Conditional Probability and Conditional Expectation

3.1. Introduction

One of the most useful concepts in probability theory is that of conditional probability and conditional expectation. The reason is twofold. First, in practice, we are often interested in calculating probabilities and expectations when some partial information is available; hence, the desired probabilities and expectations are conditional ones. Secondly, in calculating a desired probability or expectation it is often extremely useful to first "condition" on some appropriate random variable.

3.2. The Discrete Case

Recall that for any two events E and F, the conditional probability of E given F is defined, as long as $P(F) > 0$, by

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Hence, if X and Y are discrete random variables, then it is natural to define the *conditional probability mass function* of X given that $Y = y$, by

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x | Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

for all values of y such that $P\{Y = y\} > 0$. Similarly, the conditional probability distribution function of X given that $Y = y$ is defined, for all y such that $P\{Y = y\} > 0$, by

$$F_{X|Y}(x|y) = P\{X \leqslant x|Y = y\}$$
$$= \sum_{a \leqslant x} p_{X|Y}(a|y)$$

Finally, the conditional expectation of X given that $Y = y$ is defined by

$$E[X|Y = y] = \sum_x x P\{X = x|Y = y\}$$
$$= \sum_x x p_{X|Y}(x|y)$$

In other words, the definitions are exactly as before with the exception that everything is now conditional on the event that $Y = y$. If X is independent of Y, then the conditional mass function, distribution, and expectation are the same as the unconditional ones. This follows, since if X is independent of Y, then

$$p_{X|Y}(x|y) = P\{X = x|Y = y\}$$
$$= P\{X = x\}$$

Example 3.1 Suppose that $p(x, y)$, the joint probability mass function of X and Y, is given by

$$p(1, 1) = 0.5, \qquad p(1, 2) = 0.1, \qquad p(2, 1) = 0.1, \qquad p(2, 2) = 0.3$$

Calculate the conditional probability mass function of X given that $Y = 1$.

Solution: We first note that

$$p_Y(1) = \sum_x p(x, 1) = p(1, 1) + p(2, 1) = 0.6$$

Hence,

$$p_{X|Y}(1|1) = P\{X = 1|Y = 1\}$$
$$= \frac{P\{X = 1, Y = 1\}}{P\{Y = 1\}}$$
$$= \frac{p(1, 1)}{p_Y(1)}$$
$$= \frac{5}{6}$$

Similarly,

$$p_{X|Y}(2|1) = \frac{p(2,1)}{p_Y(1)} = \frac{1}{6} \quad \blacksquare$$

Example 3.2 If X_1 and X_2 are independent binomial random variables with respective parameters (n_1, p) and (n_2, p), calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$.

Solution: With $q = 1 - p$,

$$\begin{aligned}
P\{X_1 = k | X_1 + X_2 = m\} &= \frac{P\{X_1 = k, X_1 + X_2 = m\}}{P\{X_1 + X_2 = m\}} \\
&= \frac{P\{X_1 = k, X_2 = m - k\}}{P\{X_1 + X_2 = m\}} \\
&= \frac{P\{X_1 = k\}P\{X_2 = m - k\}}{P\{X_1 + X_2 = m\}} \\
&= \frac{\binom{n_1}{k} p^k q^{n_1 - k} \binom{n_2}{m-k} p^{m-k} q^{n_2 - m + k}}{\binom{n_1 + n_2}{m} p^m q^{n_1 + n_2 - m}}
\end{aligned}$$

where we have used that $X_1 + X_2$ is a binomial random variable with parameters $(n_1 + n_2, p)$ (see Example 2.44). Thus, the conditional probability mass function of X_1, given that $X_1 + X_2 = m$, is

$$P\{X_1 = k | X_1 + X_2 = m\} = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1 + n_2}{m}} \tag{3.1}$$

The distribution given by Equation (3.1), first seen in Example 2.34, is known as the *hypergeometric* distribution. It is the distribution of the number of blue balls that are chosen when a sample of m balls is randomly chosen from an urn that contains n_1 blue and n_2 red balls. (To intuitively see why the conditional distribution is hypergeometric, consider $n_1 + n_2$ independent trials that each result in a success with probability p; let X_1 represent the number of successes in the first n_1 trials and let X_2 represent the number of successes in the final n_2 trials. Because all trials have the same probability of being a success, each of the $\binom{n_1+n_2}{m}$ subsets of m trials is equally likely to be the success trials; thus, the number of the m success trials that are among the first n_1 trials is a hypergeometric random variable.) $\blacksquare$

Example 3.3 If X and Y are independent Poisson random variables with respective means λ_1 and λ_2, calculate the conditional expected value of X given that $X+Y=n$.

Solution: Let us first calculate the conditional probability mass function of X given that $X+Y=n$. We obtain

$$\begin{aligned}
P\{X=k|X+Y=n\} &= \frac{P\{X=k, X+Y=n\}}{P\{X+Y=n\}} \\
&= \frac{P\{X=k, Y=n-k\}}{P\{X+Y=n\}} \\
&= \frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}
\end{aligned}$$

where the last equality follows from the assumed independence of X and Y. Recalling (see Example 2.36) that $X+Y$ has a Poisson distribution with mean $\lambda_1+\lambda_2$, the preceding equation equals

$$\begin{aligned}
P\{X=k|X+Y=n\} &= \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1} \\
&= \frac{n!}{(n-k)!k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}
\end{aligned}$$

In other words, the conditional distribution of X given that $X+Y=n$, is the binomial distribution with parameters n and $\lambda_1/(\lambda_1+\lambda_2)$. Hence,

$$E\{X|X+Y=n\} = n\frac{\lambda_1}{\lambda_1+\lambda_2} \qquad \blacksquare$$

Example 3.4 Consider an experiment which results in one of three possible outcomes with outcome i occurring with probability $p_i, i=1,2,3, \sum_{i=1}^3 p_i = 1$. Suppose that n independent replications of this experiment are performed and let $X_i,\ i=1,2,3$, denote the number of times outcome i appears. Determine the conditional expectation of X_1 given that $X_2=m$.

Solution: For $k \leqslant n - m$,

$$P\{X_1 = k | X_2 = m\} = \frac{P\{X_1 = k, X_2 = m\}}{P\{X_2 = m\}}$$

Now if $X_1 = k$ and $X_2 = m$, then it follows that $X_3 = n - k - m$. However,

$$\begin{aligned} &P\{X_1 = k,\ X_2 = m,\ X_3 = n - k - m\} \\ &\quad = \frac{n!}{k!m!(n-k-m)!} p_1^k p_2^m p_3^{(n-k-m)} \end{aligned} \tag{3.2}$$

This follows since any particular sequence of the n experiments having outcome 1 appearing k times, outcome 2 m times, and outcome 3 $(n - k - m)$ times has probability $p_1^k p_2^m p_3^{(n-k-m)}$ of occurring. Since there are $n!/[k!m!(n-k-m)!]$ such sequences, Equation (3.2) follows.

Therefore, we have

$$P\{X_1 = k | X_2 = m\} = \frac{\dfrac{n!}{k!m!(n-k-m)!} p_1^k p_2^m p_3^{(n-k-m)}}{\dfrac{n!}{m!(n-m)!} p_2^m (1-p_2)^{n-m}}$$

where we have used the fact that X_2 has a binomial distribution with parameters n and p_2. Hence,

$$P\{X_1 = k | X_2 = m\} = \frac{(n-m)!}{k!(n-m-k)!} \left(\frac{p_1}{1-p_2}\right)^k \left(\frac{p_3}{1-p_2}\right)^{n-m-k}$$

or equivalently, writing $p_3 = 1 - p_1 - p_2$,

$$P\{X_1 = k | X_2 = m\} = \binom{n-m}{k} \left(\frac{p_1}{1-p_2}\right)^k \left(1 - \frac{p_1}{1-p_2}\right)^{n-m-k}$$

In other words, the conditional distribution of X_1, given that $X_2 = m$, is binomial with parameters $n - m$ and $p_1/(1-p_2)$. Consequently,

$$E[X_1 | X_2 = m] = (n-m)\frac{p_1}{1-p_2} \qquad \blacksquare$$

Remarks (i) The desired conditional probability in Example 3.4 could also have been computed in the following manner. Consider the $n - m$ experiments that did not result in outcome 2. For each of these experiments, the probability that outcome 1 was obtained is given by

$$P\{\text{outcome } 1|\text{not outcome } 2\} = \frac{P\{\text{outcome } 1, \text{not outcome } 2\}}{P\{\text{not outcome } 2\}}$$
$$= \frac{p_1}{1 - p_2}$$

It therefore follows that, given $X_2 = m$, the number of times outcome 1 occurs is binomially distributed with parameters $n - m$ and $p_1/(1 - p_2)$.

(ii) Conditional expectations possess all of the properties of ordinary expectations. For instance, such identities as

$$E\left[\sum_{i=1}^{n} X_i | Y = y\right] = \sum_{i=1}^{n} E[X_i | Y = y]$$

remain valid.

Example 3.5 There are n components. On a rainy day, component i will function with probability p_i; on a nonrainy day, component i will function with probability q_i, for $i = 1, \ldots, n$. It will rain tomorrow with probability α. Calculate the conditional expected number of components that function tomorrow, given that it rains.

Solution: Let

$$X_i = \begin{cases} 1, & \text{if component } i \text{ functions tomorrow} \\ 0, & \text{otherwise} \end{cases}$$

Then, with Y defined to equal 1 if it rains tomorrow, and 0 otherwise, the desired conditional expectation is obtained as follows.

$$E\left[\sum_{t=1}^{n} X_i | Y = 1\right] = \sum_{i=1}^{n} E[X_i | Y = 1]$$
$$= \sum_{i=1}^{n} p_i \quad \blacksquare$$

3.3. The Continuous Case

If X and Y have a joint probability density function $f(x, y)$, then the *conditional probability density function* of X, given that $Y = y$, is defined for all values of y

such that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

To motivate this definition, multiply the left side by dx and the right side by $(dx\, dy)/dy$ to get

$$\begin{aligned}
f_{X|Y}(x|y)\, dx &= \frac{f(x, y)\, dx\, dy}{f_Y(y)\, dy} \\
&\approx \frac{P\{x \leqslant X \leqslant x + dx,\ y \leqslant Y \leqslant y + dy\}}{P\{y \leqslant Y \leqslant y + dy\}} \\
&= P\{x \leqslant X \leqslant x + dx | y \leqslant Y \leqslant y + dy\}
\end{aligned}$$

In other words, for small values dx and dy, $f_{X|Y}(x|y)\, dx$ is approximately the conditional probability that X is between x and $x + dx$ given that Y is between y and $y + dy$.

The *conditional expectation* of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

Example 3.6 Suppose the joint density of X and Y is given by

$$f(x, y) = \begin{cases} 6xy(2 - x - y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Compute the conditional expectation of X given that $Y = y$, where $0 < y < 1$.

Solution: We first compute the conditional density

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\
&= \frac{6xy(2 - x - y)}{\int_0^1 6xy(2 - x - y)\, dx} \\
&= \frac{6xy(2 - x - y)}{y(4 - 3y)} \\
&= \frac{6x(2 - x - y)}{4 - 3y}
\end{aligned}$$

Hence,

$$
\begin{aligned}
E[X|Y=y] &= \int_0^1 \frac{6x^2(2-x-y)\,dx}{4-3y} \\
&= \frac{(2-y)2-\frac{6}{4}}{4-3y} \\
&= \frac{5-4y}{8-6y} \qquad \blacksquare
\end{aligned}
$$

Example 3.7 Suppose the joint density of X and Y is given by

$$
f(x,y) = \begin{cases} 4y(x-y)e^{-(x+y)}, & 0 < x < \infty, 0 \leqslant y \leqslant x \\ 0, & \text{otherwise} \end{cases}
$$

Compute $E[X|Y=y]$.

Solution: The conditional density of X, given that $Y = y$, is given by

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{4y(x-y)e^{-(x+y)}}{\int_y^\infty 4y(x-y)e^{-(x+y)}\,dx}, \qquad x > y \\
&= \frac{(x-y)e^{-x}}{\int_y^\infty (x-y)e^{-x}\,dx} \\
&= \frac{(x-y)e^{-x}}{\int_0^\infty we^{-(y+w)}\,dw}, \qquad x > y \qquad \text{(by letting } w = x-y\text{)} \\
&= (x-y)e^{-(x-y)}, \qquad x > y
\end{aligned}
$$

where the final equality used that $\int_0^\infty we^{-w}dw$ is the expected value of an exponential random variable with mean 1. Therefore, with W being exponential with mean 1,

$$
E[X|Y=y] = \int_y^\infty x(x-y)e^{-(x-y)}\,dx
$$

$$= \int_0^\infty (w + y)we^{-w}\, dw$$
$$= E[W^2] + yE[W]$$
$$= 2 + y \quad \blacksquare$$

Example 3.8 The joint density of X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{2}ye^{-xy}, & 0 < x < \infty, 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

What is $E[e^{X/2}|Y = 1]$?

Solution: The conditional density of X, given that $Y = 1$, is given by

$$f_{X|Y}(x|1) = \frac{f(x, 1)}{f_Y(1)}$$
$$= \frac{\frac{1}{2}e^{-x}}{\int_0^\infty \frac{1}{2}e^{-x}\, dx} = e^{-x}$$

Hence, by Proposition 2.1,

$$E[e^{X/2}|Y = 1] = \int_0^\infty e^{x/2} f_{X|Y}(x|1)\, dx$$
$$= \int_0^\infty e^{x/2}e^{-x}\, dx$$
$$= 2 \quad \blacksquare$$

3.4. Computing Expectations by Conditioning

Let us denote by $E[X|Y]$ that function of the random variable Y whose value at $Y = y$ is $E[X|Y = y]$. Note that $E[X|Y]$ is itself a random variable. An extremely important property of conditional expectation is that for all random variables X

and Y

$$E[X] = E\big[E[X|Y]\big] \tag{3.3}$$

If Y is a discrete random variable, then Equation (3.3) states that

$$E[X] = \sum_y E[X|Y = y]P\{Y = y\} \tag{3.3a}$$

while if Y is continuous with density $f_Y(y)$, then Equation (3.3) says that

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y)\, dy \tag{3.3b}$$

We now give a proof of Equation (3.3) in the case where X and Y are both discrete random variables.

Proof of Equation (3.3) When *X* and *Y* Are Discrete We must show that

$$E[X] = \sum_y E[X|Y = y]P\{Y = y\} \tag{3.4}$$

Now, the right side of the preceding can be written

$$\begin{aligned}
\sum_y E[X|Y = y]P\{Y = y\} &= \sum_y \sum_x xP\{X = x|Y = y\}P\{Y = y\} \\
&= \sum_y \sum_x x \frac{P\{X = x, Y = y\}}{P\{Y = y\}} P\{Y = y\} \\
&= \sum_y \sum_x xP\{X = x, Y = y\} \\
&= \sum_x x \sum_y P\{X = x, Y = y\} \\
&= \sum_x xP\{X = x\} \\
&= E[X]
\end{aligned}$$

and the result is obtained. ■

One way to understand Equation (3.4) is to interpret it as follows. It states that to calculate $E[X]$ we may take a weighted average of the conditional expected value of X given that $Y = y$, each of the terms $E[X|Y = y]$ being weighted by the probability of the event on which it is conditioned.

The following examples will indicate the usefulness of Equation (3.3).

Example 3.9 Sam will read either one chapter of his probability book or one chapter of his history book. If the number of misprints in a chapter of his probability book is Poisson distributed with mean 2 and if the number of misprints in his history chapter is Poisson distributed with mean 5, then assuming Sam is equally likely to choose either book, what is the expected number of misprints that Sam will come across?

Solution: Letting X denote the number of misprints and letting

$$Y = \begin{cases} 1, & \text{if Sam chooses his history book} \\ 2, & \text{if Sam chooses his probability book} \end{cases}$$

then

$$\begin{aligned} E[X] &= E[X|Y=1]P\{Y=1\} + E[X|Y=2]P\{Y=2\} \\ &= 5\left(\tfrac{1}{2}\right) + 2\left(\tfrac{1}{2}\right) \\ &= \tfrac{7}{2} \quad \blacksquare \end{aligned}$$

Example 3.10 (The Expectation of the Sum of a Random Number of Random Variables) Suppose that the expected number of accidents per week at an industrial plant is four. Suppose also that the numbers of workers injured in each accident are independent random variables with a common mean of 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of injuries during a week?

Solution: Letting N denote the number of accidents and X_i the number injured in the ith accident, $i = 1, 2, \ldots,$ then the total number of injuries can be expressed as $\sum_{i=1}^{N} X_i$. Now

$$E\left[\sum_1^N X_i\right] = E\left[E\left[\sum_1^N X_i | N\right]\right]$$

But

$$\begin{aligned} E\left[\sum_1^N X_i | N = n\right] &= E\left[\sum_1^n X_i | N = n\right] \\ &= E\left[\sum_1^n X_i\right] \qquad \text{by the independence of } X_i \text{ and } N \\ &= nE[X] \end{aligned}$$

which yields that

$$E\left[\sum_{i=1}^{N} X_i | N\right] = NE[X]$$

and thus

$$E\left[\sum_{i=1}^{N} X_i\right] = E\big[NE[X]\big] = E[N]E[X]$$

Therefore, in our example, the expected number of injuries during a week equals $4 \times 2 = 8$. ■

The random variable $\sum_{i=1}^{N} X_i$, equal to the sum of a random number N of independent and identically distributed random variables that are also independent of N, is called a *compound random variable*. As just shown in Example 3.10, the expected value of a compound random variable is $E[X]E[N]$. Its variance will be derived in Example 3.17.

Example 3.11 (The Mean of a Geometric Distribution) A coin, having probability p of coming up heads, is to be successively flipped until the first head appears. What is the expected number of flips required?

Solution: Let N be the number of flips required, and let

$$Y = \begin{cases} 1, & \text{if the first flip results in a head} \\ 0, & \text{if the first flip results in a tail} \end{cases}$$

Now

$$\begin{aligned} E[N] &= E[N|Y=1]P\{Y=1\} + E[N|Y=0]P\{Y=0\} \\ &= pE[N|Y=1] + (1-p)E[N|Y=0] \end{aligned} \tag{3.5}$$

However,

$$E[N|Y=1] = 1, \qquad E[N|Y=0] = 1 + E[N] \tag{3.6}$$

To see why Equation (3.6) is true, consider $E[N|Y=1]$. Since $Y=1$, we know that the first flip resulted in heads and so, clearly, the expected number of flips required is 1. On the other hand if $Y=0$, then the first flip resulted in tails. However, since the successive flips are assumed independent, it follows that, after the first tail, the expected additional number of flips until the first head is just $E[N]$. Hence $E[N|Y=0] = 1 + E[N]$. Substituting Equation (3.6) into Equation (3.5) yields

$$E[N] = p + (1-p)(1 + E[N])$$

or

$$E[N] = 1/p \quad \blacksquare$$

Because the random variable N is a geometric random variable with probability mass function $p(n) = p(1-p)^{n-1}$, its expectation could easily have been computed from $E[N] = \sum_1^\infty np(n)$ without recourse to conditional expectation. However, if you attempt to obtain the solution to our next example without using conditional expectation, you will quickly learn what a useful technique "conditioning" can be.

Example 3.12 A miner is trapped in a mine containing three doors. The first door leads to a tunnel that takes him to safety after two hours of travel. The second door leads to a tunnel that returns him to the mine after three hours of travel. The third door leads to a tunnel that returns him to his mine after five hours. Assuming that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until the miner reaches safety?

Solution: Let X denote the time until the miner reaches safety, and let Y denote the door he initially chooses. Now

$$\begin{aligned} E[X] &= E[X|Y=1]P\{Y=1\} + E[X|Y=2]P\{Y=2\} \\ &\quad + E[X|Y=3]P\{Y=3\} \\ &= \tfrac{1}{3}\big(E[X|Y=1] + E[X|Y=2] + E[X|Y=3]\big) \end{aligned}$$

However,

$$\begin{aligned} E[X|Y=1] &= 2, \\ E[X|Y=2] &= 3 + E[X], \\ E[X|Y=3] &= 5 + E[X], \end{aligned} \tag{3.7}$$

To understand why this is correct consider, for instance, $E[X|Y=2]$, and reason as follows. If the miner chooses the second door, then he spends three hours in the tunnel and then returns to the mine. But once he returns to the mine the problem is as before, and hence his expected additional time until safety is just $E[X]$. Hence $E[X|Y=2] = 3 + E[X]$. The argument behind the other equalities in Equation (3.7) is similar. Hence

$$E[X] = \tfrac{1}{3}\big(2 + 3 + E[X] + 5 + E[X]\big) \qquad \text{or} \qquad E[X] = 10 \quad \blacksquare$$

Example 3.13 (The Matching Rounds Problem) Suppose in Example 2.31 that those choosing their own hats depart, while the others (those without a match) put their selected hats in the center of the room, mix them up, and then reselect. Also, suppose that this process continues until each individual has his own hat.

(a) Find $E[R_n]$ where R_n is the number of rounds that are necessary when n individuals are initially present.
(b) Find $E[S_n]$ where S_n is the total number of selections made by the n individuals, $n \geqslant 2$.
(c) Find the expected number of false selections made by one of the n people, $n \geqslant 2$.

Solution: (a) It follows from the results of Example 2.31 that no matter how many people remain there will, on average, be one match per round. Hence, one might suggest that $E[R_n] = n$. This turns out to be true, and an induction proof will now be given. Because it is obvious that $E[R_1] = 1$, assume that $E[R_k] = k$ for $k = 1, \ldots, n-1$. To compute $E[R_n]$, start by conditioning on X_n, the number of matches that occur in the first round. This gives

$$E[R_n] = \sum_{i=0}^{n} E[R_n | X_n = i] P\{X_n = i\}$$

Now, given a total of i matches in the initial round, the number of rounds needed will equal 1 plus the number of rounds that are required when $n - i$ persons are to be matched with their hats. Therefore,

$$\begin{aligned}
E[R_n] &= \sum_{i=0}^{n} (1 + E[R_{n-i}]) P\{X_n = i\} \\
&= 1 + E[R_n] P\{X_n = 0\} + \sum_{i=1}^{n} E[R_{n-i}] P\{X_n = i\} \\
&= 1 + E[R_n] P\{X_n = 0\} + \sum_{i=1}^{n} (n-i) P\{X_n = i\} \\
&\qquad \text{by the induction hypothesis} \\
&= 1 + E[R_n] P\{X_n = 0\} + n(1 - P\{X_n = 0\}) - E[X_n] \\
&= E[R_n] P\{X_n = 0\} + n(1 - P\{X_n = 0\})
\end{aligned}$$

where the final equality used the result, established in Example 2.31, that $E[X_n] = 1$. Since the preceding equation implies that $E[R_n] = n$, the result is proven.

(b) For $n \geqslant 2$, conditioning on X_n, the number of matches in round 1, gives

$$\begin{aligned} E[S_n] &= \sum_{i=0}^{n} E[S_n|X_n = i]P\{X_n = i\} \\ &= \sum_{i=0}^{n} (n + E[S_{n-i}])P\{X_n = i\} \\ &= n + \sum_{i=0}^{n} E[S_{n-i}]P\{X_n = i\} \end{aligned}$$

where $E[S_0] = 0$. To solve the preceding equation, rewrite it as

$$E[S_n] = n + E[S_{n-X_n}]$$

Now, if there were *exactly* one match in each round, then it would take a total of $1 + 2 + \cdots + n = n(n + 1)/2$ selections. Thus, let us try a solution of the form $E[S_n] = an + bn^2$. For the preceding equation to be satisfied by a solution of this type, for $n \geqslant 2$, we need

$$an + bn^2 = n + E[a(n - X_n) + b(n - X_n)^2]$$

or, equivalently,

$$an + bn^2 = n + a(n - E[X_n]) + b(n^2 - 2nE[X_n] + E[X_n^2])$$

Now, using the results of Example 2.31 and Exercise 72 of Chapter 2 that $E[X_n]=\text{Var}(X_n) = 1$, the preceding will be satisfied if

$$an + bn^2 = n + an - a + bn^2 - 2nb + 2b$$

and this will be valid provided that $b = 1/2, a = 1$. That is,

$$E[S_n] = n + n^2/2$$

satisfies the recursive equation for $E[S_n]$.

The formal proof that $E[S_n] = n + n^2/2$, $n \geqslant 2$, is obtained by induction on n. It is true when $n = 2$ (since, in this case, the number of selections is twice the number of rounds and the number of rounds is a geometric random variable with parameter $p = 1/2$). Now, the recursion gives that

$$E[S_n] = n + E[S_n]P\{X_n = 0\} + \sum_{i=1}^{n} E[S_{n-i}]P\{X_n = i\}$$

Hence, upon assuming that $E[S_0] = E[S_1] = 0$, $E[S_k] = k + k^2/2$, for $k = 2, \ldots, n-1$ and using that $P\{X_n = n-1\} = 0$, we see that

$$\begin{aligned} E[S_n] &= n + E[S_n]P\{X_n = 0\} + \sum_{i=1}^{n} [n - i + (n-i)^2/2]P\{X_n = i\} \\ &= n + E[S_n]P\{X_n = 0\} + (n + n^2/2)(1 - P\{X_n = 0\}) \\ &\quad - (n+1)E[X_n] + E[X_n^2]/2 \end{aligned}$$

Substituting the identities $E[X_n] = 1$, $E[X_n^2] = 2$ in the preceding shows that

$$E[S_n] = n + n^2/2$$

and the induction proof is complete.

(c) If we let C_j denote the number of hats chosen by person j, $j = 1, \ldots, n$ then

$$\sum_{j=1}^{n} C_j = S_n$$

Taking expectations, and using the fact that each C_j has the same mean, yields the result

$$E[C_j] = E[S_n]/n = 1 + n/2$$

Hence, the expected number of false selections by person j is

$$E[C_j - 1] = n/2. \quad \blacksquare$$

Example 3.14 Independent trials, each of which is a success with probability p, are performed until there are k consecutive successes. What is the mean number of necessary trials?

Solution: Let N_k denote the number of necessary trials to obtain k consecutive successes, and let M_k denote its mean. We will obtain a recursive equation for M_k by conditioning on N_{k-1}, the number of trials needed for $k-1$ consecutive successes. This yields

$$M_k = E[N_k] = E\big[E[N_k|N_{k-1}]\big]$$

Now,

$$E[N_k|N_{k-1}] = N_{k-1} + 1 + (1-p)E[N_k]$$

where the preceding follows since if it takes N_{k-1} trials to obtain $k-1$ consecutive successes, then either the next trial is a success and we have our k in a row or it is a failure and we must begin anew. Taking expectations of both sides of the preceding yields

$$M_k = M_{k-1} + 1 + (1-p)M_k$$

or

$$M_k = \frac{1}{p} + \frac{M_{k-1}}{p}$$

Since N_1, the time of the first success, is geometric with parameter p, we see that

$$M_1 = \frac{1}{p}$$

and, recursively

$$M_2 = \frac{1}{p} + \frac{1}{p^2},$$

$$M_3 = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3}$$

and, in general,

$$M_k = \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k} \quad \blacksquare$$

Example 3.15 (Analyzing the Quick-Sort Algorithm) Suppose we are given a set of n distinct values—$x_1, \ldots, x_n$—and we desire to put these values in increasing order or, as it is commonly called, to *sort* them. An efficient procedure for accomplishing this is the quick-sort algorithm which is defined recursively as follows: When $n = 2$ the algorithm compares the two values and puts them in the appropriate order. When $n > 2$ it starts by choosing at random one of the n values—say, x_i—and then compares each of the other $n-1$ values with x_i, noting which are smaller and which are larger than x_i. Letting S_i denote the set of elements smaller than x_i, and $\bar{S}_i$ the set of elements greater than x_i, the algorithm now sorts the set S_i and the set $\bar{S}_i$. The final ordering, therefore, consists of the ordered set of the elements in S_i, then x_i, and then the ordered set of the elements in $\bar{S}_i$. For instance, suppose that the set of elements is 10, 5, 8, 2, 1, 4, 7. We start by choosing one of these values at random (that is, each of the 7 values has probability of $\frac{1}{7}$ of being chosen). Suppose, for instance, that the value 4 is chosen. We then compare 4 with each of the other six values to obtain

$$\{2, 1\},\ 4,\ \{10, 5, 8, 7\}$$

We now sort the set $\{2, 1\}$ to obtain

$$1, 2, 4, \{10, 5, 8, 7\}$$

Next we choose a value at random from $\{10, 5, 8, 7\}$—say 7 is chosen—and compare each of the other three values with 7 to obtain

$$1, 2, 4, 5, 7, \{10, 8\}$$

Finally, we sort $\{10, 8\}$ to end up with

$$1, 2, 4, 5, 7, 8, 10$$

One measure of the effectiveness of this algorithm is the expected number of comparisons that it makes. Let us denote by M_n the expected number of comparisons needed by the quick-sort algorithm to sort a set of n distinct values. To obtain a recursion for M_n we condition on the rank of the initial value selected to obtain:

$$M_n = \sum_{j=1}^{n} E[\text{number of comparisons}|\text{value selected is } j\text{th smallest}]\frac{1}{n}$$

Now if the initial value selected is the jth smallest, then the set of values smaller than it is of size $j - 1$, and the set of values greater than it is of size $n - j$. Hence, as $n - 1$ comparisons with the initial value chosen must be made, we see that

$$\begin{aligned} M_n &= \sum_{j=1}^{n} (n - 1 + M_{j-1} + M_{n-j})\frac{1}{n} \\ &= n - 1 + \frac{2}{n}\sum_{k=1}^{n-1} M_k \qquad (\text{since } M_0 = 0) \end{aligned}$$

or, equivalently,

$$nM_n = n(n-1) + 2\sum_{k=1}^{n-1} M_k$$

To solve the preceding, note that upon replacing n by $n+1$ we obtain

$$(n+1)M_{n+1} = (n+1)n + 2\sum_{k=1}^{n} M_k$$

Hence, upon subtraction,

$$(n+1)M_{n+1} - nM_n = 2n + 2M_n$$

or

$$(n+1)M_{n+1} = (n+2)M_n + 2n$$

Therefore,

$$\frac{M_{n+1}}{n+2} = \frac{2n}{(n+1)(n+2)} + \frac{M_n}{n+1}$$

Iterating this gives

$$\begin{aligned}
\frac{M_{n+1}}{n+2} &= \frac{2n}{(n+1)(n+2)} + \frac{2(n-1)}{n(n+1)} + \frac{M_{n-1}}{n} \\
&= \cdots \\
&= 2\sum_{k=0}^{n-1} \frac{n-k}{(n+1-k)(n+2-k)} \qquad \text{since } M_1 = 0
\end{aligned}$$

Hence,

$$\begin{aligned}
M_{n+1} &= 2(n+2)\sum_{k=0}^{n-1} \frac{n-k}{(n+1-k)(n+2-k)} \\
&= 2(n+2)\sum_{i=1}^{n} \frac{i}{(i+1)(i+2)}, \qquad n \geqslant 1
\end{aligned}$$

Using the identity $i/(i+1)(i+2) = 2/(i+2) - 1/(i+1)$, we can approximate M_{n+1} for large n as follows:

$$\begin{aligned}
M_{n+1} &= 2(n+2)\left[\sum_{i=1}^{n} \frac{2}{i+2} - \sum_{i=1}^{n} \frac{1}{i+1}\right] \\
&\sim 2(n+2)\left[\int_3^{n+2} \frac{2}{x}\,dx - \int_2^{n+1} \frac{1}{x}\,dx\right] \\
&= 2(n+2)[2\log(n+2) - \log(n+1) + \log 2 - 2\log 3] \\
&= 2(n+2)\left[\log(n+2) + \log\frac{n+2}{n+1} + \log 2 - 2\log 3\right] \\
&\sim 2(n+2)\log(n+2) \quad \blacksquare
\end{aligned}$$

Although we usually employ the conditional expectation identity to more easily enable us to compute an unconditional expectation, in our next example we show how it can sometimes be used to obtain the conditional expectation.

Example 3.16 In the match problem of Example 2.31 involving n, $n > 1$, individuals, find the conditional expected number of matches given that the first person did not have a match.

Solution: Let X denote the number of matches, and let X_1 equal 1 if the first person has a match and let it equal 0 otherwise. Then,

$$\begin{aligned}
E[X] &= E[X|X_1 = 0]P\{X_1 = 0\} + E[X|X_1 = 1]P\{X_1 = 1\} \\
&= E[X|X_1 = 0]\,\frac{n-1}{n} + E[X|X_1 = 1]\,\frac{1}{n}
\end{aligned}$$

But, from Example 2.31

$$E[X] = 1$$

Moreover, given that the first person has a match, the expected number of matches is equal to 1 plus the expected number of matches when $n-1$ people select among their own $n-1$ hats, showing that

$$E[X|X_1 = 1] = 2$$

Therefore, we obtain the result

$$E[X|X_1 = 0] = \frac{n-2}{n-1} \quad \blacksquare$$

3.4.1. Computing Variances by Conditioning

Conditional expectations can also be used to compute the variance of a random variable. Specifically, we can use that

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

and then use conditioning to obtain both $E[X]$ and $E[X^2]$. We illustrate this technique by determining the variance of a geometric random variable.

Example 3.17 (Variance of the Geometric Random Variable) Independent trials, each resulting in a success with probability p, are performed in sequence. Let N be the trial number of the first success. Find $\mathrm{Var}(N)$.

Solution: Let $Y = 1$ if the first trial results in a success, and $Y = 0$ otherwise.

$$\mathrm{Var}(N) = E(N^2) - (E[N])^2$$

To calculate $E[N^2]$ and $E[N]$ we condition on Y. For instance,

$$E[N^2] = E\big[E[N^2|Y]\big]$$

However,

$$E[N^2|Y = 1] = 1,$$
$$E[N^2|Y = 0] = E[(1 + N)^2]$$

These two equations are true since if the first trial results in a success, then clearly $N = 1$ and so $N^2 = 1$. On the other hand, if the first trial results in a failure, then the total number of trials necessary for the first success will equal one (the first trial that results in failure) plus the necessary number of additional trials. Since this latter quantity has the same distribution as N, we get that $E[N^2|Y=0]=E[(1+N)^2]$. Hence, we see that

$$\begin{aligned}E[N^2]&=E[N^2|Y=1]P\{Y=1\}+E[N^2|Y=0]P\{Y=0\}\\&=p+E[(1+N)^2](1-p)\\&=1+(1-p)E[2N+N^2]\end{aligned}$$

Since, as was shown in Example 3.11, $E[N]=1/p$, this yields

$$E[N^2]=1+\frac{2(1-p)}{p}+(1-p)E[N^2]$$

or

$$E[N^2]=\frac{2-p}{p^2}$$

Therefore,

$$\begin{aligned}\mathrm{Var}(N)&=E[N^2]-(E[N])^2\\&=\frac{2-p}{p^2}-\left(\frac{1}{p}\right)^2\\&=\frac{1-p}{p^2}\quad\blacksquare\end{aligned}$$

Another way to use conditioning to obtain the variance of a random variable is to apply the conditional variance formula. The conditional variance of X given that $Y=y$ is defined by

$$\mathrm{Var}(X|Y=y)=E\big[(X-E[X|Y=y])^2|Y=y\big]$$

That is, the conditional variance is defined in exactly the same manner as the ordinary variance with the exception that all probabilities are determined conditional on the event that $Y=y$. Expanding the right side of the preceding and taking expectation term by term yield that

$$\mathrm{Var}(X|Y=y)=E[X^2|Y=y]-(E[X|Y=y])^2$$

Letting $\mathrm{Var}(X|Y)$ denote that function of Y whose value when $Y=y$ is $\mathrm{Var}(X|Y=y)$, we have the following result.

Proposition 3.1 The Conditional Variance Formula

$$\operatorname{Var}(X) = E\big[\operatorname{Var}(X|Y)\big] + \operatorname{Var}\big(E[X|Y]\big) \tag{3.8}$$

Proof

$$\begin{aligned}
E\big[\operatorname{Var}(X|Y)\big] &= E\big[E[X^2|Y] - (E[X|Y])^2\big] \\
&= E\big[E[X^2|Y]\big] - E\big[(E[X|Y])^2\big] \\
&= E[X^2] - E\big[(E[X|Y])^2\big]
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{Var}(E[X|Y]) &= E\big[(E[X|Y])^2\big] - \big(E\big[E[X|Y]\big]\big)^2 \\
&= E\big[(E[X|Y])^2\big] - (E[X])^2
\end{aligned}$$

Therefore,

$$E\big[\operatorname{Var}(X|Y)\big] + \operatorname{Var}\big(E[X|Y]\big) = E[X^2] - (E[X])^2$$

which completes the proof. ■

Example 3.18 (The Variance of a Compound Random Variable) Let X_1, $X_2, \ldots$ be independent and identically distributed random variables with distribution F having mean μ and variance σ^2, and assume that they are independent of the nonnegative integer valued random variable N. As noted in Example 3.10, where its expected value was determined, the random variable $S = \sum_{i=1}^{N} X_i$ is called a compound random variable. Find its variance.

Solution: Whereas we could obtain $E[S^2]$ by conditioning on N, let us instead use the conditional variance formula. Now,

$$\begin{aligned}
\operatorname{Var}(S|N = n) &= \operatorname{Var}\left(\sum_{i=1}^{N} X_i | N = n\right) \\
&= \operatorname{Var}\left(\sum_{i=1}^{n} X_i | N = n\right) \\
&= \operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= n\sigma^2
\end{aligned}$$

By the same reasoning,

$$E[S|N = n] = n\mu$$

Therefore,

$$\text{Var}(S|N) = N\sigma^2, \qquad E[S|N] = N\mu$$

and the conditional variance formula gives that

$$\text{Var}(S) = E[N\sigma^2] + \text{Var}(N\mu) = \sigma^2 E[N] + \mu^2 \text{Var}(N)$$

If N is a Poisson random variable, then $S = \sum_{i=1}^{N} X_i$ is called a *compound Poisson* random variable. Because the variance of a Poisson random variable is equal to its mean, it follows that for a compound Poisson random variable having $E[N] = \lambda$

$$\text{Var}(S) = \lambda\sigma^2 + \lambda\mu^2 = \lambda E[X^2]$$

where X has the distribution F. ■

3.5. Computing Probabilities by Conditioning

Not only can we obtain expectations by first conditioning on an appropriate random variable, but we may also use this approach to compute probabilities. To see this, let E denote an arbitrary event and define the indicator random variable X by

$$X = \begin{cases} 1, & \text{if } E \text{ occurs} \\ 0, & \text{if } E \text{ does not occur} \end{cases}$$

It follows from the definition of X that

$$\begin{aligned} E[X] &= P(E), \\ E[X|Y = y] &= P(E|Y = y), \qquad \text{for any random variable } Y \end{aligned}$$

Therefore, from Equations (3.3a) and (3.3b) we obtain

$$\begin{aligned} P(E) &= \sum_y P(E|Y = y)P(Y = y), \qquad &\text{if } Y \text{ is discrete} \\ &= \int_{-\infty}^{\infty} P(E|Y = y) f_Y(y)\, dy, \qquad &\text{if } Y \text{ is continuous} \end{aligned}$$

Example 3.19 Suppose that X and Y are independent continuous random variables having densities f_X and f_Y, respectively. Compute $P\{X < Y\}$.

Solution: Conditioning on the value of Y yields

$$\begin{aligned}
P\{X < Y\} &= \int_{-\infty}^{\infty} P\{X < Y|Y = y\} f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} P\{X < y|Y = y\} f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} P\{X < y\} f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} F_X(y) f_Y(y)\, dy
\end{aligned}$$

where

$$F_X(y) = \int_{-\infty}^{y} f_X(x)\, dx \quad \blacksquare$$

Example 3.20 An insurance company supposes that the number of accidents that each of its policyholders will have in a year is Poisson distributed, with the mean of the Poisson depending on the policyholder. If the Poisson mean of a randomly chosen policyholder has a gamma distribution with density function

$$g(\lambda) = \lambda e^{-\lambda}, \qquad \lambda \geqslant 0$$

what is the probability that a randomly chosen policyholder has exactly n accidents next year?

Solution: Let X denote the number of accidents that a randomly chosen policyholder has next year. Letting Y be the Poisson mean number of accidents for this policyholder, then conditioning on Y yields

$$\begin{aligned}
P\{X = n\} &= \int_0^{\infty} P\{X = n|Y = \lambda\} g(\lambda)\, d\lambda \\
&= \int_0^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \lambda e^{-\lambda}\, d\lambda \\
&= \frac{1}{n!} \int_0^{\infty} \lambda^{n+1} e^{-2\lambda}\, d\lambda
\end{aligned}$$

However, because

$$h(\lambda) = \frac{2e^{-2\lambda}(2\lambda)^{n+1}}{(n+1)!}, \qquad \lambda > 0$$

is the density function of a gamma $(n+2, 2)$ random variable, its integral is 1. Therefore,

$$1 = \int_0^\infty \frac{2e^{-2\lambda}(2\lambda)^{n+1}}{(n+1)!}\, d\lambda = \frac{2^{n+2}}{(n+1)!} \int_0^\infty \lambda^{n+1} e^{-2\lambda}\, d\lambda$$

showing that

$$P\{X = n\} = \frac{n+1}{2^{n+2}} \quad \blacksquare$$

Example 3.21 Suppose that the number of people who visit a yoga studio each day is a Poisson random variable with mean λ. Suppose further that each person who visits is, independently, female with probability p or male with probability $1 - p$. Find the joint probability that exactly n women and m men visit the academy today.

Solution: Let N_1 denote the number of women, and N_2 the number of men, who visit the academy today. Also, let $N = N_1 + N_2$ be the total number of people who visit. Conditioning on N gives

$$P\{N_1 = n, N_2 = m\} = \sum_{i=0}^{\infty} P\{N_1 = n, N_2 = m | N = i\}\, P\{N = i\}$$

Because $P\{N_1 = n, N_2 = m | N = i\} = 0$ when $i \neq n + m$, the preceding equation yields that

$$P\{N_1 = n, N_2 = m\} = P\{N_1 = n, N_2 = m | N = n + m\} e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!}$$

Given that $n + m$ people visit it follows, because each of these $n + m$ is independently a woman with probability p, that the conditional probability that n of them are women (and m are men) is just the binomial probability of n

successes in $n+m$ trials. Therefore,

$$
\begin{aligned}
P\{N_1=n,\ N_2=m\} &= \binom{n+m}{n} p^n(1-p)^m e^{-\lambda}\frac{\lambda^{n+m}}{(n+m)!}\\
&= \frac{(n+m)!}{n!m!} p^n(1-p)^m e^{-\lambda p} e^{-\lambda(1-p)}\frac{\lambda^n\lambda^m}{(n+m)!}\\
&= e^{-\lambda p}\frac{(\lambda p)^n}{n!}\, e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}
\end{aligned}
$$

Because the preceding joint probability mass function factors into two products, one of which depends only on n and the other only on m, it follows that N_1 and N_2 are independent. Moreover, because

$$
\begin{aligned}
P\{N_1=n\} &= \sum_{m=0}^{\infty} P\{N_1=n,\ N_2=m\}\\
&= e^{-\lambda p}\frac{(\lambda p)^n}{n!}\sum_{m=0}^{\infty} e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!} = e^{-\lambda p}\frac{(\lambda p)^n}{n!}
\end{aligned}
$$

and, similarly,

$$
P\{N_2=m\} = e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}
$$

we can conclude that N_1 and N_2 are independent Poisson random variables with respective means λp and $\lambda(1-p)$. Therefore, this example establishes the important result that when each of a Poisson number of events is independently classified either as being type 1 with probability p or type 2 with probability $1-p$, then the numbers of type 1 and type 2 events are independent Poisson random variables. ■

Example 3.22 Let $X_1,\dots,X_n$ be independent Bernoulli random variables, with X_i having parameter p_i, $i=1,\dots,n$. That is, $P\{X_i=1\}=p_i$, $P\{X_i=0\}=q_i=1-p_i$. Suppose we want to compute the probability mass function of their sum, $X_1+\cdots+X_n$. To do so, we will recursively obtain the probability mass function of $X_1+\cdots+X_k$, first for $k=1$, then $k=2$, and on up to $k=n$. To begin, let

$$
P_k(j) = P\{X_1+\cdots+X_k=j\}
$$

and note that

$$P_k(k) = \prod_{i=1}^{k} p_i, \qquad P_k(0) = \prod_{i=1}^{k} q_i$$

For $0 < j < k$, conditioning on X_k yields the recursion

$$\begin{aligned}
P_k(j) &= P\{X_1 + \cdots + X_k = j | X_k = 1\} p_k + P\{X_1 + \cdots + X_k = j | X_k = 0\} q_k \\
&= P\{X_1 + \cdots + X_{k-1} = j - 1 | X_k = 1\} p_k \\
&\quad + P\{X_1 + \cdots + X_{k-1} = j | X_k = 0\} q_k \\
&= P\{X_1 + \cdots + X_{k-1} = j - 1\} p_k + P\{X_1 + \cdots + X_{k-1} = j\} q_k \\
&= p_k\, P_{k-1}(j-1) + q_k\, P_{k-1}(j)
\end{aligned}$$

Starting with $P_1(1) = p_1,\ P_1(0) = q_1,$ the preceding equations can be recursively solved to obtain the functions $P_2(j)$, $P_3(j)$, up to $P_n(j)$. ■

Example 3.23 (The Best Prize Problem) Suppose that we are to be presented with n distinct prizes in sequence. After being presented with a prize we must immediately decide whether to accept it or reject it and consider the next prize. The only information we are given when deciding whether to accept a prize is the relative rank of that prize compared to ones already seen. That is, for instance, when the fifth prize is presented we learn how it compares with the first four prizes already seen. Suppose that once a prize is rejected it is lost, and that our objective is to maximize the probability of obtaining the best prize. Assuming that all $n!$ orderings of the prizes are equally likely, how well can we do?

Solution: Rather surprisingly, we can do quite well. To see this, fix a value $k, 0 \leqslant k < n$, and consider the strategy that rejects the first k prizes and then accepts the first one that is better than all of those first k. Let P_k (best) denote the probability that the best prize is selected when this strategy is employed. To compute this probability, condition on X, the position of the best prize. This gives

$$\begin{aligned}
P_k(\text{best}) &= \sum_{i=1}^{n} P_k(\text{best}|X = i) P(X = i) \\
&= \frac{1}{n} \sum_{i=1}^{n} P_k(\text{best}|X = i)
\end{aligned}$$

Now, if the overall best prize is among the first k, then no prize is ever selected under the strategy considered. On the other hand, if the best prize is in position i, where $i > k$, then the best prize will be selected if the best of the first

k prizes is also the best of the first $i-1$ prizes (for then none of the prizes in positions $k+1, k+2, \ldots, i-1$ would be selected). Hence, we see that

$$\begin{aligned}
P_k(\text{best}|X=i) &= 0, \qquad \text{if } i \leqslant k \\
P_k(\text{best}|X=i) &= P\{\text{best of first } i-1 \text{ is among the first } k\} \\
&= k/(i-1), \qquad \text{if } i > k
\end{aligned}$$

From the preceding, we obtain that

$$\begin{aligned}
P_k(\text{best}) &= \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1} \\
&\approx \frac{k}{n} \int_k^{n-1} \frac{1}{x}\, dx \\
&= \frac{k}{n} \log\left(\frac{n-1}{k}\right) \\
&\approx \frac{k}{n} \log\left(\frac{n}{k}\right)
\end{aligned}$$

Now, if we consider the function

$$g(x) = \frac{x}{n} \log\left(\frac{n}{x}\right)$$

then

$$g'(x) = \frac{1}{n} \log\left(\frac{n}{x}\right) - \frac{1}{n}$$

and so

$$g'(x) = 0 \Rightarrow \log(n/x) = 1 \Rightarrow x = n/e$$

Thus, since $P_k(\text{best}) \approx g(k)$, we see that the best strategy of the type considered is to let the first n/e prizes go by and then accept the first one to appear that is better than all of those. In addition, since $g(n/e) = 1/e$, the probability that this strategy selects the best prize is approximately $1/e \approx 0.36788$.

Remark Most students are quite surprised by the size of the probability of obtaining the best prize, thinking that this probability would be close to 0 when n is large. However, even without going through the calculations, a little thought reveals that the probability of obtaining the best prize can be made to be reasonably large. Consider the strategy of letting half of the prizes go by, and then selecting the first one to appear that is better than all of those. The probability that a prize is actually selected is the probability that the overall best is among the second half and this is $1/2$. In addition, given that a prize is selected, at the time of selection that prize would have been the best of more than $n/2$ prizes to have appeared, and would thus have probability of at least $1/2$ of being the overall best. Hence, the strategy of letting the first half of all prizes go by and then accepting the first one that is better than all of those prizes results in a probability greater than $1/4$ of obtaining the best prize. ■

Example 3.24 At a party n men take off their hats. The hats are then mixed up and each man randomly selects one. We say that a match occurs if a man selects his own hat. What is the probability of no matches? What is the probability of exactly k matches?

Solution: Let E denote the event that no matches occur, and to make explicit the dependence on n, write $P_n = P(E)$. We start by conditioning on whether or not the first man selects his own hat—call these events M and M^c. Then

$$P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c)$$

Clearly, $P(E|M) = 0$, and so

$$P_n = P(E|M^c)\frac{n-1}{n} \tag{3.9}$$

Now, $P(E|M^c)$ is the probability of no matches when $n-1$ men select from a set of $n-1$ hats that does not contain the hat of one of these men. This can happen in either of two mutually exclusive ways. Either there are no matches and the extra man does not select the extra hat (this being the hat of the man that chose first), or there are no matches and the extra man does select the extra hat. The probability of the first of these events is just P_{n-1}, which is seen by regarding the extra hat as "belonging" to the extra man. Because the second event has probability $[1/(n-1)]P_{n-2}$, we have

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1}P_{n-2}$$

and thus, from Equation (3.9),

$$P_n = \frac{n-1}{n}P_{n-1} + \frac{1}{n}P_{n-2}$$

or, equivalently,

$$P_n - P_{n-1} = -\frac{1}{n}(P_{n-1} - P_{n-2}) \tag{3.10}$$

However, because P_n is the probability of no matches when n men select among their own hats, we have

$$P_1 = 0, \qquad P_2 = \tfrac{1}{2}$$

and so, from Equation (3.10),

$$P_3 - P_2 = -\frac{(P_2 - P_1)}{3} = -\frac{1}{3!} \quad \text{or} \quad P_3 = \frac{1}{2!} - \frac{1}{3!},$$

$$P_4 - P_3 = -\frac{(P_3 - P_2)}{4} = \frac{1}{4!} \quad \text{or} \quad P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

and, in general, we see that

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}$$

To obtain the probability of exactly k matches, we consider any fixed group of k men. The probability that they, and only they, select their own hats is

$$\frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_{n-k} = \frac{(n-k)!}{n!}P_{n-k}$$

where P_{n-k} is the conditional probability that the other $n-k$ men, selecting among their own hats, have no matches. Because there are $\binom{n}{k}$ choices of a set of k men, the desired probability of exactly k matches is

$$\frac{P_{n-k}}{k!} = \frac{\dfrac{1}{2!} - \dfrac{1}{3!} + \cdots + \dfrac{(-1)^{n-k}}{(n-k)!}}{k!}$$

which, for n large, is approximately equal to $e^{-1}/k!$.

Remark The recursive equation, Equation (3.10), could also have been obtained by using the concept of a cycle, where we say that the sequence of distinct individuals $i_1, i_2, \ldots, i_k$ constitutes a *cycle* if i_1 chooses i_2's hat, i_2 chooses i_3's hat, $\ldots, i_{k-1}$ chooses i_k's hat, and i_k chooses i_1's hat. Note that every individual is part of a cycle, and that a cycle of size $k = 1$ occurs when someone chooses his or her own hat. With E being, as before, the event that no matches occur, it follows upon conditioning on the size of the cycle containing a specified person, say person 1, that

$$P_n = P(E) = \sum_{k=1}^{n} P(E|C=k)P(C=k) \tag{3.11}$$

where C is the size of the cycle that contains person 1. Now call person 1 the first person, and note that $C = k$ if the first person does not choose 1's hat; the person whose hat was chosen by the first person—call this person the second person—does not choose 1's hat; the person whose hat was chosen by the second person—call this person the third person—does not choose 1's hat; ..., the person whose hat was chosen by the $(k-1)$st person does choose 1's hat. Consequently,

$$P(C=k) = \frac{n-1}{n}\,\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\,\frac{1}{n-k+1} = \frac{1}{n} \tag{3.12}$$

That is, the size of the cycle that contains a specified person is equally likely to be any of the values $1, 2, \ldots, n$. Moreover, since $C = 1$ means that 1 chooses his or her own hat, it follows that

$$P(E|C=1) = 0 \tag{3.13}$$

On the other hand, if $C = k$, then the set of hats chosen by the k individuals in this cycle is exactly the set of hats of these individuals. Hence, conditional on $C = k$, the problem reduces to determining the probability of no matches when $n - k$ people randomly choose among their own $n - k$ hats. Therefore, for $k > 1$

$$P(E|C=k) = P_{n-k}$$

Substituting (3.12), (3.13), and (3.14) back into Equation (3.11) gives

$$P_n = \frac{1}{n}\sum_{k=2}^{n} P_{n-k} \tag{3.14}$$

which is easily shown to be equivalent to Equation (3.10). ■

Example 3.25 (The Ballot Problem) In an election, candidate A receives n votes, and candidate B receives m votes where $n > m$. Assuming that all orderings are equally likely, show that the probability that A is always ahead in the count of votes is $(n-m)/(n+m)$.

Solution: Let $P_{n,m}$ denote the desired probability. By conditioning on which candidate receives the last vote counted we have

$$P_{n,m} = P\{A \text{ always ahead}|A \text{ receives last vote}\}\frac{n}{n+m} + P\{A \text{ always ahead}|B \text{ receives last vote}\}\frac{m}{n+m}$$

Now given that A receives the last vote, we can see that the probability that A is always ahead is the same as if A had received a total of $n-1$ and B a total of m votes. Because a similar result is true when we are given that B receives the last vote, we see from the preceding that

$$P_{n,m} = \frac{n}{n+m}P_{n-1,m} + \frac{m}{m+n}P_{n,m-1} \tag{3.15}$$

We can now prove that $P_{n,m} = (n-m)/(n+m)$ by induction on $n+m$. As it is true when $n+m=1$, that is, $P_{1,0}=1$, assume it whenever $n+m=k$. Then when $n+m=k+1$, we have by Equation (3.15) and the induction hypothesis that

$$\begin{aligned} P_{n,m} &= \frac{n}{n+m}\,\frac{n-1-m}{n-1+m} + \frac{m}{m+n}\,\frac{n-m+1}{n+m-1} \\ &= \frac{n-m}{n+m} \end{aligned}$$

and the result is proven. ■

The ballot problem has some interesting applications. For example, consider successive flips of a coin that always land on "heads" with probability p, and let us determine the probability distribution of the first time, after beginning, that the total number of heads is equal to the total number of tails. The probability that the first time this occurs is at time $2n$ can be obtained by first conditioning on the total number of heads in the first $2n$ trials. This yields

$$\begin{aligned} &P\{\text{first time equal} = 2n\} \\ &\quad = P\{\text{first time equal} = 2n|n \text{ heads in first } 2n\}\binom{2n}{n}p^n(1-p)^n \end{aligned}$$

Now given a total of n heads in the first $2n$ flips we can see that all possible orderings of the n heads and n tails are equally likely, and thus the preceding conditional probability is equivalent to the probability that in an election, in which each candidate receives n votes, one of the candidates is always ahead in the counting until the last vote (which ties them). But by conditioning on whomever receives the last vote, we see that this is just the probability in the ballot problem when $m = n - 1$. Hence

$$P\{\text{first time equal} = 2n\} = P_{n,n-1}\binom{2n}{n}p^n(1-p)^n$$
$$= \frac{\binom{2n}{n}p^n(1-p)^n}{2n-1}$$

Suppose now that we wanted to determine the probability that the first time there are i more heads than tails occurs after the $(2n+i)$th flip. Now, in order for this to be the case, the following two events must occur:

(a) The first $2n+i$ tosses result in $n+i$ heads and n tails; and
(b) The order in which the $n+i$ heads and n tails occur is such that the number of heads is never i more than the number of tails until after the final flip.

Now, it is easy to see that event (b) will occur if and only if the order of appearance of the $n+i$ heads and n tails is such that starting from the final flip and working backwards heads is always in the lead. For instance, if there are 4 heads and 2 tails ($n = 2,\ i = 2$), then the outcome ____TH would not suffice because there would have been 2 more heads than tails sometime before the sixth flip (since the first 4 flips resulted in 2 more heads than tails).

Now, the probability of the event specified in (a) is just the binomial probability of getting $n+i$ heads and n tails in $2n+i$ flips of the coin.

We must now determine the conditional probability of the event specified in (b) given that there are $n+i$ heads and n tails in the first $2n+i$ flips. To do so, note first that given that there are a total of $n+i$ heads and n tails in the first $2n+i$ flips, all possible orderings of these flips are equally likely. As a result, the conditional probability of (b) given (a) is just the probability that a random ordering of $n+i$ heads and n tails will, when counted in reverse order, always have more heads than tails. Since all reverse orderings are also equally likely, it follows from the ballot problem that this conditional probability is $i/(2n+i)$.

That is, we have shown that

$$P\{a\} = \binom{2n+i}{n}p^{n+i}(1-p)^n,$$
$$P\{b|a\} = \frac{i}{2n+i}$$

and so

$$P\{\text{first time heads leads by } i \text{ is after flip } 2n+i\} = \binom{2n+i}{n}p^{n+i}(1-p)^n\frac{i}{2n+i}$$

Example 3.26 Let $U_1, U_2, \dots$ be a sequence of independent uniform (0, 1) random variables, and let

$$N = \min\{n \geqslant 2: U_n > U_{n-1}\}$$

and

$$M = \min\{n \geqslant 1: U_1 + \cdots + U_n > 1\}$$

That is, N is the index of the first uniform random variable that is larger than its immediate predecessor, and M is the number of uniform random variables we need sum to exceed 1. Surprisingly, N and M have the same probability distribution, and their common mean is e!

Solution: It is easy to find the distribution of N. Since all $n!$ possible orderings of $U_1, \dots, U_n$ are equally likely, we have

$$P\{N > n\} = P\{U_1 > U_2 > \cdots > U_n\} = 1/n!$$

To show that $P\{M > n\} = 1/n!$, we will use mathematical induction. However, to give ourselves a stronger result to use as the induction hypothesis, we will prove the stronger result that for $0 < x \leqslant 1$, $P\{M(x) > n\} = x^n/n!$, $n \geqslant 1$, where

$$M(x) = \min\{n \geqslant 1: U_1 + \cdots + U_n > x\}$$

is the minimum number of uniforms that need be summed to exceed x. To prove that $P\{M(x) > n\} = x^n/n!$, note first that it is true for $n = 1$ since

$$P\{M(x) > 1\} = P\{U_1 \leqslant x\} = x$$

So assume that for all $0 < x \leqslant 1$, $P\{M(x) > n\} = x^n/n!$. To determine $P\{M(x) > n + 1\}$, condition on U_1 to obtain:

$$\begin{aligned}
P\{M(x) > n + 1\} &= \int_0^1 P\{M(x) > n + 1 | U_1 = y\}\, dy \\
&= \int_0^x P\{M(x) > n + 1 | U_1 = y\}\, dy \\
&= \int_0^x P\{M(x - y) > n\}\, dy \\
&= \int_0^x \frac{(x - y)^n}{n!}\, dy \qquad \text{by the induction hypothesis}
\end{aligned}$$

$$= \int_0^x \frac{u^n}{n!}\, du$$

$$= \frac{x^{n+1}}{(n+1)!}$$

where the third equality of the preceding follows from the fact that given $U_1 = y$, $M(x)$ is distributed as 1 plus the number of uniforms that need be summed to exceed $x - y$. Thus, the induction is complete and we have shown that for $0 < x \leqslant 1, n \geqslant 1$,

$$P\{M(x) > n\} = x^n/n!$$

Letting $x = 1$ shows that N and M have the same distribution. Finally, we have that

$$E[M] = E[N] = \sum_{n=0}^{\infty} P\{N > n\} = \sum_{n=0}^{\infty} 1/n! = e \quad \blacksquare$$

Example 3.27 Let $X_1, X_2, \ldots$ be independent continuous random variables with a common distribution function F and density $f = F'$, and suppose that they are to be observed one at a time in sequence. Let

$$N = \min\{n \geqslant 2\colon X_n = \text{second largest of } X_1, \ldots, X_n\}$$

and let

$$M = \min\{n \geqslant 2\colon X_n = \text{second smallest of } X_1, \ldots, X_n\}$$

Which random variable—X_N, the first random variable which when observed is the second largest of those that have been seen, or X_M, the first one that on observation is the second smallest to have been seen—tends to be larger?

Solution: To calculate the probability density function of X_N, it is natural to condition on the value of N; so let us start by determining its probability mass function. Now, if we let

$$A_i = \{X_i \neq \text{second largest of } X_1, \ldots, X_i\}, \qquad i \geqslant 2$$

then, for $n \geqslant 2$,

$$P\{N = n\} = P\left(A_2 A_3 \cdots A_{n-1} A_n^c\right)$$

Since the X_i are independent and identically distributed it follows that, for any $m \geqslant 1$, knowing the rank ordering of the variables $X_1, \ldots, X_m$ yields no information about the set of m values $\{X_1, \ldots, X_m\}$. That is, for instance, knowing that $X_1 < X_2$ gives us no information about the values of $\min(X_1, X_2)$ or $\max(X_1, X_2)$. It follows from this that the events $A_i,\ i \geqslant 2$ are independent. Also, since X_i is equally likely to be the largest, or the second largest, ..., or the ith largest of $X_1, \ldots, X_i$ it follows that $P\{A_i\} = (i-1)/i,\ i \geqslant 2$. Therefore, we see that

$$P\{N = n\} = \frac{1}{2}\frac{2}{3}\frac{3}{4}\cdots\frac{n-2}{n-1}\frac{1}{n} = \frac{1}{n(n-1)}$$

Hence, conditioning on N yields that the probability density function of X_N is as follows:

$$f_{X_N}(x) = \sum_{n=2}^{\infty} \frac{1}{n(n-1)} f_{X_N|N}(x|n)$$

Now, since the ordering of the variables $X_1, \ldots, X_n$ is independent of the set of values $\{X_1, \ldots, X_n\}$, it follows that the event $\{N = n\}$ is independent of $\{X_1, \ldots, X_n\}$. From this, it follows that the conditional distribution of X_N given that $N = n$ is equal to the distribution of the second largest from a set of n random variables having distribution F. Thus, using the results of Example 2.37 concerning the density function of such a random variable, we obtain that

$$\begin{aligned}
f_{X_N}(x) &= \sum_{n=2}^{\infty} \frac{1}{n(n-1)} \frac{n!}{(n-2)!1!}(F(x))^{n-2} f(x)(1 - F(x)) \\
&= f(x)(1 - F(x)) \sum_{i=0}^{\infty} (F(x))^i \\
&= f(x)
\end{aligned}$$

Thus, rather surprisingly, X_N has the same distribution as X_1, namely, F. Also, if we now let $W_i = -X_i,\ i \geqslant 1$, then W_M will be the value of the first W_i, which on observation is the second largest of all those that have been seen. Hence, by the preceding, it follows that W_M has the same distribution as W_1. That is, $-X_M$ has the same distribution as $-X_1$, and so X_M also has distribution F! In other words, whether we stop at the first random variable that is the second largest of all those presently observed, or we stop at the first one that is the second smallest of all those presently observed, we will end up with a random variable having distribution F.

Whereas the preceding result is quite surprising, it is a special case of a general result known as *Ignatov's theorem*, which yields even more surprises.

For instance, for $k \geqslant 1$, let

$$N_k = \min\{n \geqslant k\colon X_n = k\text{th largest of } X_1, \ldots, X_n\}$$

Therefore, N_2 is what we previously called N, and X_{N_k} is the first random variable that upon observation is the kth largest of all those observed up to this point. It can then be shown by a similar argument as used in the preceding that X_{N_k} has distribution function F for all k (see Exercise 82 at the end of this chapter). In addition, it can be shown that the random variables $X_{N_k},\ k \geqslant 1$ are independent. (A statement and proof of Ignatov's theorem in the case of discrete random variables are given in Section 3.6.6.) ■

The use of conditioning can also result in a more computationally efficient solution than a direct calculation. This is illustrated by our next example.

Example 3.28 Consider n independent trials in which each trial results in one of the outcomes $1, \ldots, k$ with respective probabilities $p_1, \ldots, p_k$, $\sum_{i=1}^{k} p_i = 1$. Suppose further that $n > k$, and that we are interested in determining the probability that each outcome occurs at least once. If we let A_i denote the event that outcome i does not occur in any of the n trials, then the desired probability is $1 - P(\bigcup_{i=1}^{k} A_i)$, and it can be obtained by using the inclusion–exclusion theorem as follows:

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i) - \sum_i \sum_{j>i} P(A_i A_j) + \sum_i \sum_{j>i} \sum_{k>j} P(A_i A_j A_k) - \cdots + (-1)^{k+1} P(A_1 \cdots A_k)$$

where

$$P(A_i) = (1 - p_i)^n$$

$$P(A_i A_j) = (1 - p_i - p_j)^n, \qquad i < j$$

$$P(A_i A_j A_k) = (1 - p_i - p_j - p_k)^n, \qquad i < j < k$$

The difficulty with the preceding solution is that its computation requires the calculation of $2^k - 1$ terms, each of which is a quantity raised to the power n. The preceding solution is thus computationally inefficient when k is large. Let us now see how to make use of conditioning to obtain an efficient solution.

To begin, note that if we start by conditioning on N_k (the number of times that outcome k occurs) then when $N_k > 0$ the resulting conditional probability will equal the probability that all of the outcomes $1, \ldots, k-1$ occur at least once when $n - N_k$ trials are performed, and each results in outcome i with probability $p_i/\sum_{j=1}^{k-1} p_j$, $i = 1, \ldots, k-1$. We could then use a similar conditioning step on these terms.

To follow through on the preceding idea, let $A_{m,r}$, for $m \leqslant n,\ r \leqslant k$, denote the event that each of the outcomes $1, \ldots, r$ occurs at least once when m independent trials are performed, where each trial results in one of the outcomes $1, \ldots, r$ with respective probabilities $p_1/P_r, \ldots, p_r/P_r$, where $P_r = \sum_{j=1}^{r} p_j$. Let $P(m, r) = P(A_{m,r})$ and note that $P(n, k)$ is the desired probability. To obtain an expression for $P(m, r)$, condition on the number of times that outcome r occurs. This gives

$$
\begin{aligned}
P(m, r) &= \sum_{j=0}^{m} P\{A_{m,r} | r \text{ occurs } j \text{ times}\} \binom{m}{j} \left(\frac{p_r}{P_r}\right)^j \left(1 - \frac{p_r}{P_r}\right)^{m-j} \\
&= \sum_{j=1}^{m-r+1} P(m - j, r - 1) \binom{m}{j} \left(\frac{p_r}{P_r}\right)^j \left(1 - \frac{p_r}{P_r}\right)^{m-j}
\end{aligned}
$$

Starting with

$$
\begin{aligned}
P(m, 1) &= 1, \qquad \text{if } m \geqslant 1 \\
P(m, 1) &= 0, \qquad \text{if } m = 0
\end{aligned}
$$

we can use the preceding recursion to obtain the quantities $P(m, 2)$, $m = 2, \ldots, n - (k - 2)$, and then the quantities $P(m, 3)$, $m = 3, \ldots, n - (k - 3)$, and so on, up to $P(m, k - 1)$, $m = k - 1, \ldots, n - 1$. At this point we can then use the recursion to compute $P(n, k)$. It is not difficult to check that the amount of computation needed is a polynomial function of k, which will be much smaller than 2^k when k is large. ■

As noted previously, conditional expectations given that $Y = y$ are exactly the same as ordinary expectations except that all probabilities are computed conditional on the event that $Y = y$. As such, conditional expectations satisfy all the properties of ordinary expectations. For instance, the analog of

$$
E[X] = \begin{cases} \displaystyle\sum_w E[X|W = w] P\{W = w\}, & \text{if } W \text{ is discrete} \\ \displaystyle\int_w E[X|W = w] f_W(w)\, dw, & \text{if } W \text{ is continuous} \end{cases}
$$

is that

$$
E[X|Y = y] = \begin{cases} \displaystyle\sum_w E[X|W = w, Y = y] P\{W = w | Y = y\}, & \text{if } W \text{ is discrete} \\ \displaystyle\int_w E[X|W = w, Y = y] f_{W|Y}(w|y)\, dw, & \text{if } W \text{ is continuous} \end{cases}
$$

If $E[X|Y, W]$ is defined to be that function of Y and W that, when $Y = y$ and $W = w$, is equal to $E[X|Y = y, W = w]$, then the preceding can be written as

$$E[X|Y] = E\big[E[X|Y, W]\big|Y\big]$$

Example 3.29 An automobile insurance company classifies each of its policyholders as being of one of the types $i = 1, \ldots, k$. It supposes that the numbers of accidents that a type i policyholder has in successive years are independent Poisson random variables with mean λ_i, $i = 1, \ldots, k$. The probability that a newly insured policyholder is type i is p_i, $\sum_{i=1}^{k} p_i = 1$. Given that a policyholder had n accidents in her first year, what is the expected number that she has in her second year? What is the conditional probability that she has m accidents in her second year?

Solution: Let N_i denote the number of accidents the policyholder has in year i, $i = 1, 2$. To obtain $E[N_2|N_1 = n]$, condition on her risk type T.

$$\begin{aligned}
E[N_2|N_1 = n] &= \sum_{j=1}^{k} E[N_2|T = j, N_1 = n]P\{T = j|N_1 = n\} \\
&= \sum_{j=1}^{k} E[N_2|T = j]P\{T = j|N_1 = n\} \\
&= \sum_{j=1}^{k} \lambda_j P\{T = j|N_1 = n\} \\
&= \frac{\sum_{j=1}^{k} e^{-\lambda_j}\lambda_j^{n+1} p_j}{\sum_{j=1}^{k} e^{-\lambda_j}\lambda_j^{n} p_j}
\end{aligned}$$

where the final equality used that

$$\begin{aligned}
P\{T = j|N_1 = n\} &= \frac{P\{T = j, N_1 = n\}}{P\{N_1 = n\}} \\
&= \frac{P\{N_1 = n|T = j\}P\{T = j\}}{\sum_{j=1}^{k} P\{N_1 = n|T = j\}P\{T = j\}} \\
&= \frac{p_j e^{-\lambda_j}\lambda_j^{n}/n!}{\sum_{j=1}^{k} p_j e^{-\lambda_j}\lambda_j^{n}/n!}
\end{aligned}$$

The conditional probability that the policyholder has m accidents in year 2 given that she had n in year 1 can also be obtained by conditioning on her type.

$$\begin{aligned}
P\{N_2 = m|N_1 = n\} &= \sum_{j=1}^{k} P\{N_2 = m|T = j, N_1 = n\}P\{T = j|N_1 = n\} \\
&= \sum_{j=1}^{k} e^{-\lambda_j} \frac{\lambda_j^m}{m!} P\{T = j|N_1 = n\} \\
&= \frac{\sum_{j=1}^{k} e^{-2\lambda_j} \lambda_j^{m+n} p_j}{m! \sum_{j=1}^{k} e^{-\lambda_j} \lambda_j^n p_j}
\end{aligned}$$

Another way to calculate $P\{N_2 = m|N_1 = n\}$ is first to write

$$P\{N_2 = m|N_1 = n\} = \frac{P\{N_2 = m, N_1 = n\}}{P\{N_1 = n\}}$$

and then determine both the numerator and denominator by conditioning on T. This yields

$$\begin{aligned}
P\{N_2 = m|N_1 = n\} &= \frac{\sum_{j=1}^{k} P\{N_2 = m, N_1 = n|T = j\}p_j}{\sum_{j=1}^{k} P\{N_1 = n|T = j\}p_j} \\
&= \frac{\sum_{j=1}^{k} e^{-\lambda_j} \frac{\lambda_j^m}{m!} e^{-\lambda_j} \frac{\lambda_j^n}{n!} p_j}{\sum_{j=1}^{k} e^{-\lambda_j} \frac{\lambda_j^n}{n!} p_j} \\
&= \frac{\sum_{j=1}^{k} e^{-2\lambda_j} \lambda_j^{m+n} p_j}{m! \sum_{j=1}^{k} e^{-\lambda_j} \lambda_j^n p_j} \qquad \blacksquare
\end{aligned}$$

3.6. Some Applications

3.6.1. A List Model

Consider n elements—$e_1, e_2, \ldots, e_n$—that are initially arranged in some ordered list. At each unit of time a request is made for one of these elements—e_i being requested, independently of the past, with probability P_i. After being requested the element is then moved to the front of the list. That is, for instance, if the present ordering is $e_1,\ e_2,\ e_3,\ e_4$ and if e_3 is requested, then the next ordering is $e_3,\ e_1,\ e_2,\ e_4$.

We are interested in determining the expected position of the element requested after this process has been in operation for a long time. However, before computing this expectation, let us note two possible applications of this model. In the first we have a stack of reference books. At each unit of time a book is randomly selected and is then returned to the top of the stack. In the second application we have a computer receiving requests for elements stored in its memory. The request probabilities for the elements may not be known, so to reduce the average time it takes the computer to locate the element requested (which is proportional to the position of the requested element if the computer locates the element by starting at the beginning and then going down the list), the computer is programmed to replace the requested element at the beginning of the list.

To compute the expected position of the element requested, we start by conditioning on which element is selected. This yields

$$\begin{aligned}
&E[\text{position of element requested}]\\
&\quad=\sum_{i=1}^{n} E[\text{position}|e_i \text{ is selected}]P_i\\
&\quad=\sum_{i=1}^{n} E[\text{position of } e_i|e_i \text{ is selected}]P_i\\
&\quad=\sum_{i=1}^{n} E[\text{position of } e_i\,]P_i
\end{aligned} \tag{3.16}$$

where the final equality used that the position of e_i and the event that e_i is selected are independent because, regardless of its position, e_i is selected with probability P_i.

Now

$$\text{position of } e_i = 1 + \sum_{j\neq i} I_j$$

where

$$I_j = \begin{cases} 1, & \text{if } e_j \text{ precedes } e_i \\ 0, & \text{otherwise} \end{cases}$$

and so,

$$\begin{aligned}
E[\text{position of } e_i\,] &= 1 + \sum_{j\neq i} E[I_j]\\
&= 1 + \sum_{j\neq i} P\{e_j \text{ precedes } e_i\}
\end{aligned} \tag{3.17}$$

To compute $P\{e_j \text{ precedes } e_i\}$, note that e_j will precede e_i if the most recent request for either of them was for e_j. But given that a request is for either e_i or e_j, the probability that it is for e_j is

$$P\{e_j|e_i \text{ or } e_j\} = \frac{P_j}{P_i + P_j}$$

and, thus,

$$P\{e_j \text{ precedes } e_i\} = \frac{P_j}{P_i + P_j}$$

Hence from Equations (3.16) and (3.17) we see that

$$E\{\text{position of element requested}\} = 1 + \sum_{i=1}^{n} P_i \sum_{j \neq i} \frac{P_j}{P_i + P_j}$$

This list model will be further analyzed in Section 4.8, where we will assume a different reordering rule—namely, that the element requested is moved one closer to the front of the list as opposed to being moved to the front of the list as assumed here. We will show there that the average position of the requested element is less under the one-closer rule than it is under the front-of-the-line rule.

3.6.2. A Random Graph

A graph consists of a set V of elements called nodes and a set A of pairs of elements of V called arcs. A graph can be represented graphically by drawing circles for nodes and drawing lines between nodes i and j whenever (i, j) is an arc. For instance if $V = \{1, 2, 3, 4\}$ and $A = \{(1, 2), (1, 4), (2, 3), (1, 2), (3, 3)\}$, then we can represent this graph as shown in Figure 3.1. Note that the arcs have no direction (a graph in which the arcs are ordered pairs of nodes is called a directed graph); and that in the figure there are multiple arcs connecting nodes 1 and 2, and a self-arc (called a self-loop) from node 3 to itself.

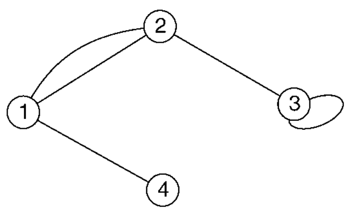

Figure 3.1. A graph.

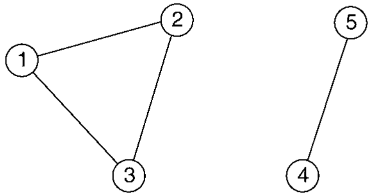

Figure 3.2. A disconnected graph.

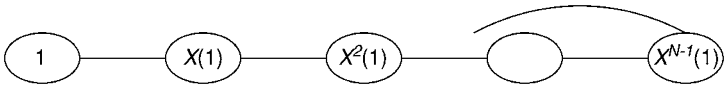

Figure 3.3.

We say that there exists a path from node i to node j, $i \neq j$, if there exists a sequence of nodes $i, i_1, \ldots, i_k, j$ such that $(i, i_1), (i_1, i_2), \ldots, (i_k, j)$ are all arcs. If there is a path between each of the $\binom{n}{2}$ distinct pair of nodes we say that the graph is *connected*. The graph in Figure 3.1 is connected but the graph in Figure 3.2 is not. Consider now the following graph where $V = \{1, 2, \ldots, n\}$ and $A = \{(i, X(i)),\ i = 1, \ldots, n\}$ where the $X(i)$ are independent random variables such that

$$P\{X(i) = j\} = \frac{1}{n}, \qquad j = 1, 2, \ldots, n$$

In other words from each node i we select at random one of the n nodes (including possibly the node i itself) and then join node i and the selected node with an arc. Such a graph is commonly referred to as a *random graph*.

We are interested in determining the probability that the random graph so obtained is connected. As a prelude, starting at some node—say, node 1—let us follow the sequence of nodes, 1, $X(1)$, $X^2(1), \ldots$, where $X^n(1) = X(X^{n-1}(1))$; and define N to equal the first k such that $X^k(1)$ is not a new node. In other words,

$$N = 1\text{st } k \text{ such that } X^k(1) \in \{1, X(1), \ldots, X^{k-1}(1)\}$$

We can represent this as shown in Figure 3.3 where the arc from $X^{N-1}(1)$ goes back to a node previously visited.

To obtain the probability that the graph is connected we first condition on N to obtain

$$P\{\text{graph is connected}\} = \sum_{k=1}^{n} P\{\text{connected}|N = k\}P\{N = k\} \tag{3.18}$$

Now given that $N = k$, the k nodes $1, X(1), \ldots, X^{k-1}(1)$ are connected to each other, and there are no other arcs emanating out of these nodes. In other words, if we regard these k nodes as being one supernode, the situation is the same as if we had one supernode and $n - k$ ordinary nodes with arcs emanating from the ordinary nodes—each arc going into the supernode with probability k/n. The solution in this situation is obtained from Lemma 3.2 by taking $r = n - k$.

Lemma 3.1 Given a random graph consisting of nodes $0, 1, \ldots, r$ and r arcs—namely, (i, Y_i), $i = 1, \ldots, r$, where

$$Y_i = \begin{cases} j & \text{with probability } \dfrac{1}{r+k}, \qquad j = 1, \ldots, r \\[2ex] 0 & \text{with probability } \dfrac{k}{r+k} \end{cases}$$

then

$$P\{\text{graph is connected}\} = \frac{k}{r+k}$$

[In other words, for the preceding graph there are $r + 1$ nodes—r ordinary nodes and one supernode. Out of each ordinary node an arc is chosen. The arc goes to the supernode with probability $k/(r+k)$ and to each of the ordinary ones with probability $1/(r+k)$. There is no arc emanating out of the supernode.]

Proof The proof is by induction on r. As it is true when $r = 1$ for any k, assume it true for all values less than r. Now in the case under consideration, let us first condition on the number of arcs (j, Y_j) for which $Y_j = 0$. This yields

$$P\{\text{connected}\}$$
$$= \sum_{i=0}^{r} P\{\text{connected} | i \text{ of the } Y_j = 0\} \binom{r}{i} \left(\frac{k}{r+k}\right)^i \left(\frac{r}{r+k}\right)^{r-i} \tag{3.19}$$

Now given that exactly i of the arcs are into the supernode (see Figure 3.4), the situation for the remaining $r - i$ arcs which do not go into the supernode is the same as if we had $r - i$ ordinary nodes and one supernode with an arc going out of each of the ordinary nodes—into the supernode with probability i/r and into each ordinary node with probability $1/r$. But by the induction hypothesis the probability that this would lead to a connected graph is i/r.

Hence,

$$P\{\text{connected} | i \text{ of the } Y_j = 0\} = \frac{i}{r}$$

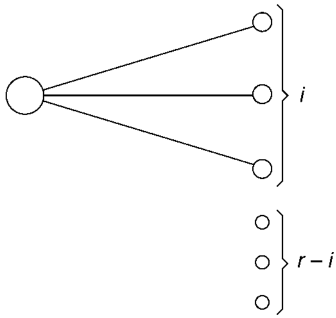

Figure 3.4. The situation given that i of the r arcs are into the supernode.

and from Equation (3.19)

$$\begin{aligned}P\{\text{connected}\} &= \sum_{i=0}^{r} \frac{i}{r}\binom{r}{i}\left(\frac{k}{r+k}\right)^{i}\left(\frac{r}{r+k}\right)^{r-i} \\ &= \frac{1}{r}E\left[\text{binomial}\left(r,\ \frac{k}{r+k}\right)\right] \\ &= \frac{k}{r+k}\end{aligned}$$

which completes the proof of the lemma. ■

Hence as the situation given $N=k$ is exactly as described by Lemma 3.2 when $r=n-k$, we see that, for the original graph,

$$P\{\text{graph is connected}|N=k\} = \frac{k}{n}$$

and, from Equation (3.18),

$$P\{\text{graph is connected}\} = \frac{E(N)}{n} \tag{3.20}$$

To compute $E(N)$ we use the identity

$$E(N) = \sum_{i=1}^{\infty} P\{N \geqslant i\}$$

which can be proved by defining indicator variables I_i, $i \geqslant 1$, by

$$I_i = \begin{cases} 1, & \text{if } i \leqslant N \\ 0, & \text{if } i > N \end{cases}$$

Hence,

$$N = \sum_{i=1}^{\infty} I_i$$

and so

$$\begin{aligned} E(N) &= E\left[\sum_{i=1}^{\infty} I_i\right] \\ &= \sum_{i=1}^{\infty} E[I_i] \\ &= \sum_{i=1}^{\infty} P\{N \geqslant i\} \end{aligned} \tag{3.21}$$

Now the event $\{N \geqslant i\}$ occurs if the nodes $1, X(1), \ldots, X^{i-1}(1)$ are all distinct. Hence,

$$\begin{aligned} P\{N \geqslant i\} &= \frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-i+1)}{n} \\ &= \frac{(n-1)!}{(n-i)!n^{i-1}} \end{aligned}$$

and so, from Equations (3.20) and (3.21),

$$\begin{aligned} P\{\text{graph is connected}\} &= (n-1)!\sum_{i=1}^{n} \frac{1}{(n-i)!n^i} \\ &= \frac{(n-1)!}{n^n}\sum_{j=0}^{n-1}\frac{n^j}{j!} \qquad (\text{by } j = n - i) \end{aligned} \tag{3.22}$$

We can also use Equation (3.22) to obtain a simple approximate expression for the probability that the graph is connected when n is large. To do so, we first note that if X is a Poisson random variable with mean n, then

$$P\{X < n\} = e^{-n}\sum_{j=0}^{n-1}\frac{n^j}{j!}$$

Since a Poisson random variable with mean n can be regarded as being the sum of n independent Poisson random variables each with mean 1, it follows from the central limit theorem that for n large such a random variable has approximately a normal distribution and as such has probability $\frac{1}{2}$ of being less than its mean. That is, for n large

$$P\{X < n\} \approx \tfrac{1}{2}$$

and so for n large,

$$\sum_{j=0}^{n-1} \frac{n^j}{j!} \approx \frac{e^n}{2}$$

Hence from Equation (3.22), for n large,

$$P\{\text{graph is connected}\} \approx \frac{e^n (n-1)!}{2n^n}$$

By employing an approximation due to Stirling which states that for n large

$$n! \approx n^{n+1/2} e^{-n} \sqrt{2\pi}$$

We see that, for n large,

$$P\{\text{graph is connected}\} \approx \sqrt{\frac{\pi}{2(n-1)}}\, e \left(\frac{n-1}{n}\right)^n$$

and as

$$\lim_{n\to\infty} \left(\frac{n-1}{n}\right)^n = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = e^{-1}$$

We see that, for n large,

$$P\{\text{graph is connected}\} \approx \sqrt{\frac{\pi}{2(n-1)}}$$

Now a graph is said to consist of r connected components if its nodes can be partitioned into r subsets so that each of the subsets is connected and there are no arcs between nodes in different subsets. For instance, the graph in Figure 3.5 consists of three connected components—namely, $\{1, 2, 3\}$, $\{4, 5\}$, and $\{6\}$. Let C denote the number of connected components of our random graph and let

$$P_n(i) = P\{C = i\}$$

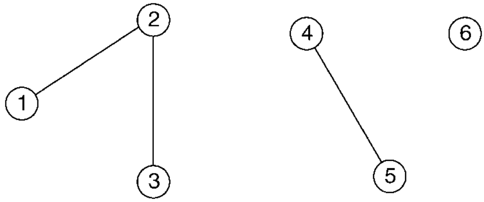

Figure 3.5. A graph having three connected components.

where we use the notation $P_n(i)$ to make explicit the dependence on n, the number of nodes. Since a connected graph is by definition a graph consisting of exactly one component, from Equation (3.22) we have

$$\begin{aligned} P_n(1) &= P\{C = 1\} \\ &= \frac{(n-1)!}{n^n} \sum_{j=0}^{n-1} \frac{n^j}{j!} \end{aligned} \tag{3.23}$$

To obtain $P_n(2)$, the probability of exactly two components, let us first fix attention on some particular node—say, node 1. In order that a given set of $k-1$ other nodes—say, nodes $2, \ldots, k$—will along with node 1 constitute one connected component, and the remaining $n-k$ a second connected component, we must have

(i) $X(i) \in \{1, 2, \ldots, k\}$, for all $i = 1, \ldots, k$.
(ii) $X(i) \in \{k+1, \ldots, n\}$, for all $i = k+1, \ldots, n$.
(iii) The nodes $1, 2, \ldots, k$ form a connected subgraph.
(iv) The nodes $k+1, \ldots, n$ form a connected subgraph.

The probability of the preceding occurring is clearly

$$\left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} P_k(1) P_{n-k}(1)$$

and because there are $\binom{n-1}{k-1}$ ways of choosing a set of $k-1$ nodes from the nodes 2 through n, we have

$$P_n(2) = \sum_{k=1}^{n-1} \binom{n-1}{k-1} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} P_k(1) P_{n-k}(1)$$

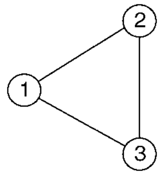

Figure 3.6. A cycle.

and so $P_n(2)$ can be computed from Equation (3.23). In general, the recursive formula for $P_n(i)$ is given by

$$P_n(i) = \sum_{k=1}^{n-i+1} \binom{n-1}{k-1} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} P_k(1)P_{n-k}(i-1)$$

To compute $E[C]$, the expected number of connected components, first note that every connected component of our random graph must contain exactly one cycle [a cycle is a set of arcs of the form $(i, i_1),\ (i_1, i_2), \ldots,\ (i_{k-1}, i_k),\ (i_k, i)$ for distinct nodes $i, i_1, \ldots, i_k$]. For example, Figure 3.6 depicts a cycle.

The fact that every connected component of our random graph must contain exactly one cycle is most easily proved by noting that if the connected component consists of r nodes, then it must also have r arcs and, hence, must contain exactly one cycle (why?). Thus, we see that

$$\begin{aligned}
E[C] &= E[\text{number of cycles}] \\
&= E\left[\sum_S I(S)\right] \\
&= \sum_S E[I(S)]
\end{aligned}$$

where the sum is over all subsets $S \subset \{1, 2, \ldots, n\}$ and

$$I(S) = \begin{cases} 1, & \text{if the nodes in } S \text{ are all the nodes of a cycle} \\ 0, & \text{otherwise} \end{cases}$$

Now, if S consists of k nodes, say $1, \ldots, k$, then

$$\begin{aligned}
E[I(S)] &= P\{1, X(1), \ldots, X^{k-1}(1) \text{ are all distinct and contained in} \\
&\qquad 1, \ldots, k \text{ and } X^k(1) = 1\} \\
&= \frac{k-1}{n}\frac{k-2}{n} \cdots \frac{1}{n}\frac{1}{n} = \frac{(k-1)!}{n^k}
\end{aligned}$$

Hence, because there are $\binom{n}{k}$ subsets of size k we see that

$$E[C] = \sum_{k=1}^{n} \binom{n}{k} \frac{(k-1)!}{n^k}$$

3.6.3. Uniform Priors, Polya's Urn Model, and Bose–Einstein Statistics

Suppose that n independent trials, each of which is a success with probability p, are performed. If we let X denote the total number of successes, then X is a binomial random variable such that

$$P\{X = k|p\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n$$

However, let us now suppose that whereas the trials all have the same success probability p, its value is not predetermined but is chosen according to a uniform distribution on (0, 1). (For instance, a coin may be chosen at random from a huge bin of coins representing a uniform spread over all possible values of p, the coin's probability of coming up heads. The chosen coin is then flipped n times.) In this case, by conditioning on the actual value of p, we have that

$$\begin{aligned} P\{X = k\} &= \int_0^1 P\{X = k|p\} f(p)\, dp \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\, dp \end{aligned}$$

Now, it can be shown that

$$\int_0^1 p^k (1-p)^{n-k}\, dp = \frac{k!(n-k)!}{(n+1)!} \tag{3.24}$$

and thus

$$\begin{aligned} P\{X = k\} &= \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} \\ &= \frac{1}{n+1}, \qquad k = 0, 1, \ldots, n \end{aligned} \tag{3.25}$$

In other words, each of the $n + 1$ possible values of X is equally likely.

As an alternate way of describing the preceding experiment, let us compute the conditional probability that the $(r+1)$st trial will result in a success given a total of k successes (and $r-k$ failures) in the first r trials.

$$
\begin{aligned}
P\{&(r+1)\text{st trial is a success} | k \text{ successes in first } r\} \\
&= \frac{P\{(r+1)\text{st is a success},\ k \text{ successes in first } r \text{ trials}\}}{P\{k \text{ successes in first } r \text{ trials}\}} \\
&= \frac{\int_0^1 P\{(r+1)\text{st is a success},\ k \text{ in first } r | p\}\, dp}{1/(r+1)} \\
&= (r+1) \int_0^1 \binom{r}{k} p^{k+1} (1-p)^{r-k}\, dp \\
&= (r+1) \binom{r}{k} \frac{(k+1)!(r-k)!}{(r+2)!} \qquad \text{by Equation (3.24)} \\
&= \frac{k+1}{r+2} \qquad (3.26)
\end{aligned}
$$

That is, if the first r trials result in k successes, then the next trial will be a success with probability $(k+1)/(r+2)$.

It follows from Equation (3.26) that an alternative description of the stochastic process of the successive outcomes of the trials can be described as follows: There is an urn which initially contains one white and one black ball. At each stage a ball is randomly drawn and is then replaced along with another ball of the same color. Thus, for instance, if of the first r balls drawn, k were white, then the urn at the time of the $(r+1)$th draw would consist of $k+1$ white and $r-k+1$ black, and thus the next ball would be white with probability $(k+1)/(r+2)$. If we identify the drawing of a white ball with a successful trial, then we see that this yields an alternate description of the original model. This latter urn model is called *Polya's urn model*.

Remarks (i) In the special case when $k = r$, Equation (3.26) is sometimes called Laplace's rule of succession, after the French mathematician Pierre de Laplace. In Laplace's era, this "rule" provoked much controversy, for people attempted to employ it in diverse situations where its validity was questionable. For instance, it was used to justify such propositions as "If you have dined twice at a restaurant and both meals were good, then the next meal also will be good with probability $\frac{3}{4}$," and "Since the sun has risen the past 1,826,213 days, so will it rise tomorrow with probability 1,826,214/1,826,215." The trouble with such claims resides in the fact that it is not at all clear the situation they are describing can be modeled as consisting of independent trials having a common probability of success which is itself uniformly chosen.

(ii) In the original description of the experiment, we referred to the successive trials as being independent, and in fact they are independent when the success probability is known. However, when p is regarded as a random variable, the successive outcomes are no longer independent because knowing whether an outcome is a success or not gives us some information about p, which in turn yields information about the other outcomes.

The preceding can be generalized to situations in which each trial has more than two possible outcomes. Suppose that n independent trials, each resulting in one of m possible outcomes $1, \ldots, m$, with respective probabilities $p_1, \ldots, p_m$ are performed. If we let X_i denote the number of type i outcomes that result in the n trials, $i = 1, \ldots, m$, then the vector $X_1, \ldots, X_m$ will have the multinomial distribution given by

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m | \mathbf{p}\} = \frac{n!}{x_1! \cdots x_m!} p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}$$

where $x_1, \ldots, x_m$ is any vector of nonnegative integers that sum to n. Now let us suppose that the vector $\mathbf{p} = (p_1, \ldots, p_m)$ is not specified, but instead is chosen by a "uniform" distribution. Such a distribution would be of the form

$$f(p_1, \ldots, p_m) = \begin{cases} c, & 0 \leqslant p_i \leqslant 1, i = 1, \ldots, m, \ \sum_1^m p_i = 1 \\ 0, & \text{otherwise} \end{cases}$$

The preceding multivariate distribution is a special case of what is known as the *Dirichlet distribution*, and it is not difficult to show, using the fact that the distribution must integrate to 1, that $c = (m-1)!$.

The unconditional distribution of the vector $\mathbf{X}$ is given by

$$P\{X_1 = x_1, \ldots, X_m = x_m\} = \iint \cdots \int P\{X_1 = x_1, \ldots, X_m = x_m | p_1, \ldots, p_m\}$$
$$\times f(p_1, \ldots, p_m)\, dp_1 \cdots dp_m = \frac{(m-1)!n!}{x_1! \cdots x_m!} \underset{\substack{0 \leqslant p_i \leqslant 1 \\ \sum_1^m p_i = 1}}{\iint \cdots \int} p_1^{x_1} \cdots p_m^{x_m}\, dp_1 \cdots dp_m$$

Now it can be shown that

$$\underset{\substack{0 \leqslant p_i \leqslant 1 \\ \sum_1^m p_i = 1}}{\iint \cdots \int} p_1^{x_1} \cdots p_m^{x_m}\, dp_1 \cdots dp_m = \frac{x_1! \cdots x_m!}{\left(\sum_1^m x_i + m - 1\right)!} \tag{3.27}$$

and thus, using the fact that $\sum_1^m x_i = n$, we see that

$$P\{X_1 = x_1, \ldots, X_m = x_m\} = \frac{n!(m-1)!}{(n+m-1)!} = \binom{n+m-1}{m-1}^{-1} \tag{3.28}$$

Hence, all of the $\binom{n+m-1}{m-1}$ possible outcomes [there are $\binom{n+m-1}{m-1}$ possible non-negative integer valued solutions of $x_1 + \cdots + x_m = n$] of the vector $(X_1, \ldots, X_m)$ are equally likely. The probability distribution given by Equation (3.28) is sometimes called the *Bose–Einstein distribution*.

To obtain an alternative description of the foregoing, let us compute the conditional probability that the $(n+1)$st outcome is of type j if the first n trials have resulted in x_i type i outcomes, $i = 1, \ldots, m$, $\sum_1^m x_i = n$. This is given by

$$\begin{aligned} &P\{(n+1)\text{st is } j | x_i \text{ type } i \text{ in first } n,\ i = 1, \ldots, m\} \\ &= \frac{P\{(n+1)\text{st is } j,\ x_i \text{ type } i \text{ in first } n,\ i = 1, \ldots, m\}}{P\{x_i \text{ type } i \text{ in first } n,\ i = 1, \ldots, m\}} \\ &= \frac{\dfrac{n!(m-1)!}{x_1! \cdots x_m!} \displaystyle\iint \cdots \int p_1^{x_1} \cdots p_j^{x_j+1} \cdots p_m^{x_m}\, dp_1 \cdots dp_m}{\dbinom{n+m-1}{m-1}^{-1}} \end{aligned}$$

where the numerator is obtained by conditioning on the **p** vector and the denominator is obtained by using Equation (3.28). By Equation (3.27), we have that

$$\begin{aligned} &P\{(n+1)\text{st is } j | x_i \text{ type } i \text{ in first } n,\ i = 1, \ldots, m\} \\ &= \frac{\dfrac{(x_j+1)n!(m-1)!}{(n+m)!}}{\dfrac{(m-1)!n!}{(n+m-1)!}} \\ &= \frac{x_j+1}{n+m} \end{aligned} \tag{3.29}$$

Using Equation (3.29), we can now present an urn model description of the stochastic process of successive outcomes. Namely, consider an urn that initially contains one of each of m types of balls. Balls are then randomly drawn and are replaced along with another of the same type. Hence, if in the first n drawings there have been a total of x_j type j balls drawn, then the urn immediately before the $(n+1)$st draw will contain x_j+1 type j balls out of a total of $m+n$, and so the probability of a type j on the $(n+1)$st draw will be given by Equation (3.29).

Remark Consider a situation where n particles are to be distributed at random among m possible regions; and suppose that the regions appear, at least before the experiment, to have the same physical characteristics. It would thus seem that the most likely distribution for the number of particles that fall into each of the regions is the multinomial distribution with $p_i \equiv 1/m$. (This, of course, would correspond to each particle, independent of the others, being equally likely to fall in any of the m regions.) Physicists studying how particles distribute themselves observed the behavior of such particles as photons and atoms containing an even number of elementary particles. However, when they studied the resulting data, they were amazed to discover that the observed frequencies did not follow the multinomial distribution but rather seemed to follow the Bose–Einstein distribution. They were amazed because they could not imagine a physical model for the distribution of particles which would result in all possible outcomes being equally likely. (For instance, if 10 particles are to distribute themselves between two regions, it hardly seems reasonable that it is just as likely that both regions will contain 5 particles as it is that all 10 will fall in region 1 or that all 10 will fall in region 2.)

However, from the results of this section we now have a better understanding of the cause of the physicists' dilemma. In fact, two possible hypotheses present themselves. First, it may be that the data gathered by the physicists were actually obtained under a variety of different situations, each having its own characteristic $\mathbf{p}$ vector which gave rise to a uniform spread over all possible $\mathbf{p}$ vectors. A second possibility (suggested by the urn model interpretation) is that the particles select their regions sequentially and a given particle's probability of falling in a region is roughly proportional to the fraction of the landed particles that are in that region. (In other words, the particles presently in a region provide an "attractive" force on elements that have not yet landed.)

3.6.4. Mean Time for Patterns

Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a sequence of independent and identically distributed discrete random variables such that

$$p_i = P\{X_j = i\}$$

For a given subsequence, or *pattern*, $i_1, \ldots, i_n$ let $T = T(i_1, \ldots, i_n)$ denote the number of random variables that we need to observe until the pattern appears. For instance, if the subsequence of interest is $3, 5, 1$ and the sequence is $\mathbf{X} = (5, 3, 1, 3, 5, 3, 5, 1, 6, 2, \ldots)$ then $T = 8$. We want to determine $E[T]$.

To begin, let us consider whether the pattern has an overlap, where we say that the pattern $i_1, i_2, \ldots, i_n$ has an overlap if for some k, $1 \leqslant k < n$, the sequence of its final k elements is the same as that of its first k elements. That is, it has an overlap if for some $1 \leqslant k < n$,

$$(i_{n-k+1}, \ldots, i_n) = (i_1, \ldots, i_k), \qquad k < n$$

For instance, the pattern 3, 5, 1 has no overlaps, whereas the pattern 3, 3, 3 does.

Case 1: The pattern has no overlaps.
In this case we will argue that T will equal $j + n$ if and only if the pattern does not occur within the first j values, and the next n values are $i_1, \ldots, i_n$. That is,

$$T = j + n \Leftrightarrow \{T > j,\ (X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)\} \tag{3.30}$$

To verify (3.30), note first that $T = j + n$ clearly implies both that $T > j$ and that $(X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)$. On the other hand, suppose that

$$T > j \quad \text{and} \quad (X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n) \tag{3.31}$$

Let $k < n$. Because $(i_1, \ldots, i_k) \neq (i_{n-k+1}, \ldots, i_n)$, it follows that $T \neq j + k$. But (3.31) implies that $T \leqslant j + n$, so we can conclude that $T = j + n$. Thus we have verified (3.30).

Using (3.30), we see that

$$P\{T = j + n\} = P\{T > j,\ (X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)\}$$

However, whether $T > j$ is determined by the values $X_1, \ldots, X_j$, and is thus independent of $X_{j+1}, \ldots, X_{j+n}$. Consequently,

$$\begin{aligned} P\{T = j + n\} &= P\{T > j\}P\{(X_{j+1}, \ldots, X_{j+n}) = (i_1, \ldots, i_n)\} \\ &= P\{T > j\}p \end{aligned}$$

where

$$p = p_{i_1} p_{i_2} \cdots p_{i_n}$$

Summing both sides of the preceding over all j yields

$$1 = \sum_{j=0}^{\infty} P\{T = j + n\} = p \sum_{j=0}^{\infty} P\{T > j\} = pE[T]$$

or

$$E[T] = \frac{1}{p}$$

Case 2: The pattern has overlaps.
For patterns having overlaps there is a simple trick that will enable us to obtain $E[T]$ by making use of the result for nonoverlapping patterns. To make the analy-

sis more transparent, consider a specific pattern, say $\mathbf{P} = (3, 5, 1, 3, 5)$. Let x be a value that does not appear in the pattern, and let T_x denote the time until the pattern $\mathbf{P}_x = (3, 5, 1, 3, 5, x)$ appears. That is, T_x is the time of occurrence of the new pattern that puts x at the end of the original pattern. Because x did not appear in the original pattern it follows that the new pattern has no overlaps; thus,

$$E[T_x] = \frac{1}{p_x p}$$

where $p = \prod_{j=1}^{n} p_{i_j} = p_3^2 p_5^2 p_1$. Because the new pattern can occur only after the original one, write

$$T_x = T + A$$

where T is the time at which the pattern $\mathbf{P} = (3, 5, 1, 3, 5)$ occurs, and A is the additional time after the occurrence of the pattern $\mathbf{P}$ until $\mathbf{P}_x$ occurs. Also, let $E[T_x|i_1, \ldots i_r]$ denote the expected additional time after time r until the pattern $\mathbf{P}_x$ appears given that the first r data values are $i_1, \ldots, i_r$. Conditioning on X, the next data value after the occurrence of the pattern $(3, 5, 1, 3, 5)$, gives that

$$E[A|X = i] = \begin{cases} 1 + E[T_x|3, 5, 1], & \text{if } i = 1 \\ 1 + E[T_x|3], & \text{if } i = 3 \\ 1, & \text{if } i = x \\ 1 + E[T_x], & \text{if } i \neq 1, 3, x \end{cases}$$

Therefore,

$$\begin{aligned} E[T_x] &= E[T] + E[A] \\ &= E[T] + 1 + E[T_x|3, 5, 1]p_1 + E[T_x|3]p_3 + E[T_x](1 - p_1 - p_3 - p_x) \end{aligned} \tag{3.32}$$

But

$$E[T_x] = E[T(3, 5, 1)] + E[T_x|3, 5, 1]$$

giving that

$$E[T_x|3, 5, 1] = E[T_x] - E[T(3, 5, 1)]$$

Similarly,

$$E[T_x|3] = E[T_x] - E[T(3)]$$

Substituting back into Equation (3.32) gives

$$p_x E[T_x] = E[T] + 1 - p_1 E[T(3, 5, 1)] - p_3 E[T(3)]$$

But, by the result in the nonoverlapping case,

$$E[T(3,5,1)] = \frac{1}{p_3 p_5 p_1}, \qquad E[T(3)] = \frac{1}{p_3}$$

yielding the result

$$E[T] = p_x E[T_x] + \frac{1}{p_3 p_5} = \frac{1}{p} + \frac{1}{p_3 p_5}$$

For another illustration of the technique, let us reconsider Example 3.14, which is concerned with finding the expected time until n consecutive successes occur in independent Bernoulli trials. That is, we want $E[T]$, when the pattern is $\mathbf{P} = (1, 1, \dots, 1)$. Then, with $x \neq 1$ we consider the nonoverlapping pattern $\mathbf{P}_x = (1, \dots, 1, x)$, and let T_x be its occurrence time. With A and X as previously defined, we have that

$$E[A|X=i] = \begin{cases} 1 + E[A], & \text{if } i = 1 \\ 1, & \text{if } i = x \\ 1 + E[T_x], & \text{if } i \neq 1, x \end{cases}$$

Therefore,

$$E[A] = 1 + E[A]p_1 + E[T_x](1 - p_1 - p_x)$$

or

$$E[A] = \frac{1}{1-p_1} + E[T_x]\frac{1-p_1-p_x}{1-p_1}$$

Consequently,

$$\begin{aligned} E[T] &= E[T_x] - E[A] \\ &= \frac{p_x E[T_x] - 1}{1 - p_1} \\ &= \frac{(1/p_1)^n - 1}{1 - p_1} \end{aligned}$$

where the final equality used that $E[T_x] = \frac{1}{p_1^n p_x}$.

The mean occurrence time of any overlapping pattern $\mathbf{P} = (i_1, \dots, i_n)$ can be obtained by the preceding method. Namely, let T_x be the time until the nonoverlapping pattern $\mathbf{P}_x = (i_1, \dots, i_n, x)$ occurs; then use the identity

$$E[T_x] = E[T] + E[A]$$

to relate $E[T]$ and $E[T_x] = \frac{1}{p\,p_x}$; then condition on the next data value after **P** occurs to obtain an expression for $E[A]$ in terms of quantities of the form

$$E[T_x|i_1,\ldots,i_r] = E[T_x] - E[T(i_1,\ldots,i_r)]$$

If $(i_1,\ldots,i_r)$ is nonoverlapping, use the nonoverlapping result to obtain $E[T(i_1,\ldots,i_r)]$; otherwise, repeat the process on the subpattern $(i_1,\ldots,i_r)$.

Remark We can utilize the preceding technique even when the pattern $i_1,\ldots,i_n$ includes all the distinct data values. For instance, in coin tossing the pattern of interest might be h, t, h. Even in such cases, we should let x be a data value that is not in the pattern and use the preceding technique (even though $p_x = 0$). Because p_x will appear only in the final answer in the expression $p_x E[T_x] = \frac{p_x}{p_x p}$, by interpreting this fraction as $1/p$ we obtain the correct answer. (A rigorous approach, yielding the same result, would be to reduce one of the positive p_i by ϵ, take $p_x = \epsilon$, solve for $E[T]$, and then let ϵ go to 0.) ■

3.6.5. The *k*-Record Values of Discrete Random Variables

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables whose set of possible values is the positive integers, and let $P\{X = j\}$, $j \geqslant 1$, denote their common probability mass function. Suppose that these random variables are observed in sequence, and say that X_n is a *k-record value* if

$$X_i \geqslant X_n \qquad \text{for exactly } k \text{ of the values } i,\ i = 1,\ldots,n$$

That is, the nth value in the sequence is a k-record value if exactly k of the first n values (including X_n) are at least as large as it. Let $\mathbf{R}_k$ denote the ordered set of k-record values.

It is a rather surprising result that not only do the sequences of k-record values have *the same probability distributions* for all k, these sequences are also *independent* of each other. This result is known as Ignatov's theorem.

Ignatov's Theorem $\mathbf{R}_k$, $k \geqslant 1$, are independent and identically distributed random vectors.

Proof Define a series of subsequences of the data sequence $X_1, X_2, \ldots$ by letting the ith subsequence consist of all data values that are at least as large as i, $i \geqslant 1$. For instance, if the data sequence is

$$2, 5, 1, 6, 9, 8, 3, 4, 1, 5, 7, 8, 2, 1, 3, 4, 2, 5, 6, 1, \ldots$$

then the subsequences are as follows:

$$\begin{array}{ll} \geqslant 1: & 2,5,1,6,9,8,3,4,1,5,7,8,2,1,3,4,2,5,6,1,\ldots \\ \geqslant 2: & 2,5,6,9,8,3,4,5,7,8,2,3,4,2,5,6,\ldots \\ \geqslant 3: & 5,6,9,8,3,4,5,7,8,3,4,5,6,\ldots \end{array}$$

and so on.

Let X^i_j be the jth element of subsequence i. That is, X^i_j is the jth data value that is at least as large as i. An important observation is that i is a k-record value if and only if $X^i_k = i$. That is, i will be a k-record value if and only if the kth value to be at least as large as i is equal to i. (For instance, for the preceding data, since the fifth value to be at least as large as 3 is equal to 3 it follows that 3 is a five-record value.) Now, it is not difficult to see that, independent of which values in the first subsequence are equal to 1, the values in the second subsequence are independent and identically distributed according to the mass function

$$P\{\text{value in second subsequence} = j\} = P\{X = j | X \geqslant 2\}, \qquad j \geqslant 2$$

Similarly, independent of which values in the first subsequence are equal to 1 and which values in the second subsequence are equal to 2, the values in the third subsequence are independent and identically distributed according to the mass function

$$P\{\text{value in third subsequence} = j\} = P\{X = j | X \geqslant 3\}, \qquad j \geqslant 3$$

and so on. It therefore follows that the events $\{X^i_j = i\}$, $i \geqslant 1$, $j \geqslant 1$, are independent and

$$P\{i \text{ is a } k\text{-record value}\} = P\{X^i_k = i\} = P\{X = i | X \geqslant i\}$$

It now follows from the independence of the events $\{X^i_k = i\}$, $i \geqslant 1$, and the fact that $P\{\text{i is a } k\text{-record value}\}$ does not depend on k, that $\mathbf{R}_k$ has the same distribution for all $k \geqslant 1$. In addition, it follows from the independence of the events $\{X^i_k = 1\}$, that the random vectors $\mathbf{R}_k$, $k \geqslant 1$, are also independent. ■

Suppose now that the X_i, $i \geqslant 1$ are independent finite-valued random variables with probability mass function

$$p_i = P\{X = i\}, \qquad i = 1, \ldots, m$$

and let

$$T = \min\{n: \ X_i \geqslant X_n \text{ for exactly } k \text{ of the values } i, \ i = 1, \ldots, n\}$$

denote the first *k-record index*. We will now determine its mean.

Proposition 3.3 Let $\lambda_i = p_i / \sum_{j=i}^{m} p_j,\ i = 1, \ldots, m$. Then

$$E[T] = k + (k-1)\sum_{i=1}^{m-1} \lambda_i$$

Proof To begin, suppose that the observed random variables $X_1\ X_2, \ldots$ take on one of the values $i, i+1, \ldots, m$ with respective probabilities

$$P\{X = j\} = \frac{p_j}{p_i + \cdots + p_m}, \qquad j = i, \ldots, m$$

Let T_i denote the first k-record index when the observed data have the preceding mass function, and note that since the each data value is at least i it follows that the k-record value will equal i, and T_i will equal k, if $X_k = i$. As a result,

$$E[T_i | X_k = i] = k$$

On the other hand, if $X_k > i$ then the k-record value will exceed i, and so all data values equal to i can be disregarded when searching for the k-record value. In addition, since each data value greater than i will have probability mass function

$$P\{X = j | X > i\} = \frac{p_j}{p_{i+1} + \cdots + p_m}, \qquad j = i+1, \ldots, m$$

it follows that the total number of data values greater than i that need be observed until a k-record value appears has the same distribution as T_{i+1}. Hence,

$$E[T_i | X_k > i] = E[T_{i+1} + N_i | X_k > i]$$

where T_{i+1} is the total number of variables greater than i that we need observe to obtain a k-record, and N_i is the number of values equal to i that are observed in that time. Now, given that $X_k > i$ and that $T_{i+1} = n$ $(n \geqslant k)$ it follows that the time to observe T_{i+1} values greater than i has the same distribution as the number of trials to obtain n successes given that trial k is a success and that each trial is independently a success with probability $1 - p_i / \sum_{j \geqslant i} p_j = 1 - \lambda_i$. Thus, since the number of trials needed to obtain a success is a geometric random variable with mean $1/(1-\lambda_i)$, we see that

$$E[T_i | T_{i+1}, X_k > i] = 1 + \frac{T_{i+1} - 1}{1 - \lambda_i} = \frac{T_{i+1} - \lambda_i}{1 - \lambda_i}$$

Taking expectations gives that

$$E[T_i | X_k > i] = E\left[\frac{T_{i+1} - \lambda_i}{1 - \lambda_i} \middle| X_k > i\right] = \frac{E[T_{i+1}] - \lambda_i}{1 - \lambda_i}$$

Thus, upon conditioning on whether $X_k = i$, we obtain

$$\begin{aligned} E[T_i] &= E[T_i|X_k = i]\lambda_i + E[T_i|X_k > i](1 - \lambda_i) \\ &= (k-1)\lambda_i + E[T_{i+1}] \end{aligned}$$

Starting with $E[T_m] = k$, the preceding gives that

$$\begin{aligned} E[T_{m-1}] &= (k-1)\lambda_{m-1} + k \\ E[T_{m-2}] &= (k-1)\lambda_{m-2} + (k-1)\lambda_{m-1} + k \\ &= (k-1)\sum_{j=m-2}^{m-1} \lambda_j + k \\ E[T_{m-3}] &= (k-1)\lambda_{m-3} + (k-1)\sum_{j=m-2}^{m-1} \lambda_j + k \\ &= (k-1)\sum_{j=m-3}^{m-1} \lambda_j + k \end{aligned}$$

In general,

$$E[T_i] = (k-1)\sum_{j=i}^{m-1} \lambda_j + k$$

and the result follows since $T = T_1$. ■

3.7. An Identity for Compound Random Variables

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, and let $S_n = \sum_{i=1}^n X_i$ be the sum of the first n of them, $n \geqslant 0$, where $S_0 = 0$. Recall that if N is a nonnegative integer valued random variable that is independent of the sequence $X_1, X_2, \ldots$ then

$$S_N = \sum_{i=1}^{N} X_i$$

is said to be a *compound random variable*, with the distribution of N called the *compounding distribution*. In this subsection we will first derive an identity involving such random variables. We will then specialize to where the X_i are positive integer valued random variables, prove a corollary of the identity, and then use this corollary to develop a recursive formula for the probability mass function of S_N, for a variety of common compounding distributions.

To begin, let M be a random variable that is independent of the sequence $X_1, X_2, \ldots$, and which is such that

$$P\{M = n\} = \frac{nP\{N = n\}}{E[N]}, \qquad n = 1, 2, \ldots$$

Proposition 3.4 The Compound Random Variable Identity

For any function h

$$E[S_N h(S_N)] = E[N]E[X_1 h(S_M)]$$

Proof

$$\begin{aligned}
E[S_N h(S_N)] &= E[\sum_{i=1}^{N} X_i h(S_N)] \\
&= \sum_{n=0}^{\infty} E[\sum_{i=1}^{N} X_i h(S_N) | N = n] P\{N = n\} \\
&\quad \text{(by conditioning on } N\text{)} \\
&= \sum_{n=0}^{\infty} E[\sum_{i=1}^{n} X_i h(S_n) | N = n] P\{N = n\} \\
&= \sum_{n=0}^{\infty} E[\sum_{i=1}^{n} X_i h(S_n)] P\{N = n\} \\
&\quad \text{(by independence of } N \text{ and } X_1, \ldots, X_n\text{)} \\
&= \sum_{n=0}^{\infty} \sum_{i=1}^{n} E[X_i h(S_n)] P\{N = n\}
\end{aligned}$$

Now, because $X_1, \ldots, X_n$ are independent and identically distributed, and $h(S_n) = h(X_1 + \cdots + X_n)$ is a symmetric function of $X_1, \ldots, X_n$, it follows that the distribution of $X_i h(S_n)$ is the same for all $i = 1, \ldots, n$. Therefore, continuing the preceding string of equalities yields

$$\begin{aligned}
E[S_N h(S_N)] &= \sum_{n=0}^{\infty} n E[X_1 h(S_n)] P\{N = n\} \\
&= E[N] \sum_{n=0}^{\infty} E[X_1 h(S_n)] P\{M = n\} \qquad \text{(definition of } M\text{)}
\end{aligned}$$

$$= E[N] \sum_{n=0}^{\infty} E[X_1 h(S_n)|M = n] P\{M = n\}$$

(independence of M and $X_1, \ldots, X_n$)

$$= E[N] \sum_{n=0}^{\infty} E[X_1 h(S_M)|M = n] P\{M = n\}$$

$$= E[N] E[X_1 h(S_M)]$$

which proves the proposition. ■

Suppose now that the X_i are positive integer valued random variables, and let

$$\alpha_j = P\{X_1 = j\}, \qquad j > 0$$

The successive values of $P\{S_N = k\}$ can often be obtained from the following corollary to Proposition 3.4.

Corollary 3.5

$$P\{S_N = 0\} = P\{N = 0\}$$

$$P\{S_N = k\} = \frac{1}{k} E[N] \sum_{j=1}^{k} j\alpha_j P\{S_{M-1} = k - j\}, \qquad k > 0$$

Proof For k fixed, let

$$h(x) = \begin{cases} 1, & \text{if } x = k \\ 0, & \text{if } x \neq k \end{cases}$$

and note that $S_N h(S_N)$ is either equal to k if $S_N = k$ or is equal to 0 otherwise. Therefore,

$$E[S_N h(S_N)] = k P\{S_N = k\}$$

and the compound identity yields

$$k P\{S_N = k\} = E[N] E[X_1 h(S_M)]$$

$$= E[N] \sum_{j=1}^{\infty} E[X_1 h(S_M))|X_1 = j] \alpha_j$$

$$= E[N] \sum_{j=1}^{\infty} j E[h(S_M)|X_1 = j]\alpha_j$$

$$= E[N] \sum_{j=1}^{\infty} j P\{S_M = k|X_1 = j\}\alpha_j \tag{3.33}$$

Now,

$$\begin{aligned} P\{S_M = k|X_1 = j\} &= P\left\{\sum_{i=1}^{M} X_i = k \middle| X_1 = j\right\} \\ &= P\left\{j + \sum_{i=2}^{M} X_i = k \middle| X_1 = j\right\} \\ &= P\left\{j + \sum_{i=2}^{M} X_i = k\right\} \\ &= P\left\{j + \sum_{i=1}^{M-1} X_i = k\right\} \\ &= P\{S_{M-1} = k - j\} \end{aligned}$$

The next to last equality followed because $X_2, \ldots, X_M$ and $X_1, \ldots, X_{M-1}$ have the same joint distribution; namely that of $M - 1$ independent random variables that all have the distribution of X_1, where $M - 1$ is independent of these random variables. Thus the proof follows from Equation (3.33). ■

When the distributions of $M - 1$ and N are related, the preceding corollary can be a useful recursion for computing the probability mass function of S_N, as is illustrated in the following subsections.

3.7.1. Poisson Compounding Distribution

If N is the Poisson distribution with mean λ, then

$$\begin{aligned} P\{M - 1 = n\} &= P\{M = n + 1\} \\ &= \frac{(n + 1)P\{N = n + 1\}}{E[N]} \end{aligned}$$

$$= \frac{1}{\lambda}(n+1)e^{-\lambda}\frac{\lambda^{n+1}}{(n+1)!}$$

$$= e^{-\lambda}\frac{\lambda^n}{n!}$$

Consequently, $M - 1$ is also Poisson with mean λ. Thus, with

$$P_n = P\{S_N = n\}$$

the recursion given by Corollary 3.5 can be written

$$P_0 = e^{-\lambda}$$

$$P_k = \frac{\lambda}{k}\sum_{j=1}^{k} j\,\alpha_j\,P_{k-j}, \qquad k > 0$$

Remark When the X_i are identically 1, the preceding recursion reduces to the well-known identity for a Poisson random variable having mean λ:

$$P\{N = 0\} = e^{-\lambda}$$

$$P\{N = n\} = \frac{\lambda}{n}P\{N = n-1\}, \qquad n \geqslant 1$$

Example 3.30 Let S be a compound Poisson random variable with $\lambda = 4$ and

$$P\{X_i = i\} = 1/4, \qquad i = 1, 2, 3, 4$$

Let us use the recursion given by Corollary 3.5 to determine $P\{S = 5\}$. It gives

$$P_0 = e^{-\lambda} = e^{-4}$$

$$P_1 = \lambda\alpha_1 P_0 = e^{-4}$$

$$P_2 = \frac{\lambda}{2}(\alpha_1 P_1 + 2\alpha_2 P_0) = \frac{3}{2}e^{-4}$$

$$P_3 = \frac{\lambda}{3}(\alpha_1 P_2 + 2\alpha_2 P_1 + 3\alpha_3 P_0) = \frac{13}{6}e^{-4}$$

$$P_4 = \frac{\lambda}{4}(\alpha_1 P_3 + 2\alpha_2 P_2 + 3\alpha_3 P_1 + 4\alpha_4 P_0) = \frac{73}{24}e^{-4}$$

$$P_5 = \frac{\lambda}{5}(\alpha_1 P_4 + 2\alpha_2 P_3 + 3\alpha_3 P_2 + 4\alpha_4 P_1 + 5\alpha_5 P_0) = \frac{501}{120}e^{-4} \qquad \blacksquare$$

3.7.2. Binomial Compounding Distribution

Suppose that N is a binomial random variable with parameters r and p. Then,

$$\begin{aligned}
P\{M-1=n\} &= \frac{(n+1)P\{N=n+1\}}{E[N]} \\
&= \frac{n+1}{rp}\binom{r}{n+1}p^{n+1}(1-p)^{r-n-1} \\
&= \frac{n+1}{rp}\frac{r!}{(r-1-n)!(n+1)!}\,p^{n+1}(1-p)^{r-1-n} \\
&= \frac{(r-1)!}{(r-1-n)!n!}\,p^{n}(1-p)^{r-1-n}
\end{aligned}$$

Thus, $M-1$ is a binomial random variable with parameters $r-1,\ p$.

Fixing p, let $N(r)$ be a binomial random variable with parameters r and p, and let

$$P_r(k) = P\{S_{N(r)} = k\}$$

Then, Corollary 3.5 yields that

$$\begin{aligned}
P_r(0) &= (1-p)^r \\
P_r(k) &= \frac{rp}{k}\sum_{j=1}^{k} j\,\alpha_j\,P_{r-1}(k-j), \qquad k>0
\end{aligned}$$

For instance, letting k equal 1, then 2, and then 3 gives

$$\begin{aligned}
P_r(1) &= rp\,\alpha_1(1-p)^{r-1} \\
P_r(2) &= \frac{rp}{2}\,[\alpha_1 P_{r-1}(1) + 2\alpha_2 P_{r-1}(0)] \\
&= \frac{rp}{2}\Big[(r-1)p\alpha_1^2(1-p)^{r-2} + 2\alpha_2(1-p)^{r-1}\Big] \\
P_r(3) &= \frac{rp}{3}\Big[\alpha_1 P_{r-1}(2) + 2\alpha_2 P_{r-1}(1) + 3\alpha_3 P_{r-1}(0)\Big] \\
&= \frac{\alpha_1 rp}{3}\frac{(r-1)p}{2}\Big[(r-2)p\alpha_1^2(1-p)^{r-3} + 2\alpha_2(1-p)^{r-2}\Big] \\
&\quad + \frac{2\alpha_2 rp}{3}(r-1)p\,\alpha_1(1-p)^{r-2} + \alpha_3 rp(1-p)^{r-1}
\end{aligned}$$

3.7.3. A Compounding Distribution Related to the Negative Binomial

Suppose, for a fixed value of p, $0 < p < 1$, the compounding random variable N has a probability mass function

$$P\{N = n\} = \binom{n+r-1}{r-1} p^r (1-p)^n, \qquad n = 0, 1, \ldots$$

Such a random variable can be thought of as being the number of failures that occur before a total of r successes have been amassed when each trial is independently a success with probability p. (There will be n such failures if the rth success occurs on trial $n + r$. Consequently, $N + r$ is a negative binomial random variable with parameters r and p.) Using that the mean of the negative binomial random variable $N + r$ is $E[N + r] = r/p$, we see that $E[N] = r\frac{1-p}{p}$.

Regard p as fixed, and call N an NB(r) random variable. The random variable $M - 1$ has probability mass function

$$\begin{aligned}
P\{M-1=n\} &= \frac{(n+1)P\{N=n+1\}}{E[N]} \\
&= \frac{(n+1)p}{r(1-p)} \binom{n+r}{r-1} p^r (1-p)^{n+1} \\
&= \frac{(n+r)!}{r!n!} p^{r+1} (1-p)^n \\
&= \binom{n+r}{r} p^{r+1} (1-p)^n
\end{aligned}$$

In other words, $M - 1$ is an NB$(r + 1)$ random variable.

Letting, for an NB(r) random variable N,

$$P_r(k) = P\{S_N = k\}$$

Corollary 3.5 yields that

$$P_r(0) = p^r$$

$$P_r(k) = \frac{r(1-p)}{kp} \sum_{j=1}^{k} j\,\alpha_j\, P_{r+1}(k-j), \qquad k > 0$$

Thus,

$$\begin{aligned}
P_r(1) &= \frac{r(1-p)}{p} \alpha_1 P_{r+1}(0) \\
&= rp^r(1-p)\alpha_1,
\end{aligned}$$

$$P_r(2) = \frac{r(1-p)}{2p}\left[\alpha_1 P_{r+1}(1) + 2\alpha_2 P_{r+1}(0)\right]$$
$$= \frac{r(1-p)}{2p}\left[\alpha_1^2 (r+1) p^{r+1}(1-p) + 2\alpha_2 p^{r+1}\right]$$
$$P_r(3) = \frac{r(1-p)}{3p}\left[\alpha_1 P_{r+1}(2) + 2\alpha_2 P_{r+1}(1) + 3\alpha_3 P_{r+1}(0)\right]$$

and so on.

Exercises

1. If X and Y are both discrete, show that $\sum_x p_{X|Y}(x|y) = 1$ for all y such that $p_Y(y) > 0$.

***2.** Let X_1 and X_2 be independent geometric random variables having the same parameter p. Guess the value of

$$P\{X_1 = i | X_1 + X_2 = n\}$$

> **Hint:** Suppose a coin having probability p of coming up heads is continually flipped. If the second head occurs on flip number n, what is the conditional probability that the first head was on flip number i, $i = 1, \ldots, n-1$?

Verify your guess analytically.

3. The joint probability mass function of X and Y, $p(x, y)$, is given by

$$\begin{array}{lll} p(1,1) = \frac{1}{9}, & p(2,1) = \frac{1}{3}, & p(3,1) = \frac{1}{9}, \\ p(1,2) = \frac{1}{9}, & p(2,2) = 0, & p(3,2) = \frac{1}{18}, \\ p(1,3) = 0, & p(2,3) = \frac{1}{6}, & p(3,3) = \frac{1}{9} \end{array}$$

Compute $E[X|Y = i]$ for $i = 1, 2, 3$.

4. In Exercise 3, are the random variables X and Y independent?

5. An urn contains three white, six red, and five black balls. Six of these balls are randomly selected from the urn. Let X and Y denote respectively the number of white and black balls selected. Compute the conditional probability mass function of X given that $Y = 3$. Also compute $E[X|Y = 1]$.

***6.** Repeat Exercise 5 but under the assumption that when a ball is selected its color is noted, and it is then replaced in the urn before the next selection is made.

7. Suppose $p(x, y, z)$, the joint probability mass function of the random variables X, Y, and Z, is given by

$$\begin{aligned} p(1,1,1) &= \tfrac{1}{8}, & p(2,1,1) &= \tfrac{1}{4}, \\ p(1,1,2) &= \tfrac{1}{8}, & p(2,1,2) &= \tfrac{3}{16}, \\ p(1,2,1) &= \tfrac{1}{16}, & p(2,2,1) &= 0, \\ p(1,2,2) &= 0, & p(2,2,2) &= \tfrac{1}{4} \end{aligned}$$

What is $E[X|Y=2]$? What is $E[X|Y=2, Z=1]$?

8. An unbiased die is successively rolled. Let X and Y denote, respectively, the number of rolls necessary to obtain a six and a five. Find (a) $E[X]$, (b) $E[X|Y=1]$, (c) $E[X|Y=5]$.

9. Show in the discrete case that if X and Y are independent, then

$$E[X|Y=y] = E[X] \qquad \text{for all } y$$

10. Suppose X and Y are independent continuous random variables. Show that

$$E[X|Y=y] = E[X] \qquad \text{for all } y$$

11. The joint density of X and Y is

$$f(x,y) = \frac{(y^2 - x^2)}{8} e^{-y}, \qquad 0 < y < \infty, \quad -y \leqslant x \leqslant y$$

Show that $E[X|Y=y] = 0$.

12. The joint density of X and Y is given by

$$f(x,y) = \frac{e^{-x/y} e^{-y}}{y}, \qquad 0 < x < \infty, \quad 0 < y < \infty$$

Show $E[X|Y=y] = y$.

***13.** Let X be exponential with mean $1/\lambda$; that is,

$$f_X(x) = \lambda e^{-\lambda x}, \qquad 0 < x < \infty$$

Find $E[X|X > 1]$.

14. Let X be uniform over (0, 1). Find $E[X|X < \frac{1}{2}]$.

15. The joint density of X and Y is given by

$$f(x,y) = \frac{e^{-y}}{y}, \qquad 0 < x < y, \quad 0 < y < \infty$$

Compute $E[X^2|Y=y]$.

16. The random variables X and Y are said to have a bivariate normal distribution if their joint density function is given by

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \right.$$
$$\left. \times \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right] \right\}$$

for $-\infty < x < \infty,\ -\infty < y < \infty$, where $\sigma_x,\ \sigma_y,\ \mu_x,\ \mu_y$, and ρ are constants such that $-1 < \rho < 1,\ \sigma_x > 0, \sigma_y > 0, -\infty < \mu_x < \infty, -\infty < \mu_y < \infty$.

(a) Show that X is normally distributed with mean μ_x and variance σ_x^2, and Y is normally distributed with mean μ_y and variance σ_y^2.
(b) Show that the conditional density of X given that $Y = y$ is normal with mean $\mu_x + (\rho\sigma_x/\sigma_y)(y - \mu_y)$ and variance $\sigma_x^2(1-\rho^2)$.

The quantity ρ is called the correlation between X and Y. It can be shown that

$$\begin{aligned}\rho &= \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x\sigma_y} \\ &= \frac{\text{Cov}(X,Y)}{\sigma_x\sigma_y}\end{aligned}$$

17. Let Y be a gamma random variable with parameters (s, α). That is, its density is

$$f_Y(y) = Ce^{-\alpha y}y^{s-1}, \qquad y > 0$$

where C is a constant that does not depend on y. Suppose also that the conditional distribution of X given that $Y = y$ is Poisson with mean y. That is,

$$P\{X = i|Y = y\} = e^{-y}y^i/i!, \qquad i \geqslant 0$$

Show that the conditional distribution of Y given that $X = i$ is the gamma distribution with parameters $(s + i, \alpha + 1)$.

18. Let $X_1, \ldots, X_n$ be independent random variables having a common distribution function that is specified up to an unknown parameter θ. Let $T = T(\mathbf{X})$ be a function of the data $\mathbf{X} = (X_1, \ldots, X_n)$. If the conditional distribution of $X_1, \ldots, X_n$ given $T(\mathbf{X})$ does not depend on θ then $T(\mathbf{X})$ is said to be a *sufficient statistic* for θ. In the following cases, show that $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for θ.

(a) The X_i are normal with mean θ and variance 1.
(b) The density of X_i is $f(x) = \theta e^{-\theta x}$, $x > 0$.
(c) The mass function of X_i is $p(x) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$, $0 < \theta < 1$.
(d) The X_i are Poisson random variables with mean θ.

***19.** Prove that if X and Y are jointly continuous, then

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y)\, dy$$

20. An individual whose level of exposure to a certain pathogen is x will contract the disease caused by this pathogen with probability $P(x)$. If the exposure level of a randomly chosen member of the population has probability density function f, determine the conditional probability density of the exposure level of that member given that he or she

(a) has the disease.
(b) does not have the disease.
(c) Show that when $P(x)$ increases in x, then the ratio of the density of part (a) to that of part (b) also increases in x.

21. Consider Example 3.12 which refers to a miner trapped in a mine. Let N denote the total number of doors selected before the miner reaches safety. Also, let T_i denote the travel time corresponding to the ith choice, $i \geqslant 1$. Again let X denote the time when the miner reaches safety.

(a) Give an identity that relates X to N and the T_i.
(b) What is $E[N]$?
(c) What is $E[T_N]$?
(d) What is $E[\sum_{i=1}^{N} T_i | N = n]$?
(e) Using the preceding, what is $E[X]$?

22. Suppose that independent trials, each of which is equally likely to have any of m possible outcomes, are performed until the same outcome occurs k consecutive times. If N denotes the number of trials, show that

$$E[N] = \frac{m^k - 1}{m - 1}$$

Some people believe that the successive digits in the expansion of $\pi = 3.14159\ldots$ are "uniformly" distributed. That is, they believe that these digits have all the appearance of being independent choices from a distribution that is equally likely to be any of the digits from 0 through 9. Possible evidence against this hypothesis is the fact that starting with the 24,658,601st digit there is a run of nine successive 7s. Is this information consistent with the hypothesis of a uniform distribution?

To answer this, we note from the preceding that if the uniform hypothesis were correct, then the expected number of digits until a run of nine of the same value occurs is

$$(10^9 - 1)/9 = 111{,}111{,}111$$

Thus, the actual value of approximately 25 million is roughly 22 percent of the theoretical mean. However, it can be shown that under the uniformity assumption the standard deviation of N will be approximately equal to the mean. As a result, the observed value is approximately 0.78 standard deviations less than its theoretical mean and is thus quite consistent with the uniformity assumption.

***23.** A coin having probability p of coming up heads is successively flipped until two of the most recent three flips are heads. Let N denote the number of flips. (Note that if the first two flips are heads, then $N = 2$.) Find $E[N]$.

24. A coin, having probability p of landing heads, is continually flipped until at least one head and one tail have been flipped.

(a) Find the expected number of flips needed.
(b) Find the expected number of flips that land on heads.
(c) Find the expected number of flips that land on tails.
(d) Repeat part (a) in the case where flipping is continued until a total of at least two heads and one tail have been flipped.

25. A gambler wins each game with probability p. In each of the following cases, determine the expected total number of wins.

(a) The gambler will play n games; if he wins X of these games, then he will play an additional X games before stopping.
(b) The gambler will play until he wins; if it takes him Y games to get this win, then he will play an additional Y games.

26. You have two opponents with whom you alternate play. Whenever you play A, you win with probability p_A; whenever you play B, you win with probability p_B, where $p_B > p_A$. If your objective is to minimize the number of games you need to play to win two in a row, should you start with A or with B?

Hint: Let $E[N_i]$ denote the mean number of games needed if you initially play i. Derive an expression for $E[N_A]$ that involves $E[N_B]$; write down the equivalent expression for $E[N_B]$ and then subtract.

27. A coin that comes up heads with probability p is continually flipped until the pattern T, T, H appears. (That is, you stop flipping when the most recent flip lands heads, and the two immediately preceding it lands tails.) Let X denote the number of flips made, and find $E[X]$.

28. Polya's urn model supposes that an urn initially contains r red and b blue balls. At each stage a ball is randomly selected from the urn and is then returned along with m other balls of the same color. Let X_k be the number of red balls drawn in the first k selections.

(a) Find $E[X_1]$.
(b) Find $E[X_2]$.
(c) Find $E[X_3]$.
(d) Conjecture the value of $E[X_k]$, and then verify your conjecture by a conditioning argument.
(e) Give an intuitive proof for your conjecture.

Hint: Number the initial r red and b blue balls, so the urn contains one type i red ball, for each $i = 1, \ldots, r$; as well as one type j blue ball, for each $j = 1, \ldots, b$. Now suppose that whenever a red ball is chosen it is returned along with m others of the same type, and similarly whenever a blue ball is chosen it is returned along with m others of the same type. Now, use a symmetry argument to determine the probability that any given selection is red.

29. Two players take turns shooting at a target, with each shot by player i hitting the target with probability p_i, $i = 1, 2$. Shooting ends when two consecutive shots hit the target. Let μ_i denote the mean number of shots taken when player i shoots first, $i = 1, 2$.

(a) Find μ_1 and μ_2.
(b) Let h_i denote the mean number of times that the target is hit when player i shoots first, $i = 1, 2$. Find h_1 and h_2.

30. Let X_i, $i \geqslant 0$ be independent and identically distributed random variables with probability mass function

$$p(j) = P\{X_i = j\}, \qquad j = 1, \ldots, m, \qquad \sum_{j=1}^{m} P(j) = 1$$

Find $E[N]$, where $N = \min\{n > 0\colon\ X_n = X_0\}$.

31. Each element in a sequence of binary data is either 1 with probability p or 0 with probability $1 - p$. A maximal subsequence of consecutive values having identical outcomes is called a run. For instance, if the outcome sequence is $1, 1, 0, 1, 1, 1, 0$, the first run is of length 2, the second is of length 1, and the third is of length 3.

(a) Find the expected length of the first run.
(b) Find the expected length of the second run.

32. Independent trials, each resulting in success with probability p, are performed.

(a) Find the expected number of trials needed for there to have been both at least n successes and at least m failures.

Hint: Is it useful to know the result of the first $n + m$ trials?

(b) Find the expected number of trials needed for there to have been either at least n successes and at least m failures.

Hint: Make use of the result from part (a).

33. If R_i denotes the random amount that is earned in period i, then $\sum_{i=1}^{\infty} \beta^{i-1} R_i$, where $0 < \beta < 1$ is a specified constant, is called the total discounted reward with discount factor β. Let T be a geometric random variable with parameter $1 - \beta$ that is independent of the R_i. Show that the expected total discounted reward is equal to the expected total (undiscounted) reward earned by time T. That is, show that

$$E\left[\sum_{i=1}^{\infty} \beta^{i-1} R_i\right] = E\left[\sum_{i=1}^{T} R_i\right]$$

34. A set of n dice is thrown. All those that land on six are put aside, and the others are again thrown. This is repeated until all the dice have landed on six. Let N denote the number of throws needed. (For instance, suppose that $n = 3$ and that on the initial throw exactly two of the dice land on six. Then the other die will be thrown, and if it lands on six, then $N = 2$.) Let $m_n = E[N]$.

(a) Derive a recursive formula for m_n and use it to calculate m_i, $i = 2, 3, 4$ and to show that $m_5 \approx 13.024$.
(b) Let X_i denote the number of dice rolled on the ith throw. Find $E[\sum_{i=1}^{N} X_i]$.

35. Consider n multinomial trials, where each trial independently results in outcome i with probability p_i, $\sum_{i=1}^{k} p_i = 1$. With X_i equal to the number of trials that result in outcome i, find $E[X_1 | X_2 > 0]$.

36. Let $p_0 = P\{X = 0\}$ and suppose that $0 < p_0 < 1$. Let $\mu = E[X]$ and $\sigma^2 = \text{Var}(X)$. Find (a) $E[X | X \neq 0]$ and (b) $\text{Var}(X | X \neq 0)$.

37. A manuscript is sent to a typing firm consisting of typists A, B, and C. If it is typed by A, then the number of errors made is a Poisson random variable with mean 2.6; if typed by B, then the number of errors is a Poisson random variable with mean 3; and if typed by C, then it is a Poisson random variable with mean 3.4. Let X denote the number of errors in the typed manuscript. Assume that each typist is equally likely to do the work.

(a) Find $E[X]$.
(b) Find $\text{Var}(X)$.

38. Let U be a uniform $(0, 1)$ random variable. Suppose that n trials are to be performed and that conditional on $U = u$ these trials will be independent with a common success probability u. Compute the mean and variance of the number of successes that occur in these trials.

39. A deck of n cards, numbered 1 through n, is randomly shuffled so that all $n!$ possible permutations are equally likely. The cards are then turned over one at a time until card number 1 appears. These upturned cards constitute the first cycle. We now determine (by looking at the upturned cards) the lowest numbered card that has not yet appeared, and we continue to turn the cards face up until that card appears. This new set of cards represents the second cycle. We again determine the lowest numbered of the remaining cards and turn the cards until it appears, and so on until all cards have been turned over. Let m_n denote the mean number of cycles.

(a) Derive a recursive formula for m_n in terms of $m_k,\ k = 1, \dots, n-1$.
(b) Starting with $m_0 = 0$, use the recursion to find m_1, m_2, m_3, and m_4.
(c) Conjecture a general formula for m_n.
(d) Prove your formula by induction on n. That is, show it is valid for $n = 1$, then assume it is true for any of the values $1, \dots, n-1$ and show that this implies it is true for n.
(e) Let X_i equal 1 if one of the cycles ends with card i, and let it equal 0 otherwise, $i=1, \dots, n$. Express the number of cycles in terms of these X_i.
(f) Use the representation in part (e) to determine m_n.
(g) Are the random variables $X_1, \dots, X_n$ independent? Explain.
(h) Find the variance of the number of cycles.

40. A prisoner is trapped in a cell containing three doors. The first door leads to a tunnel that returns him to his cell after two days of travel. The second leads to a tunnel that returns him to his cell after three days of travel. The third door leads immediately to freedom.

(a) Assuming that the prisoner will always select doors 1, 2, and 3 with probabilities 0.5, 0.3, 0.2, what is the expected number of days until he reaches freedom?
(b) Assuming that the prisoner is always equally likely to choose among those doors that he has not used, what is the expected number of days until he reaches freedom? (In this version, for instance, if the prisoner initially tries door 1, then when he returns to the cell, he will now select only from doors 2 and 3.)
(c) For parts (a) and (b) find the variance of the number of days until the prisoner reaches freedom.

41. A rat is trapped in a maze. Initially it has to choose one of two directions. If it goes to the right, then it will wander around in the maze for three minutes and will then return to its initial position. If it goes to the left, then with probability $\frac{1}{3}$ it will depart the maze after two minutes of traveling, and with probability $\frac{2}{3}$ it will return to its initial position after five minutes of traveling. Assuming that the rat is at all times equally likely to go to the left or the right, what is the expected number of minutes that it will be trapped in the maze?

42. A total of 11 people, including you, are invited to a party. The times at which people arrive at the party are independent uniform (0, 1) random variables.

(a) Find the expected number of people who arrive before you.
(b) Find the variance of the number of people who arrive before you.

43. The number of claims received at an insurance company during a week is a random variable with mean μ_1 and variance σ_1^2. The amount paid in each claim is a random variable with mean μ_2 and variance σ_2^2. Find the mean and variance of the amount of money paid by the insurance company each week. What independence assumptions are you making? Are these assumptions reasonable?

44. The number of customers entering a store on a given day is Poisson distributed with mean $\lambda = 10$. The amount of money spent by a customer is uniformly distributed over (0, 100). Find the mean and variance of the amount of money that the store takes in on a given day.

45. An individual traveling on the real line is trying to reach the origin. However, the larger the desired step, the greater is the variance in the result of that step. Specifically, whenever the person is at location x, he next moves to a location having mean 0 and variance βx^2. Let X_n denote the position of the individual after having taken n steps. Supposing that $X_0 = x_0$, find

(a) $E[X_n]$;
(b) $\text{Var}(X_n)$.

46. (a) Show that

$$\text{Cov}(X, Y) = \text{Cov}(X, E[Y \mid X])$$

(b) Suppose, that, for constants a and b,

$$E[Y \mid X] = a + bX$$

Show that

$$b = \text{Cov}(X, Y)/\text{Var}(X)$$

***47.** If $E[Y \mid X] = 1$, show that

$$\text{Var}(X\,Y) \geqslant \text{Var}(X)$$

48. Give another proof of the result of Example 3.17 by computing the moment generating function of $\sum_{i=1}^{N} X_i$ and then differentiating to obtain its moments.

Hint: Let

$$\phi(t) = E\left[\exp\left(t \sum_{i=1}^{N} X_i\right)\right]$$
$$= E\left[E\left[\exp\left(t \sum_{i=1}^{N} X_i\right) \middle| N\right]\right]$$

Now,

$$E\left[\exp\left(t \sum_{i=1}^{N} X_i\right) \middle| N = n\right] = E\left[\exp\left(t \sum_{i=1}^{n} X_i\right)\right] = (\phi_X(t))^n$$

since N is independent of the Xs where $\phi_X(t) = E[e^{tX}]$ is the moment generating function for the Xs. Therefore,

$$\phi(t) = E\left[(\phi_X(t))^N\right]$$

Differentiation yields

$$\phi'(t) = E\left[N(\phi_X(t))^{N-1}\phi'_X(t)\right],$$
$$\phi''(t) = E\left[N(N-1)(\phi_X(t))^{N-2}(\phi'_X(t))^2 + N(\phi_X(t))^{N-1}\phi''_X(t)\right]$$

Evaluate at $t = 0$ to get the desired result.

49. A and B play a series of games with A winning each game with probability p. The overall winner is the first player to have won two more games than the other.

(a) Find the probability that A is the overall winner.
(b) Find the expected number of games played.

50. There are three coins in a barrel. These coins, when flipped, will come up heads with respective probabilities 0.3, 0.5, 0.7. A coin is randomly selected from among these three and is then flipped ten times. Let N be the number of heads obtained on the ten flips. Find

(a) $P\{N = 0\}$.
(b) $P\{N = n\}, n = 0, 1, \ldots, 10$.
(c) Does N have a binomial distribution?
(d) If you win \$1 each time a head appears and you lose \$1 each time a tail appears, is this a fair game? Explain.

51. Do Exercise 50 under the assumption that each time a coin is flipped, it is then put back in the barrel and another coin is randomly selected. Does N have a binomial distribution now?

52. Suppose that X and Y are independent random variables with probability density functions f_X and f_Y. Determine a one-dimensional integral expression for $P\{X + Y < x\}$.

***53.** Suppose X is a Poisson random variable with mean λ. The parameter λ is itself a random variable whose distribution is exponential with mean 1. Show that $P\{X = n\} = (\frac{1}{2})^{n+1}$.

54. A coin is randomly selected from a group of ten coins, the nth coin having a probability $n/10$ of coming up heads. The coin is then repeatedly flipped until a head appears. Let N denote the number of flips necessary. What is the probability distribution of N? Is N a geometric random variable? When would N be a geometric random variable; that is, what would have to be done differently?

55. Suppose in Exercise 42 that, aside from yourself, the number of other people who are invited is a Poisson random variable with mean 10.

(a) Find the expected number of people who arrive before you.
(b) Find the probability that you are the nth person to arrive.

56. Data indicate that the number of traffic accidents in Berkeley on a rainy day is a Poisson random variable with mean 9, whereas on a dry day it is a Poisson random variable with mean 3. Let X denote the number of traffic accidents tomorrow. If it will rain tomorrow with probability 0.6, find

(a) $E[X]$;
(b) $P\{X = 0\}$;
(c) $\mathrm{Var}(X)$.

57. The number of storms in the upcoming rainy season is Poisson distributed but with a parameter value that is uniformly distributed over $(0, 5)$. That is, Λ is uniformly distributed over $(0, 5)$, and given that $\Lambda = \lambda$, the number of storms is Poisson with mean λ. Find the probability there are at least three storms this season.

58. A collection of n coins is flipped. The outcomes are independent, and the ith coin comes up heads with probability α_i, $i = 1, \ldots, n$. Suppose that for some value of j, $1 \leqslant j \leqslant n$, $\alpha_j = \frac{1}{2}$. Find the probability that the total number of heads to appear on the n coins is an even number.

59. Let A and B be mutually exclusive events of an experiment. If independent replications of the experiment are continually performed, what is the probability that A occurs before B?

***60.** Two players alternate flipping a coin that comes up heads with probability p. The first one to obtain a head is declared the winner. We are interested in the probability that the first player to flip is the winner. Before determining this probability, which we will call $f(p)$, answer the following questions.

(a) Do you think that $f(p)$ is a monotone function of p? If so, is it increasing or decreasing?
(b) What do you think is the value of $\lim_{p\to 1} f(p)$?
(c) What do you think is the value of $\lim_{p\to 0} f(p)$?
(d) Find $f(p)$.

61. Suppose in Exercise 29 that the shooting ends when the target has been hit twice. Let m_i denote the mean number of shots needed for the first hit when player i shoots first, $i = 1, 2$. Also, let P_i, $i = 1, 2$, denote the probability that the first hit is by player 1, when player i shoots first.

(a) Find m_1 and m_2.
(b) Find P_1 and P_2.

For the remainder of the problem, assume that player 1 shoots first.

(c) Find the probability that the final hit was by 1.
(d) Find the probability that both hits were by 1.
(e) Find the probability that both hits were by 2.
(f) Find the mean number of shots taken.

62. A, B, and C are evenly matched tennis players. Initially A and B play a set, and the winner then plays C. This continues, with the winner always playing the waiting player, until one of the players has won two sets in a row. That player is then declared the overall winner. Find the probability that A is the overall winner.

63. Suppose there are n types of coupons, and that the type of each new coupon obtained is independent of past selections and is equally likely to be any of the n types. Suppose one continues collecting until a complete set of at least one of each type is obtained.

(a) Find the probability that there is exactly one type i coupon in the final collection.

> **Hint:** Condition on T, the number of types that are collected before the first type i appears.

(b) Find the expected number of types that appear exactly once in the final collection.

64. A and B roll a pair of dice in turn, with A rolling first. A's objective is to obtain a sum of 6, and B's is to obtain a sum of 7. The game ends when either player reaches his or her objective, and that player is declared the winner.

(a) Find the probability that A is the winner.

(b) Find the expected number of rolls of the dice.
(c) Find the variance of the number of rolls of the dice.

65. The number of red balls in an urn that contains n balls is a random variable that is equally likely to be any of the values $0, 1, \ldots, n$. That is,

$$P\{i \text{ red}, n - i \text{ non-red}\} = \frac{1}{n+1}, \qquad i = 0, \ldots, n$$

The n balls are then randomly removed one at a time. Let Y_k denote the number of red balls in the first k selections, $k = 1, \ldots, n$.

(a) Find $P\{Y_n = j\}$, $j = 0, \ldots, n$.
(b) Find $P\{Y_{n-1} = j\}$, $j = 0, \ldots, n$.
(c) What do you think is the value of $P\{Y_k = j\}$, $j = 0, \ldots, n$?
(d) Verify your answer to part (c) by a backwards induction argument. That is, check that your answer is correct when $k = n$, and then show that whenever it is true for k it is also true for $k - 1$, $k = 1, \ldots, n$.

66. The opponents of soccer team A are of two types: either they are a class 1 or a class 2 team. The number of goals team A scores against a class i opponent is a Poisson random variable with mean λ_i, where $\lambda_1 = 2$, $\lambda_2 = 3$. This weekend the team has two games against teams they are not very familiar with. Assuming that the first team they play is a class 1 team with probability 0.6 and the second is, independently of the class of the first team, a class 1 team with probability 0.3, determine

(a) the expected number of goals team A will score this weekend.
(b) the probability that team A will score a total of five goals.

***67.** A coin having probability p of coming up heads is continually flipped. Let $P_j(n)$ denote the probability that a run of j successive heads occurs within the first n flips.

(a) Argue that

$$P_j(n) = P_j(n-1) + p^j(1-p)[1 - P_j(n-j-1)]$$

(b) By conditioning on the first non-head to appear, derive another equation relating $P_j(n)$ to the quantities $P_j(n-k)$, $k = 1, \ldots, j$.

68. In a knockout tennis tournament of 2^n contestants, the players are paired and play a match. The losers depart, the remaining 2^{n-1} players are paired, and they play a match. This continues for n rounds, after which a single player remains unbeaten and is declared the winner. Suppose that the contestants are numbered 1 through 2^n, and that whenever two players contest a match, the lower numbered one wins with probability p. Also suppose that the pairings of the remaining players are always done at random so that all possible pairings for that round are equally likely.

(a) What is the probability that player 1 wins the tournament?
(b) What is the probability that player 2 wins the tournament?

Hint: Imagine that the random pairings are done in advance of the tournament. That is, the first-round pairings are randomly determined; the 2^{n-1} first-round pairs are then themselves randomly paired, with the winners of each pair to play in round 2; these 2^{n-2} groupings (of four players each) are then randomly paired, with the winners of each grouping to play in round 3, and so on. Say that players i and j are scheduled to meet in round k if, provided they both win their first $k-1$ matches, they will meet in round k. Now condition on the round in which players 1 and 2 are scheduled to meet.

69. In the match problem, say that $(i, j), i < j$, is a pair if i chooses j's hat and j chooses i's hat.

(a) Find the expected number of pairs.
(b) Let Q_n denote the probability that there are no pairs, and derive a recursive formula for Q_n in terms of $Q_j, j < n$.

Hint: Use the cycle concept.

(c) Use the recursion of part (b) to find Q_8.

70. Let N denote the number of cycles that result in the match problem.

(a) Let $M_n = E[N]$, and derive an equation for M_n in terms of $M_1, \ldots, M_{n-1}$.
(b) Let C_j denote the size of the cycle that contains person j. Argue that

$$N = \sum_{j=1}^{n} 1/C_j$$

and use the preceding to determine $E[N]$.
(c) Find the probability that persons $1, 2, \ldots, k$ are all in the same cycle.
(d) Find the probability that $1, 2, \ldots, k$ is a cycle.

71. Use the equation following (3.14) to obtain Equation (3.10).

Hint: First multiply both sides of Equation (3.14) by n, then write a new equation by replacing n with $n-1$, and then subtract the former from the latter.

72. In Example 3.25 show that the conditional distribution of N given that $U_1 = y$ is the same as the conditional distribution of M given that $U_1 = 1 - y$. Also, show that

$$E[N|U_1 = y] = E[M|U_1 = 1 - y] = 1 + e^y$$

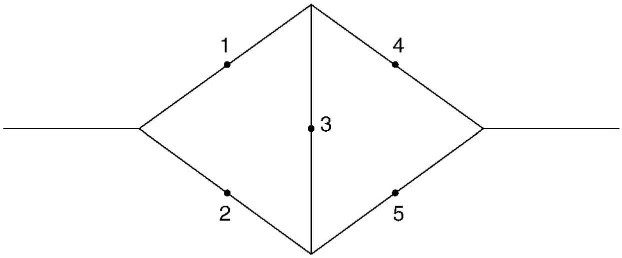

Figure 3.7.

***73.** Suppose that we continually roll a die until the sum of all throws exceeds 100. What is the most likely value of this total when you stop?

74. There are five components. The components act independently, with component i working with probability p_i, $i = 1, 2, 3, 4, 5$. These components form a system as shown in Figure 3.7.

The system is said to work if a signal originating at the left end of the diagram can reach the right end, where it can pass through a component only if that component is working. (For instance, if components 1 and 4 both work, then the system also works.) What is the probability that the system works?

75. This problem will present another proof of the ballot problem of Example 3.24.

(a) Argue that

$$P_{n,m} = 1 - P\{A \text{ and } B \text{ are tied at some point}\}$$

(b) Explain why

$$P\{A \text{ receives first vote and they are eventually tied}\}$$
$$= P\{B \text{ receives first vote and they are eventually tied}\}$$

Hint: Any outcome in which they are eventually tied with A receiving the first vote corresponds to an outcome in which they are eventually tied with B receiving the first vote. Explain this correspondence.

(c) Argue that $P\{\text{eventually tied}\} = 2m/(n + m)$, and conclude that $P_{n,m} = (n - m)/(n + m)$.

76. Consider a gambler who on each bet either wins 1 with probability 18/38 or loses 1 with probability 20/38. (These are the probabilities if the bet is that a roulette wheel will land on a specified color.) The gambler will quit either when he or she is winning a total of 5 or after 100 plays. What is the probability he or she plays exactly 15 times?

77. Show that

(a) $E[XY|Y=y]=yE[X|Y=y]$
(b) $E[g(X,Y)|Y=y]=E[g(X,y)|Y=y]$
(c) $E[XY]=E[YE[X|Y]]$

78. In the ballot problem (Example 3.24), compute $P\{A$ is never behind$\}$.

79. An urn contains n white and m black balls which are removed one at a time. If $n>m$, show that the probability that there are always more white than black balls in the urn (until, of course, the urn is empty) equals $(n-m)/(n+m)$. Explain why this probability is equal to the probability that the set of withdrawn balls always contains more white than black balls. [This latter probability is $(n-m)/(n+m)$ by the ballot problem.]

80. A coin that comes up heads with probability p is flipped n consecutive times. What is the probability that starting with the first flip there are always more heads than tails that have appeared?

81. Let $X_i, i\geqslant 1$, be independent uniform $(0,1)$ random variables, and define N by

$$N=\min\{n\colon X_n<X_{n-1}\}$$

where $X_0=x$. Let $f(x)=E[N]$.

(a) Derive an integral equation for $f(x)$ by conditioning on X_1.
(b) Differentiate both sides of the equation derived in part (a).
(c) Solve the resulting equation obtained in part (b).
(d) For a second approach to determining $f(x)$ argue that

$$P\{N\geqslant k\}=\frac{(1-x)^{k-1}}{(k-1)!}$$

(e) Use part (d) to obtain $f(x)$.

82. Let $X_1, X_2, \ldots$ be independent continuous random variables with a common distribution function F and density $f=F'$, and for $k\geqslant 1$ let

$$N_k=\min\{n\geqslant k\colon X_n=k\text{th largest of } X_1,\ldots,X_n\}$$

(a) Show that $P\{N_k=n\}=\frac{k-1}{n(n-1)}, n\geqslant k$.
(b) Argue that

$$f_{X_{N_k}}(x)=f(x)(\bar{F}(x))^{k-1}\sum_{i=0}^{\infty}\binom{i+k-2}{i}(F(x))^i$$

(c) Prove the following identity:

$$a^{1-k} = \sum_{i=0}^{\infty} \binom{i+k-2}{i} (1-a)^i, \qquad 0 < a < 1, k \geqslant 2$$

Hint: Use induction. First prove it when $k = 2$, and then assume it for k. To prove it for $k + 1$, use the fact that

$$\sum_{i=1}^{\infty} \binom{i+k-1}{i} (1-a)^i = \sum_{i=1}^{\infty} \binom{i+k-2}{i} (1-a)^i + \sum_{i=1}^{\infty} \binom{i+k-2}{i-1} (1-a)^i$$

where the preceding used the combinatorial identity

$$\binom{m}{i} = \binom{m-1}{i} + \binom{m-1}{i-1}$$

Now, use the induction hypothesis to evaluate the first term on the right side of the preceding equation.

(d) Conclude that X_{N_k} has distribution F.

83. An urn contains n balls, with ball i having weight $w_i, i = 1, \ldots, n$. The balls are withdrawn from the urn one at a time according to the following scheme: When S is the set of balls that remains, ball i, $i \in S$, is the next ball withdrawn with probability $w_i / \sum_{j \in S} w_j$. Find the expected number of balls that are withdrawn before ball i, $i = 1, \ldots, n$.

84. In the list example of Section 3.6.1 suppose that the initial ordering at time $t = 0$ is determined completely at random; that is, initially all $n!$ permutations are equally likely. Following the front-of-the-line rule, compute the expected position of the element requested at time t.

Hint: To compute $P\{e_j \text{ precedes } e_i \text{ at time } t\}$ condition on whether or not either e_i or e_j has ever been requested prior to t.

85. In the list problem, when the P_i are known, show that the best ordering (best in the sense of minimizing the expected position of the element requested) is to place the elements in decreasing order of their probabilities. That is, if $P_1 > P_2 > \cdots > P_n$, show that $1, 2, \ldots, n$ is the best ordering.

86. Consider the random graph of Section 3.6.2 when $n = 5$. Compute the probability distribution of the number of components and verify your solution by using it to compute $E[C]$and then comparing your solution with

$$E[C] = \sum_{k=1}^{5} \binom{5}{k} \frac{(k-1)!}{5^k}$$

87. (a) From the results of Section 3.6.3 we can conclude that there are $\binom{n+m-1}{m-1}$ nonnegative integer valued solutions of the equation $x_1 + \cdots + x_m = n$. Prove this directly.

(b) How many positive integer valued solutions of $x_1 + \cdots + x_m = n$ are there?

Hint: Let $y_i = x_i - 1$.

(c) For the Bose–Einstein distribution, compute the probability that exactly k of the X_i are equal to 0.

88. In Section 3.6.3, we saw that if U is a random variable that is uniform on (0, 1) and if, conditional on $U = p$, X is binomial with parameters n and p, then

$$P\{X = i\} = \frac{1}{n+1}, \qquad i = 0, 1, \ldots, n$$

For another way of showing this result, let $U, X_1, X_2, \ldots, X_n$ be independent uniform (0, 1) random variables. Define X by

$$X = \#i\colon X_i < U$$

That is, if the $n+1$ variables are ordered from smallest to largest, then U would be in position $X+1$.

(a) What is $P\{X = i\}$?

(b) Explain how this proves the result of Exercise 88.

89. Let $I_1, \ldots, I_n$ be independent random variables, each of which is equally likely to be either 0 or 1. A well-known nonparametric statistical test (called the signed rank test) is concerned with determining $P_n(k)$ defined by

$$P_n(k) = P\left\{\sum_{j=1}^{n} j I_j \leqslant k\right\}$$

Justify the following formula:

$$P_n(k) = \tfrac{1}{2} P_{n-1}(k) + \tfrac{1}{2} P_{n-1}(k-n)$$

90. The number of accidents in each period is a Poisson random variable with mean 5. With $X_n, n \geqslant 1$, equal to the number of accidents in period n, find $E[N]$ when

(a) $N = \min(n\colon X_{n-2} = 2, X_{n-1} = 1, X_n = 0)$;
(b) $N = \min(n\colon X_{n-3} = 2, X_{n-2} = 1, X_{n-1} = 0, X_n = 2)$.

91. Find the expected number of flips of a coin, which comes up heads with probability p, that are necessary to obtain the pattern h, t, h, h, t, h, t, h.

92. The number of coins that Josh spots when walking to work is a Poisson random variable with mean 6. Each coin is equally likely to be a penny, a nickel, a dime, or a quarter. Josh ignores the pennies but picks up the other coins.

(a) Find the expected amount of money that Josh picks up on his way to work.
(b) Find the variance of the amount of money that Josh picks up on his way to work.
(c) Find the probability that Josh picks up exactly 25 cents on his way to work.

***93.** Consider a sequence of independent trials, each of which is equally likely to result in any of the outcomes $0, 1, \ldots, m$. Say that a round begins with the first trial, and that a new round begins each time outcome 0 occurs. Let N denote the number of trials that it takes until all of the outcomes $1, \ldots, m-1$ have occurred in the same round. Also, let T_j denote the number of trials that it takes until j distinct outcomes have occurred, and let I_j denote the jth distinct outcome to occur. (Therefore, outcome I_j first occurs at trial T_j.)

(a) Argue that the random vectors $(I_1, \ldots, I_m)$ and $(T_1, \ldots, T_m)$ are independent.
(b) Define X by letting $X = j$ if outcome 0 is the jth distinct outcome to occur. (Thus, $I_X = 0$.) Derive an equation for $E[N]$ in terms of $E[T_j]$, $j = 1, \ldots, m-1$ by conditioning on X.
(c) Determine $E[T_j]$, $j = 1, \ldots, m-1$.

Hint: See Exercise 42 of Chapter 2.

(d) Find $E[N]$.

94. Let N be a hypergeometric random variable having the distribution of the number of white balls in a random sample of size r from a set of w white and b blue balls. That is,

$$P\{N = n\} = \frac{\binom{w}{n}\binom{b}{r-n}}{\binom{w+b}{r}}$$

where we use the convention that $\binom{m}{j} = 0$ if either $j < 0$ or $j > m$. Now, consider a compound random variable $S_N = \sum_{i=1}^{N} X_i$, where the X_i are positive integer valued random variables with $\alpha_j = P\{X_i = j\}$.

(a) With M as defined as in Section 3.7, find the distribution of $M - 1$.
(b) Suppressing its dependence on b, let $P_{w,r}(k) = P\{S_N = k\}$, and derive a recursion equation for $P_{w,r}(k)$.
(c) Use the recursion of (b) to find $P_{w,r}(2)$.