6.4 Random Variables

What are Random Variables?

A **random variable** for an experiment with a sample space S is a function that assigns a number to each element of S. Typically instead of using f to stand for such a function we use X (at first, a random variable was conceived of as a variable related to an experiment, explaining the use of X, but it is very helpful in understanding the mathematics to realize it actually is a function on the sample space).

For example, if we consider the process of flipping a coin n times, we have the set of all sequences of n Hs and Ts as our sample space. The "number of heads" random variable takes a sequence and tells us how many heads are in that sequence. Somebody might say "Let X be the number of heads in 5 flips of a coin." In that case $X(HTHHT) = 3$ while $X(THTHT) = 2$. It may be rather jarring to see X used to stand for a function, but it is the notation most people use.

For a sequence of hashes of n keys into a table with k locations, we might have a random variable X_i which is the number of keys that are hashed to location i of the table, or a random variable X that counts the number of collisions (hashes to a location that already has at least one key). For an n question test on which each answer is either right or wrong (a short answer, True-False or multiple choice test for example) we could have a random variable that gives the number of right answers in a particular sequence of answers to the test. For a meal at a restaurant we might have a random variable that gives the price of any particular sequence of choices of menu items.

Exercise 6.4-1 Give several random variables that might be of interest to a doctor whose sample space is her patients.

Exercise 6.4-2 If you flip a coin six times, how many heads do you expect?

A doctor might be interested in patients' ages, weights, temperatures, blood pressures, cholesterol levels, etc.

For Exercise 6.4-2, in six flips of a coin, it is natural to expect three heads. We might argue that if we average the number of heads over all possible outcomes, the average should be half the number of flips. Since the probability of any given sequence equals that of any other, it is reasonable to say that this average is what we expect. Thus we would say we expect the number of heads to be half the number of flips. We will explore this more formally later.

Binomial Probabilities

When we study an independent trials process with two outcomes at each stage, it is traditional to refer to those outcomes as successes and failures. When we are flipping a coin, we are often interested in the number of heads. When we are analyzing student performance on a test, we are interested in the number of correct answers. When we are analyzing the outcomes in drug trials, we are interested in the number of trials where the drug was successful in treating the disease. This suggests a natural random variable associated with an independent trials process with two outcomes at each stage, namely the number of successes in n trials. We will analyze in general the probability of exactly k successes in n independent trials with probability p of success (and thus probability $1-p$ of failure) on each trial. It is standard to call such an independent trials process a *Bernoulli trials process*.

Exercise 6.4-3 Suppose we have 5 Bernoulli trials with probability p of success on each trial. What is the probability of success on the first three trials and failure on the last two? Failure on the first two trials and success on the last three? Success on trials 1, 3, and 5, and failure on the other two? Success on any particular three trials, and failure on the other two?

Since the probability of a sequence of outcomes is the product of the probabilities of the individual outcomes, the probability of any sequence of 3 successes and 2 failures is $p^3(1-p)^2$. More generally, in n Bernoulli trials, the probability of a given sequence of k successes and $n-k$ failures is $p^k(1-p)^{n-k}$. However this is not the probability of having k successes, because many different sequences could have k successes.

How many sequences of n successes and failures have exactly k successes? The number of ways to choose the k places out of n where the successes occur is $\binom{n}{k}$, so the number of sequences with k successes is $\binom{n}{k}$. This paragraph and the last together give us Theorem 6.7.

Theorem 6.7 *The probability of having exactly k successes in a sequence of n independent trials with two outcomes and probability p of success on each trial is*

$$P(\textit{exactly } k \textit{ successes}) = \binom{n}{k} p^k (1-p)^{n-k}$$

.

Proof: The proof follows from the two paragraphs preceding the theorem.■

Because of the connection between these probabilities and the binomial coefficients, the probabilities of Theorem 6.7 are called **binomial probabilities**, or the **binomial probability distribution**.

Exercise 6.4-4 A student takes a ten question objective test. Suppose that a student who knows 80% of the course material has probability .8 of success an any question, independently of how the student did on any other problem. What is the probability that this student earns a grade of 80 or better?

Exercise 6.4-5 Recall the primality testing algorithm from Section 2.4. Here we said that we could, by choosing a random number less than or equal to n, perform a test on n that, if n was not prime, would certify this fact with probability 1/2. Suppose we perform 20 of these tests. It is reasonable to assume that each of these tests is independent of the rest of them. What is the probability that a non-prime number is certified to be non-prime?

Since a grade of 80 or better on a ten question test corresponds to 8, 9, or 10 successes in ten trials, in Exercise 6.4-4 we have

$$P(80 \text{ or better}) = \binom{10}{8}(.8)^8(.2)^2 + \binom{10}{9}(.8)^9(.2)^1 + (.8)^{10}.$$

Some work with a calculator gives us that this sum is approximately .678.

In Exercise 6.4-5, we will first compute the probability that a non-prime number is not certified to be non-prime. If we think of success as when the number is certified non-prime and failure when it isn't, then we see that the only way to fail to certify a number is to have 20 failures. Using our formula we see that the probability that a non-prime number is not certified non-prime is just $\binom{20}{20}(.5)^{20} = 1/1048576$. Thus the chance of this happening is less than one in a million, and the chance of certifying the non-prime as non-prime is 1 minus this. Therefore the probability that a non-prime number will be certified non-prime is 1048575/1048576, which is more than .999999, so a non-prime number is almost sure to be certified non-prime.

A Taste of Generating Functions We note a nice connection between the probability of having exactly k successes and the binomial theorem. Consider, as an example, the polynomial $(H+T)^3$. Using the binomial theorem, we get that this is

$$(H+T)^3 = \binom{3}{0}H^3 + \binom{3}{1}H^2T + \binom{3}{2}HT^2 + \binom{3}{3}T^3.$$

We can interpret this as telling us that if we flip a coin three times, with outcomes heads or tails each time, then there are $\binom{3}{0} = 1$ way of getting 3 heads, $\binom{3}{2} = 3$ ways of getting two heads and one tail, $\binom{3}{1} = 3$ ways of getting one head and two tails and $\binom{3}{3} = 1$ way of getting 3 tails.

Similarly, if we replace H and T by px and $(1-p)y$ we would get the following:

$$(px+(1-p)y)^3 = \binom{3}{0}p^3x^3 + \binom{3}{1}p^2(1-p)x^2y + \binom{3}{2}p(1-p)^2xy^2 + \binom{3}{3}(1-p)^3y^3.$$

Generalizing this to n repeated trials where in each trial the probability of success is p, we see that by taking $(px+(1-p)y)^n$ we get

$$(px+(1-p)y)^n = \sum_{k=0}^{k}\binom{n}{k}p^k(1-p)^{n-k}x^ky^{n-k}.$$

Taking the coefficient of x^ky^{n-k} from this sum, we get exactly the result of Theorem 6.7. This connection is a simple case of a very powerful tool known as *generating functions*. We say that the polynomial $(px+(1-p)y)^n$ *generates* the binomial probabilities. In fact, we don't even need the y, because

$$(px+1-p)^n = \sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i}x^i.$$

In general, the *generating function* for the sequence $a_0, a_1, a_2, \ldots, a_n$ is $\sum_{i=1}^{n} a_ix^i$, and the *generating function* for an infinite sequence $a_0, a_1, a_2, \ldots, a_n, \ldots$ is the infinite series $\sum_{i=1}^{\infty} a_ix^i$.

Expected Value

In Exercise 6.4-4 and Exercise 6.4-2 we asked about the value you would expect a random variable(in these cases, a test score and the number of heads in six flips of a coin) to have. We haven't yet defined what we mean by the value we expect, and yet it seems to make sense in the places we asked about it. If we say we expect 1 head if we flip a coin twice, we can explain our reasoning by taking an average. There are four outcomes, one with no heads, two with one head, and one with two heads, giving us an average of

$$\frac{0+1+1+2}{4} = 1.$$

Notice that using averages compels us to have some expected values that are impossible to achieve. For example in three flips of a coin the eight possibilities for the number of heads are 0, 1, 1, 1, 2, 2, 2, 3, giving us for our average

$$\frac{0+1+1+1+2+2+2+3}{8} = 1.5.$$

Exercise 6.4-6 An interpretation in games and gambling makes it clear that it makes sense to expect a random variable to have a value that is not one of the possible outcomes. Suppose that I proposed the following game. You pay me some money, and then you flip three coins. I will pay you one dollar for every head that comes up. Would you play this game if you had to pay me \$2.00? How about if you had to pay me \$1? How much do you think it should cost, in order for this game to be fair?

Since you expect to get 1.5 heads, you expect to make \$1.50. Therefore, it is reasonable to play this game as long as the cost is at most \$1.50.

Certainly averaging our variable over all elements of our sample space by adding up one result for each element of the sample space as we have done above is impractical even when we are talking about something as simple as ten flips of a coin. However we can ask how many times each possible number of heads arises, and then multiply the number of heads by the number of times it arises to get an average number of heads of

$$\frac{0\binom{10}{0}+1\binom{10}{1}+2\binom{10}{2}+\cdots+9\binom{10}{9}+10\binom{10}{10}}{1024} = \frac{\sum_{i=0}^{10} i\binom{10}{i}}{1024}. \tag{6.22}$$

Thus we wonder whether we have seen a formula for $\sum_{i=0}^{n} i\binom{n}{i}$. Perhaps we have, but in any case the binomial theorem and a bit of calculus or a proof by induction show that

$$\sum_{i=0}^{n} i\binom{n}{i} = 2^{n-1}n,$$

giving us $512 \cdot 10/1024 = 5$ for the fraction in Equation 6.22. If you are asking "Does it have to be that hard?" then good for you. Once we know a bit about the theory of expected values of random variables, computations like this will be replaced by far simpler ones.

Besides the nasty computations that a simple question lead us to, the average value of a random variable on a sample space need not have anything to do with the result we expect. For instance if we replace heads and tails with right and wrong, we get the sample space of possible results that a student will get when taking a ten question test with probability .9 of getting the right answer on any one question. Thus if we compute the average number of right answers in all the possible patterns of test results we get an average of 5 right answers. This is not the number of right answers we expect because averaging has nothing to do with the underlying process that gave us our probability! If we analyze the ten coin flips a bit more carefully, we can resolve this disconnection. We can rewrite Equation 6.22 as

$$0\frac{\binom{10}{0}}{1024}+1\frac{\binom{10}{1}}{1024}+2\frac{\binom{10}{2}}{1024}+\cdots+9\frac{\binom{10}{9}}{1024}+10\frac{\binom{10}{10}}{1024}=\sum_{i=0}^{10} i\frac{\binom{10}{i}}{1024}. \tag{6.23}$$

In Equation 6.23 we see we can compute the average number of heads by multiplying each value of our "number of heads" random variable by the probability that we have that value for our random variable, and then adding the results. This gives us a "weighted average" of the values of our random variable, each value weighted by its probability. Because the idea of weighting a random variable by its probability comes up so much in Probability Theory, there is a special notation that has developed to use this weight in equations. We use $P(X = x_i)$ to stand for the probability that the random variable X equals the value x_i. We call the function that assigns $P(x_i)$ to the event $P(X = x_i)$ the *distribution function* of the random variable X. Thus, for example, the binomial probability distribution is the distribution function for the "number of successes" random variable in Bernoulli trials.

We define the **expected value** or **expectation** of a random variable X whose values are the set $\{x_1, x_2, \ldots x_k\}$ to be

$$E(X) = \sum_{i=1}^{k} x_i P(X = x_i).$$

Then for someone taking a ten-question test with probability .9 of getting the correct answer on each question, the expected number of right answers is

$$\sum_{i=0}^{10} i\binom{10}{i}(.9)^i(.1)^{10-i}.$$

In the end of section exercises we will show a technique (that could be considered an application of generating functions) that allows us to compute this sum directly by using the binomial theorem and calculus. We now proceed to develop a less direct but easier way to compute this and many other expected values.

Exercise 6.4-7 Show that if a random variable X is defined on a sample space S (you may assume X has values x_1, x_2, $\ldots x_k$ as above) then the expected value of X is given by

$$E(X) = \sum_{s:s\in S} X(s)P(s).$$

(In words, we take each member of the sample space, compute its probability, multiply the probability by the value of the random variable and add the results.)

In Exercise 6.4-7 we asked for a proof of a fundamental lemma

Lemma 6.8 *If a random variable X is defined on a (finite) sample space S, then its expected value is given by*

$$E(X) = \sum_{s:s\in S} X(s)P(s).$$

Proof: Assume that the values of the random variable are x_1, x_2, $\ldots x_k$. Let F_i stand for the event that the value of X is x_i, so that $P(F_i) = P(X = x_i)$. Then, in the sum on the right-hand side of the equation in the statement of the lemma, we can take the items in the sample space, group them together into the events F_i and and rework the sum into the definition of expectation, as follows:

$$\begin{aligned}
\sum_{s:s\in S} X(s)P(s) &= \sum_{i=1}^{k} \sum_{s:s\in F_i} X(s)P(s) \\
&= \sum_{i=1}^{k} \sum_{s:s\in F_i} x_i P(s) \\
&= \sum_{i=1}^{k} x_i \sum_{s:s\in F_i} P(s) \\
&= \sum_{i=1}^{k} x_i P(F_i) \\
&= \sum_{i=1}^{k} x_i P(X = x_i) = E(X).
\end{aligned}$$

■

The proof of the lemma need not be so formal and symbolic as what we wrote; in English, it simply says that when we compute the sum in the Lemma, we can group together all elements of the sample space that have X-value x_i and add up their probabilities; this gives us $x_i P(x_i)$, which leads us to the definition of the expected value of X.

Expected Values of Sums and Numerical Multiples

Another important point about expected value follows naturally from what we think about when we use the word "expect" in English. If a paper grader expects to earn ten dollars grading papers today and expects to earn twenty dollars grading papers tomorrow, then she expects to earn thirty dollars grading papers in these two days. We could use X_1 to stand for the amount of money she makes grading papers today and X_2 to stand for the amount of money she makes grading papers tomorrow, so we are saying

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

This formula holds for any sum of a pair of random variables, and more generally for any sum of random variables on the same sample space.

Theorem 6.9 *Suppose X and Y are random variables on the (finite) sample space S. Then*

$$E(X + Y) = E(X) + E(Y).$$

Proof: From Lemma 6.8 we may write

$$E(X+Y) = \sum_{s:s\in S} (X(s)+Y(s))P(s) = \sum_{s:s\in S} X(s)P(s) + \sum_{s:s\in S} Y(s)P(s) = E(X) + E(Y).$$

■

If we double the credit we give for each question on a test, we would expect students' scores to double. Thus our next theorem should be no surprise. In it we use the notation cX for the random variable we get from X by multiplying all its values by the number c.

Theorem 6.10 *Suppose X is a random variable on a sample space S. Then for any number c, $E(cX) = cE(X)$.*

Proof: Left as a problem.■

Theorems 6.9 and 6.10 are very useful in proving facts about random variables. Taken together, they are typically called *linearity of expectation*. (The idea that the expectation of a sum is the same as the sum of expectations is called the *additivity of expectation*.) The idea of linearity will often allow us to work with expectations much more easily than if we had to work with the underlying probabilities.

For example, on one flip of a coin, our expected number of heads is .5. Suppose we flip a coin n times and let X_i be the number of heads we see on flip i, so that X_i is either 0 or 1. (For example in five flips of a coin, $X_2(HTHHT) = 0$ while $X_3(HTHHT) = 1$.) Then X, the total number of heads in n flips is given by

$$X = X_1 + X_2 + \cdots X_n, \tag{6.24}$$

the sum of the number of heads on the first flip, the number on the second, and so on through the number of heads on the last flip. But the expected value of each X_i is .5. We can take the expectation of both sides of Equation 6.24 and apply Lemma 6.9 repeatedly (or use induction) to get that

$$\begin{aligned} E(X) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= .5 + .5 + \cdots + .5 \\ &= .5n \end{aligned}$$

Thus in n flips of a coin, the expected number of heads is $.5n$. Compare the ease of this method with the effort needed earlier to deal with the expected number of heads in ten flips! Dealing with probability .9 or, in general with probability p poses no problem.

Exercise 6.4-8 Use the additivity of expectation to determine the expected number of correct answers a student will get on an n question "fill in the blanks" test if he or she knows 90% of the material in the course and the questions on the test are an accurate and uniform sampling of the material in the course.

In Exercise 6.4-8, since the questions sample the material in the course accurately, the most natural probability for us to assign to the event that the student gets a correct answer on a given question is .9. We can let X_i be the number of correct answers on question i (that is, either 1 or 0 depending on whether or not the student gets the correct answer). Then the expected number of right answers is the expected value of the sum of the variables X_i. From Theorem 6.9 see that in n trials with probability .9 of success, we expect to have $.9n$ successes. This gives that the expected number of right answers on a ten question test with probability .9 of getting each question right is 9, as we expected. This is a special case of our next theorem, which is proved by the same kind of computation.

Theorem 6.11 *In a Bernoulli trials process, in which each experiment has two outcomes and probability p of success, the expected number of successes is np.*

Proof: Let X_i be the number of successes in the ith of n independent trials. The expected number of successes on the ith trial (i.e. the expected value of X_i) is, by definition,

$$p \cdot 1 + (1-p) \cdot 0 = p.$$

The number of successes X in all n trials is the sum of the random variables X_i. Then by Theorem 6.9 the expected number of successes in n independent trials is the sum of the expected values of the n random variables X_i and this sum is np.■

The Number of Trials until the First Success

Exercise 6.4-9 How many times do you expect to have to flip a coin until you first see a head? Why? How many times to you expect to have to roll two dice until you see a sum of seven? Why?

Our intuition suggests that we should have to flip a coin twice to see a head. However we could conceivably flip a coin forever without seeing a head, so should we really expect to see a head in two flips? The probability of getting a seven on two dice is 1/6. Does that mean we should expect to have to roll the dice six times before we see a seven?

In order to analyze this kind of question we have to realize that we are stepping out of the realm of independent trials processes on finite sample spaces. We will consider the process of repeating independent trials with probability p of success until we have a success and then stopping. Now the possible outcomes of our multistage process are the infinite set

$$\{S, FS, FFS, \ldots, F^iS, \ldots\},$$

in which we have used the notation F^iS to stand for the sequence of i failures followed by a success. Since we have an infinite sequence of outcomes, it makes sense to think about whether we can assign an infinite sequence of probability weights to its members so that the resulting sequence of probabilities adds to one. If so, then all our definitions make sense, and in fact the proofs of all our theorems remain valid.[5] There is only one way to assign weights that is consistent with our knowledge of (finite) independent trials processes, namely

$$P(S) = p, \quad P(FS) = (1-p)p, \quad \ldots, \quad P(F^iS) = (1-p)^ip, \quad \ldots.$$

Thus we have to hope these weights add to one; in fact their sum is

$$\sum_{i=0}^{\infty}(1-p)^ip = p\sum_{i=0}^{\infty}(1-p)^i = p\frac{1}{1-(1-p)} = \frac{p}{p} = 1.$$

[5]for those who are familiar with the concept of convergence for infinite sums (i.e. infinite series), it is worth noting that it is the fact that probability weights cannot be negative and must add to one that makes all the sums we need to deal with for all the theorems we have proved so far converge. That doesn't mean all sums we might want to deal with will converge; some random variables defined on the sample space we have described will have infinite expected value. However those we need to deal with for the expected number of trials until success do converge.

Therefore we have a legitimate assignment of probabilities and the set of sequences

$$\{F, FS, FFS, FFFS, \ldots, F^iS, \ldots\}$$

is a sample space with these probability weights. This probability distribution, $P(F^iS) = (1-p)^i p$, is called a *geometric* distribution because of the geometric series we used in proving the probabilities sum to 1.

Theorem 6.12 *Suppose we have a sequence of trials in which each trial has two outcomes, success and failure, and where at each step the probability of success is p. Then the expected number of trials until the first success is $1/p$.*

Proof:

We consider the random variable X which is i if the first success is on trial i. (In other words, $X(F^{i-1}S)$ is i.) The probability that the first success is on trial i is $(1-p)^{i-1}p$, since in order for this to happen there must be $i-1$ failures followed by 1 success. The expected number of trials is the expected value of X, which is, by the definition of expected value and the previous two sentences,

$$\begin{aligned}
E[\text{number of trials}] &= \sum_{i=0}^{\infty} p(1-p)^{i-1} i \\
&= p \sum_{i=0}^{\infty} (1-p)^{i-1} i \\
&= \frac{p}{1-p} \sum_{i=0}^{\infty} (1-p)^i i \\
&= \frac{p}{1-p} \frac{1-p}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

To go from the third to the fourth line we used the fact that

$$\sum_{j=0}^{\infty} j x^j = \frac{x}{(1-x)^2}, \tag{6.25}$$

true for x with absolute value less than one. We proved a finite version of this equation as Theorem 4.6; the infinite version is even easier to prove. ■

Applying this theorem, we see that the expected number of times you need to flip a coin until you get heads is 2, and the expected number of times you need to roll two dice until you get a seven is 6.

Important Concepts, Formulas, and Theorems

1. *Random Variable.* A *random variable* for an experiment with a sample space S is a function that assigns a number to each element of S.

2. Bernoulli Trials Process. An independent trials process with two outcomes, success and failure, at each stage and probability p of success and $1-p$ of failure at each stage is called a *Bernoulli trials process.*

3. Probability of a Sequence of Bernoulli Trials. In n Bernoulli trials with probability p of success, the probability of a given sequence of k successes and $n-k$ failures is $p^k(1-p)^{n-k}$.

4. *The Probability of k Successes in n Bernoulli Trials* The probability of having exactly k successes in a sequence of n independent trials with two outcomes and probability p of success on each trial is

$$P(\text{exactly } k \text{ successes}) = \binom{n}{k} p^k (1-p)^{n-k}$$

.

5. Binomial Probability Distribution. The probabilities of of k successes in n Bernoulli trials, $\binom{n}{k}p^k(1-p)^{n-k}$, are called *binomial probabilities*, or the *binomial probability distribution.*

6. *Generating Function.* The *generating function* for the sequence $a_0, a_1, a_2, \ldots, a_n$ is $\sum_{i=1}^{n} a_i x^i$, and the *generating function* for an infinite sequence $a_0, a_1, a_2, \ldots, a_n, \ldots$ is the infinite series $\sum_{i=1}^{\infty} a_i x^i$. The polynomial $(px + 1 - p)^n$ is the generating function for the binomial probabilities for n Bernoulli trials with probability p of success.

7. *Distribution Function.* We call the function that assigns $P(x_i)$ to the event $P(X = x_i)$ the *distribution function* of the random variable X.

8. *Expected Value.* We define the *expected value* or *expectation* of a random variable X whose values are the set $\{x_1, x_2, \ldots x_k\}$ to be

$$E(X) = \sum_{i=1}^{k} x_i P(X = x_i).$$

9. *Another Formula for Expected Values.* If a random variable X is defined on a (finite) sample space S, then its expected value is given by

$$E(X) = \sum_{s:s\in S} X(s)P(s).$$

10. *Expected Value of a Sum.* Suppose X and Y are random variables on the (finite) sample space S. Then

$$E(X + Y) = E(X) + E(Y).$$

This is called the *additivity of expectation.*

11. *Expected Value of a Numerical Multiple.* Suppose X is a random variable on a sample space S. Then for any number c, $E(cX) = cE(X)$. This result and the additivity of expectation together are called the *linearity of expectation.*

12. *Expected Number of Successes in Bernoulli Trials.* In a Bernoulli trials process, in which each experiment has two outcomes and probability p of success, the expected number of successes is np.

13. *Expected Number of Trials Until Success.* Suppose we have a sequence of trials in which each trial has two outcomes, success and failure, and where at each step the probability of success is p. Then the expected number of trials until the first success is $1/p$.

Problems

1. Give several random variables that might be of interest to someone rolling five dice (as one does, for example, in the game Yatzee).

2. Suppose I offer to play the following game with you if you will pay me some money. You roll a die, and I give you a dollar for each dot that is on top. What is the maximum amount of money a rational person might be willing to pay me in order to play this game?

3. How many sixes do we expect to see on top if we roll 24 dice?

4. What is the expected sum of the tops of n dice when we roll them?

5. In an independent trials process consisting of six trials with probability p of success, what is the probability that the first three trials are successes and the last three are failures? The probability that the last three trials are successes and the first three are failures? The probability that trials 1, 3, and 5 are successes and trials 2, 4, and 6 are failures? What is the probability of three successes and three failures?

6. What is the probability of exactly eight heads in ten flips of a coin? Of eight or more heads?

7. How many times do you expect to have to role a die until you see a six on the top face?

8. Assuming that the process of answering the questions on a five-question quiz is an independent trials process and that a student has a probability of .8 of answering any given question correctly, what is the probability of a sequence of four correct answers and one incorrect answer? What is the probability that a student answers exactly four questions correctly?

9. What is the expected value of the constant random variable X that has $X(s) = c$ for every member s of the sample space? We frequently just use c to stand for this random variable, and thus this question is asking for $E(c)$.

10. Someone is taking a true-false test and guessing when they don't know the answer. We are going to compute a score by subtracting a percentage of the number of incorrect answers from the number of correct answers. When we convert this "corrected score" to a percentage score we want its expected value to be the percentage of the material being tested that the test-taker knows. How can we do this?

11. Do Problem 10 of this section for the case that someone is taking a multiple choice test with five choices for each answer and guesses randomly when they don't know the answer.

12. Suppose we have ten independent trials with three outcomes called good, bad, and indifferent, with probabilities p, q, and r, respectively. What is the probability of three goods, two bads, and five indifferents? In n independent trials with three outcomes A, B, and C, with probabilities p, q, and r, what is the probability of i As, j Bs, and k Cs? (In this problem we assume $p + q + r = 1$ and $i + j + k = n$.)

13. In as many ways as you can, prove that

$$\sum_{i=0}^{n} i\binom{n}{i} = 2^{n-1}n.$$

14. Prove Theorem 6.10.

15. Two nickels, two dimes, and two quarters are in a cup. We draw three coins, one after the other, without replacement. What is the expected amount of money we draw on the first draw? On the second draw? What is the expected value of the total amount of money we draw? Does this expected value change if we draw the three coins all together?

16. In this exercise we will evaluate the sum

$$\sum_{i=0}^{10} i\binom{10}{i}(.9)^i(.1)^{10-i}$$

that arose in computing the expected number of right answers a person would have on a ten question test with probability .9 of answering each question correctly. First, use the binomial theorem and calculus to show that

$$10(.1+x)^9 = \sum_{i=0}^{10} i\binom{10}{i}(.1)^{10-i}x^{i-1}$$

Substituting in $x = .9$ gives us almost the sum we want on the right hand side of the equation, except that in every term of the sum the power on .9 is one too small. Use some simple algebra to fix this and then explain why the expected number of right answers is 9.

17. Give an example of two random variables X and Y such that $E(XY) \neq E(X)E(Y)$. Here XY is the random variable with $(XY)(s) = X(s)Y(s)$.

18. Prove that if X and Y are independent in the sense that the event that $X = x$ and the event that $Y = y$ are independent for each pair of values x of X and y of Y, then $E(XY) = E(X)E(Y)$. See Exercise 6-17 for a definition of XY.

19. Use calculus and the sum of a geometric series to show that

$$\sum_{j=0}^{\infty} jx^j = \frac{x}{(1-x)^2}$$

as in Equation 6.25.

20. Give an example of a random variable on the sample space $\{S, FS, FFS, \ldots, F^iS, \ldots\}$ with an infinite expected value.