19 Random Variables

Thus far, we have focused on probabilities of events. For example, we computed the probability that you win the Monty Hall game or that you have a rare medical condition given that you tested positive. But, in many cases we would like to know more. For example, *how many* contestants must play the Monty Hall game until one of them finally wins? *How long* will this condition last? *How much* will I lose gambling with strange dice all night? To answer such questions, we need to work with random variables.

19.1 Random Variable Examples

Definition 19.1.1. A *random variable* R on a probability space is a total function whose domain is the sample space.

The codomain of R can be anything, but will usually be a subset of the real numbers. Notice that the name “random variable” is a misnomer; random variables are actually functions.

For example, suppose we toss three independent, unbiased coins. Let C be the number of heads that appear. Let $M = 1$ if the three coins come up all heads or all tails, and let $M = 0$ otherwise. Now every outcome of the three coin flips uniquely determines the values of C and M. For example, if we flip heads, tails, heads, then $C = 2$ and $M = 0$. If we flip tails, tails, tails, then $C = 0$ and $M = 1$. In effect, C counts the number of heads, and M indicates whether all the coins match.

Since each outcome uniquely determines C and M, we can regard them as functions mapping outcomes to numbers. For this experiment, the sample space is:

$$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Now C is a function that maps each outcome in the sample space to a number as follows:

$$\begin{array}{rclcrcl} C(HHH) &=& 3 & \qquad & C(THH) &=& 2 \\ C(HHT) &=& 2 & & C(THT) &=& 1 \\ C(HTH) &=& 2 & & C(TTH) &=& 1 \\ C(HTT) &=& 1 & & C(TTT) &=& 0. \end{array}$$

Similarly, M is a function mapping each outcome another way:

$$\begin{array}{rclcrcl} M(HHH) & = & 1 & \qquad & M(THH) & = & 0 \\ M(HHT) & = & 0 & & M(THT) & = & 0 \\ M(HTH) & = & 0 & & M(TTH) & = & 0 \\ M(HTT) & = & 0 & & M(TTT) & = & 1. \end{array}$$

So C and M are random variables.

19.1.1 Indicator Random Variables

An *indicator random variable* is a random variable that maps every outcome to either 0 or 1. Indicator random variables are also called *Bernoulli variables*. The random variable M is an example. If all three coins match, then $M = 1$; otherwise, $M = 0$.

Indicator random variables are closely related to events. In particular, an indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0. For example, the indicator M partitions the sample space into two blocks as follows:

$$\underbrace{HHH \quad TTT}_{M=1} \quad \underbrace{HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH}_{M=0}.$$

In the same way, an event E partitions the sample space into those outcomes in E and those not in E. So E is naturally associated with an indicator random variable, I_E, where $I_E(\omega) = 1$ for outcomes $\omega \in E$ and $I_E(\omega) = 0$ for outcomes $\omega \notin E$. This means that event E is the same as the event $[I_E = 1]$. For example the variable M above is really just the indicator variable I_E, where E is the event that all three coins match.

19.1.2 Random Variables and Events

There is a strong relationship between events and more general random variables as well. A random variable that takes on several values partitions the sample space into several blocks. For example, C partitions the sample space as follows:

$$\underbrace{TTT}_{C=0} \quad \underbrace{TTH \quad THT \quad HTT}_{C=1} \quad \underbrace{THH \quad HTH \quad HHT}_{C=2} \quad \underbrace{HHH}_{C=3}.$$

Each block is a subset of the sample space and is therefore an event. So the assertion that $C = 2$ defines the event

$$[C = 2] = \{THH, HTH, HHT\},$$

and this event has probability

$$\Pr[C = 2] = \Pr[THH] + \Pr[HTH] + \Pr[HHT] = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = 3/8.$$

Likewise $[M = 1]$ is the event $\{TTT, HHH\}$ and has probability $1/4$.

More generally, any assertion about the values of random variables defines an event. For example, the assertion that $C \leq 1$ defines

$$[C \leq 1] = \{TTT, TTH, THT, HTT\},$$

and so $\Pr[C \leq 1] = 1/2$.

Another example is the assertion that $C \cdot M$ is an odd number. If you think about it for a minute, you'll realize that this is an obscure way of saying that all three coins came up heads, namely,

$$[C \cdot M \text{ is odd}] = \{HHH\}.$$

19.2 Independence

The notion of independence carries over from events to random variables as well. Random variables R_1 and R_2 are *independent* iff for all x_1, x_2, the two events

$$[R_1 = x_1] \quad \text{and} \quad [R_2 = x_2]$$

are independent.

For example, are C and M independent? Intuitively, the answer should be "no." The number of heads C completely determines whether all three coins match; that is, whether $M = 1$. But, to verify this intuition, we must find some $x_1, x_2 \in \mathbb{R}$ such that:

$$\Pr[C = x_1 \text{ AND } M = x_2] \neq \Pr[C = x_1] \cdot \Pr[M = x_2].$$

One appropriate choice of values is $x_1 = 2$ and $x_2 = 1$. In this case, we have:

$$\Pr[C = 2 \text{ AND } M = 1] = 0 \neq \frac{1}{4} \cdot \frac{3}{8} = \Pr[M = 1] \cdot \Pr[C = 2].$$

The first probability is zero because we never have exactly two heads ($C = 2$) when all three coins match ($M = 1$). The other two probabilities were computed earlier.

On the other hand, let H_1 be the indicator variable for the event that the first flip is a Head, so

$$[H_1 = 1] = \{HHH, HTH, HHT, HTT\}.$$

Then H_1 is independent of M, since

$$\Pr[M = 1] = 1/4 = \Pr\big[M = 1 \mid H_1 = 1\big] = \Pr\big[M = 1 \mid H_1 = 0\big]$$
$$\Pr[M = 0] = 3/4 = \Pr\big[M = 0 \mid H_1 = 1\big] = \Pr\big[M = 0 \mid H_1 = 0\big]$$

This example is an instance of:

Lemma 19.2.1. *Two events are independent iff their indicator variables are independent.*

The simple proof is left to Problem 19.1.

Intuitively, the independence of two random variables means that knowing some information about one variable doesn't provide any information about the other one. We can formalize what "some information" about a variable R is by defining it to be the value of some quantity that depends on R. This intuitive property of independence then simply means that functions of independent variables are also independent:

Lemma 19.2.2. *Let R and S be independent random variables, and f and g be functions such that* $\text{domain}(f) = \text{codomain}(R)$ *and* $\text{domain}(g) = \text{codomain}(S)$. *Then $f(R)$ and $g(S)$ are independent random variables.*

The proof is another simple exercise left to Problem 19.36.

As with events, the notion of independence generalizes to more than two random variables.

Definition 19.2.3. Random variables $R_1, R_2, \ldots, R_n$ are *mutually independent* iff for all $x_1, x_2, \ldots, x_n$, the n events

$$[R_1 = x_1], [R_2 = x_2], \ldots, [R_n = x_n]$$

are mutually independent. They are *k-way independent* iff every subset of k of them are mutually independent.

Lemmas 19.2.1 and 19.2.2 both extend straightforwardly to k-way independent variables.

19.3 Distribution Functions

A random variable maps outcomes to values. The probability density function, $\text{PDF}_R(x)$, of a random variable R measures the probability that R takes the value x, and the closely related cumulative distribution function $\text{CDF}_R(x)$ measures the probability that $R \leq x$. Random variables that show up for different spaces of outcomes often wind up behaving in much the same way because they have the same probability of taking different values, that is, because they have the same pdf/cdf.

Definition 19.3.1. Let R be a random variable with codomain V. The *probability density function* of R is a function $\text{PDF}_R : V \to [0, 1]$ defined by:

$$\text{PDF}_R(x) ::= \begin{cases} \Pr[R = x] & \text{if } x \in \text{range}(R), \\ 0 & \text{if } x \notin \text{range}(R). \end{cases}$$

If the codomain is a subset of the real numbers, then the *cumulative distribution function* is the function $\text{CDF}_R : \mathbb{R} \to [0, 1]$ defined by:

$$\text{CDF}_R(x) ::= \Pr[R \leq x].$$

A consequence of this definition is that

$$\sum_{x \in \text{range}(R)} \text{PDF}_R(x) = 1.$$

This is because R has a value for each outcome, so summing the probabilities over all outcomes is the same as summing over the probabilities of each value in the range of R.

As an example, suppose that you roll two unbiased, independent, 6-sided dice. Let T be the random variable that equals the sum of the two rolls. This random variable takes on values in the set $V = \{2, 3, \ldots, 12\}$. A plot of the probability density function for T is shown in Figure 19.1. The lump in the middle indicates that sums close to seven are the most likely. The total area of all the rectangles is 1 since the dice must take on exactly one of the sums in $V = \{2, 3, \ldots, 12\}$.

The cumulative distribution function for T is shown in Figure 19.2: The height of the ith bar in the cumulative distribution function is equal to the *sum* of the heights of the leftmost i bars in the probability density function. This follows from the definitions of pdf and cdf:

$$\text{CDF}_R(x) = \Pr[R \leq x] = \sum_{y \leq x} \Pr[R = y] = \sum_{y \leq x} \text{PDF}_R(y).$$

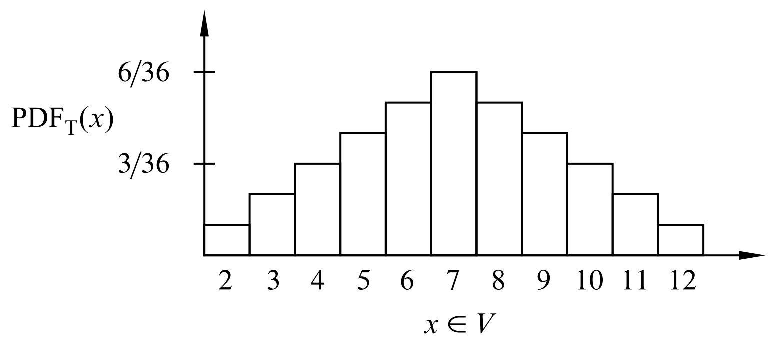


Figure 19.1 The probability density function for the sum of two 6-sided dice.

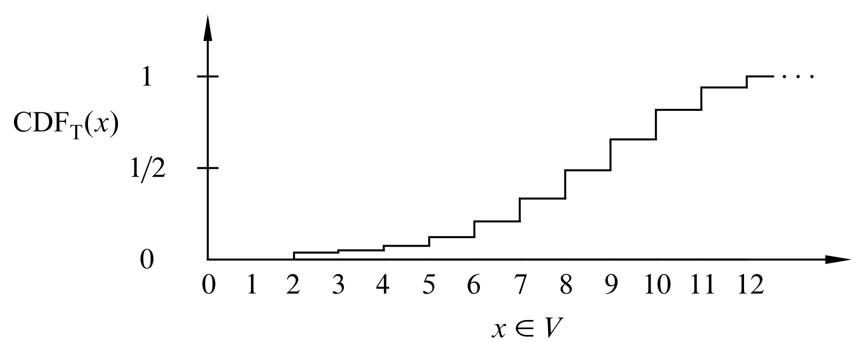


Figure 19.2 The cumulative distribution function for the sum of two 6-sided dice.

It also follows from the definition that

$$\lim_{x \to \infty} \mathrm{CDF}_R(x) = 1 \text{ and } \lim_{x \to -\infty} \mathrm{CDF}_R(x) = 0.$$

Both PDF_R and CDF_R capture the same information about R, so take your choice. The key point here is that neither the probability density function nor the cumulative distribution function involves the sample space of an experiment.

One of the really interesting things about density functions and distribution functions is that many random variables turn out to have the *same* pdf and cdf. In other words, even though R and S are different random variables on different probability spaces, it is often the case that

$$\mathrm{PDF}_R = \mathrm{PDF}_S.$$

In fact, some pdf's are so common that they are given special names. For example, the most important distributions in computer science arguably are the *Bernoulli distribution*, the *Uniform distribution*, the *Binomial distribution*, and the *Geometric distribution*. We look more closely at these common distributions in the next several sections.

19.3.1 Bernoulli Distributions

A Bernoulli distribution is the distribution function for a Bernoulli variable. Specifically, the *Bernoulli distribution* has a probability density function of the form $f_p : \{0, 1\} \to [0, 1]$ where

$$\begin{aligned} f_p(0) &= p, \quad \text{and} \\ f_p(1) &= q, \end{aligned}$$

for some $p \in [0, 1]$ with $q ::= 1 - p$. The corresponding cumulative distribution function is $F_p : \mathbb{R} \to [0, 1]$ where

$$F_p(x) ::= \begin{cases} 0 & \text{if } x < 0 \\ p & \text{if } 0 \le x < 1 \\ 1 & \text{if } 1 \le x. \end{cases}$$

19.3.2 Uniform Distributions

A random variable that takes on each possible value in its codomain with the same probability is said to be *uniform*. If the codomain V has n elements, then the *uniform distribution* has a pdf of the form

$$f : V \to [0, 1]$$

where

$$f(v) = \frac{1}{n}$$

for all $v \in V$.

If the elements of V in increasing order are $a_1, a_2, \ldots, a_n$, then the cumulative distribution function would be $F : \mathbb{R} \to [0, 1]$ where

$$F(x) ::= \begin{cases} 0 & \text{if } x < a_1 \\ k/n & \text{if } a_k \leq x < a_{k+1} \text{ for } 1 \leq k < n \\ 1 & \text{if } a_n \leq x. \end{cases}$$

Uniform distributions come up all the time. For example, the number rolled on a fair die is uniform on the set $\{1, 2, \ldots, 6\}$. An indicator variable is uniform when its pdf is $f_{1/2}$.

19.3.3 The Numbers Game

Enough definitions—let's play a game! We have two envelopes. Each contains an integer in the range $0, 1, \ldots, 100$, and the numbers are distinct. To win the game, you must determine which envelope contains the larger number. To give you a fighting chance, we'll let you peek at the number in one envelope selected at random. Can you devise a strategy that gives you a better than 50% chance of winning?

For example, you could just pick an envelope at random and guess that it contains the larger number. But this strategy wins only 50% of the time. Your challenge is to do better.

So you might try to be more clever. Suppose you peek in one envelope and see the number 12. Since 12 is a small number, you might guess that the number in the other envelope is larger. But perhaps we've been tricky and put small numbers in *both* envelopes. Then your guess might not be so good!

An important point here is that the numbers in the envelopes may *not* be random. We're picking the numbers and we're choosing them in a way that we think will defeat your guessing strategy. We'll only use randomization to choose the numbers if that serves our purpose: making you lose!

Intuition Behind the Winning Strategy

People are surprised when they first learn that there is a strategy that wins more than 50% of the time, regardless of what numbers we put in the envelopes.

Suppose that you somehow knew a number x that was in between the numbers in the envelopes. Now you peek in one envelope and see a number. If it is bigger than x, then you know you're peeking at the higher number. If it is smaller than x, then you're peeking at the lower number. In other words, if you know a number x between the numbers in the envelopes, then you are certain to win the game.

The only flaw with this brilliant strategy is that you do *not* know such an x. This sounds like a dead end, but there's a cool way to salvage things: try to *guess* x! There is some probability that you guess correctly. In this case, you win 100% of the time. On the other hand, if you guess incorrectly, then you're no worse off than before; your chance of winning is still 50%. Combining these two cases, your overall chance of winning is better than 50%.

Many intuitive arguments about probability are wrong despite sounding persuasive. But this one goes the other way: it may not convince you, but it's actually correct. To justify this, we'll go over the argument in a more rigorous way—and while we're at it, work out the optimal way to play.

Analysis of the Winning Strategy

For generality, suppose that we can choose numbers from the integer interval $[0..n]$. Call the lower number L and the higher number H.

Your goal is to guess a number x between L and H. It's simplest if x does not equal L or H, so you should select x at random from among the half-integers:

$$\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots, \frac{2n-1}{2}$$

But what probability distribution should you use?

The uniform distribution—selecting each of these half-integers with equal probability—turns out to be your best bet. An informal justification is that if we figured out that you were unlikely to pick some number—say $50\frac{1}{2}$—then we'd always put 50 and 51 in the envelopes. Then you'd be unlikely to pick an x between L and H and would have less chance of winning.

After you've selected the number x, you peek into an envelope and see some number T. If $T > x$, then you guess that you're looking at the larger number. If $T < x$, then you guess that the other number is larger.

All that remains is to determine the probability that this strategy succeeds. We can do this with the usual four step method and a tree diagram.

Step 1: Find the sample space.
You either choose x too low ($< L$), too high ($> H$), or just right ($L < x < H$). Then you either peek at the lower number ($T = L$) or the higher number ($T = H$). This gives a total of six possible outcomes, as show in Figure 19.3.

Step 2: Define events of interest.
The four outcomes in the event that you win are marked in the tree diagram.

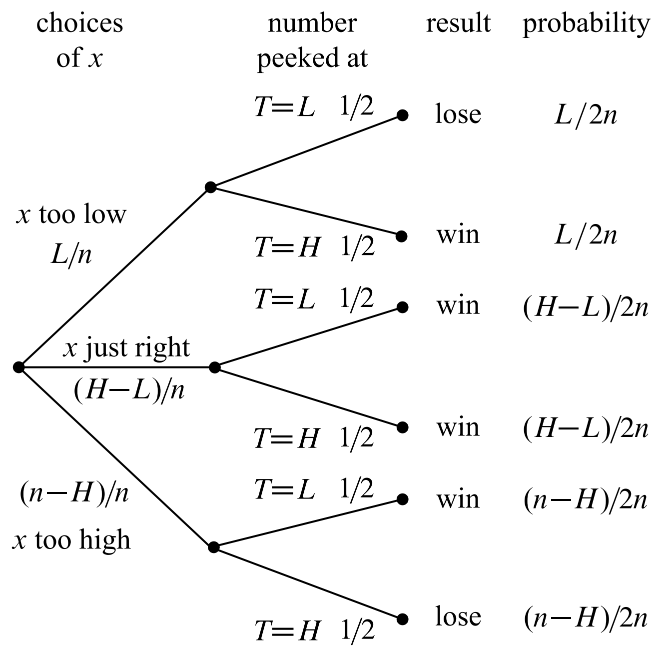


Figure 19.3 The tree diagram for the numbers game.

Step 3: Assign outcome probabilities.
First, we assign edge probabilities. Your guess x is too low with probability L/n, too high with probability $(n-H)/n$, and just right with probability $(H-L)/n$. Next, you peek at either the lower or higher number with equal probability. Multiplying along root-to-leaf paths gives the outcome probabilities.

Step 4: Compute event probabilities.
The probability of the event that you win is the sum of the probabilities of the four outcomes in that event:

$$\begin{aligned}
\Pr[\text{win}] &= \frac{L}{2n} + \frac{H-L}{2n} + \frac{H-L}{2n} + \frac{n-H}{2n} \\
&= \frac{1}{2} + \frac{H-L}{2n} \\
&\geq \frac{1}{2} + \frac{1}{2n}
\end{aligned}$$

The final inequality relies on the fact that the higher number H is at least 1 greater than the lower number L since they are required to be distinct.

Sure enough, you win with this strategy more than half the time, regardless of the numbers in the envelopes! So with numbers chosen from the range $0, 1, \ldots, 100$, you win with probability at least $1/2 + 1/200 = 50.5\%$. If instead we agree to stick to numbers $0, \ldots, 10$, then your probability of winning rises to 55%. By Las Vegas standards, those are great odds.

Randomized Algorithms

The best strategy to win the numbers game is an example of a *randomized algorithm*—it uses random numbers to influence decisions. Protocols and algorithms that make use of random numbers are very important in computer science. There are many problems for which the best known solutions are based on a random number generator.

For example, the most commonly-used protocol for deciding when to send a broadcast on a shared bus or Ethernet is a randomized algorithm known as *exponential backoff*. One of the most commonly-used sorting algorithms used in practice, called *quicksort*, uses random numbers. You'll see many more examples if you take an algorithms course. In each case, randomness is used to improve the probability that the algorithm runs quickly or otherwise performs well.

19.3.4 Binomial Distributions

The third commonly-used distribution in computer science is the *binomial distribution*. The standard example of a random variable with a binomial distribution is the number of heads that come up in n independent flips of a coin. If the coin is fair, then the number of heads has an *unbiased binomial distribution*, specified by the pdf $f_n : [0..n] \to [0, 1]$:

$$f_n(k) ::= \binom{n}{k} 2^{-n}.$$

This is because there are $\binom{n}{k}$ sequences of n coin tosses with exactly k heads, and each such sequence has probability 2^{-n}.

A plot of $f_{20}(k)$ is shown in Figure 19.4. The most likely outcome is $k = 10$ heads, and the probability falls off rapidly for larger and smaller values of k. The falloff regions to the left and right of the main hump are called the *tails of the distribution*.

In many fields, including Computer Science, probability analyses come down to getting small bounds on the tails of the binomial distribution. In the context of a problem, this typically means that there is very small probability that something *bad* happens, which could be a server or communication link overloading or a randomized algorithm running for an exceptionally long time or producing the wrong result.

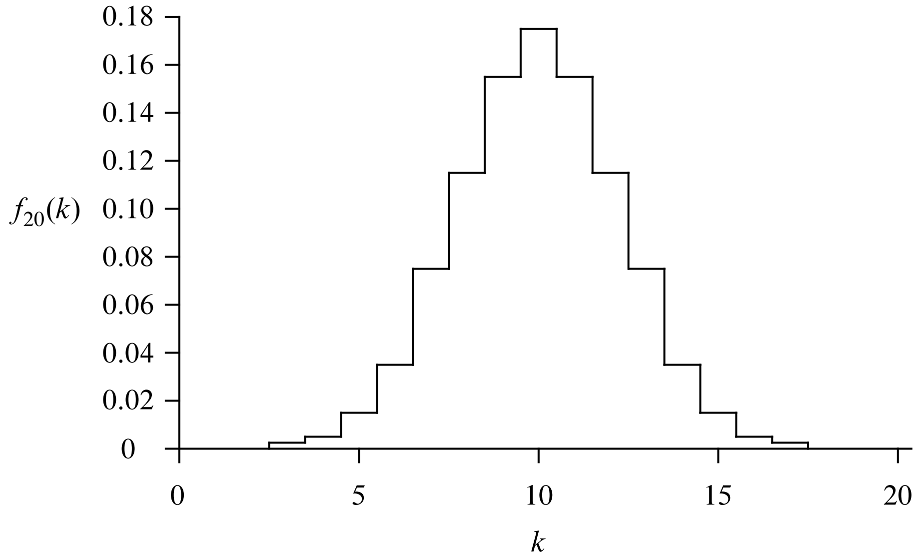


Figure 19.4 The pdf for the unbiased binomial distribution for $n = 20$, $f_{20}(k)$.

The tails do get small very fast. For example, the probability of flipping at most 25 heads in 100 tosses is less than 1 in 3,000,000. In fact, the tail of the distribution falls off so rapidly that the probability of flipping exactly 25 heads is nearly twice the probability of flipping exactly 24 heads *plus* the probability of flipping exactly 23 heads *plus* ... the probability of flipping no heads.

The General Binomial Distribution

If the coins are biased so that each coin is heads with probability p and tails with probability $q ::= 1 - p$, then the number of heads has a *general binomial density function* specified by the pdf $f_{n,p} : [0..n] \to [0, 1]$ where

$$f_{n,p}(k) = \binom{n}{k} p^k q^{n-k}. \tag{19.1}$$

for some $n \in \mathbb{N}^+$ and $p \in [0, 1]$. This is because there are $\binom{n}{k}$ sequences with k heads and $n - k$ tails, but now $p^k q^{n-k}$ is the probability of each such sequence.

For example, the plot in Figure 19.5 shows the probability density function $f_{n,p}(k)$ corresponding to flipping $n = 20$ independent coins that are heads with probability $p = 0.75$. The graph shows that we are most likely to get $k = 15$ heads, as you might expect. Once again, the probability falls off quickly for larger and smaller values of k.

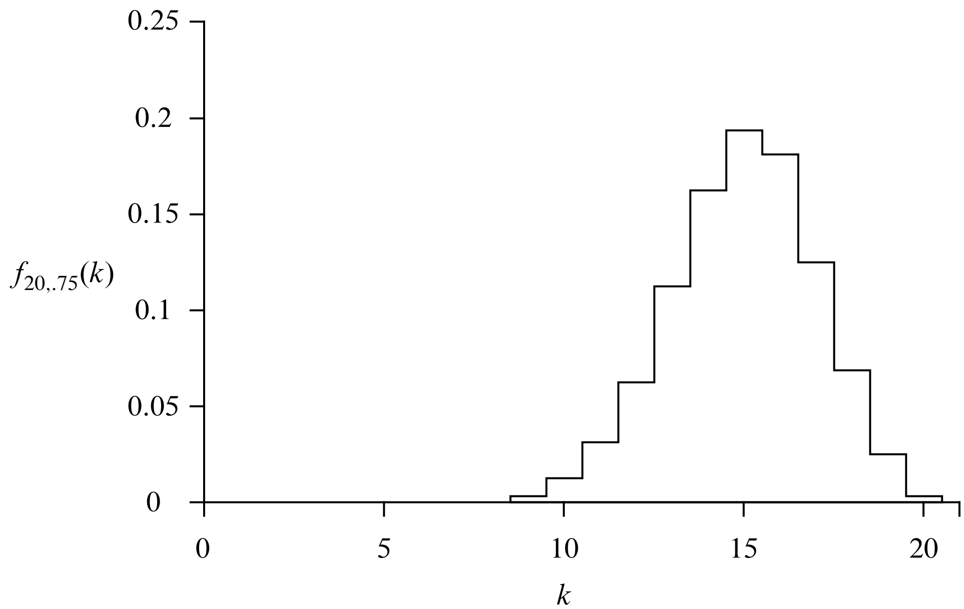


Figure 19.5 The pdf for the general binomial distribution $f_{n,p}(k)$ for $n = 20$ and $p = .75$.

19.4 Great Expectations

The *expectation* or *expected value* of a random variable in simple cases is just an average value. For example, the first thing you typically want to know when you see your grade on an exam is the average score of the class. This average score is the same as the expected score of a random student.

In general, the expected value of a random variable is the sum of all it possible values when each value is weighted according to its probability. To make this work, we need to be able to add values and multiply them by probabilities. This will certainly be possible if the values are real numbers; for technical reasons, we focus on *nonnegative* real values. Now we can define expected value formally as follows:

Definition 19.4.1. If R is a nonnegative real-valued random variable defined on a sample space $\mathcal{S}$, then the expectation of R is

$$\text{Ex}[R] ::= \sum_{\omega \in \mathcal{S}} R(\omega) \Pr[\omega]. \tag{19.2}$$

The expectation of a random variable is also known as its *mean*.

From now on, we will assume our *random variables are nonnegative real-valued* unless we explcitly say otherwise.

Let's work through some examples.

19.4.1 The Expected Value of a Uniform Random Variable

Rolling a 6-sided die provides an example of a uniform random variable. Let R be the value that comes up when you roll a fair 6-sided die. Then by (19.2), the expected value of R is

$$\text{Ex}[R] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

This calculation shows that the name "expected" value is a little misleading; the random variable might *never* actually take on that value. No one expects to roll a $3\frac{1}{2}$ on an ordinary die!

In general, if R_n is a random variable with a uniform distribution on $\{a_1, a_2, \ldots, a_n\}$, then the expectation of R_n is simply the average of the a_i's:

$$\text{Ex}[R_n] = \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

19.4.2 The Expected Value of a Reciprocal Random Variable

Define a random variable S to be the reciprocal of the value that comes up when you roll a fair 6-sided die. That is, $S = 1/R$ where R is the value that you roll. Now,

$$\text{Ex}[S] = \text{Ex}\left[\frac{1}{R}\right] = \frac{1}{1} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{49}{120}.$$

Notice that

$$\text{Ex}\left[1/R\right] \neq 1/\text{Ex}[R].$$

Assuming that these two quantities are equal is a common mistake.

19.4.3 The Expected Value of an Indicator Random Variable

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

Lemma 19.4.2. *If I_A is the indicator random variable for event A, then*

$$\text{Ex}[I_A] = \Pr[A].$$

Proof.

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

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

19.4.4 Alternate Definition of Expectation

There is another standard way to define expectation:

Theorem 19.4.3. *For any random variable R,*

$$\text{Ex}[R] = \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]. \tag{19.3}$$

The proof of Theorem 19.4.3, like many of the elementary proofs about expectation in this chapter, follows by regrouping of terms in equation (19.2):

Proof. Suppose R is defined on a sample space $\mathcal{S}$. Then,

$$\begin{aligned}\text{Ex}[R] &::= \sum_{\omega \in \mathcal{S}} R(\omega) \Pr[\omega] \\ &= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R = x]} R(\omega) \Pr[\omega] \\ &= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R = x]} x \Pr[\omega] && \text{(def of the event } [R = x]) \\ &= \sum_{x \in \text{range}(R)} x \left(\sum_{\omega \in [R = x]} \Pr[\omega] \right) && \text{(factoring } x \text{ from the inner sum)} \\ &= \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]. && \text{(def of } \Pr[R = x])\end{aligned}$$

The first equality follows because the events $[R = x]$ for $x \in \text{range}(R)$ partition the sample space $\mathcal{S}$, so summing over the outcomes in $[R = x]$ for $x \in \text{range}(R)$ is the same as summing over $\mathcal{S}$. ■

In general, equation (19.3) is more useful than the defining equation (19.2) for calculating expected values. It also 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, summing over all outcomes as in equation (19.2) allows easier proofs of some basic properties of expectation.

Notice that the order in which terms appear in the sums (19.3) and (19.2) is not specified, and the proof of Theorem 19.4.3—and lots of proofs below—involve regrouping the terms in sums. This is OK because of a well-known property of countable sums of nonnegative real numbers:

Theorem 19.4.4. *A countable sum of nonnegative real numbers converges to the same value, or else always diverges, regardless of the order in which the numbers are summed.*

In fact as long as reordering terms in the infinite sum (19.2) for expectation preserves convergence, we can allow random variables R taking negative as well as positive values. In this case, $\text{Ex}[R]$ will be well-defined and will have all the basic properties we establish below for nonnegative variables. But reordering does not preserve convergence for arbitrary sums of positive and negative values (see Problems 14.14 and 14.16), and there is no useful definition of the expectation for *arbitrary* real-valued random variables.

19.4.5 Conditional Expectation

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 19.4.5. The *conditional expectation* $\text{Ex}[R \mid A]$ of a random variable R given event A is:

$$\text{Ex}[R \mid A] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr\left[R = r \mid A\right]. \tag{19.4}$$

For example, we can compute the expected value of a roll of a fair die, given 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 (19.4),

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

Conditional expectation is useful in dividing complicated expectation calculations into simpler cases. We can find a desired expectation by calculating the conditional expectation in each simple case and averaging them, weighing each case by its probability.

For example, suppose that 49.6% 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 person? 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. Then

$$\begin{aligned}
\text{Ex}[H] &= \text{Ex}[H \mid M] \Pr[M] + \text{Ex}[H \mid F] \Pr[F] \\
&= (5 + 11/12) \cdot 0.496 + (5 + 5/12) \cdot (1 - 0.496) \\
&= 5.6646 \ldots.
\end{aligned}$$

which is a little less than 5' 8."

This method is justified by:

Theorem 19.4.6 (Law of Total Expectation)**.** *Let R be a random variable on a sample space $\mathcal{S}$, and suppose that $A_1, A_2, \ldots,$ is a partition of $\mathcal{S}$. Then*

$$\text{Ex}[R] = \sum_i \text{Ex}[R \mid A_i] \Pr[A_i].$$

Proof.

$$\begin{aligned}
\text{Ex}[R] &= \sum_{r \in \text{range}(R)} r \cdot \Pr[R = r] && \text{(by 19.3)} \\
&= \sum_r r \cdot \sum_i \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(Law of Total Probability)} \\
&= \sum_r \sum_i r \cdot \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(distribute constant } r\text{)} \\
&= \sum_i \sum_r r \cdot \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(exchange order of summation)} \\
&= \sum_i \Pr[A_i] \sum_r r \cdot \Pr\left[R = r \mid A_i\right] && \text{(factor constant } \Pr[A_i]\text{)} \\
&= \sum_i \Pr[A_i] \, \text{Ex}[R \mid A_i]. && \text{(Def 19.4.5 of cond. expectation)}
\end{aligned}$$

■

19.4.6 Geometric Distributions

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?

This will be easy to figure out using the Law of Total Expectation, Theorem 19.4.6. Specifically, we want to find $\text{Ex}[C]$ where C is the number of hours until the first crash. We'll do this by conditioning on whether or not the crash occurs in the first hour.

So define A to be the event that the system fails on the first step and $\overline{A}$ to be the complementary event that the system does not fail on the first step. Then the mean time to failure $\text{Ex}[C]$ is

$$\text{Ex}[C] = \text{Ex}[C \mid A]\Pr[A] + \text{Ex}[C \mid \overline{A}]\Pr[\overline{A}]. \tag{19.5}$$

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

$$\text{Ex}[C \mid A] = 1. \tag{19.6}$$

Since $\overline{A}$ is the condition that the system does *not* crash on the first step, conditioning on $\overline{A}$ is equivalent to taking a first step without failure and then starting over without conditioning. Hence,

$$\text{Ex}[C \mid \overline{A}] = 1 + \text{Ex}[C]. \tag{19.7}$$

Plugging (19.6) and (19.7) into (19.5):

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

Then, rearranging terms gives

$$1 = \text{Ex}[C] - q\,\text{Ex}[C] = p\,\text{Ex}[C],$$

and thus

$$\text{Ex}[C] = 1/p.$$

The general principle here is well-worth remembering.

Mean Time to Failure

If a system independently fails at each time step with probability p, then the expected number of steps up to the first failure is $1/p$.

So, 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 further example, suppose a couple insists on having children until they get a boy, then how many baby girls should they expect before their first boy? Assume for simplicity that there is a 50% chance that a child will be a boy and that the genders of siblings are mutually independent.

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 boy?” In this case, a crash corresponds to having a boy, so we should set $p = 1/2$. By the preceding analysis, the couple should expect a baby boy after having $1/p = 2$ children. Since the last of these will be a boy, they should expect just one girl. So even in societies where couples pursue this commitment to boys, the expected population will divide evenly between boys and girls.

There is a simple intuitive argument that explains the mean time to failure formula (19.8). Suppose the system is restarted after each failure. This makes the mean time to failure the same as the mean time between successive repeated failures. Now if the probability of failure at a given step is p, then after n steps we expect to have pn failures. Now the average number of steps between failures is, by definition, equal to $n/pn = 1/p$.

For the record, we’ll state a formal version of this result. A random variable like C that counts steps to first failure is said to have a *geometric distribution* with parameter p.

Definition 19.4.7. A random variable C has a *geometric distribution* with parameter p iff $\text{codomain}(C) = \mathbb{Z}^+$ and

$$\Pr[C = i] = q^{i-1} p.$$

Lemma 19.4.8. *If a random variable C has a geometric distribution with parameter p, then*

$$\text{Ex}[C] = \frac{1}{p}. \tag{19.8}$$

19.4.7 Expected Returns in Gambling Games

Some of the most interesting examples of expectation can be explained in terms of gambling games. For straightforward games where you win w dollars with probability p and you lose x dollars with probability $q = 1 - p$, it is easy to compute your *expected return* or *winnings*. It is simply

$$pw - qx \text{ dollars}.$$

For example, if you are flipping a fair coin and you win \$1 for heads and you lose \$1 for 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.

Splitting the Pot

We'll now look at a different game which is fair—but only on first analysis.

It's late on a Friday night in your neighborhood hangout when two new biker dudes, Eric and Nick, stroll over and propose a simple wager. Each player will put \$2 on the bar and secretly write "heads" or "tails" on their napkin. Then you will flip a fair coin. The \$6 on the bar will then be "split"—that is, be divided equally—among the players who correctly predicted the outcome of the coin toss. Pot splitting like this is a familiar feature in poker games, betting pools, and lotteries.

This sounds like a fair game, but after your regrettable encounter with strange dice (Section 17.3), you are definitely skeptical about gambling with bikers. So before agreeing to play, you go through the four-step method and write out the tree diagram to compute your expected return. The tree diagram is shown in Figure 19.6.

The "payoff" values in Figure 19.6 are computed by dividing the \$6 pot[1] 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 Eric guessed wrong, then your payoff is

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

In the case that everyone is wrong, you all agree to split the pot and so, again, your payoff is zero.

To compute your expected return, you use equation (19.3):

$$\begin{aligned}
\text{Ex}[\text{payoff}] &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} \\
&\qquad + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + 0 \cdot \frac{1}{8} \\
&= 0.
\end{aligned}$$

[1] The money invested in a wager is commonly referred to as the *pot*.

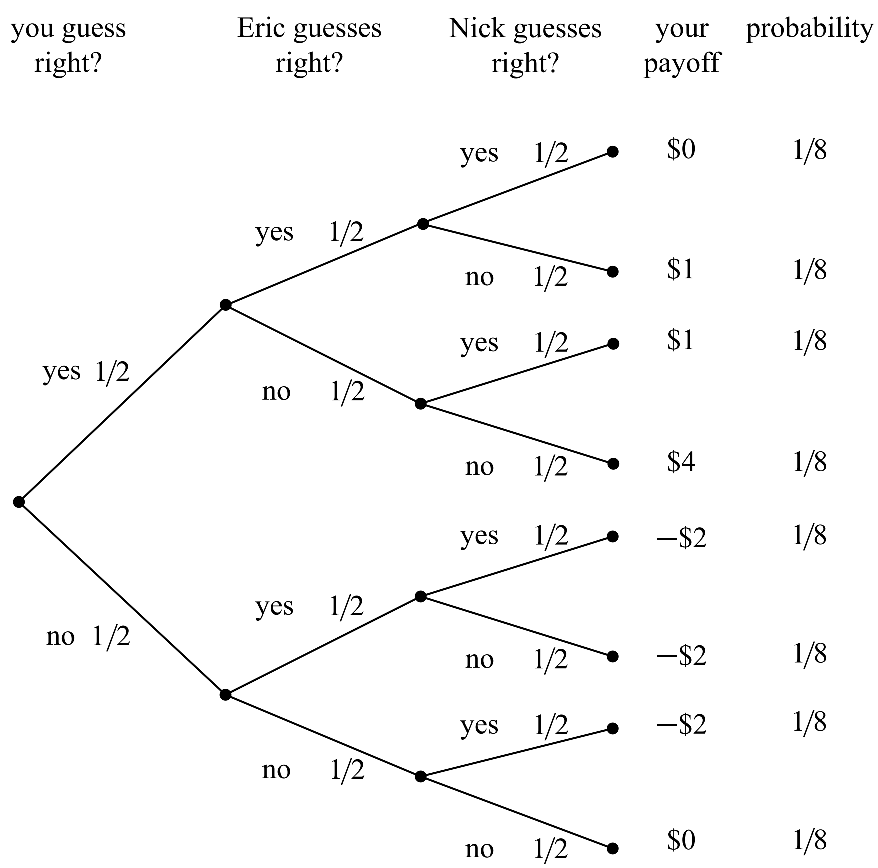


Figure 19.6 The tree diagram for the game where three players each wager $2 and then guess the outcome of a fair coin toss. The winners split the pot.

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.

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. As Nick and Eric are consoling you on your “bad luck,” you do a back-of-the-envelope calculation and decide that the probability of losing \$500 in 1000 fair \$2 wagers is very, very small.

Now it is possible of course that you are very, very unlucky. But it is more likely that something fishy is going on. Somehow the tree diagram in Figure 19.6 is not a good model of the game.

The “something” that’s fishy is the opportunity that Nick and Eric have to collude against you. The fact that the coin flip is fair certainly means that each of Nick and Eric can only guess the outcome of the coin toss with probability $1/2$. But when you look back at the previous 1000 bets, you notice that Eric and Nick never made the same guess. In other words, Nick always guessed “tails” when Eric guessed “heads,” and vice-versa. Modelling this fact now results in a slightly different tree diagram, as shown in Figure 19.7.

The payoffs for each outcome are the same in Figures 19.6 and 19.7, 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. Since Nick and Eric are always guessing differently, one of them will always get a share of the pot. As you might imagine, this is not good for you!

When we use equation (19.3) to compute your expected return in the collusion scenario, we find that

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

So watch out for these biker dudes! By colluding, Nick and Eric have made it so that you expect to lose \$.50 every time you play. No wonder you lost \$500 over the course of 1000 wagers.

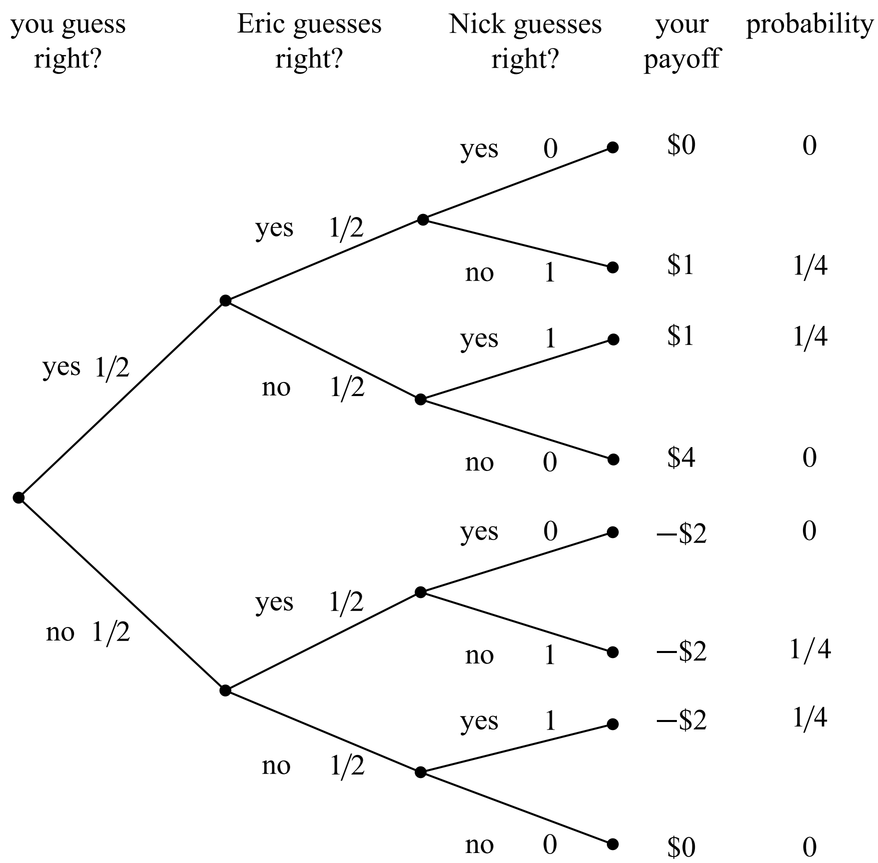


Figure 19.7 The revised tree diagram reflecting the scenario where Nick always guesses the opposite of Eric.

How to Win the Lottery

Similar opportunities to collude arise in many betting games. For example, consider the typical weekly football betting pool, where each participant wagers \$10 and the participants that pick the most games correctly split a large pot. The pool seems fair if you think of it as in Figure 19.6. But, in fact, if two or more players collude by guessing differently, they can get an "unfair" advantage at your expense!

In some cases, the collusion is inadvertent and you can profit from it. For example, many years ago, a former MIT Professor of Mathematics named Herman Chernoff figured out a way to make money by playing the state lottery. This was surprising since the state usually takes a large share of the wagers before paying the winners, and so the expected return from a lottery ticket is typically pretty poor. So how did Chernoff find a way to make money? It turned out to be easy!

In a typical state 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 states divides 1/2 of the money collected among the people who guessed correctly and spends the other half redecorating the governor's residence.

This is a lot like the game you played with Nick and Eric, 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; they pick the same numbers not on purpose as in the previous game with Nick and Eric, but based on the Red Sox winning average or today's date. The result is as though the players were intentionally colluding 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! By analyzing actual state lottery data, he determined that he could win an average of 7 cents on the dollar. In other words, his expected return was not $-\$.50$ as you might think, but $+\$.07$.[2] Inadvertent collusion often arises in betting pools and is a phenomenon that you can take advantage of.

[2] Most lotteries now offer randomized tickets to help smooth out the distribution of selected sequences.

19.5 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 19.5.1. *For any random variables R_1 and R_2,*

$$\operatorname{Ex}[R_1 + R_2] = \operatorname{Ex}[R_1] + \operatorname{Ex}[R_2].$$

Proof. Let $T ::= R_1 + R_2$. The proof follows straightforwardly by rearranging terms in equation (19.2) in the definition of expectation:

$$\begin{aligned}
\operatorname{Ex}[T] &::= \sum_{\omega \in \mathcal{S}} T(\omega) \cdot \Pr[\omega] \\
&= \sum_{\omega \in \mathcal{S}} (R_1(\omega) + R_2(\omega)) \cdot \Pr[\omega] && \text{(def of } T) \\
&= \sum_{\omega \in \mathcal{S}} R_1(\omega) \Pr[\omega] + \sum_{\omega \in \mathcal{S}} R_2(\omega) \Pr[\omega] && \text{(rearranging terms)} \\
&= \operatorname{Ex}[R_1] + \operatorname{Ex}[R_2]. && \text{(by (19.2))}
\end{aligned}$$

■

Essentially the same proof implies:

Theorem 19.5.2. *For random variables R_1, R_2 and constants $a_1, a_2 \in \mathbb{R}$,*

$$\operatorname{Ex}[a_1 R_1 + a_2 R_2] = a_1 \operatorname{Ex}[R_1] + a_2 \operatorname{Ex}[R_2].$$

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

Corollary 19.5.3 (Linearity of Expectation)**.** *For any random variables $R_1, \ldots, R_k$ and constants $a_1, \ldots, a_k \in \mathbb{R}$,*

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

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.

As an example, let's compute the expected value of the sum of two fair dice.

19.5.1 Expected Value of Two Dice

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

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:

$$\text{Ex}[R_1 + R_2] = \text{Ex}[R_1] + \text{Ex}[R_2] = 3.5 + 3.5 = 7.$$

Assuming that the dice were independent, we could use a tree diagram to prove that this expected sum is seven, but this would be a bother since there are 36 cases. And without assuming independence, it's not apparent how to apply the tree diagram approach at all. But notice that we did *not* have to assume that the two dice were independent. For example, suppose the roll of the second die was forced to match the roll of the first die. Then the expected sum of two dice remains equal to seven because the second die is still fair.

19.5.2 Sums of Indicator Random Variables

Linearity of expectation is especially useful when you have a sum of indicator random variables. As an example, suppose 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 using equation (19.2) or (19.3). But linearity of expectation lets us sidestep this issue.

We'll use a standard, useful trick to apply linearity, namely, we'll 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 the ith man gets his own hat, and $G_i = 0$ otherwise. The number of men that get their own hat is then the sum of these indicator random variables:

$$G = G_1 + G_2 + \cdots + G_n. \tag{19.9}$$

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 again, we don't need to worry about this dependence, since linearity holds regardless.

Since G_i is an indicator random variable, we know from Lemma 19.4.2 that

$$\text{Ex}[G_i] = \Pr[G_i = 1] = 1/n. \tag{19.10}$$

By Linearity of Expectation and equation (19.9), this means that

$$\begin{aligned}
\text{Ex}[G] &= \text{Ex}[G_1 + G_2 + \cdots + G_n] \\
&= \text{Ex}[G_1] + \text{Ex}[G_2] + \cdots + \text{Ex}[G_n] \\
&= \overbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}^{n} \\
&= 1.
\end{aligned}$$

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, regardless of the number of men with hats!

More generally, Linearity of Expectation provides a very good method for computing the expected number of events that will happen.

Theorem 19.5.4. *Given any collection of events $A_1, A_2, \ldots, A_n$, the expected number of events that will occur is*

$$\sum_{i=1}^{n} \Pr[A_i].$$

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 random variable for A_i, where $R_i(\omega) = 1$ if $w \in A_i$ and $R_i(\omega) = 0$ if $w \notin A_i$. Let $R = R_1 + R_2 + \cdots + R_n$. Then

$$\begin{aligned}
\text{Ex}[R] &= \sum_{i=1}^{n} \text{Ex}[R_i] && \text{(by Linearity of Expectation)} \\
&= \sum_{i=1}^{n} \Pr[R_i = 1] && \text{(by Lemma 19.4.2)} \\
&= \sum_{i=1}^{n} \Pr[A_i]. && \text{(def of indicator variable)}
\end{aligned}$$

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.

19.5.3 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?

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

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

Applying equation (19.3), this means that

$$\text{Ex}[J] = \sum_{k=0}^{n} k \Pr[J = k] = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}. \tag{19.11}$$

This sum looks a tad nasty, but linearity of expectation leads to an easy derivation of a simple closed form. We just express J as a sum of indicator random variables, which is easy. Namely, let J_i be the indicator random variable for the ith coin coming up heads, that is,

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

Then the number of heads is simply

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

By Theorem 19.5.4,

$$\text{Ex}[J] = \sum_{i=1}^{n} \Pr[J_i] = pn. \tag{19.12}$$

That really was easy. If we flip n mutually independent 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 19.5.4 do not assume any independence.

If you are not yet convinced that Linearity of Expectation and Theorem 19.5.4 are powerful tools, consider this: without even trying, we have used them to prove a complicated looking identity, namely,

$$\sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = pn, \tag{19.13}$$

which follows by combining equations (19.11) and (19.12) (see also Exercise 19.31).

The next section has an even more convincing illustration of the power of linearity to solve a challenging problem.

19.5.4 The Coupon Collector Problem

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

There are different colored Racin' Rocket cars. The color of car awarded to us by the kind server at the Taco Bell register appears to be selected uniformly and independently at random. What is the expected number of kid's meals that we must purchase in order to acquire at least one of each color 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? 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 colors 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:

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

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.

In the general case there are n colors of Racin' Rockets that we're collecting. 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 rule. This means that

$$\text{Ex}[X_k] = \frac{n}{n-k}.$$

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

$$\begin{aligned}
\text{Ex}[T] &= \text{Ex}[X_0 + X_1 + \cdots + X_{n-1}] \\
&= \text{Ex}[X_0] + \text{Ex}[X_1] + \cdots + \text{Ex}[X_{n-1}] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \cdots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1} \\
&= n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&= n\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} + \frac{1}{n}\right) \\
&= nH_n \qquad (19.14) \\
&\sim n \ln n.
\end{aligned}$$

Cool! It's those Harmonic Numbers again.

We can use equation (19.14) to answer some concrete questions. For example, the 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:

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

19.5.5 Infinite Sums

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

Theorem 19.5.5 (Linearity of Expectation)**.** *Let $R_0, R_1, \ldots,$ be random variables such that*

$$\sum_{i=0}^{\infty} \text{Ex}[\,|R_i|\,]$$

converges. Then

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

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:

$$\begin{aligned}
\sum_{i=0}^{\infty} \text{Ex}[R_i] &= \sum_{i=0}^{\infty} \sum_{s \in \mathcal{S}} R_i(s) \cdot \Pr[s] && \text{(Def. 19.4.1)} \\
&= \sum_{s \in \mathcal{S}} \sum_{i=0}^{\infty} R_i(s) \cdot \Pr[s] && \text{(exchanging order of summation)} \\
&= \sum_{s \in \mathcal{S}} \left[\sum_{i=0}^{\infty} R_i(s)\right] \cdot \Pr[s] && \text{(factoring out } \Pr[s]) \\
&= \sum_{s \in \mathcal{S}} T(s) \cdot \Pr[s] && \text{(Def. of } T) \\
&= \text{Ex}[T] && \text{(Def. 19.4.1)} \\
&= \text{Ex}\left[\sum_{i=0}^{\infty} R_i\right]. && \text{(Def. of } T). \quad \blacksquare
\end{aligned}$$

19.5.6 A Gambling Paradox

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.

The casino gets its advantage from the green slots, which make the probability of both red and black each less than 1/2. In the US, a roulette wheel has 2 green slots among 18 black and 18 red slots, so the probability of red is $18/38 \approx 0.473$. In Europe, where roulette wheels have only 1 green slot, the odds for red are a little better—that is, $18/37 \approx 0.486$—but still less than even.

Of course you can't expect to win playing roulette, even if you had the good fortune to gamble against a *fair* roulette wheel. To prove this, note that with a fair wheel, you are equally likely win or lose each bet, so your expected win on any spin is zero. Therefore if you keep betting, your expected win is the sum of your expected wins on each bet: still zero.

Even so, gamblers regularly try to develop betting strategies to win at roulette despite the bad odds. A well known strategy of this kind is *bet doubling*, where you bet, say, \$10 on red and keep doubling the bet until a red comes up. This means you stop playing if red comes up on the first spin, and you leave the casino with a \$10 profit. If red does not come up, you bet \$20 on the second spin. Now if the second spin comes up red, you get your \$20 bet plus \$20 back and again walk away with a net profit of $\$20 - 10 = \$10$. If red does not come up on the second spin, you next bet \$40 and walk away with a net win of $\$40 - 20 - 10 = \$10$ if red comes up on on the third spin, and so on.

Since we've reasoned that you can't even win against a fair wheel, this strategy against an unfair wheel shouldn't work. But wait a minute! There is a 0.486 probability of red appearing on each spin of the wheel, so the mean time until a red occurs is less than three. What's more, red will come up *eventually* with probability one, and as soon as it does, you leave the casino \$10 ahead. In other words, by bet doubling you are *certain* to win \$10, and so your expectation is \$10, not zero!

Something's wrong here.

19.5.7 Solution to the Paradox

The argument claiming the expectation is zero against a fair wheel is flawed by an implicit, invalid use of linearity of expectation for an infinite sum.

To explain this carefully, let B_n be the number of dollars you win on your nth bet, where B_n is defined to be zero if red comes up before the nth spin of the wheel. Now the dollar amount you win in any gambling session is

$$\sum_{n=1}^{\infty} B_n,$$

and your expected win is

$$\text{Ex}\left[\sum_{n=1}^{\infty} B_n\right]. \tag{19.15}$$

Moreover, since we're assuming the wheel is fair, it's true that $\text{Ex}[B_n] = 0$, so

$$\sum_{n=1}^{\infty} \text{Ex}[B_n] = \sum_{n=1}^{\infty} 0 = 0. \tag{19.16}$$

The flaw in the argument that you can't win is the implicit appeal to linearity of expectation to conclude that the expectation (19.15) equals the sum of expectations in (19.16). This is a case where linearity of expectation fails to hold—even though the expectation (19.15) is 10 and the sum (19.16) of expectations converges. The problem is that the expectation of the sum of the absolute values of the bets diverges, so the condition required for infinite linearity fails. In particular, under bet doubling your nth bet is $10 \cdot 2^{n-1}$ dollars while the probability that you will make an nth bet is 2^{-n}. So

$$\text{Ex}[|B_n|] = 10 \cdot 2^{n-1} 2^{-n} = 5.$$

Therefore the sum

$$\sum_{n=1}^{\infty} \text{Ex}[|B_n|] = 5 + 5 + 5 + \cdots$$

diverges rapidly.

So the presumption that you can't beat a fair game, and the argument we offered to support this presumption, are mistaken: by bet doubling, you can be sure to walk away a winner. Probability theory has led to an apparently absurd conclusion.

But probability theory shouldn't be rejected because it leads to this absurd conclusion. If you only had a finite amount of money to bet with—say enough money to make k bets before going bankrupt—then it would be correct to calculate your expection by summing $B_1 + B_2 + \cdots + B_k$, and your expectation would be zero for the fair wheel and negative against an unfair wheel. In other words, in order to follow the bet doubling strategy, you need to have an infinite bankroll. So it's absurd to assume you could actually follow a bet doubling strategy, and we needn't be concerned when an absurd assumption leads to an absurd conclusion.

19.5.8 Expectations of Products

While the expectation of a sum is the sum of the expectations, the same is usually not true for products. For example, suppose that we roll a fair 6-sided die and denote the outcome with the random variable R. Does $\text{Ex}[R \cdot R] = \text{Ex}[R] \cdot \text{Ex}[R]$?

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

$$\begin{aligned}\text{Ex}\left[R^2\right] &= \sum_{\omega\in\mathcal{S}} R^2(\omega)\Pr[w] = \sum_{i=1}^{6} i^2 \cdot \Pr[R_i = i] \\ &= \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} = 15\ 1/6 \neq 12\ 1/4.\end{aligned}$$

That is,

$$\text{Ex}[R \cdot R] \neq \text{Ex}[R] \cdot \text{Ex}[R].$$

So the expectation of a product is not always equal to the product of the expectations.

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

Theorem 19.5.6. *For any two* independent *random variables R_1, R_2,*

$$\text{Ex}[R_1 \cdot R_2] = \text{Ex}[R_1] \cdot \text{Ex}[R_2].$$

The proof follows by rearrangement of terms in the sum that defines $\text{Ex}[R_1 \cdot R_2]$. Details appear in Problem 19.29.

Theorem 19.5.6 extends routinely to a collection of mutually independent variables.

Corollary 19.5.7. *[Expectation of Independent Product]*
If random variables $R_1, R_2, \ldots, R_k$ are mutually independent, then

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

19.6 Really Great Expectations

Making independent tosses of a fair coin until some desired pattern comes up is a simple process you should feel in some command of by now, right? So how about a bet about the simplest such process—tossing until a head comes up? Ok, you're wary of betting with us, but how about this: we'll let *you set the odds*.

19.6.1 Repeating Yourself

Here's the bet: you make independent tosses of a fair coin until a head comes up. Then you will repeat the process. If a second head comes up in the same or fewer tosses than the first, you have to start over yet again. You keep starting over until you finally toss a run of tails longer than your first one. The payment rules are that you will pay me 1 cent each time you start over. When you win by finally getting a run of tails longer than your first one, I will pay you some generous amount. Notice by the way that you're certain to win—whatever your initial run of tails happened to be, a longer run will eventually occur again with probability 1!

For example, if your first tosses are TTTH, then you will keep tossing until you get a run of 4 tails. So your winning flips might be

$$\texttt{TTTHTHTTHHTTHTHTTTHTHHHTTTT}.$$

In this run there are 10 heads, which means you had to start over 9 times. So you would have paid me 9 cents by the time you finally won by tossing 4 tails. Now you've won, and I'll pay you generously —how does 25 cents sound? Maybe you'd rather have $1? How about $1000?

Of course there's a trap here. Let's calculate your expected winnings.

Suppose your initial run of tails had length k. After that, each time a head comes up, you have to start over and try to get $k+1$ tails in a row. If we regard your getting $k+1$ tails in a row as a "failed" try, and regard your having to start over because a head came up too soon as a "successful" try, then the number of times you have to start over is the number of tries till the first failure. So the expected number of tries will be the mean time to failure, which is 2^{k+1}, because the probability of tossing $k+1$ tails in a row is $2^{-(k+1)}$.

Let T be the length of your initial run of tails. So $T = k$ means that your initial tosses were $\texttt{T}^k\texttt{H}$. Let R be the number of times you repeat trying to beat your original run of tails. The number of cents you expect to finish with is the number of cents in my generous payment minus $\text{Ex}[R]$. It's now easy to calculate $\text{Ex}[R]$ by conditioning on the value of T:

$$\text{Ex}[R] = \sum_{k\in\mathbb{N}} \text{Ex}[R \mid T = k]\cdot\Pr[T = k] = \sum_{k\in\mathbb{N}} 2^{k+1}\cdot 2^{-(k+1)} = 1+1+1+\cdots = \infty.$$

So you can expect to pay me an infinite number of cents before winning my "generous" payment. No amount of generosity can make this bet fair! In fact this particular example is a special case of an astonishingly general one: the expected waiting time for *any* random variable to achieve a larger value remains infinite.

Problems for Section 19.2

Practice Problems

Problem 19.1.
Let I_A and I_B be the indicator variables for events A and B. Prove that I_A and I_B are independent iff A and B are independent.

Hint: Let $A^1 ::= A$ and $A^0 ::= \overline{A}$, so the event $[I_A = c]$ is the same as A^c for $c \in \{0, 1\}$; likewise for B^1, B^0.

Homework Problems

Problem 19.2.
Let R, S and T be random variables with the same codomain V.

(a) Suppose R is uniform—that is,

$$\Pr[R = b] = \frac{1}{|V|},$$

for all $b \in V$—and R is independent of S. Originally this text had the following argument:

> The probability that $R = S$ is the same as the probability that R takes whatever value S happens to have, therefore
>
> $$\Pr[R = S] = \frac{1}{|V|}. \tag{19.17}$$

Are you convinced by this argument? Write out a careful proof of (19.17).

Hint: The event $[R = S]$ is a disjoint union of events

$$[R = S] = \bigcup_{b \in V} [R = b \text{ AND } S = b].$$

(b) Let $S \times T$ be the random variable giving the values of S and T.[3] Now suppose R has a uniform distribution, and R is independent of $S \times T$. How about this argument?

[3]That is, $S \times T : \mathcal{S} \to V \times V$ where

$$(S \times T)(\omega) ::= (S(\omega), T(\omega))$$

for every outcome $\omega \in \mathcal{S}$.

> The probability that $R = S$ is the same as the probability that R equals the first coordinate of whatever value $S \times T$ happens to have, and this probability remains equal to $1/|V|$ by independence. Therefore the event $[R = S]$ is independent of $[S = T]$.

Write out a careful proof that $[R = S]$ is independent of $[S = T]$.

(c) Let $V = \{1, 2, 3\}$ and (R, S, T) take the following triples of values with equal probability,

$$(1, 1, 1), (2, 1, 1), (1, 2, 3), (2, 2, 3), (1, 3, 2), (2, 3, 2).$$

Verify that

1. R is independent of $S \times T$,
2. The event $[R = S]$ is not independent of $[S = T]$.
3. S and T have a uniform distribution.

Problem 19.3.
Let R, S and T be mutually independent indicator variables.

In general, the event that $S = T$ is not independent of $R = S$. We can explain this intuitively as follows: suppose for simplicity that S is uniform, that is, equally likely to be 0 or 1. This implies that S is equally likely as not to equal R, that is $\Pr[R = S] = 1/2$; likewise, $\Pr[S = T] = 1/2$.

Now suppose further that both R and T are more likely to equal 1 than to equal 0. This implies that $R = S$ makes it more likely than not that $S = 1$, and knowing that $S = 1$, makes it more likely than not that $S = T$. So knowing that $R = S$ makes it more likely than not that $S = T$, that is, $\Pr\left[S = T \mid R = S\right] > 1/2$.

Now prove rigorously (without any appeal to intuition)

Lemma 19.6.1. *Events $[R = S]$ and $[S = T]$ are independent iff either R is uniform*[4]*, or T is uniform, or S is constant*[5].

[4] That is, $\Pr[R = 1] = 1/2$.
[5] That is, $\Pr[S = 1]$ is one or zero.

Problems for Section 19.3

Practice Problems

Problem 19.4.
Suppose R, S and T be mutually independent random variables on the same probability space with uniform distribution on the range $\{1, 2, 3\}$.

Let $M = \max\{R, S, T\}$. Compute the values of the probability density function PDF_M of M.

Class Problems

Guess the Bigger Number Game

Team 1:

- Write two different integers between 0 and 7 on separate pieces of paper.
- Put the papers face down on a table.

Team 2:

- Turn over one paper and look at the number on it.
- Either stick with this number or switch to the other (unseen) number.

Team 2 wins if it chooses the larger number; else, Team 1 wins.

Problem 19.5.
The analysis in Section 19.3.3 implies that Team 2 has a strategy that wins 4/7 of the time no matter how Team 1 plays. Can Team 2 do better? The answer is "no," because Team 1 has a strategy that guarantees that it wins at least 3/7 of the time, no matter how Team 2 plays. Describe such a strategy for Team 1 and explain why it works.

Problem 19.6.
Suppose you have a biased coin that has probability p of flipping heads. Let J be the number of heads in n independent coin flips. So J has the general binomial

distribution:

$$\text{PDF}_J(k) = \binom{n}{k} p^k q^{n-k}$$

where $q ::= 1 - p$.

(a) Show that

$$\begin{aligned} \text{PDF}_J(k-1) &< \text{PDF}_J(k) && \text{for } k < np + p, \\ \text{PDF}_J(k-1) &> \text{PDF}_J(k) && \text{for } k > np + p. \end{aligned}$$

(b) Conclude that the maximum value of PDF_J is asymptotically equal to

$$\frac{1}{\sqrt{2\pi npq}}.$$

Hint: For the asymptotic estimate, it's ok to assume that np is an integer, so by part (a), the maximum value is $\text{PDF}_J(np)$. Use Stirling's Formula.

Problem 19.7.

Let $R_1, R_2, \ldots, R_m$, be mutually independent random variables with uniform distribution on $[1..n]$. Let $M ::= \max\{R_i \mid i \in [1..m]\}$.

(a) Write a formula for $\text{PDF}_M(1)$.

(b) More generally, write a formula for $\Pr[M \leq k]$.

(c) For $k \in [1..n]$, write a formula for $\text{PDF}_M(k)$ in terms of expressions of the form "$\Pr[M \leq j]$" for $j \in [1..n]$.

Homework Problems

Problem 19.8.

An over-caffeinated sailor of Tech Dinghy wanders along Seaside Boulevard. In each step, the sailor randomly moves one unit left or right with equal probability.

We let the sailor's initial position be designated location zero, with successive positions to the right labelled 1,2,..., and positions to the left labelled -1,-2,.... Let L_t be the random variable giving the sailor's location after t steps. Before he starts, the sailor is known to be at location zero, so

$$\text{PDF}_{L_0}(n) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

After one step, the sailor is equally likely to be at location 1 or -1, so

$$\text{PDF}_{L_1}(n) = \begin{cases} 1/2 & \text{if } n = \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Give the distributions PDF_{L_t} for $t = 2, 3, 4$ by filling in the table of probabilities below, where omitted entries are 0. For each row, write all the nonzero entries so they have the same denominator.

	location								
	-4	-3	-2	-1	0	1	2	3	4
initially					1				
after 1 step				$1/2$	0	$1/2$			
after 2 steps			?	?	?	?	?		
after 3 steps		?	?	?	?	?	?	?	
after 4 steps	?	?	?	?	?	?	?	?	?

(b) Help the staff of the Sailing Pavilion locate the sailor by answering the following questions. Provide your derivations and reasoning.

(i) What is the final location of a t-step walk that moves right exactly i times?

(ii) How many different length-t walks are there that end at that location?

(iii) What is the probability that the sailor ends at this location?

(iv) Let $B_t ::= (L_t + t)/2$. Conclude that B_t has an unbiased binomial distribution.

Problems for Section 19.4

Practice Problems

Problem 19.9.
Bruce Lee, on a movie that didn't go public, is practicing by breaking 5 boards with his fists. He is able to break a board with probability 0.8—he is practicing with his left fist, that's why it's not 1—and he breaks each board independently.

(a) What is the probability that Bruce breaks exactly 2 out of the 5 boards that are placed before him?

(b) What is the probability that Bruce breaks at most 3 out of the 5 boards that are placed before him?

(c) What is the expected number of boards Bruce will break?

Problem 19.10.
A news article reporting on the departure of a school official from California to Alabama dryly commented that this move would raise the average IQ in both states. Explain.

Class Problems

Problem 19.11.
Here's a dice game with maximum payoff k: make three independent rolls of a fair die, and if you roll a six

- no times, then you lose 1 dollar;
- exactly once, then you win 1 dollar;
- exactly twice, then you win 2 dollars;
- all three times, then you win k dollars.

For what value of k is this game fair?[6]

Problem 19.12. (a) Suppose we flip a fair coin and let $N_{\texttt{TT}}$ be the number of flips until the first time two consecutive Tails appear. What is $\text{Ex}[N_{\texttt{TT}}]$?

Hint: Let D be the tree diagram for this process. Explain why D can be described by the tree in Figure 19.8. Use the **Law of Total Expectation** 19.4.6.

(b) Let $N_{\texttt{TH}}$ be the number of flips until a Tail immediately followed by a Head comes up. What is $\text{Ex}[N_{\texttt{TH}}]$?

(c) Suppose we now play a game: flip a fair coin until either TT or TH occurs. You win if TT comes up first, and lose if TH comes up first. Since TT takes 50% longer on average to turn up, your opponent agrees that he has the advantage. So you tell him you're willing to play if you pay him \$5 when he wins, and he pays you with a mere 20% premium—that is \$6—when you win.

If you do this, you're sneakily taking advantage of your opponent's untrained intuition, since you've gotten him to agree to unfair odds. What is your expected profit per game?

[6] This game is actually offered in casinos with $k = 3$, where it is called Carnival Dice.

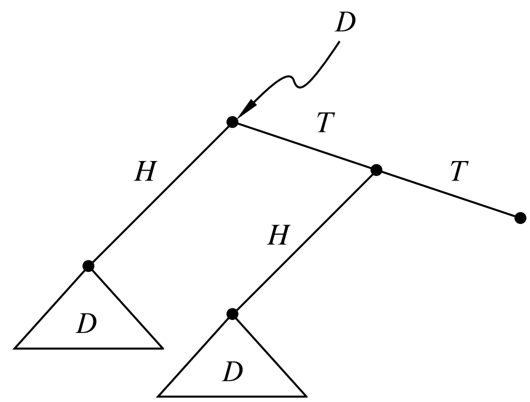


Figure 19.8 Sample space tree for coin toss until two consecutive tails.

Problem 19.13.
Ben Bitdiddle is asked to analyze a game in which a fair coin is tossed until the first time a head turns up. If this head occurs on the nth toss, and n is odd, then he wins $\$2^n/n$, but if n is even, he loses $\$2^n/n$. Ben observes that the expected dollar win from this game is

$$(1/2)\cdot 2 - (1/4)\cdot 2 + (1/8)\cdot 8/3 + \cdots \pm (1/2^n)\cdot 2^n/n = 1 - 1/2 + 1/3 - 1/4 + \cdots \pm 1/n.$$

which is the alternating harmonic series—a series that converges to a definite real number $r > 0$. Since $r > 0$, Ben concludes that it's to his advantage to play this game, but as usual, his shoot-from-the-hip analysis is off the mark. Explain.

Problem 19.14.
Let T be a positive integer valued random variable such that

$$\mathrm{PDF}_T(n) = \frac{1}{an^2},$$

where

$$a ::= \sum_{n \in \mathbb{Z}^+} \frac{1}{n^2}.$$

(a) Prove that $\mathrm{Ex}[T]$ is infinite.

(b) Prove that $\mathrm{Ex}[\sqrt{T}]$ is finite.

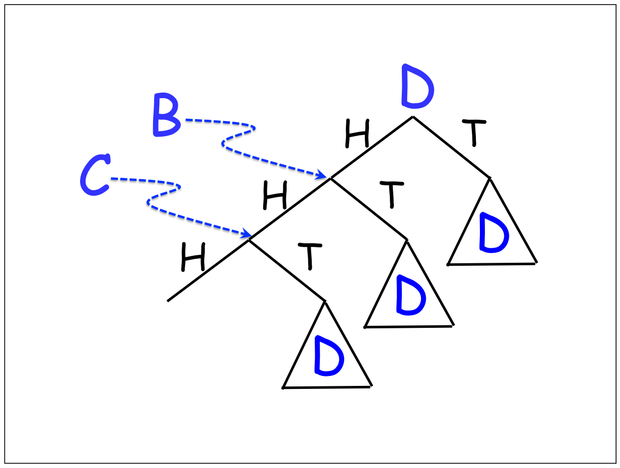


Figure 19.9 Outcome Tree for Flipping Until HHH

Exam Problems

Problem 19.15.
A record of who beat whom in a round-robin tournament can be described with a *tournament digraph*, where the vertices correspond to players and there is an edge $\langle x \rightarrow y \rangle$ iff x beat y in their game. A *ranking* of the players is a path that includes all the players. A tournament digraph may in general have one or more rankings.[7]

Suppose we construct a random tournament digraph by letting each of the players in a match be equally likely to win and having results of all the matches be mutually independent. Find a formula for the expected number of rankings in a random 10-player tournament. Conclude that there is a 10-vertex tournament digraph with more than 7000 rankings.

This problem is an instance of the *probabilistic method*. It uses probability to prove the existence of an object without constructing it.

Problem 19.16.
A coin with probability p of flipping Heads and probability $q ::= 1 - p$ of flipping tails is repeatedly flipped until three consecutive Heads occur. The outcome tree D for this setup is illustrated in Figure 19.9.

Let $e(S)$ be the expected number of flips starting at the root of subtree S of D.

[7] It has a unique ranking iff it is a DAG, see Problem 10.10.

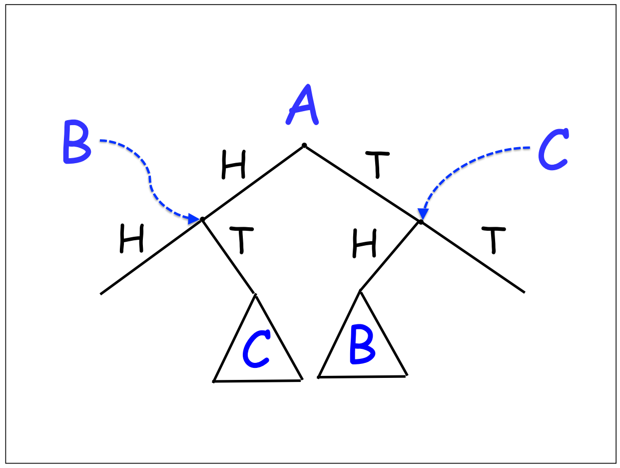


Figure 19.10 Outcome Tree for Flipping Until HH or TT

So we're interested in finding $e(D)$.

Write a small system of equations involving $e(D)$, $e(B)$, and $e(C)$ that could be solved to find $e(D)$. *You do **not** need to solve the equations.*

Problem 19.17.
A coin with probability p of flipping Heads and probability $q ::= 1 - p$ of flipping tails is repeatedly flipped until two consecutive flips match—that is, until HH or TT occurs. The outcome tree A for this setup is illustrated in Figure 19.10.

Let $e(T)$ be the expected number of flips starting at the root of subtree T of A. So we're interested in finding $e(A)$.

Write a small system of equations involving $e(A)$, $e(B)$, and $e(C)$ that could be solved to find $e(A)$. *You do **not** need to solve the equations.*

Homework Problems

Problem 19.18.
We are given a random vector of n distinct numbers. We then determine the maximum of these numbers using the following procedure:

Pick the first number. Call it the *current maximum*. Go through the rest of the vector (in order) and each time we come across a number (call it x) that exceeds our current maximum, we update the current maximum with x.

What is the expected number of times we update the current maximum?

Hint: Let X_i be the indicator variable for the event that the ith element in the vector is larger than all the previous elements.

Problem 19.19.
A fair six-sided die is repeatedly thrown until a six appears. We are interested in the expected time for a six to appear under certain conditions.

A natural probability space $\mathcal{S}$ modelling this situation is the set of finite strings $s \in [1..5]^*6$ of integers from one to six that end at the first occurrence of a six, with $\Pr[s] ::= (1/6)^{|s|}$. Let T be the random variable equal to the number of throws until six appears, namely, $T(s)$ is the length $|s|$ of s.

(a) What is the expected time $\text{Ex}[T]$ till a six is thrown?

Let V be the event that all the dice throws are eVen. That is, $V = \{2, 4\}^*6$ is the event that all throws are 2's and 4's until the first 6.

(b) Prove that $\Pr[V] = 1/4$.

Hint: $V = 2V \cup 4V \cup \{6\}$.

(c) Use the definition of $\text{Ex}[T \mid V]$ as a sum over $s \in V$ to compute the expected time $\text{Ex}[T \mid V]$ till a six is thrown given that all throws are even.

(d) Given that all throws are even, the only possible throws are two, four and six, so we might as well just consider a three-sided die with sides two, four and six. By Mean Time to Failure, the expected time till a six is thrown by a three-sided die is 1/(1/3) = 3, so $\text{Ex}[T \mid V] = 3$, contradicting part (c)! Explain.[8]

Problems for Section 19.6

Class Problems

Problem 19.20.
You have a biased coin with nonzero probability $p < 1$ of tossing a Head. You toss until a Head comes up. Then, similar to the example in Section 19.6, you keep tossing until you get another Head preceded by a run of consecutive Tails

[8] If you're thrown by this, you are not alone. There are several websites devoted to explanations of this seductive problem. In fact, when it came up at the MIT Theory of Computation faculty lunch in April 2018, several attendees confidently defended the mistaken reasoning.

whose length is within 10 of your original run. That is, if you began by tossing k tails followed by a Head, then you continue tossing until you get a run of at least $\max\{k - 10, 0\}$ consecutive Tails.

(a) Let H be the number of Heads that you toss until you get the required run of Tails. Prove that the expected value of H is infinite.

(b) Let $r < 1$ be a positive real number. Instead of waiting for a run of Tails of length $k - 10$ when your original run was length k, just wait for a run of length at least rk. Show that in this case, the expected number of Heads is finite.

Exam Problems

Problem 19.21.
You have a random process for generating a positive integer K. The behavior of your process each time you use it is (mutually) independent of all its other uses. You use your process to generate an integer, and then use your procedure repeatedly until you generate an integer as big as your first one. Let R be the number of additional integers you have to generate.

(a) State and briefly explain a simple closed formula for $\text{Ex}[R \mid K = k]$ in terms of $\Pr[K \geq k]$.

Suppose $\Pr[K = k] = \Theta(k^{-4})$.

(b) Show that $\Pr[K \geq k] = \Theta(k^{-3})$.

(c) Show that $\text{Ex}[R]$ is infinite.

Problems for Section 19.6

Practice Problems

Problem 19.22.
MIT students sometimes delay doing laundry until they finish their problem sets. Assume all random values described below are mutually independent.

(a) A *busy* student must complete 3 problem sets before doing laundry. Each problem set requires 1 day with probability $2/3$ and 2 days with probability $1/3$. Let B be the number of days a busy student delays laundry. What is $\text{Ex}[B]$?

Example: If the first problem set requires 1 day and the second and third problem sets each require 2 days, then the student delays for $B = 5$ days.

(b) A *relaxed* student rolls a fair, 6-sided die in the morning. If he rolls a 1, then he does his laundry immediately (with zero days of delay). Otherwise, he delays for one day and repeats the experiment the following morning. Let R be the number of days a relaxed student delays laundry. What is $\text{Ex}[R]$?

Example: If the student rolls a 2 the first morning, a 5 the second morning, and a 1 the third morning, then he delays for $R = 2$ days.

(c) Before doing laundry, an *unlucky* student must recover from illness for a number of days equal to the product of the numbers rolled on two fair, 6-sided dice. Let U be the expected number of days an unlucky student delays laundry. What is $\text{Ex}[U]$?

Example: If the rolls are 5 and 3, then the student delays for $U = 15$ days.

(d) A student is *busy* with probability $1/2$, *relaxed* with probability $1/3$, and *unlucky* with probability $1/6$. Let D be the number of days the student delays laundry. What is $\text{Ex}[D]$?

Problem 19.23.
Each Math for Computer Science final exam will be graded according to a rigorous procedure:

- With probability $4/7$ the exam is graded by a *TA*,with probability $2/7$ it is graded by a *lecturer*, and with probability $1/7$, it is accidentally dropped behind the radiator and arbitrarily given a score of 84.

- TAs score an exam by scoring each problem individually and then taking the sum.

  - There are ten true/false questions worth 2 points each. For each, full credit is given with probability $3/4$, and no credit is given with probability $1/4$.
  - There are four questions worth 15 points each. For each, the score is determined by rolling two fair dice, summing the results, and adding 3.
  - The single 20 point question is awarded either 12 or 18 points with equal probability.

- Lecturers score an exam by rolling a fair die twice, multiplying the results, and then adding a "general impression"score.
  - With probability $4/10$, the general impression score is 40.
  - With probability $3/10$, the general impression score is 50.
  - With probability $3/10$, the general impression score is 60.

Assume all random choices during the grading process are independent.

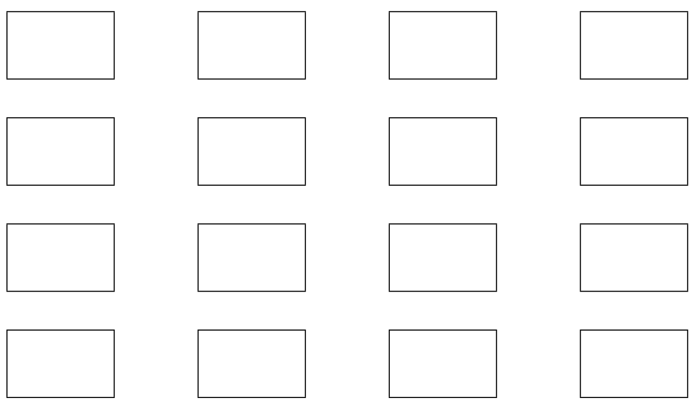
(a) What is the expected score on an exam graded by a TA?

(b) What is the expected score on an exam graded by a lecturer?

(c) What is the expected score on a Math for Computer Science final exam?

Class Problems

Problem 19.24.
A classroom has sixteen desks in a 4×4 arrangement as shown below.



If there is a girl in front, behind, to the left, or to the right of a boy, then the two *flirt*. One student may be in multiple flirting couples; for example, a student in a corner of the classroom can flirt with up to two others, while a student in the center can flirt with as many as four others. Suppose that desks are occupied mutually independently by boys and girls with equal probability. What is the expected number of flirting couples? *Hint:* Linearity.

Problem 19.25.
A *literal* is a propositional variable P or its negation $\overline{P}$, where as usual "$\overline{P}$" abbreviates "NOT(P)." A *3-clause* is an OR of three literals from three different variables. For example,

$$P_1 \text{ OR } P_2 \text{ OR } \overline{P_3}$$

is a 3-clause, but $P_1 \text{ OR } \overline{P_1} \text{ OR } P_2$ is not because P_1 appears twice. A *3-CNF* is a formula that is an AND of 3-clauses. For example,

$$(P_1 \text{ OR } P_2 \text{ OR } \overline{P_3}) \text{ AND } (\overline{P_1} \text{ OR } P_3 \text{ OR } \overline{P_4}) \text{ AND } (P_2 \text{ OR } P_3 \text{ OR } \overline{P_4})$$

is a 3-CNF.

Suppose that G is a 3-CNF with seven 3-clauses. Assign true/false values to the variables in G independently and with equal probability.

(a) What is the probability that the nth clause is true?

(b) What is the expected number of true 3-clauses in G?

(c) Use the fact that the answer to part (b) is greater than six to conclude G must be satisfiable.

Problem 19.26.
A *literal* is a propositional variable or its negation. A k*-clause* is an OR of k literals, with no variable occurring more than once in the clause. For example,

$$P \text{ OR } \overline{Q} \text{ OR } \overline{R} \text{ OR } V,$$

is a 4-clause, but

$$\overline{V} \text{ OR } \overline{Q} \text{ OR } \overline{X} \text{ OR } V,$$

is not, since V appears twice.

Let $\mathcal{S}$ be a set of n distinct k-clauses involving v variables. The variables in different k-clauses may overlap or be completely different, so $k \le v \le nk$.

A random assignment of true/false values will be made independently to each of the v variables, with true and false assignments equally likely. Write formulas in n, k and v in answer to the first two parts below.

(a) What is the probability that any particular k-clause in $\mathcal{S}$ is true under the random assignment?

(b) What is the expected number of true k-clauses in $\mathcal{S}$?

(c) A set of propositions is *satisfiable* iff there is an assignment to the variables that makes all of the propositions true. Use your answer to part (b) to prove that if $n < 2^k$, then $\mathcal{S}$ is satisfiable.

Problem 19.27.
There are n students who are both taking Math for Computer Science (MCS) and Introduction to Signal Processing (SP) this term. To make it easier on themselves, the Professors in charge of these classes have decided to randomly permute their class lists and then assign students grades based on their rank in the permutation (just as many students have suspected). Assume the permutations are equally likely and independent of each other. What is the expected number of students that have in rank in SP that is higher by k than their rank in MCS?

Hint: Let X_r be the indicator variable for the rth ranked student in CS having a rank in SP of at least $r + k$.

Problem 19.28.
A man has a set of n keys, one of which fits the door to his apartment. He tries the keys randomly until he finds the key that fits. Let T be the number of times he tries keys until he finds the right key.

(a) Suppose each time he tries a key that does not fit the door, he simply puts it back. This means he might try the same ill-fitting key several times before he finds the right key. What is $\text{Ex}[T]$?

Hint: Mean time to failure.

Now suppose he throws away each ill-fitting key that he tries. That is, he chooses keys randomly from *among those he has not yet tried*. This way he is sure to find the right key within n tries.

(b) If he hasn't found the right key yet and there are m keys left, what is the probability that he will find the right key on the next try?

(c) Given that he did not find the right key on his first $k - 1$ tries, verify that the probability that he does not find it on the kth trial is given by

$$\Pr\left[T > k \mid T > k - 1\right] = \frac{n - k}{n - (k - 1)}.$$

(d) Prove that

$$\Pr[T > k] = \frac{n - k}{n}. \tag{19.18}$$

Hint: This can be argued directly, but if you don't see how, induction using part (c) will work.

(e) Conclude that in this case

$$\text{Ex}[T] = \frac{n+1}{2}.$$

Problem 19.29.
Justify each line of the following proof that if R and S are *independent* random variables, then

$$\text{Ex}[R \cdot S] = \text{Ex}[R] \cdot \text{Ex}[S].$$

Proof.

$$\begin{aligned}
&\text{Ex}[R \cdot S] \\
&= \sum_{t \in \text{range}(R \cdot S)} t \cdot \Pr[R \cdot S = t] \\
&= \sum_{r \in \text{range}(R),\ s \in \text{range}(S)} rs \cdot \Pr[R = r \text{ and } S = s] \\
&= \sum_{r \in \text{range}(R)} \left(\sum_{s \in \text{range}(S)} rs \cdot \Pr[R = r \text{ and } S = s] \right) \\
&= \sum_{r \in \text{range}(R)} \left(\sum_{s \in \text{range}(S)} rs \cdot \Pr[R = r] \cdot \Pr[S = s] \right) \\
&= \sum_{r \in \text{range}(R)} \left(r \Pr[R = r] \cdot \sum_{s \in \text{range}(S)} s \Pr[S = s] \right) \\
&= \sum_{r \in \text{range}(R)} r \Pr[R = r] \cdot \text{Ex}[S] \\
&= \text{Ex}[S] \cdot \sum_{r \in \text{range}(R)} r \Pr[R = r] \\
&= \text{Ex}[S] \cdot \text{Ex}[R].
\end{aligned}$$

■

Problem 19.30.
A gambler bets on the toss of a fair coin: if the toss is Heads, the gambler gets back the amount he bet along with an additional the amount equal to his bet. Otherwise he loses the amount bet. For example, the gambler bets \$10 and wins, he gets back \$20 for a net profit of \$10. If he loses, he gets back nothing for a net profit of $-\$10$—that is, a net loss of \$10.

Gamblers often try to develop betting strategies to beat the odds is such a game. A well known strategy of this kind is *bet doubling*, namely, bet \$10 on red, and keep doubling the bet until a red comes up. So if the gambler wins his first \$10 bet, he stops playing and leaves with his \$10 profit. If he loses the first bet, he bets \$20 on the second toss. Now if the second toss is Heads, he gets his \$20 bet plus \$20 back and again walks away with a net profit of $20 - 10 = \$10$. If he loses the second toss, he bets \$40 on the third toss, and so on.

You would think that any such strategy will be doomed: in a fair game your expected win by definition is zero, so no strategy should have nonzero expectation. We can make this reasoning more precise as follows:

> Let W_n be a random variable equal to the amount won in the nth coin toss. So with the bet doubling strategy starting with a \$10 bet, $W_1 = \pm 10$ with equal probability. If the betting ends before the nth bet, define $W_n = 0$. So W_2 is zero with probability 1/2, is 10 with probability 1/4, and is -10 with probability 1/4. Now letting W be the amount won when the gambler stops betting, we have
>
> $$W = W_1 + W_2 + \cdots + W_n + \cdots .$$
>
> Furthermore, since each toss is fair,
>
> $$\text{Ex}[W_n] = 0$$
>
> for all $n > 0$. Now by linearity of expectation, we have
>
> $$\text{Ex}[W] = \text{Ex}[W_1] + \text{Ex}[W_2] + \cdots + \text{Ex}[W_n] + \cdots = 0 + 0 + \cdots + 0 + \cdots = 0, \tag{19.19}$$
>
> confirming that with fair tosses, the expected win is zero.

But wait a minute!

(a) Explain why the gambler is certain to win eventually if he keeps betting.

(b) Prove that when the gambler finally wins a bet, his net profit is \$10.

(c) Since the gambler's profit is always \$10 when he wins, and he is certain to win, his expected profit is also \$10. That is

$$\text{Ex}[W] = 10,$$

contradicting (19.19). So what's wrong with the reasoning that led to the false conclusion (19.19)?

Homework Problems

Problem 19.31.
Applying linearity of expectation to the binomial distribution $f_{n,p}$ immediately yielded the identity 19.13:

$$\text{Ex}[f_{n,p}] ::= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = pn. \tag{19.20}$$

Though it might seem daunting to prove this equation without appeal to linearity, it is, after all, pretty similar to the binomial identity, and this connection leads to an immediate alternative algebraic derivation.

(a) Starting with the binomial identity for $(x+y)^n$, prove that

$$xn(x+y)^{n-1} = \sum_{k=0}^{n} k \binom{n}{k} x^k y^{n-k}. \tag{19.21}$$

(b) Now conclude equation (19.20).

Problem 19.32.
Short-term Capital Management (STCM) wants you to invest in a fund with the following rules: you invest one million dollars in their Forward Looking Internet Package (FLIP). Each year, the money in your FLIP account will double or halve with equal probability, and each year STCM will pay you a dividend equal to 10% of the money in your account.

(a) What is the expected number of dollars in your account at the end of k years? Write a simple formula in terms of k.

Hint: \$1,000,000 is in the account the end of year zero. Let X_i be 2 or $1/2$ depending on what happens to your money at the end of the ith year. So the amount of money in the account at the end of year one is $X_1 \cdot \$1,000,000$ and the dividend paid is $(1/10)X_1 \cdot \$1,000,000$.

(b) Give a closed form numerical expression for the expected total number of dollars in dividend payments you will receive by the end of the 10th year. You do *not* need to evaluate your expression.

(c) Adam Smith does his own analysis of your account. He lets $Y_i = 1$ if the money doubles at the end of year i and $Y_i = -1$ otherwise. Then the money in your account after year k is

$$10^6 2^{Y_1} 2^{Y_2} \cdots 2^{Y_k} = 10^6 2^{Y_1+Y_2+\cdots+Y_k}.$$

But $\text{Ex}[Y_i] = 0$, so

$$2^{\text{Ex}[Y_1+Y_2+\cdots+Y_k]} = 2^{\text{Ex}[Y_1]+\text{Ex}[Y_2]+\cdots+\text{Ex}[Y_k]} = 2^{k\cdot 0} = 2^0 = 1.$$

In other words, the expected amount of money in your account forever remains the same as your original investment.

What is wrong with Adam Smith's analysis?

Problem 19.33.

A coin will be flipped repeatedly until the sequence TTH (tail/tail/head) comes up. Successive flips are independent, and the coin has probability p of coming up heads. Let $N_{\texttt{TTH}}$ be the number of coin flips until TTH first appears. What value of p minimizes $\text{Ex}[N_{\texttt{TTH}}]$?

Problem 19.34.

(A true story from World War Two.)

The army needs to test n soldiers for a disease. There is a blood test that accurately determines when a blood sample contains blood from a diseased soldier. The army presumes, based on experience, that the fraction of soldiers with the disease is approximately equal to some small number p.

Approach (1) is to test blood from each soldier individually; this requires n tests. Approach (2) is to randomly group the soldiers into g groups of k soldiers, where $n = gk$. For each group, blend the k blood samples of the people in the group, and test the blended sample. If the group-blend is free of the disease, we are done with that group after one test. If the group-blend tests positive for the disease, then someone in the group has the disease, and we to test all the people in the group for a total of $k + 1$ tests on that group.

Since the groups are chosen randomly, each soldier in the group has the disease with probability p, and it is safe to assume that whether one soldier has the disease is independent of whether the others do.

(a) What is the expected number of tests in Approach (2) as a function of the number of soldiers n, the disease fraction p, and the group size k?

(b) Show how to choose k so that the expected number of tests using Approach (2) is approximately $n\sqrt{p}$. *Hint:* Since p is small, you may assume that $(1-p)^k \approx 1$ and $\ln(1-p) \approx -p$.

(c) What fraction of the work does Approach (2) expect to save over Approach (1) in a million-strong army of whom approximately 1% are diseased?

(d) Can you come up with a better scheme by using multiple levels of grouping, that is, groups of groups?

Problem 19.35.
A wheel-of-fortune has the numbers from 1 to $2n$ arranged in a circle. The wheel has a spinner, and a spin randomly determines the two numbers at the opposite ends of the spinner. How would you arrange the numbers on the wheel to maximize the expected value of:

(a) the sum of the numbers chosen? What is this maximum?

(b) the product of the numbers chosen? What is this maximum?

Hint: For part (b), verify that the sum of the products of numbers oppposite each other is maximized when successive integers are on the opposite ends of the spinner, that is, 1 is opposite 2, 3 is opposite 4, 5 is opposite 6, ...

Problem 19.36.
Let R and S be independent random variables, and f and g be any functions such that $\text{domain}(f) = \text{codomain}(R)$ and $\text{domain}(g) = \text{codomain}(S)$. Prove that $f(R)$ and $g(S)$ are also independent random variables.

Hint: The event $[f(R) = a]$ is the disjoint union of all the events $[R = r]$ for r such that $f(r) = a$.

Problem 19.37.
Peeta bakes between 1 and $2n$ loaves of bread to sell every day. Each day he rolls a fair, n-sided die to get a number from 1 to n, then flips a fair coin. If the coin is heads, he bakes m loaves of bread , where m is the number on the die that day, and if the coin is tails, he bakes $2m$ loaves.

(a) For any positive integer $k \le 2n$, what is the probability that Peeta will make k loaves of bread on any given day?

Hint: Express your solution by cases.

(b) What is the expected number of loaves that Peeta would bake on any given day?

(c) Continuing this process, Peeta bakes bread every day for 30 days. What is the expected total number of loaves that Peeta would bake?

Exam Problems

Problem 19.38.
A box initially contains n balls, all colored black. A ball is drawn from the box at random.

- If the drawn ball is black, then a biased coin with probability, $p > 0$, of coming up heads is flipped. If the coin comes up heads, a white ball is put into the box; otherwise the black ball is returned to the box.
- If the drawn ball is white, then it is returned to the box.

This process is repeated until the box contains n white balls.

Let D be the number of balls drawn until the process ends with the box full of white balls. Prove that $\text{Ex}[D] = nH_n/p$, where H_n is the nth Harmonic number.

Hint: Let D_i be the number of draws after the ith white ball until the draw when the $(i+1)$st white ball is put into the box.

Problem 19.39.
A gambler bets \$10 on "red" at a roulette table (the odds of red are 18/38, slightly less than even) to win \$10. If he wins, he gets back twice the amount of his bet, and he quits. Otherwise, he doubles his previous bet and continues.

For example, if he loses his first two bets but wins his third bet, the total spent on his three bets is $10 + 20 + 40$ dollars, but he gets back $2 \cdot 40$ dollars after his win on the third bet, for a net profit of \$10.

(a) What is the expected number of bets the gambler makes before he wins?

(b) What is his probability of winning?

(c) What is his expected final profit (amount won minus amount lost)?

(d) You can beat a biased game by bet doubling, but bet doubling is not feasible because it requires an infinite bankroll. Verify this by proving that the expected size of the gambler's last bet is infinite.

Problem 19.40.
Six pairs of cards with ranks 1–6 are shuffled and laid out in a row, for example,

$$\boxed{1}\,\boxed{2}\,\boxed{3}\,\boxed{3}\,\boxed{4}\,\boxed{6}\,\boxed{1}\,\boxed{4}\,\boxed{5}\,\boxed{5}\,\boxed{2}\,\boxed{6}$$

In this case, there are two adjacent pairs with the same value, the two 3's and the two 5's. What is the expected number of adjacent pairs with the same value?

Problem 19.41.
There are six kinds of cards, three of each kind, for a total of eighteen cards. The cards are randonly shuffled and laid out in a row, for example,

$$\boxed{1}\,\boxed{2}\,\boxed{5}\,\boxed{5}\,\boxed{5}\,\boxed{1}\,\boxed{4}\,\boxed{6}\,\boxed{2}\,\boxed{6}\boxed{6}\,\boxed{2}\,\boxed{1}\,\boxed{4}\,\boxed{3}\,\boxed{3}\,\boxed{3}\,\boxed{4}$$

In this case, there are two adjacent triples of the same kind, the three 3's and the three 5's.

(a) Derive a formula for the probability that the 4th, 5th, and 6th consecutive cards will be the same kind—that is, all 1's or all 2's or... all 6's?

(b) Let $p ::= \Pr[\text{4th, 5th and 6th cards match}]$—that is, p is the correct answer to part (a). Write a simple formula for the expected number of matching triples in terms of p.