Chapter 20

Random Variables

So far we focused on probabilities of events —that you win the Monty Hall game; that you have a rare medical condition, given that you tested positive; .... Now we focus on quantitative questions: *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 playing 6.042 games all day? Random variables are the mathematical tool for addressing such questions.

20.1 Random Variable Examples

Definition 20.1.1. 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:

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

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

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

So C and M are random variables.

20.1.1 Indicator Random Variables

An *indicator random variable* is a random variable that maps every outcome to either 0 or 1. These 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 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 $I_E(p) = 1$ for outcomes $p \in E$ and $I_E(p) = 0$ for outcomes $p \notin E$. Thus, $M = I_F$ where F is the event that all three coins match.

20.1.2 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. Thus, we can regard an equation or inequality involving a random variable as an event. For example, the event that $C = 2$ consists of the outcomes THH, HTH, and HHT. The event $C \leq 1$ consists of the outcomes TTT, TTH, THT, and HTT.

Naturally enough, we can talk about the probability of events defined by properties of random variables. For example,

$$\begin{aligned}\Pr\{C=2\} &= \Pr\{THH\}+\Pr\{HTH\}+\Pr\{HHT\}\\ &= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}.\end{aligned}$$

20.1.3 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 in the codomain of R_1, and x_2 in the codomain of R_2, we have:

$$\Pr\{R_1=x_1 \text{ AND } R_2=x_2\}=\Pr\{R_1=x_1\}\cdot\Pr\{R_2=x_2\}.$$

As with events, we can formulate independence for random variables in an equivalent and perhaps more intuitive way: random variables R_1 and R_2 are independent if for all x_1 and x_2

$$\Pr\{R_1=x_1 \mid R_2=x_2\}=\Pr\{R_1=x_1\}.$$

whenever the lefthand conditional probability is defined, that is, whenever $\Pr\{R_2=x_2\}>0$.

As an 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 event that the first flip is a Head, so

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

Then H_1 is independent of M, since

$$\begin{aligned}\Pr\{M=1\}=1/4=\Pr\{M=1 \mid H_1=1\}=\Pr\{M=1 \mid H_1=0\}\\ \Pr\{M=0\}=3/4=\Pr\{M=0 \mid H_1=1\}=\Pr\{M=0 \mid H_1=0\}\end{aligned}$$

This example is an instance of a simple lemma:

Lemma 20.1.2. *Two events are independent iff their indicator variables are independent.*

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

Definition 20.1.3. Random variables $R_1, R_2, \ldots, R_n$ are *mutually independent* iff

$$\begin{aligned}&\Pr\{R_1 = x_1 \text{ AND } R_2 = x_2 \text{ AND } \cdots \text{ AND } R_n = x_n\}\\ &\quad = \Pr\{R_1 = x_1\} \cdot \Pr\{R_2 = x_2\} \cdots \Pr\{R_n = x_n\}.\end{aligned}$$

for all $x_1, x_2, \ldots, x_n$.

It is a simple exercise to show that the probability that any *subset* of the variables takes a particular set of values is equal to the product of the probabilities that the individual variables take their values. Thus, for example, if $R_1, R_2, \ldots, R_{100}$ are mutually independent random variables, then it follows that:

$$\Pr\{R_1 = 7 \text{ AND } R_7 = 9.1 \text{ AND } R_{23} = \pi\} = \Pr\{R_1 = 7\} \cdot \Pr\{R_7 = 9.1\} \cdot \Pr\{R_{23} = \pi\}.$$

20.2 Probability Distributions

A random variable maps outcomes to values, but random variables that show up for different spaces of outcomes wind up behaving in much the same way because they have the same probability of taking any given value. Namely, random variables on different probability spaces may wind up having the same probability density function.

Definition 20.2.1. Let R be a random variable with codomain V. The *probability density function (pdf)* of R is a function $\text{PDF}_R : V \to [0, 1]$ defined by:

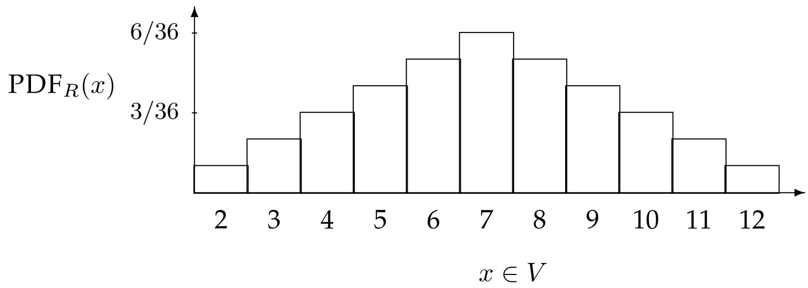
$$\text{PDF}_R(x) ::= \begin{cases} \Pr\{R = x\} & \text{if } x \in \text{range}(R), \\ 0 & \text{if } x \notin \text{range}(R). \end{cases}$$

A consequence of this definition is that

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

This follows 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, let's return to the experiment of rolling two fair, independent dice. As before, let T be the total 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 is shown below:

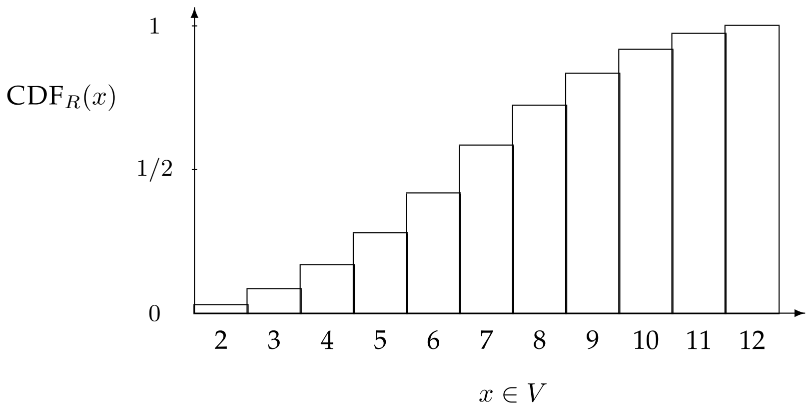


The lump in the middle indicates that sums close to 7 are the most likely. The total area of all the rectangles is 1 since the dice must take on exactly one of the sums in $V = \{2, 3, \ldots, 12\}$.

A closely-related idea is the *cumulative distribution function (cdf)* for a random variable R whose codomain is real numbers. This is a function $\text{CDF}_R : \mathbb{R} \to [0, 1]$ defined by:

$$\text{CDF}_R(x) = \Pr\{R \leq x\}$$

As an example, the cumulative distribution function for the random variable T is shown below:



The height of the i-th 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:

$$\begin{aligned}
\text{CDF}_R(x) &= \Pr\{R \leq x\} \\
&= \sum_{y \leq x} \Pr\{R = y\} \\
&= \sum_{y \leq x} \text{PDF}_R(y)
\end{aligned}$$

In summary, $\text{PDF}_R(x)$ measures the probability that $R = x$ and $\text{CDF}_R(x)$ measures the probability that $R \leq x$. Both the PDF_R and CDF_R capture the same information about the random variable R— you can derive one from the other —but sometimes one is more convenient. The key point here is that neither the probability density function nor the cumulative distribution function involves the sample space of an experiment.

We'll now look at three important distributions and some applications.

20.2.1 Bernoulli Distribution

Indicator random variables are perhaps the most common type because of their close association with events. The probability density function of an indicator random variable, B, is always

$$\begin{aligned} \text{PDF}_B(0) &= p \\ \text{PDF}_B(1) &= 1 - p \end{aligned}$$

where $0 \leq p \leq 1$. The corresponding cumulative distribution function is:

$$\begin{aligned} \text{CDF}_B(0) &= p \\ \text{CDF}_B(1) &= 1 \end{aligned}$$

20.2.2 Uniform Distribution

A random variable that takes on each possible value with the same probability is called *uniform*. For example, the probability density function of a random variable U that is uniform on the set $\{1, 2, \ldots, N\}$ is:

$$\text{PDF}_U(k) = \frac{1}{N}$$

And the cumulative distribution function is:

$$\text{CDF}_U(k) = \frac{k}{N}$$

Uniform distributions come up all the time. For example, the number rolled on a fair die is uniform on the set $\{1, 2, \ldots, 6\}$.

20.2.3 The Numbers Game

Let's play a game! I 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, I'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 the left envelope and see the number 12. Since 12 is a small number, you might guess that that other number is larger. But perhaps I'm sort of 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. I'm picking the numbers and I'm choosing them in a way that I think will defeat your guessing strategy. I'll only use randomization to choose the numbers if that serves *my* end: making you lose!

Intuition Behind the Winning Strategy

Amazingly, there is a strategy that wins more than 50% of the time, regardless of what numbers I put in the envelopes!

Suppose that you somehow knew a number x *between* my lower number and higher numbers. Now you peek in an envelope and see one or the other. 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 my lower and higher numbers, then you are certain to win the game.

The only flaw with this brilliant strategy is that you do *not* know x. Oh well.

But what if you 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%!

Informal arguments about probability, like this one, often sound plausible, but do not hold up under close scrutiny. In contrast, this argument sounds completely implausible— but is actually correct!

Analysis of the Winning Strategy

For generality, suppose that I can choose numbers from the set $\{0, 1, \ldots, n\}$. Call the lower number L and the higher number H.

Your goal is to guess a number x between L and H. To avoid confusing equality cases, you select x at random from among the half-integers:

$$\left\{\frac{1}{2},\ 1\frac{1}{2},\ 2\frac{1}{2},\ \ldots,\ n-\frac{1}{2}\right\}$$

But what probability distribution should you use?

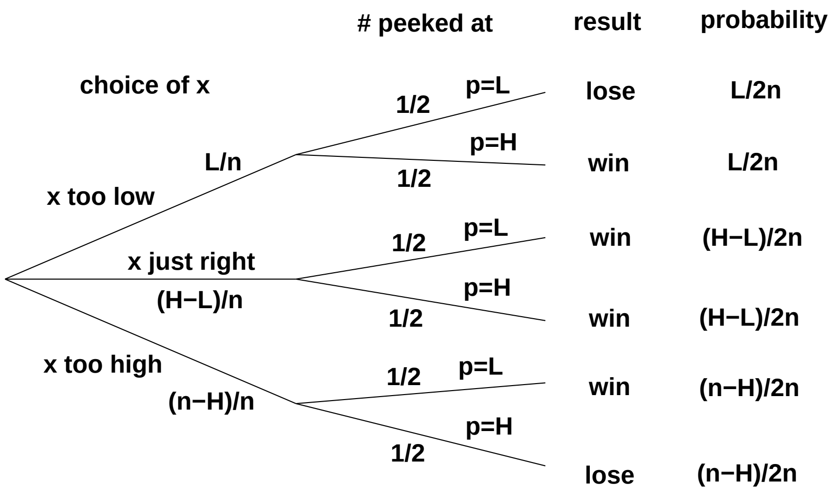
The uniform distribution turns out to be your best bet. An informal justification is that if I figured out that you were unlikely to pick some number— say $50\frac{1}{2}$— then I'd always put 50 and 51 in the evelopes. Then you'd be unlikely to pick an x between L and H and would have less chance of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

After you've selected the number x, you peek into an envelope and see some number p. If $p > x$, then you guess that you're looking at the larger number. If

$p < x$, 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 ($p = L$) or the higher number ($p = H$). This gives a total of six possible outcomes.



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

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! For example, if I choose numbers in the range $0, 1, \ldots, 100$, then you win with probability at least $\frac{1}{2} + \frac{1}{200} = 50.5\%$. Even better, if I'm allowed only numbers in the range $0, \ldots, 10$, then your probability of winning rises to 55%! By Las Vegas standards, those are great odds!

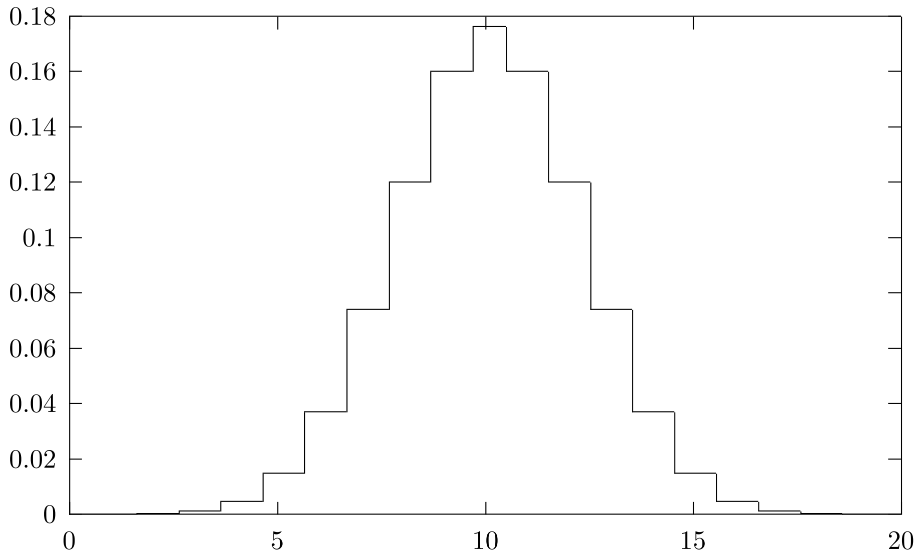
20.2.4 Binomial Distribution

The *binomial distribution* plays an important role in Computer Science as it does in most other sciences. 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; call this random variable H_n. If the coin is fair, then H_n has an *unbiased binomial density function*:

$$\mathrm{PDF}_{H_n}(k) = \binom{n}{k} 2^{-n}.$$

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

Here is a plot of the unbiased probability density function $\mathrm{PDF}_{H_n}(k)$ corresponding to $n = 20$ coins flips. The most likely outcome is $k = 10$ heads, and the probability falls off rapidly for larger and smaller values of k. These falloff regions to the left and right of the main hump are usually 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.

As an example, we can calculate the probability of flipping at most 25 heads in 100 tosses of a fair coin and see that it is very small, namely, 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 fewer than 25 heads! That is, the probability of flipping exactly 25 heads —small as it is —is still nearly twice as large as the probability of flipping exactly 24 heads *plus* the probability of flipping exactly 23 heads *plus* . . . the probability of flipping no heads.

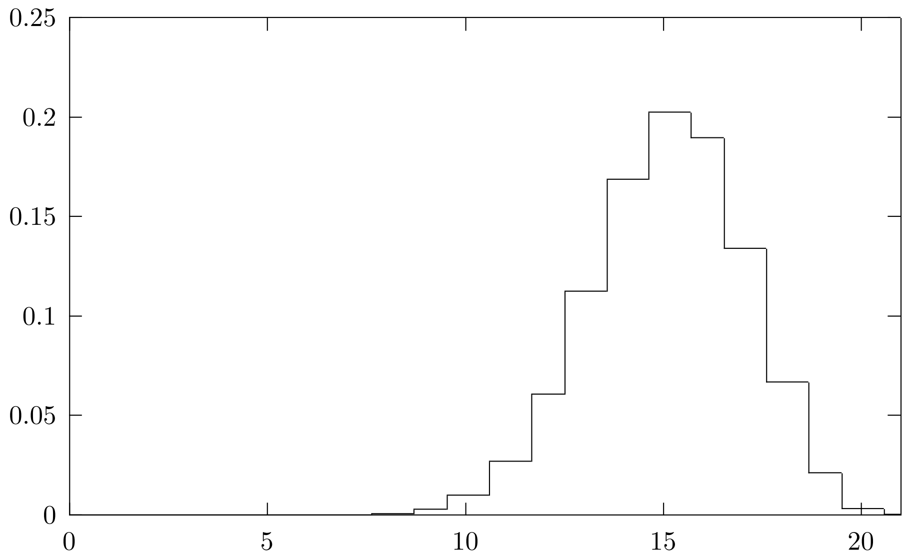
The General Binomial Distribution

Now let J be the number of heads that come up on n independent coins, each of which is heads with probability p. Then J has a *general binomial density function*:

$$\text{PDF}_J(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

As before, there are $\binom{n}{k}$ sequences with k heads and $n-k$ tails, but now the probability of each such sequence is $p^k(1-p)^{n-k}$.

As an example, the plot below shows the probability density function $\text{PDF}_J(k)$ corresponding to flipping $n = 20$ independent coins that are heads with probabilty $p = 0.75$. The graph shows that we are most likely to get around $k = 15$ heads, as you might expect. Once again, the probability falls off quickly for larger and smaller values of k.



20.2.5 Problems

Class Problems

Guess the Bigger Number Game
Team 1: • Write different integers between 0 and 7 on two 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 unseen other number. Team 2 wins if it chooses the larger number.

Problem 20.1.
In section 20.2.3, Team 2 was shown to have 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 20.2.
Suppose X_1, X_2, and X_3 are three mutually independent random variables, each having the uniform distribution

$$\Pr\{X_i = k\} \text{ equal to } 1/3 \text{ for each of } k = 1, 2, 3.$$

Let M be another random variable giving the maximum of these three random variables. What is the density function of M?

Homework Problems

Problem 20.3.
A drunken sailor wanders along main street, which conveniently consists of the points along the x axis with integral coordinates. In each step, the sailor moves one unit left or right along the x axis. A particular *path* taken by the sailor can be

described by a sequence of "left" and "right" steps. For example, ⟨left,left,right⟩ describes the walk that goes left twice then goes right.

We model this scenario with a random walk graph whose vertices are the integers and with edges going in each direction between consecutive integers. All edges are labelled $1/2$.

The sailor begins his random walk at the origin. This is described by an initial distribution which labels the origin with probability 1 and all other vertices with probability 0. After one step, the sailor is equally likely to be at location 1 or -1, so the distribution after one step gives label $1/2$ to the vertices 1 and -1 and labels all other vertices with probability 0.

(a) Give the distributions after the 2nd, 3rd, and 4th step 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)

1. What is the final location of a t-step path that moves right exactly i times?
2. How many different paths are there that end at that location?
3. What is the probability that the sailor ends at this location?

(c) Let L be the random variable giving the sailor's location after t steps, and let $B ::= (L+t)/2$. Use the answer to part (b) to show that B has an unbiased binomial density function.

(d) Again let L be the random variable giving the sailor's location after t steps, where t is even. Show that

$$\Pr\left\{|L| < \frac{\sqrt{t}}{2}\right\} < \frac{1}{2}\,.$$

So there is a better than even chance that the sailor ends up at least $\sqrt{t}/2$ steps from where he started.

Hint: Work in terms of B. Then you can use an estimate that bounds the binomial distribution. Alternatively, observe that the origin is the most likely final location and then use the asymptotic estimate

$$\Pr\{L = 0\} = \Pr\{B = t/2\} \sim \sqrt{\frac{2}{\pi t}}.$$

20.3 Average & Expected Value

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

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

Definition 20.3.1.

$$\begin{aligned} \mathrm{E}[R] &::= \sum_{x \in \text{range}(R)} x \cdot \Pr\{R = x\} \qquad (20.1) \\ &= \sum_{x \in \text{range}(R)} x \cdot \mathrm{PDF}_R(x). \end{aligned}$$

Let's work through an example. Let R be the number that comes up on a fair, six-sided die. Then by (20.1), the expected value of R is:

$$\begin{aligned} \mathrm{E}[R] &= \sum_{k=1}^{6} k \cdot \frac{1}{6} \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{7}{2} \end{aligned}$$

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

There is an even simpler formula for expectation:

Theorem 20.3.2. *If R is a random variable defined on a sample space, $\mathcal{S}$, then*

$$\mathrm{E}[R] = \sum_{\omega \in \mathcal{S}} R(\omega) \Pr\{\omega\} \qquad (20.2)$$

The proof of Theorem 20.3.2, like many of the elementary proofs about expectation in this chapter, follows by judicious regrouping of terms in the defining sum (20.1):

Proof.

$$\begin{aligned}
\mathrm{E}[R] &::= \sum_{x \in \mathrm{range}(R)} x \cdot \Pr\{R = x\} && \text{(Def 20.3.1 of expectation)} \\
&= \sum_{x \in \mathrm{range}(R)} x \left(\sum_{\omega \in [R=x]} \Pr\{\omega\} \right) && \text{(def of } \Pr\{R = x\}\text{)} \\
&= \sum_{x \in \mathrm{range}(R)} \sum_{\omega \in [R=x]} x \Pr\{\omega\} && \text{(distributing } x \text{ over the inner sum)} \\
&= \sum_{x \in \mathrm{range}(R)} \sum_{\omega \in [R=x]} R(\omega) \Pr\{\omega\} && \text{(def of the event } [R = x]\text{)} \\
&= \sum_{\omega \in \mathcal{S}} R(\omega) \Pr\{\omega\}
\end{aligned}$$

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

In general, the defining sum (20.1) is better for calculating expected values and has the advantage that it does not depend on the sample space, but only on the density function of the random variable. On the other hand, the simpler sum over all outcomes (20.2)is sometimes easier to use in proofs about expectation.

20.3.1 Expected Value of an Indicator Variable

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

Lemma 20.3.3. *If I_A is the indicator random variable for event A, then*

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

Proof.

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

■

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

20.3.2 Conditional Expectation

Just like event probabilities, expectations can be conditioned on some event.

Definition 20.3.4. The *conditional expectation*, $\mathrm{E}[R \mid A]$, of a random variable, R, given event, A, is:

$$\mathrm{E}[R \mid A] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr\{R = r \mid A\}. \tag{20.3}$$

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

For example, we can compute the expected value of a roll of a fair die, *given*, for example, that the number rolled is at least 4. We do this by letting R be the outcome of a roll of the die. Then by equation (20.3),

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

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

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

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

which is a little less that $5' \, 8''$.

The Law of *Total Expectation* justifies this method.

Theorem 20.3.5. *Let $A_1, A_2, \ldots$ be a partition of the sample space. Then*

Rule (Law of Total Expectation)**.**

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

Proof.

$$\begin{aligned}
\mathrm{E}[R] &::= \sum_{r \in \mathrm{range}(R)} r \cdot \Pr\{R = r\} && \text{(Def 20.3.1 of expectation)} \\
&= \sum_r r \cdot \sum_i \Pr\{R = r \mid A_i\} \Pr\{A_i\} && \text{(Law of Total Probability)} \\
&= \sum_r \sum_i r \cdot \Pr\{R = r \mid A_i\} \Pr\{A_i\} && \text{(distribute constant } r\text{)} \\
&= \sum_i \sum_r r \cdot \Pr\{R = r \mid A_i\} \Pr\{A_i\} && \text{(exchange order of summation)} \\
&= \sum_i \Pr\{A_i\} \sum_r r \cdot \Pr\{R = r \mid A_i\} && \text{(factor constant } \Pr\{A_i\}\text{)} \\
&= \sum_i \Pr\{A_i\} \mathrm{E}[R \mid A_i]. && \text{(Def 20.3.4 of cond. expectation)}
\end{aligned}$$

■

20.3.3 Mean Time to Failure

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

If we let C be the number of hours until the crash, then the answer to our problem is $\mathrm{E}[C]$. Now the probability that, for $i > 0$, the first crash occurs in the ith hour is the probability that it does not crash in each of the first $i - 1$ hours and it does crash in the ith hour, which is $(1 - p)^{i-1}p$. So from formula (20.1) for expectation, we have

$$\begin{aligned}
\mathrm{E}[C] &= \sum_{i \in \mathbb{N}} i \cdot \Pr\{R = i\} \\
&= \sum_{i \in \mathbb{N}^+} i(1-p)^{i-1}p \\
&= p \sum_{i \in \mathbb{N}^+} i(1-p)^{i-1} \\
&= p \frac{1}{(1-(1-p))^2} && \text{(by (17.1))} \\
&= \frac{1}{p}
\end{aligned}$$

A simple alternative derivation that does not depend on the formula (17.1) (which you remembered, right?) is based on conditional expectation. Given that the computer crashes in the first hour, the expected number of hours to the first crash is obviously 1! On the other hand, given that the computer does not crash in the first hour, then the expected total number of hours till the first crash is the expectation of one plus the number of additional hours to the first crash. So,

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

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

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 really wants to have a baby girl. For simplicity assume there is a 50% chance that each child they have is a girl, and the genders of their children are mutually independent. If the couple insists on having children until they get a girl, then how many baby boys should they expect first?

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

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

20.3.4 Linearity of Expectation

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

Theorem 20.3.6. *For any random variables R_1 and R_2,*

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

Proof. Let $T ::= R_1 + R_2$. The proof follows straightforwardly by rearranging terms in the sum (20.2)

$$\begin{aligned}
\mathrm{E}[T] &= \sum_{\omega \in \mathcal{S}} T(\omega) \cdot \Pr\{\omega\} && \text{(Theorem 20.3.2)}\\
&= \sum_{\omega \in \mathcal{S}} (R_1(\omega) + R_2(\omega)) \cdot \Pr\{\omega\} && \text{(def of } T\text{)}\\
&= \sum_{\omega \in \mathcal{S}} R_1(\omega)\Pr\{\omega\} + \sum_{\omega \in \mathcal{S}} R_2(\omega)\Pr\{\omega\} && \text{(rearranging terms)}\\
&= \mathrm{E}[R_1] + \mathrm{E}[R_2]. && \text{(Theorem 20.3.2)}
\end{aligned}$$

■

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

Theorem 20.3.7 (Linearity of Expectation)**.** *For random variables R_1, R_2 and constants $a_1, a_2 \in \mathbb{R}$,*

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

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

Corollary 20.3.8. *For any random variables $R_1, \ldots, R_k$ and constants $a_1, \ldots, a_k \in \mathbb{R}$,*

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

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

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:

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

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

The Hat-Check Problem

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

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

The trick is to express G as a sum of indicator variables. In particular, let G_i be an indicator for the event that the ith man gets his own hat. That is, $G_i = 1$ if he gets his own hat, and $G_i = 0$ otherwise. The number of men that get their own hat is the sum of these indicators:

$$G = G_1 + G_2 + \cdots + G_n. \tag{20.4}$$

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

Now since G_i is an indicator, we know $1/n = \Pr\{G_i = 1\} = \text{E}[G_i]$ by Lemma 20.3.3. Now we can take the expected value of both sides of equation (20.4) and apply linearity of expectation:

$$\begin{aligned}
\text{E}[G] &= \text{E}[G_1 + G_2 + \cdots + G_n] \\
&= \text{E}[G_1] + \text{E}[G_2] + \cdots + \text{E}[G_n] \\
&= \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = n\left(\frac{1}{n}\right) = 1.
\end{aligned}$$

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!

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 that come up heads?

Let J be the number of heads after the flips, so J has the (n,p)-binomial distribution. Now let I_k be the indicator for the kth coin coming up heads. By Lemma 20.3.3, we have

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

But

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

so by linearity

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

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

The Coupon Collector Problem

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

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

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

A clever application of linearity of expectation leads to a simple solution to the coupon collector problem. Suppose there are five different types of Racin' Rocket, and I 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 I get a new kind of car. For example, the middle segment ends when I get a red car for the first time. In this way, we can break the problem of collecting every type of car into stages. Then we can analyze each stage individually and assemble the results using linearity of expectation.

Let's return to the general case where I'm collecting n Racin' Rockets. Let X_k be the length of the kth segment. The total number of kid's meals I 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, I have k different types of car, and the segment ends when I acquire a new type. When I own k types, each kid's meal contains a type that I 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 I get a new kind of car is $n/(n-k)$ by the "mean time to failure" formula. So we have:

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

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

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

Let's use this general solution 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:

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

20.3.5 The Expected Value of a Product

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

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

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

$$\mathrm{E}[R_1 \cdot R_2] = \mathrm{E}[R_1] \cdot \mathrm{E}[R_2] = 3.5 \cdot 3.5 = 12.25. \tag{20.5}$$

Here the first equality holds because the dice are independent.

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

$$\begin{aligned}
\mathrm{E}[R_1 \cdot R_2] &= \mathrm{E}\left[R_1^2\right] \\
&= \sum_{i=1}^{6} i^2 \cdot \Pr\left\{R_1^2 = i^2\right\} \\
&= \sum_{i=1}^{6} i^2 \cdot \Pr\{R_1 = i\} \\
&= \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} \\
&= 15\ 1/6 \\
&\neq 12\ 1/4 \\
&= \mathrm{E}[R_1] \cdot \mathrm{E}[R_2].
\end{aligned}$$

Theorem 20.3.9. *For any two* independent *random variables R_1, R_2,*

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

Proof. The event $[R_1 \cdot R_2 = r]$ can be split up into events of the form $[R_1 = r_1 \text{ and } R_2 = r_2]$ where $r_1 \cdot r_2 = r$. So

$$\begin{aligned}
&\mathrm{E}[R_1 \cdot R_2] \\
&::= \sum_{r \in \text{range}(R_1 \cdot R_2)} r \cdot \Pr\{R_1 \cdot R_2 = r\} \\
&= \sum_{r_i \in \text{range}(R_i)} r_1 r_2 \cdot \Pr\{R_1 = r_1 \text{ and } R_2 = r_2\} \\
&= \sum_{r_1 \in \text{range}(R_1)} \sum_{r_2 \in \text{range}(R_2)} r_1 r_2 \cdot \Pr\{R_1 = r_1 \text{ and } R_2 = r_2\} && \text{(ordering terms in the sum)} \\
&= \sum_{r_1 \in \text{range}(R_1)} \sum_{r_2 \in \text{range}(R_2)} r_1 r_2 \cdot \Pr\{R_1 = r_1\} \cdot \Pr\{R_2 = r_2\} && \text{(indep. of } R_1, R_2\text{)} \\
&= \sum_{r_1 \in \text{range}(R_1)} \left(r_1 \Pr\{R_1 = r_1\} \cdot \sum_{r_2 \in \text{range}(R_2)} r_2 \Pr\{R_2 = r_2\} \right) && \text{(factoring out } r_1 \Pr\{R_1 = r_1\}\text{)} \\
&= \sum_{r_1 \in \text{range}(R_1)} r_1 \Pr\{R_1 = r_1\} \cdot \mathrm{E}[R_2] && \text{(def of } \mathrm{E}[R_2]\text{)} \\
&= \mathrm{E}[R_2] \cdot \sum_{r_1 \in \text{range}(R_1)} r_1 \Pr\{R_1 = r_1\} && \text{(factoring out } \mathrm{E}[R_2]\text{)} \\
&= \mathrm{E}[R_2] \cdot \mathrm{E}[R_1]. && \text{(def of } \mathrm{E}[R_1]\text{)}
\end{aligned}$$

■

Theorem 20.3.9 extends routinely to a collection of mutually independent variables.

Corollary 20.3.10. *If random variables $R_1, R_2, \ldots, R_k$ are mutually independent, then*

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

20.3.6 Problems

Practice Problems

Problem 20.4.
MIT students sometimes delay laundry for a few days. 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 $\mathrm{E}[B]$?

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 $\mathrm{E}[R]$?

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 $\mathrm{E}[U]$?

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 $\mathrm{E}[D]$?

Problem 20.5.
Each 6.042 final exam will be graded according to a rigorous procedure:

- With probability $\frac{4}{7}$ the exam is graded by a *TA*,with probability $\frac{2}{7}$ it is graded by a *lecturer*, and with probability $\frac{1}{7}$, 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 $\frac{3}{4}$, and no credit is given with probability $\frac{1}{4}$.
  - 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 $\frac{4}{10}$, the general impression score is 40.
  - With probability $\frac{3}{10}$, the general impression score is 50.
  - With probability $\frac{3}{10}$, 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 6.042 final exam?

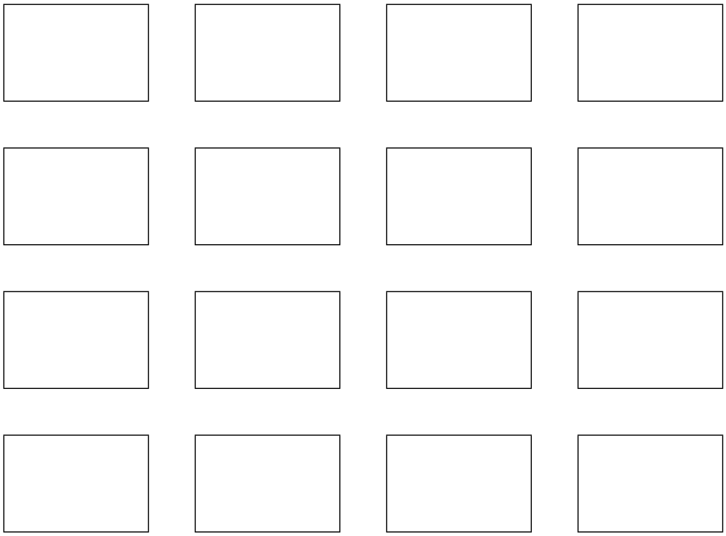
Class Problems

Problem 20.6.
Let's see what it takes to make Carnival Dice fair. Here's the game with payoff parameter k: make three independent rolls of a fair die. If you roll a six

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

For what value of k is this game fair?

Problem 20.7.
A classroom has sixteen desks arranged as shown below.



If there is a girl in front, behind, to the left, or to the right of a boy, then the two of them *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 by boys and girls with equal probability and mutually independently. What is the expected number of flirting couples? *Hint:* Linearity.

Problem 20.8.
Here are seven propositions:

$$\begin{array}{ccccc}
x_1 & \text{OR} & x_3 & \text{OR} & \overline{x_7} \\
\overline{x_5} & \text{OR} & x_6 & \text{OR} & x_7 \\
x_2 & \text{OR} & \overline{x_4} & \text{OR} & x_6 \\
\overline{x_4} & \text{OR} & x_5 & \text{OR} & \overline{x_7} \\
x_3 & \text{OR} & \overline{x_5} & \text{OR} & \overline{x_8} \\
x_9 & \text{OR} & \overline{x_8} & \text{OR} & x_2 \\
\overline{x_3} & \text{OR} & x_9 & \text{OR} & x_4
\end{array}$$

Note that:

1. Each proposition is the disjunction (OR) of three terms of the form x_i or the form $\overline{x_i}$.

2. The variables in the three terms in each proposition are all different.

Suppose that we assign true/false values to the variables $x_1, \ldots, x_9$ independently and with equal probability.

(a) What is the expected number of true propositions?

Hint: Let T_i be an indicator for the event that the i-th proposition is true.

(b) Use your answer to prove that for *any* set of 7 propositions satisfying the conditions 1. and 2., there is an assignment to the variables that makes all 7 of the propositions true.

Problem 20.9. (a) Suppose we flip a fair coin until two Tails in a row come up. What is the expected number, N_{TT}, of flips we perform? *Hint:* Let D be the tree diagram for this process. Explain why $D = H \cdot D + T \cdot (H \cdot D + T)$. Use the Law of Total Expectation 20.3.5

(b) Suppose we flip a fair coin until a Tail immediately followed by a Head come up. What is the expected number, N_{TH}, of flips we perform?

(c) Suppose we now play a game: flip a fair coin until either `TT` or `TH` first occurs. You win if `TT` comes up first, 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, but he merely pays you a 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?

Problem 20.10.
Justify each line of the following proof that if R_1 and R_2 are *independent*, then

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

Proof.

$$
\begin{aligned}
&\mathrm{E}\left[R_1 \cdot R_2\right] \\
&= \sum_{r \in \mathrm{range}(R_1 \cdot R_2)} r \cdot \Pr\left\{R_1 \cdot R_2 = r\right\} \\
&= \sum_{r_i \in \mathrm{range}(R_i)} r_1 r_2 \cdot \Pr\left\{R_1 = r_1 \text{ and } R_2 = r_2\right\} \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \sum_{r_2 \in \mathrm{range}(R_2)} r_1 r_2 \cdot \Pr\left\{R_1 = r_1 \text{ and } R_2 = r_2\right\} \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \sum_{r_2 \in \mathrm{range}(R_2)} r_1 r_2 \cdot \Pr\left\{R_1 = r_1\right\} \cdot \Pr\left\{R_2 = r_2\right\} \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \left(r_1 \Pr\left\{R_1 = r_1\right\} \cdot \sum_{r_2 \in \mathrm{range}(R_2)} r_2 \Pr\left\{R_2 = r_2\right\} \right) \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} r_1 \Pr\left\{R_1 = r_1\right\} \cdot \mathrm{E}\left[R_2\right] \\
&= \mathrm{E}\left[R_2\right] \cdot \sum_{r_1 \in \mathrm{range}(R_1)} r_1 \Pr\left\{R_1 = r_1\right\} \\
&= \mathrm{E}\left[R_2\right] \cdot \mathrm{E}\left[R_1\right].
\end{aligned}
$$

■

Problem 20.11.
Here are seven propositions:

$$
\begin{array}{rcrcr}
x_1 & \vee & x_3 & \vee & \neg x_7 \\
\neg x_5 & \vee & x_6 & \vee & x_7 \\
x_2 & \vee & \neg x_4 & \vee & x_6 \\
\neg x_4 & \vee & x_5 & \vee & \neg x_7 \\
x_3 & \vee & \neg x_5 & \vee & \neg x_8 \\
x_9 & \vee & \neg x_8 & \vee & x_2 \\
\neg x_3 & \vee & x_9 & \vee & x_4
\end{array}
$$

Note that:

1. Each proposition is the OR of three terms of the form x_i or the form $\neg x_i$.
2. The variables in the three terms in each proposition are all different.

Suppose that we assign true/false values to the variables $x_1, \ldots, x_9$ independently and with equal probability.

(a) What is the expected number of true propositions?

(b) Use your answer to prove that there exists an assignment to the variables that makes *all* of the propositions true.

Problem 20.12.
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 the last 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 20.13.
A gambler bets \$10 on "red" at a roulette table (the odds of red are 18/38 which 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.

(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) The fact that the gambler's expected profit is positive, despite the fact that the game is biased against him, is known as the *St. Petersberg paradox*. The paradox arises from an unrealistic, implicit assumption about the gambler's money. Explain.

Hint: What is the expected size of his last bet?

Homework Problems

Problem 20.14.
Let R and S be independent random variables, and f and g be any functions such that domain $(f) =$ codomain (R) and domain $(g) =$ codomain (S). Prove that $f(R)$ and $g(S)$ are independent random variables. *Hint:* The event $[f(R) = a]$ is the disjoint union of all the events $[R = r]$ for r such that $f(r) = a$.

MIT OpenCourseWare
http://ocw.mit.edu

6.042J / 18.062J Mathematics for Computer Science
Spring 2010

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.