Chapter 5

Important Distributions and Densities

5.1 Important Distributions

In this chapter, we describe the discrete probability distributions and the continuous probability densities that occur most often in the analysis of experiments. We will also show how one simulates these distributions and densities on a computer.

Discrete Uniform Distribution

In Chapter 1, we saw that in many cases, we assume that all outcomes of an experiment are equally likely. If X is a random variable which represents the outcome of an experiment of this type, then we say that X is uniformly distributed. If the sample space S is of size n, where $0 < n < \infty$, then the distribution function $m(\omega)$ is defined to be $1/n$ for all $\omega \in S$. As is the case with all of the discrete probability distributions discussed in this chapter, this experiment can be simulated on a computer using the program **GeneralSimulation**. However, in this case, a faster algorithm can be used instead. (This algorithm was described in Chapter 1; we repeat the description here for completeness.) The expression

$$1 + \lfloor n\,(rnd) \rfloor$$

takes on as a value each integer between 1 and n with probability $1/n$ (the notation $\lfloor x \rfloor$ denotes the greatest integer not exceeding x). Thus, if the possible outcomes of the experiment are labelled $\omega_1\ \omega_2,\ \ldots,\ \omega_n$, then we use the above expression to represent the subscript of the output of the experiment.

If the sample space is a countably infinite set, such as the set of positive integers, then it is not possible to have an experiment which is uniform on this set (see Exercise 3). If the sample space is an uncountable set, such as the set of real numbers, then we use continuous density functions (see Section 5.2).

Binomial Distribution

The binomial distribution with parameters n, p, and k was defined in Chapter 3. It is the distribution of the random variable which counts the number of heads which occur when a coin is tossed n times, assuming that on any one toss, the probability that a head occurs is p. The distribution function is given by the formula

$$b(n, p, k) = \binom{n}{k} p^k q^{n-k} \ ,$$

where $q = 1 - p$.

One straightforward way to simulate a binomial random variable X is to compute the sum of n independent $0-1$ random variables, each of which take on the value 1 with probability p. This method requires n calls to a random number generator to obtain one value of the random variable. When n is relatively large (say at least 30), the Central Limit Theorem (see Chapter 9) implies that the binomial distribution is well-approximated by the corresponding normal density function (which is defined in Section 5.2) with parameters $\mu = np$ and $\sigma = \sqrt{npq}$. Thus, in this case we can compute a value Y of a normal random variable with these parameters, and if $-1/2 \leq Y < n + 1/2$, we can use the value

$$\lfloor Y + 1/2 \rfloor$$

to represent the random variable X. If $Y < -1/2$ or $Y > n + 1/2$, we reject Y and compute another value. We will see in the next section how we can quickly simulate normal random variables.

Geometric Distribution

Consider a Bernoulli trials process continued for an infinite number of trials; for example, a coin tossed an infinite sequence of times. We showed in Section 2.2 how to assign a probability measure to the infinite tree. Thus, we can determine the distribution for any random variable X relating to the experiment provided $P(X = a)$ can be computed in terms of a finite number of trials. For example, let T be the number of trials up to and including the first success. Then

$$\begin{aligned} P(T = 1) &= p \ , \\ P(T = 2) &= qp \ , \\ P(T = 3) &= q^2 p \ , \end{aligned}$$

and in general,

$$P(T = n) = q^{n-1} p \ .$$

To show that this is a distribution, we must show that

$$p + qp + q^2 p + \cdots = 1 \ .$$

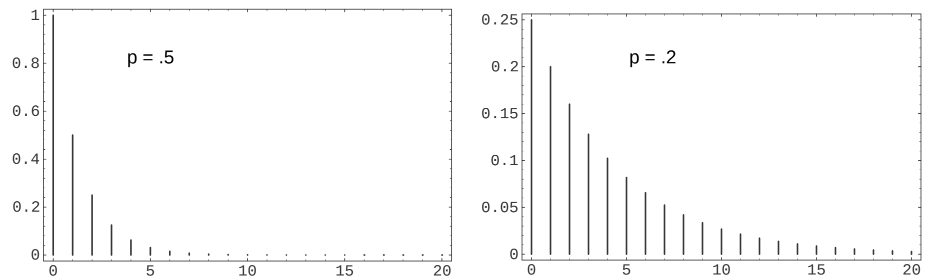

Figure 5.1: Geometric distributions.

The left-hand expression is just a geometric series with first term p and common ratio q, so its sum is

$$\frac{p}{1-q}$$

which equals 1.

In Figure 5.1 we have plotted this distribution using the program **GeometricPlot** for the cases $p = .5$ and $p = .2$. We see that as p decreases we are more likely to get large values for T, as would be expected. In both cases, the most probable value for T is 1. This will always be true since

$$\frac{P(T = j+1)}{P(T = j)} = q < 1 \ .$$

In general, if $0 < p < 1$, and $q = 1 - p$, then we say that the random variable T has a *geometric distribution* if

$$P(T = j) = q^{j-1}p \ ,$$

for $j = 1,\ 2,\ 3,\ \ldots$.

To simulate the geometric distribution with parameter p, we can simply compute a sequence of random numbers in $[0, 1)$, stopping when an entry does not exceed p. However, for small values of p, this is time-consuming (taking, on the average, $1/p$ steps). We now describe a method whose running time does not depend upon the size of p. Let X be a geometrically distributed random variable with parameter p, where $0 < p < 1$. Now, define Y to be the smallest integer satisfying the inequality

$$1 - q^Y \geq rnd \ . \tag{5.1}$$

Then we have

$$\begin{aligned} P(Y = j) &= P\Big(1 - q^j \geq rnd > 1 - q^{j-1}\Big) \\ &= q^{j-1} - q^j \\ &= q^{j-1}(1-q) \\ &= q^{j-1}p \ . \end{aligned}$$

Thus, Y is geometrically distributed with parameter p. To generate Y, all we have to do is solve Equation 5.1 for Y. We obtain

$$Y = \left\lfloor \frac{\log(1 - rnd)}{\log\ q} \right\rfloor .$$

Since $\log(1 - rnd)$ and $\log(rnd)$ are identically distributed, Y can also be generated using the equation

$$Y = \left\lfloor \frac{\log\ rnd}{\log\ q} \right\rfloor .$$

Example 5.1 The geometric distribution plays an important role in the theory of queues, or waiting lines. For example, suppose a line of customers waits for service at a counter. It is often assumed that, in each small time unit, either 0 or 1 new customers arrive at the counter. The probability that a customer arrives is p and that no customer arrives is $q = 1 - p$. Then the time T until the next arrival has a geometric distribution. It is natural to ask for the probability that no customer arrives in the next k time units, that is, for $P(T > k)$. This is given by

$$\begin{aligned} P(T > k) = \sum_{j=k+1}^{\infty} q^{j-1}p &= q^k(p + qp + q^2p + \cdots) \\ &= q^k . \end{aligned}$$

This probability can also be found by noting that we are asking for no successes (i.e., arrivals) in a sequence of k consecutive time units, where the probability of a success in any one time unit is p. Thus, the probability is just q^k, since arrivals in any two time units are independent events.

It is often assumed that the length of time required to service a customer also has a geometric distribution but with a different value for p. This implies a rather special property of the service time. To see this, let us compute the conditional probability

$$P(T > r + s \,|\, T > r) = \frac{P(T > r + s)}{P(T > r)} = \frac{q^{r+s}}{q^r} = q^s .$$

Thus, the probability that the customer's service takes s more time units is independent of the length of time r that the customer has already been served. Because of this interpretation, this property is called the "memoryless" property, and is also obeyed by the exponential distribution. (Fortunately, not too many service stations have this property.) □

Negative Binomial Distribution

Suppose we are given a coin which has probability p of coming up heads when it is tossed. We fix a positive integer k, and toss the coin until the kth head appears. We let X represent the number of tosses. When $k = 1$, X is geometrically distributed.

For a general k, we say that X has a negative binomial distribution. We now calculate the probability distribution of X. If $X = x$, then it must be true that there were exactly $k - 1$ heads thrown in the first $x - 1$ tosses, and a head must have been thrown on the xth toss. There are

$$\binom{x-1}{k-1}$$

sequences of length x with these properties, and each of them is assigned the same probability, namely

$$p^{k-1}q^{x-k} .$$

Therefore, if we define

$$u(x, k, p) = P(X = x) ,$$

then

$$u(x, k, p) = \binom{x-1}{k-1} p^k q^{x-k} .$$

One can simulate this on a computer by simulating the tossing of a coin. The following algorithm is, in general, much faster. We note that X can be understood as the sum of k outcomes of a geometrically distributed experiment with parameter p. Thus, we can use the following sum as a means of generating X:

$$\sum_{j=1}^{k} \left\lfloor \frac{\log\ rnd_j}{\log\ q} \right\rfloor .$$

Example 5.2 A fair coin is tossed until the second time a head turns up. The distribution for the number of tosses is $u(x, 2, p)$. Thus the probability that x tosses are needed to obtain two heads is found by letting $k = 2$ in the above formula. We obtain

$$u(x, 2, 1/2) = \binom{x-1}{1} \frac{1}{2^x} ,$$

for $x = 2, 3, \ldots$.

In Figure 5.2 we give a graph of the distribution for $k = 2$ and $p = .25$. Note that the distribution is quite asymmetric, with a long tail reflecting the fact that large values of x are possible. □

Poisson Distribution

The Poisson distribution arises in many situations. It is safe to say that it is one of the three most important discrete probability distributions (the other two being the uniform and the binomial distributions). The Poisson distribution can be viewed as arising from the binomial distribution or from the exponential density. We shall now explain its connection with the former; its connection with the latter will be explained in the next section.

Suppose that we have a situation in which a certain kind of occurrence happens at random over a period of time. For example, the occurrences that we are interested

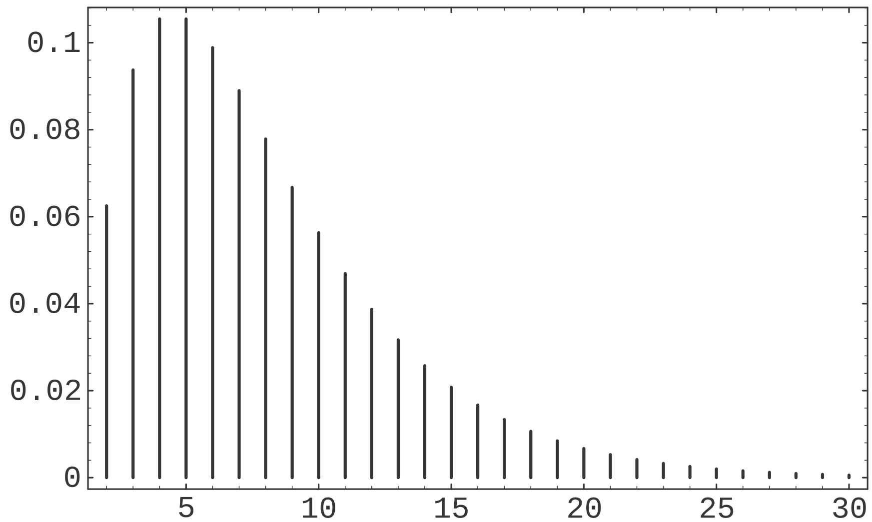

Figure 5.2: Negative binomial distribution with $k = 2$ and $p = .25$.

in might be incoming telephone calls to a police station in a large city. We want to model this situation so that we can consider the probabilities of events such as more than 10 phone calls occurring in a 5-minute time interval. Presumably, in our example, there would be more incoming calls between 6:00 and 7:00 P.M. than between 4:00 and 5:00 A.M., and this fact would certainly affect the above probability. Thus, to have a hope of computing such probabilities, we must assume that the average rate, i.e., the average number of occurrences per minute, is a constant. This rate we will denote by λ. (Thus, in a given 5-minute time interval, we would expect about 5λ occurrences.) This means that if we were to apply our model to the two time periods given above, we would simply use different rates for the two time periods, thereby obtaining two different probabilities for the given event.

Our next assumption is that the number of occurrences in two non-overlapping time intervals are independent. In our example, this means that the events that there are j calls between 5:00 and 5:15 P.M. and k calls between 6:00 and 6:15 P.M. on the same day are independent.

We can use the binomial distribution to model this situation. We imagine that a given time interval is broken up into n subintervals of equal length. If the subintervals are sufficiently short, we can assume that two or more occurrences happen in one subinterval with a probability which is negligible in comparison with the probability of at most one occurrence. Thus, in each subinterval, we are assuming that there is either 0 or 1 occurrence. This means that the sequence of subintervals can be thought of as a sequence of Bernoulli trials, with a success corresponding to an occurrence in the subinterval.

To decide upon the proper value of p, the probability of an occurrence in a given subinterval, we reason as follows. On the average, there are λt occurrences in a time interval of length t. If this time interval is divided into n subintervals, then we would expect, using the Bernoulli trials interpretation, that there should be np occurrences. Thus, we want

$$\lambda t = np \ ,$$

so

$$p = \frac{\lambda t}{n} \ .$$

We now wish to consider the random variable X, which counts the number of occurrences in a given time interval. We want to calculate the distribution of X. For ease of calculation, we will assume that the time interval is of length 1; for time intervals of arbitrary length t, see Exercise 11. We know that

$$P(X = 0) = b(n, p, 0) = (1 - p)^n = \Bigl(1 - \frac{\lambda}{n}\Bigr)^n \ .$$

For large n, this is approximately $e^{-\lambda}$. It is easy to calculate that for any fixed k, we have

$$\frac{b(n, p, k)}{b(n, p, k - 1)} = \frac{\lambda - (k - 1)p}{kq}$$

which, for large n (and therefore small p) is approximately λ/k. Thus, we have

$$P(X = 1) \approx \lambda e^{-\lambda} \ ,$$

and in general,

$$P(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda} \ . \tag{5.2}$$

The above distribution is the Poisson distribution. We note that it must be checked that the distribution given in Equation 5.2 really *is* a distribution, i.e., that its values are non-negative and sum to 1. (See Exercise 12.)

The Poisson distribution is used as an approximation to the binomial distribution when the parameters n and p are large and small, respectively (see Examples 5.3 and 5.4). However, the Poisson distribution also arises in situations where it may not be easy to interpret or measure the parameters n and p (see Example 5.5).

Example 5.3 A typesetter makes, on the average, one mistake per 1000 words. Assume that he is setting a book with 100 words to a page. Let S_{100} be the number of mistakes that he makes on a single page. Then the exact probability distribution for S_{100} would be obtained by considering S_{100} as a result of 100 Bernoulli trials with $p = 1/1000$. The expected value of S_{100} is $\lambda = 100(1/1000) = .1$. The exact probability that $S_{100} = j$ is $b(100, 1/1000, j)$, and the Poisson approximation is

$$\frac{e^{-.1}(.1)^j}{j!}.$$

In Table 5.1 we give, for various values of n and p, the exact values computed by the binomial distribution and the Poisson approximation. □

j	Poisson $\lambda = .1$	Binomial $n = 100$ $p = .001$	Poisson $\lambda = 1$	Binomial $n = 100$ $p = .01$	Poisson $\lambda = 10$	Binomial $n = 1000$ $p = .01$
0	.9048	.9048	.3679	.3660	.0000	.0000
1	.0905	.0905	.3679	.3697	.0005	.0004
2	.0045	.0045	.1839	.1849	.0023	.0022
3	.0002	.0002	.0613	.0610	.0076	.0074
4	.0000	.0000	.0153	.0149	.0189	.0186
5			.0031	.0029	.0378	.0374
6			.0005	.0005	.0631	.0627
7			.0001	.0001	.0901	.0900
8			.0000	.0000	.1126	.1128
9					.1251	.1256
10					.1251	.1257
11					.1137	.1143
12					.0948	.0952
13					.0729	.0731
14					.0521	.0520
15					.0347	.0345
16					.0217	.0215
17					.0128	.0126
18					.0071	.0069
19					.0037	.0036
20					.0019	.0018
21					.0009	.0009
22					.0004	.0004
23					.0002	.0002
24					.0001	.0001
25					.0000	.0000

Table 5.1: Poisson approximation to the binomial distribution.

Example 5.4 In his book,[1] Feller discusses the statistics of flying bomb hits in the south of London during the Second World War.

Assume that you live in a district of size 10 blocks by 10 blocks so that the total district is divided into 100 small squares. How likely is it that the square in which you live will receive no hits if the total area is hit by 400 bombs?

We assume that a particular bomb will hit your square with probability 1/100. Since there are 400 bombs, we can regard the number of hits that your square receives as the number of *successes* in a Bernoulli trials process with $n = 400$ and $p = 1/100$. Thus we can use the Poisson distribution with $\lambda = 400 \cdot 1/100 = 4$ to approximate the probability that your square will receive j hits. This probability is $p(j) = e^{-4}4^j/j!$. The expected number of squares that receive exactly j hits is then $100 \cdot p(j)$. It is easy to write a program **LondonBombs** to simulate this situation and compare the expected number of squares with j hits with the observed number. In Exercise 26 you are asked to compare the actual observed data with that predicted by the Poisson distribution.

In Figure 5.3, we have shown the simulated hits, together with a spike graph showing both the observed and predicted frequencies. The observed frequencies are shown as squares, and the predicted frequencies are shown as dots. □

If the reader would rather not consider flying bombs, he is invited to instead consider an analogous situation involving cookies and raisins. We assume that we have made enough cookie dough for 500 cookies. We put 600 raisins in the dough, and mix it thoroughly. One way to look at this situation is that we have 500 cookies, and after placing the cookies in a grid on the table, we throw 600 raisins at the cookies. (See Exercise 22.)

Example 5.5 Suppose that in a certain fixed amount A of blood, the average human has 40 white blood cells. Let X be the random variable which gives the number of white blood cells in a random sample of size A from a random individual. We can think of X as binomially distributed with each white blood cell in the body representing a trial. If a given white blood cell turns up in the sample, then the trial corresponding to that blood cell was a success. Then p should be taken as the ratio of A to the total amount of blood in the individual, and n will be the number of white blood cells in the individual. Of course, in practice, neither of these parameters is very easy to measure accurately, but presumably the number 40 is easy to measure. But for the average human, we then have $40 = np$, so we can think of X as being Poisson distributed, with parameter $\lambda = 40$. In this case, it is easier to model the situation using the Poisson distribution than the binomial distribution. □

To simulate a Poisson random variable on a computer, a good way is to take advantage of the relationship between the Poisson distribution and the exponential density. This relationship and the resulting simulation algorithm will be described in the next section.

[1] ibid., p. 161.

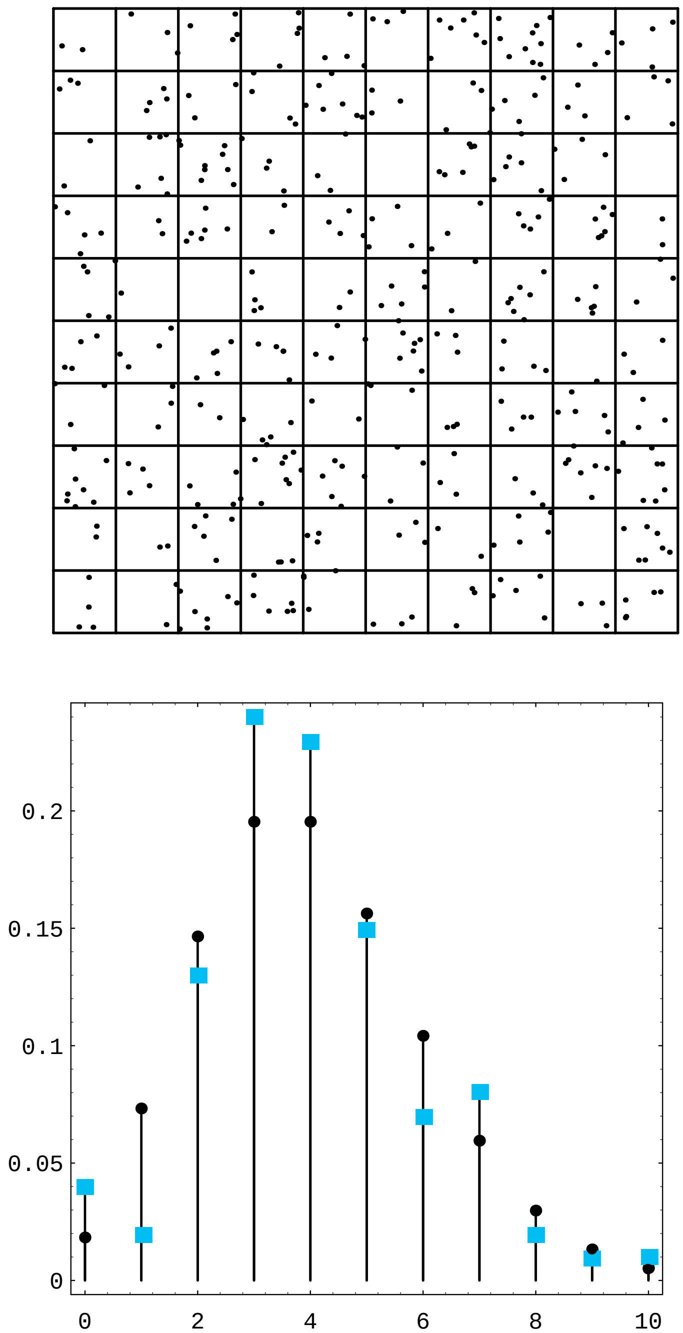

Figure 5.3: Flying bomb hits.

Hypergeometric Distribution

Suppose that we have a set of N balls, of which k are red and $N - k$ are blue. We choose n of these balls, without replacement, and define X to be the number of red balls in our sample. The distribution of X is called the hypergeometric distribution. We note that this distribution depends upon three parameters, namely N, k, and n. There does not seem to be a standard notation for this distribution; we will use the notation $h(N, k, n, x)$ to denote $P(X = x)$. This probability can be found by noting that there are

$$\binom{N}{n}$$

different samples of size n, and the number of such samples with exactly x red balls is obtained by multiplying the number of ways of choosing x red balls from the set of k red balls and the number of ways of choosing $n - x$ blue balls from the set of $N - k$ blue balls. Hence, we have

$$h(N, k, n, x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \ .$$

This distribution can be generalized to the case where there are more than two types of objects. (See Exercise 40.)

If we let N and k tend to ∞, in such a way that the ratio k/N remains fixed, then the hypergeometric distribution tends to the binomial distribution with parameters n and $p = k/N$. This is reasonable because if N and k are much larger than n, then whether we choose our sample with or without replacement should not affect the probabilities very much, and the experiment consisting of choosing with replacement yields a binomially distributed random variable (see Exercise 44).

An example of how this distribution might be used is given in Exercises 36 and 37. We now give another example involving the hypergeometric distribution. It illustrates a statistical test called Fisher's Exact Test.

Example 5.6 It is often of interest to consider two traits, such as eye color and hair color, and to ask whether there is an association between the two traits. Two traits are associated if knowing the value of one of the traits for a given person allows us to predict the value of the other trait for that person. The stronger the association, the more accurate the predictions become. If there is no association between the traits, then we say that the traits are independent. In this example, we will use the traits of gender and political party, and we will assume that there are only two possible genders, female and male, and only two possible political parties, Democratic and Republican.

Suppose that we have collected data concerning these traits. To test whether there is an association between the traits, we first assume that there is no association between the two traits. This gives rise to an "expected" data set, in which knowledge of the value of one trait is of no help in predicting the value of the other trait. Our collected data set usually differs from this expected data set. If it differs by quite a bit, then we would tend to reject the assumption of independence of the traits. To

	Democrat	Republican	
Female	24	4	28
Male	8	14	22
	32	18	50

Table 5.2: Observed data.

	Democrat	Republican	
Female	s_{11}	s_{12}	t_{11}
Male	s_{21}	s_{22}	t_{12}
	t_{21}	t_{22}	n

Table 5.3: General data table.

nail down what is meant by "quite a bit," we decide which possible data sets differ from the expected data set by at least as much as ours does, and then we compute the probability that any of these data sets would occur under the assumption of independence of traits. If this probability is small, then it is unlikely that the difference between our collected data set and the expected data set is due entirely to chance.

Suppose that we have collected the data shown in Table 5.2. The row and column sums are called marginal totals, or marginals. In what follows, we will denote the row sums by t_{11} and t_{12}, and the column sums by t_{21} and t_{22}. The ijth entry in the table will be denoted by s_{ij}. Finally, the size of the data set will be denoted by n. Thus, a general data table will look as shown in Table 5.3. We now explain the model which will be used to construct the "expected" data set. In the model, we assume that the two traits are independent. We then put t_{21} yellow balls and t_{22} green balls, corresponding to the Democratic and Republican marginals, into an urn. We draw t_{11} balls, without replacement, from the urn, and call these balls females. The t_{12} balls remaining in the urn are called males. In the specific case under consideration, the probability of getting the actual data under this model is given by the expression

$$\frac{\binom{32}{24}\binom{18}{4}}{\binom{50}{28}} ,$$

i.e., a value of the hypergeometric distribution.

We are now ready to construct the expected data set. If we choose 28 balls out of 50, we should expect to see, on the average, the same percentage of yellow balls in our sample as in the urn. Thus, we should expect to see, on the average, $28(32/50) = 17.92 \approx 18$ yellow balls in our sample. (See Exercise 36.) The other expected values are computed in exactly the same way. Thus, the expected data set is shown in Table 5.4. We note that the value of s_{11} determines the other three values in the table, since the marginals are all fixed. Thus, in considering the possible data sets that could appear in this model, it is enough to consider the various possible values of s_{11}. In the specific case at hand, what is the probability

	Democrat	Republican	
Female	18	10	28
Male	14	8	22
	32	18	50

Table 5.4: Expected data.

of drawing exactly a yellow balls, i.e., what is the probability that $s_{11} = a$? It is

$$\frac{\binom{32}{a}\binom{18}{28-a}}{\binom{50}{28}} . \tag{5.3}$$

We are now ready to decide whether our actual data differs from the expected data set by an amount which is greater than could be reasonably attributed to chance alone. We note that the expected number of female Democrats is 18, but the actual number in our data is 24. The other data sets which differ from the expected data set by more than ours correspond to those where the number of female Democrats equals 25, 26, 27, or 28. Thus, to obtain the required probability, we sum the expression in (5.3) from $a = 24$ to $a = 28$. We obtain a value of .000395. Thus, we should reject the hypothesis that the two traits are independent. □

Finally, we turn to the question of how to simulate a hypergeometric random variable X. Let us assume that the parameters for X are N, k, and n. We imagine that we have a set of N balls, labelled from 1 to N. We decree that the first k of these balls are red, and the rest are blue. Suppose that we have chosen m balls, and that j of them are red. Then there are $k - j$ red balls left, and $N - m$ balls left. Thus, our next choice will be red with probability

$$\frac{k-j}{N-m} .$$

So at this stage, we choose a random number in $[0, 1]$, and report that a red ball has been chosen if and only if the random number does not exceed the above expression. Then we update the values of m and j, and continue until n balls have been chosen.

Benford Distribution

Our next example of a distribution comes from the study of leading digits in data sets. It turns out that many data sets that occur "in real life" have the property that the first digits of the data are not uniformly distributed over the set $\{1, 2, \ldots, 9\}$. Rather, it appears that the digit 1 is most likely to occur, and that the distribution is monotonically decreasing on the set of possible digits. The Benford distribution appears, in many cases, to fit such data. Many explanations have been given for the occurrence of this distribution. Possibly the most convincing explanation is that this distribution is the only one that is invariant under a change of scale. If one thinks of certain data sets as somehow "naturally occurring," then the distribution should be unaffected by which units are chosen in which to represent the data, i.e., the distribution should be invariant under change of scale.

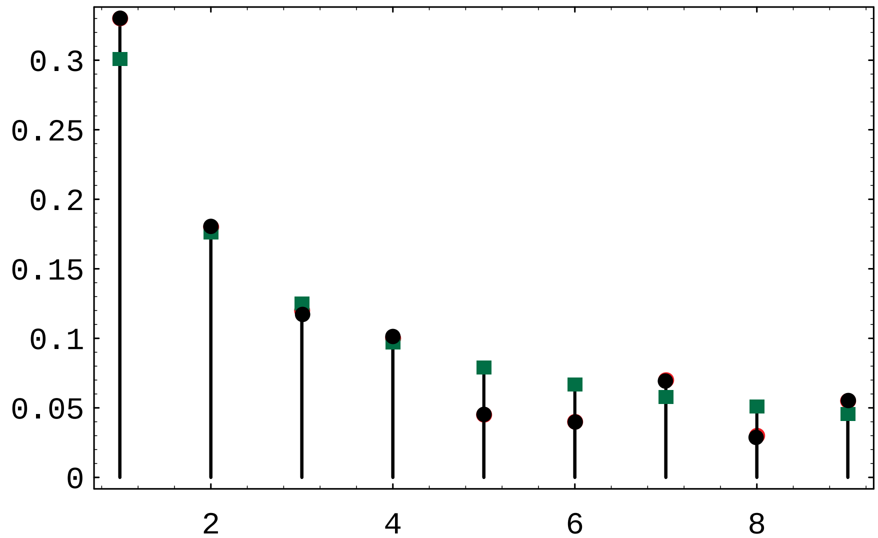

Figure 5.4: Leading digits in President Clinton's tax returns.

Theodore Hill[2] gives a general description of the Benford distribution, when one considers the first d digits of integers in a data set. We will restrict our attention to the first digit. In this case, the Benford distribution has distribution function

$$f(k) = \log_{10}(k+1) - \log_{10}(k) ,$$

for $1 \le k \le 9$.

Mark Nigrini[3] has advocated the use of the Benford distribution as a means of testing suspicious financial records such as bookkeeping entries, checks, and tax returns. His idea is that if someone were to "make up" numbers in these cases, the person would probably produce numbers that are fairly uniformly distributed, while if one were to use the actual numbers, the leading digits would roughly follow the Benford distribution. As an example, Negrini analyzed President Clinton's tax returns for a 13-year period. In Figure 5.4, the Benford distribution values are shown as squares, and the President's tax return data are shown as circles. One sees that in this example, the Benford distribution fits the data very well.

This distribution was discovered by the astronomer Simon Newcomb who stated the following in his paper on the subject: "That the ten digits do not occur with equal frequency must be evident to anyone making use of logarithm tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9."[4]

[2]T. P. Hill, "The Significant Digit Phenomenon," *American Mathematical Monthly,* vol. 102, no. 4 (April 1995), pgs. 322-327.

[3]M. Nigrini, "Detecting Biases and Irregularities in Tabulated Data," working paper

[4]S. Newcomb, "Note on the frequency of use of the different digits in natural numbers," *American Journal of Mathematics,* vol. 4 (1881), pgs. 39-40.

Exercises

1 For which of the following random variables would it be appropriate to assign a uniform distribution?

(a) Let X represent the roll of one die.

(b) Let X represent the number of heads obtained in three tosses of a coin.

(c) A roulette wheel has 38 possible outcomes: 0, 00, and 1 through 36. Let X represent the outcome when a roulette wheel is spun.

(d) Let X represent the birthday of a randomly chosen person.

(e) Let X represent the number of tosses of a coin necessary to achieve a head for the first time.

2 Let n be a positive integer. Let S be the set of integers between 1 and n. Consider the following process: We remove a number from S and write it down. We repeat this until S is empty. The result is a permutation of the integers from 1 to n. Let X denote this permutation. Is X uniformly distributed?

3 Let X be a random variable which can take on countably many values. Show that X cannot be uniformly distributed.

4 Suppose we are attending a college which has 3000 students. We wish to choose a subset of size 100 from the student body. Let X represent the subset, chosen using the following possible strategies. For which strategies would it be appropriate to assign the uniform distribution to X? If it is appropriate, what probability should we assign to each outcome?

(a) Take the first 100 students who enter the cafeteria to eat lunch.

(b) Ask the Registrar to sort the students by their Social Security number, and then take the first 100 in the resulting list.

(c) Ask the Registrar for a set of cards, with each card containing the name of exactly one student, and with each student appearing on exactly one card. Throw the cards out of a third-story window, then walk outside and pick up the first 100 cards that you find.

5 Under the same conditions as in the preceding exercise, can you describe a procedure which, if used, would produce each possible outcome with the same probability? Can you describe such a procedure that does not rely on a computer or a calculator?

6 Let $X_1,\ X_2,\ \ldots,\ X_n$ be n mutually independent random variables, each of which is uniformly distributed on the integers from 1 to k. Let Y denote the minimum of the X_i's. Find the distribution of Y.

7 A die is rolled until the first time T that a six turns up.

(a) What is the probability distribution for T?

(b) Find $P(T > 3)$.

(c) Find $P(T > 6|T > 3)$.

8 If a coin is tossed a sequence of times, what is the probability that the first head will occur after the fifth toss, given that it has not occurred in the first two tosses?

9 A worker for the Department of Fish and Game is assigned the job of estimating the number of trout in a certain lake of modest size. She proceeds as follows: She catches 100 trout, tags each of them, and puts them back in the lake. One month later, she catches 100 more trout, and notes that 10 of them have tags.

(a) Without doing any fancy calculations, give a rough estimate of the number of trout in the lake.

(b) Let N be the number of trout in the lake. Find an expression, in terms of N, for the probability that the worker would catch 10 tagged trout out of the 100 trout that she caught the second time.

(c) Find the value of N which maximizes the expression in part (b). This value is called the *maximum likelihood estimate* for the unknown quantity N. *Hint*: Consider the ratio of the expressions for successive values of N.

10 A census in the United States is an attempt to count everyone in the country. It is inevitable that many people are not counted. The U. S. Census Bureau proposed a way to estimate the number of people who were not counted by the latest census. Their proposal was as follows: In a given locality, let N denote the actual number of people who live there. Assume that the census counted n_1 people living in this area. Now, another census was taken in the locality, and n_2 people were counted. In addition, n_{12} people were counted both times.

(a) Given N, n_1, and n_2, let X denote the number of people counted both times. Find the probability that $X = k$, where k is a fixed positive integer between 0 and n_2.

(b) Now assume that $X = n_{12}$. Find the value of N which maximizes the expression in part (a). *Hint*: Consider the ratio of the expressions for successive values of N.

11 Suppose that X is a random variable which represents the number of calls coming in to a police station in a one-minute interval. In the text, we showed that X could be modelled using a Poisson distribution with parameter λ, where this parameter represents the average number of incoming calls per minute. Now suppose that Y is a random variable which represents the number of incoming calls in an interval of length t. Show that the distribution of Y is given by

$$P(Y = k) = e^{-k\lambda}\frac{(\lambda t)^k}{k!}\ ,$$

i.e., Y is Poisson with parameter λt. *Hint*: Suppose a Martian were to observe the police station. Let us also assume that the basic time interval used on Mars is exactly t Earth minutes. Finally, we will assume that the Martian understands the derivation of the Poisson distribution in the text. What would she write down for the distribution of Y?

12 Show that the values of the Poisson distribution given in Equation 5.2 sum to 1.

13 The Poisson distribution with parameter $\lambda = .3$ has been assigned for the outcome of an experiment. Let X be the outcome function. Find $P(X = 0)$, $P(X = 1)$, and $P(X > 1)$.

14 On the average, only 1 person in 1000 has a particular rare blood type.

(a) Find the probability that, in a city of 10,000 people, no one has this blood type.

(b) How many people would have to be tested to give a probability greater than 1/2 of finding at least one person with this blood type?

15 Write a program for the user to input n, p, j and have the program print out the exact value of $b(n, p, k)$ and the Poisson approximation to this value.

16 Assume that, during each second, a Dartmouth switchboard receives one call with probability .01 and no calls with probability .99. Use the Poisson approximation to estimate the probability that the operator will miss at most one call if she takes a 5-minute coffee break.

17 The probability of a royal flush in a poker hand is $p = 1/649{,}740$. How large must n be to render the probability of having no royal flush in n hands smaller than $1/e$?

18 A baker blends 600 raisins and 400 chocolate chips into a dough mix and, from this, makes 500 cookies.

(a) Find the probability that a randomly picked cookie will have no raisins.

(b) Find the probability that a randomly picked cookie will have exactly two chocolate chips.

(c) Find the probability that a randomly chosen cookie will have at least two bits (raisins or chips) in it.

19 The probability that, in a bridge deal, one of the four hands has all hearts is approximately 6.3×10^{-12}. In a city with about 50,000 bridge players the resident probability expert is called on the average once a year (usually late at night) and told that the caller has just been dealt a hand of all hearts. Should she suspect that some of these callers are the victims of practical jokes?

20 An advertiser drops 10,000 leaflets on a city which has 2000 blocks. Assume that each leaflet has an equal chance of landing on each block. What is the probability that a particular block will receive no leaflets?

21 In a class of 80 students, the professor calls on 1 student chosen at random for a recitation in each class period. There are 32 class periods in a term.

(a) Write a formula for the exact probability that a given student is called upon j times during the term.

(b) Write a formula for the Poisson approximation for this probability. Using your formula estimate the probability that a given student is called upon more than twice.

22 Assume that we are making raisin cookies. We put a box of 600 raisins into our dough mix, mix up the dough, then make from the dough 500 cookies. We then ask for the probability that a randomly chosen cookie will have 0, 1, 2, ... raisins. Consider the cookies as trials in an experiment, and let X be the random variable which gives the number of raisins in a given cookie. Then we can regard the number of raisins in a cookie as the result of $n = 600$ independent trials with probability $p = 1/500$ for success on each trial. Since n is large and p is small, we can use the Poisson approximation with $\lambda = 600(1/500) = 1.2$. Determine the probability that a given cookie will have at least five raisins.

23 For a certain experiment, the Poisson distribution with parameter $\lambda = m$ has been assigned. Show that a most probable outcome for the experiment is the integer value k such that $m - 1 \leq k \leq m$. Under what conditions will there be two most probable values? *Hint*: Consider the ratio of successive probabilities.

24 When John Kemeny was chair of the Mathematics Department at Dartmouth College, he received an average of ten letters each day. On a certain weekday he received no mail and wondered if it was a holiday. To decide this he computed the probability that, in ten years, he would have at least 1 day without any mail. He assumed that the number of letters he received on a given day has a Poisson distribution. What probability did he find? *Hint*: Apply the Poisson distribution twice. First, to find the probability that, in 3000 days, he will have at least 1 day without mail, assuming each year has about 300 days on which mail is delivered.

25 Reese Prosser never puts money in a 10-cent parking meter in Hanover. He assumes that there is a probability of .05 that he will be caught. The first offense costs nothing, the second costs 2 dollars, and subsequent offenses cost 5 dollars each. Under his assumptions, how does the expected cost of parking 100 times without paying the meter compare with the cost of paying the meter each time?

Number of deaths	Number of corps with x deaths in a given year
0	144
1	91
2	32
3	11
4	2

Table 5.5: Mule kicks.

26 Feller[5] discusses the statistics of flying bomb hits in an area in the south of London during the Second World War. The area in question was divided into $24 \times 24 = 576$ small areas. The total number of hits was 537. There were 229 squares with 0 hits, 211 with 1 hit, 93 with 2 hits, 35 with 3 hits, 7 with 4 hits, and 1 with 5 or more. Assuming the hits were purely random, use the Poisson approximation to find the probability that a particular square would have exactly k hits. Compute the expected number of squares that would have 0, 1, 2, 3, 4, and 5 or more hits and compare this with the observed results.

27 Assume that the probability that there is a significant accident in a nuclear power plant during one year's time is .001. If a country has 100 nuclear plants, estimate the probability that there is at least one such accident during a given year.

28 An airline finds that 4 percent of the passengers that make reservations on a particular flight will not show up. Consequently, their policy is to sell 100 reserved seats on a plane that has only 98 seats. Find the probability that every person who shows up for the flight will find a seat available.

29 The king's coinmaster boxes his coins 500 to a box and puts 1 counterfeit coin in each box. The king is suspicious, but, instead of testing all the coins in 1 box, he tests 1 coin chosen at random out of each of 500 boxes. What is the probability that he finds at least one fake? What is it if the king tests 2 coins from each of 250 boxes?

30 (From Kemeny[6]) Show that, if you make 100 bets on the number 17 at roulette at Monte Carlo (see Example 6.13), you will have a probability greater than 1/2 of coming out ahead. What is your expected winning?

31 In one of the first studies of the Poisson distribution, von Bortkiewicz[7] considered the frequency of deaths from kicks in the Prussian army corps. From the study of 14 corps over a 20-year period, he obtained the data shown in Table 5.5. Fit a Poisson distribution to this data and see if you think that the Poisson distribution is appropriate.

[5]ibid., p. 161.
[6]Private communication.
[7]L. von Bortkiewicz, *Das Gesetz der Kleinen Zahlen* (Leipzig: Teubner, 1898), p. 24.

32 It is often assumed that the auto traffic that arrives at the intersection during a unit time period has a Poisson distribution with expected value m. Assume that the number of cars X that arrive at an intersection from the north in unit time has a Poisson distribution with parameter $\lambda = m$ and the number Y that arrive from the west in unit time has a Poisson distribution with parameter $\lambda = \bar{m}$. If X and Y are independent, show that the total number $X + Y$ that arrive at the intersection in unit time has a Poisson distribution with parameter $\lambda = m + \bar{m}$.

33 Cars coming along Magnolia Street come to a fork in the road and have to choose either Willow Street or Main Street to continue. Assume that the number of cars that arrive at the fork in unit time has a Poisson distribution with parameter $\lambda = 4$. A car arriving at the fork chooses Main Street with probability 3/4 and Willow Street with probability 1/4. Let X be the random variable which counts the number of cars that, in a given unit of time, pass by Joe's Barber Shop on Main Street. What is the distribution of X?

34 In the appeal of the *People v. Collins* case (see Exercise 4.1.28), the counsel for the defense argued as follows: Suppose, for example, there are 5,000,000 couples in the Los Angeles area and the probability that a randomly chosen couple fits the witnesses' description is 1/12,000,000. Then the probability that there are two such couples given that there is at least one is not at all small. Find this probability. (The California Supreme Court overturned the initial guilty verdict.)

35 A manufactured lot of brass turnbuckles has S items of which D are defective. A sample of s items is drawn without replacement. Let X be a random variable that gives the number of defective items in the sample. Let $p(d) = P(X = d)$.

(a) Show that
$$p(d) = \frac{\binom{D}{d}\binom{S-D}{s-d}}{\binom{S}{s}} .$$
Thus, X is hypergeometric.

(b) Prove the following identity, known as *Euler's formula*:
$$\sum_{d=0}^{\min(D,s)} \binom{D}{d}\binom{S-D}{s-d} = \binom{S}{s} .$$

36 A bin of 1000 turnbuckles has an unknown number D of defectives. A sample of 100 turnbuckles has 2 defectives. The *maximum likelihood estimate* for D is the number of defectives which gives the highest probability for obtaining the number of defectives observed in the sample. Guess this number D and then write a computer program to verify your guess.

37 There are an unknown number of moose on Isle Royale (a National Park in Lake Superior). To estimate the number of moose, 50 moose are captured and tagged. Six months later 200 moose are captured and it is found that 8 of these were tagged. Estimate the number of moose on Isle Royale from these data, and then verify your guess by computer program (see Exercise 36).

38 A manufactured lot of buggy whips has 20 items, of which 5 are defective. A random sample of 5 items is chosen to be inspected. Find the probability that the sample contains exactly one defective item

(a) if the sampling is done with replacement.

(b) if the sampling is done without replacement.

39 Suppose that N and k tend to ∞ in such a way that k/N remains fixed. Show that

$$h(N, k, n, x) \to b(n, k/N, x) \ .$$

40 A bridge deck has 52 cards with 13 cards in each of four suits: spades, hearts, diamonds, and clubs. A hand of 13 cards is dealt from a shuffled deck. Find the probability that the hand has

(a) a distribution of suits 4, 4, 3, 2 (for example, four spades, four hearts, three diamonds, two clubs).

(b) a distribution of suits 5, 3, 3, 2.

41 Write a computer algorithm that simulates a hypergeometric random variable with parameters N, k, and n.

42 You are presented with four different dice. The first one has two sides marked 0 and four sides marked 4. The second one has a 3 on every side. The third one has a 2 on four sides and a 6 on two sides, and the fourth one has a 1 on three sides and a 5 on three sides. You allow your friend to pick any of the four dice he wishes. Then you pick one of the remaining three and you each roll your die. The person with the largest number showing wins a dollar. Show that you can choose your die so that you have probability 2/3 of winning no matter which die your friend picks. (See Tenney and Foster.[8])

43 The students in a certain class were classified by hair color and eye color. The conventions used were: Brown and black hair were considered dark, and red and blonde hair were considered light; black and brown eyes were considered dark, and blue and green eyes were considered light. They collected the data shown in Table 5.6. Are these traits independent? (See Example 5.6.)

44 Suppose that in the hypergeometric distribution, we let N and k tend to ∞ in such a way that the ratio k/N approaches a real number p between 0 and 1. Show that the hypergeometric distribution tends to the binomial distribution with parameters n and p.

[8] R. L. Tenney and C. C. Foster, *Non-transitive Dominance*, Math. Mag. 49 (1976) no. 3, pgs. 115-120.

	Dark Eyes	Light Eyes	
Dark Hair	28	15	43
Light Hair	9	23	32
	37	38	75

Table 5.6: Observed data.

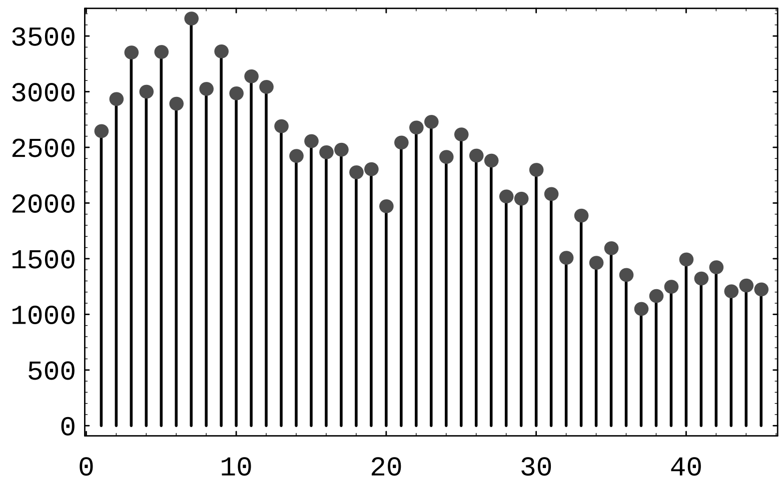

Figure 5.5: Distribution of choices in the Powerball lottery.

45 (a) Compute the leading digits of the first 100 powers of 2, and see how well these data fit the Benford distribution.

(b) Multiply each number in the data set of part (a) by 3, and compare the distribution of the leading digits with the Benford distribution.

46 In the Powerball lottery, contestants pick 5 different integers between 1 and 45, and in addition, pick a bonus integer from the same range (the bonus integer can equal one of the first five integers chosen). Some contestants choose the numbers themselves, and others let the computer choose the numbers. The data shown in Table 5.7 are the contestant-chosen numbers in a certain state on May 3, 1996. A spike graph of the data is shown in Figure 5.5.

The goal of this problem is to check the hypothesis that the chosen numbers are uniformly distributed. To do this, compute the value v of the random variable χ^2 given in Example 5.10. In the present case, this random variable has 44 degrees of freedom. One can find, in a χ^2 table, the value $v_0 = 59.43$, which represents a number with the property that a χ^2-distributed random variable takes on values that exceed v_0 only 5% of the time. Does your computed value of v exceed v_0? If so, you should reject the hypothesis that the contestants' choices are uniformly distributed.

Integer	Times Chosen	Integer	Times Chosen	Integer	Times Chosen
1	2646	2	2934	3	3352
4	3000	5	3357	6	2892
7	3657	8	3025	9	3362
10	2985	11	3138	12	3043
13	2690	14	2423	15	2556
16	2456	17	2479	18	2276
19	2304	20	1971	21	2543
22	2678	23	2729	24	2414
25	2616	26	2426	27	2381
28	2059	29	2039	30	2298
31	2081	32	1508	33	1887
34	1463	35	1594	36	1354
37	1049	38	1165	39	1248
40	1493	41	1322	42	1423
43	1207	44	1259	45	1224

Table 5.7: Numbers chosen by contestants in the Powerball lottery.

5.2 Important Densities

In this section, we will introduce some important probability density functions and give some examples of their use. We will also consider the question of how one simulates a given density using a computer.

Continuous Uniform Density

The simplest density function corresponds to the random variable U whose value represents the outcome of the experiment consisting of choosing a real number at random from the interval $[a, b]$.

$$f(\omega) = \begin{cases} 1/(b-a), & \text{if } a \le \omega \le b, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to simulate this density on a computer. We simply calculate the expression

$$(b-a)rnd + a \ .$$

Exponential and Gamma Densities

The exponential density function is defined by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } 0 \le x < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Here λ is any positive constant, depending on the experiment. The reader has seen this density in Example 2.17. In Figure 5.6 we show graphs of several exponential densities for different choices of λ. The exponential density is often used to

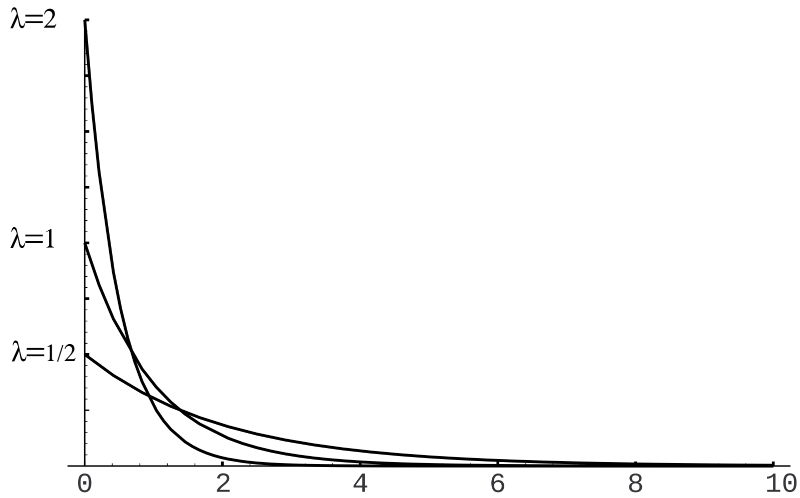

Figure 5.6: Exponential densities.

describe experiments involving a question of the form: How long until something happens? For example, the exponential density is often used to study the time between emissions of particles from a radioactive source.

The cumulative distribution function of the exponential density is easy to compute. Let T be an exponentially distributed random variable with parameter λ. If $x \geq 0$, then we have

$$\begin{aligned} F(x) &= P(T \leq x) \\ &= \int_0^x \lambda e^{-\lambda t}\, dt \\ &= 1 - e^{-\lambda x} \ . \end{aligned}$$

Both the exponential density and the geometric distribution share a property known as the "memoryless" property. This property was introduced in Example 5.1; it says that

$$P(T > r + s \,|\, T > r) = P(T > s) \ .$$

This can be demonstrated to hold for the exponential density by computing both sides of this equation. The right-hand side is just

$$1 - F(s) = e^{-\lambda s} \ ,$$

while the left-hand side is

$$\frac{P(T > r + s)}{P(T > r)} \quad = \quad \frac{1 - F(r + s)}{1 - F(s)}$$

$$\begin{aligned} &= \frac{e^{-\lambda(r+s)}}{e^{-\lambda r}} \\ &= e^{-\lambda s} \ . \end{aligned}$$

There is a very important relationship between the exponential density and the Poisson distribution. We begin by defining $X_1,\ X_2,\ \ldots$ to be a sequence of independent exponentially distributed random variables with parameter λ. We might think of X_i as denoting the amount of time between the ith and $(i+1)$st emissions of a particle by a radioactive source. (As we shall see in Chapter 6, we can think of the parameter λ as representing the reciprocal of the average length of time between emissions. This parameter is a quantity that might be measured in an actual experiment of this type.)

We now consider a time interval of length t, and we let Y denote the random variable which counts the number of emissions that occur in the time interval. We would like to calculate the distribution function of Y (clearly, Y is a discrete random variable). If we let S_n denote the sum $X_1 + X_2 + \cdots + X_n$, then it is easy to see that

$$P(Y = n) = P(S_n \le t \text{ and } S_{n+1} > t) \ .$$

Since the event $S_{n+1} \le t$ is a subset of the event $S_n \le t$, the above probability is seen to be equal to

$$P(S_n \le t) - P(S_{n+1} \le t) \ . \tag{5.4}$$

We will show in Chapter 7 that the density of S_n is given by the following formula:

$$g_n(x) = \begin{cases} \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This density is an example of a gamma density with parameters λ and n. The general gamma density allows n to be any positive real number. We shall not discuss this general density.

It is easy to show by induction on n that the cumulative distribution function of S_n is given by:

$$G_n(x) = \begin{cases} 1 - e^{-\lambda x}\left(1 + \frac{\lambda x}{1!} + \cdots + \frac{(\lambda x)^{n-1}}{(n-1)!}\right), & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Using this expression, the quantity in (5.4) is easy to compute; we obtain

$$e^{-\lambda t}\frac{(\lambda t)^n}{n!} \ ,$$

which the reader will recognize as the probability that a Poisson-distributed random variable, with parameter λt, takes on the value n.

The above relationship will allow us to simulate a Poisson distribution, once we have found a way to simulate an exponential density. The following random variable does the job:

$$Y = -\frac{1}{\lambda}\log(rnd) \ . \tag{5.5}$$

Using Corollary 5.2 (below), one can derive the above expression (see Exercise 3). We content ourselves for now with a short calculation that should convince the reader that the random variable Y has the required property. We have

$$\begin{aligned} P(Y \le y) &= P\Big(-\frac{1}{\lambda}\log(rnd) \le y\Big) \\ &= P(\log(rnd) \ge -\lambda y) \\ &= P(rnd \ge e^{-\lambda y}) \\ &= 1 - e^{-\lambda y}\ . \end{aligned}$$

This last expression is seen to be the cumulative distribution function of an exponentially distributed random variable with parameter λ.

To simulate a Poisson random variable W with parameter λ, we simply generate a sequence of values of an exponentially distributed random variable with the same parameter, and keep track of the subtotals S_k of these values. We stop generating the sequence when the subtotal first exceeds λ. Assume that we find that

$$S_n \le \lambda < S_{n+1}\ .$$

Then the value n is returned as a simulated value for W.

Example 5.7 (Queues) Suppose that customers arrive at random times at a service station with one server, and suppose that each customer is served immediately if no one is ahead of him, but must wait his turn in line otherwise. How long should each customer expect to wait? (We define the waiting time of a customer to be the length of time between the time that he arrives and the time that he begins to be served.)

Let us assume that the interarrival times between successive customers are given by random variables X_1, X_2, ..., X_n that are mutually independent and identically distributed with an exponential cumulative distribution function given by

$$F_X(t) = 1 - e^{-\lambda t}.$$

Let us assume, too, that the service times for successive customers are given by random variables $Y_1, Y_2, \ldots, Y_n$ that again are mutually independent and identically distributed with another exponential cumulative distribution function given by

$$F_Y(t) = 1 - e^{-\mu t}.$$

The parameters λ and μ represent, respectively, the reciprocals of the average time between arrivals of customers and the average service time of the customers. Thus, for example, the larger the value of λ, the smaller the average time between arrivals of customers. We can guess that the length of time a customer will spend in the queue depends on the relative sizes of the average interarrival time and the average service time.

It is easy to verify this conjecture by simulation. The program **Queue** simulates this queueing process. Let $N(t)$ be the number of customers in the queue at time t.

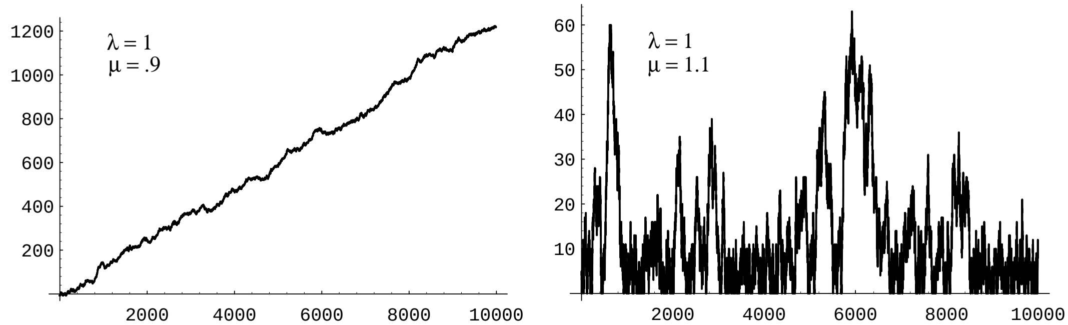

Figure 5.7: Queue sizes.

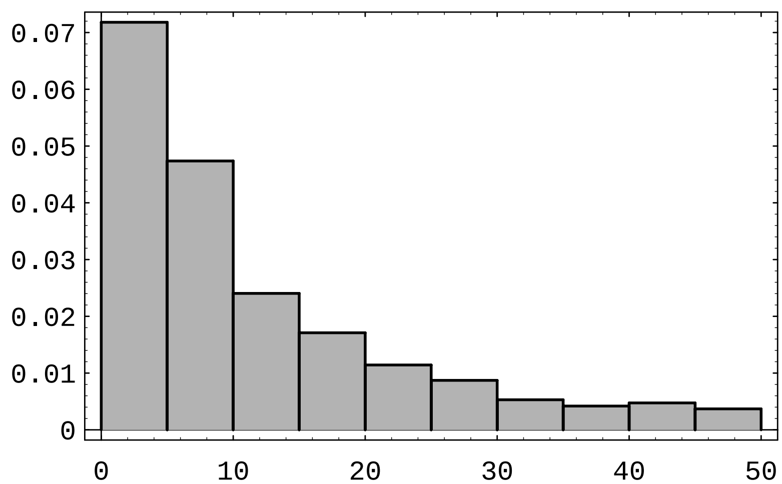

Figure 5.8: Waiting times.

Then we plot $N(t)$ as a function of t for different choices of the parameters λ and μ (see Figure 5.7).

We note that when $\lambda < \mu$, then $1/\lambda > 1/\mu$, so the average interarrival time is greater than the average service time, i.e., customers are served more quickly, on average, than new ones arrive. Thus, in this case, it is reasonable to expect that $N(t)$ remains small. However, if $\lambda > \mu$ then customers arrive more quickly than they are served, and, as expected, $N(t)$ appears to grow without limit.

We can now ask: How long will a customer have to wait in the queue for service? To examine this question, we let W_i be the length of time that the ith customer has to remain in the system (waiting in line and being served). Then we can present these data in a bar graph, using the program **Queue**, to give some idea of how the W_i are distributed (see Figure 5.8). (Here $\lambda = 1$ and $\mu = 1.1$.)

We see that these waiting times appear to be distributed exponentially. This is always the case when $\lambda < \mu$. The proof of this fact is too complicated to give here, but we can verify it by simulation for different choices of λ and μ, as above. □

Functions of a Random Variable

Before continuing our list of important densities, we pause to consider random variables which are functions of other random variables. We will prove a general theorem that will allow us to derive expressions such as Equation 5.5.

Theorem 5.1 Let X be a continuous random variable, and suppose that $\phi(x)$ is a strictly increasing function on the range of X. Define $Y = \phi(X)$. Suppose that X and Y have cumulative distribution functions F_X and F_Y respectively. Then these functions are related by

$$F_Y(y) = F_X(\phi^{-1}(y)).$$

If $\phi(x)$ is strictly decreasing on the range of X, then

$$F_Y(y) = 1 - F_X(\phi^{-1}(y)) \ .$$

Proof. Since ϕ is a strictly increasing function on the range of X, the events $(X \leq \phi^{-1}(y))$ and $(\phi(X) \leq y)$ are equal. Thus, we have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\phi(X) \leq y) \\ &= P(X \leq \phi^{-1}(y)) \\ &= F_X(\phi^{-1}(y)) \ . \end{aligned}$$

If $\phi(x)$ is strictly decreasing on the range of X, then we have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\phi(X) \leq y) \\ &= P(X \geq \phi^{-1}(y)) \\ &= 1 - P(X < \phi^{-1}(y)) \\ &= 1 - F_X(\phi^{-1}(y)) \ . \end{aligned}$$

This completes the proof. □

Corollary 5.1 Let X be a continuous random variable, and suppose that $\phi(x)$ is a strictly increasing function on the range of X. Define $Y = \phi(X)$. Suppose that the density functions of X and Y are f_X and f_Y, respectively. Then these functions are related by

$$f_Y(y) = f_X(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y) \ .$$

If $\phi(x)$ is strictly decreasing on the range of X, then

$$f_Y(y) = -f_X(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y) \ .$$

Proof. This result follows from Theorem 5.1 by using the Chain Rule. □

If the function ϕ is neither strictly increasing nor strictly decreasing, then the situation is somewhat more complicated but can be treated by the same methods. For example, suppose that $Y = X^2$, Then $\phi(x) = x^2$, and

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(-\sqrt{y} \le X \le +\sqrt{y}) \\ &= P(X \le +\sqrt{y}) - P(X \le -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y})\ . \end{aligned}$$

Moreover,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy}(F_X(\sqrt{y}) - F_X(-\sqrt{y})) \\ &= \Bigl(f_X(\sqrt{y}) + f_X(-\sqrt{y})\Bigr) \frac{1}{2\sqrt{y}}\ . \end{aligned}$$

We see that in order to express F_Y in terms of F_X when $Y = \phi(X)$, we have to express $P(Y \le y)$ in terms of $P(X \le x)$, and this process will depend in general upon the structure of ϕ.

Simulation

Theorem 5.1 tells us, among other things, how to simulate on the computer a random variable Y with a prescribed distribution function F. We assume that $F(y)$ is strictly increasing for those values of y where $0 < F(y) < 1$. For this purpose, let U be a random variable which is uniformly distributed on $[0, 1]$. Then U has distribution function $F_U(u) = u$. Now, if F is the prescribed distribution function for Y, then to write Y in terms of U we first solve the equation

$$F(y) = u$$

for y in terms of u. We obtain $y = F^{-1}(u)$. Note that since F is an increasing function this equation always has a unique solution (see Figure 5.9). Then we set $Z = F^{-1}(U)$ and obtain, by Theorem 5.1,

$$F_Z(y) = F_U(F(y)) = F(y)\ ,$$

since $F_U(u) = u$. Therefore, Z and Y have the same distribution function. Summarizing, we have the following.

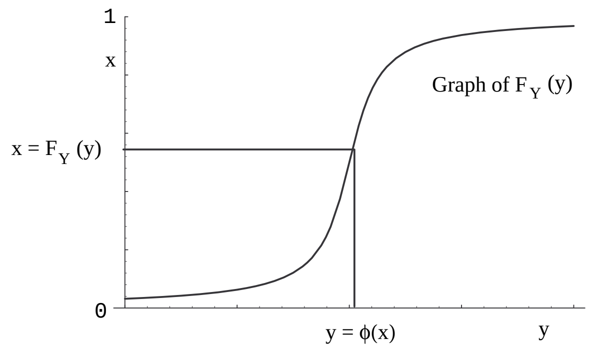

Figure 5.9: Converting a uniform distribution F_U into a prescribed distribution F_Y.

Corollary 5.2 If $F(y)$ is a given distribution function that is strictly increasing when $0 < F(y) < 1$ and if U is a random variable with uniform distribution on $[0, 1]$, then

$$Y = F^{-1}(U)$$

has the distribution $F(y)$. □

Thus, to simulate a random variable with a given distribution F we need only set $Y = F^{-1}(\text{rnd})$.

Normal Density

We now come to the most important density function, the normal density function. We have seen in Chapter 3 that the binomial distribution functions are bell-shaped, even for moderate size values of n. We recall that a binomially-distributed random variable with parameters n and p can be considered to be the sum of n mutually independent 0-1 random variables. A very important theorem in probability theory, called the Central Limit Theorem, states that under very general conditions, if we sum a large number of mutually independent random variables, then the distribution of the sum can be closely approximated by a certain specific continuous density, called the normal density. This theorem will be discussed in Chapter 9.

The normal density function with parameters μ and σ is defined as follows:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} .$$

The parameter μ represents the "center" of the density (and in Chapter 6, we will show that it is the average, or expected, value of the density). The parameter σ is a measure of the "spread" of the density, and thus it is assumed to be positive. (In Chapter 6, we will show that σ is the standard deviation of the density.) We note that it is not at all obvious that the above function is a density, i.e., that its

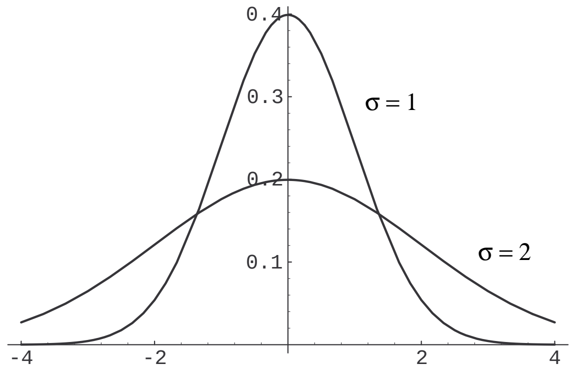

Figure 5.10: Normal density for two sets of parameter values.

integral over the real line equals 1. The cumulative distribution function is given by the formula

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(u-\mu)^2/2\sigma^2}\, du\ .$$

In Figure 5.10 we have included for comparison a plot of the normal density for the cases $\mu = 0$ and $\sigma = 1$, and $\mu = 0$ and $\sigma = 2$.

One cannot write F_X in terms of simple functions. This leads to several problems. First of all, values of F_X must be computed using numerical integration. Extensive tables exist containing values of this function (see Appendix A). Secondly, we cannot write F_X^{-1} in closed form, so we cannot use Corollary 5.2 to help us simulate a normal random variable. For this reason, special methods have been developed for simulating a normal distribution. One such method relies on the fact that if U and V are independent random variables with uniform densities on $[0, 1]$, then the random variables

$$X = \sqrt{-2\log U}\cos 2\pi V$$

and

$$Y = \sqrt{-2\log U}\sin 2\pi V$$

are independent, and have normal density functions with parameters $\mu = 0$ and $\sigma = 1$. (This is not obvious, nor shall we prove it here. See Box and Muller.[9])

Let Z be a normal random variable with parameters $\mu = 0$ and $\sigma = 1$. A normal random variable with these parameters is said to be a *standard* normal random variable. It is an important and useful fact that if we write

$$X = \sigma Z + \mu\ ,$$

then X is a normal random variable with parameters μ and σ. To show this, we will use Theorem 5.1. We have $\phi(z) = \sigma z + \mu$, $\phi^{-1}(x) = (x - \mu)/\sigma$, and

$$F_X(x) \quad = \quad F_Z\left(\frac{x-\mu}{\sigma}\right),$$

[9] G. E. P. Box and M. E. Muller, *A Note on the Generation of Random Normal Deviates*, Ann. of Math. Stat. 29 (1958), pgs. 610-611.

$$\begin{aligned} f_X(x) &= f_Z\left(\frac{x-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \ . \end{aligned}$$

The reader will note that this last expression is the density function with parameters μ and σ, as claimed.

We have seen above that it is possible to simulate a standard normal random variable Z. If we wish to simulate a normal random variable X with parameters μ and σ, then we need only transform the simulated values for Z using the equation $X = \sigma Z + \mu$.

Suppose that we wish to calculate the value of a distribution function for the normal random variable X, with parameters μ and σ. We can reduce this calculation to one concerning the standard normal random variable Z as follows:

$$\begin{aligned} F_X(x) &= P(X \le x) \\ &= P\left(Z \le \frac{x-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{x-\mu}{\sigma}\right) \ . \end{aligned}$$

This last expression can be found in a table of values of the distribution function for a standard normal random variable. Thus, we see that it is unnecessary to make tables of normal distribution functions with arbitrary μ and σ.

The process of changing a normal random variable to a standard normal random variable is known as standardization. If X has a normal distribution with parameters μ and σ and if

$$Z = \frac{X-\mu}{\sigma} \ ,$$

then Z is said to be the standardized version of X.

The following example shows how we use the standardized version of a normal random variable X to compute specific probabilities relating to X.

Example 5.8 Suppose that X is a normally distributed random variable with parameters $\mu = 10$ and $\sigma = 3$. Find the probability that X is between 4 and 16.

To solve this problem, we note that $Z = (X - 10)/3$ is the standardized version of X. So, we have

$$\begin{aligned} P(4 \le X \le 16) &= P(X \le 16) - P(X \le 4) \\ &= F_X(16) - F_X(4) \\ &= F_Z\left(\frac{16-10}{3}\right) - F_Z\left(\frac{4-10}{3}\right) \\ &= F_Z(2) - F_Z(-2) \ . \end{aligned}$$

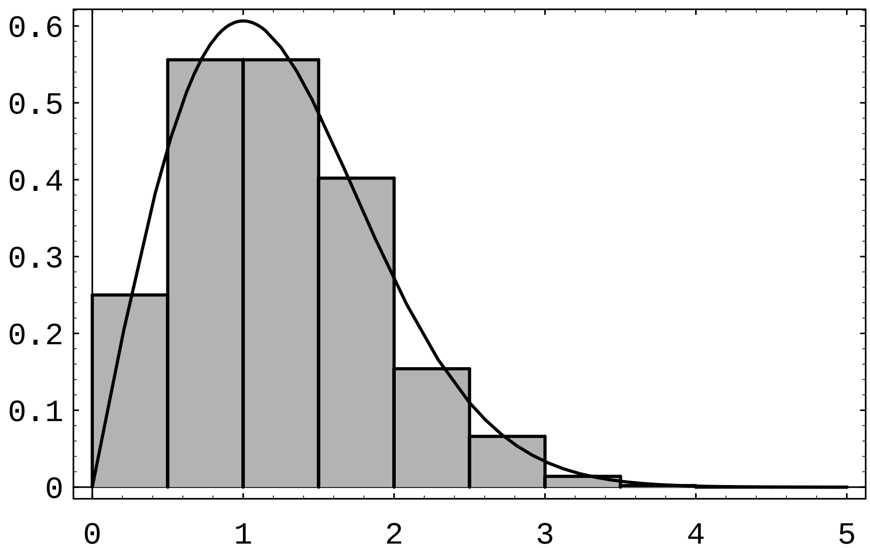

Figure 5.11: Distribution of dart distances in 1000 drops.

This last expression can be evaluated by using tabulated values of the standard normal distribution function (see 12.3); when we use this table, we find that $F_Z(2) = .9772$ and $F_Z(-2) = .0228$. Thus, the answer is .9544.

In Chapter 6, we will see that the parameter μ is the mean, or average value, of the random variable X. The parameter σ is a measure of the spread of the random variable, and is called the standard deviation. Thus, the question asked in this example is of a typical type, namely, what is the probability that a random variable has a value within two standard deviations of its average value. □

Maxwell Density

Example 5.9 Suppose that we drop a dart on a large table top, which we consider as the xy-plane, and suppose that the x and y coordinates of the dart point are independent and have a normal distribution with parameters $\mu = 0$ and $\sigma = 1$. What is the distribution of the distance of the point from the origin?

This problem arises in physics when it is assumed that a moving particle in R^n has components of the velocity that are mutually independent and normally distributed and it is desired to find the distribution of the speed of the particle. Maxwell calculated the density for the speed in the case that $n = 3$.

We can simulate this experiment for the case $n = 2$ by picking independently a pair of coordinates (x, y), each from a normal distribution with $\mu = 0$ and $\sigma = 1$ on $(-\infty, \infty)$, calculating the distance $r = \sqrt{x^2 + y^2}$ of the point (x, y) from the origin, repeating this process a large number of times, and then presenting the results in a bar graph. The results are shown in Figure 5.11.

	Female	Male	
A	37	56	93
B	63	60	123
C	47	43	90
Below C	5	8	13
	152	167	319

Table 5.8: Calculus class data.

We have also plotted the theoretical density

$$f(r) = re^{-r^2/2} .$$

This will be derived in Chapter 7; see Example 7.7. This density is called the (two-dimensional) Maxwell density. □

Chi-Squared Density

We return to the problem of independence of traits discussed in Example 5.6. It is frequently the case that we have two traits, each of which have several different values. As was seen in the example, quite a lot of calculation was needed even in the case of two values for each trait. We now give another method for testing independence of traits, which involves much less calculation.

Example 5.10 Suppose that we have the data shown in Table 5.8 concerning grades and gender of students in a Calculus class. We can use the same sort of model in this situation as was used in Example 5.6. We imagine that we have an urn with 319 balls of two colors, say blue and red, corresponding to females and males, respectively. We now draw 93 balls, without replacement, from the urn. These balls correspond to the grade of A. We continue by drawing 123 balls, which correspond to the grade of B. When we finish, we have four sets of balls, with each ball belonging to exactly one set. (We could have stipulated that the balls were of four colors, corresponding to the four possible grades. In this case, we would draw a subset of size 152, which would correspond to the females. The balls remaining in the urn would correspond to the males. The choice does not affect the final determination of whether we should reject the hypothesis of independence of traits.)

The expected data set can be determined in exactly the same way as in Example 5.6. If we do this, we obtain the expected values shown in Table 5.9. Even if the traits are independent, we would still expect to see some differences between the numbers in corresponding boxes in the two tables. However, if the differences are large, then we might suspect that the two traits are not independent. In Example 5.6, we used the probability distribution of the various possible data sets to compute the probability of finding a data set that differs from the expected data set by at least as much as the actual data set does. We could do the same in this case, but the amount of computation is enormous.

	Female	Male	
A	44.3	48.7	93
B	58.6	64.4	123
C	42.9	47.1	90
Below C	66.2	6.8	13
	152	167	319

Table 5.9: Expected data.

Instead, we will describe a single number which does a good job of measuring how far a given data set is from the expected one. To quantify how far apart the two sets of numbers are, we could sum the squares of the differences of the corresponding numbers. (We could also sum the absolute values of the differences, but we would not want to sum the differences.) Suppose that we have data in which we expect to see 10 objects of a certain type, but instead we see 18, while in another case we expect to see 50 objects of a certain type, but instead we see 58. Even though the two differences are about the same, the first difference is more surprising than the second, since the expected number of outcomes in the second case is quite a bit larger than the expected number in the first case. One way to correct for this is to divide the individual squares of the differences by the expected number for that box. Thus, if we label the values in the eight boxes in the first table by O_i (for observed values) and the values in the eight boxes in the second table by E_i (for expected values), then the following expression might be a reasonable one to use to measure how far the observed data is from what is expected:

$$\sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} \ .$$

This expression is a random variable, which is usually denoted by the symbol χ^2, pronounced "ki-squared." It is called this because, under the assumption of independence of the two traits, the density of this random variable can be computed and is approximately equal to a density called the chi-squared density. We choose not to give the explicit expression for this density, since it involves the gamma function, which we have not discussed. The chi-squared density is, in fact, a special case of the general gamma density.

In applying the chi-squared density, tables of values of this density are used, as in the case of the normal density. The chi-squared density has one parameter n, which is called the number of degrees of freedom. The number n is usually easy to determine from the problem at hand. For example, if we are checking two traits for independence, and the two traits have a and b values, respectively, then the number of degrees of freedom of the random variable χ^2 is $(a-1)(b-1)$. So, in the example at hand, the number of degrees of freedom is 3.

We recall that in this example, we are trying to test for independence of the two traits of gender and grades. If we assume these traits are independent, then the ball-and-urn model given above gives us a way to simulate the experiment. Using a computer, we have performed 1000 experiments, and for each one, we have

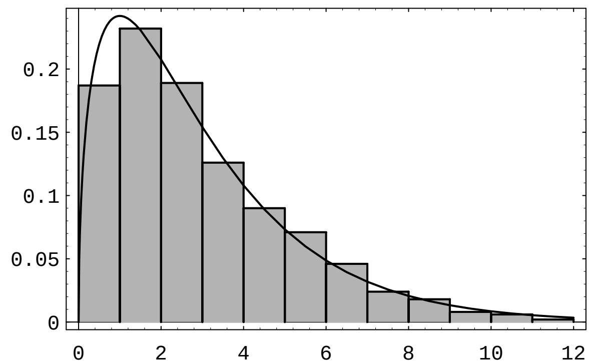

Figure 5.12: Chi-squared density with three degrees of freedom.

calculated a value of the random variable χ^2. The results are shown in Figure 5.12, together with the chi-squared density function with three degrees of freedom.

As we stated above, if the value of the random variable χ^2 is large, then we would tend not to believe that the two traits are independent. But how large is large? The actual value of this random variable for the data above is 4.13. In Figure 5.12, we have shown the chi-squared density with 3 degrees of freedom. It can be seen that the value 4.13 is larger than most of the values taken on by this random variable.

Typically, a statistician will compute the value v of the random variable χ^2, just as we have done. Then, by looking in a table of values of the chi-squared density, a value v_0 is determined which is only exceeded 5% of the time. If $v \geq v_0$, the statistician rejects the hypothesis that the two traits are independent. In the present case, $v_0 = 7.815$, so we would not reject the hypothesis that the two traits are independent. □

Cauchy Density

The following example is from Feller.[10]

Example 5.11 Suppose that a mirror is mounted on a vertical axis, and is free to revolve about that axis. The axis of the mirror is 1 foot from a straight wall of infinite length. A pulse of light is shown onto the mirror, and the reflected ray hits the wall. Let ϕ be the angle between the reflected ray and the line that is perpendicular to the wall and that runs through the axis of the mirror. We assume that ϕ is uniformly distributed between $-\pi/2$ and $\pi/2$. Let X represent the distance between the point on the wall that is hit by the reflected ray and the point on the wall that is closest to the axis of the mirror. We now determine the density of X.

[10]W. Feller, *An Introduction to Probability Theory and Its Applications,*, vol. 2, (New York: Wiley, 1966)

Let B be a fixed positive quantity. Then $X \geq B$ if and only if $\tan(\phi) \geq B$, which happens if and only if $\phi \geq \arctan(B)$. This happens with probability

$$\frac{\pi/2 - \arctan(B)}{\pi} \ .$$

Thus, for positive B, the distribution function of X is

$$F(B) = 1 - \frac{\pi/2 - \arctan(B)}{\pi} \ .$$

Therefore, the density function for positive B is

$$f(B) = \frac{1}{\pi(1 + B^2)} \ .$$

Since the physical situation is symmetric with respect to $\phi = 0$, it is easy to see that the above expression for the density is correct for negative values of B as well.

The Law of Large Numbers, which we will discuss in Chapter 8, states that in many cases, if we take the average of independent values of a random variable, then the average approaches a specific number as the number of values increases. It turns out that if one does this with a Cauchy-distributed random variable, the average does not approach any specific number. □

Exercises

1 Choose a number U from the unit interval $[0, 1]$ with uniform distribution. Find the distribution and density for the random variables

(a) $Y = U + 2$.

(b) $Y = U^3$.

2 Choose a number U from the interval $[0, 1]$ with uniform distribution. Find the distribution and density for the random variables

(a) $Y = 1/(U + 1)$.

(b) $Y = \log(U + 1)$.

3 Use Corollary 5.2 to derive the expression for the random variable given in Equation 5.5. *Hint*: The random variables $1 - rnd$ and rnd are identically distributed.

4 Suppose we know a random variable Y as a function of the uniform random variable U: $Y = \phi(U)$, and suppose we have calculated the distribution function $F_Y(y)$ and thence the density $f_Y(y)$. How can we check whether our answer is correct? An easy simulation provides the answer: Make a bar graph of $Y = \phi(rnd)$ and compare the result with the graph of $f_Y(y)$. These graphs should look similar. Check your answers to Exercises 1 and 2 by this method.

5 Choose a number U from the interval $[0, 1]$ with uniform distribution. Find the distribution and density for the random variables

(a) $Y = |U - 1/2|$.

(b) $Y = (U - 1/2)^2$.

6 Check your results for Exercise 5 by simulation as described in Exercise 4.

7 Explain how you can generate a random variable whose distribution function is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ x^2, & \text{if } 0 \le x \le 1, \\ 1, & \text{if } x > 1. \end{cases}$$

8 Write a program to generate a sample of 1000 random outcomes each of which is chosen from the distribution given in Exercise 7. Plot a bar graph of your results and compare this empirical density with the density for the distribution given in Exercise 7.

9 Let U, V be random numbers chosen independently from the interval $[0, 1]$ with uniform distribution. Find the distribution and density of each of the variables

(a) $Y = U + V$.

(b) $Y = |U - V|$.

10 Let U, V be random numbers chosen independently from the interval $[0, 1]$. Find the distribution and density for the random variables

(a) $Y = \max(U, V)$.

(b) $Y = \min(U, V)$.

11 Write a program to simulate the random variables of Exercises 9 and 10 and plot a bar graph of the results. Compare the resulting empirical density with the density found in Exercises 9 and 10.

12 A number U is chosen at random in the interval $[0, 1]$. Find the probability that

(a) $R = U^2 < 1/4$.

(b) $S = U(1 - U) < 1/4$.

(c) $T = U/(1 - U) < 1/4$.

13 Find the distribution function F and the density function f for each of the random variables R, S, and T in Exercise 12.

14 A point P in the unit square has coordinates X and Y chosen at random in the interval $[0, 1]$. Let D be the distance from P to the nearest edge of the square, and E the distance to the nearest corner. What is the probability that

(a) $D < 1/4$?

(b) $E < 1/4$?

15 In Exercise 14 find the distribution F and density f for the random variable D.

16 Let X be a random variable with density function

$$f_X(x) = \begin{cases} cx(1-x), & \text{if } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) What is the value of c?

(b) What is the distribution function F_X for X?

(c) What is the probability that $X < 1/4$?

17 Let X be a random variable with distribution function

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \sin^2(\pi x/2), & \text{if } 0 \le x \le 1, \\ 1, & \text{if } 1 < x. \end{cases}$$

(a) What is the density function f_X for X?

(b) What is the probability that $X < 1/4$?

18 Let X be a random variable with distribution function F_X, and let $Y = X + b$, $Z = aX$, and $W = aX + b$, where a and b are any constants. Find the distribution functions F_Y, F_Z, and F_W. *Hint*: The cases $a > 0$, $a = 0$, and $a < 0$ require different arguments.

19 Let X be a random variable with density function f_X, and let $Y = X + b$, $Z = aX$, and $W = aX + b$. Find the density functions f_Y, f_Z, and f_W. (See Exercise 18.)

20 Let X be a random variable uniformly distributed over $[c, d]$, and let $Y = aX + b$. For what choice of a and b is Y uniformly distributed over $[0, 1]$?

21 Let X be a random variable with distribution function F strictly increasing on the range of X. Let $Y = F(X)$. Show that Y is uniformly distributed in the interval $[0, 1]$. (The formula $X = F^{-1}(Y)$ then tells us how to construct X from a uniform random variable Y.)

22 Let X be a random variable with distribution function F. The *median* of X is the value m for which $F(m) = 1/2$. Then $X < m$ with probability 1/2 and $X > m$ with probability 1/2. Find m if X is

(a) uniformly distributed over the interval $[a, b]$.

(b) normally distributed with parameters μ and σ.

(c) exponentially distributed with parameter λ.

Test Score	Letter grade
$\mu + \sigma < x$	A
$\mu < x < \mu + \sigma$	B
$\mu - \sigma < x < \mu$	C
$\mu - 2\sigma < x < \mu - \sigma$	D
$x < \mu - 2\sigma$	F

Table 5.10: Grading on the curve.

23 Let X be a random variable with density function f_X. The *mean* of X is the value $\mu = \int x f_x(x)\,dx$. Then μ gives an average value for X (see Section 6.3). Find μ if X is distributed uniformly, normally, or exponentially, as in Exercise 22.

24 Let X be a random variable with density function f_X. The *mode* of X is the value M for which $f(M)$ is maximum. Then values of X near M are most likely to occur. Find M if X is distributed normally or exponentially, as in Exercise 22. What happens if X is distributed uniformly?

25 Let X be a random variable normally distributed with parameters $\mu = 70$, $\sigma = 10$. Estimate

(a) $P(X > 50)$.

(b) $P(X < 60)$.

(c) $P(X > 90)$.

(d) $P(60 < X < 80)$.

26 Bridies' Bearing Works manufactures bearing shafts whose diameters are normally distributed with parameters $\mu = 1$, $\sigma = .002$. The buyer's specifications require these diameters to be $1.000 \pm .003$ cm. What fraction of the manufacturer's shafts are likely to be rejected? If the manufacturer improves her quality control, she can reduce the value of σ. What value of σ will ensure that no more than 1 percent of her shafts are likely to be rejected?

27 A final examination at Podunk University is constructed so that the test scores are approximately normally distributed, with parameters μ and σ. The instructor assigns letter grades to the test scores as shown in Table 5.10 (this is the process of "grading on the curve").

What fraction of the class gets A, B, C, D, F?

28 (Ross[11]) An expert witness in a paternity suit testifies that the length (in days) of a pregnancy, from conception to delivery, is approximately normally distributed, with parameters $\mu = 270$, $\sigma = 10$. The defendant in the suit is able to prove that he was out of the country during the period from 290 to 240 days before the birth of the child. What is the probability that the defendant was in the country when the child was conceived?

[11] S. Ross, *A First Course in Probability Theory,* 2d ed. (New York: Macmillan, 1984).

29 Suppose that the time (in hours) required to repair a car is an exponentially distributed random variable with parameter $\lambda = 1/2$. What is the probability that the repair time exceeds 4 hours? If it exceeds 4 hours what is the probability that it exceeds 8 hours?

30 Suppose that the number of years a car will run is exponentially distributed with parameter $\mu = 1/4$. If Prosser buys a used car today, what is the probability that it will still run after 4 years?

31 Let U be a uniformly distributed random variable on $[0, 1]$. What is the probability that the equation

$$x^2 + 4Ux + 1 = 0$$

has two distinct real roots x_1 and x_2?

32 Write a program to simulate the random variables whose densities are given by the following, making a suitable bar graph of each and comparing the exact density with the bar graph.

(a) $f_X(x) = e^{-x}$ on $[0, \infty)$ (but just do it on $[0, 10]$).

(b) $f_X(x) = 2x$ on $[0, 1]$.

(c) $f_X(x) = 3x^2$ on $[0, 1]$.

(d) $f_X(x) = 4|x - 1/2|$ on $[0, 1]$.

33 Suppose we are observing a process such that the time between occurrences is exponentially distributed with $\lambda = 1/30$ (i.e., the average time between occurrences is 30 minutes). Suppose that the process starts at a certain time and we start observing the process 3 hours later. Write a program to simulate this process. Let T denote the length of time that we have to wait, after we start our observation, for an occurrence. Have your program keep track of T. What is an estimate for the average value of T?

34 Jones puts in two new lightbulbs: a 60 watt bulb and a 100 watt bulb. It is claimed that the lifetime of the 60 watt bulb has an exponential density with average lifetime 200 hours ($\lambda = 1/200$). The 100 watt bulb also has an exponential density but with average lifetime of only 100 hours ($\lambda = 1/100$). Jones wonders what is the probability that the 100 watt bulb will outlast the 60 watt bulb.

If X and Y are two independent random variables with exponential densities $f(x) = \lambda e^{-\lambda x}$ and $g(x) = \mu e^{-\mu x}$, respectively, then the probability that X is less than Y is given by

$$P(X < Y) = \int_0^\infty f(x)(1 - G(x))\,dx,$$

where $G(x)$ is the distribution function for $g(x)$. Explain why this is the case. Use this to show that

$$P(X < Y) = \frac{\lambda}{\lambda + \mu}$$

and to answer Jones's question.

35 Consider the simple queueing process of Example 5.7. Suppose that you watch the size of the queue. If there are j people in the queue the next time the queue size changes it will either decrease to $j-1$ or increase to $j+1$. Use the result of Exercise 34 to show that the probability that the queue size decreases to $j-1$ is $\mu/(\mu+\lambda)$ and the probability that it increases to $j+1$ is $\lambda/(\mu+\lambda)$. When the queue size is 0 it can only increase to 1. Write a program to simulate the queue size. Use this simulation to help formulate a conjecture containing conditions on μ and λ that will ensure that the queue will have times when it is empty.

36 Let X be a random variable having an exponential density with parameter λ. Find the density for the random variable $Y = rX$, where r is a positive real number.

37 Let X be a random variable having a normal density and consider the random variable $Y = e^X$. Then Y has a *log normal* density. Find this density of Y.

38 Let X_1 and X_2 be independent random variables and for $i = 1, 2$, let $Y_i = \phi_i(X_i)$, where ϕ_i is strictly increasing on the range of X_i. Show that Y_1 and Y_2 are independent. Note that the same result is true without the assumption that the ϕ_i's are strictly increasing, but the proof is more difficult.