XI

Probability (TO DO)

Part XI: Contents

40 Random variables (TO DO)

write chapter

Having properly developed the Lebesgue measure and the integral on it, we can now proceed to develop random variables.

§40.1 Random variables

With all this set-up, random variables are going to be really quick to define.

Definition 40.1.1. A (real) **random variable** X on a probability space $\Omega = (\Omega, \mathscr{A}, \mu)$ is a measurable function $X \colon \Omega \to \mathbb{R}$, where $\mathbb{R}$ is equipped with the Borel σ-algebra.

In particular, addition of random variables, etc. all makes sense, as we can just add. Also, we can integrate X over Ω, by previous chapter.

Definition 40.1.2 (First properties of random variables)**.** Given a random variable X, the **expected value** of X is defined by the Lebesgue integral

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mu.$$

Confusingly, the letter μ is often used for expected values.

The **kth moment** of X is defined as $\mathbb{E}[X^k]$, for each positive integer $k \geq 1$. The **variance** of X is then defined as

$$\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

> **Question 40.1.3.** Show that $\mathbf{1}_A$ is a random variable (just check that it is Borel measurable), and its expected value is $\mu(A)$.

An important property of expected value you probably already know:

Theorem 40.1.4 (Linearity of expectation)
If X and Y are random variables on Ω then

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Proof. $\mathbb{E}[X+Y] = \int_\Omega X(\omega) + Y(\omega)\, d\mu = \int_\Omega X(\omega)\, d\mu + \int_\Omega Y(\omega)\, d\mu = \mathbb{E}[X] + \mathbb{E}[Y]$. □

Note that X and Y do not have to be "independent" here: a notion we will define shortly.

§40.2 Distribution functions

§40.3 Examples of random variables

§40.4 Characteristic functions

§40.5 Independent random variables

§40.6 A few harder problems to think about

Problem 40A (Equidistribution)**.** Let X_1, X_2, … be i.i.d. uniform random variables on $[0,1]$. Show that almost surely the X_i are equidistributed, meaning that

$$\lim_{N\to\infty} \frac{\#\{1 \le i \le N \mid a \le X_i(\omega) \le b\}}{N} = b - a \qquad \forall 0 \le a < b \le 1$$

holds for almost all choices of ω.

Problem 40B (Side length of triangle independent from median)**.** Let X_1, Y_1, X_2, Y_2, X_3, Y_3 be six independent standard Gaussians. Define triangle ABC in the Cartesian plane by $A = (X_1, Y_1)$, $B = (X_2, Y_2)$, $C = (X_3, Y_3)$. Prove that the length of side BC is independent from the length of the A-median.

41 Large number laws (TO DO)

write chapter

§41.1 Notions of convergence

§41.1.i Almost sure convergence

Definition 41.1.1. Let X, X_n be random variables on a probability space Ω. We say X_n **converges almost surely** to X if

$$\mu\left(\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\right) = 1.$$

This is a very strong notion of convergence: it says in almost every *world*, the values of X_n converge to X. In fact, it is almost better for me to give a *non-example*.

Example 41.1.2 (Non-example of almost sure convergence)
Imagine an immortal skeleton archer is practicing shots, and on the nth shot, he scores a bulls-eye with probability $1-\frac{1}{n}$ (which tends to 1 because the archer improves over time). Let $X_n \in \{0, 1, \dots, 10\}$ be the score of the nth shot.

Although the skeleton is gradually approaching perfection, there are *almost no worlds* in which the archer misses only finitely many shots: that is

$$\mu\left(\omega \in \Omega : \lim_n X_n(\omega) = 10\right) = 0.$$

§41.1.ii Convergence in probability

Therefore, for many purposes we need a weaker notion of convergence.

Definition 41.1.3. Let X, X_n be random variables on a probability space Ω. We say X_n **converges in probability** to X if if for every $\varepsilon > 0$ and $\delta > 0$, we have

$$\mu\left(\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\right) \ge 1 - \delta$$

for n large enough (in terms of ε and δ).

In this sense, our skeleton archer does succeed: for any $\delta > 0$, if $n > \delta^{-1}$ then the skeleton archer does hit a bulls-eye in a $1 - \delta$ fraction of the worlds. In general, you can think of this as saying that for any $\delta > 0$, the chance of an ε-anomaly event at the nth stage eventually drops below δ.

Remark 41.1.4 — To mask δ from the definition, this is sometimes written instead as: for all ε

$$\lim_{n\to\infty} \mu\left(\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\right) = 1.$$

I suppose it doesn't make much difference, though I personally don't like the asymmetry.

§41.1.iii Convergence in law

§41.2 Weak law of large numbers

write

As the name implies, this is a direct corollary of the strong law of large numbers. Nevertheless, the proof of this law is simpler, and some applications only require the weak law.

§41.2.i Application: Weierstrass approximation

§41.3 Strong law of large numbers

§41.3.i Motivation: Biased random walk

Consider a random walk defined as follows:

- Let $X_0 = 1$.
- For each $i \geq 1$, define X_i to be $X_{i-1} - 1$ with probability $p = 0.6$ or $X_{i-1} + 1$ with probability $1 - p = 0.4$.

Then we can ask: What's the expected number of steps until some X_i equals 0?

A naive attempt might be the following.

> Let $f(i)$ be the expected number of steps starting to reach 0 starting from $X_0 = i$.
>
> Then:
>
> - $f(0) = 0$,
> - $f(1) = 1 + 0.6f(0) + 0.4f(2)$,
> - $f(2) = 1 + 0.6f(1) + 0.4f(3)$,
> - $\vdots$

This isn't getting anywhere because there are infinitely many terms. A better attempt is the following:

> Let the answer be x. If we start from $X_0 = 2$, let i be the first time such that $X_i = 1$ and j be the first time after i such that $X_j = 0$. Then
>
> $$\mathbb{E}[i] = \mathbb{E}[j - i] = x.$$
>
> Therefore,
>
> $$x = 1 + 0.6 \cdot 0 + 0.4 \cdot (2x)$$
>
> Solving the equation, we get $x = 5$.

It gives the correct result — however, the same method gives $x = -5$ when the probability of going down is $p = 0.4$, which is clearly nonsense.

What went wrong? The problem is when $p = 0.4$, there is a nonzero probability[1] that the sequence never reaches 0, so the expected value is undefined and we're subtracting ∞ from ∞ in the proof.

In this case, the strong law of large numbers can help us patch this hole, by showing that in almost every world, the sequence X_i eventually reaches 0.

[1] Preview: Using martingale theory next chapter, you will be able to prove the probability the sequence never reaches 0 is exactly $1 - \frac{0.4}{0.6}$.

§41.3.ii Statement

Theorem 41.3.1 (Strong law of large numbers)
Let X_1, X_2, ... be i.i.d. random variables with mean 0. Define the partial mean

$$M_n = \frac{X_1 + \cdots + X_n}{n}.$$

Then, in almost every world, $M_n \to 0$.

In other words, M_n converges almost surely to 0.

The requirement that the mean is 0 is only to simplify the proof, as long as the mean exists, we can subtract the mean from the random variables and apply the theorem.

Example 41.3.2 (The hypothesis $\mathbb{E}[X_i] = 0$ is important)
Consider an example where $M_n \to 0$ does not hold — this is a minor variation of the St. Petersburg paradox.

Let the distribution of each X_i be as follows:

$$X_i = \begin{cases} 1 & \text{with probability } \frac{1}{4} \\ -1 & \text{with probability } \frac{1}{4} \\ 2 & \text{with probability } \frac{1}{8} \\ -2 & \text{with probability } \frac{1}{8} \\ 4 & \text{with probability } \frac{1}{16} \\ -4 & \text{with probability } \frac{1}{16} \\ & \vdots \end{cases}$$

Formally, X_i takes each of the value in $\{2^k, -2^k\}$ with probability 2^{-k-2}.

In this case, the mean $\mathbb{E}[X_i] = \int_\Omega X_i(\omega)$ is actually undefined. Furthermore, as symmetric as the distribution may look, in almost no world will M_n converge to 0.

Intuitively, you can see why:

- Within the first 16 values, on average there's one X_i with $|X_i| = 4$, this will skew M_{16} by $\frac{1}{4}$.
- Within the first 32 values, on average there's one X_i with $|X_i| = 8$, this will skew M_{32} by $\frac{1}{4}$.
- Et cetera.

In other words, just like our skeleton archer, there are almost no worlds in which the M_n got skewed by more than $\frac{1}{4}$ only finitely many times.

§41.3.iii Proof for finite-variance case

In practice, most distribution we ever come across has finite variance, it may be better to give a counterexample.

Example 41.3.3 (A distribution with finite mean but infinite variance)
Taking $Y_i = \operatorname{sgn}(X_i)\sqrt{|X_i|}$ where X_i is the St. Petersburg paradox example above suffices. If you do the math, you will see $\mathbb{E}[Y_i] = 0$, but $\mathbb{E}[Y_i^2] = \infty$.

We will give a proof when $\mathbb{E}[X_i^2]$ is finite first.

First, we define a seemingly unrelated series as follows:

$$T_n = X_1 + \frac{X_2}{2} + \frac{X_3}{3} + \cdots + \frac{X_n}{n}.$$

This step is a bit difficult to motivate. On the positive side, it is easy to show the following:

Claim 41.3.4. In almost every world, the sequence T_n converges.

That is the same as saying the series

$$X_1 + \frac{X_2}{2} + \frac{X_3}{3} + \cdots$$

converges.

The key idea is to show that the total variance of the summands are finite. Indeed:

$$\begin{aligned} \operatorname{Var}[X_1] + \operatorname{Var}\left[\frac{X_2}{2}\right] + \operatorname{Var}\left[\frac{X_3}{3}\right] + \cdots &= \operatorname{Var}[X_1] + \frac{1}{4}\operatorname{Var}[X_2] + \frac{1}{9}\operatorname{Var}[X_3] + \cdots \\ &= \operatorname{Var}[X_1] \cdot \left(1 + \frac{1}{4} + \frac{1}{9} + \cdots\right) \end{aligned}$$

which is finite.

Why should finite total variance imply almost surely convergence? Intuitively, we recall:

Theorem 41.3.5 (Chebyshev's inequality)
Let X be a random variable with mean 0 and variance σ^2. Then

$$\Pr[|X| \ge k\sigma] \le \frac{1}{k^2}.$$

Or equivalently we can write it in the following form, which avoid the $\sqrt{-}$ implicit in the definition of σ:

$$\Pr[|X| \ge a] \le \frac{1}{a^2}\operatorname{Var}[X].$$

So if we look at, say, T_{1000} and T_{2000}:

$$\operatorname{Var}[T_{2000} - T_{1000}] = \sum_{i=1001}^{2000} \frac{\operatorname{Var}[X_i]}{i^2}$$

Because $\sum_{i=1}^{\infty} \frac{\operatorname{Var}[X_i]}{i^2}$ is finite, we expect $\sum_{i=1001}^{2000} \frac{\operatorname{Var}[X_i]}{i^2}$ to be very small, which means T_{2000} should deviate very little from T_{1000}.

To show convergence, we need something stronger, however.

Theorem 41.3.6 (Kolmogorov's inequality)

Let $X_1, \ldots, X_n$ be independent random variables with mean 0. Define $S_i = X_1 + \cdots + X_i$ for each $1 \le i \le n$. Then

$$\Pr[|S_i| \ge a \text{ for any } 1 \le i \le n] \le \frac{1}{a^2} \operatorname{Var}[S_n].$$

You can see why this theorem is stronger — with Chebyshev's inequality, we can only show

$$\Pr[|S_n| \ge a] \le \frac{1}{a^2} \operatorname{Var}[S_n].$$

So, with the same right hand side, we can also bound the probability of $|S_1| \ge a \vee |S_2| \ge a \vee \cdots$ for free!

Proof. Define A_i be the event that i is the smallest value such that $|S_i| \ge a$. Then the left hand side above equals

$$\Pr[|S_i| \ge a \text{ for any } 1 \le i \le n] = \Pr[A_1] + \Pr[A_2] + \cdots + \Pr[A_n].$$

Intuitively, if the events $|S_i| \ge a$ were independent, the best we can do is to use Chebyshev's inequality to bound individual probability values:

$$\Pr[|S_i| \ge a] \le \frac{1}{a^2} \operatorname{Var}[S_i]$$

However, they're very much not independent — in fact, they are positively correlated! For example, we have:

$$\mathbb{E}[S_n \mid S_1 = a] = a$$

because $\mathbb{E}[X_2 + \cdots + X_n] = 0$. So if each X_i is symmetrically distributed, $\Pr[S_n \ge a \mid S_1 = a] \ge \frac{1}{2}$, which is much larger than $\frac{1}{a^2} \operatorname{Var}[S_n]$ for large a.

Here is the formal proof. For each $1 \le i \le n$, we have

$$\mathbb{E}[S_i^2 \mid A_i] \ge a^2$$

which is equivalent to

$$\Pr[A_i] \le \frac{1}{a^2} \mathbb{E}[S_i^2 \cdot \mathbf{1}_{A_i}]$$

and

$$\begin{aligned}\mathbb{E}[S_n^2 \cdot \mathbf{1}_{A_i}] &= \mathbb{E}[(S_i + (S_n - S_i))^2 \cdot \mathbf{1}_{A_i}] \\ &= \mathbb{E}[S_i^2 \cdot \mathbf{1}_{A_i}] + \mathbb{E}[S_i \cdot \mathbf{1}_{A_i}(S_n - S_i)] + \mathbb{E}[(S_n - S_i)^2 \cdot \mathbf{1}_{A_i}]\end{aligned}$$

The middle term $\mathbb{E}[S_i \cdot \mathbf{1}_{A_i}(S_n - S_i)]$ is 0 because $S_i \cdot \mathbf{1}_{A_i}$ and $S_n - S_i = X_{i+1} + \cdots + X_n$ are independent and $\mathbb{E}[X_{i+1} + \cdots + X_n] = 0$, and the last term is ≥ 0.

Combining the inequalities, we get

$$a^2 \Pr[A_i] \le \mathbb{E}[S_n^2 \cdot \mathbf{1}_{A_i}].$$

Summing over all i gives the final result. ☐

Generalizing:

Corollary 41.3.7

Let $X_1, \dots$ be independent random variables with mean 0. Define S_i as above. Then

$$\Pr[|S_i| \ge a \text{ for any } 1 \le i] \le \frac{1}{a^2} \sum_{1 \le i} \operatorname{Var}[X_i].$$

Proof. The event

$$|S_i| \ge a \text{ for any } 1 \le i \le n$$

is a subset of the event

$$|S_i| \ge a \text{ for any } 1 \le i \le n+1$$

therefore we have

$$\Pr[|S_i| \ge a \text{ for any } 1 \le i] = \lim_{n \to \infty} \Pr[|S_i| \ge a \text{ for any } 1 \le i \le n].$$

Applying Kolmogorov's inequality on each $\Pr[|S_i| \ge a \text{ for any } 1 \le i \le n]$, we get the result. □

Now, the idea is to apply this on the *tails* of the sequence

$$X_1, \frac{X_2}{2}, \frac{X_3}{3}, \dots$$

By the corollary, we know that for every $\varepsilon > 0$, in almost every world, there exists n_ε such that for arbitrary $i \ge n_\varepsilon$, $|T_i - T_{n_\varepsilon}| < \frac{\varepsilon}{2}$. By triangle inequality, for arbitrary $i, j \ge n_\varepsilon$, then $|T_i - T_j| < \varepsilon$.

By Cauchy's criterion for convergence, this implies the sequence T_n is convergent in almost every world.

Finally, here is the relation with the original goal:

Claim 41.3.8 (Relation with the original series)**.** In every world where T_n converges, then M_n converges to 0.

Proof. Just a bit of algebraic manipulation. We try to write M_n in terms of T_n.

We have

$$X_n = n \cdot (T_n - T_{n-1})$$

so

$$\begin{aligned} M_n &= \frac{(T_1 - T_0) + 2(T_2 - T_1) + \dots + n(T_n - T_{n-1})}{n} \\ &= \frac{nT_n - (T_0 + T_1 + \dots + T_{n-1})}{n} \\ &= T_n - \frac{T_0 + T_1 + \dots + T_{n-1}}{n}. \end{aligned}$$

Now this is easy: given T_n converges, $\frac{T_0+T_1+\dots+T_{n-1}}{n}$ must also converge to the same value (Cesàro mean), so $M_n \to 0$ as required. □

Exercise 41.3.9. The converse is not true: if $M_n \to 0$, T_n does not necessarily converge. Find a counterexample. (Write T_n in terms of M_n, and see what happens.)

§41.3.iv The general proof

write

The basic idea is to truncate the value of each X_i so that each of them has finite variance.

§41.4 A few harder problems to think about

Problem 41A (Quantifier hell)**.** In the definition of convergence in probability suppose we allowed $\delta = 0$ (rather than $\delta > 0$). Show that the modified definition is equivalent to almost sure convergence.

Problem 41B (Almost sure convergence is not topologizable)**.** Consider the space of all random variables on $\Omega = [0, 1]$. Prove that it's impossible to impose a metric on this space which makes the following statement true:

> A sequence X_1, X_2, ..., of random variables converges almost surely to X if and only if X_i converge to X in the metric.

42 Stopped martingales (TO DO)

§42.1 How to make money almost surely

We now take our newfound knowledge of measure theory to a casino.

Here's the most classical example that shows up: a casino lets us play a game where we can bet any amount of on a fair coin flip, but with bad odds: we win \$$n$ if the coin is heads, but lose \$$2n$ if the coin is tails, for a value of n of our choice. This seems like a game that no one in their right mind would want to play.

Well, if we have unbounded time and money, we actually can almost surely make a profit.

Example 42.1.1 (Being even greedier than 18th century France)
In the game above, we start by betting \$1.

- If we win, we leave having made \$1.
- If we lose, we then bet \$10 instead, and
  - If we win, then we leave having made \$10 − \$2 = \$8, and
  - If we lose then we bet \$100 instead, and
    * If we win, we leave having made \$1000 − \$20 − \$2 = \$978, and
    * If we lose then we bet \$1000 instead, and so on. . .

Since the coin will almost surely show heads eventually, we make money whenever that happens. In fact, the expected amount of time until a coin shows heads is only 2 flips! What could go wrong?

This chapter will show that under sane conditions such as "finite time" or "finite money", one cannot actually make money in this way — the *optional stopping theorem.* This will give us an excuse to define conditional probabilities, and then talk about martingales (which generalize the fair casino).

Once we realize that trying to extract money from Las Vegas is a lost cause, we will stop gambling and then return to solving math problems, by showing some tricky surprises, where problems that look like they have nothing to do with gambling can be solved by considering a suitable martingale.

In everything that follows, $\Omega = (\Omega, \mathscr{A}, \mu)$ is a probability space.

§42.2 Sub-σ-algebras and filtrations

Prototypical example for this section: σ-algebra generated by a random variable, and coin flip filtration.

We considered our Ω as a space of worlds, equipped with a σ-algebra $\mathscr{A}$ that lets us integrate over Ω. However, it is a sad fact of life that at any given time, you only know partial information about the world. For example, at the time of writing, we know that the world did not end in 2012 (see https://en.wikipedia.org/wiki/2012_phenomenon), but the fate of humanity in future years remains at slightly uncertain.

Let's write this measure-theoretically: we could consider

$$\begin{aligned} \Omega &= A \sqcup B \\ A &= \{\omega \text{ for which world ends in 2012}\} \\ B &= \{\omega \text{ for which world does not end in 2012}\} . \end{aligned}$$

We will assume that A and B are measurable sets, that is, $A, B \in \mathscr{A}$. That means we could have good fun arguing about what the values of $\mu(A)$ and $\mu(B)$ should be ("a priori probability that the world ends in 2012"), but let's move on to a different silly example.

We will now introduce a new notion that we will need when we define conditional probabilities later.

Definition 42.2.1. Let $\Omega = (\Omega, \mathscr{A}, \mu)$ be a probability space. A **sub-σ-algebra** $\mathscr{F}$ on Ω is exactly what it sounds like: a σ-algebra $\mathscr{F}$ on the set Ω such that each $A \in \mathscr{F}$ is measurable (i.e., $\mathscr{F} \subseteq \mathscr{A}$).

The motivation is that $\mathscr{F}$ is the σ-algebra of sets which let us ask questions about some piece of information. For example, in the 2012 example we gave above, we might take $\mathscr{F} = \{\varnothing, A, B, \Omega\}$, which are the sets we care about if we are thinking only about 2012.

Here are some more serious examples.

Example 42.2.2 (Examples of sub-σ-algebras)

(a) Let $X : \Omega \to \{0, 1, 2\}$ be a random variable taking on one of three values. If we're interested in X then we could define

$$\begin{aligned} A &= \{\omega \mid X(\omega) = 1\} \\ B &= \{\omega \mid X(\omega) = 2\} \\ C &= \{\omega \mid X(\omega) = 3\} \end{aligned}$$

then we could write

$$\mathscr{F} = \{\varnothing,\ A,\ B,\ C,\ A \cup B,\ B \cup C,\ C \cup A,\ \Omega\} .$$

This is a sub-σ-algebra on $\mathscr{F}$ that lets us ask questions about X like "what is the probability $X \neq 3$", say.

(b) Now suppose $Y : \Omega \to [0, 1]$ is another random variable. If we are interested in Y, the $\mathscr{F}$ that captures our curiosity is

$$\mathscr{F} = \{Y^{\text{pre}}(B) \mid B \subseteq [0, 1] \text{ is measurable }\} .$$

You might notice a trend here which we formalize now:

Definition 42.2.3. Let $X : \Omega \to \mathbb{R}$ be a random variable. The **sub-σ-algebra generated by X** is defined by

$$\sigma(X) := \{X^{\text{pre}}(B) \mid B \subseteq \mathbb{R} \text{ is measurable }\} .$$

If $X_1, \ldots$ is a sequence (finite or infinite) of random variables, the sub-σ-algebra generated by them is the smallest σ-algebra which contains $\sigma(X_i)$ for each i.

Finally, we can put a lot of these together — since we're talking about time, we learn more as we grow older, and this can be formalized.

Definition 42.2.4. A **filtration** on $\Omega = (\Omega, \mathscr{A}, \mu)$ is a nested sequence[1]

$$\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \dots$$

of sub-σ-algebras on Ω.

Example 42.2.5 (Filtration)

Suppose you're bored in an infinitely long class and start flipping a fair coin to pass the time. (Accordingly, we could let $\Omega = \{H, T\}^\infty$ consist of infinite sequences of heads H and tails T.) We could let $\mathscr{F}_n$ denote the sub-σ-algebra generated by the values of the first n coin flips. So:

- $\mathscr{F}_0 = \{\varnothing, \Omega\}$,
- $\mathscr{F}_1 = \{\varnothing, \text{first flip } H, \text{first flip } T, \Omega\}$,
- $\mathscr{F}_2 = \{\varnothing, \text{first flips } HH, \text{second flip } T, \Omega, \text{first flip and second flip differ}, \dots\}$.
- and so on, with $\mathscr{F}_n$ being the measurable sets "determined" only by the first n coin flips.

Exercise 42.2.6. In the previous example, compute the cardinality $|\mathscr{F}_n|$ for each integer n.

More importantly,

X is $\mathscr{F}$-measurable if X is determined only by the information given in $\mathscr{F}$.

Example 42.2.7

In the example above, let X_3 be the value of the third coin flip. Then:

- X_3 is not $\mathscr{F}_2$-measurable. (That is, we don't know X_3 from the knowledge of the first 2 coin flips.)
- But it is $\mathscr{F}_3$-measurable.

Exercise 42.2.8. Check this! (Recall that a function is measurable if it lifts open sets to measurable sets. So you need to show e.g. $X_3^{\text{pre}}(\{H\}) \notin \mathscr{F}_2$.)

So, not only can we formalize partial information about the world, we can also formalize what it means for something to only depend on that partial information.

[1]For convenience, we will restrict ourselves to $\mathbb{Z}_{\geq 0}$-indexed filtrations, though really any index set is okay.

§42.3 Conditional expectation

Prototypical example for this section: $\mathbb{E}(X \mid X+Y)$ *for* X *and* Y *distributed over* $[0,1]$.

We'll need the definition of conditional probability to define a martingale, but this turns out to be surprisingly tricky. Let's consider the following simple example to see why.

Example 42.3.1 (Why high-school methods aren't enough here)

Suppose we have two independent random variables X, Y distributed uniformly over $[0,1]$ (so we may as well take $\Omega = [0,1]^2$). We might try to ask the question:

> *"what is the expected value of X given that $X+Y=0.6$"?*

Intuitively, we know the answer has to be 0.3. However, if we try to write down a definition, we quickly run into trouble. Ideally we want to say something like

$$\mathbb{E}[X \text{ given } X+Y=0.6] = \frac{\int_S X}{\int_S 1} \text{ where } S = \{\omega \in \Omega \mid X(\omega)+Y(\omega)=0.6\}.$$

The problem is that S is a set of measure zero, so we quickly run into $\frac{0}{0}$, meaning a definition of this shape will not work out.

The way that this is typically handled in measure theory is to use the notion of sub-σ-algebra that we defined.

But first, we should explain what $\mathbb{E}(X \mid X+Y)$ means first — why are we conditioning on another random variable instead of an event?

To motivate conditioning on a random variable, consider the following situation. Suppose that the weather tomorrow depends on the weather today and the random fluctuations. So we may have statements such as:

$$\Pr(\text{it rains tomorrow} \mid \text{it rains today}) = 0.6,$$
$$\Pr(\text{it rains tomorrow} \mid \text{it doesn't rain today}) = 0.3.$$

This is the standard conditional probability: $\Pr(A \mid B) = \frac{\Pr(A \wedge B)}{\Pr(B)}$.

Note that "the weather today" is itself a random variable.

Let Z be the weather forecast tonight's prediction of the probability, suppose it works as above. Then Z is a random real variable, defined by:

$$Z\colon \Omega \to \mathbb{R}$$
$$Z(\omega) = \Pr(\text{it rains tomorrow} \mid \text{weather today} = \text{weather today}(\omega))$$

It would only be reasonable to write

$$Z = \Pr(\text{it rains tomorrow} \mid \text{weather today}).$$

We're conditioning on a random variable, and $\Pr(\cdots \mid \cdots)$ is itself a random variable instead of a single value in $\mathbb{R}$, but that's perfectly okay.

Similarly, if Ω is finite and every subset of it is measurable, for random real variables X and Y it would be sensible for us to define random real variable $Z = \mathbb{E}(X \mid Y)$ by

$$Z\colon \Omega \to \mathbb{R}$$
$$Z(\omega) = \mathbb{E}[X \mid Y = Y(\omega)].$$

Example 42.3.2

Let X and Y be the result of rolling two dices. Then:

- $\mathbb{E}[X \mid X+Y=3] = 1.5$, as you can easily calculate.
- $\mathbb{E}(X \mid X+Y)$ is a random variable, which would takes the value 1.5 in any world whether $X+Y=3$.
- More generally, we have in fact

$$\mathbb{E}(X \mid X+Y) = \frac{X+Y}{2}.$$

Notice how the random variable $\mathbb{E}(X \mid X+Y)$ depends only on the value of $X+Y$ — by definition.

Of course, as we explained earlier, this naive attempts will give us division-by-zero everywhere for the continuous case — so, enters the sub-σ-algebra.

Proposition 42.3.3 (Conditional expectation definition)

Let $X \colon \Omega \to \mathbb{R}$ be an *absolutely integrable* random variable (meaning $\mathbb{E}[|X|] < \infty$) over a probability space Ω, and let $\mathscr{F}$ be a sub-σ-algebra on it.

Then there exists a function $\eta \colon \Omega \to \mathbb{R}$ satisfying the following two properties:

- η is $\mathscr{F}$-measurable (that is, measurable as a function $(\Omega, \mathscr{F}, \mu) \to \mathbb{R}$); and
- for any set $A \in \mathscr{F}$ we have $\mathbb{E}[\eta \cdot \mathbf{1}_A] = \mathbb{E}[X \cdot \mathbf{1}_A]$.

Moreover, this random variable is unique up to almost sureness.

Proof. Omitted, but relevant buzzword used is "Radon-Nikodym derivative". □

Definition 42.3.4. Let η be as in the previous proposition.

- We denote η by $\mathbb{E}(X \mid \mathscr{F})$ and call it the **conditional expectation** of X with respect to $\mathscr{F}$.
- If Y is a random variable then $\mathbb{E}(X \mid Y)$ denotes $\mathbb{E}(X \mid \sigma(Y))$, i.e. the conditional expectation of X with respect to the σ-algebra generated by Y.

Example 42.3.5

As we can expect, $\eta = \frac{X+Y}{2}$ satisfies the condition of $\mathbb{E}(X \mid X+Y)$ above.

The way to motivate doing all this is the following. We want to be able to say something like:

$$\mathbb{E}[X \mid X+Y=0.6] = \lim_{\varepsilon \to 0} \mathbb{E}[X \mid 0.6-\varepsilon < X+Y < 0.6+\varepsilon]$$

Unfortunately, this setup does not work in general where $\mathscr{F}$ might not be generated by just one random real variable. Let's see how the definition above helps us.

- Let $A = \{\omega \in \Omega \mid 0.6-\varepsilon < X(\omega)+Y(\omega) < 0.6+\varepsilon\}$, this set certainly belongs to the sub-σ-algebra generated by $X+Y$ (because it is $(X+Y)^{\text{pre}}((0.6-\varepsilon, 0.6+\varepsilon))$).

- Recall that η is $\mathscr{F}$-measurable means η only depends on the information in $\mathscr{F} = \sigma(X+Y)$, that is, on $X+Y$. This makes sense.
- Look at the right hand side:

$$\mathbb{E}[X \mid 0.6 - \varepsilon < X + Y < 0.6 + \varepsilon] = \mathbb{E}[X \cdot \mathbf{1}_A].$$

The law of total expectation says that $\mathbb{E}[\mathbb{E}(X \mid Y)] = \mathbb{E}[X]$. So, intuitively, the second property above simply requires this law to hold over all set $A \in \mathscr{F}$.

In our case, we have the following:

$$\mathbb{E}[\eta \mid 0.6 - \varepsilon < X + Y < 0.6 + \varepsilon] = \mathbb{E}[X \mid 0.6 - \varepsilon < X + Y < 0.6 + \varepsilon].$$

More fine print:

Remark 42.3.6 (This notation is terrible) — The notation $\mathbb{E}(X \mid \mathscr{F})$ is admittedly confusing, since it is actually an entire function $\Omega \to \mathbb{R}$, rather than just a real number like $\mathbb{E}[X]$ — though, as you can see, it has its merits. For this reason I try to be careful to remember to use parentheses rather than square brackets for conditional expectations; not everyone does this.

Abuse of Notation 42.3.7. In addition, when we write $Y = \mathbb{E}(X \mid \mathscr{F})$, there is some abuse of notation happening here since $\mathbb{E}(X \mid \mathscr{F})$ is defined only up to some reasonable uniqueness (i.e. up to measure zero changes). So this really means that "Y satisfies the hypothesis of Proposition 42.3.3", but this is so pedantic that no one bothers.

For example, in the example above, if we change η to be

$$\eta(\omega) = \begin{cases} 0 & \text{if } X(\omega) + Y(\omega) = 0.6 \\ \frac{X(\omega)+Y(\omega)}{2} & \text{otherwise} \end{cases}$$

then $\mathbb{E}[\eta \cdot \mathbf{1}_A] = \mathbb{E}[X \cdot \mathbf{1}_A]$ still holds for every set A, but now it seems to be saying $\mathbb{E}[X \mid X + Y = 0.6] = 0$?

Nevertheless, we must agree that we must sacrifice a measure zero set, since otherwise if we have

$$T(\omega) = \begin{cases} 1 & \text{if } Y(\omega) > 0.5 \text{ or } (Y(\omega) = 0.5 \text{ and } X(\omega) \in \mathbb{Q}) \\ 0 & \text{otherwise} \end{cases}$$

then it is certainly measurable (i.e. a random variable), $\mathbb{E}[T \mid Y = 0.4] = 0$ and $\mathbb{E}[T \mid Y = 0.6] = 1$, but what is $\mathbb{E}[T \mid Y = 0.5]$? (You may argue it should be 0, but what if $\mathbb{Q}$ is changed to something more non-measurable? Besides, why should the conditional expectation change when we only modify T on a probability-zero set anyway?)

properties

§42.4 Supermartingales

Prototypical example for this section: Visiting a casino is a supermartingale, assuming house odds.

Definition 42.4.1. Let X_0, X_1, ... be a sequence of random variables on a probability space Ω, and let $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots$ be a filtration.

Then $(X_n)_{n\ge 0}$ is a **supermartingale** with respect to $(\mathscr{F}_n)_{n\ge 0}$ if the following conditions hold:

- X_n is absolutely integrable for every n;
- X_n is measurable with respect to $\mathscr{F}_n$; and
- for each $n = 1, 2, \dots$ the inequality

  $$\mathbb{E}(X_n \mid \mathscr{F}_{n-1}) \le X_{n-1}$$

  holds for all $\omega \in \Omega$.

In a **submartingale** the inequality $\le$ is replaced with $\ge$, and in a **martingale** it is replaced by $=$.

what's the etymology of the term?

Abuse of Notation 42.4.2 (No one uses that filtration thing anyways)**.** We will always take $\mathscr{F}_n$ to be the σ-algebra generated by the previous variables X_0, X_1, …, X_{n-1}, and do so without further comment. Nonetheless, all the results that follow hold in the more general setting of a supermartingale with respect to some filtration.

We will prove all our theorems for supermartingales; the analogous versions for submartingales can be obtained by replacing $\le$ with $\ge$ everywhere (since X_n is a martingale iff $-X_n$ is a supermartingale) and for martingales by replacing $\le$ with $=$ everywhere (since X_n is a martingale iff it is both a supermartingale and a submartingale).

Let's give examples.

Example 42.4.3 (Supermartingales)

(a) **Random walks**: an ant starts at the position 0 on the number line. Every minute, it flips a fair coin and either walks one step left or one step right. If X_t is the position at the tth time, then X_t is a martingale, because

$$\mathbb{E}(X_t \mid X_0, \dots, X_{t-1}) = \frac{(X_{t-1}+1) + (X_{t-1}-1)}{2} = X_{t-1}.$$

(b) **Casino game**: Consider a gambler using the strategy described at the beginning of the chapter. This is a martingale, since every bet the gambler makes has expected value 0.

(c) **Multiplying independent variables**: Let X_1, X_2, …, be independent (not necessarily identically distributed) integrable random variables with mean 1. Then the sequence Y_1, Y_2, … defined by

$$Y_n := X_1 X_2 \cdots X_n$$

is a martingale; as $\mathbb{E}(Y_n \mid Y_1, \dots, Y_{n-1}) = \mathbb{E}[Y_n] \cdot Y_{n-1} = Y_{n-1}$.

(d) **Iterated blackjack**: Suppose one shows up to a casino and plays infinitely many games of blackjack. If X_t is their wealth at time t, then X_t is a supermartingale. This is because each game has negative expected value (house edge).

Example 42.4.4 (Frivolous/inflamamtory example — real life is a supermartingale)
Let X_t be your happiness on day t of your life. Life has its ups and downs, so it is not the case that $X_t \le X_{t-1}$ for every t. For example, you might win the lottery one day.

However, on any given day, many things can go wrong (e.g. zombie apocalypse), and by Murphy's Law this is more likely than things going well. Also, as you get older, you have an increasing number of responsibilities and your health gradually begins to deteriorate.

Thus it seems that

$$\mathbb{E}(X_t \mid X_0, \dots, X_{t-1}) \le X_{t-1}$$

is a reasonable description of the future — *in expectation*, each successive day is slightly worse than the previous one. (In particular, if we set $X_t = -\infty$ on death, then as long as you have a positive probability of dying, the displayed inequality is obviously true.)

Before going on, we will state without proof one useful result: if a martingale is bounded, then it will almost certainly converge.

Theorem 42.4.5 (Doob's martingale convergence theorem)
Let $X_0, \dots$ be a supermartingale on a probability space Ω such that

$$\sup_{n \in \mathbb{Z}_{\ge 0}} \mathbb{E}\left[|X_n|\right] < \infty.$$

Then, there exists a random variable $X_\infty \colon \Omega \to \mathbb{R}$ such that

$$X_n \xrightarrow{\text{a.s.}} X_\infty.$$

§42.5 Optional stopping

Prototypical example for this section: Las Vegas.

In the first section we described how to make money almost surely. The key advantage the gambler had was the ability to quit whenever he wanted (equivalently, an ability to control the size of the bets; betting \$0 forever is the same as quitting.) Let's formalize a notion of stopping time.

The idea is we want to define a function $\tau \colon \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ such that

- $\tau(\omega)$ specifies the index after which we *stop* the martingale. Note that the decisions to stop after time n must be made with only the information available at that time — i.e., with respect to $\mathscr{F}_n$ of the filtration.
- $X_{\tau \wedge n}$ is the random value representing the value at time n of the stopped martingale, where if n is *after* the stopping time, we just take it to be the our currently value after we leave.

  So for example in a world ω where we stopped at time 3, then $X_{\tau\wedge 0}(\omega) = X_0(\omega)$, $X_{\tau\wedge 1}(\omega) = X_1(\omega)$, $X_{\tau\wedge 2}(\omega) = X_2(\omega)$, $X_{\tau\wedge 3}(\omega) = X_3(\omega)$, but then

  $$X_3(\omega) = X_{\tau\wedge 4}(\omega) = X_{\tau\wedge 5}(\omega) = X_{\tau\wedge 6}(\omega) = \dots$$

  since we have stopped — the value stops changing.

- X_τ denotes the eventual value after we stop (or the limit X_∞ if we never stop).

Here's the compiled machine code.

Definition 42.5.1. Let $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots$ be a filtration on a probability space Ω.

- A **stopping time** is a function

  $$\tau \colon \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$$

  with the property that for each integer n, the set

  $$\{\omega \in \Omega \mid \tau(\omega) = n\}$$

  is $\mathscr{F}_n$-measurable (i.e., is in $\mathscr{F}_n$).

- For each $n \ge 0$ we define $X_{\tau \wedge n} \colon \Omega \to \mathbb{R}$ by

  $$X_{\tau \wedge n}(\omega) = X_{\min\{\tau(\omega), n\}}(\omega)$$

- Finally, we let the eventual outcome be denoted by

  $$X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \tau(\omega) \neq \infty \\ \lim_{n \to \infty} X_n(\omega) & \tau(\omega) = \infty \text{ and } \lim_{n\to\infty} X_n(\omega) \text{ exists} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

  We require that the "undefined" case occurs only for a set of measure zero (for example, if Theorem 42.4.5 applies). Otherwise we don't allow X_τ to be defined.

Proposition 42.5.2 (Stopped supermartingales are still supermartingales)
Let X_0, X_1, ... be a supermartingale. Then the sequence

$$X_{\tau \wedge 0}, \; X_{\tau \wedge 1}, \; \dots$$

is itself a supermartingale.

Proof. We have almost everywhere the inequalities

$$\begin{aligned} \mathbb{E}\left(X_{\tau \wedge n} \mid \mathscr{F}_{n-1}\right) &= \mathbb{E}\left(X_{n-1} + \mathbf{1}_{\tau(\omega)=n-1}(X_n - X_{n-1}) \mid \mathscr{F}_{n-1}\right) \\ &= \mathbb{E}\left(X_{n-1} \mid \mathscr{F}_{n-1}\right) + \mathbb{E}\left(\mathbf{1}_{\tau(\omega)=n-1} \cdot (X_n - X_{n-1}) \mid \mathscr{F}_{n-1}\right) \\ &= X_{n-1} + \mathbf{1}_{\tau(\omega)=n-1} \cdot \mathbb{E}\left(X_n - X_{n-1} \mid \mathscr{F}_{n-1}\right) \le X_{n-1} \end{aligned}$$

as functions from $\Omega \to \mathbb{R}$. ∎

Theorem 42.5.3 (Doob's optional stopping theorem)
Let $X_0, X_1, \dots$ be a supermartingale on a probability space Ω, with respect to a filtration $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots$. Let τ be a stopping time with respect to this filtration. Suppose that *any* of the following hypotheses are true, for some constant C:

(a) **Finite time**: $\tau(\omega) \le C$ for almost all ω.

(b) **Finite money**: for each $n \ge 1$, $|X_{\tau \wedge n}(\omega)| \le C$ for almost all ω.

(c) **Finite bets**: we have $\mathbb{E}[\tau] < \infty$, and for each $n \ge 1$, the conditional expectation

$$\mathbb{E}\left(|X_n - X_{n-1}| \mid \mathscr{F}_n\right)$$

takes on values at most C for almost all $\omega \in \Omega$ satisfying $\tau(\omega) \ge n$.

Then X_τ is well-defined almost everywhere, and more importantly,

$$\mathbb{E}[X_\tau] \le \mathbb{E}[X_0].$$

The last equation can be cheekily expressed as "the only winning move is not to play".

do later tonight

Proof. □

Exercise 42.5.4. Conclude that going to Las Vegas with the strategy described in the first section is a really bad idea. What goes wrong?

While this is useful to make us stop gambling, it doesn't allow us to compute anything — we don't know anything about $\mathbb{E}[X_\tau]$ other than it's $\le \mathbb{E}[X_0]$. However:

Corollary 42.5.5
With the same hypothesis as above:

- If $X_0, X_1, \dots$ is a submartingale, then $\mathbb{E}[X_\tau] \ge \mathbb{E}[X_0]$.
- If it is a martingale, then $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.

Proof. If $X_0, X_1, \dots$ is a submartingale, then $Y_0, Y_1, \dots$ defined by $Y_i = -X_i$ is a supermartingale, and the hypothesis is still satisfied. Apply the theorem to Y_τ we get the result.

If $X_0, X_1, \dots$ is a martingale, then it is both a supermartingale and a submartingale, the result follows immediately. □

This finally let us calculate something — if we can compute $\mathbb{E}[X_0]$ and write the result as $\mathbb{E}[X_\tau]$ for some martingale, then we can solve the problem!

§42.6 Fun applications of optional stopping (TO DO)

We now give three problems which showcase some of the power of the results we have developed so far.

§42.6.i The ballot problem

Suppose Alice and Bob are racing in an election; Alice received a votes total while Bob received b votes total, and $a > b$. If the votes are chosen in random order, one could ask: what is the probability that Alice remains strictly ahead of Bob in the election?

Proposition 42.6.1 (Ballot problem)
This occurs with probability $\frac{a-b}{a+b}$.

We should try to model this as a martingale. A natural way to do it is the random walk, as in Example 42.4.3:

$$X_0 = 0$$
$$X_i = X_{i-1} + \begin{cases} 1 \text{ with probability } \frac{1}{2} \\ -1 \text{ otherwise.} \end{cases}$$

Here, each 1 represents Alice getting a vote, and each -1 represents Bob getting a vote.

Then, we need to compute

$$\Pr[X_i > 0 \text{ for all } 1 \le i \le a+b \mid X_{a+b} = a-b].$$

While this is natural, the fact that the probability is conditioned on $X_{a+b} = a - b$ makes us unable to apply the optional stopping theorem.

Instead, the following will work: We start with all the votes, and *remove* them in random order.

$$(A_0, B_0) = (a, b),$$
$$(A_i, B_i) = \begin{cases} (A_{i-1} - 1, B_{i-1}) \text{ with probability } \frac{A_{i-1}}{A_{i-1}+B_{i-1}} \\ (A_{i-1}, B_{i-1} - 1) \text{ otherwise} \end{cases} \quad \text{for } 1 \le i \le a+b.$$

Of course, here A_i represents the number of votes Alice has left, and B_i represents the number of votes Bob has left.

Now we just need to compute

$$\Pr[A_i > B_i \text{ for all } 0 \le i \le a+b-1].$$

There's no longer any conditional expectation!

Next, we need to construct a martingale. We could try to define $X_i = A_i - B_i$ similar to above, but that will not be a martingale.

Here is the trick: we *modify* X_i to make it a martingale. (More examples where a sequence of random variables is modified to create a martingale can be found in Problem 42A.)

Example 42.6.2

Suppose $a = 2$ and $b = 1$. Then:

$$(A_0, B_0) = (2, 1),$$

$$(A_1, B_1) = \begin{cases} (1,1) \text{ with probability } \frac{2}{3} \\ (2,0) \text{ otherwise.} \end{cases}$$

So $\mathbb{E}[A_0 - B_0] = 1$ while $\mathbb{E}[A_1 - B_1] = \frac{2}{3} \cdot 0 + \frac{1}{3} \cdot 2 = \frac{2}{3}$, if we define X_i as above of course it wouldn't be a martingale.

However, it's not difficult to find ways to modify it to form a martingale. For example:

- If we define $X_1 = A_1 - B_1 + \frac{1}{3}$, then $\mathbb{E}[X_1] = 1$, so we're fine.
- Similarly, we can also define $X_1 = \frac{3}{2} \cdot (A_1 - B_1)$.
- Or $X_1 = \frac{9}{4} \cdot (A_1 - B_1)^2$.
- ... et cetera...

We need to make $\mathbb{E}[X_i \mid X_0 = x_0, X_1 = x_1, \dots, X_{i-1} = x_{i-1}] = x_{i-1}$ for *every* i and every $(x_0, x_1, \dots, x_{i-1})$. (You can try to find out yourself what modification will work yourself before continue reading.)

Turns out, the following will work:

$$X_i = \frac{A_i - B_i}{a + b - i} \text{ for } 0 \le i \le a + b - 1.$$

We cannot extend it to $i \ge a+b$, but it is fine. (If you're worried, just define $X_i = X_{a+b-1}$ for $i \ge a + b$.)

Exercise 42.6.3. Check that it works. (The math is very similar to the problem about Pólya's urn in Problem 42A.)

For the stopping time, there is only one natural way to define it: τ is $a + b - 1$ if Alice remains ahead of Bob i.e. $X_i > 0$ for every $0 \le i \le a + b - 1$, otherwise τ is the smallest i such that $X_i = 0$.

Then, the optional stopping theorem states:

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

$\mathbb{E}[X_0]$ is easy to calculate, it is $\frac{A_0 - B_0}{a+b} = \frac{a-b}{a+b}$. What is $\mathbb{E}[X_\tau]$?

- If Alice remains ahead of Bob, $X_\tau = X_{a+b-1} = 1$.
- Otherwise, $X_\tau = 0$.

Therefore $\mathbb{E}[X_\tau]$ is exactly the probability we need to calculate, so we're done.

Remark 42.6.4 (This is cheating a little) — Note that you can equivalently write

$$X_i = \frac{A_i - B_i}{A_i + B_i}.$$

Which is exactly the form of the answer.

That is to say, if you already know the form of the answer, martingale theory can help you to check it. But if you don't...

§42.6.ii ABRACADABRA

To be written.
https://www.jeremykun.com/2014/03/03/martingales-and-the-optional-stopping-theorem/

§42.6.iii USA TST 2018

To be written.
https://aops.com/community/p9513099

§42.7 A few harder problems to think about

Problem 42A (Examples of martingales)**.** We give some more examples of martingales.

(a) **(Simple random walk)** Let X_1, X_2, ... be i.i.d. random variables which equal $+1$ with probability $1/2$, and -1 with probability $1/2$. Prove that

$$Y_n = (X_1 + \cdots + X_n)^2 - n$$

is a martingale.

(b) **(de Moivre's martingale)** Fix real numbers p and q such that $p, q > 0$ and $p+q = 1$. Let X_1, X_2, ... be i.i.d. random variables which equal $+1$ with probability p, and -1 with probability q. Show that

$$Y_n = \left(qp^{-1}\right)^{X_1+X_2+\cdots+X_n}$$

is a martingale.

(c) **(Pólya's urn)** An urn contains one red and one blue marble initially. Every minute, a marble is randomly removed from the urn, and two more marbles of the same color are added to the urn. Thus after n minutes, the urn will have $n+2$ marbles.

Let r_n denote the fraction of marbles which are red. Show that r_n is a martingale.

Problem 42B. A deck has 52 cards; of them 26 are red and 26 are black. The cards are drawn and revealed one at a time. At any point, if there is at least one card remaining in the deck, you may stop the dealer; you win if (and only if) the next card in the deck is red. If all cards are dealt, then you lose. Across all possible strategies, determine the maximal probability of winning.

Problem 42C (Wald's identity)**.** Let μ be a real number. Let X_1, X_2, ... be independent random variables on a probability space Ω with mean μ. Finally let $\tau \colon \Omega \to \{1, 2, \dots\}$ be a stopping time such that $\mathbb{E}[\tau] < \infty$, and the event $\tau = n$ depends only on X_1, ..., X_n.

Prove that

$$\mathbb{E}[X_1 + X_2 + \cdots + X_\tau] = \mu\mathbb{E}[\tau].$$

Problem 42D (Unbiased drunkard's walk)**.** An ant starts at 0 on a number line, and walks left or right one unit with probability $1/2$. It stops once it reaches either -17 or $+8$.

(a) Find the probability it reaches $+8$ before -17.

(b) Find the expected value of the amount of time it takes to reach either endpoint.

Problem 42E (Biased drunkard's walk)**.** Let $0 < p < 1$ be a real number. An ant starts at 0 on a number line, and walks left or right one unit with probability p. It stops once it reaches either -17 or $+8$. Find the probability it reaches $+8$ first.

Problem 42F. The number 1 is written on a blackboard. Every minute, if the number a is written on the board, it's erased and replaced by a real number in the interval $[0, 2.01a]$ selected uniformly at random. What is the probability that the resulting sequence of numbers approaches 0?