75 or more heads is the same as the probability of flipping 25 or fewer tails. By the above analysis, this is also extremely small. In ch 16

---

chapter 17 Expectation

~~16.3 Average & Expected Value~~ 17.1 Definitions and Examples

The *expectation* of a random variable is its average value, where each value is weighted according to the probability that it comes up. The expectation is also called the *expected value* or the *mean* of the random variable.

For example, suppose we select a student uniformly at random from the class, and let R be the student's quiz score. Then $\mathrm{E}[R]$ is just the class average —the first thing everyone wants to know after getting their test back! For similar reasons, the first thing you usually want to know about a random variable is its expected value.

— INSERT A goes here.

INSERT A

The expectation or expected value of a random variable is a single number that tells you a lot about the behavior of the variable. Roughly, the expectation is the average value of the random variable where each value is weighted according to its probability. Formally, the expected value (also known as the ~~XXX~~ average or mean) of a random variable is defined as follows.

Definition A1: If R is a random variable defined on a sample space S, then

$$\mathrm{Ex}[R] ::= \sum_{w \in S} R(w) \Pr(w). \quad (16.5)$$

For example, suppose S is the set of students in a class, and we ~~XXX~~ select a student uniformly at random. Let

R be the selected student's exam score.
Then Ex[R] is just the class average
— the first thing everyone wants to know
after getting their test back! For
similar reasons, the first thing
you usually want to know about a random
variable is its expected value.

Let's work through some examples.

17.1.1 The Expected Value of a Uniform Random Variable

roll a fair 6-sided die.

Let R be the value that comes up when you
Then the expected value of R is

$$E[R] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \ldots$$

Definition 16.3.1. If R is a random variable, then

$$\mathrm{E}[R] ::= \sum_{x \in \mathrm{range}(R)} x \cdot \Pr\{R = x\} \tag{16.4}$$

$$= \sum_{x \in \mathrm{range}(R)} x \cdot \mathrm{PDF}_R(x).$$

Let's work through an example. Let R be the number that comes up on a fair, six-sided die. Then by (16.4), the expected value of R is:

$$\mathrm{E}[R] = \sum_{k=1}^{6} k \cdot \frac{1}{6}$$

$$\mathrm{Ex}[R] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= \frac{7}{2}.$$

This calculation shows that the name "expected value" is a little misleading; the random variable might *never* actually take on that value. You don't ever expect to roll a $3\frac{1}{2}$ on an ordinary die!

There is an even simpler formula for expectation:

— INSERT B goes here —

— INSERT C goes here —

(text on pp 1057-1058)

INSERT D goes here

INSERT B

~~Note that~~ Also note that the mean of a random variable is not the same as the median. The median is the midpoint of a distribution.

Definition A2: If R is a random variable, then the median[1] of R is the value $x \in \text{Range}(R)$ such that

$$\Pr[R \leq x] \leq \frac{1}{2} \text{ and}$$
$$\Pr[R > x] < \frac{1}{2}.$$

~~In other words, the median is the value x such that the random variable is at most x at most half of the time~~

For example, for a single roll of a fair 6-sided die, the median is 4. ~~This is the. Since the expectation is much more~~

[1] Some texts define the median to be the value of $x \in \text{Range}(R)$ for which $\Pr[R \leq x] < \frac{1}{2}$ and $\Pr[R > x] \leq \frac{1}{2}$. The difference in definitions is minor and not important.

In this text, we will not devote much attention on the median. Rather, we will focus on the expected value, which is much more interesting and useful.

Rolling a 6-sided die provides an example of a uniform random variable. In general, if R_n is a random variable with a uniform distribution on $\{1, 2, \ldots, n\}$, then

$$\begin{aligned} Ex[R_n] &= \sum_{i=1}^{n} \frac{i}{n} \\ &= \frac{n(n+1)}{2n} \\ &= \frac{n+1}{2}. \end{aligned}$$

17.1.3 Alternate Definitions

There are several equivalent ways to define expectation.

Theorem 16.3.2. *If R is a random variable defined on a sample space S, then*

David: we'll need a macro for Ex

$$\mathrm{E}[R] = \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x] \qquad (16.4)$$

The proof of Theorem 16.3.2, like many of the elementary proofs about expectation in this chapter, follows by judicious regrouping of terms in the defining sum (16.4):

Proof.

$$\begin{aligned}
\mathrm{E}[R] &::= \sum_{x \in \text{range}(R)} x \cdot \Pr\{R = x\} && \text{(Def 16.3.1 of expectation)}\\
&= \sum_{x \in \text{range}(R)} x \left(\sum_{\omega \in [R=x]} \Pr\{\omega\} \right) && \text{(def of } \Pr\{R = x\}\text{)}\\
&= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R=x]} x \Pr\{\omega\} && \text{(distributing } x \text{ over the inner sum)}\\
&= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R=x]} R(\omega) \Pr\{\omega\} && \text{(def of the event } [R = x]\text{)}\\
\mathrm{E}[R] &::= \sum_{\omega \in S} R(\omega) \Pr\{\omega\}
\end{aligned}$$

DAVID: reverse the order of these — go bottom/up.

The first equality follows because the events $[R = x]$ for $x \in \text{range}(R)$ partition the

sample space S, so summing over the outcomes in $[R = x]$ for $x \in \text{range}(R)$ is the same as summing over S. ■

In general, the defining sum (16.4) is better for calculating expected values and has the advantage that it does not depend on the sample space, but only on the density function of the random variable. On the other hand, the simpler sum over all outcomes (16.5) is sometimes easier to use in proofs about expectation.

16.3.1 Expected Value of an Indicator Variable

The expected value of an indicator random variable for an event is just the probability of that event.

Lemma 16.3.3. *If I_A is the indicator random variable for event A, then*

$$\mathrm{E}[I_A] = \Pr\{A\}.$$

INSERT D

range of the random variable is $\mathbb{N}$, as we will see from the following corollaries.

Corollary A4: If the range of a random variable R is $\mathbb{N}$, then

$$Ex[R] = \sum_{i=1}^{\infty} i \Pr[R=i] \quad \&$$

$$= \sum_{i=0}^{\infty} \Pr[R>i]. \quad (\text{eqn A6})$$

Proof: The first equality follows directly from Theorem 16.3.2 and the fact that range $(R) = \mathbb{N}$. The second equality follows from the following analysis:

$$\Pr[R>0] = \Pr[R=1] + \Pr[R=2] + \Pr[R=3] + \cdots$$

$$\Pr[R>1] = \Pr[R=2] + \Pr[R=3] + \cdots$$

$$\Pr[R>2] = \Pr[R=3] + \cdots$$

$$\vdots$$

$$\sum_{i=0}^{\infty} \Pr[R>i] = 1 \cdot \Pr[R=1] + 2\Pr[R=2] + 3\Pr[R=3] + \cdots$$

$$= \sum_{i=1}^{\infty} i \Pr[R=i]. \quad \square$$

end of insert C

Proof.

$$\begin{aligned}
\mathrm{E}[I_A] &= 1 \cdot \Pr\{I_A = 1\} + 0 \cdot \Pr\{I_A = 0\} \\
&= \Pr\{I_A = 1\} \\
&= \Pr\{A\}. && \text{(def of } I_A\text{)}
\end{aligned}$$

■

For example, if A is the event that a coin with bias p comes up heads, then $\mathrm{E}[I_A] = \Pr\{I_A = 1\} = p$.

~~16.3.2~~ 17.1.7 Conditional Expectation

this is INSERT X (goes to E-10)

Just like event probabilities, expectations can be conditioned on some event. Given a random variable R, the expected value of R conditioned on an event A is the (probability-weighted) average value of R over outcomes in A. More formally:

Definition 16.3.4. The *conditional expectation*, $\mathrm{E}[R \mid A]$, of a random variable, R, given

event A is:

$$\mathrm{E}[R \mid A] ::= \sum_{r \in \mathrm{range}(R)} r \cdot \Pr\{R = r \mid A\}. \tag{16.6}$$

In other words, it is the average value of the variable R when values are weighted by their conditional probabilities given A.

For example, we can compute the expected value of a roll of a fair die, *given*, for example, that the number rolled is at least 4. We do this by letting R be the outcome of a roll of the die. Then by equation (16.6),

$$\mathrm{E}[R \mid R \geq 4] = \sum_{i=1}^{6} i \cdot \Pr\{R = i \mid R \geq 4\} = 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot \tfrac{1}{3} + 5 \cdot \tfrac{1}{3} + 6 \cdot \tfrac{1}{3} = 5.$$

The power of conditional expectation is that it lets us divide complicated expectation calculations into simpler cases. We can find the desired expectation by calculating the conditional expectation in each simple case and averaging them, weighing each case by its probability.

Insert 4 cont

For example, suppose that 49.8% of the people in the world are male and the rest female —which is more or less true. Also suppose the expected height of a randomly chosen male is 5′ 11″, while the expected height of a randomly chosen female is 5′ 5″. What is the expected height of a randomly chosen individual? We can calculate this by averaging the heights of men and women. Namely, let H be the height (in feet) of a randomly chosen person, and let M be the event that the person is male and F the event that the person is female. ~~We have~~ Then

$$\begin{aligned} \mathrm{E}[H] &= \mathrm{E}[H \mid M]\Pr\{M\} + \mathrm{E}[H \mid F]\Pr\{F\} \\ &= (5 + 11/12)\cdot 0.498 + (5 + 5/12)\cdot 0.502 \\ &= 5.665 \end{aligned}$$

which is a little less ~~that~~ than 5′ 8″.

This method is justified by the

~~The~~ Law of *Total Expectation* ~~justifies this method~~.

Theorem 16.3.5. *Let $A_1, A_2, \ldots$ be a partition of the sample space. Then*

Rule (Law of Total Expectation).

$$\mathrm{E}[R] = \sum_i \mathrm{E}[R \mid A_i] \Pr\{A_i\}.$$

Proof.

$$\begin{aligned}
\mathrm{E}[R] &= \sum_{r \in \mathrm{range}(R)} r \cdot \Pr\{R = r\} && \text{(Def 16.3.1 of expectation)} \\
&= \sum_r r \cdot \sum_i \Pr\{R = r \mid A_i\} \Pr\{A_i\} && \text{(Law of Total Probability)} \\
&= \sum_r \sum_i r \cdot \Pr\{R = r \mid A_i\} \Pr\{A_i\} && \text{(distribute constant } r\text{)} \\
&= \sum_i \sum_r r \cdot \Pr\{R = r \mid A_i\} \Pr\{A_i\} && \text{(exchange order of summation)} \\
&= \sum_i \Pr\{A_i\} \sum_r r \cdot \Pr\{R = r \mid A_i\} && \text{(factor constant } \Pr\{A_i\}\text{)} \\
&= \sum_i \Pr\{A_i\} \mathrm{E}[R \mid A_i]. && \text{(Def 16.3.4 of cond. expectation)}
\end{aligned}$$

■

The mean time to failure is a critical parameter in the design of most any system. For example, suppose that a

17.1.4

16.3.3 Mean Time to Failure

A computer program crashes at the end of each hour of use with probability p, if it has not crashed already. What is the expected time until the program crashes?

If we let C be the number of hours until the crash, then the answer to our problem is $\mathrm{E}[C]$. Now the probability that, for $i > 0$, the first crash occurs in the ith hour is the probability that it does not crash in each of the first $i-1$ hours and it does crash in the ith.

$\mathrm{Ex}[C]$. C is a random variable with values in $\mathbb{N}$ and so we can use Corollary A4 to determine that

$$\mathrm{Ex}[C] = \sum_{i=0}^{\infty} \Pr[C > i] \quad \text{eqn (16.7)}$$

ith hour, which is $(1-p)^{i-1}p$. So from formula (16.4) for expectation, we have

$$\begin{aligned}
\mathrm{E}[C] &= \sum_{i \in \mathbb{N}} i \cdot \Pr\{R = i\} \\
&= \sum_{i \in \mathbb{N}^+} i(1-p)^{i-1}p \\
&= p \sum_{i \in \mathbb{N}^+} i(1-p)^{i-1} \\
&= p \frac{1}{(1-(1-p))^2} \qquad \text{(by (13.1))} \\
&= \frac{1}{p}
\end{aligned}$$

A simple alternative derivation that does not depend on the formula (13.1) (which you remembered, right?) is based on conditional expectation. Given that the computer crashes in the first hour, the expected number of hours to the first crash is obviously 1! On the other hand, given that the computer does not crash in the first hour, then the expected total number of hours till the first crash is the expectation of one plus the

number of additional hours to the first crash. So,

$$\mathrm{E}[C] = p \cdot 1 + (1-p)\,\mathrm{E}[C+1] = p + \mathrm{E}[C] - p\,\mathrm{E}[C] + 1 - p,$$

from which we immediately calculate that $\mathrm{E}[C] = 1/p$.

EDITING NOTE: There is a useful trick for calculating expectations of nonegative integer valued variables:

Lemma 16.3.6. *If R is a nonegative integer-valued random variable, then:*

$$\mathrm{E}[R] = \sum_{i \in \mathbb{N}} \Pr\{R > i\} \tag{16.7}$$

Proof. Consider the sum:

$$\begin{array}{lllll} \Pr\{R=1\} & + \Pr\{R=2\} & + \Pr\{R=3\} & + \cdots \\ & + \Pr\{R=2\} & + \Pr\{R=3\} & + \cdots \\ & & + \Pr\{R=3\} & + \cdots \\ & & & + \cdots \end{array}$$

The successive columns sum to $1 \cdot \Pr\{R = 1\}$, $2 \cdot \Pr\{R = 2\}$, $3 \cdot \Pr\{R = 3\}$, ... Thus, the whole sum is equal to:

$$\sum_{i \in \mathbb{N}} i \cdot \Pr\{R = i\}$$

which equals $\mathrm{E}[R]$ by (16.4). On the other hand, the successive rows sum to $\Pr\{R > 0\}$, $\Pr\{R > 1\}$, $\Pr\{R > 2\}$, ... Thus, the whole sum is also equal to:

$$\sum_{i \in \mathbb{N}} \Pr\{R > i\},$$

which therefore must equal $\mathrm{E}[R]$ as well. ■

Now $\Pr\{C > i\}$ is easy to evaluate: a crash happens later than the ith hour iff the system did not crash during the first i hours, which happens with probability $(1 - p)^i$.

Plugging this into (16.7) gives:

$$\begin{aligned} \mathbf{E}[C] &= \sum_{i \in \mathbb{N}} (1-p)^i \\ &= \frac{1}{1-(1-p)} \qquad \text{(sum of geometric series)} \\ &= \frac{1}{p} \end{aligned}$$

The general principle here is well-worth remembering: if a system fails at each time step with probability p, then the expected number of steps up to the first failure is $1/p$.

For example, if there is a 1% chance that the program crashes at the end of each hour, then the expected time until the program crashes is $1/0.01 = 100$ hours.

As a related example, suppose a couple really wants to have a baby girl. For simplicity, assume that there is a 50% chance that each child they have is a girl, and that the genders of their children are mutually independent. If the couple insists on having children until

they get a girl, then how many baby boys should they expect first?

This is really a variant of the previous problem. The question, "How many hours until the program crashes?" is mathematically the same as the question, "How many children must the couple have until they get a girl?" In this case, a crash corresponds to having a girl, so we should set $p = 1/2$. By the preceding analysis, the couple should expect a baby girl after having $1/p = 2$ children. Since the last of these will be the girl, they should expect just one boy.

Something to think about: If every couple follows the strategy of having children until they get a girl, what will eventually happen to the fraction of girls born in this world?

INSERT E goes here

INSERT E

17.1.5 Dealing with Infinity

The analysis of the mean time to failure was easy enough. But if you think about it further, you might start to wonder about the case when the computer program never fails. For example, what if the program runs forever? How do we handle outcomes with an infinite value?

These are good questions and we wonder about them too. Indeed, mathematicians have gone to a lot of work to reason about sample spaces with an infinite number of outcomes or outcomes with infinite value. To keep matters simple in this text, we will follow the common

convention of ignoring the contributions of outcomes that have probability zero when computing expected values. This means that we can safely ignore the "never-fail" outcome, because it has probability

$$\lim_{n\to\infty} (1-p)^n = 0.$$

In general, when we are computing expectations for infinite sample spaces, we will generally focus our attention on a subset of outcomes that occur with collective probability one. For the most part, this will allow us to ignore[1] the "infinite" outcomes because they will typically happen with probability zero.

1 If this still bothers you, you might consider taking a course on measure theory.

This assumption does not mean that the expected value of a random variable is always finite, however. Indeed, there are many examples where the expected value is infinite. And where infinity raises its ugly head, trouble is sure to follow. Let's see an example.

17.1.6 Pitfall: Computing Expectations by Sampling

Suppose that

E-4

So we expect to have 1 boy! Most people guess more.

Questions?

Q: So if you use this algorithm, what is the expected number of children you have?

A: 2. One boy & then the one girl.

Q: How about this question. Suppose you keep having babies until you get at least one boy & one girl. How many children do you expect to have?

A: 3 1st kid, then exp number kids to get baby of other sex.

Any questions about expectation?

Next we're going to look at a nasty example that you see a lot in experimental work.

Suppose you are trying to estimate a parameter such as the average delay across a communications channel. So you set up an experiment to measure how long it takes to send a test packet from one end to the other & you run the experiment 100 times & record latency each time.

You assume that there is some probability distribution function for the probability that a packet has a certain amount of delay.

You record the latency, rounded to the nearest millisecond for each of the hundred experiments, and then compute the average

of the 100 measurements. Suppose that this average is 8.3 ms.

Because you are careful, you repeat the entire process twice more, and get an averages of 7.8 ms and 7.9 ms. You conclude that the average latency across the channel is

$$\frac{7.8 + 7.9 + 8.3}{3} = 8 \text{ ms.}$$

You might be right but you might also be horribly wrong. In fact, the expected latency might well be infinite. Here's how.

Let D be a random variable that denotes the time it takes for the packet to cross the channel. Suppose that

$$f(x) = \Pr(D = i) = \frac{1}{i} - \frac{1}{i+1}$$

$$= \frac{1}{i(i+1)}$$

$$\Pr(D=i) = \begin{cases} 0 & \text{for } i=0 \\ \frac{1}{i} - \frac{1}{i+1} & \text{for } i \in \mathbb{N}^+ . \end{cases} \qquad \text{(eqn T1)}$$

It is easy to check that

$$\sum_{i=0}^{\infty} \Pr(D=i) = \left(1 - \tfrac{1}{2}\right) + \left(\tfrac{1}{2} - \tfrac{1}{3}\right) + \left(\tfrac{1}{3} - \tfrac{1}{4}\right) + \cdots = 1$$

and so D is, in fact, a random variable.

From Equation T1, it we might expect that D is likely to be small. Indeed, D=1 with probability 1/2, D=2 with probability 1/6, and so forth. So if we took 100 samples of D, about 50 would be 1ms, about 16 would be 2ms, and very few would be large. In summary, it might well be the case that the

average of the 100 measurements would be under 10ms, just as in our example.

~~But reasoning this way is not very precise and it can lead~~

This sort of reasoning and the ~~exact experimental computation~~ calculation of expected values by averaging ~~A~~ experimental values is very common in practice. It ~~a~~ can easily lead to incorrect conclusions, however. For example, using Corollary A4, we can quickly (and accurately) determine that

$$\begin{aligned} Ex[D] &= \sum_{i=1}^{\infty} i \Pr(D=i) \\ &= \sum_{i=1}^{\infty} i \left(\frac{1}{i} - \frac{1}{i+1}\right) \\ &= \sum_{i=1}^{\infty} i \left(\frac{1}{i(i+1)}\right) \\ &= \sum_{i=1}^{\infty} \frac{1}{i+1} \\ &= \infty . \end{aligned}$$

In other words, the expected time to cross the communication channel is infinite! This result is a far cry from the 10ms that we calculated. what went wrong?

It is true that most of the time, the value of D will be small. But sometimes D will be very large and this ~~leads to~~ happens with sufficient probability that the expected value of D ~~is~~ is unbounded. In fact, if you ~~repeat~~ keep repeating the experiment, you are likely to see ~~1000 times, the average observed latency is likely to be a little longer~~ some outcomes and averages that are much larger than 10 ms. In practice, such "outliers" are sometimes discarded, which masks the true behavior of D.

In general, the best way to compute an ~~expected value~~ expected value in practice is to first use the experimental data to ~~com~~

figure out the distribution as best you can, and then to use ~~Definition a Lemma Definition~~ Theorem 16.3.2 or Corollary A4 to compute its expectation. This method will help you identify cases where the expectation is infinite, ~~A straight average~~ and will generally be more accurate than a simple averaging of the data.

17.2 Expected Returns in Gambling Games

~~Suppose~~ Some of the most interesting examples of expectation ~~sums~~ can be ~~explained~~ best explained in terms of gambling games. For ~~straightforward~~ games where you win ~~\$B~~ \$A with probability p and you lose \$B with probability 1-p, it is easy to compute your expected return or winnings. It is simply

$$p \cdot A - (1-p) \cdot B .$$

For example, if you are flipping a fair coin and you win \$1 ~~for~~ if heads ~~comes up~~ and you lose \$1 ~~for~~ if tails; then your expected winnings

17.1.07 Conditional Expectation

— INSERT X goes here —
(text from pp 1058-1059)

As another example, consider the channel latency problem from section 17.1.6. The expected latency for this problem was infinite. But what if we look at the expected latency conditioned on the latency not exceeding n. Then

$$\begin{aligned}
Ex[D] &= \sum_{i=1}^{\infty} i \Pr(D = i \mid D \le n), \\
&= \sum_{i=1}^{\infty} i \frac{\Pr(D = i \wedge D \le n)}{\Pr(D \le n)} \\
&= \sum_{i=1}^{n} i \frac{i \Pr(D = i)}{\Pr(D \le n)} \\
&= \frac{1}{\Pr(D \le n)} \sum_{i=1}^{n} i \left(\frac{1}{i(i+1)}\right) \\
&= \frac{1}{\Pr(D \le n)} \sum_{i=1}^{n} \frac{1}{i+1} \\
&= \frac{1}{\Pr(D \le n)} \left(H_{n+1} - 1\right),
\end{aligned}$$

where H_{n+1} is the $(n+1)$st Harmonic number

$$H_{n+1} = \ln(n+1) + \gamma + \varepsilon(n)$$

and $\varepsilon(n) \leq$. ~~To compute Pr.~~ The second equality follows from the definition of conditional expectation, ~~and~~ the third equality follows from the fact the $Pr(O = i \wedge D \leq n) = 0$ for $i > n$, and the fourth equality follows from the definition of D in Equation T1.

To compute ~~the probability~~ $Pr(D \leq n)$, we observe that

$$Pr(D \leq n) = 1 - Pr(D > n)$$

$$= 1 - \sum_{i=n+1}^{\infty} \frac{1}{i} - \frac{1}{i+1}$$

$$= 1 - \left[\left(\frac{1}{n+1} - \frac{1}{n+2}\right) + \left(\frac{1}{n+2} - \frac{1}{n+3}\right) + \left(\frac{1}{n+3} - \frac{1}{n+4}\right) + \cdots\right]$$

$$= 1 - \frac{1}{n+1}$$

$$= \frac{n}{n+1} .$$

Hence,

$$Ex[D] = \frac{n+1}{n}(H_{n+1} - 1).$$

For $n = 1000$, this is about 6.5. This explains why the expected value of D appears to be finite when you try to evaluate it experimentally. If you compute 100 samples of D, it is likely that all of them will be at most 1000ms. If you condition on not having any outcomes greater than 1000ms, then the conditional expected value will be about 6.5, which would be a commonly observed result in practice. Yet we know that $E[D]$ is infinite. For this reason, expectations computed in practice are often really just conditional expectations where the condition is that rare "outlier" sample points are eliminated from the analysis.

The Law of Total Expectation

— INSERT Y goes here —

(text from pp 1059 – 1061)

~~For example~~ As a more interesting application of the Law of Total Expectation, let's take another look at the mean time to failure of a system that fails with probability p at each step. We'll define A_1 to be the event that the system fails on the first step and A_2 to be the complementary event (namely, that the system does not fail on the first step. Then the mean time to failure $Ex[C]$ is

$$Ex[C] = Ex[C|A_1]\Pr[A_1] + Ex[C|A_2]\Pr[A_2] \quad \text{(eqn 3)}$$

$$= 1 \cdot p + (Ex[C] + 1)(1-p)$$

$$= p + (1-p)Ex[C] + 1 - p$$

$$= 1 + (1-p)Ex[C] .$$

Rearranging terms, we find that

$$Ex[C] - (1-p)Ex[C] = 1$$

$$1 = Ex[C] - (1-p)Ex[C]$$

$$= pEx[C]$$

and thus that

$$Ex[C] = 1/p ,$$

as expected.

Since A_1 is the ~~system~~ condition that the system crashes on the first step, we know that

$$Ex[C|A_1] = 1. \qquad \text{(Eqn 1)}$$

since A_2 is the condition that the system does not crash on the first step, conditioning on A_2 is equivalent to taking a first step without failure and then starting over without conditioning. Hence,

$$Ex[C|A_2] = 1 + Ex[C]. \qquad \text{(Eqn 2)}$$

Plugging Equations 1 and 2 into Equation 3, we find that

$$\begin{aligned} Ex[C] &= 1 \cdot p + (1 + Ex[C])(1-p) \\ &= p + 1 - p + (1-p)Ex[C] \\ &= 1 + (1-p)Ex[C], \end{aligned}$$

Rearranging terms, we find that

$$1 = \mathrm{Ex}[c] - (1-p)\,\mathrm{Ex}[c]$$

$$= p\,\mathrm{Ex}[c],$$

and thus that

$$\mathrm{Ex}[c] = 1/p,$$

as expected.

We will use this sort of analysis extensively in chapter 19 when we examine the expected behavior of random walks.

~~FOE~~

17.1.8 Expectations of Functions of Random Variables

Expectations can also be defined for functions of any random variable R : S → V

Expectations can also be defined for functions of any random variable $R : S \to V$, that is any total function $f : V \to \mathbb{R}$, that is possible to define.

Definition #P7: Let $R : S \to V$ be a random variable and $f : V \to \mathbb{R}$ be a total function on the range of R. Then

$$Ex[f(R)] = \sum_{w \in S} f(R(w)) \Pr(w). \quad \text{(eqn P8)}$$

&

Equivalently,

$$Ex[f(R)] = \sum_{r \in \text{Range}(R)} f(r) \Pr(R = r). \quad \text{(eqn P9)}$$

For example, suppose that R is the value obtained by rolling a fair 6-sided die. Then

$$Ex[1/R] = \frac{1}{1} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6}$$

$$= \frac{49}{120}.$$

figure out the distribution as best you can, and then to use ~~a lemma~~ ~~Definition~~ Theorem 16.3.2 or Corollary A4 to compute its expectation. This method will help you identify cases where the expectation is infinite, ~~a straight average~~ and will generally be more accurate than a simple averaging of the data.

17.2 Expected Returns in Gambling Games

~~Suppose~~ Some of the most interesting examples of expectation ~~one~~ can be ~~at~~ best explained in terms of gambling games. For straightforward games where you win ~~#B~~ #A with probability p and you lose #B with probability 1-p, it is easy to compute your expected return or winnings. It is simply

$$p \cdot A - (1-p) \cdot B .$$

For example, if you are flipping a fair coin and you win #1 ~~for~~ if heads ~~comes up~~ and you lose #1 ~~for~~ if tails; then your expected winnings

are

$$\frac{1}{2} \cdot 1 \$ - \left(1 - \frac{1}{2}\right) \cdot 1 = 0.$$

In such cases, the game is said to be fair since your expected return is zero.

Some gambling games are more complicated, and thus more interesting. For example, consider the following game where the winners split a pot. This sort of game is representative of many poker games, betting pools and lotteries.

17.2.1 Splitting the Pot

~~Much has happened since your last encounter as~~

After your last encounter with biker dude, one thing led to another and you have dropped out of school and become a ~~joined the local chapter of~~ Hell's Angels. ~~One night~~ It's late on a Friday night and, a nostalgic ~~sad~~ feeling for the old days, you drop by your old ~~favorite~~ ~~pub where you~~ hangout, where you encounter two of your former TA's, Eric and Nick. Eric and

Nick propose that you join them in a simple wager. Each player will put $2 on the bar and secretly write "heads" or "tails" on their napkin. Then one player will flip a fair coin. The $6 on the bar[1] will then be divided evenly among the players who correctly predicted the outcome of the coin toss.

After your life-altering encounter with a strange dice, you are a little more than skeptical. So Eric and Nick agree to let you be the one to flip the coin. This certainly seems fair. How can you lose?

But you have learned your lesson and so before agreeing, you write out the tree diagram and compute your expected return. The tree diagram is shown in Figure E.1.

---

1 The money invested in a wager is commonly refered to as the pot.

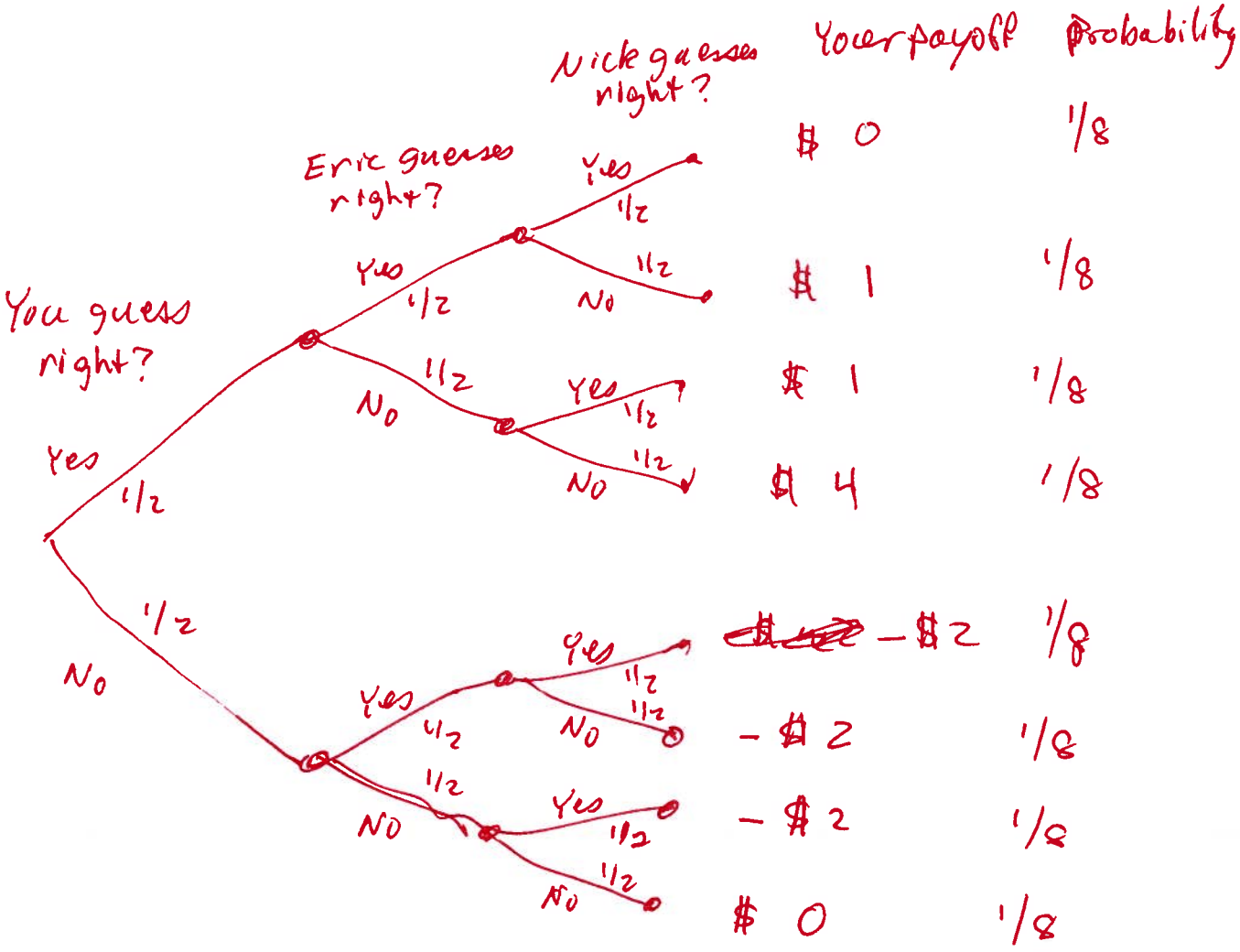

Figure E1: The tree diagram for the game where 3 players each wager $2, and then guess the outcome of a fair coin toss. The winners split the pot.

The "payoff" values in Figure E1 are computed by dividing the $6 pot among those players who guessed correctly and then subtracting the $2 that you put into the pot at the beginning. For example, if all three players guessed correctly, then your payoff is $0, since you just get back your $2 wager. If you and Nick guess correctly and Nick guessed wrong, then your payoff is

$$\frac{6}{2} - 2 = 1.$$

— so again, your payoff is zero.

In the case that everyone is wrong, you all agree to split the pot and

To compute your expected return, you use Equation 16.5 in the definition of expected value. This yields

$$\text{Ex}[\text{payoff}] = 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{1}{8} + 4 \cdot \tfrac{1}{8} + (-2) \cdot \tfrac{1}{8} + (-2) \cdot \tfrac{1}{8} + (-2) \cdot \tfrac{1}{8} + 0 \cdot \tfrac{1}{8}$$

$$= 0.$$

This confirms that the game is fair. So, for old time's sake, you break your solemn vow to never ever engage in strange gambling games.

17.2.2 The Impact of Collusion

Needless to say, things are not turning out well for you. The more times you play the game, the more money you seem to be losing. After 1000 wagers, you have lost over $500. How can this be? As Nick and Eric are consoling you on your "bad luck," you do a back-of-the-napkin calculation using the bounds on the tails of the binomial distribution from chapter 16 that suggests that the probability of losing $500 in 1000 wagers is less than the probability of a martian waltzing in and handing you one of those golden disks. How can this be?

Something must be wrong with the free

Good If it is possible that you are truly very very unlucky. But it is more likely

Well something is wrong with ~~your~~ the tree diagram in Figure E1. And that something just might have something to do with the possibility that Nick and Eric are collaborating against you. # To be sure, Nick and Eric can only guess the outcome of the coin toss with probability 1/2, but what if Nick and Eric always guess differently? In other words, what if ~~Eric~~ Nick always guesses ~~Eats~~ "tails" when Eric guesses "heads," and vice-versa. This would result in ~~the~~ a slightly different tree diagram, as shown in Figure E2.

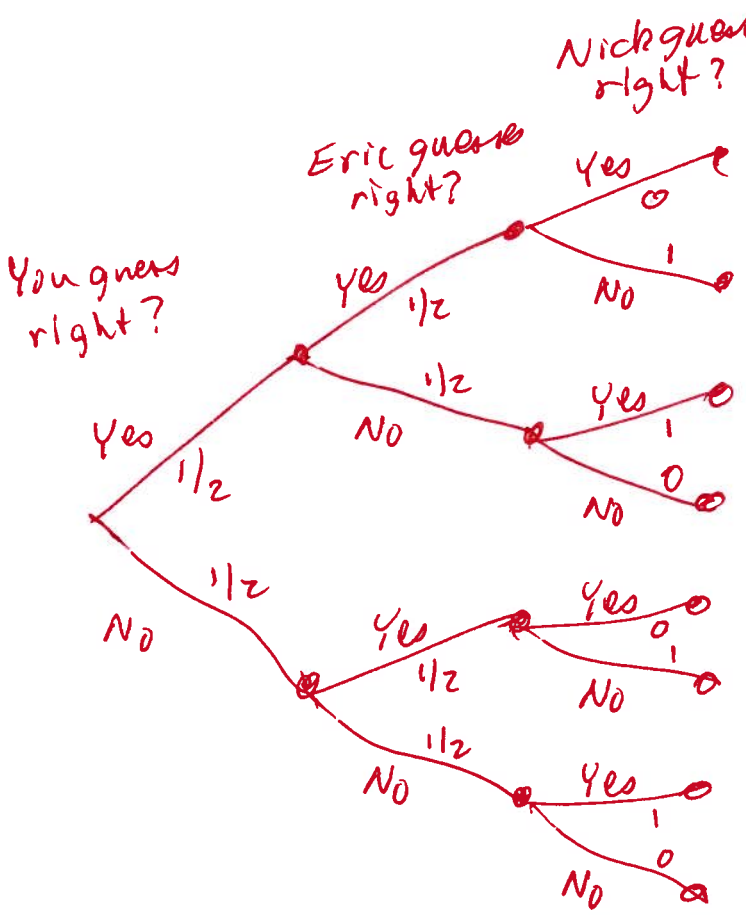

Figure E2: The revised tree diagram reflecting the scenario where Nick always guesses the opposite of Eric.

~~The difference between the tree diagrams in Figures E1 and E2~~

The payoffs for each outcome are the same in Figures E1 and E2, but the probabilities of the outcomes are different. For example, it is no longer possible for all three players to guess correctly, since Nick and Eric are always guessing differently. More importantly, the outcome where your payoff is $4 is also no longer possible. ~~Let's see what happens~~ When we use Equation 16.5 to compute your expected return, we find that

$$\begin{aligned} \text{Ex}[\text{payoff}] &= 0 \cdot 0 + 1 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{4} + 4 \cdot 0 + (-2) \cdot 0 \\ &\quad + (-2) \cdot \tfrac{1}{4} + (-2) \cdot \tfrac{1}{4} + 0 \cdot 0 \\ &= -\tfrac{1}{2}. \end{aligned}$$

are always guessing differently, one of them will always get a share of the pot. ~~This might not be good for you~~ As you might imagine, this is not good for you!

This is very bad indeed. By collaborating, ~~Because Nick~~ Nick and Eric have ~~made~~ reduced made it ~~the~~ so that you expect to lose $.50 every time you play. No wonder you lost ~~over~~ $500 ~~for~~ over the course of 1000 wagers. Maybe it would be a good idea to go back to school — your Hell's Angels buds may not be too happy that you lost ~~the~~ their ~~them~~ $500.

17.2.3 How to Win the Lottery

Similar opportunities to "collaborate" arise in many ~~betting~~ ~~go~~ betting games. For example, consider the ~~weekly football~~ typical weekly football betting pool, where each participant wagers $10 and the participants that ~~guesses~~ pick the most games correctly split a large pot. ~~It~~ ~~you~~ the pool seems fair if you think of it as in Figure E1. But, in fact, if two or more players collaborate ~~as in E1~~ by guessing ~~opposite~~

differently, they can get an "unfair" advantage at your expense!

In some cases, the collaboration is inadvertant and you can profit from it. For example, many years ago, a former MIT Professor of mathematics ~~figured~~ named Herman Chernoff figured out a way to make money by playing the state lottery. This was surprising since state lotteries typically have very poor expected returns. That's because the state usually takes a large share of the wagers before distributing the rest of the pot among the winners. ~~So how did he do it? That means that anyone~~ Hence, anyone who buys a lottery ticket is expected to lose money. So how did Chernoff find a way to make money? ~~To find~~ It turned out to be easy!

~~We'll see this same idea in Chapter 18 when we~~

In a typical state lottery:

the money it takes in. But Chernoff figured out a way to win! Here are rules for a typical lottery:

- All players pay $1 to play and select 4 numbers from 1 to 36,
- The state draws 4 numbers from 1 to 36 uniformly at random,
- The state divides 1/2 the money collected among the people who guessed correctly and spends the other half repairing the Big Dig.

This is a lot like our betting game, except that there are more players and more choices. Chernoff discovered that a small set of numbers was selected by a large fraction of the population apparently many people think the same way, not on purpose as in the previous game with Angelina and Arvind, e.g., based on Manny's batting average or today's date. It was as if the players were collaborating to lose! If any one of them guessed correctly, then they'd have to split the pot with many other players. By selecting numbers uniformly at random, Chernoff was unlikely to get one of these favored sequences. So if he won, he'd likely get the whole pot! Thus, in this case people unknowingly collaborate to lose! By analyzing actual state lottery data, he determined that he could win an average of 7 cents on the dollar (based on a profile of bets) this way!

There is another variation of the accidental collusion with a betting pool. For example, how many people ever participated in a Super Bowl betting pool where the goal was to get closest to the total number of points scored in the game? Suppose the average Super Bowl has a total of 30 points scored and everyone knows this and 50 people are in the pool. Then most people will guess around 30 points. Where should you guess? Well, you should guess just outside of this range because you get to cover a lot more ground and you don't share the pot if you win. Of course, if you are in a pool with 6.042 students and they all know this strategy, then maybe you should guess 30 points after all.

2 Equivalent Definitions of Expectation

There are some other ways of writing the definition of expectation. Sometimes using one of these other formulations can make computing an expectation a lot easier. One option is to group together all outcomes on which the random variable takes on the same value.

Theorem 1.

$$\mathrm{Ex}(R) = \sum_{x \in \mathrm{range}(R)} x \cdot \Pr\{R = x\}$$

17.3 Expectations of Sums

16.3.4 Linearity of Expectation

Expected values obey a simple, very helpful rule called *Linearity of Expectation*. Its simplest form says that the expected value of a sum of random variables is the sum of the expected values of the variables.

Theorem 16.3.7. *For any random variables R_1 and R_2,*

$$\mathrm{E}[R_1 + R_2] = \mathrm{E}[R_1] + \mathrm{E}[R_2].$$

Proof. Let $T ::= R_1 + R_2$. The proof follows straightforwardly by rearranging terms in

$$\begin{aligned}
\mathrm{E}[T] &= \sum_{\omega \in S} T(\omega) \cdot \Pr\{\omega\} && \text{(Theorem 16.3.2)} \\
&= \sum_{\omega \in S} (R_1(\omega) + R_2(\omega)) \cdot \Pr\{\omega\} && \text{(def of } T\text{)} \\
&= \sum_{\omega \in S} R_1(\omega) \Pr\{\omega\} + \sum_{\omega \in S} R_2(\omega) \Pr\{\omega\} && \text{(rearranging terms)} \\
&= \mathrm{E}[R_1] + \mathrm{E}[R_2]. && \text{(Theorem 16.3.2)}
\end{aligned}$$

■

A small extension of this proof, which we leave to the reader, implies

Theorem 16.3.8 (Linearity of Expectation). *For random variables R_1, R_2 and constants $a_1, a_2 \in \mathbb{R}$,*

$$\mathrm{E}[a_1 R_1 + a_2 R_2] = a_1 \mathrm{E}[R_1] + a_2 \mathrm{E}[R_2].$$

In other words, expectation is a linear function. A routine induction extends the result to more than two variables:

Corollary 16.3.9. *For any random variables $R_1, \ldots, R_k$ and constants $a_1, \ldots, a_k \in \mathbb{R}$,*

$$\mathrm{E}\left[\sum_{i=1}^{k} a_i R_i\right] = \sum_{i=1}^{k} a_i \mathrm{E}\left[R_i\right].$$

The great thing about linearity of expectation is that *no independence is required*. This is really useful, because dealing with independence is a pain, and we often need to work with random variables that are not independent.

EDITING NOTE: Even when the random variables *are* independent, we know from previous experience that proving independence requires a lot of work.

Expected Value of Two Dice

What is the expected value of the sum of two fair dice?

no # ⟵ Let the random variable R_1 be the number on the first die, and let R_2 be the number on the second die. We observed earlier that the expected value of one die is 3.5. We can find the expected value of the sum using linearity of expectation:

$$\mathrm{E}[R_1 + R_2] = \mathrm{E}[R_1] + \mathrm{E}[R_2] = 3.5 + 3.5 = 7.$$

Notice that we did *not* have to assume that the two dice were independent. The expected sum of two dice is 7, even if they are glued together (provided each individual die remains fair after the gluing). Proving that this expected sum is 7 with a tree diagram would be a bother: there are 36 cases. And if we did not assume that the dice were independent, the job would be really tough!

17.3.2 Sums of Indicator Variables

Linearity of expectation is especially useful when you have a sum of indicator random variables. ~~For example, consider the following problem~~ As an example, suppose there

The Hat-Check Problem

There is a dinner party where n men check their hats. The hats are mixed up during dinner, so that afterward each man receives a random hat. In particular, each man gets his own hat with probability $1/n$. What is the expected number of men who get their own hat?

Letting G be the number of men that get their own hat, we want to find the expectation of G. But all we know about G is that the probability that a man gets his own hat back is $1/n$. There are many different probability distributions of hat permutations with this property, so we don't know enough about the distribution of G to calculate its expectation directly. But linearity of expectation makes the problem really easy.

The trick is to express G as a sum of indicator variables. In particular, let G_i be an indicator for the event that the ith man gets his own hat. That is, $G_i = 1$ if he gets his

own hat, and $G_i = 0$ otherwise. The number of men that get their own hat is the sum of these indicators:

$$G = G_1 + G_2 + \cdots + G_n. \tag{16.8}$$

These indicator variables are *not* mutually independent. For example, if $n - 1$ men all get their own hats, then the last man is certain to receive his own hat. But, since we plan to use linearity of expectation, we don't have worry about independence!

Now since G_i is an indicator, we know $1/n = \Pr\{G_i = 1\} = \mathrm{E}[G_i]$ by Lemma 16.3.3.

Now we can take the expected value of both sides of equation (16.8) and apply linearity of expectation:

$$\begin{aligned}
\mathrm{E}[G] &= \mathrm{E}[G_1 + G_2 + \cdots + G_n] \\
&= \mathrm{E}[G_1] + \mathrm{E}[G_2] + \cdots + \mathrm{E}[G_n] \\
&= \underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}} = 1. \\
&= 1.
\end{aligned}$$

So even though we don't know much about how hats are scrambled, we've figured out that on average, just one man gets his own hat back!

INSERT G goes here

17.3.3 The Expectation of a Binomial Distribution

Suppose that we independently flip n biased coins, each with probability p of coming up heads. What is the expected number of heads?

INSERT H goes here

Let J be the number of heads after the flips, so J has the (n, p)-binomial distribution.

Now let I_k be the indicator for the kth coin coming up heads. By Lemma 16.3.3, we have

$$\mathrm{E}[I_k] = p.$$

But

$$J = \sum_{k=1}^{n} I_k,$$

This is just a special case of the hat-check problem, with appetizers in place of hats. In the hat-check problem, we assumed only that each man received his own hat with probability $1/n$. Beyond that, we made no assumptions about how the hats could be permuted. This problem is a special case because we happen to know that appetizers are cyclically shifted relative to their initial position. This means that either everyone gets their original appetizer back, or no one does. But our previous analysis still holds: the *expected* number of people that get their own appetizer back is 1.

The nice thing about solving problems with linearity of expectations is that you don't need to know very much about the underlying distribution to compute interesting facts about it. Because you don't use much information about the distribution, the calculations are also usually much easier using this method.

In fact, linearity of expectations provides a very general method for computing the expected number of events that will happen.

Theorem 1. *Given any collection of n events $A_1, \ldots, A_n \subseteq S$, the expected number of events that will occur is $\sum_{1 \le i \le n} \Pr(A_i)$.*

For example, A_i could be the event that the ith man gets the right hat back. But in general, it could be any subset of the sample space, and we are asking for the expected number of events that will contain a random sample point.

Proof. Define R_i to be the indicator variable for A_i, where $R_i(w) = 1$ if $w \in A_i$, and $R_i(w) = 0$ if $w \notin A_i$. Let $R = R_1 + R_2 + \cdots + R_n$. Then

$$\begin{aligned}
\mathrm{Ex}[R] &= \sum_{i=1}^{n} \mathrm{Ex}[R_i] && \text{(by Linearity of Expectation)} \\
&= \sum_{i=1}^{n} \Pr(R_i = 1) && \text{(by Lemma 16.3.3)} \\
&= \sum_{i=1}^{n} \sum_{w \in A_i} \Pr(w) && \text{(definition of indicator variable)} \\
&= \sum_{i=1}^{n} \Pr(A_i).
\end{aligned}$$

□

So whenever you are asked for the expected number of events that occur, all you have to do is sum the probabilities that each event occurs. Independence is not needed.

As a final example, suppose you flip N fair coins. Let R be the number of heads, and R_i the event that the ith coin is a head. Then $\mathrm{Ex}(R) = \sum_i \mathrm{Ex}(R)_i = \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} = \frac{N}{2}$.

This is the easy way. You could also solve this the hard way, assuming the coins are

INSERT H

Let J be the random variable denoting the number of heads. Then J has a binomial distribution with parameters n, p, and

$$\Pr[J=k] = \binom{n}{k} k^p (n-k)^{1-p}.$$

Applying Equation 16.4, this means that

$$E_x[J] = \sum_{k=0}^{n} k \Pr[J=k]$$

$$= \sum_{k=0}^{n} k \binom{n}{k} k^p (n-k)^{1-p}. \quad (\text{Eqn T7})$$

Let's try

Ouch! This is one nasty looking sum.

another approach.

Since we have just learned about linearity of expectations, sums of indicator random variables, may be will be helpful. But Theorem T3

how do we express J as a sum of indicator random variables? It turns out to be easy. Let J_i be the indicator random variable for the ith coin.

In particular, define

$$J_i = \begin{cases} 1 & \text{if the ith coin is heads} \\ 0 & \text{if the ith coin is tails.} \end{cases}$$

Then, the number of heads is simply

$$J = J_1 + J_2 + \cdots + J_n .$$

By Theorem T3,

$$Ex[J] = \sum_{i=1}^{n} Pr(J_i)$$

$$= np . \qquad \text{(Eqn T8)}$$

That really was easy. If we flip n mutually independ coins, we expect to get pn heads. Hence, the expected value of a binomial distribution with parameters n and p is simply pn.

But what if the coins are not mutually independent? It doesn't matter — the answer is still pn because Linearity of Expectation and Theorem T3 do not assume any independence.

If you are not yet convinced that linearity of expectation and Theorem T3 are simple and powerful tools, consider this: we have used them to prove a without even trying,

very complicated identity, ~~by combining~~ namely[1]

~~Equations I 7 and I 8 yields that~~

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (n-k)^{1-p} = pn .$$

~~This follows directly from Equat~~

If you are still not convinced, then take a look at the next problem.

---

1 This follows ~~directly from~~ by combining Equations I 7 and I 8.

so by linearity

$$\mathrm{E}[J] = \mathrm{E}\left[\sum_{k=1}^{n} I_k\right] = \sum_{k=1}^{n} \mathrm{E}[I_k] = \sum_{k=1}^{n} p = pn.$$

In short, the expectation of an (n, p)-binomially distributed variable is pn.

17.3.4 *The Coupon Collector Problem*

Every time I purchase a kid's meal at Taco Bell, I am graciously presented with a miniature "Racin' Rocket" car together with a launching device which enables me to project my new vehicle across any tabletop or smooth floor at high velocity. Truly, my delight knows no bounds.

There are n different types of Racin' Rocket car (blue, green, red, gray, etc.). The type of car awarded to me each day by the kind woman at the Taco Bell register appears to be selected uniformly and independently at random. What is the expected number of kid's meals that I must purchase in order to acquire at least one of each type of Racin'

Rocket car?

The same mathematical question shows up in many guises: for example, what is the expected number of people you must poll in order to find at least one person with each possible birthday? Here, instead of collecting Racin' Rocket cars, you're collecting birthdays. The general question is commonly called the *coupon collector problem* after yet another interpretation.

A clever application of linearity of expectation leads to a simple solution to the coupon collector problem. Suppose there are five different types of Racin' Rocket cars, and we receive this sequence:

blue green green red blue orange blue orange gray

Let's partition the sequence into 5 segments:

$$\underbrace{\text{blue}}_{X_0}\ \underbrace{\text{green}}_{X_1}\ \underbrace{\text{green red}}_{X_2}\ \underbrace{\text{blue orange}}_{X_3}\ \underbrace{\text{blue orange gray}}_{X_4}$$

The rule is that a segment ends whenever we get a new kind of car. For example, the middle segment ends when we get a red car for the first time. In this way, we can break the problem of collecting every type of car into stages. Then we can analyze each stage individually and assemble the results using linearity of expectation.

Let's return to the general case where we're collecting n Racin' Rockets. Let X_k be the length of the kth segment. The total number of kid's meals we must purchase to get all n Racin' Rockets is the sum of the lengths of all these segments:

$$T = X_0 + X_1 + \cdots + X_{n-1}$$

Now let's focus our attention on X_k, the length of the kth segment. At the beginning of segment k, we have k different types of car, and the segment ends when we acquire a new type. When we own k types, each kid's meal contains a type that we already have with probability k/n. Therefore, each meal contains a new type of car with probability

$1 - k/n = (n-k)/n$. Thus, the expected number of meals until we get a new kind of car is $n/(n-k)$ by the "mean time to failure" formula in Equation T10. This means that So we have:

$$\mathrm{E}[X_k] = \frac{n}{n-k}.$$

Linearity of expectation, together with this observation, solves the coupon collector problem:

$$\begin{aligned}
\mathrm{E}[T] &= \mathrm{E}[X_0 + X_1 + \cdots + X_{n-1}] \\
&= \mathrm{E}[X_0] + \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_{n-1}] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \cdots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1} \\
&= n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&= n\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} + \frac{1}{n}\right) \\
&= nH_n \qquad \text{(eqn T11)} \\
&\sim n \ln(n).
\end{aligned}$$

we can use Equation T11

Let's use this general solution to answer some concrete questions. For example, the

Wow! It's those Harmonic numbers again!

expected number of die rolls required to see every number from 1 to 6 is:

$$6H_6 = 14.7\ldots$$

And the expected number of people you must poll to find at least one person with each possible birthday is:

$$365H_{365} = 2364.6\ldots$$

EDITING NOTE: unedited from F02

Let A_i be the event that coin i comes up heads. Since the coin is fair, $\Pr\{A_i\} = 1/2$. Since there are N coins in all, there are N such events. By linearity of expectation (Theorem 16.3.9), the expected number of events that occur —the number of coins that come up heads —is $N(1/2) = N/2$.

Let's try to solve the same problem the hard way. In this case, assume that the coins are fair. Let the random variable R be the number of heads. We want to compute the

expected value of R.

$$\begin{aligned} \mathrm{E}[R] &= \sum_{i=0}^{N} i \cdot \Pr\{R = i\} \\ &= \sum_{i=0}^{N} i \binom{N}{i} 2^{-N} \end{aligned}$$

The first equation follows from the definition of expectation. In the second step, we evaluate $\Pr\{R = i\}$. An outcome of tossing the N coins can be represented by a length N sequence of H's and T's. An H in position i indicates that the ith coin is heads, and a T indicates that the ith coin is tails. The sample space consists of all 2^N such sequences. The outcomes are equiprobable, and so each has probability 2^{-N}. The number of outcomes with exactly i heads is the number of length N sequences with i H's, which is $\binom{N}{i}$. Therefore, $\Pr\{R = i\} = \binom{N}{i} 2^{-N}$.

The answer from linearity of expectation and from the hard way must be the same,

so we can equate the two results to obtain a neat identity.[1]

$$\sum_{i=0}^{N} i\binom{N}{i}2^{-N} = \frac{N}{2}$$

$$\sum_{i=0}^{N} i\binom{N}{i} = N2^{N-1}$$

The expected number of heads is $N/2$, even if some coins are glued together.

We can extend this reasoning to n tosses of a coin with probability p of a head, rather than $1/2$. If we do this, we get the generalized combinatorial identity:

$$\sum_{i=0}^{N} i\binom{N}{i}p^i(1-p)^{N-i} = Np$$

Here, the p^i factor gives the probabilities for the heads and the $(1-p)^{N-i}$ factor gives the probabilities for the tails. The right-hand side is the sum of N terms, each giving the probability of a particular A_i, which is p. The total is Np. For example, consider an

[1] The identity also has a simple combinatorial proof given in Problem ??.

ordinary die. Let A_1 be the event that the value is odd, A_2 the event that the value is 1, 2, or 3, and A_3 the event that the value is 4, 5, or 6. These events are not mutually independent. However, the expected number of these events that occur is still obtainable by adding $\Pr\{A_1\} + \Pr\{A_2\} + \Pr\{A_3\}$, which yields 3/2.

The Number-Picking Game

Here is a game that you and I could play that reveals a strange property of expectation.

First, you think of a probability density function on the natural numbers. Your distribution can be absolutely anything you like. For example, you might choose a uniform distribution on $1, 2, \ldots, 6$, like the outcome of a fair die roll. Or you might choose a binomial distribution on $0, 1, \ldots, n$. You can even give every natural number a non-zero probability, provided that the sum of all probabilities is 1.

Next, I pick a random number z according to your distribution. Then, you pick a

random number y_1 according to the same distribution. If your number is bigger than mine ($y_1 > z$), then the game ends. Otherwise, if our numbers are equal or mine is bigger ($z \geq y_1$), then you pick a new number y_2 with the same distribution, and keep picking values y_3, y_4, etc. until you get a value that is strictly bigger than my number, z. What is the expected number of picks that you must make?

Certainly, you always need at least one pick, so the expected number is greater than one. An answer like 2 or 3 sounds reasonable, though one might suspect that the answer depends on the distribution. Let's find out whether or not this intuition is correct.

The number of picks you must make is a natural-valued random variable, so from formula (16.7) we have:

$$\mathrm{E}\left[\#\text{ picks by you}\right] = \sum_{k \in \mathbb{N}} \Pr\left\{(\#\text{ picks by you}) > k\right\} \tag{16.9}$$

Suppose that I've picked my number z, and you have picked k numbers $y_1, y_2, \ldots, y_k$.

There are two possibilities:

- If there is a unique largest number among our picks, then my number is as likely to be it as any one of yours. So with probability $1/(k+1)$ my number is larger than all of yours, and you must pick again.

- Otherwise, there are several numbers tied for largest. My number is as likely to be one of these as any of your numbers, so with probability greater than $1/(k+1)$ you must pick again.

In both cases, with probability at least $1/(k+1)$, you need more than k picks to beat me. In other words:

$$\Pr\{(\#\text{ picks by you}) > k\} \geq \frac{1}{k+1} \tag{16.10}$$

This suggests that in order to minimize your rolls, you should choose a distribution such that ties are very rare. For example, you might choose the uniform distribution on

$\{1, 2, \ldots, 10^{100}\}$. In this case, the probability that you need more than k picks to beat me is very close to $1/(k+1)$ for moderate values of k. For example, the probability that you need more than 99 picks is almost exactly 1%. This sounds very promising for you; intuitively, you might expect to win within a reasonable number of picks on average!

Unfortunately for intuition, there is a simple proof that the expected number of picks that you need in order to beat me is *infinite*, regardless of the distribution! Let's plug (16.10) into (16.9):

$$\begin{aligned} \mathrm{E}[\#\text{ picks by you}] &= \sum_{k \in \mathbb{N}} \frac{1}{k+1} \\ &= \infty \end{aligned}$$

end of text for problem scatter

This phenomenon can cause all sorts of confusion! For example, suppose you have a communication network where each packet of data has a $1/k$ chance of being delayed by k or more steps. This sounds good; there is only a 1% chance of being delayed by 100

or more steps. But the *expected* delay for the packet is actually infinite!

There is a larger point here as well: not every random variable has a well-defined expectation. This idea may be disturbing at first, but remember that an expected value is just a weighted average. And there are many sets of numbers that have no conventional average either, such as:

$$\{1, -2, 3, -4, 5, -6, \ldots\}$$

Strictly speaking, we should qualify virtually all theorems involving expectation with phrases such as "...provided all expectations exist." But we're going to leave that assumption implicit.

Random variables with infinite or ill-defined expectations are more the exception than the rule, but they do creep in occasionally.

■

David: This whole section 16.4 moves to page 1109 as INSERT Z. It becomes section 17.5.

17.5
~~16.4~~ Expectation ~~of a~~ Quotient~~s~~

INSERT P goes here

17.5.1 ~~16.4.1~~ A RISC Paradox

data in Figure P4 is representative of data in ~~this paper~~

The ~~following data is taken from~~ a paper by some famous professors. They wanted to show that programs on a RISC processor are generally shorter than programs on a CISC processor. For this purpose, they applied a RISC compiler and then a CISC compiler to some benchmark source programs and made a table of compiled program lengths.

Benchmark	RISC	CISC	CISC/RISC
E-string search	150	120	0.8
F-bit test	120	180	1.5
Ackerman	150	300	2.0
Rec 2-sort	2800	1400	0.5
Average			1.2

the program length for ~~the~~ ~~data~~

Figure P4: ~~Data~~ ~~that~~ ~~to complete~~ used benchmark problems using RISC and CISC compilers.

in Figure P4

Each row contains the data for one benchmark. The numbers in the second and third columns are program lengths for each type of compiler. The fourth column contains the ratio of the CISC program length to the RISC program length. Averaging this ratio over

INSERT P

If S and T are random variables, we know from ~~section~~ linearity of Expectation that

$$Ex[S+T] = Ex[S] + Ex[T].$$

If S and T are independent, we know from Theorem 16.4.4, that

$$Ex[S \cdot T] = Ex[S]\,Ex[T].$$

Is it also true that

$$Ex[S/T] = Ex[S] / Ex[T] \; ? \quad (\text{eqn P1})$$

of course, we have to worry about the situation when $T=0$, but what if we assume that T is always positive? ~~Let's see~~ As we will soon see, Equation P1 is usually ~~false~~ not true, but let's see if we can prove it anyway...

P2

False Claim ~~#~~: If S and T are independent random variables with $T>0$, then

$$Ex[S/T] = Ex[S]/Ex[T].$$

— INSERT Q goes here —

(Text on p 1092)

Here is a counterexample. ~~Let~~ Define T so that

~~T = 1 with probability 1/2~~
~~2~~

$\Pr[T=1] = 1/2$ and $\Pr[T=2] = 1/2$.

~~$\Pr[T=1] = 1/2$~~

Then

$$Ex[T] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2}$$

$$= \frac{3}{2}.$$

and

$$\frac{1}{Ex[T]} = \frac{2}{3}$$

~~and~~

~~$Ex\left[\frac{1}{T}\right] = 1$~~

$$Ex\left[\frac{1}{T}\right] = \frac{1}{1} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}$$

$$= \frac{3}{4}.$$

$$\neq \frac{1}{Ex[\frac{1}{T}]}.$$

~~In fact, it is~~ Clearly

This means that Claim 16.4.1 is also false since we could define $S = 1$ with probability 1. In fact, both Claims 16.4.1 and 16.4.2 are untrue for most all choices of S and T. Unfortunately, ~~both~~ the fact that they are false does not keep them from being widely used in practice! Let's see an example.

all benchmarks gives the value 1.2 in the lower right. The authors conclude that CISC programs are 20% longer on average.

However, some critics of their paper took the same data and argued this way: redo the final column, taking the other ratio, RISC/CISC instead of CISC/RISC,

Benchmark	RISC	CISC	RISC/CISC
E-string search	150	120	1.25
F-bit test	120	180	0.67
Ackerman	150	300	0.5
Rec 2-sort	2800	1400	2.0
Average			1.1

From this table, we would conclude that RISC programs are 10% longer than CISC programs on average! We are using the same reasoning as in the paper, so this conclusion is equally justifiable— yet the result is opposite! What is going on?

16.4.2 A Probabilistic Interpretation

To resolve these contradictory conclusions, we can model the RISC vs. CISC debate with the machinery of probability theory.

Let the sample space be the set of benchmark programs. Let the random variable R be the length of the compiled RISC program, and let the random variable C be the length of the compiled CISC program. We would like to compare the average length, $\mathrm{E}[R]$, of a RISC program to the average length, $\mathrm{E}[C]$, of a CISC program.

To compare average program lengths, we must assign a probability to each sample point; in effect, this assigns a "weight" to each benchmark. One might like to weigh benchmarks based on how frequently similar programs arise in practice. Lacking such data, however, we will assign all benchmarks equal weight; that is, our sample space is uniform.

In terms of our probability model, the paper computes C/R for each sample point, and then averages to obtain $\mathrm{E}[C/R] = 1.2$. This much is correct. The authors then conclude that "CISC programs are 20% longer on average"; that is, they conclude that

$\mathrm{E}[C] = 1.2\,\mathrm{E}[R]$.

There in lies the problem. The authors have implicitly used False Claim 16.4.1 to assume that $E_X[C/R] = E_X[C] / E_X[R]$.

~~This is why the~~ By using the same false logic, ~~this is why~~ the critics can arrive at the opposite conclusion; namely that RISC programs are 10% longer on average.

Similarly, the critics calculation correctly showed that $\mathrm{E}[R/C] = 1.1$. They then concluded that $\mathrm{E}[R] = 1.1\ \mathrm{E}[C]$, that is, a RISC program is 10% longer than a CISC program on average.

These arguments make a natural assumption, namely, that

False Claim 16.4.1. *If S and T are independent random variables with $T > 0$, then*

$$\mathrm{E}\left[\frac{S}{T}\right] = \frac{\mathrm{E}[S]}{\mathrm{E}[T]}.$$

In other words False Claim 16.4.1 simply generalizes the rule for expectation of a product to a rule for the expectation of a quotient. But the rule for requires independence, and we surely don't expect C and R to be independent: large source programs will lead to large compiled programs, so when the RISC program is large, so the CISC would be too.

However, we can easily compensate for this kind of dependence: we should compare

the lengths of the programs *relative to the size of the source code*. While the lengths of C and R are dependent, it's more plausible that their *relative* lengths will be independent. So we really want to divide the second and third entries in each row of the table by a "normalizing factor" equal to the length of the benchmark program in the first entry of the row.

But note that normalizing this way will have no effect on the fourth column! That's because the normalizing factors applied to the second and and third entries of the rows will cancel. So the independence hypothesis of False Claim 16.4.1 may be justified, in which case the authors' conclusions would be justified. But then, so would the contradictory conclusions of the critics. Something must be wrong! Maybe it's False Claim 16.4.1 (duh!), so let's try and prove it.

This is Insert Q & goes to page P-2

False proof.

$$\mathrm{E}\left[\frac{S}{T}\right] = \mathrm{E}\left[S \cdot \frac{1}{T}\right]$$

$$= \mathrm{E}[S] \cdot \mathrm{E}\left[\frac{1}{T}\right] \qquad \text{(independence of } S \text{ and } T\text{)} \qquad (16.11)$$

$$= \mathrm{E}[S] \cdot \frac{1}{\mathrm{E}[T]}. \qquad (16.12)$$

$$= \frac{\mathrm{E}[S]}{\mathrm{E}[T]}.$$

□

Note that line (16.11) uses the fact that if S and T are independent, then so are S and ~~This is not~~ $1/T$. This holds because functions of independent random variables yield independent random variables, as ~~shown in Problem ??~~. It is a fact that needs proof, which we will leave to the reader, but it is not the bug. ■

~~But this proof is bogus!~~ The bug is in line (16.12), which assumes

claim

False ~~Theorem~~ 16.4.2.

$$\mathrm{E}\left[\frac{1}{T}\right] = \frac{1}{\mathrm{E}[T]}.$$

Here is a counterexample:

Example. Suppose $T = 1$ with probability $1/2$ and $T = 2$ with probability $1/2$. Then

$$\begin{aligned}
\frac{1}{\mathrm{E}[T]} &= \frac{1}{1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2}} \\
&= \frac{2}{3} \\
&\neq \frac{3}{4} \\
&= \frac{1}{1} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \\
&= \mathrm{E}\left[\frac{1}{T}\right].
\end{aligned}$$

The two quantities are not equal, so False Claim 16.4.2 really is false.

Unfortunately, the fact that Claim 16.4.1 and 16.4.2 are false does not mean that they are never used!

this is insert Q
d at desktop 1096

~~16.4.3~~ The Proper Quotient ← subsubsection

We can compute $\mathrm{E}[R]$ and $\mathrm{E}[C]$ as follows:

$$\begin{aligned}
\mathrm{E}[R] &= \sum_{i \in \text{Range(R)}} i \cdot \Pr\{R = i\} \\
&= \frac{150}{4} + \frac{120}{4} + \frac{150}{4} + \frac{2800}{4} \\
&= 805
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[C] &= \sum_{i \in \text{Range(C)}} i \cdot \Pr\{C = i\} \\
&= \frac{120}{4} + \frac{180}{4} + \frac{300}{4} + \frac{1400}{4} \\
&= 500
\end{aligned}$$

Now since $\mathrm{E}[R] / \mathrm{E}[C] = 1.61$, we conclude that the average RISC program is 61% longer than the average CISC program. This is a third answer, completely different from

the other two! Furthermore, this answer makes RISC look really bad in terms of code length. This one is the correct conclusion, under our assumption that the benchmarks deserve equal weight. Neither of the earlier results were correct—not surprising since both were based on the same false Claim.

end of insert a

16.4.4 A Simpler Example ← subsubsection

The source of the problem is clearer in the following, simpler example. Suppose the data were as follows.

Benchmark	Processor A	Processor B	B/A	A/B
Problem 1	2	1	1/2	2
Problem 2	1	2	2	1/2
Average			1.25	1.25

Now the data for the processors A and B is exactly symmetric; the two processors are equivalent. Yet, from the third column we would conclude that Processor B programs are 25% longer on average, and from the fourth column we would conclude that Pro-

cessor A programs are 25% longer on average. Both conclusions are obviously wrong.

The moral is that one must be very careful in summarizing data, we must not take an average of ratios blindly!

INSERT Q goes here (from PP 1094-1095.)
end of text moving to Section 17.5

EDITING NOTE.

Infinite Linearity of Expectation

We know that expectation is linear over finite sums. It's useful to extend this result to infinite summations. This works as long as we avoid sums whose values may depend on the order of summation.

17.3.5 Infinite Sums

Linearity of expectation also works for an infinite number of random variables provided that the variables satisfy some stringent absolute convergence criteria.

Convergence Conditions for Infinite Linearity

Theorem 16.4.3. *[Linearity of Expectation] Let $R_0, R_1, \ldots$, be random variables such that*

$$\sum_{i=0}^{\infty} \mathrm{E}\left[|R_i|\right]$$

converges. Then

$$\mathrm{E}\left[\sum_{i=0}^{\infty} R_i\right] = \sum_{i=0}^{\infty} \mathrm{E}\left[R_i\right].$$

Proof. Let $T ::= \sum_{i=0}^{\infty} R_i$.

We leave it to the reader to verify that, under the given convergence hypothesis, all the sums in the following derivation are absolutely convergent, which justifies rearranging

them as follows:

$$\begin{aligned}
\sum_{i=0}^{\infty} \mathrm{E}[R_i] &= \sum_{i=0}^{\infty} \sum_{s \in \mathcal{S}} R_i(s) \cdot \Pr\{s\} && \text{(Def. 16.5)} \\
&= \sum_{s \in \mathcal{S}} \sum_{i=0}^{\infty} R_i(s) \cdot \Pr\{s\} && \text{(exchanging order of summation)} \\
&= \sum_{s \in \mathcal{S}} \left[\sum_{i=0}^{\infty} R_i(s) \right] \cdot \Pr\{s\} && \text{(factoring out } \Pr\{s\}) \\
&= \sum_{s \in \mathcal{S}} T(s) \cdot \Pr\{s\} && \text{(Def. of } T) \\
&= \mathrm{E}[T] && \text{(Def. 16.5)} \\
&= \mathrm{E}\left[\sum_{i=0}^{\infty} R_i \right]. && \text{(Def. of } T).
\end{aligned}$$

■

Note that the finite linearity of expectation we established in Corollary 16.3.9 follows as a special case of Theorem 16.4.3: since $\mathrm{E}[R_i]$ is finite, so is $\mathrm{E}[|R_i|]$, and therefore so is their sum for $0 \leq i \leq n$. Hence the convergence hypothesis of Theorem 16.4.3 is trivially

satisfied if there are only finitely many R_i's.

Exercise: Show that linearity of expectation fails for the sum of two variables, one with expectation $+\infty$ and the other with $-\infty$.

A Paradox

move this to the problems

One of the simplest casino bets is on "red" or "black" at the roulette table. In each play at roulette, a small ball is set spinning around a roulette wheel until it lands in a red, black, or green colored slot. The payoff for a bet on red or black matches the bet; for example, if you bet \$10 on red and the ball lands in a red slot, you get back your original \$10 bet plus another matching \$10.

In the US, a roulette wheel has 2 green slots among 18 black and 18 red slots, so the probability of red is $p ::= 18/38 \approx 0.473$. In Europe, where roulette wheels have only 1 green slot, the odds for red are a little better —that is, $p = 18/37 \approx 0.486$—but still less

than even. To make the game fair, we might agree to ignore green, so that $p = 1/2$.

There is a notorious gambling strategy which seems to guarantee a profit at roulette: bet \$10 on red, and keep doubling the bet until a red comes up. This strategy implies that a player will leave the game as a net winner of \$10 as soon as the red first appears. Of course the player may need an awfully large bankroll to avoid going bankrupt before red shows up—but we know that the mean time until a red occurs is $1/p$, so it seems possible that a moderate bankroll might actually work out. (In this setting, a "win" on red corresponds to a "failure" in a mean-time-to-failure situation.)

Suppose we have the good fortune to gamble against a fair roulette wheel. In this case, our expected win on any spin is zero, since at the ith spin we are equally likely to win or lose $10 \cdot 2^{i-1}$ dollars. So our expected win after any finite number of spins remains zero, and therefore our expected win using this gambling strategy is zero. This is just what we should have anticipated in a fair game.

But wait a minute. As long as there is a fixed, positive probability of red appearing on each spin of the wheel—even if the wheel is unfair—it's *certain* that red will eventually come up. So with probability one, we leave the casino having won \$10, and our expected dollar win is obviously \$10, not zero!

Something's wrong here. What?

Solution to the Paradox

The expected amount won is indeed \$10.

The argument claiming the expectation is zero is flawed by an invalid use of linearity of expectation for an infinite sum. To pinpoint this flaw, let's first make the sample space explicit: a sample point is a sequence $B^n R$ representing a run of $n \geq 0$ black spins terminated by a red spin. Since the wheel is fair, the probability of $B^n R$ is $2^{-(n+1)}$.

Let C_i be the number of dollars won on the ith spin. So $C_i = 10 \cdot 2^{i-1}$ when red comes

up for the first time on the ith spin, that is, at precisely one sample point, namely $B^{i-1}R$. Similarly, $C_i = -10 \cdot 2^{i-1}$ when the first red spin comes up after the ith spin, namely, at the sample points $B^n R$ for $n \geq i$. Finally, we will define C_i by convention to be zero at sample points in which the session ends before the ith spin, that is, at points $B^n R$ for $n < i - 1$.

The dollar amount won in any gambling session is the value of the sum $\sum_{i=1}^{\infty} C_i$. At any sample point $B^n R$, the value of this sum is

$$10 \cdot -(1 + 2 + 2^2 + \cdots + 2^{n-1}) + 10 \cdot 2^n = 10,$$

which trivially implies that its expectation is 10 as well. That is, the amount we are *certain* to leave the casino with, as well as expectation of the amount we win, is \$10.

Moreover, our reasoning that $\mathrm{E}[C_i] = 0$ is sound, so

$$\sum_{i=1}^{\infty} \mathrm{E}[C_i] = \sum_{i=1}^{\infty} 0 = 0.$$

The flaw in our argument is the claim that, since the expectation at each spin was zero, therefore the final expectation would also be zero. Formally, this corresponds to concluding that

$$\mathrm{E}\left[\text{amount won}\right] = \mathrm{E}\left[\sum_{i=1}^{\infty} C_i\right] = \sum_{i=1}^{\infty} \mathrm{E}\left[C_i\right] = 0.$$

The flaw lies exactly in the second equality. This is a case where linearity of expectation fails to hold—even though both $\sum_{i=1}^{\infty} \mathrm{E}\left[C_i\right]$ and $\mathrm{E}\left[\sum_{i=1}^{\infty} C_i\right]$ are finite—because the convergence hypothesis needed for linearity is false. Namely, the sum

$$\sum_{i=1}^{\infty} \mathrm{E}\left[|C_i|\right]$$

does not converge. In fact, the expected value of $|C_i|$ is 10 because $|C_i| = 10 \cdot 2^i$ with probability 2^{-i} and otherwise is zero, so this sum rapidly approaches infinity.

Probability theory truly leads to this apparently paradoxical conclusion: a game allowing an unbounded—even though always finite—number of "fair" moves may not

be fair in the end. In fact, our reasoning leads to an even more startling conclusion: even against an *unfair* wheel, as long as there is some fixed positive probability of red on each spin, we are certain to win $10!

This is clearly a case where naive intuition is unreliable: we don't expect to beat a fair game, and we do expect to lose when the odds are against us. Nevertheless, the "paradox" that in fact we always win by bet-doubling cannot be denied.

But remember that from the start we chose to assume that no one goes bankrupt while executing our bet-doubling strategy. This assumption is crucial, because the expected loss while waiting for the strategy to produce its ten dollar profit is actually infinite! So it's not surprising, after all, that we arrived at an apparently paradoxical conclusion from an unrealistic assumption.

This example also serves a warning that in making use of infinite linearity of expectation, the convergence hypothesis which justifies it had better be checked.

end of text moving to problem section

For WALD'S theorem see F02 ln11-12.

17.4 Expectations of Products ← section

16.4.5 The Expected Value of a Product

While the expectation of a sum is the sum of the expectations, the same is usually not true for products. But it is true in an important special case, namely, when the random variables are *independent*.

→ For example, suppose that we roll a fair 6-sided die and denote the outcome with the random variable R.

For example, suppose we throw two *independent*, fair dice and multiply the numbers that come up. What is the expected value of this product?

We know that $Ex[R] = 3\frac{1}{2}$. Th

Let random variables R_1 and R_2 be the numbers shown on the two dice. We can compute the expected value of the product as follows:

$$\mathrm{E}[R_1 \cdot R_2] = \mathrm{E}[R_1] \cdot \mathrm{E}[R_2] = 3.5 \cdot 3.5 = 12.25. \tag{16.13}$$

Does $Ex[R \cdot R] = Ex[R] \cdot Ex[R]$?

We know that $Ex[R] = 3\frac{1}{2}$ and thus $Ex[R]^2 = 12\frac{1}{4}$. Let's compute $Ex[R^2]$ also to see if we get the same result.

Here the first equality holds because the dice are independent.

At the other extreme, suppose the second die is always the same as the first. Now $R_1 = R_2$, and we can compute the expectation, $\mathrm{E}\left[R_1^2\right]$, of the product of the dice explicitly, confirming that it is not equal to the product of the expectations.

$$\mathrm{E}[R_1 \cdot R_2] = \mathrm{E}\left[R_1^2\right]$$

$$\mathrm{Ex}[R^2] = \sum_{i=1}^{6} i^2 \cdot \Pr\{R_1^2 = i^2\}$$

$$= \sum_{i=1}^{6} i^2 \cdot \Pr[R_1 = i]$$

$$= \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6}$$

$$= 15\ 1/6$$

$$\neq 12\ 1/4 .$$

$$\neq \mathrm{E}[R_1] \cdot \mathrm{E}[R_2]$$

$$\sum_{w \in S} R^2(w) \Pr[w]$$

Hence,

$$\mathrm{Ex}[R^2] \neq \mathrm{Ex}[R]^2$$

and so the expectation of a product is not always equal to the product of the expectation.

There is a special case when such a relationship does hold; namely, when the random variables in the product are independent.

Theorem 16.4.4. *For any two* independent *random variables* R_1, R_2,

$$\mathrm{E}[R_1 \cdot R_2] = \mathrm{E}[R_1] \cdot \mathrm{E}[R_2].$$

Proof. The event $[R_1 \cdot R_2 = r]$ can be split up into events of the form $[R_1 = r_1 \text{ and } R_2 =$

$r_2]$ where $r_1 \cdot r_2 = r$. So

Ex

$$\mathrm{E}[R_1 \cdot R_2]$$

$$::= \sum_{r \in \text{range}(R_1 \cdot R_2)} r \cdot \Pr\{R_1 \cdot R_2 = r\}$$

$$= \sum_{r_i \in \text{range}(R_i)} r_1 r_2 \cdot \Pr\{R_1 = r_1 \text{ and } R_2 = r_2\}$$

$$= \sum_{r_1 \in \text{range}(R_1)} \sum_{r_2 \in \text{range}(R_2)} r_1 r_2 \cdot \Pr\{R_1 = r_1 \text{ and } R_2 = r_2\} \quad \text{(ordering terms in the sum)}$$

$$= \sum_{r_1 \in \text{range}(R_1)} \sum_{r_2 \in \text{range}(R_2)} r_1 r_2 \cdot \Pr\{R_1 = r_1\} \cdot \Pr\{R_2 = r_2\} \quad \text{(indep. of } R_1, R_2\text{)}$$

$$= \sum_{r_1 \in \text{range}(R_1)} \left(r_1 \Pr\{R_1 = r_1\} \cdot \sum_{r_2 \in \text{range}(R_2)} r_2 \Pr\{R_2 = r_2\} \right) \quad \text{(factoring out } r_1 \Pr\{R_1 = r_1\}\text{)}$$

Ex

$$= \sum_{r_1 \in \text{range}(R_1)} r_1 \Pr\{R_1 = r_1\} \cdot \mathrm{E}[R_2] \quad \text{(def of } \mathrm{E}[R_2]\text{)}$$

Ex

$$= \mathrm{E}[R_2] \cdot \sum_{r_1 \in \text{range}(R_1)} r_1 \Pr\{R_1 = r_1\} \quad \text{(factoring out } \mathrm{E}[R_2]\text{)}$$

Ex Ex

$$= \mathrm{E}[R_2] \cdot \mathrm{E}[R_1]\,. \quad \text{(def of } \mathrm{E}[R_1]\text{)}$$

■

For example, let R_1 and R_2 be random variables denoting the result of rolling two independent and fair 6-sided dice. Then

$$\mathrm{Ex}[R_1 \cdot R_2] = \mathrm{Ex}[R_1]\,\mathrm{Ex}[R_2]$$
$$= 3\tfrac{1}{2} \cdot 3\tfrac{1}{2}$$
$$= 12\tfrac{1}{4}\,.$$

by induction random

Theorem 16.4.4 extends routinely to a collection of mutually independent variables.

Corollary 16.4.5. *If random variables $R_1, R_2, \ldots, R_k$ are mutually independent, then*

$$\mathrm{E}\left[\prod_{i=1}^{k} R_i\right] = \prod_{i=1}^{k} \mathrm{E}[R_i].$$

Ex Ex

Class Problems

Homework Problems

Practice Problems

Class Problems

Homework Problems

— INSERT B goes here —
(This is the old 16.4 on pp 1087–1096)

— INSERT K goes here —

17.6 Problems