Notes 23

Recall that a *random variable* is a function R from the sample space of a probability distribution to the reals.

In this lecture, we study the important notion of *expectation* of a random variable. Expectation condenses the entire random variable down to a single number, an "average". Obviously, loses lots of information by doing this.

1 Definition and Simple Examples

Fix a probability distribution Pr on sample space S, and let R be a random variable, $R : S \to \mathbb{R}$.

Definition 1.1 *The* **expected value** *or* **expectation** *or* **mean** *of random variable R is*

$$E[R] = \sum_{s \in S} R(s) \cdot \Pr(s)$$

This can be stated equivalently (also used often as the definition):

Theorem 1.2 $E[R] = \sum_{r \in \mathbb{R}} r \cdot \Pr(R = r)$

Proof. These are equivalent, because the second is just obtained from the first by grouping together the points that have the same R-value.
Formally,

$$\begin{aligned}
E[R] &= \sum_{s \in S} R(s) \cdot \Pr(s) \\
&= \sum_{r \in \mathbb{R}} \sum_{s \mid R(s) = r} R(s) \cdot \Pr(s) \\
&= \sum_{r \in \mathbb{R}} \sum_{s \mid R(s) = r} r \cdot \Pr(s) \\
&= \sum_{r \in \mathbb{R}} r \cdot \sum_{s \mid R(s) = r} \Pr(s) \\
&= \sum_{r \in \mathbb{R}} r \cdot \Pr(R = r)
\end{aligned}$$

■

Example 1 Roll 1 six-sided die.
Sample space S is the six sides.
Random variable R is the number of spots.
The mean (calculated by either definition above) is:

$$E[R] = \sum_{i=1}^{6} i\left(\frac{1}{6}\right) = \frac{1}{6}\frac{6 \cdot 7}{2} = 3.5$$

Recall than the *indicator variable* for an event A is the random variable that's 1 for elements of A, 0 for elements of $\bar{A}$.

Example 2 If I is the indicator variable for an event A, then, calculating using the second definition:

$$\begin{aligned} E[I] &= 1 \cdot \Pr(I = 1) + 0 \cdot \Pr(I = 0) \\ &= 1 \cdot \Pr(A) + 0 \cdot \Pr(\overline{A}) \\ &= \Pr[A] \end{aligned}$$

Very roughly speaking, the expected value is supposed to indicate what we "expect" the value of the random variable to be if we select a sample point according to the probability distribution.

What does this mean?

It doesn't really say anything about what will happen on any particular trial.
It does tell us about what should happen when there are a large number of repeated trials.
Next week, we will study a theorem known as the *Law of Large Numbers*, which says that, in a large number of experiments, the probability is very high that the *average* outcome is close to the expected value.

Example 3 In one die roll, there is no reason to expect an outcome near 3.5.
But over many die rolls, the outcomes will almost certainly average to around 3.5.

Note that the mean doesn't tell much about the distribution.
Completely different distributions can have the same mean:

Example 4 Suppose the die has 3 sides containing 1 spot, 3 sides with 6 spots.
Mean is still 3.5. Very different distribution.

The expectation is the weighted average of all possible values of the random variable.
The concept of expectation also arises for continuous random variables, only there, the sum is replaced by an integral.

Expectation is often just called the *average*.
But be careful – this is the *mean*, not the *median*:

Definition 1.3 *The* **median** *of R is the unique value $r \in Range(R)$ such that* $\Pr(R < r) \leq 1/2$ *and* $\Pr(R > r) < 1/2$.
(Sometimes the $\leq$ and $<$ are switched.)

Example 5 Ordinary die.
The median is 4, because the probability that the number of dots is strictly less than 4 is exactly 1/2, while the probability that the number of dots is strictly greater than 4 is less than 1/2.
(Sometimes say 3.)

The median and mean can be very far apart.

Example 6 $2n$-sided die, with n sides having 0, n having 100.
Mean is 50.
Median is 100 (or 0).

2 Computing the Mean—a Shortcut

If the range of a random variable is the natural numbers, there is a different way to express the mean.
This can sometimes help in computing it.

Theorem 2.1 *If $R : S \to \mathbb{N}$, then*

$$E[R] = \sum_{i=0}^{\infty} \Pr(R > i).$$

Proof.

$$\begin{array}{rcllllll}
\sum_{i=0}^{\infty} \Pr(R > i) & = & \Pr(R > 0)+ & \Pr(R > 1)+ & \Pr(R > 2)+ & \Pr(R > 3)+ & \Pr(R > 4) + \dots \\
& = & & \Pr(R = 1)+ & \Pr(R = 2)+ & \Pr(R = 3)+ & \Pr(R = 4) + \dots \\
& & & & \Pr(R = 2)+ & \Pr(R = 3)+ & \Pr(R = 4) + \dots \\
& & & & & \Pr(R = 3)+ & \Pr(R = 4) + \dots \\
& & & & & & \Pr(R = 4) + \dots \\
& = & \sum_{i=1}^{\infty} i \Pr(R = i) & & & & \\
& = & E[R] & & & &
\end{array}$$

■

We give an example to show the use of this shortcut.

Suppose we are given a system that might fail at any of the discrete times $1, 2, \ldots$.
Failure at each time step is independent, probability p.
Let R be the number of non-failing times that pass prior to the first failing time.
Then R is distributed according to the *negative binomial distribution*:

$$\Pr(R = k) = p(1-p)^k,$$

since this is the probability that the first k steps are non-failing ("successful") and the $(k+1)^{st}$ is a failure.

What is our sample space?
Finite sequences of successes followed by one failure,
together with one infinite sequence, of all successes.
So it's countable.

The probability of a sequence of k successes followed by 1 failure is as given above.
The probability of the sequence of all successes is 0.
(Possible event, but probability is 0 – that actually makes sense.)

The sum of all these probabilities is:

$$p + p(1-p) + p(1-p)^2 + \ldots = \frac{p}{1-(1-p)} = 1$$

as needed.

The random variable R associates k with the sequence of k successes and 1 failure.
(Associate some arbitrary default with the infinite sequence – it won't matter, for our purposes, because that has probability 0 anyway).

What is the expected value of R?
(This is usually called the *mean time to failure* (MTF).

Use the shortcut to calculate.

The event $R > i$ occurs when the first $i+1$ trials are all successes.
Thus,

$$\Pr(R > i) = (1-p)^{i+1}$$

Therefore, using the shortcut sum,

$$E[R] = (1-p) + (1-p)^2 + (1-p)^3 + \cdots = \frac{1-p}{1-(1-p)} = \frac{1-p}{p}$$

It follows that the expected number of successes before the first failure is $\frac{1-p}{p}$.
This means that the failure occurs at step $1 + \frac{1-p}{p} = \frac{1}{p}$.

Example 7 Consider a couple that wants to have a girl.
Assume girls and boys are equally likely.
What is the expected number of boys before the first girl?

Here, $p = 1/2$.
The girl is analogous to a system failure (not to be taken literally).
From the above analysis, the expected birth-number of the girl is:

$$E[R] = \frac{1}{p} = \frac{1}{1/2} = 2$$

Thus, you expect only one boy before the first girl.
So the average family size is 2!
This might seem counterintuitive, since you would expect the attempts to have a girl to lead to long sequences of boys;
however, these sequences are counterbalanced by the chance of having a girl on the first try.

3 Linearity of Expectation

3.1 The Results

It turns out that expectation is a "linear function":

Theorem 3.1 *For any random variables R, R_1, R_2 and any real constant a, the following hold:*

1. $E[aR] = aE[R]$
2. $E[R_1 + R_2] = E[R_1] + E[R_2]$
3. $E[R + a] = E[R] + a$

Note that we do not need to assume independence of the random variables.

Example 8 Roll two 6-sided dice.
Let R_1 and R_2 be the outcome of the 1st and 2nd die, respectively.
Also, let $R = R_1 + R_2$ be the sum.
Then

$$E[R] = E[R_1] + E[R_2] = 3.5 + 3.5 = 7$$

Proof. We first show $E[aR] = aE[R]$.
Can be done using either definition of expectation.
I'll use the first definition, since it's easier.

$$\begin{aligned} E[aR] &= \sum_{s \in S} aR(s) \Pr(s) \\ &= a \sum_{s \in S} R(s) \Pr(s) \\ &= aE[R] \end{aligned}$$

Using the second definition, it's a little harder:

$$\begin{aligned} E[aR] &= \sum_{r} r \Pr(aR = r) \\ &= \sum_{r} r \Pr(R = r/a) \\ &= \sum_{r'} ar' \Pr(R = r') \qquad \text{(substitute } r' = r/a\text{)} \\ &= a \sum_{r'} r' \Pr(R = r') \\ &= aE[R] \end{aligned}$$

We next show $E[R_1 + R_2] = E[R_1] + E[R_2]$:

$$\begin{aligned} E[R_1 + R_2] &= \sum_{s \in S} (R_1(s) + R_2(s)) \Pr(s) \\ &= \sum_{s \in S} R_1(s) \Pr(s) + \sum_{s \in S} R_2(s) \Pr(s) \\ &= E[R_1] + E[R_2] \end{aligned}$$

Finally, we show $E[R + a] = E[R] + a$.

$$\begin{aligned} E[R + a] &= \sum_{s \in S} (R(s) + a) \Pr(s) \\ &= \sum_{s \in S} R(s) \Pr(s) + \sum_{s \in S} a \Pr(s) \\ &= E[R] + \sum_{s \in S} a \Pr(s) \\ &= E[R] + a \sum_{s \in S} \Pr(s) \\ &= E[R] + a \end{aligned}$$

■

LTTR to show the last two parts using the second definition.

Corollary 3.2 $E[a_1R_1 + a_2R_2] = a_1E[R_1] + a_2E[R_2]$.

Proof.

$$\begin{aligned} E[a_1R_1 + a_2R_2] &= E[a_1R_1] + E[a_2R_2] \\ &= a_1E[R_1] + a_2E[R_2] \end{aligned}$$

■

Corollary 3.3 $E[R_1 + R_2 + R_3 + \cdots + R_N] = E[R_1] + E[R_2] + E[R_3] + \ldots E[R_N]$

Proof. By induction, based on Theorem 3.1, part 2. ■

3.2 Simple Examples

Example 9 (Hat Check Problem)

n men check their hats at a restaurant.
The perverse hat check person scrambles the hats randomly (random permutation).
What is the expected number of men to get their own hat?

Let the random variable R denote the number of men to get their own hat back.
$\Pr(R = k)$ is difficult to calculate directly.

However, we can get some help by expressing R in the form of a sum of auxiliary random variables.
Namely, define R_i as

$$R_i = \begin{cases} 1 & \text{if the ith man gets his hat} \\ 0 & \text{otherwise} \end{cases},$$

That is, the indicator variable for the event that man i gets his hat back.

Then clearly $R = R_1 + R_2 + R_3 + \cdots + R_n$. Obviously, the R_i are not independent.
For example, if $R_1 = R_2 = \ldots R_{n-1} = 1$, then $R_n = 1$.
Nevertheless, we can apply linearity of expectation.

Note, $E[R_i] = 1 \cdot \Pr(R_i = 1) + 0 \cdot \Pr(R_i = 0) = 1/n$.
This is because there is a 1 in n chance that man i gets his own hat.

$$E[R] = E[R_1] + E[R_2] + \cdots + E[R_n] = n(1/n) = 1$$

So, you expect one man to get his hat back.

Example 10 (Chinese appetizer problem)

n people at a circular table.
n different appetizers on a Lazy Susan.
The Lazy Susan is randomly rotated.
What is the expected number of people who get the same appetizer?

Answer: 1.
Same reasoning as above.
The expectation for each individual person is 1/n.
Therefore, the expectation for n people is 1.

Note that the probability space is very different from the hat check problem.
There are only n distinct outcomes, instead of the $n!$ in the hat check problem.
Either everyone gets the same appetizer back (probability $1/n$) or no one does.
Nevertheless, we can solve both problems easily by using linearity of expectation.
We didn't need to know much about the actual distribution.
Also, we don't need independence.

3.3 Combinations of Events

Can use linearity to calculate how many of an arbitrary set of events are likely to occur.
The events needn't be independent.

Theorem 3.4 *Given any collection of n events, $A_1, A_2, A_3, \ldots A_n$. Then the expected number of these events that occur is*

$$\sum_i Pr(A_i).$$

Just add up the probabilities!

Proof. Let R_i be the indicator variable for A_i.
That is,

$$R_i = \begin{cases} 1 & \text{if event } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Then the total number of the events that occur is:

$$R = R_1 + R_2 + \cdots + R_n$$

We want the expected value of R.

$$\begin{aligned} E[R] &= \sum_i E[R_i] \\ &= \sum_i \Pr(A_i) \end{aligned}$$

■

Example 11 Suppose I flip n coins, independently.
Let $R =$ total number of heads.
What is the expected number of heads, $E[R]$?

Can calculate directly:

$$E[R] = \sum_{i=0}^{n} i \Pr(R = i) = \sum_{i=0}^{n} i \binom{n}{i} 2^{-n}$$

Using the fact (LTTR) that

$$\sum_{i=0}^{n} i \binom{n}{i} = n2^{n-1}$$

we see that this is $n/2$.

Let's get a simpler proof using combinations of events.

Define A_i to be the event that the ith flip yields heads.
$R(s)$ is just the total number of events A_i containing the sample point s.

By the theorem above, $E[R] = \sum_i \Pr(A_i)$.
Since each $\Pr(A_i) = 1/2$, this yields $n/2$.

Can see this as a probabilistic proof of the combinatorial identity:

$$\sum_{i=0}^{n} i \binom{n}{i} 2^{-n} = n/2$$

The two sides compute the same expectation, so they must be equal.

Besides being easier, the expectation argument is also more powerful.
It doesn't assume independent distributions on the coins.

We can extend the preceding to show that if we conduct n independent trials, each succeeding with probability p, then the expected number of successes is np.
That is, the mean of the binomial distribution is np.
(Because the probability of k successes is given by the binomial term $\binom{n}{k}p^k(1-p)^{n-k}$.)

Example 12 Suppose I make one random choice, to either use an all heads coin or an all tails coin in my n coin flips.
Now there are only two sample points, one with all heads in the sequence and one with all tails.

But for each event A_i, the probability is still $1/2$.
So the theorem still applies, yielding $n/2$ as the expected number of heads.

Example 13 Consider a 6-sided die.
A_1 is the event that the value is odd.
A_2 is the event that the value is 1, 2 or 3
A_3 is the event that the value is 4, 5, or 6

Not independent.
In fact, the last two exclude each other.

But we can still calculate the expected number of these events that hold as $\Pr(A_1)+\Pr(A_2)+\Pr(A_3) = 1/2 + 1/2 + 1/2 = 3/2$.

4 Expectations of Products

Now consider the expectation of the *product* of two random variables.

We might hope for the identity:
$E[R_1R_2] = E[R_1] \cdot E[R_2]$

But this does not hold in general.
However, it does hold in the important special case when R_1 and R_2 are independent:

Theorem 4.1 *Let R_1 and R_2 be independent random variables.*
Then $E[R_1R_2] = E[R_1] \cdot E[R_2]$.

Proof. We have

$$\begin{aligned}
E[R_1R_2] &= \sum_r r \cdot Pr(R_1R_2 = r) \\
&= \sum_r r \sum_{r_1,r_2|r_1r_2=r} Pr(R_1 = r_1 \text{ and } R_2 = r_2) \\
&= \sum_r \sum_{r_1,r_2|r_1r_2=r} rPr(R_1 = r_1 \text{ and } R_2 = r_2) \\
&= \sum_r \sum_{r_1,r_2|r_1r_2=r} r_1r_2Pr(R_1 = r_1 \text{ and } R_2 = r_2) \\
&= \sum_{r_1,r_2} r_1r_2Pr(R_1 = r_1 \text{ and } R_2 = r_2) \\
&= \sum_{r_1,r_2} r_1r_2Pr(R_1 = r_1) \cdot Pr(R_2 = r_2).
\end{aligned}$$

The first equality arises by decomposing the probability of an event into the sum of the probabilities of disjoint events whose union is the original event.
The next 3 steps are just calculations.
The last equality follows from the independence of R_1 and R_2.

Now, we can factor this expression as

$$\left(\sum_{r_1} r_1 \cdot Pr(R_1 = r_1)\right)\left(\sum_{r_2} r_2 \cdot Pr(R_2 = r_2)\right).$$

which is equal to $E[R_1]E[R_2]$.
Therefore,

$$E[R_1R_2] = E[R_1] \cdot E[R_2].$$

■

Example 14 Rolling 2 fair, 6-sided dice independently.
Let R_1, R_2 be values of the first and second dice, respectively.
Since these are independent, we have that

$$E[R_1R_2] = E[R_1] \cdot E[R_2] = \frac{7}{2} \cdot \frac{7}{2} = \frac{49}{4}.$$

Now, what if the two dice are rigged so that they always give the same value as one another? In this case, R_1 and R_2 always assume the same values, and we have

$$E[R_1R_2] = \sum_{i=1}^{6} i^2 \cdot \frac{1}{6} = \frac{91}{6} \neq E[R_1]^2 = (\frac{7}{2})^2.$$

So this is a case of two dependent random variables for which the expectation of their product is not the product of their expectations.

Finally, by induction on n, it is easy to prove the following corollary of Theorem 4.1

Theorem 4.2 *If $R_1, \ldots, R_n$ are mutually independent, then*

$$E[R_1 R_2 \cdots R_n] = (E[R_1])(E[R_2]) \cdots (E[R_n]).$$

5 Conditional Expectation

Can also define a conditional version of expectation, conditioned on a particular event:

Definition 5.1 *The conditional expectation of R given an event A is defined as*

$$E[R \mid A] = \sum_{s \in A} R(s) \Pr(s \mid A)$$

.

Theorem 5.2

$$E[R \mid A] = \sum_{r} r \cdot Pr(R = r \mid A).$$

Proof. Analogous to the unconditional version of this statement, in Theorem 1.2. ∎

Example 15 Let R be the number of spots we see in a typical die roll. Then

$$E[R \mid R \leq 4] = 2.5$$

An important use of conditional expectation is in doing a "case analysis" to compute an expectation.
Depends on a theorem that breaks down the expectation:

Theorem 5.3 *Let $\{A_i\}$ be a set of disjoint events that exhaustively partition the sample space. Then for any random variable R,*

$$E[R] = \sum_{i} E[R \mid A_i] \Pr(A_i)$$

The proof of this theorem depends on the following basic conditional probability lemma:

Lemma 5.4 *If $\{A_i\}$ partition the sample space, then*

$$\Pr(B) = \sum_i \Pr(B|A_i)\Pr(A_i)$$

.

Proof. RHS is just $\sum_i \Pr(B \cap A)$, which adds up to $\Pr(B)$ by definition of a probability measure. ■ We are now ready to prove the theorem (5.3):

Proof.

$$\begin{aligned} E[R] &= \sum_r r \cdot \Pr(R = r) \\ &= \sum_r r \cdot \left(\sum_i \Pr(R = r \mid A_i)\Pr(A_i) \right) \\ &= \sum_i \left(\sum_r r \cdot \Pr(R = r \mid A_i) \right) \Pr(A_i) \\ &= \sum_i E[R \mid A_i]\Pr(A_i) \end{aligned}$$

The first equation follows from the lemma just above. ■

Example 16 (Assembly plants)

Suppose there are two computer assembly plants F and S.

Plant F is fast but sloppy:
They make 70% of the computers, but a computer from their plant has a 2% probability of breaking down on any given day of use.

Plant S is slow and careful:
They make 30% of the computers, but the breakdown probability is only 1%.

These computers get randomly mixed, so I buy one at random from the store.
What is the expected number of days my computer works?

Let D be the random variable giving the number of days my computer works.
Let I_S be the event that I bought the computer from S, and let I_F be that I bought it from F.

We decompose this problem using conditional expectations.
That is, we calculate separately the condition expectation of the number of working days for S and for F,
then combine using the conditional expectation theorem.

The expected number of working days for a computer from S is obtained from an MTF calculation, that is, a negative binomial distribution with failure probability $p = .01$.
Recall this is just

$$E[D \mid I_S] = \frac{1-p}{p} = 99$$

Similarly, the expected number of working days for a computer from F is

$$E[D \mid I_F] = \frac{1-.02}{.02} = 49$$

.
Now to calculate the expected number of working days for the computer when we don't know which factory it came from, we use the conditional expectation theorem:

$$E[D] = E[D|I_S]\Pr(I_S) + E[D|I_F]\Pr(I_F) = 99(.30) + 49(.70) = 64$$

6 Wald's Theorem

The final topic is a theorem that allows another kind of combination of experiments – one in which a probability distribution determines how many experiments are performed.
Begin with an example:

Example 17 Suppose I flip 3 standard coins and then roll a die for each coin that comes up heads.
How many total spots do I expect to see?
Well, if H is a random variable denoting the number of heads I see, and $D_1, D_2, \ldots$ are random variables for die rolls, then the random variable describing the number of spots can be written as

$$N = \sum_{i=1}^{H} D_i.$$

This notation is a bit strange, because the upper summation index is a random variable.
It makes sense, though – for each element of the sample space, I calculate the value of the random variable N by summing up as indicated, based on the values of the other random variables at the same point.

How can I compute the expectation? One way is by conditional expectations, conditioning on the number of coins that come up heads:

$$\begin{aligned}
E[N] &= \sum_{k=0}^{3} E[N \mid H = k] \Pr(H = k) \\
&= \sum_{k=0}^{3} E[\sum_{i=1}^{k} D_i] \Pr(H = k) \\
&= \frac{1}{8}0 + \frac{3}{8}(E[D_1]) + \frac{3}{8}(E[D_1] + E[D_2]) \\
&\quad + \frac{1}{8}(E[D_1] + E[D_2] + E[D_3]) \\
&= \frac{3}{8}3.5 + \frac{3}{8}7 + \frac{1}{8}10.5 \\
&= \frac{21}{4}
\end{aligned}$$

The first equality uses the conditional expectation theorem.
The second specializes the sum to the case where the upper bound is k.

Is there an easier way to figure this out?
Notice that $E[H] = \frac{7}{2}$, and $E[D] = \frac{3}{2}$, so the expectation turns out to be $E[H] \cdot E[D]$.
But why?

Doesn't follow from linearity of expectation (rule for expectation sums), because we don't know how many copies of $E[D]$ to add up.

Doesn't follow from the product rule.
That would apply if I multiplied the number of heads by the outcome of rolling one die.

But it is true.
Need a new result called *Wald's Theorem*:

Theorem 6.1 *Let Q and R be independent random variables, where Q takes on natural number values.*
Let T be the sum of Q samples of R, that is:

$$T = \sum_{i=1}^{Q} R_i.$$

Then $E[T] = E[Q] \cdot E[R]$.

Proof. Picture the random experiment as follows.
First we generate an infinite number of samples of R, call them $R_1, R_2, \ldots$; then we use Q

to decide how many of these copies to add up in computing T.
This experiment works on a different sample space from the original, but gives the same distribution.
Then:

$$
\begin{aligned}
E[T] &= \sum_k E[T \mid Q = k] \cdot \Pr(Q = k) \\
&= \sum_k E[\sum_{i=1}^{k} R_i] \cdot \Pr(Q = k) \qquad (*) \\
&= \sum_k \left(\sum_{i=1}^{k} E[R_i] \right) \cdot \Pr(Q = k) \\
&= \sum_k \left(\sum_{i=1}^{k} E[R] \right) \cdot \Pr(Q = k) \\
&= \sum_k (kE[R]) \cdot \Pr(Q = k) \\
&= E[R] \sum_k k \cdot \Pr(Q = k) \\
&= E[R]E[Q]
\end{aligned}
$$

■ One of the lines in this proof – needs a little closer attention.
Namely, look at the line labeled (*), where we say that

$$E[\sum_{i=1}^{Q} R_i \mid Q = k] = E[\sum_{i=1}^{k} R_i].$$

Let's look at this more closely.
In fact, this is the place where we use the independence of the random variables Q and R.

A bit more carefully:
The LHS can be rewritten as

$$E[\sum_{i=1}^{k} R_i \mid Q = k]$$

We want to show this is equal to the RHS.
Now, independence of R and Q can be used to show that, for any r:

$$\Pr(\sum_{i=1}^{k} R_i = r | Q = k) = \Pr(\sum_{i=1}^{k} R_i = r)$$

This can be used to show that the conditional expectation is the same as the unconditional expectation:

$$E[\sum_{i=1}^{k} R_i \mid Q = k] = E[\sum_{i=1}^{k} R_i]$$

as needed.

If Q and R are dependent, the theorem might not hold.

We close this section with an application of Wald's Theorem to *branching processes*. Branching processes are used to model population growth, nuclear chain reactions, mutations, and contagious diseases.

Example 18 (Contagious disease)

Suppose someone is infected with a new, incurable disease.
Every day, he encounters n healthy people and infects each of them with probability p.
What is the expected number of infected people after k days?

Let S_k be a random variable giving the number infected after k days.
We can write a recurrence for S_{k+1} in terms of S_k.

Let the sick people after day k be numbered $1, \ldots, S_k$.
Let H_i be a random variable giving the number of healthy people infected by sick person i in one day.
Then H_i has a binomial distribution with parameters n and p.
Then:

$$S_{k+1} = S_k + \sum_{i=1}^{S_k} H_i$$

It follows from linearity of expectation that

$$E[S_{+1}] = E[S_k] + E[\sum_{i=1}^{S_k} H_i].$$

Then by Wald's theorem,

$$E[S_{k+1}] = E[S_k] + E[S_k] \cdot E[H]$$

where H is a variable with the binomial distribution with parameters n and p.
Thus, $E[H] = np$.

So

$$E[S_{k+1}] = E[S_k] + E[S_k] \cdot np = (1 + np)E[S_k]$$

Also, $E[S_0] = 1$.
So we have a recurrence involving these expectations.
Writing e_i for $E[S_i]$, we have:

$$e_{k+1} = (1 + np)e_k$$

and

$$e_0 = 1$$

It follows that $e_k = (1 + np)^k$.

Eventually, we will have more sick people than people. What erroneous assumption is in our model? That each sick person meets a different collection of 10 healthy people. Once there are enough sick people, it will become less likely for a sick person to meet so many healthy people. Also, many sick people will encounter the same healthy person. But our "approximation" will be valid as long as the sick population is not too big. Now suppose any sick person recovers from the disease each day with probability r. Let B_i be the indicator variable for whether the i^{th} sick person gets better. Then $E[B_i] = r$. Then

$$S_{k+1} = S_k + \sum_{i=1}^{S_k} (H_i - B_i)$$

It follows that

$$E[S_{k+1}] = E[S_k] + E[S_k] \cdot E[H] - E[S_k]E[B_i]$$

and so

$$\begin{aligned} e_{k+1} &= e_k(1 + np - r) \\ &= (1 + np - r)^k \end{aligned}$$

Therefore, the progress of the disease is determined entirely by the parameter $(1+np-r)$. If this exceeds 1, it looks like we have an epidemic on our hands (but expectation isn't telling the whole story: there's a chance the first person gets better without making anyone sick). If less than 1, it looks pretty innocuous. In fact, next time we will see that if less than 1, the disease almost certainly dies out after a certain number of days.

7 Extra Problems

7.1 Coupon Collecting

Suppose there are n different baseball cards. I buy them one at a time and stop when I get one of each. How long do I expect to continue?

Let W_k be the time I wait until I get the k^{th} card, and let $X_k = W_{k+1} - W_k$. Then clearly,

$$W_n = \sum X_k$$

and therefore the waiting time is

$$\sum E[X_k].$$

To determine $E[X_k]$, consider the "experiment" being performed during this k^{th} time interval. I have k cards, and I want to find the $(k+1)^{st}$. Each time I buy a card, my probability of getting a new card is $p = (n-k)/n$. Each attempt is independent, so I have a sequence of Bernoulli Trials. By the previous example on negative binomial distributions ("mean time to failure," only now it is success) my expected waiting time is

$$\begin{aligned}\frac{1}{1-p} &= \frac{1}{1-k/n}\\ &= \frac{n}{n-k}\end{aligned}$$

It follows that the expected waiting time to get all the cards is

$$\begin{aligned}\sum_{k=1}^{n-1} \frac{n}{n-k} &= n \sum_{k=1}^{n-1} \frac{1}{k}\\ &\approx n \ln n\end{aligned}$$

where the last approximation comes from our recognition of the *Harmonic series.*

(So to get all 1000 distinct baseball cards, we expect to buy about 7000).

We were actually a bit sloppy in our analysis of the coupon collecting problem. When we talked about the experiment in the k^{th} time interval, we referred to the trials being done in this interval. But which trials in our sequence of purchases make up the k^{th} experiment? The ones after we got k cards. It follows that we are actually conditioning on a particular value of W_k. In other words, we are computing $E[X_k]$ using conditional expectations! What saves us is that $E[X_k]$ turns out to be independent of W_k:

$$\begin{aligned}E[X_k] &= \sum_z E[X_k \mid W_k = z] \cdot \Pr[W_k = z]\\ &= \sum_z \frac{n}{n-k} \Pr[W_k = z]\\ &= \frac{n}{n-k} \sum_z \Pr[W_k = z]\\ &= \frac{n}{n-k}\end{aligned}$$

7.2 Carnival Game

Let's consider the following carnival game.

- Player picks $A \in [1, 6]$.
- Roll 3 six-sided dice. (Assume fair and mutually independent).

- If the player matches any of the dice, he wins \$1. Otherwise, he loses \$1.

Let R denote the player's winning. Then, $E[R] = 1 \cdot \Pr(win) - 1 \cdot \Pr(lose)$.

One initial approach may be as follows. The probability of die 1 matching is 1/6. Similary, the probability of die 2 and die 3 matching is 1/6. Therefore, the probability of getting one match is 3/6. Thus, the game is fair, and $E[R] = 0$.

This analysis is wrong. $Pr(win) \neq 1/2$. What was the problem? Well, we imagined an indicator variable for each die: I_k says die k is a winner. But then we claimed that $I = I_1 + I_2 + I_3$ is an indicator variable for whether the player wins. This is wrong, since I isn't an indicator for anything: its value may be 2 or 3. $\sum I_k$ actually measures the *expected number of matching dice*, a different quantity. Indicator variables need to be $0, 1$.

The correct method is to apply inclusion/exclusion.

$$\begin{aligned}
Pr(R=1) &= \Pr(D_1 = A \vee D_2 = A \vee D_3 = A) \\
&= Pr(D_1 = A) + \Pr(D_2 = A) + \Pr(D_3 = A) - \Pr(D_1 = A \wedge D_2 = A) \\
&\quad - Pr(D_1 = A \wedge D_3 = A) - \Pr(D_2 = A \wedge D_3 = A) + \Pr(D_1 = A \wedge D_2 = A \wedge D_3 = A) \\
&= 1/6 + 1/6 + 1/6 - 1/36 - 1/36 - 1/36 + 1/216 \\
&= .421
\end{aligned}$$

Another way to calculate this probability is to find the probability of a loss. $Pr(lose) = (5/6)^3 = .579 \implies \Pr(win) = 1 - .579 = .421$.

Thus, the game is an expected loss. Another variant of the game is to pay \$2 if two dice match. Again, this is a losing game.

7.3 A two-player game

Here's a game you don't want to play unless you know what you're doing. Two players, A and B, agree on a probability distribution $f(x)$ for a random variable X over $\mathbb{N}$. That is, $f(x) = Pr(X = x)$.

Now player A picks a number a at random according to f. Then player B independently picks a number b according to f; if $b \leq a$, then B repeats this until she gets a number $b > a$. Each time through the loop, B pays A \$1.

What is B's expected loss? The answer is ∞.

To get some intuition for this, consider first the finite case: $f(x)$ is non-zero only for a finite set of numbers. Let t be the largest such number. Then with positive probability, A will pick t, in which case B can never pick a bigger number and so her loss will be infinite.

More precisely, if R is a random variable denoting the number of times B pays \$1, then we have

$$ER = \sum_{x \in \mathbb{N}} Pr(R > x).$$

But $Pr(R > x) \geq f(t)$ for all $x \geq t$, so this sum is infinite.

Now let's consider the case of an arbitrary distribution function f. Here, there may be no "biggest number" t, so we have to use a different argument.

Consider the game as follows: A draws the first number, and then B draws infinitely many more numbers, paying \$1 until a number comes up that's bigger than the first number. So B will pay at least \$m if among the first m numbers, the first is biggest.

For a fixed m, let A denote this event: that the first of m numbers drawn according to f is the biggest. We claim that $Pr(A) = \frac{1}{m}$, regardless of f. Intuitively, this is because only the order in which the numbers arrive matters. We can prove this as follows:

The sample space is the set S of all m-vectors of numbers. Let's say that two such vectors are equivalent if one is a permutation of the other. This gives a partition of S into equivalence classes: $S = \cup_i M_i$ and $\sum_i Pr(M_i) = 1$.

Now for each i, we have $Pr(A \mid M_i) \geq \frac{1}{m}$, since one we've fixed the m numbers, all orders are equally likely. (We don't necessarily have equality since numbers can be repeated.) Now we have

$$\begin{aligned} Pr(A) &= \sum_i Pr(M_i) \cdot Pr(A \mid M_i) \\ &\geq \frac{1}{m} \sum_i Pr(M_i) \\ &= \frac{1}{m}. \end{aligned}$$

Finally, we can compute the expected loss:

$$ER = \sum_{m \in \mathbb{N}} Pr(R > m) \geq \sum_{m \in \mathbb{N}} \frac{1}{m} = \infty.$$

7.4 Another two-player game

You remember Monty Hall? Well, not satisfied with the money *Let's Make a Deal* is making in syndication, he's turned to other sorts of games. Here's one:

There are two envelopes, one containing twice as much money as the other. He gives you one, but he does not allow you to open it. He then asks whether you want to keep the envelopes, or switch. Should you?

Here's some incorrect reasoning: lets say you have $\$x$ in your current envelope; so with probability $\frac{1}{2}$ the other envelope has $\$(2x)$, and with probability $\frac{1}{2}$ the other envelope has $\$x/2$. So your expected gain on swapping is $x/2 - x/4 = x/4$ — that is, you should always swap!

But this argument is incorrect, and the reason is that the probability is not 50-50 that the other envelope has either twice or half as much. For example, suppose I randomly put \$1 and \$2 in two envelopes. You open one and it contains \$2. What is the probability that the other envelope contains \$4? It's zero.

A correct argument, which gives the intuitively reasonable answer that it makes no difference whether you switch, is the following. One envelope has $\$y$ and the other has $\$2y$. So if you open an envelope and see $\$x$, then either $x = y$ with probability $\frac{1}{2}$, or $x = 2y$ with probability $\frac{1}{2}$. So in the first case your gain on switching is x, and in the second your gain is $-x$. Thus your expected gain is zero.

Now Monty allows you to open the envelope in hand, revealing $\$x$. Again he asks whether you want to keep the money, or switch. Given this extra information - the actual number x - should you switch or not. Intuitively, it would seem that knowing the number x does not give you any edge and so it shouldn't matter whether you switch or not. But, in actuality there is a strategy which allows you to pick the larger of the two sums with probability better than $\frac{1}{2}$. The strategy is as follows - choose any distribution which has continuous support over the reals, like the normal distribution, and pick a number r from it, if $r > x$ then switch, otherwise stay. Note that whenever r is smaller than the lesser of the two sums or bigger than the greater of the two then we end up with the larger sum with probability $\frac{1}{2}$, but whenever r lies in between the two sums then we always end up with the larger sum. Thus we do better than $\frac{1}{2}$, since there is a non-zero probability of r landing exactly between the two sums. Note, however, that if we use the normal distribution to pick r then if the sums in question grow very large our probability of winning the larger amount goes to $\frac{1}{2}$. In fact it can be shown that there is no distribution that can guarantee us a winning probability greater than $\frac{1}{2}$ by any fixed constant.

Of course, all this assumes that Monty offers the switch regardless of how much money he sees you holding. Where do we use this fact?