mathbb{E}[X \mid \mathcal{G}]$ means a $\mathcal{G}$-measurable random variable Y such that $\mathbb{E}[Y\delta_B] = \mathbb{E}[X\delta_B]$ for all $B \in \mathcal{G}$; the uniqueness lemma follows from the same argument as in Lemma 1.1.

Lemma 2.3: Existence for Non-negative Random Variables

If X is a $[0,\infty]$-valued random variable, then $\mathbb{E}[X \mid \mathcal{G}]$ exists.

Proof. Recall that for a $[0,\infty]$-valued random variable X, there exists a sequence of non-negative simple random variables $\{X_n\}_{n\geq 1}$ such that $X_n \nearrow X$. Let $Y_n = \mathbb{E}[X_n \mid \mathcal{G}]$. For any n, $X_n - X_{n-1}$ is a non-negative simple random variable. Therefore, by Lemma 2.1 and Lemma 2.2,

$$\begin{aligned}
Y_n - Y_{n-1} &= \mathbb{E}[X_n \mid \mathcal{G}] - \mathbb{E}[X_{n-1} \mid \mathcal{G}] \\
&= \mathbb{E}[X_n - X_{n-1} \mid \mathcal{G}] \\
&\geq 0 \text{ a.s.}
\end{aligned}$$

Let $A_n = \{\omega : Y_n(\omega) \geq Y_{n-1}(\omega)\}$ so that $\mathbb{P}[A_n] = 1$. Let $A = \bigcap_{n=1}^{\infty} A_n \in \mathcal{G}$. Then $\mathbb{P}[A] = 1$. On A, Y_n is an increasing sequence of non-negative simple random variables. Let $Y = \lim_{n\to\infty} Y_n$ on A (this may be ∞ in certain regions) and 0 on A^c.

Claim: $Y = \mathbb{E}[X \mid \mathcal{G}]$. Since Y_n is $\mathcal{G}$-measurable for all n, so is Y. Now take any $B \in \mathcal{G}$ and note

$$\mathbb{E}[Y\delta_B] = \mathbb{E}[Y\delta_{B\cap A}] \quad (\text{since } \mathbb{P}[A^c] = 0).$$

On $B \cap A$, $Y_n \nearrow Y$, so by the monotone convergence theorem,

$$\begin{aligned}
\mathbb{E}[Y\delta_{B\cap A}] &= \lim_{n\to\infty} \mathbb{E}[Y_n\delta_{B\cap A}] \\
&= \lim_{n\to\infty} \mathbb{E}[X_n\delta_{B\cap A}],
\end{aligned}$$

where the last step holds because $Y_n = \mathbb{E}[X_n \mid \mathcal{G}]$. But again by the monotone convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}[X_n\delta_{B\cap A}] = \mathbb{E}[X\delta_{B\cap A}] = \mathbb{E}[X\delta_B].$$

Therefore, $Y = \mathbb{E}[X \mid \mathcal{G}]$. □

Lemma 2.4: Linearity for Non-negative Random Variables

Linearity holds in the following sense: if X_1, X_2 are $[0,\infty]$-valued random variables, then for all $a, b \geq 0$,

$$\mathbb{E}[aX_1 + bX_2 \mid \mathcal{G}] = a\,\mathbb{E}[X_1 \mid \mathcal{G}] + b\,\mathbb{E}[X_2 \mid \mathcal{G}] \quad \text{a.e.}$$

Proof. The argument is the same as Lemma 2.1, using uniqueness. □

We can now prove Theorem 1.1.

Proof of Theorem 1.1. We know uniqueness by Lemma 1.1, so we need only show existence. Take any integrable X and decompose it into positive and negative parts: $X = X^+ - X^-$, so $|X| = X^+ + X^-$. Since X is integrable, we know $\mathbb{E}[|X|] < \infty$, which implies $\mathbb{E}[X^+] < \infty$ and $\mathbb{E}[X^-] < \infty$. Let $Y_1 = \mathbb{E}[X^+ \mid \mathcal{G}]$ and $Y_2 = \mathbb{E}[X^- \mid \mathcal{G}]$ (which exist by Lemma 2.3). We have

$$\mathbb{E}[Y_1] = \mathbb{E}[Y_1 \delta_\Omega] = \mathbb{E}\big[X^+ \delta_\Omega\big] = \mathbb{E}\big[X^+\big] < \infty$$

and similarly $\mathbb{E}[Y_2] = \mathbb{E}[X^-] < \infty$. Thus, Y_1, Y_2 are integrable random variables. In particular, they are finite a.s. Let $Y = Y_1 - Y_2$ on the set $\{\omega : Y_1(\omega), Y_2(\omega) < \infty\}$, and 0 on the complement.

Claim: $Y = \mathbb{E}[X \mid \mathcal{G}]$. To see this, note Y is integrable and $\mathcal{G}$-measurable since Y_1 and Y_2 are integrable and $\mathcal{G}$-measurable. Now take any $B \in \mathcal{G}$. Then

$$\begin{aligned}
\mathbb{E}[Y \delta_B] &= \mathbb{E}[Y_1 \delta_B] - \mathbb{E}[Y_2 \delta_B] \\
&= \mathbb{E}\big[X^+ \delta_B\big] - \mathbb{E}\big[X^- \delta_B\big] \\
&= \mathbb{E}\big[(X^+ - X^-)\delta_B\big] \\
&= \mathbb{E}[X \delta_B].
\end{aligned}$$

This proves the claim. □

Lemma 2.5: Integrability of Conditional Expectation

If X is integrable, so is $\mathbb{E}[X \mid \mathcal{G}]$. Moreover,

$$\big|\mathbb{E}[X \mid \mathcal{G}]\big| \leq \mathbb{E}[|X| \mid \mathcal{G}] \text{ a.s.}$$

Proof. Note that $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}]$ by the proof of Theorem 1.1. Thus

$$\begin{aligned}
\big|\mathbb{E}[X \mid \mathcal{G}]\big| &= \big|\mathbb{E}\big[X^+ \mid \mathcal{G}\big] - \mathbb{E}\big[X^- \mid \mathcal{G}\big]\big| \\
&\leq \big|\mathbb{E}\big[X^+ \mid \mathcal{G}\big]\big| + \big|\mathbb{E}\big[X^- \mid \mathcal{G}\big]\big| \\
&= \mathbb{E}\big[X^+ \mid \mathcal{G}\big] + \mathbb{E}\big[X^- \mid \mathcal{G}\big] \qquad \text{(since } X^+, X^- \text{ are non-negative)} \\
&= \mathbb{E}\big[X^+ + X^- \mid \mathcal{G}\big] \\
&= \mathbb{E}[|X| \mid \mathcal{G}].
\end{aligned}$$

Thus, $\mathbb{E}\big[\big|\mathbb{E}[X \mid \mathcal{G}]\big|\big] \leq \mathbb{E}[\mathbb{E}[|X| \mid \mathcal{G}]] = \mathbb{E}[|X|] < \infty$. □

2.3 Properties of Conditional Expectation

We have a lot of the same properties for conditional expectation that we do for unconditional expectation. However, since conditional expectation is a random variable rather than a number, it is important that we still prove them. Moreover, the results hold almost surely. We present the properties largely without commentary due to the similarity to unconditional expectation.

Lemma 2.6: Linearity of Conditional Expectation

If X_1, X_2 are integrable random variables, then for all $a, b \in \mathbb{R}$,

$$\mathbb{E}[aX_1 + bX_2 \mid \mathcal{G}] = a\,\mathbb{E}[X_1 \mid \mathcal{G}] + b\,\mathbb{E}[X_2 \mid \mathcal{G}].$$

Proof. Same as before. □

Lemma 2.7: Stability of Constants

If X is constant, then $\mathbb{E}[X \mid \mathcal{G}] = X$.

Proof. Trivial. □

Lemma 2.8: Pulling Out Independent Factors

If X is independent of $\mathcal{G}$, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$ a.s.

Proof. Take any $B \in \mathcal{G}$. Then $\mathbb{E}[X\delta_B] = \mathbb{E}[X]\,\mathbb{E}[\delta_B]$ by independence. This becomes $\mathbb{E}[\mathbb{E}[X]\delta_B]$ by linearity, so $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$ a.s. □

Lemma 2.9: Stability

If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$ a.s.

Proof. For any $B \in \mathcal{G}$, we have $\mathbb{E}[X\delta_B] = \mathbb{E}[X\delta_B]$ (yes, this is a tautology; we write it here to make the defining relation of conditional expectation explicit). Since X is $\mathcal{G}$-measurable, we have $X = \mathbb{E}[X \mid \mathcal{G}]$. □

Lemma 2.10: Monotonicity of Conditional Expectation

If $X \leq Y$ a.s., then $\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]$ a.s.

Proof. The conditional expectation of non-negative random variables is non-negative. Considering $Y - X$ yields the lemma. □

Lemma 2.11: Monotone Convergence Theorem for Conditional Expectation

Let $\{X_n\}_{n\geq 1}$ be a sequence of $[0,\infty]$-valued random variables increasing to X. Then $\mathbb{E}[X \mid \mathcal{G}] = \lim_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}]$ a.s.

Proof. Let $Y_n = \mathbb{E}[X_n \mid \mathcal{G}]$. Take any $B \in \mathcal{G}$. Then by the ordinary monotone convergence theorem, $\mathbb{E}[X\delta_B] = \lim_{n\to\infty} \mathbb{E}[X_n\delta_B]$. But this is $\lim_{n\to\infty} \mathbb{E}[Y_n\delta_B]$. Since Y_n is an increasing sequence a.e., we have by Lemma 2.10 that $Y = \lim_{n\to\infty} Y_n$ exists a.e. Thus, again applying the monotone convergence theorem, we get $\mathbb{E}[Y_n\delta_B] \to \mathbb{E}[Y\delta_B]$ as $n \to \infty$. Hence $Y = \mathbb{E}[X \mid \mathcal{G}]$. □

Lemma 2.12: Fatou's Lemma for Conditional Expectation

If $\{X_n\}_{n\geq 1}$ is a sequence of non-negative random variables, then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \mid \mathcal{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}] \text{ a.s.}$$

Proof. For each n, let $Y_n = \inf_{k\geq n} X_k$. Then $\{Y_n\}$ is an increasing sequence of non-negative random variables with

$$\lim_{n\to\infty} Y_n = Y = \liminf_{n\to\infty} X_n.$$

By Lemma 2.11, we have

$$\begin{aligned}
\mathbb{E}[Y \mid \mathcal{G}] &= \mathbb{E}\Big[\lim_{n\to\infty} Y_n \mid \mathcal{G}\Big] \\
&= \lim_{n\to\infty} \mathbb{E}[Y_n \mid \mathcal{G}] \\
&\leq \lim_{n\to\infty} \inf_{k\geq n} \mathbb{E}[X_k \mid \mathcal{G}] \qquad \text{(by Lemma 2.10)} \\
&= \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}].
\end{aligned}$$

□

Lemma 2.13: Dominated Convergence Theorem for Conditional Expectation

Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables converging a.e. to X. Suppose that there is an integrable random variable Z such that for all n, $|X_n| \leq Z$ a.e. Then

$$\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}] \text{ a.e.}$$

Proof. Note that $X_n + Z$ is a non-negative random variable for all n. Also, $X_n + Z \to X + Z$ a.e. Thus, by Lemma 2.12,

$$\begin{aligned}
\mathbb{E}[X + Z \mid \mathcal{G}] &\leq \liminf_{n\to\infty} \mathbb{E}[X_n + Z \mid \mathcal{G}] \\
&= \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}] + \mathbb{E}[Z \mid \mathcal{G}].
\end{aligned}$$

(Note that everything is finite because X_n, Z are integrable). Subtracting off $\mathbb{E}[Z \mid \mathcal{G}]$ from both sides yields $\mathbb{E}[X \mid \mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}]$. Similarly, $Z - X_n$ is a non-negative random variable for all n, and it converges to $Z - X$ a.e. Using Lemma 2.12 again shows $\mathbb{E}[X \mid \mathcal{G}] \geq \limsup_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}]$ a.s. Thus, $\mathbb{E}[X \mid \mathcal{G}] = \lim_{n\to\infty} \mathbb{E}[X_n \mid \mathcal{G}]$ a.s. □

Lemma 2.14: Conditional Expectation of L^1 Metric

Using the same conditions as in Lemma 2.13, $\mathbb{E}[|X_n - X| \mid \mathcal{G}] \to 0$ a.s.

Proof. We have $|X_n - X| \to 0$ a.s., and $|X_n - X| \leq 2Z$. Apply Lemma 2.13. □

2.4 Properties of Conditional Expectation With No Unconditional Analogue

First, notice that unconditional expectation yields a *number*, while conditional expectation is a *random variable*. This permits us to consider more types of convergence. Here we present L^1 convergence of conditional expectation.

Lemma 2.15: L^1 Convergence for Conditional Expectation

Using the same conditions as Lemma 2.13, we have $\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}]$ in L^1.

Proof. We have

$$\begin{aligned}
\mathbb{E}\big[\big|\mathbb{E}[X_n \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}]\big|\big] &= \mathbb{E}\big[\big|\mathbb{E}[X_n - X \mid \mathcal{G}]\big|\big] \\
&\leq \mathbb{E}[\mathbb{E}[|X_n - X| \mid \mathcal{G}]] \\
&= \mathbb{E}[|X_n - X|] \to 0,
\end{aligned}$$

where the last step follows by Lemma 2.13. □

We now present two very important theorems that will have a lot of application later on.

Theorem 2.1: Tower Property

If $\mathcal{G}' \subseteq \mathcal{G} \subseteq \mathcal{F}$ are σ-algebras, then for any integrable X,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}'] = \mathbb{E}[X \mid \mathcal{G}'] \text{ a.e.}$$

Proof. Let $Y = \mathbb{E}[X \mid \mathcal{G}]$ and $Z = \mathbb{E}[X \mid \mathcal{G}']$. Take any $B \in \mathcal{G}'$, so that it is also an element of $\mathcal{G}$. Thus $\mathbb{E}[Z\delta_B] = \mathbb{E}[X\delta_B] = \mathbb{E}[Y\delta_B]$, and Z is $\mathcal{G}'$-measurable. Hence $Z = \mathbb{E}[Y \mid \mathcal{G}']$ a.e. □

The following theorem has conditions which are *crucial* for the validity of the theorem. Be sure to verify the conditions any time you use the theorem!

Theorem 2.2: Pulling Out Measurable Random Variables

Suppose that X, Y are random variables such that

1. *Y is $\mathcal{G}$-measurable,*
2. *X is integrable, and*
3. *XY is integrable.*

Then $\mathbb{E}[XY \mid \mathcal{G}] = Y\,\mathbb{E}[X \mid \mathcal{G}]$ a.e.

Proof. First, let us prove the identity

$$\mathbb{E}[XY] = \mathbb{E}[Y\,\mathbb{E}[X \mid \mathcal{G}]]. \tag{3}$$

To prove (3), let us first assume that $Y = \delta_B$ for some $B \in \mathcal{G}$. Then (3) follows from the definition of $\mathbb{E}[X \mid \mathcal{G}]$. From this, it follows that (3) holds when Y is simple by applying Lemma 2.6.

Next, suppose that X and Y are non-negative for which the hypotheses in the theorem hold. Let Y_n be a sequence of $\mathcal{G}$-measurable non-negative simple random variables increasing to Y. Then $XY_n \nearrow XY$, and

$$\mathbb{E}[X \mid \mathcal{G}]Y_n \nearrow \mathbb{E}[X \mid \mathcal{G}]Y$$

so $\mathbb{E}[XY_n] = \mathbb{E}[Y_n\,\mathbb{E}[X \mid \mathcal{G}]]$. Thus, by the ordinary monotone convergence theorem,

$$\begin{aligned}
\mathbb{E}[XY] &= \lim_{n\to\infty} \mathbb{E}[XY_n] \\
&= \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]Y_n] \\
&= \mathbb{E}[Y\,\mathbb{E}[X \mid \mathcal{G}]].
\end{aligned}$$

So (3) is true for non-negative random variables.

Finally, take any arbitrary X, Y satisfying the hypotheses in the theorem. Note that $|XY| = (X^+ + X^-)(Y^+ + Y^-) = X^+Y^+ + X^-Y^+ + X^+Y^- + X^-Y^-$. So, integrability of XY implies $\mathbb{E}[|XY|] < \infty$, and so each of the terms X^+Y^+, X^+Y^-, X^-Y^+, and X^-Y^- are integrable. Moreover, integrability of X implies that X^+ and X^- are integrable. Thus, we can apply the case for non-negative random variables. We have $\mathbb{E}[X^+Y^+] = \mathbb{E}[Y^+\,\mathbb{E}[X^+ \mid \mathcal{G}]]$, and so on. Putting everything together and applying Lemma 2.6 yields $\mathbb{E}[XY] = \mathbb{E}[Y\,\mathbb{E}[X \mid \mathcal{G}]]$. This completes the proof of (3).

Let us now prove the main result. Take any X, Y satisfying the hypotheses, and take any $B \in \mathcal{G}$. Then the pair $(X, Y\delta_B)$ also satisfy the hypotheses of the theorem. Hence, by (3),

$$\begin{aligned}\mathbb{E}[XY\delta_B] &= \mathbb{E}[Y\delta_B\,\mathbb{E}[X \mid \mathcal{G}]] \\ &= \mathbb{E}[Y\,\mathbb{E}[X \mid \mathcal{G}]\delta_B].\end{aligned}$$

Since this holds for all $B \in \mathcal{G}$, we have that $\mathbb{E}[XY \mid \mathcal{G}] = Y\,\mathbb{E}[X \mid \mathcal{G}]$ a.e. □

In the next class, we will prove a very important property of conditional expectation called Jensen's inequality, and we will then move on to martingales.

3 January 11

3.1 Jensen's Inequality

There is one more property of conditional expectation that we will need and that is very important: Jensen's inequality for conditional expectation. Before stating and proving Jensen's inequality, we need some lemmas.

Lemma 3.1: Supremum is Convex

The pointwise supremum of any collection of convex functions defined on an interval is convex.

Proof. Let $\mathcal{A}$ be a collection of convex functions on an interval I. Let $g(x) = \sup_{f \in \mathcal{A}} f(x)$. Take any $x, y \in I$ and $t \in [0,1]$. Then for any $f \in \mathcal{A}$,

$$\begin{aligned}f(tx + (1-t)y) &\leq tf(x) + (1-t)f(y) \\ &\leq tg(x) + (1-t)g(y).\end{aligned}$$

Taking the supremum over $f \in \mathcal{A}$ on the left yields

$$g(tx + (1-t)y) \leq tg(x) + (1-t)g(y).$$

□

A particularly important consequence of Lemma 3.1 is that the supremum over any collection of linear functions is convex. A converse of this statement is also true.

Lemma 3.2: Convex Functions are Suprema Over Countable Collections

Let I be an interval and $\phi : I \to \mathbb{R}$ be a convex function. Then there are sequences a_n and b_n such that

$$\phi(x) = \sup_n (a_n x + b_n)$$

for all $x \in I$.

Proof. For any $x \in I$, the graph of ϕ has a tangent line at $(x, \phi(x))$ that lies entirely below the graph (this is the supporting hyperplane theorem from convex analysis).[1] Using this fact, choose, for every $x \in I$, some $a_x, b_x \in \mathbb{R}$ such that $y \mapsto a_x y + b_x$ is a tangent line at $(x, \phi(x))$. That is, $a_x x + b_x = \phi(x)$

[1] To see this, we know by convexity of ϕ that $(\phi(x+t) - \phi(x))/t$ is a non-decreasing function of $t \in \mathbb{R}$. So $a = \lim_{t \searrow 0}(\phi(x+t) - \phi(x))/t$ and $b = \lim_{t \nearrow 0}(\phi(x+t) - \phi(x))/t$ both exist. Now any line through $(x, \phi(x))$ with slope in $[b, a]$ is a tangent line lying entirely below the graph of ϕ.

and $a_x y + b_x \le \phi(y)$ for every $y \in I$. Now define $g : I \to \mathbb{R}$ as $g(x) = \sup_{y \in \mathbb{Q} \cap I}(a_y x + b_y)$. Then by Lemma 3.1, we know g is a convex function. Moreover, we know $g \le \phi$ everywhere since the tangent lines all entirely below ϕ. For $x \in \mathbb{Q}$,

$$\begin{aligned} g(x) &= \sup_{y \in \mathbb{Q} \cap I}(a_y x + b_y) \\ &\ge a_x x + b_x \\ &= \phi(x). \end{aligned}$$

So, for all $x \in \mathbb{Q} \cap I$, we have $g(x) = \phi(x)$. But convex functions are continuous, so since g and ϕ are both convex, we conclude $g = \phi$. □

We are now ready to state and prove Jensen's inequality for conditional expectation.

Theorem 3.1: Jensen's Inequality for Conditional Expectation

Let X be an integrable random variable taking value in some interval I (finite or infinite). Let $\phi : I \to \mathbb{R}$ be a convex function such that $\phi(X)$ is integrable. Then for any sub-σ-algebra $\mathcal{G}$, the inequality

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

holds a.s.

Proof. Let ϕ and the sequences a_n and b_n be as in Lemma 3.2. Then for any n, we have

$$\begin{aligned} \mathbb{E}[\phi(X) \mid \mathcal{G}] &\ge \mathbb{E}[a_n X + b_n \mid \mathcal{G}] \\ &= a_n \mathbb{E}[X \mid \mathcal{G}] + b_n. \end{aligned}$$

Take the supremum over n on the right. This gives

$$\mathbb{E}[\phi(X) \mid \mathcal{G}] \ge \phi(\mathbb{E}[X \mid \mathcal{G}]) \text{ a.s.}$$

□

An important thing to note is that $\mathbb{E}[X \mid \mathcal{G}] \in I$ a.s. This is true because I is an interval; this is not true in general!

3.2 Martingales

Though martingales are applied in many areas of math, it is developed only in probability. We first define some preliminary vocabulary.

Definition 3.1: Filtrations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A monotone sequence

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$$

of sub-σ-algebras is called a ***filtration****.*

A sequence of random variables $\{X_n\}_{n \ge 0}$ is said to be **adapted** to a filtration if X_n is $\mathcal{F}_n$-measurable for every n. Often when we are given such a sequence of random variables, the filtration is clear from context, so we may omit the phrase.

Example 3.1: Filtration Generated by Sequence

Let $\{X_n\}_{n\geq 0}$ be an arbitrary sequence of random variables, and let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Then $\{X_n\}_{n\geq 0}$ is adapted to $\{\mathcal{F}_n\}_{n\geq 0}$.

We are now ready to define a martingale.

Definition 3.2: Martingales

*Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. We say that $\{X_n\}_{n\geq 0}$ is a **martingale** if*

1. $\mathbb{E}[|X_n|] < \infty$ *for each n, and*
2. $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ *a.s. for all n (this is the **martingale property**).*

We often think of n as time, so if $\{X_n\}_{n\geq 0}$ represents (random) measurements over time, then the martingale property essentially says that our best prediction of the measurement at time $n+1$ given all information up to time n is the measurement X_n itself.

Example 3.2: Simplest Non-trivial Martingale

Let $X_1, X_2, \ldots$ be independent random variables with mean 0.[a] Let $S_0 = 0$ and $S_n = \sum_{i=1}^n X_i$ for $n \geq 1$ (this defines a **simple random walk**). Let $\mathcal{F}_0$ be the trivial σ-algebra, and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for $n \geq 1$.

Claim: The sequence $\{S_n\}_{n\geq 0}$ is a martingale adapted to $\{\mathcal{F}_n\}_{n\geq 0}$.

Proof: For any $n \geq 1$, $|S_n| \leq |X_1| + \cdots + |X_n|$, so $\mathbb{E}[|S_n|] < \infty$. This shows the first condition of Definition 3.2. For the second condition, first compute

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[S_n + X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[S_n \mid \mathcal{F}_n] + \mathbb{E}[X_{n+1} \mid \mathcal{F}_n].$$

Now notice that, since S_n is $\mathcal{F}_n$-measurable, we have $\mathbb{E}[S_n \mid \mathcal{F}_n] = S_n$ a.s. Also, since X_{n+1} is independent of $\mathcal{F}_n$, we have $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1}] = 0$ a.s. This shows the second condition of Definition 3.2.

[a] If $X_1, X_2, \ldots$ are independent are integrable, but not necessarily having mean 0, then we can subtract off the mean from each X_i to obtain a martingale by the same argument: $S_n = \sum_{i=1}^n (X_i - \mu_i)$, where $\mu_i = \mathbb{E}[X_i]$, is a martingale.

Example 3.3: Extending the Simplest Non-trivial Martingale

Let $X_1, X_2, \ldots$ be independent square-integrable random variables. Let $Z_0 = 0$ and

$$Z_n = \left(\sum_{i=1}^n (X_i - \mu_i)\right)^2 - \sum_{i=1}^n \sigma_i^2,$$

where $\sigma_i^2 = \mathrm{Var}[X_i]$.

Claim: The sequence $\{Z_n\}_{n\geq 0}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}_{n\geq 0}$ as defined in Example 3.1.

Proof: For notational simplicity, let us assume $\mu_i = 0$ and $\sigma_i^2 = 1$ for all i (the general case is similar). Then $Z_n = S_n^2 - n$, and

$$\begin{aligned}\mathbb{E}[|Z_n|] &\le \mathbb{E}\big[S_n^2\big] + n \\ &= Zn.\end{aligned}$$

Clearly, Z_n is $\mathcal{F}_n$-measurable. Let us verify the martingale property. We have

$$\begin{aligned}\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}\big[S_{n+1}^2 \mid \mathcal{F}_n\big] - (n+1) \\ &= \mathbb{E}\big[(S_n + X_{n+1})^2 \mid \mathcal{F}_n\big] - (n+1) \\ &= \mathbb{E}\big[S_n^2 + 2X_{n+1}S_n + X_{n+1}^2 \mid \mathcal{F}_n\big] - (n+1).\end{aligned}$$

Since S_n^2 is $\mathcal{F}_n$-measurable, we know $\mathbb{E}\big[S_n^2 \mid \mathcal{F}_n\big] = S_n^2$ a.s. Next, since S_n is $\mathcal{F}_n$-measurable, X_{n+1} is integrable, and $X_{n+1}S_n$ is integrable[a], we have $\mathbb{E}[X_{n+1}S_n \mid \mathcal{F}_n] = S_n\,\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]$ a.s. by Theorem 2.2. But X_{n+1} is independent of $\mathcal{F}_n$, so $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1}] = 0$ a.s. Finally, again by independence, we have $\mathbb{E}\big[X_{n+1}^2 \mid \mathcal{F}_n\big] = \mathbb{E}\big[X_{n+1}^2\big] = 1$ a.s. Thus,

$$\mathbb{E}\big[S_n^2 + 2X_{n+1}S_n + X_{n+1}^2 \mid \mathcal{F}_n\big] = S_n^2 + 1$$

a.s. Thus, $\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = S_n^2 + 1 - (n+1) = S_n^2 - n = Z_n$.

[a] Notice $\mathbb{E}[|X_{n+1}S_n|] \le \sqrt{\mathbb{E}\big[X_{n+1}^2\big]\,\mathbb{E}[S_n^2]} < \infty$ by Cauchy-Schwarz

Note that Example 3.3 defines a sort of "weird" martingale, in that Z_n is defined by taking a sum, *then* squaring it, rather than taking a sum of squares. Doing the latter would reduce to Example 3.2.

Example 3.4: Martingale from Moment-Generating Function

Let $X_1, X_2, \ldots$ be i.i.d. random variables. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Suppose that $\mathbb{E}\big[e^{\theta X_i}\big] < \infty$ for some nonzero $\theta \in \mathbb{R}$. Let $m(\theta) = \mathbb{E}\big[e^{\theta X_i}\big]$, and let

$$M_n = \frac{e^{\theta S_n}}{m(\theta)^n},$$

where $S_n = X_1 + \cdots + X_n$. Then $\{M_n\}$ is a martingale.

Proof: Exercise.

3.3 Stopping Times and Stopped σ-Algebras

Suppose that we are invested in the stock market. Our portfolio consists of some stock whose price fluctuates up and down over time, and suppose we have a complete history of the prices of this stock. We want to make a decision whether to sell some amount of stock, and indeed, we will sell some our stock at time T, where T is a random variable. Intuitively, we call T a stopping time if, whether we sell at a particular time $T = n$ or not, the decision to do so is a function of all the prices of the stock that we've observed up until time n. Let us describe stopping times more precisely.

Definition 3.3: Stopping Times

Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration of σ-algebras, and let T be a random variable taking value in $\{0,1,2,\ldots\}\cup\{\infty\}$. Then T is called a ***stopping time*** *with respect to the filtration $\{\mathcal{F}_n\}_{n\geq 0}$ if, for all $n \in \{0,1,2,\ldots\}$, the event $\{T = n\} \in \mathcal{F}_n$.*

For instance, if $\{\mathcal{F}_n\}_{n\geq 0}$ is a filtration of the form presented in Example 3.1, then a stopping time is essentially a random variable such that, no matter where you are in the sample space, the decision to stop at time n depends only on historical observations up to time n.

Example 3.5: Stopping Times for Set Inclusion

Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Take any Borel set $A \subseteq \mathbb{R}$. Let $T = \inf\{n \geq 0 \mid X_n \in A\}$, with the infimum defined as ∞ if there does not exist n such that $X_n \in A$.[a] It's easy to check that T is measurable.

Claim: T is a stopping time with respect to $\{\mathcal{F}_n\}_{n\geq 0}$.

Proof: Take any $n \in \{0,1,2,\ldots\}$. Then the event $\{T = n\} = \{X_n \in A \mid X_j \notin A \text{ for } j < n\} \in \mathcal{F}_n$. This inclusion is true since $\mathcal{F}_n \supseteq \sigma(X_0,\ldots,X_n)$ (recall that filtrations are monotone sequences of σ-algebras).

[a] This is defined pointwise: for every $\omega \in \Omega$, we define $T(\omega) = \inf\{n \geq 0 \mid X_n(\omega) \in A\}$.

Intuitively, Example 3.5 gives the first time that the sequence $\{X_n\}_{n\geq 0}$ "hits" the set A.

Example 3.6: Stopping Times for Random Walk

Let $S_0, S_1, S_2, \ldots$ be a simple symmetric random walk (SSRW) on $\mathbb{Z}$ starting at 0, that is, $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ where the X_i are i.i.d. and $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = 1/2$. Take any integers $a < 0 < b$. Let $T = \inf\{n \mid S_n = a \text{ or } S_n = b\}$, which is (intuitively) the "first time" that S_n is a or b. Then by Example 3.5 (using $\{S_n\}_{n\geq 0}$ as the sequence adapted to $\{\mathcal{F}_n\}_{n\geq 0}$), we know T is a stopping time.

We now present the more confusing notion of stopped σ-algebras.

Definition 3.4: Stopped σ-Algebras

Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration and T be a stopping time adapted to this filtration. Then the ***stopped σ-algebra*** *$\mathcal{F}_T$ is defined as the set of all $A \in \mathcal{F}$ (recall $\mathcal{F}$ is the σ-algebra of the probability space) such that, for all $n \in \{0,1,2,\ldots\}$, the event $A \cap \{T = n\} \in \mathcal{F}_n$.*

Note that it's easy to check that $\mathcal{F}_T$ is, indeed, a σ-algebra.[2] Intuitively, events in $\mathcal{F}_T$ are events that can be described by events up to the stopping time.

[2] Clearly $\varnothing \in \mathcal{F}_T$. We have $A \in \mathcal{F}_T \implies A^c \in \mathcal{F}_T$ since $A^c \cap \{T = n\} = \{T = n\} \setminus (A \cap \{T = n\})$. Also, $A_1, A_2, \ldots \in \mathcal{F}_T$ implies their union lies in $\mathcal{F}_T$, which is easy to check.

4 January 13

4.1 Stopped σ-Algebras Continued

Example 4.1: Stopped σ-Algebra for Random Walk

Consider Example 3.6, where $\{S_n\}_{n\geq 0}$ is an SSRW on $\mathbb{Z}$ starting at 0 and $T = \inf\{n \mid S_n = a \text{ or } S_n = b\}$, where $a < 0 < b$. Let $A = \{S_n \geq 0 \text{ for all } n \leq T\}$.

Claim: $A \in \mathcal{F}_T$.

Proof: Note that $A \cap \{T = n\}$ can be written in a way without involving T, namely,

$$A \cap \{T = n\} = \{S_0 \geq 0, \ldots, S_n \geq 0\} \cap \{S_0 \in (a, b), \ldots, S_{n-1} \in (a, b), S_n \in \{a, b\}\},$$

which belongs to $\mathcal{F}_n$.

Lemma 4.1: Monotonicity of Stopped σ-Algebras

If S and T are stopping times with respect to $\{\mathcal{F}_n\}_{n\geq 0}$, and $S \leq T$ always, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.

Proof. Take any $A \in \mathcal{F}_S$ and any n. Then

$$\begin{aligned} A \cap \{T = n\} &= A \cap \{S \leq n\} \cap \{T = n\} \\ &= \left(\bigcup_{k=0}^{n} A \cap \{S = k\}\right) \cap \{T = n\}, \end{aligned}$$

which belongs to $\mathcal{F}_n$ since each $A \cap \{S = k\}$ belongs to $\mathcal{F}_k \subseteq \mathcal{F}_n$, and $\{T = n\}$ belongs to $\mathcal{F}_n$. $\square$

To modify Lemma 4.1 to include the case $S \leq T$ a.s., some technicalities would need to be addressed (e.g., changing S and T on a set of measure 0, considering things like completions of σ-algebras, etc.). We will not address these since it almost never appears in practice.

Lemma 4.2: Constants Are Stopping Times

A random variable T that is always equal to some fixed non-negative integer N is a stopping time with respect to any filtration $\{\mathcal{F}_n\}_{n\geq 0}$, and $\mathcal{F}_T = \mathcal{F}_N$.

Proof. Trivial. $\square$

Lemma 4.3: Conditioning Martingales

If $\{X_n\}_{n\geq 0}$ is a martingale with respect to $\{\mathcal{F}_n\}_{n\geq 0}$, then for all $m \leq n$, we have

$$\mathbb{E}[X_n \mid \mathcal{F}_m] = X_m \text{ a.s.}$$

Proof. By induction. We know the result for $n = 1$, and then we apply Theorem 2.1. For instance,

$$\begin{aligned} \mathbb{E}[X_n \mid \mathcal{F}_{n-2}] &= \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-2}] \\ &= \mathbb{E}[X_{n-1} \mid \mathcal{F}_{n-2}] \\ &= X_{n-2}. \end{aligned}$$

Consequently, $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for all n because $\mathbb{E}[X_n] = \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{F}_0]] = \mathbb{E}[X_0]$. $\square$

4.2 Optional Stopping Theorem

Before covering the very important result called the optional stopping theorem, we need a few new pieces of vocabulary. A stopping time is called **bounded** if there is an integer N such that $T \leq N$ always. We also have the notion of a stopped random variable.

Definition 4.1: Stopped Random Variable

*Given any sequence of random variables $\{X_n\}_{n\geq 0}$ and any finite, non-negative-integer-valued T, the **stopped random variable** X_T is defined as*

$$X_T(\omega) = X_{T(\omega)}(\omega).$$

We generally assume $\{X_n\}_{n\geq 0}$ is a martingale and T is a stopping time with $T < \infty$ always.

Example 4.2: Stopped Simple Symmetric Random Walk

If S_n is an SSRW with $S_0 = 0$, and $T = \inf\{n \mid S_n = a \text{ or } S_n = b\}$ with $a < 0 < b$, then the stopped random variable $S_T \in \{a, b\}$. We will show later that $T < \infty$ a.s.

We now state the optional stopping theorem.

Theorem 4.1: Optional Stopping Theorem

Let $\{X_n\}_{n\geq 0}$ be a martingale adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Let S and T be bounded stopping times with respect to this filtration such that $S \leq T$ always. Then X_S and X_T are integrable, and $\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S$ a.s. In particular, $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Proof. Let N be an integer such that $S \leq T \leq N$ always. Such an N exists since we have assumed S and T to be bounded stopping times. First, let us check that X_S is $\mathcal{F}_S$-measurable. Take any $0 \leq n \leq N$ and $A \in \mathcal{B}(\mathbb{R})$. Then $\{X_S \in A\} \cap \{S = n\} = \{X_n \in A\} \cap \{S = n\}$. We see $\{X_n \in A\}$ and $\{S = n\}$ are each in $\mathcal{F}_n$, so $\{X_S \in A\} \cap \{S = n\} \in \mathcal{F}_n$. Thus, $\{X_S \in A\} \in \mathcal{F}_S$ for all $A \in \mathcal{B}(\mathbb{R})$, which means X_S is $\mathcal{F}_S$-measurable.

Now we note that $|X_S|$ and $|X_T|$ are bounded by

$$|X_0| + |X_1| + \cdots + |X_N|.$$

Thus, X_S and X_T are integrable. Now take any $A \in \mathcal{F}_S$. We want to show that $\mathbb{E}[X_S \delta_A] = \mathbb{E}[X_T \delta_A]$. We now observe that

$$\begin{aligned}
\mathbb{E}[X_N \delta_A] &= \mathbb{E}\left[X_n \delta_A \sum_{n=0}^{N} \delta_{\{T=n\}}\right] \\
&= \sum_{n=0}^{N} \mathbb{E}\big[X_n \delta_{\{T=n\}} \delta_A\big].
\end{aligned}$$

Take any n and note that $A \cap \{T = n\} \in \mathcal{F}_n$ since $A \in \mathcal{F}_S \subseteq \mathcal{F}_T$. Then

$$\begin{aligned}
\mathbb{E}\big[X_N \delta_{\{T=n\}} \delta_A\big] &= \mathbb{E}\big[X_N \delta_{\{T=n\}\cap A}\big] \\
&= \mathbb{E}\big[\mathbb{E}[X_N \mid \mathcal{F}_n] \delta_{\{T=n\}\cap A}\big] \\
&= \mathbb{E}\big[X_n \delta_{\{T=n\}\cap A}\big].
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}[X_N\delta_A] &= \sum_{n=0}^{N}\mathbb{E}\big[X_n\delta_{\{T=n\}\delta_A}\big] \\
&= \mathbb{E}\left[\left(\sum_{n=0}^{N}X_n\delta_{\{T=n\}}\right)\delta_A\right]. \qquad (4)
\end{aligned}$$

Now, $\sum_{n=0}^{N} X_n(\omega)\delta_{\{T(\omega)=n\}} = X_{T(\omega)}(\omega) = X_T(\omega)$, so (4) becomes $\mathbb{E}[X_T\delta_A]$. But by exactly the same steps with S instead of T, we also have $\mathbb{E}[X_N\delta_A] = \mathbb{E}[X_S\delta_A]$. Thus, $\mathbb{E}[X_T\delta_A] = \mathbb{E}[X_S\delta_A]$ for all $A \in \mathcal{F}_S$. Hence, $\mathbb{E}[X_T \mid \mathcal{F}_S] = X_S$ a.s.

For the second claim, take $S \equiv 0$. Then $\mathcal{F}_S = \mathcal{F}_0$ and $X_S = X_0$. Thus, $\mathbb{E}[X_T \mid \mathcal{F}_0] = X_0$, which implies $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. □

We'll now discuss a "trick" which lets us work with bounded stopping times even if we have an unbounded stopping time. Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration and T be a stopping time with respect to this filtration. For any n, define the random variable $T \wedge n$ given by

$$(T \wedge n)(\omega) = \min\{T(\omega), n\}.$$

If $T \wedge n$ is a stopping time, then this is certainly a bounded stopping time. Indeed, the following lemma shows it is a stopping time.

Lemma 4.1: $T \wedge n$ is a Stopping Time

The random variable $T \wedge n$ is also a stopping time with respect to $\{\mathcal{F}_k\}_{k\geq 0}$.

Proof. We check the event $\{T \wedge n = m\}$ belongs to $\mathcal{F}_m$, which follows the computation

$$\{T \wedge n = m\} = \begin{cases} \{T = m\} & \text{if } m < n \\ \varnothing & \text{if } m > n \\ \{T \geq n\} & \text{if } m = n \end{cases}.$$

All of these events are in $\mathcal{F}_m$. □

4.3 Making Things Concrete

Let's return to the running example of an SSRW.

Example 4.3: Gambler's Ruin Problem

Let $\{S_n\}_{n\geq 0}$ be a SSRW starting at 0, and let $a < 0 < b$ be two integers. Let $T = \inf\{n \mid S_n = a \text{ or } S_n = b\}$.

Question: What is $\mathbb{P}[S_T = b]$?

We can think of this as modeling the following scenario: suppose a gambler starts with x dollars. At each turn, the gambler wins or loses one dollar with equal probability. The target is to reach some amount y dollars, where $y \geq x$. The question is thus, "What is the chance of reaching the target before going broke, i.e., reaching 0?" This is the same problem as above with $a = -x$ and $b = y - x$.

Answer: Note that S_T can only be a or b. Thus,

$$\begin{aligned}\mathbb{E}[S_T] &= a\mathbb{P}[S_T = a] + b\mathbb{P}[S_T = b] \\ &= a(1 - \mathbb{P}[S_T = b]) + b\mathbb{P}[S_T = b] \\ &= a + (b-a)\mathbb{P}[S_T = b]. \end{aligned} \tag{5}$$

Thus, it suffices to evaluate $\mathbb{E}[S_T]$. It will turn out that $\mathbb{E}[S_T] = \mathbb{E}[S_0] = 0$, but we need to do some work to apply Theorem 4.1 since T is not a bounded stopping time. We approach this by applying Lemma 4.1 and taking $n \to \infty$.

Take any n. Then $T \wedge n$ is a bounded stopping time, so by Theorem 4.1, we know $\mathbb{E}[S_{T\wedge n}] = \mathbb{E}[S_0] = 0$. We wish now to show that $\mathbb{E}[S_{T\wedge n}] \to \mathbb{E}[S_T]$ as $n \to \infty$. First, we claim that $T < \infty$ a.s.

Divide up the non-negative integers into disjoint blocks of length $|a| + b$. If the walk takes all positive steps in a block $\{k(|a|+b), \ldots, (k+1)(|a|+b) - 1\}$, then $T \le (k+1)(|a|+b) - 1$ because either the walk is already outside (a, b) before this block, or if not, then it gets outside (a, b) by the end of this block. But it's easy to prove that such a block exists with probability 1. Thus, $\mathbb{P}[T < \infty] = 1$.

Now if $\omega \in \Omega$ is such that $T(\omega) < \infty$, then $S_{T\wedge n}(\omega) = S_{T(\omega)\wedge n}(\omega)$, which tends to $S_{T(\omega)}(\omega)$ as $n \to \infty$. This is precisely $S_T(\omega)$. Thus, $S_{T\wedge n} \to S_T$ a.s. as $n \to \infty$. So we need only show that we can bring the limit inside of expectation.

Note that up to time T, we always have $S_n \in [a, b]$. Also, $T \wedge n \le T$. Thus, $|S_{T\wedge n}| \le |a| + b$. The dominated convergence theorem then gives us $\mathbb{E}[S_{T\wedge n}] \to \mathbb{E}[S_T]$ as $n \to \infty$. Hence, $\mathbb{E}[S_T] = 0$, so (5) implies $\mathbb{P}[S_T = b] = -a/(b-a)$. In the gambler's ruin problem, we have $\mathbb{P}[\text{reaching target}] = x/y$.

Question: What is $\mathbb{E}[T]$?

Answer: Use the martingale $S_n^2 - n$ (see Example 3.3). By Theorem 4.1, we know that for all n,

$$\mathbb{E}\big[S_{T\wedge n}^2 - (T \wedge n)\big] = \mathbb{E}\big[S_{T\wedge 0}^2 - (T \wedge 0)\big] = 0.$$

This shows that $\mathbb{E}\big[S_{T\wedge n}^2\big] = \mathbb{E}[T \wedge n]$. From before, we know $T < \infty$ a.s., so $T \wedge n \to T$ a.s. as $n \to \infty$ and $S_{T\wedge n} \to S_T$ a.s. as $n \to \infty$. Again, we have $|S_{T\wedge n}^2| \le (|a| + b)^2$. So, by the dominated convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}\big[S_{T\wedge n}^2\big] = \mathbb{E}\big[S_T^2\big].$$

Also, $0 \le T \wedge n \nearrow T$, so by the monotone convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}[T \wedge n] = \mathbb{E}[T].$$

Hence,

$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}\left[S_T^2\right] \\ &= a^2\mathbb{P}[S_T = a] + b^2\mathbb{P}[S_T = b] \\ &= a^2(1 - \mathbb{P}[S_T = b]) + b^2\mathbb{P}[S_T = b] \\ &= a^2\left(1 - \frac{-a}{b-a}\right) + b^2\frac{-a}{b-a} \\ &= \frac{a^2 b}{b-a} - \frac{b^2 a}{b-a} \\ &= -ab.\end{aligned}$$

To summarize, we have $\mathbb{E}[T] = -ab$. In gambler's ruin, we thus have $\mathbb{E}[T] = x(y-x)$.

5 January 18

5.1 Submartinagles and Supermartingales

The notions of submartingales and supermartinagles are inspired by the notions of subharmonic and superharmonic functions (recall from analysis that subharmonic functions satisfy $\Delta u \geq 0$, and superharmonic functions satisfy $\Delta u \leq 0$). Let $\{\mathcal{F}_n\}_{n\geq 0}$ be a filtration of σ-algebras. Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables adapted to this filtration.

Definition 5.1: Submartingales and Supermartingales

We say that $\{X_n\}_{n\geq 0}$ is a ***submartingale*** *if each X_n is integrable and, for all n,*

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n \text{ a.s.}$$

Similarly, $\{X_n\}_{n\geq 0}$ is a ***supermartingale*** *if each X_n is integrable and, for all n,*

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n \text{ a.s.}$$

These are more general than martingales, since notice that a martingale is both a submartingale and a supermartinagle. We now remark that it is easy to produce a submartingale from a martingale by applying a convex function.

Lemma 5.1: Convex Functions of Martingales Are Submartingales

Let ϕ be a convex function such that $\phi(X_n)$ is integrable for all n, where $\{X_n\}_{n\geq 0}$ is a martingale. Then $\{\phi(X_n)\}_{n\geq 0}$ is a submartingale.

Proof. The proof follows readily by Theorem 3.1:

$$\mathbb{E}[\phi(X_{n+1}) \mid \mathcal{F}_n] \geq \phi(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]) = \phi(X_n).$$

∎

Lemma 5.2: Concave Functions of Martingales Are Supermartingales

Let ϕ be a concave function such that $\phi(X_n)$ is integrable for all n, where $\{X_n\}_{n\geq 0}$ is a martingale. Then $\{\phi(X_n)\}_{n\geq 0}$ is a supermartingale.

Some examples of submartingales include X_n^2, $|X_n|$, and $e^{\theta X_n}$, where X_n is a martingale, provided that they are integrable.

Lemma 5.3: Convex Non-Decreasing Functions Preserve Submartingale

If $\{X_n\}_{n\geq 0}$ is a submartingale, and ϕ is a non-decreasing convex function such that $\phi(X_n)$ is integrable for all n, then $\{\phi(X_n)\}$ is a submartingale.

For instance, if X_n is a submartingale, then X_n^+ is a submartingale, and $e^{\theta X_n}$ is a submartingale for $\theta \geq 0$, provided it is integrable. However, X_n^2 and $|X_n|$ may not be submartingales.

We next describe a simple decomposition of submartingales. For this, we define a new term: a process $\{X_n\}_{n\geq 0}$ adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$ is called **predictable** if X_n is $\mathcal{F}_{n-1}$-measurable for $n \geq 1$.

Theorem 5.1: Doob Decomposition

Suppose $\{X_n\}_{n\geq 0}$ is a submartingale adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Then X_n may be decomposed as a sum of a martingale and an in increasing predictable process.

Proof. Define

$$M_n = \sum_{k=0}^{n-1}(X_{k+1} - X_k - \mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_k]) \quad \text{and} \quad A_n = \sum_{k=0}^{n-1}\mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_k].$$

Then $X_n = X_0 + M_n + A_n$. By the submartingale property,

$$\mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_k] = \mathbb{E}[X_{k+1} \mid \mathcal{F}_k] - X_k \geq 0,$$

so $\{A_n\}_{n\geq 0}$ is an increasing predictable process. Now note that

$$\mathbb{E}[(X_{k+1} - X_k) - \mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_k] \mid \mathcal{F}_k] = \mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_k] - \mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_k] = 0.$$

Hence,

$$\begin{aligned}\mathbb{E}[M_n \mid \mathcal{F}_{n-1}] &= \sum_{k=0}^{n-2}(X_{k+1} - X_k - \mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_k]) + \mathbb{E}[X_n - X_{n-1} - \mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-1}] \\ &= M_{n-1}.\end{aligned}$$

Thus, $\{M_n\}_{n\geq 0}$ is a martingale. $\square$

Note that if $\{X_n\}_{n\geq 0}$ is not a submartingale, then we can still decompose it into a martingale and a predictable process. However, the remainder part $\{A_n\}_{n\geq 0}$ will not necessarily be an increasing process.

We also have an optional stopping time theorem for submartingales and supermartingales. The proof is almost identical to Theorem 4.1, except with inequalities instead of equalities.

Theorem 5.2: Optional Stopping Theorem for Sub- and Supermartingales

Let $\{X_n\}_{n\geq 0}$ be a process adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Let S and T be bounded stopping times with respect to this filtration such that $S \leq T$ always. If $\{X_n\}$ is a submartingale (supermartingale), then X_S and X_T are integrable, and $\mathbb{E}[X_T \mid \mathcal{F}_S] \geq X_S$ ($\mathbb{E}[X_T \mid \mathcal{F}_S] \leq X_S$) a.s. In particular, $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$ ($\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$).

Proposition 5.1: Stopped Process Submartingale

Let $\{X_n\}_{n\geq 0}$ be any sequence of integrable random variables adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Let T be a stopping time for this filtration. Suppose that for all n, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$ a.s. on the set $\{T > n\}$, i.e.,

$$\mathbb{P}[\{\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] < X_n\} \cap \{T > n\}] = 0.$$

In words, X_n is a submartingale up to time T. Then $\{X_{T\wedge n}\}_{n\geq 0}$ is a submartingale adapted to $\{\mathcal{F}_n\}_{n\geq 0}$. The same holds for supermartingales and martingales, mutatis mutandis.

Proof. First, $X_{T\wedge n}$ is $\mathcal{F}_n$-measurable by Lemma 4.1, and it is integrable since $T \wedge n \leq n$, and the X_k are integrable. So we can compute

$$\begin{aligned}\mathbb{E}\big[X_{T\wedge(n+1)} \mid \mathcal{F}_n\big] &= \sum_{i=0}^{n} \mathbb{E}\big[X_{T\wedge(n+1)}\delta_{\{T=i\}} \mid \mathcal{F}_n\big] + \mathbb{E}\big[X_{T\wedge(n+1)}\delta_{\{T>n\}} \mid \mathcal{F}_n\big] \\ &= \sum_{i=0}^{n} \mathbb{E}\big[X_i\delta_{\{T=i\}} \mid \mathcal{F}_n\big] + \mathbb{E}\big[X_{n+1}\delta_{\{T>n\}} \mid \mathcal{F}_n\big].\end{aligned}$$

Each of the terms in the sum is $\mathcal{F}_i$-measurable, and hence $\mathcal{F}_n$-measurable. Hence, it becomes

$$\sum_{i=0}^{n} X_i\delta_{\{T=i\}} + \mathbb{E}\big[X_{n+1}\delta_{\{T>n\}} \mid \mathcal{F}_n\big].$$

Now, $\{T > n\} = \{T \leq n\}^c = (\{T = 0\} \cup \cdots \cup \{T = n\})^c \in \mathcal{F}_n$, so we can pull the indicator out of the second term and use the given conditions (namely, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$ a.s. on the set $\{T > n\}$) to get

$$\sum_{i=0}^{n} X_i\delta_{\{T=i\}} + \delta_{\{T>n\}}\,\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq \delta_{\{T>n\}}X_n \text{ a.s.}$$

So we get

$$\mathbb{E}\big[X_{T\wedge(n+1)} \mid \mathcal{F}_n\big] \geq \sum_{i=0}^{n} X_i\delta_{\{T=n\}} + \delta_{\{T>n\}}X_n = X_{T\wedge n},$$

which shows that $\{X_{T\wedge n}\}_{n\geq 0}$ is a submartingale. □

We can generally apply Proposition 5.1 when processes "behave nicely" up to some time, so we can extract the martingale up to that time and work with that and take a limit. We see an example application below.

Example 5.1: Gambling Problem

Suppose we start with $x > a$ dollars, and at each turn, we win or lose 1 dollar with equal probability, as in Example 4.3. However, suppose that if we have a total exceeding a dollars, then we have to pay b dollars in tax for each turn.

Question: Let $T = \min\{n \mid X_n \leq a\}$, i.e., the first turn in which our total drops below a dollars. Can we give an upper bound on $\mathbb{E}[T]$?

Answer: The event $\{T > n\} = \{X_1 > a,\, X_2 > a,\, \ldots,\, X_n > a\}$. So $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n - b$ on the set $\{T > n\}$.[a] Define $Y_n = X_n + nb$. Then

$$\begin{aligned}\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] + (n+1)b \\ &= X_n - b + (n+1)b && \text{on } \{T > n\} \\ &= X_n + nb && \text{on } \{T > n\} \\ &= Y_n && \text{on } \{T > n\},\end{aligned}$$

so by Proposition 5.1, $\{Y_{T\wedge n}\}_{n\geq 0}$ is a martingale. By Theorem 4.1, we have $\mathbb{E}[Y_{T\wedge n}] = \mathbb{E}[Y_0] = x$. But $\mathbb{E}[Y_{T\wedge n}] = \mathbb{E}[X_{T\wedge n}] + b\,\mathbb{E}[T \wedge n]$.[b] This implies $b\,\mathbb{E}[T \wedge n] \leq x - a$ for all n. The monotone convergence theorem now implies $\mathbb{E}[T] \leq (x-a)/b$.

[a] Intuitively, this is because $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, so $\mathbb{E}[X_{n-1} \mid X_1 = x_1, \ldots, X_n = x_n] = x_n - b$ if $\{X_1 = x_1, \ldots, X_n = x_n\} \subseteq \{T > n\}$.

[b] If $T > n$, then $X_{T\wedge n} = X_n > a$. If $T \leq n$, then $X_{T\wedge n} = X_T \geq a$ since $T = \min\{n \mid X_n \leq a\} = \min\{n \mid X_n = a\}$.

6 January 20

6.1 Optimal Stopping Times

Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence $X_1, \ldots, X_N$ of integrable random variables, and a filtration $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_N$ of sub-σ-algebras of $\mathcal{F}$. Here, we do *not* assume that X_n is $\mathcal{F}_n$-measurable. Given this setup, we have what is known as a finite horizon problem.

Definition 6.1: Finite Horizon Optimal Stopping Problem

A ***finite horizon optimal stopping problem*** *asks for the stopping time for the filtration* $\{\mathcal{F}_n\}_{1\leq n\leq N}$ *which maximizes* $\mathbb{E}[X_T]$.

The idea of a finite horizon problem is to think of the filtration $\mathcal{F}_n$ as providing information about our problem up to time n, and X_n is the reward obtained if we execute some action at time n. We want to maximize our expected reward when our decisions can be made only on prior information. Incredibly, finite horizon optimal stopping problems can be solved in this completely general setup.

The first thing to do is to compute our expected reward if we execute the action at time n, given all information up to then. If we are able to compute these, then presumably we can pick the optimal stopping time. This is made concrete by the following lemma.

Lemma 6.1: Optimal Stopping Given Prior Knowledge is Optimal

Let $Y_n = \mathbb{E}[X_n \mid \mathcal{F}_n]$.[a] *For any stopping time* T *for the filtration* $\{\mathcal{F}_n\}_{1\leq n\leq N}$ *(taking value in* $\{1, \ldots, N\}$*), we have* $\mathbb{E}[X_T] = \mathbb{E}[Y_T]$.

[a] This is not just X_n since we do not assume $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$.

Proof. Note that

$$\mathbb{E}[Y_T] = \sum_{i=1}^{N} \mathbb{E}\big[Y_n \delta_{\{T=n\}}\big] = \sum_{n=1}^{N} \mathbb{E}\big[X_n \delta_{\{T=n\}}\big]$$

because $\{T = n\} \in \mathcal{F}_n$ and $Y_n = \mathbb{E}[X_n \mid \mathcal{F}_n]$. The last expression is $\mathbb{E}[X_T]$. $\square$

We next "propagate back" the conditional expected rewards using what is known as the **Snell envelope** of $Y_1, \ldots, Y_N$. We define $V_N = Y_N$ and, by backward induction,

$$V_n = \max\{Y_n, \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]\}$$

for $n = N-1, N-2, \ldots, 1$. Since each V_n is integrable and $\mathcal{F}_n$-measurable, this definition is well defined and gives that $\{V_n\}_{1\leq n\leq N}$ is a supermartingale adapted to $\{\mathcal{F}_n\}_{1\leq n\leq N}$. Indeed, $V_n \geq \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]$ a.s., and it is the "smallest" supermartingale that dominates Y_n.

Finally, we define $\tau = \min\{n \mid Y_n = V_n\}$, which will be our optimal stopping time. Notice that τ will exist, since we know that $Y_N = V_N$. It is also a stopping time for $\{\mathcal{F}_n\}_{1\leq n\leq N}$, since

$$\{\tau = n\} = \{V_n = Y_n\} \cap \{V_k \neq Y_k \text{ for all } k < n\},$$

which certainly belongs to $\mathcal{F}_n$. The intuition behind τ is that it is the first time for which our best expected reward given information up to time n is obtained when we execute the decision *now* (i.e., at time n) rather than waiting another time step. Since we are not guaranteed another opportunity for this to occur, we should execute the decision. Let us now formally prove that τ gives the optimal stopping time.

Theorem 6.1: Optimal Stopping Theorem

The stopping time τ maximizes $\mathbb{E}[X_\tau]$ among all stopping times adapted to the filtration $\{\mathcal{F}_n\}_{1\leq n\leq N}$. Moreover, this maximum value is equal to $\mathbb{E}[V_1]$.

Proof. Let T be any other stopping time. By Lemma 6.1, we have $\mathbb{E}[X_\tau] = \mathbb{E}[Y_\tau]$ and $\mathbb{E}[X_T] = \mathbb{E}[Y_T]$, so it suffices to check that $\mathbb{E}[Y_\tau] \geq \mathbb{E}[Y_T]$. If $\tau(\omega) > n$ for some ω, then $V_n > Y_n$, which implies that $V_n(\omega) = \mathbb{E}[V_{n+1} \mid \mathcal{F}_n](\omega)$. Thus, $V_n = \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]$ a.s. on the set $\{\tau > n\}$. We thus have by Proposition 5.1 that $\{V_{\tau\wedge n}\}_{1\leq n\leq N}$ is a martingale adapted to $\{\mathcal{F}_n\}_{1\leq n\leq N}$. Moreover, $Y_\tau = V_\tau$ by the definition of τ. Hence,

$$\mathbb{E}[Y_\tau] = \mathbb{E}[V_\tau] = \mathbb{E}[V_{\tau\wedge N}] = \mathbb{E}[V_{\tau\wedge 1}] = \mathbb{E}[V_1].$$

But, since $\{V_n\}_{1\leq n\leq N}$ is a supermartingale, T is a bounded stopping time, and $V_n \geq Y_n$ for all n, Theorem 5.2 implies that

$$\mathbb{E}[V_1] \geq \mathbb{E}[V_T] \geq \mathbb{E}[Y_T],$$

as desired. $\square$

Let us now look at a particularly insightful example that formalizes a standard example from an undergraduate introduction to probability.

Example 6.1: Interviews

Suppose N candidates appear for an interview in a random order. They are ranked $r_1, \dots, r_N$, which is a uniformly random permutation of $1, \dots, N$. Upon interviewing a candidate, the hiring manager must either hire the candidate immediately or move on to the next candidate with no possibility of hiring them later. At each time step, the hiring manager knows only the relative ranking of the candidates they have interviewed.

Question: What is the hiring manager's optimal strategy in order to maximize chance of choosing the best candidate?

Answer: Up to time n, the hiring manager knows the relative ranks of the first n candidates. Let's call them $r_1^n, \dots, r_n^n$. For instance, if $N = 5$ and $(r_1, \dots, r_N) = (2, 4, 1, 3, 5)$, then $(r_1^3, r_2^3, r_3^3) = (2, 3, 1)$ (*not* $(2, 4, 1)$ since they are *relative* ranks). So, the proper σ-algebras to consider are $\mathcal{F}_n = \sigma(r_1^n, \dots, r_n^n)$. It is easy to see that $\mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_N$ since $(r_1^n, \dots, r_n^n)$ is a measurable function of $(r_1^{n+1}, \dots, r_{n+1}^{n+1})$. Mathematically, the goal is now to find the stopping time τ that maximizes $\mathbb{P}[r_\tau = 1]$.[a]

We can put this in our framework for optimal stopping times by taking $X_n = \delta_{\{r_n = 1\}}$. Then $\mathbb{E}[X_T] = \mathbb{P}[r_T = 1]$ for all stopping times T. Now, compute $Y_n = \mathbb{E}[X_n \mid \mathcal{F}_n] = \mathbb{P}[r_n = 1 \mid \mathcal{F}_n]$. We claim that

$$Y_n = \begin{cases} 0 & \text{if } r_n^n \neq 1 \\ n/N & \text{if } r_n^n = 1 \end{cases}$$

The proof of the claim is as follows. The value for Y_n when $r_n^n \neq 1$ is clear, so suppose that we are given $r_n = 1$. The ranks of the other candidates form a uniformly random permutation of $2, \dots, N$. So, for any permutation $\sigma_1, \dots, \sigma_{n-1}$ of $2, \dots, N$, we have

$$\mathbb{P}\big[r_1^n = \sigma_1, \dots, r_{n-1}^n = \sigma_{n-1}, r_n^n = 1 \mid r_n = 1\big] = \frac{1}{(n-1)!}.$$

By Bayes's rule,

$$\begin{aligned} &\mathbb{P}\big[r_n = 1 \mid r_1^n = \sigma_1, \dots, r_{n-1}^n = \sigma_{n-1}, r_n^n = 1\big] \\ &= \frac{\mathbb{P}\big[r_1^n = \sigma_1, \dots, r_{n-1}^n = \sigma_{n-1}, r_n^n = 1 \mid r_n = 1\big]\mathbb{P}[r_n = 1]}{\mathbb{P}\big[r_1^n = \sigma_1, \dots, r_{n-1}^n = \sigma_{n-1}, r_n^n = 1\big]} \\ &= \frac{1}{(n-1)!}\frac{1/N}{1/n!} = \frac{n}{N}. \end{aligned}$$

In order to use this fact that $Y_n = n/N\delta_{\{r_n^n = 1\}}$, we will also check that Y_n is independent of $\mathcal{F}_{n-1}$ for $n \geq 2$. To see this, observe that Y_n can take only wo values, namely, 0 and n/N. So it suffices to check that $\mathbb{P}[\{Y_n = n/N\} \cap A] = \mathbb{P}[Y_n = n/N]\mathbb{P}[A]$ for all $A \in \mathcal{F}_{n-1}$. Notice that the event $\{Y_n = n/N\} = \{r_n^n = 1\}$. Now, any A in $\mathcal{F}_{n-1}$ is a disjoint union of events like $\{r_1^{n-1} = \sigma_1, \dots, r_{n-1}^{n-1} = \sigma_{n-1}\}$ for permutates $(\sigma_1, \dots, \sigma_{n-1})$ of $(1, \dots, n-1)$. So it suffices to prove independence for A of this form. We have

$$\begin{aligned} &\mathbb{P}\big[r_n^n = 1, r_1^{n-1} = \sigma_1, \dots, r_{n-1}^{n-1} = \sigma_{n-1}\big] \\ &= \mathbb{P}\big[r_n^n = 1, r_1^n = \sigma_1 + 1, \dots, r_{n-1}^n = \sigma_{n-1} + 1\big] \\ &= \frac{1}{n!} = \frac{1}{n}\frac{1}{(n-1)!} \\ &= \mathbb{P}[r_n^n = 1]\mathbb{P}\big[r_1^{n-1} = \sigma_1, \dots, r_{n-1}^{n-1} = \sigma_{n-1}\big]. \end{aligned}$$

This tells us that whatever we know about the first $n-1$ candidates, we know nothing about the relative rank of the nth candidate.

Define now $V_N = Y_N$ and $V_n = \max\{Y_n, \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]\}$ by backwards recursion. As a corollary of the previous result, for all $n \le N-1$, the conditional expectation $\mathbb{E}[V_{n+1} \mid \mathcal{F}_n]$ is a nonrandom number, which we will denote by v_n^N. To see this, first notice $\mathbb{E}[V_N \mid \mathcal{F}_{N-1}] = \mathbb{E}[Y_N \mid \mathcal{F}_{N-1}]$ is a constant, since Y_N is independent of $\mathcal{F}_{N-1}$. Proceeding by backwards induction, suppose that $\mathbb{E}[V_{n+1} \mid \mathcal{F}_n] = v_n^N$ is nonrandom. Then

$$\begin{aligned}
\mathbb{E}[V_n \mid \mathcal{F}_{n-1}] &= \mathbb{E}[\max\{Y_n, \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]\} \mid \mathcal{F}_{n-1}] \\
&= \mathbb{E}\big[\max\{Y_n, v_n^N\} \mid \mathcal{F}_{n-1}\big] \\
&= \text{constant, since } Y_n \text{ independent of } \mathcal{F}_{n-1} \\
&= \mathbb{E}\big[\max\{Y_n, v_n^N\}\big].
\end{aligned}$$

The last step follows since $v_n^{N-1} = \mathbb{E}\big[\max\{Y_n, v_n^N\}\big]$. We now define the optimal stopping rule $\tau = \min\{n \mid V_n = Y_n\} = \min\{n \mid Y_n \ge \mathbb{E}[V_{n+1} \mid \mathcal{F}_n]\} = \min\{n \mid Y_n \ge v_n^N\}$ (we define $v_N^N = 0$). Recall that $Y_n = n/N \delta_{\{r_n^n = 1\}}$. Let $t_N = \min\{n \mid n/N \ge v_n^N\}$. So $\tau = \min\{n \ge t_N \mid r_n^n = 1\}$. Thus, the best thing to do is to wait until some time t_N, at which point if we have a candidate with $r_n^n = 1$, we will hire that candidate. Let us now find t_N.

We recall $Y_n = n/N$ with probability $1/n$, and 0 with probability $1 - 1/n$. If $n < t_N$, then $n/N < v_n^N$, and so $v_{n-1}^N = v_n^N$. If $n \ge t_N$, then $n/N \ge v_n^N$, so

$$\begin{aligned}
v_{n-1}^N &= \frac{n}{N}\mathbb{P}\Big[Y_n = \frac{n}{N}\Big] + v_n^N \mathbb{P}[Y_n = 0] \\
&= \frac{n}{N}\frac{1}{n} + \left(1 - \frac{1}{n}\right)v_n^N \\
&= \frac{1}{N} + \left(1 - \frac{1}{n}\right)v_n^N \\
&= \frac{1}{N} + \left(1 - \frac{1}{n}\right)\frac{1}{N} + \left(1 - \frac{1}{n+1}\right)v_{n+1}^N \\
&= \frac{1}{N} + \frac{1}{N}\frac{n-1}{n} + \frac{n-1}{n+1}v_{n+1}^N \\
&\vdots \\
&= \frac{1}{N}\left(1 + n\sum_{k=n+1}^{N-1}\frac{1}{k}\right).
\end{aligned}$$

Thus, t_N is characterized as the unique n such that

$$\frac{1}{N}\left(1 + (n-1)\sum_{k=n}^{N-1}\frac{1}{k}\right) > \frac{n}{N} \ge \frac{1}{N}\left(1 + n\sum_{k=n+1}^{N-1}\frac{1}{k}\right).$$

For large N, the upper and lower bounds are approximately $(n/N)\log(N/n)$. So, when N large, this shows that $t_N/N \approx (t_N/N)\log(N/t_N)$, which implies $t_N \approx N/e$.

Question: What is the probability of choosing the best candidate with the optimal rule?

Answer: From the proof of Theorem 6.1, we know that

$$\begin{aligned}\mathbb{E}[X_\tau] &= \mathbb{E}[Y_\tau] = \mathbb{E}[V_\tau] = \mathbb{E}[V_{\tau\wedge N}] \\ &= \mathbb{E}[V_{\tau\wedge 1}] = \mathbb{E}[V_1] = \mathbb{E}\big[\max\{Y_1, v_1^N\}\big] = v_1^N,\end{aligned}$$

where the last step holds since $Y_1 \leq v_1^N$, because $t_N \geq 1$. Since $v_1^N = v_2^N = \cdots = v_{t_N-1}^N$ and $v_{t_N-1}^N = N^{-1}\Big(1 + t_N \sum_{k=t_N}^{N-1} k^{-1}\Big)$, we have that for large N, we have that

$$v_{t_N-1}^N \approx \frac{t_N}{N}\log\left(\frac{N}{t_N}\right) \approx \frac{1}{e}.$$

[a] This is not the same as minimizing the rank of the hired candidate!

7 January 25

7.1 Convergence of Martingales

The main goal of this section is to prove the following result and its analogues. The proof involves a very new idea (by Doob), so we will build up the theory before proving it.

Theorem 7.1: Submartingale Convergence Theorem

Let $\{X_n\}_{n\geq 0}$ be a submartingale adapted to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Suppose that $\sup_{n\geq 0}\mathbb{E}[X_n^+] < \infty$. Then there exists an integrable, a.s. finite random variable X such that $X_n \to X$ a.s.

Let us now develop some of the theory.

Definition 7.1: Upcrossing

*Fix a sample point $\omega \in \Omega$, and let $[a,b]$ be a bounded interval. **Upcrossings** of $[a,b]$ by $\{X_n\}_{n\geq 0}$ are sequences of indices, defined inductively. The first upcrossing starts at the first k such that $X_k \leq a$ and stops at the first $j > k$ where $X_j \geq b$. The next upcrossing starts at the first $k' > j$ where $X_{k'} \leq a$ and stops at the first $j' > k$ where $X_{j'} \geq b$, and so on.*

We now denote by U_m the number of *complete* upcrossings of $[a,b]$ by time m. Notice that this is a random variable, since the number of upcrossings depend on the sample point.

Lemma 7.1: Doob's Upcrossing Lemma

We have

$$\mathbb{E}[U_m] \leq \frac{\mathbb{E}[(X_m - a)^+] - \mathbb{E}[(X_0 - a)^+]}{b-a}.$$

Proof. First, define a new process $Y_n = a + (X_n - a)^+ = \max\{a, X_n\}$. Notice that this is a convex, increasing function of X_n. Hence, $\{Y_n\}_{n\geq 0}$ is also a submartingale adapted to $\{\mathcal{F}_n\}_{n\geq 0}$. Moreover, $\{Y_n\}_{n\geq 0}$ has exactly the same upcrossings of $[a,b]$ as $\{X_n\}_{n\geq 0}$. The new process has the extra feature that, whenever $k, k+1, \ldots, k+\ell$ is an upcrossing, we know

$$Y_j - Y_k \geq 0 \quad \text{for all } j \in \{k, k+1, \ldots, k+\ell\}.$$

Moreover, $Y_{n+\ell} - Y_k \geq b - a$ (this is true also for the original process). Thus, if we sum up the increments $Y_{j+1} - Y_j$ only for those j such that $j, j+1$ are both in the same upcrossing, then sum up to time m is $\geq U_m(b-a)$ because the last (possibly incomplete) upcrossing contributes ≥ 0. For the original process, the last upcrossing may contribute < 0 if it's incomplete. But this sum is also $\sum_{n=1}^{m} Z_n(Y_n - Y_{n-1})$, where $Z_n = \delta_{\{n-1,n \text{ in same upcrossing}\}}$.[3] This event is determined by $Y_0, \ldots, Y_{n-1}$. That is to say, Z_n is $\mathcal{F}_{n-1}$-measurable. Thus,

$$\begin{aligned}
(b-a)\,\mathbb{E}[U_m] &\leq \sum_{n=1}^{m} \mathbb{E}[Z_n(Y_n - Y_{n-1})] \\
&= \sum_{n=1}^{m} \mathbb{E}[\mathbb{E}[Z_n(Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}]] \\
&= \sum_{n=1}^{m} \mathbb{E}[Z_n\,\mathbb{E}[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}]].
\end{aligned}$$

Since $\{Y_n\}_{n\geq 0}$ is a submartingale, $\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}] \geq Y_{n-1}$, which implies $\mathbb{E}[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}] \geq 0$. Hence, the sum can be bounded above by

$$\begin{aligned}
\sum_{n=1}^{m} \mathbb{E}[\mathbb{E}[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}]] &= \sum_{n=1}^{m} \mathbb{E}[Y_n - Y_{n-1}] \\
&= \mathbb{E}[Y_m - Y_0] \\
&= \mathbb{E}\big[a + (X_m - a)^+ - (a + (X_0 - a)^+)\big] \\
&= \mathbb{E}\big[(X_m - a)^+\big] - \mathbb{E}\big[(X_0 - a)^+\big].
\end{aligned}$$

Rearranging gives us the desired inequality. □

Let us now prove Theorem 7.1.

Proof of Theorem 7.1. Take any interval $[a,b]$ with $a < b$. Let U_m be the number of complete upcrossings of $[a,b]$ by time m of the process $\{X_n\}_{n\geq 0}$. Then $U_m \nearrow U$, the total number of upcrossings of $[a,b]$ (this may be ∞). Since

$$\mathbb{E}\big[(X_m - a)^+\big] \leq \mathbb{E}\big[X_m^+ + a^+\big] \leq \sup_{m\geq 0} \mathbb{E}\big[X_m^+\big] < \infty,$$

we can apply the monotone convergence theorem and Lemma 7.1 to deduce $\mathbb{E}[U] = \lim_{m\to\infty} \mathbb{E}[U_m] < \infty$. Thus, $\mathbb{P}[U < \infty] = 1$. So, with probability 1, the process $\{X_n\}_{n\geq 0}$ upcrosses $[a,b]$ finitely many times. Thus, the probability that this process upcrosses $[a,b]$ finitely many times for all $a, b \in \mathbb{Q}$ with $a < b$ is 1. But if for some $\omega \in \Omega$, the $\lim X_n(\omega)$ does not exist in $[-\infty, \infty]$ (in the extended real line), then $\liminf X_n(\omega) < \limsup X_n(\omega)$, which implies that there is some rational interval $[a.b]$ such that $X_n(\omega)$ upcrosses $[a,b]$ infinitely many times.[4] Thus, with probability 1,

$$\lim_{n\to\infty} X_n$$

exists in $[-\infty, \infty]$. We thus need only check that $X = \lim_{n\to\infty} X_n \delta_{\{\text{limit exists}\}}$ is finite a.s. By Fatou's lemma, we know

$$\mathbb{E}\big[X^+\big] = \mathbb{E}\big[\lim X_n^+\big] \leq \liminf \mathbb{E}\big[X_n^+\big] < \infty,$$

[3]If $Y_{n-1} \geq b$, then the last upcrossing before time n has already ended, and so $n, n-1$ cannot be in the same upcrossing. If $Y_{n-1} < b$ and $n-1$ is part of an upcrossing, then n is part of an upcrossing and n must also be a part of the same upcrossing, whether or not $Y_n \geq b$. Finally, if $Y_{n-1} < b$ and $n-1$ is not a part of an upcrossing, then n and $n-1$ cannot be part of the same upcrossing.

[4]This must be checked by cases, of which there are four. Namely, each combination of $\liminf X_n(\omega)$ and $\limsup X_n(\omega)$ being finite or infinite should be checked.

where the last step holds since $\sup \mathbb{E}[X_n^+] < \infty$. Also, since $\{X_n\}$ is a submartingale,

$$\mathbb{E}\big[X^-\big] = \mathbb{E}\big[X_n^+\big] - \mathbb{E}\big[X_n^-\big] \le \mathbb{E}\big[X_n^+\big] - \mathbb{E}[X_0].$$

Hence, $\sup \mathbb{E}[X_n^-] < \infty$ and so $\mathbb{E}[X^-] < \infty$ by Fatou's lemma once again. Thus, $\mathbb{E}[|X|] < \infty$. □

8 January 27

8.1 Applying the Submartingale Convergence Theorem

We will discuss Pólya's urn, which is a precursor to models of social networks.

Example 8.1: Pólya's Urn

Suppose there is an urn of infinite capacity. In the beginning, it has one white ball and one black ball. At each time, a ball is picked uniformly at random from the urn and replaced back into it, together with one additional ball of the same color. Equivalently, we choose a color with probability proportional to the number of balls of the same color in the urn, and put an additional ball of the chosen color. Let W_n be the proportion of white balls at time n (starting with $W_0 = 1/2$).

Question: What is the limiting behavior of W_n as $n \to \infty$?

Answer: Let us first check that $\{W_n\}_{n\ge 0}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}_{n\ge 0}$, where $\mathcal{F}_n$ is the σ-algebra generated by all random variables up to time n. For instance, you can generate the process using a sequence of i.i.d. $U[0,1)$ random variables $U_1, U_2, \ldots$ as follows: at time n, you add a white ball if $U_{n+1} \le W_n$, and you add a black ball if $U_{n+1} > W_n$ to obtain the configuration at time $n+1$. Then $\mathcal{F}_n = \sigma(U_1, \ldots, U_n)$. Now, note that at time n is $n+2$. We want to show that $\mathbb{E}[W_{n+1} \mid \mathcal{F}_n] = W_n$. Let N_n denote the number of white balls at time n, which is $(n+2)W_n$. We thus have

$$\mathbb{E}[W_{n+1} \mid \mathcal{F}_n] = \frac{1}{n+3}\,\mathbb{E}[N_{n+1} \mid \mathcal{F}_n].$$

Given $\mathcal{F}_n$, we can write N_{n+1} as

$$N_{n+1} = \begin{cases} N_n + 1 & \text{with probability } W_n \\ N_n & \text{with probability } 1 - W_n. \end{cases}$$

So, the conditional expectation becomes

$$\begin{aligned} \mathbb{E}[N_{n+1} \mid \mathcal{F}_n] &= (N_n + 1)W_n + N_n(1 - W_n) \\ &= ((n+2)W_n + 1)W_n + (n+2)W_n(1 - W_n) \\ &= (n+2)W_n^2 + W_n + (n+2)W_n - (n+2)W_n^2 \\ &= (n+3)W_n, \end{aligned}$$

which proves the martingale property. Now note that $W_n \in [0,1]$ always. So $\mathbb{E}[W_n^+] \le 1$ for all n. Thus, by Theorem 7.1, $W = \lim_{n\to\infty} W_n$ exists a.s. We now check that $W \sim U[0,1]$.[a] We show this by induction that $N_n - 1 \sim U\{0, 1, \ldots, n\}$. This clearly holds for $n = 1$, so suppose that it holds for some n. Then take any $k \in \{0, 1, \ldots, n+1\}$. We have

$$\begin{aligned}
\mathbb{P}[N_{n+1} - 1 = k] &= \sum_{j=0}^{n} \mathbb{P}[N_{n+1} - 1 = k \mid N_n - 1 = j]\mathbb{P}[N_n - 1 = j] \\
&= \frac{1}{n+1} \sum_{j=0}^{n} \mathbb{P}[N_{n+1} = k+1 \mid N_n = j+1] \\
&= \frac{1}{n+1} \big(\mathbb{P}[N_{n+1} = k+1 \mid N_n = k] + \mathbb{P}[N_{n+1} = k+1 \mid N_n = k+1] \big) \\
&= \frac{1}{n+1} \left(\frac{k}{n+2} + 1 - \frac{k+1}{n+2} \right) \\
&= \frac{1}{n+2},
\end{aligned}$$

which establishes the distribution of $N_n - 1$. So

$$W_n = \frac{N_n}{n+2} \sim U\left\{ \frac{1}{n+2}, \ldots, \frac{n+1}{n+2} \right\} \xrightarrow{d} U[0,1] \quad \text{as } n \to \infty,$$

and the result follows since $W_n \to W$ a.s.

[a] In general, the limit is a beta distribution, with the parameters depending on the initial configuration and the number of balls added at each step.

8.2 More Convergence of Martingales

Before proceeding to our next convergence result, we review the concept of uniformly integrable sequences of random variables.

Definition 8.1: Uniformly Integrable Sequence of Random Variables

A sequence of integrable random variables $\{Y_n\}_{n \geq 0}$ is called ***uniformly integrable*** *if for every $\varepsilon > 0$, there is a $k > 0$ such that*

$$\mathbb{E}\big[|Y_n|\delta_{\{|Y_n|>k\}}\big] < \varepsilon$$

for every n.

Uniform integrability gives us a useful convergence result that lets us conclude L^1 convergence from almost sure convergence.

Lemma 8.1: Almost Sure Convergence to L^1 Convergence

If $\{Y_n\}$ is uniformly integrable and $Y_n \to Y$ a.s., then Y is also integrable and $Y_n \to Y$ in L^1.

Proof. Take any $\varepsilon > 0$. Find k such that $\mathbb{E}\big[|Y_n|\delta_{\{|Y_n|>k\}}\big] < \varepsilon$. Then for all n,

$$\mathbb{E}[|Y_n|] = \mathbb{E}\big[|Y_n|\delta_{\{|Y_n|\leq k\}}\big] + \mathbb{E}\big[|Y_n|\delta_{\{|Y_n|>k\}}\big] \leq k + \varepsilon.$$

By Fatou's lemma, we thus have $\mathbb{E}[|Y|] \leq \liminf \mathbb{E}[|Y_n|] \leq k + \varepsilon < \infty$. Now, for every L, we have $|Y|\delta_{\{|Y|>K\}} \leq |Y|$ and $|Y|\delta_{\{|Y|>L\}} \to 0$ as $L \to \infty$, since $\mathbb{P}[|Y| < \infty] = 1$. Thus, by the dominated convergence theorem,

$$\mathbb{E}\big[|Y|\delta_{\{|Y|>L\}}\big] \to 0$$

as $L \to \infty$. This shows that choosing k large enough, we can also ensure that $\mathbb{E}\big[|Y|\delta_{\{|Y|>k\}}\big] < \varepsilon$. So, fix k and ε. Define $\phi : \mathbb{R} \to \mathbb{R}$ which takes value $-k$ in $(-\infty, k]$, value x in $(-k, k)$, and value k in $[k, \infty)$. Then ϕ is a bounded and continuous function, and

$$|\phi(x) - x| \leq \delta_{\{|x|>k\}}.$$

Thus,

$$\begin{aligned}
\mathbb{E}[|Y_n - Y|] &\leq \mathbb{E}[|Y_n - \phi(Y_n)|] + \mathbb{E}[|\phi(Y_n) - \phi(Y)|] + \mathbb{E}[|\phi(Y) - Y|] \\
&\leq \mathbb{E}\big[|Y_n|\delta_{\{|Y_n|>k\}}\big] + \mathbb{E}\big[|Y|\delta_{\{|Y|>k\}}\big] \\
&\leq 2\varepsilon.
\end{aligned}$$

The second inequality follows since $|\phi(Y_n) - \phi(Y)| \to 0$ a.s. and is uniformly bounded by $2k$ (apply the dominated convergence theorem). Hence, $\limsup|Y_n - Y| \leq 2\varepsilon$ holds for every $\varepsilon > 0$, whence we conclude $Y_n \to Y$ in L^1. □

We can also state a result which establishes a necessary and sufficient condition for uniform integrability.

Lemma 8.2: Uniform Integrability Necessary and Sufficient Condition

A sequence $\{Y_n\}_{n\geq 1}$ is uniformly integrable if and only if $\sup \mathbb{E}[|Y_n|] < \infty$ and, for every $\varepsilon > 0$, there is a $\delta > 0$ such that $\mathbb{P}[A] < \delta$ implies $\mathbb{E}[|Y_n|\delta_A] < \varepsilon$ for every n.

Proof. Suppose first that $\{Y_n\}$ is uniformly integrable. We saw in the proof of Lemma 8.1 that $\sup \mathbb{E}[|Y_n|] < \infty$, so take any $\varepsilon > 0$. Find k sufficiently large so that

$$\mathbb{E}\big[|Y_n|\delta_{\{|Y_n|>k\}}\big] < \varepsilon/2$$

for all n. Let $\delta = \varepsilon/(2k)$. If A is any event with $\mathbb{P}[A] < \delta$, then for all n,

$$\begin{aligned}
\mathbb{E}[|Y_n|\delta_A] &= \mathbb{E}\big[|Y_n|\delta_A\delta_{\{|Y_n|\leq k\}}\big] + \mathbb{E}\big[|Y_n|\delta_A\delta_{\{|Y_n|>k\}}\big] \\
&\leq k\mathbb{P}[A] + \mathbb{E}\big[|Y_n|\delta_{\{|Y_n|>k\}}\big] \\
&< k\delta + \varepsilon/2 \\
&= \varepsilon.
\end{aligned}$$

□

We now state and prove the next convergence result.

Theorem 8.1: Lévy's Downwards Convergence Theorem

Let X be an integrable random variable defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_0 \supseteq \mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \cdots$ be a sequence of sub-σ-algebras of $\mathcal{F}$. Let $\mathcal{F}^* = \bigcap_{n=0}^{\infty} \mathcal{F}_n$. Then

$$\mathbb{E}[X \mid \mathcal{F}_n] \to \mathbb{E}[X \mid \mathcal{F}^*]$$

a.s. and in L^1 as $n \to \infty$.

Proof. Let $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$. Consider the time-reversed finite sequence $X_n, X_{n-1}, X_{n-2}, \dots, X_0$. Fix some interval $[a, b]$, where $a < b$, and let U_n be the number of complete upcrossings of $[a, b]$ by this finite sequence. Note that

$$\begin{aligned}
\mathbb{E}[X_j \mid \mathcal{F}_{j+1}] &= \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_j] \mid \mathcal{F}_{j+1}] \\
&= \mathbb{E}[X \mid \mathcal{F}_{j+1}] \\
&= X_{j+1},
\end{aligned}$$

which shows that the reverse sequence is a martingale with respect to the filtration $\mathcal{F}_n \subseteq \mathcal{F}_{n-1} \subseteq \cdots \subseteq \mathcal{F}_0$. Thus, by Lemma 7.1,

$$\begin{aligned}
\mathbb{E}[U_n] &\leq \frac{\mathbb{E}[(X_0 - a)^+] - \mathbb{E}[(X_n - a)^+]}{b - a} \\
&\leq \frac{\mathbb{E}[(X_0 - a)^+]}{b - a} \\
&= \frac{\mathbb{E}[(\mathbb{E}[X \mid \mathcal{F}_0] - a)^+]}{b - a} \\
&\leq \frac{\mathbb{E}[\mathbb{E}[(X - a)^+ \mid \mathcal{F}_0]]}{b - a} \quad \text{(by Theorem 3.1)} \\
&= \frac{\mathbb{E}[(X - a)^+]}{b - a}.
\end{aligned}$$

We can check that $U_0 \leq U_1 \leq U_2 \leq \cdots$. So let $U = \lim_{n\to\infty} U_n$. Then $\mathbb{E}[U] = \lim_{n\to\infty} \mathbb{E}[U_n] \leq \mathbb{E}[(X - a)^+]/(b - a)$, which is finite. So $\mathbb{P}[U < \infty] = 1$. Since $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$, this shows that $X^* = \lim_{n\to\infty} X_n$ exists a.s. We claim that X^* is integrable. This follows by Fatou's lemma, since

$$\mathbb{E}[|X^*|] = \mathbb{E}[\lim |X_n|] \leq \liminf \mathbb{E}[|X_n|],$$

and we know $\mathbb{E}[|X_n|] = \mathbb{E}[|\mathbb{E}[X \mid \mathcal{F}_n]|] \leq \mathbb{E}[\mathbb{E}[|X| \mid \mathcal{F}_n]] = \mathbb{E}[|X|]$ is integrable.

From here, we will show $X^* = \mathbb{E}[X \mid \mathcal{F}^*]$ and that $X_n \to X^*$ in L^1. To this end, we show □