Chapter 6
Standard Discrete Distributions

A few special discrete distributions arise very frequently in applications. Either the underlying probability mechanism of a problem is such that one of these distributions is truly the correct distribution for that problem or the problem may be such that one of these distributions is a very good choice to model that problem. We present these distributions and study their basic properties in this chapter; they deserve the special attention because of their importance in applications. The special distributions we present are the discrete uniform, binomial, geometric, negative binomial, hypergeometric, and Poisson. Benford’s distribution is also covered briefly. A few other special distributions are covered in the chapter exercises.

6.1 Introduction to Special Distributions

We first provide the pmfs of these special distributions and a quick description of the contexts where they are relevant. We will then study these distributions in detail in later sections.

The Discrete Uniform Distribution. The discrete uniform distribution represents a finite number of equally likely values. The simplest real-life example is the face obtained when a fair die is rolled once. It can also occur in some other physical phenomena, particularly when the number of possible values is small and the scientist feels that they are just equally likely. If we let the values of the random variable be 1, 2, . . . , n, then the pmf of the discrete uniform distribution is

\[ p(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n, \]

We sometimes write \( X \sim \text{Unif}\{1, 2, \ldots, n\} \).

The Binomial Distribution. The binomial distribution represents a sequence of independent coin-tossing experiments. Suppose a coin with probability \( p, 0 < p < 1 \), for heads in a single trial is tossed independently a prespecified number of times, say \( n \) times, \( n \geq 1 \). Let \( X \) be the number of times in these \( n \) tosses that a head is obtained. Then the pmf of \( X \) is

\[ P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots, n, \]
the \( \binom{n}{x} \) term giving the choice of the \( x \) tosses out of the \( n \) tosses in which the heads occur.

Coin tossing, of course, is just an artifact. Suppose a trial can result in only one of two outcomes, called a success (S) or a failure (F), the probability of obtaining a success being \( p \) in any trial. Such a trial is called a Bernoulli trial. Suppose a Bernoulli trial is repeated independently a prespecified number of times, say \( n \) times. Let \( X \) be the number of times in the \( n \) trials that a success is obtained. Then \( X \) has the pmf given above, and we say that \( X \) has a binomial distribution with parameters \( n \) and \( p \) and write \( X \sim \text{Bin}(n, p) \).

**The Geometric Distribution.** Suppose a coin with probability \( p; 0 < p < 1 \) for heads in a single trial is tossed repeatedly until a head is obtained for the first time. Assume that the tosses are independent. Let \( X \) be the number of the toss at which the very first head is obtained. Then the pmf of \( X \) is

\[
P(X = x) = p(1 - p)^{x-1}, x = 1, 2, 3, \ldots.
\]

We say that \( X \) has a geometric distribution with parameter \( p \), and we will write \( X \sim \text{Geo}(p) \). The distinction between the binomial distribution and the geometric distribution is that in the binomial case the number of tosses is prespecified, but in the geometric case the number of tosses actually performed when the experiment ends is a random variable. A geometric distribution measures a waiting time for the first success in a sequence of independent Bernoulli trials, each with the same success probability \( p \); i.e., the coin cannot change from one toss to another.

**The Negative Binomial Distribution.** The negative binomial distribution is a generalization of a geometric distribution when we repeatedly toss a coin with probability \( p \) for heads, independently, until a total number of \( r \) heads has been obtained, where \( r \) is some fixed integer \( \geq 1 \). The case \( r = 1 \) corresponds to the geometric distribution. Let \( X \) be the number of the first toss at which the \( r \)th success is obtained. Then the pmf of \( X \) is

\[
P(X = x) = \binom{x-1}{r-1} p^r (1 - p)^{x-r}, x = r, r + 1, \ldots,
\]

the term \( \binom{x-1}{r-1} \) simply giving the choice of the \( r - 1 \) tosses among the first \( x - 1 \) tosses where the first \( r - 1 \) heads were obtained. We say that \( X \) has a negative binomial distribution with parameters \( r \) and \( p \), and we will write \( X \sim \text{NB}(r, p) \).

**The Hypergeometric Distribution.** The hypergeometric distribution also represents the number of successes in a prespecified number of Bernoulli trials, but the trials happen to be dependent. A typical example is that of a finite population in which there are in all \( N \) objects, of which some \( D \) are of type I and the other \( N - D \) are of type II. A sample without replacement of size \( n \), \( 1 \leq n < N \), is chosen at random from the population. Thus, the selected sampling units are necessarily different. Let
Let $X$ be the number of units or individuals of type I among the $n$ units chosen. Then the pmf of $X$ is

$$P(X = x) = \binom{n}{x} \binom{D}{x} \binom{N-D}{n-x} / \binom{N}{n},$$

$n - N + D \leq x \leq D$; note that, trivially, $x$ is also $\geq 0$ and $\leq n$. An example would be that of a pollster polling $n = 100$ people from a population of 10,000 people, where $D = 5500$ are in favor of some proposition and the remaining $N - D = 4500$ are against it. The number of individuals in the sample who are in favor of the proposition then has the pmf above. We say that such an $X$ has a hypergeometric distribution with parameters $n, D, N$, and we will write $X \sim \text{Hypergeo}(n, D, N)$.

**The Poisson Distribution.** The Poisson distribution is perhaps the most used and useful distribution for modeling nonnegative integer-valued random variables. Unlike the first four distributions above, we cannot say that a Poisson distribution is necessarily the correct distribution for some integer-valued random variable. Rather, a Poisson distribution is chosen by a scientist as his or her model for the distribution of an integer-valued random variable. But the choice of the Poisson distribution as a model is frequently extremely successful in describing and predicting how the random variable behaves. The Poisson distribution also arises, as a mathematical fact, as the limiting distribution of numerous integer-valued random variables when in some sense a sequence of Bernoulli trials makes it increasingly harder to obtain a success; i.e., the number of times a very rare event happens if we observe the process for a long time often has an approximately Poisson distribution.

The pmf of a Poisson distribution with parameter $\lambda$ is

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \ldots;$$

by using the power series expansion of $e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$, it follows that this is indeed a valid pmf.

Three specific situations where a Poisson distribution is almost routinely adopted as a model are the following:

(a) The number of times a specific event happens in a specified period of time; e.g., the number of phone calls received by someone over a 24 hour period.

(b) The number of times a specific event or phenomenon is observed in a specified amount of area or volume; e.g., the number of bacteria of a certain kind in one liter of a sample of water, the number of misprints per page of a book, etc.

(c) The number of times a success is obtained when a Bernoulli trial with success probability $p$ is repeated independently $n$ times, with $p$ being small and $n$ being large, such that the product $np$ has a moderate value, say between .5 and 10.

We now treat these distributions in greater detail one at a time.
6.2 Discrete Uniform Distribution

**Definition 6.1.** The discrete uniform distribution on \( \{1, 2, \ldots, n\} \) is defined by the pmf \( P(X = x) = \frac{1}{n}, x = 1, 2, \ldots, n \), and zero otherwise. Of course, the set of values can be any finite set; we take the values to be \( 1, 2, \ldots, n \) for convenience.

Clearly, for any given integer \( k, 1 \leq k \leq n \), \( F(k) = P(X \leq k) = \frac{k}{n} \). The first few moments are found easily. For example,

\[
\mu = E(X) = \sum_{x=1}^{n} xp(x) = \frac{n}{n} \sum_{x=1}^{n} x - \frac{1}{n} \sum_{x=1}^{n} x = \frac{n(n+1)}{2}.
\]

Similarly,

\[
E(X^2) = \sum_{x=1}^{n} x^2 p(x) = \frac{1}{n} \sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}.
\]

Therefore,

\[
\sigma^2 = \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2 - 1}{12}.
\]

It follows from the trivial symmetric nature of the discrete uniform distribution that \( E(X - \mu)^3 = 0 \). We can also find \( E(X - \mu)^4 \) in closed form. For this, the only additional fact that we need is that \( \sum_{x=1}^{n} x^4 = \frac{n(n+1)^2}{2} \). Then, by expanding \( (X - \mu)^4 \), after some algebra it follows that

\[
E(X - \mu)^4 = \frac{(3n^2 - 7)(n^2 - 1)}{240}.
\]

The moment information about the discrete uniform distribution is collected together in the theorem below.

**Theorem 6.1.** Let \( X \sim \text{Unif}\{1, 2, \ldots, n\} \). Then,

\[
\mu = E(X) = \frac{n + 1}{2}; \quad \sigma^2 = \text{Var}(X) = \frac{n^2 - 1}{12}; \quad E(X - \mu)^3 = 0;
\]

\[
E(X - \mu)^4 = \frac{(3n^2 - 7)(n^2 - 1)}{240}.
\]
Corollary 6.1. The skewness and the kurtosis of the discrete uniform distribution are

\[ \beta = 0; \gamma = -\frac{6}{5} \frac{n^2 + 1}{n^2 - 1}. \]

6.3 Binomial Distribution

We start with a few examples.

Example 6.1 (Heads in Coin Tosses). Suppose a fair coin is tossed ten times, independently, and suppose \( X \) is the number of times in the ten tosses that a head is obtained. Then \( X \sim Bin(n, p) \) with \( n = 10, p = \frac{1}{2} \). Therefore,

\[ P(X = x) = \binom{10}{x} \left( \frac{1}{2} \right)^{10}, x = 0, 1, 2, \ldots, 10. \]

Converting to decimals, the pmf of \( X \) is

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X = x) )</td>
<td>.0010</td>
<td>.0098</td>
<td>.0439</td>
<td>.1172</td>
<td>.2051</td>
<td>.2461</td>
<td>.2051</td>
<td>.1172</td>
<td>.0439</td>
<td>.0098</td>
<td>.0010</td>
</tr>
</tbody>
</table>

Note that the pmf is symmetric about \( x = 5 \) and that \( P(X = x) \) increases from \( x = 0 \) to \( x = 5 \) and then decreases from \( x = 5 \) to \( x = 10 \) symmetrically.

Example 6.2 (Guessing on a Multiple-Choice Exam). A multiple-choice test with 20 questions has five possible answers for each question. A completely unprepared student picks the answer for each question at random and independently. Suppose \( X \) is the number of questions that the student answers correctly.

We identify each question with a Bernoulli trial and a correct answer as a success. Since there are 20 questions and the student picks his answer at random from five choices, \( X \sim Bin(n, p) \), with \( n = 20, p = \frac{1}{5} = .2 \). We can now answer any question we want about \( X \).

For example,

\[ P(\text{The student gets every answer wrong}) = P(X = 0) = .8^{20} = .0115, \]

while

\[ P(\text{The student gets every answer right}) = P(X = 20) = .2^{20} = 1.05 \times 10^{-14}, \]

a near impossibility. Suppose the instructor has decided that it will take at least 13 correct answers to pass this test. Then,
\begin{align*}
P(\text{The student will pass}) &= \sum_{x=13}^{20} \binom{20}{x} \cdot 2^x \cdot 0.8^{20-x} = 0.00015, \\
\end{align*}

still a very small probability.

**Example 6.3** *(To Cheat or Not to Cheat).* Ms. Smith drives into town once a week to buy groceries. In the past she parked her car at a lot for five dollars, but she decided that for the next five weeks she will park at the fire hydrant and risk getting tickets with fines of 25 dollars per offense. If the probability of getting a ticket is 0.1, what is the probability that she will pay more in fines in five weeks than she would pay in parking fees if she had opted not to park by the fire hydrant?

Suppose that \( X \) is the number of weeks among the next five weeks in which she gets a ticket. Then, \( X \sim Bin(5, 0.1) \). Ms. Smith’s parking fees would have been 25 dollars for the five weeks combined if she did not park by the hydrant. Thus, the required probability is

\[
P(25X > 25) = P(X > 1) = 1 - [P(X = 0) + P(X = 1)]
= 1 - \left[ 0.9^5 + \binom{5}{1} \cdot 0.1 \cdot 0.9^4 \right] = 0.0815.
\]

So the chances are quite low that Ms. Smith will pay more in tickets by breaking the law than she would pay by paying the parking fees.

**Example 6.4.** Suppose a fair coin is tossed \( n = 2m \) times. What is the probability that the number of heads obtained will be an even number?

Since \( X = \) the number of heads \( \sim Bin(2m, \frac{1}{2}) \), we want to find

\[
P(X = 0) + P(X = 2) + \cdots + P(X = 2m) = \sum_{x=0}^{m} \binom{2m}{2x} \frac{1}{2^{2m}} = \frac{2^{2m-1}}{2^m}
= \frac{1}{2}
\]

on using the identity that, for any \( n, \)

\[
\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = 2^{n-1}.
\]

Thus, with a fair coin, the chances of getting an even number of heads in an even number of tosses are \( \frac{1}{2} \). The same is true also if the number of tosses is odd and is proved similarly.

**Example 6.5** *(Flush in Poker).* A flush in five-card poker is five cards of the same suit but not in a sequence. We saw in Chapter 1 that the probability of obtaining a flush in five-card poker is 0.00197.
Suppose someone plays poker once a week every week for a year, and each time that he plays, he plays four deals. Let $X$ be the number of times he obtains a flush during the year. Assuming that decks are always well shuffled between plays, $X \sim \text{Bin}(n, p)$, where $n = 52 \times 4 = 208$ and $p = .00197$. Then, $P(X \geq 1) = 1 - (1 - .00197)^{208} = .3365$. So there is about a one in three chance that the player will obtain a flush within a year.

In this example, $n$ was large and $p$ was small. In such cases, the $\text{Bin}(n, p)$ distribution can be well approximated by a Poisson distribution with $\lambda = np$. If we do the approximation, we will get $P(X \geq 1) \approx 1 - e^{-208 \times .00197} = 1 - e^{-40976} = .3362$, clearly a very close approximation to the exact value .3365. We will discuss Poisson approximations of binomials in greater detail later in this chapter.

**Example 6.6 (A Stock Inventory Example).** This example takes a little more careful reading because the formulation is a little harder. Here is the problem. Professor Rubin loves diet soda. Twice a day he drinks an 8 oz. can of diet soda, and each time he reaches at random into one of two brown bags containing Diet Coke and Diet Pepsi, respectively. One box of soda picked up at a supermarket has six soda cans. How many boxes of each type of soda should professor Rubin buy per to be 90% sure that he will not find a brown bag empty when he reaches into it?

Let $X$ = the number of times Professor Rubin reaches to find a Diet Coke; then, $X \sim \text{Bin}(n, p)$ with $n = 14$ and $p = .5$. Since $p = .5$, $n - X$ is also distributed as the same binomial, namely $\text{Bin}(n, p)$, with $n = 14$ and $p = .5$. Suppose Professor Rubin has $N$ sodas of each type in stock. We want $P(X > N) + P(n - X > N) \leq .1$. Now,

$$P(X > N) + P(n - X > N) = 2 \sum_{x=N+1}^{n} \binom{n}{x} (.5)^x$$

$$= 2 \sum_{x=N+1}^{14} \binom{14}{x} (.5)^{14} = g(N),$$

say. By computing it, we find that $g(9) = .18$ and $g(10) = .06 < .1$. Therefore, Professor Rubin needs to have ten sodas of each type (that is, two boxes of each type of soda) in stock each week.

**Example 6.7 (Flukes are Easier in the Short Run).** Suppose two tennis players, A and B, will play an odd number of games, and whoever wins a majority of the games will be the winner. Suppose that A is a better player, and A has a probability of .6 of winning any single game. If B were to win this tournament, it might be considered a fluke.

Suppose that they were to play three games. Let $X$ be the number of games won by B. Under the usual assumptions of independence, $X \sim \text{Bin}(n, p)$ with $n = 3, p = .4$. Thus, the chances of B winning the tournament are

$$P(X \geq 2) = 3(.4)^2(.6) + .4^3 = .352.$$
Suppose next that they were to play nine games. Now, $X \sim Bin(n, p)$ with $n = 9$, $p = .4$, so the chances of B winning the tournament are

$$P(X \geq 5) = \sum_{x=5}^{9} \binom{9}{x} (.4)^x (.6)^{9-x} = .2665.$$

We see that the chances of B winning the tournament go down when they play more games. This is because a weaker player can get lucky in the short run, but the luck will run out in the long run.

Some key mathematical facts about a binomial distribution are given in the following theorem.

**Theorem 6.2.** Let $X \sim Bin(n, p)$. Then,

(a) $\mu = E(X) = np; \sigma^2 = Var(X) = np(1 - p)$.

(b) The mgf of $X$ equals $\psi(t) = (pe^t + 1 - p)^n$ at any $t$.

(c) $E[(X - \mu)^3] = np(1 - 3p + 2p^2)$.

(d) $E[(X - \mu)^4] = np(1 - p)[1 + 3(n - 2)p(1 - p)]$.

**Proof.** By writing $X$ as $X = \sum_{i=1}^{n} I_{A_i}$, where $A_i$ is the event of a success on the $i$th Bernoulli trial, it follows readily that $E(X) = \sum_{i=1}^{n} P(A_i) = np$ and $Var(X) = \sum_{i=1}^{n} Var(I_{A_i}) = \sum_{i=1}^{n} P(A_i)(1 - P(A_i)) = np(1 - p)$.

The mgf expression also follows immediately from this representation using the indicator variables $I_{A_i}$, as each indicator variable has the mgf $(pe^t + 1 - p)$, and they are independent.

Parts (c) and (d) follow on differentiating $\psi(t)$ three and four times, respectively, thus obtaining $E(X^3)$ and $E(X^4)$ as the third and fourth derivatives of $\psi(t)$ at zero, and finally plugging them into the binomial expansion $E[(X - \mu)^3] = E(X^3) - 3\mu E(X^2) + 2\mu^3$ and a similar expansion for $E[(X - \mu)^4]$. This tedious algebra is omitted.

**Corollary 6.2.** Let $\beta = \beta(n, p)$ be the skewness and $\gamma = \gamma(n, p)$ be the kurtosis of $X$. Then $\beta, \gamma \to 0$ for any $p$ as $n \to \infty$.

The corollary follows by directly using the definitions $\beta = \frac{E[(X-\mu)^3]}{\sigma^3}$ and $\gamma = \frac{E[(X-\mu)^4]}{\sigma^4} - 3$ and plugging in the formulas from the theorem above.

Thus, whatever $p, 0 < p < 1$, the binomial distribution becomes nearly symmetric and normal-like as $n$ gets large.

Mean absolute deviations, whenever they can be found in closed form, are appealing measures of variability. Remarkably, an exact formula for the mean absolute deviation of a general binomial distribution exists and is quite classic. Several different versions of it have been derived by various authors, including Poincaré (1896) and Feller (1968); Diaconis and Zabell (1991) is an authoritative exposition of the problem. Another interesting question is, which value in a general binomial distribution has the largest probability? That is, what is the mode of the distribution? The next result summarizes the answers to these questions.
Theorem 6.3 (Mean Absolute Deviation and Mode). Let $X \sim \text{Bin}(n, p)$. Let $\nu$ denote the smallest integer $> np$ and let $m = \lfloor np + p \rfloor$. Then,

(a) $E|X - np| = 2\nu(1 - p)P(X = \nu)$.

(b) The mode of $X$ equals $m$. In particular, if $np$ is an integer, then the mode is exactly $np$; if $np$ is not an integer, then the mode is one of the two integers just below and just above $np$.

Proof. Suppose first that $m \geq 1$. Part (b) can be proved by looking at the ratio $\frac{P(X=k+1)}{P(X=k)}$ and on observing that this ratio is $\geq 1$ for $k \leq m - 1$. If $n, p$ are such that $m$ is zero, then $P(X = k)$ can be directly verified to be maximized at $k = 0$. This is a standard technique for finding the maximum of a unimodal function of an integer argument. Part (a) requires nontrivial calculations; see Diaconis and Zabell (1991).

Remark 6.1. It follows from this theorem that the mode of a binomial distribution need not be the integer closest to the mean $np$. The modal value maintains a gentle oscillatory nature as $n$ increases and $p$ is held fixed; a plot when $p = .5$ is given in Figure 6.1 to illustrate this oscillation.

6.4 Geometric and Negative Binomial Distributions

Again, it is helpful to begin with some examples.

Example 6.8 (Family Planning). In some economically disadvantaged countries, a male child is considered necessary to help with physical work and family finances. Suppose a couple will have children until they have had two boys. Let $X$ be the number of children they will have. Then, $X \sim \text{NB}(r, p)$, with $r = 2, p = .5$ (assumed). Thus, $X$ has the pmf

$$P(X = x) = (x - 1)(.5)^x, x = 2, 3, \ldots$$
For example, $P(\text{The couple will have at least one girl}) = P(X \geq 3) = 1 - P(X = 2) = 1 - .25 = .75$. The probabilities of some values of $X$ are given in the following table:

<table>
<thead>
<tr>
<th>$x$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x)$</td>
<td>.25</td>
<td>.25</td>
<td>.1875</td>
<td>.125</td>
<td>.0781</td>
<td>.0469</td>
<td>.0273</td>
</tr>
</tbody>
</table>

For example, $P(X \geq 6) = 1 - P(X \leq 5) = 1 - (.25 + .25 + .1875 + .125) \approx .19$. It is surprising that nearly 19% of such couples will have six or more children!

**Example 6.9 (Meeting Someone with the Same Birthday).** Suppose you were born on October 15. How many different people do you have to meet before you find someone who was also born on October 15? Under the usual conditions of equally likely birthdays and independence of the birthdays of all people you will meet, the number of people $X$ you have to meet to find the first person with the same birthday as yours is geometric; i.e., $X \sim Geo(p)$ with $p = \frac{1}{365}$. The pmf of $X$ is $P(X = x) = p(1 - p)^{x-1}$. Thus, for any given $k$,

$$P(X > k) = \sum_{x=k+1}^{\infty} p(1 - p)^{x-1} = p \sum_{x=k}^{\infty} (1 - p)^x = (1 - p)^k.$$  

For example, the chance that you will have to meet more than 1000 people to find someone with the same birthday as yours is $(364/365)^{1000} = .064$. But, of course, you will usually not ask people you meet what their birthday is, so it may be hard to verify experimentally that you should not need to meet 1000 people.

**Example 6.10.** Suppose a door-to-door salesman makes an actual sale in 25% of the visits he makes. He is supposed to make at least two sales per day. How many visits should he plan on making to be 90% sure of making at least two sales?

Let $X$ be the visit at which the second sale is made. Then, $X \sim NB(r, p)$ with $r = 2$, $p = .25$. Therefore, $X$ has the pmf $P(X = x) = (x - 1)(.25)^2(.75)^{x-2}$, $x = 2, 3, \ldots$. Summing, for any given $k$, $P(X > k) = \sum_{x=k+1}^{\infty} (x - 1)(.25)^2(.75)^{x-2} = \frac{k+3}{3}(3/4)^k$ (try to derive this). We want $\frac{k+3}{3}(3/4)^k \leq .1$. By computing this directly, we find that $P(X > 15) < 1$ but $P(X > 14) > .1$. So, the salesman should plan on making 15 visits.

**Example 6.11 (Lack of Memory of Geometric Distribution).** Let $X \sim Geo(p)$, and suppose $m$ and $n$ are given positive integers. Then, $X$ has the interesting property

$$P(X > m + n | X > n) = P(X > m).$$

That is, suppose you are waiting for some event to happen for the first time. You have tried, say, 20 times, and you still have not succeeded. You may feel that it is due anytime now. The lack of memory property would say that $P(X > 30 | X > 20) = P(X > 10)$. That is, the chance that it will take another ten tries is the same as what it would be if you had just started, and forget that you have already been patient for a long time and have tried hard for a success.
The proof is simple. Indeed,

\[
P(X > m + n | X > n) = \frac{P(X > m + n)}{P(X > n)} = \frac{\sum_{x>m+n} p(1-p)^{x-1}}{\sum_{x>n} p(1-p)^{x-1}}
\]

\[
= \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = P(X > m).
\]

We now give some important formulas for the geometric and negative binomial distributions.

**Theorem 6.4.**

(a) Let \( X \sim \text{Geo}(p) \). Let \( q = 1 - p \). Then,

\[
E(X) = \frac{1}{p}; \ Var(X) = \frac{q}{p^2}.
\]

(b) Let \( X \sim \text{NB}(r, p), r \geq 1 \). Then,

\[
E(X) = \frac{r}{p}; \ Var(X) = \frac{rq}{p^2}.
\]

Furthermore, the mgf and the (probability) generating function of \( X \) equal

\[
\psi(t) = \left( \frac{pe^t}{1 - qe^t} \right)^r, t < \log \left( \frac{1}{q} \right);
\]

\[
G(s) = \left( \frac{ps}{1 - qs} \right)^r, s < \frac{1}{q}.
\]

**Proof.** The formula for the mean and the variance of the geometric distribution follows by simply performing the sums. For example,

\[
E(X) = \sum_{x \geq 1} xpq^{x-1} = p \sum_{x \geq 1} xq^{x-1} = p \times \frac{1}{(1-q)^2} = p \times \frac{1}{p^2} = \frac{1}{p}.
\]

To find the variance, find the second moment by summing \( \sum_{x \geq 1} x^2 pq^{x-1} \), and then plug into the variance formula \( \text{Var}(X) = E(X^2) - [E(X)]^2 \). It would be easier to find the second moment by first finding the factorial moment \( E[X(X - 1)] \) and then use the fact that \( E(X^2) = E[X(X - 1)] + E(X) \). We omit the algebra.

The mean and the variance for the general negative binomial follow from the geometric case on using the very useful representation

\[
X = X_1 + X_2 + \cdots + X_r,
\]

where \( X_i \) is the geometric random variable measuring the number of additional trials needed to obtain the \( i \)th success after the \((i - 1)\)th success has been obtained. Thus, the \( X_i \) are independent, and each is distributed as \( \text{Geo}(p) \). So, their
variance can be obtained by summing the variances of $X_1, X_2, \ldots, X_r$, which gives $\text{Var}(X) = \sum_{i=1}^{r} \frac{q}{p} = \frac{rq}{p^2}$, and the expectation of course also adds up, to give $E(X) = \frac{r}{p}$.

The formula for the mgf of the geometric distribution is immediately obtained by summing \( \sum_{x \geq 1} e^{tx}pq^{x-1} = \frac{p}{q} \sum_{x \geq 1} (qe^t)^x = \frac{p}{q} \frac{qe^t}{1-qe^t} = \frac{pe^t}{1-qe^t} \). The formula for the negative binomial distribution follows from this formula by representing $X$ as $X_1 + X_2 + \cdots + X_r$ as above. Finally, the (probability) generating function is derived by following exactly the same steps.

### 6.5 Hypergeometric Distribution

As we mentioned, the hypergeometric distribution arises when sampling without replacement from a finite population consisting of elements of just two types. Here are some illustrative examples.

**Example 6.12 (Gender Discrimination).** From a pool of five male and five female applicants, three were selected and all three happened to be men. Is there a priori evidence of gender discrimination?

If we let $X$ be the number of female applicants selected, then $X \sim \text{Hypergeo}(n, D, N)$, with $n = 3, D = 5, N = 10$. Therefore, $P(X = 0) = \frac{\binom{D}{0} \binom{N-D}{n}}{\binom{N}{n}} = \frac{\binom{5}{3}}{\binom{10}{3}} = \frac{1}{12}$.

So, if selection was done at random, which should be the policy if all applicants are equally qualified, then selecting no women is a low-probability event. There might be some a priori evidence of gender discrimination.

**Example 6.13 (Bridge).** Suppose North and South together received no aces at all in three consecutive bridge plays. Is there a reason to suspect that the distribution of cards is not being done at random?

Let $X$ be the number of aces in the hands of North and South combined in one play. Then, $P(X = 0) = \frac{\binom{48}{13} \binom{35}{13}}{\binom{52}{13}} = \frac{46}{833} = .0552$.

Therefore, the probability of North and South not receiving any aces for three consecutive plays is $(.0552)^3 = .00017$, which is very small. Either an extremely rare event has happened or the distribution of cards has not been random. Statisticians call this sort of calculation a *p-value calculation* and use it to assess doubt about some proposition, in this case randomness of the distribution of the cards.
Example 6.14 (A Classic Example: Capture-Recapture). An ingenious use of the hypergeometric distribution in estimating the size of a finite population is the capture-recapture method. It was originally used for estimating the total number of fish in a body of water, such as a pond. Let \( N \) be the number of fish in the pond. In this method, a certain number of fish, say \( D \) of them are initially captured and tagged with a safe mark or identification device and then returned to the water. Then, a second sample of \( n \) fish is recaptured from the water. Assuming that the fish population has not changed in any way in the intervening time and that the initially captured fish remixed with the fish population homogeneously, the number of fish in the second sample, say \( X \), that bear the mark is a hypergeometric random variable, namely \( X \sim \text{Hypergeo}(n, D, N) \). We will shortly see that the expected value of a hypergeometric random variable is \( np \). If we set as a formalism \( X = np \) and solve for \( N \), we get \( N = \frac{nD}{X} \). This is an estimate of the total number of fish in the pond. Although the idea is extremely original, this estimate can run into various kinds of difficulties if, for example, the first catch of fish clusters around after being returned, hides, or if the fish population has changed between the two catches due to death or birth, and of course if \( X \) turns out to be zero. Modifications of this estimate (known as the Petersen estimate) are widely used in wildlife estimation, taking a census, and by the government for estimating tax fraud and the number of people afflicted with some infection.

The mean and variance of a hypergeometric distribution are given in the next result.

Theorem 6.5. Let \( X \sim \text{Hypergeo}(n, D, N) \) and let \( p = \frac{D}{N} \). Then,

\[
E(X) = np; \quad \text{Var}(X) = np(1 - p) \left( \frac{N - n}{N - 1} \right).
\]

We will not prove this result, as it involves the standard indicator variable argument we are familiar with and some routine algebra. Two points worth mentioning are that although sampling is without replacement in the hypergeometric case, so the Bernoulli trials are not independent, the same formula for the mean as in the binomial case holds. But the variance is smaller than in the binomial case because the extra factor \( \frac{N-n}{N-1} < 1 \). Sampling without replacement makes the composition of the sample more like the composition of the entire population, and this reduces the variance around the population mean. The factor \( \frac{N-n}{N-1} \) is often called the finite population correction factor.

Problems that should truly be modeled as hypergeometric distribution problems are often analyzed as if they were binomial distribution problems. That is, the fact that samples have been taken without replacement is ignored, and one pretends that the successive draws are independent. When does it not matter that the dependence between the trials is ignored? Intuitively, we would think that if the population size \( N \) was large and neither \( D \) nor \( N - D \) was small, the trials would act like they are independent. The following theorem justifies this intuition.
Theorem 6.6 (Convergence of Hypergeometric to Binomial). Let $X = X_N \sim \text{Hypergeo}(n, D, N)$, where $D = D_N$ and $N$ are such that $N \to \infty$, $D_N \to p$, $0 < p < 1$. Then, for any fixed $n$ and for any fixed $x$,

$$P(X = x) = \binom{D}{x} \binom{N - D}{n - x} \binom{N}{n} \to \binom{n}{x} p^x (1 - p)^{n-x}$$

as $N \to \infty$.

This is proved by using Stirling’s approximation (which says that as $k \to \infty$, $k! \sim e^{-k} k^{k+1/2} \sqrt{2\pi}$) for each factorial term in $P(X = x)$ and then doing some algebra.

### 6.6 Poisson Distribution

As mentioned before, Poisson distributions arise as counts of events in fixed periods of time, fixed amounts of area or space, and as limits of binomial distributions for large $n$ and small $p$. The first thing to note, before we can work out examples, is that the single parameter $\lambda$ of a Poisson distribution is its mean; quite remarkably, $\lambda$ is also the variance of the distribution. We will write $X \sim \text{Poi}(\lambda)$ to denote a Poisson random variable. The distribution was introduced by Siméon Poisson (1838).

Theorem 6.7. Let $X \sim \text{Poi}(\lambda)$. Then,

(a) $E(X) = \text{Var}(X) = \lambda$.
(b) $E(X - \lambda)^3 = \lambda$; $E(X - \lambda)^4 = 3\lambda^2 + \lambda$.
(c) The mgf of $X$ equals

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

Proof. Although parts (a) and (b) can be proved directly, it is most efficient to derive them from the mgf. So, we first prove part (c):

$$\psi(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} P(X = x) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} [\lambda e^t]^x / x! = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

Therefore,

$$\psi'(t) = e^{\lambda(e^t - 1)} \lambda e^t,$$

$$\psi''(t) = \lambda e^t (1 + \lambda e^t) e^{\lambda(e^t - 1)},$$

$$\psi^{(3)}(t) = \lambda e^t (1 + 3\lambda e^t + \lambda^2 e^{2t}) e^{\lambda(e^t - 1)},$$

$$\psi^{(4)}(t) = \lambda e^t (1 + 7\lambda e^t + 6\lambda^2 e^{2t} + \lambda^3 e^{3t}) e^{\lambda(e^t - 1)}.$$
From these, by using the fact that \( E(X^k) = \psi^{(k)}(0) \), we get

\[
E(X) = \lambda; \ E(X^2) = \lambda + \lambda^2; \ E(X^3) = \lambda(1 + 3\lambda + \lambda^2);
\]

\[
E(X^4) = \lambda(1 + 7\lambda + 6\lambda^2 + \lambda^3).
\]

The formulas in parts (a) and (b) now follow by simply plugging in the expressions given above for the first four moments of \( X \).

**Corollary 6.3.** The skewness and kurtosis of \( X \) equal

\[
\beta = \frac{1}{\sqrt{\lambda}}; \ \gamma = \frac{1}{\lambda}.
\]

The corollary follows immediately by using the definitions of skewness and kurtosis.

Let us now see some illustrative examples. The appendix gives a table of Poisson probabilities for \( \lambda \) between .5 and 5. These may be used instead of manually calculating the probabilities whenever the required probability can be obtained from the table given in the appendix.

**Example 6.15 (Events over Time).** April receives three phone calls at her home on average per day. On what percentage of days does she receive no phone calls? More than five phone calls?

Because the number of calls received in a 24 hour period counts the occurrences of an event in a fixed time period, we model \( X = \) number of calls received by April on one day as a Poisson random variable with mean 3. Then,

\[
P(X = 0) = e^{-3} = .0498; \ P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{x=0}^{5} e^{-3} \frac{3^x}{x!}
\]

\[
= 1 - .9161 = .0839.
\]

Thus, she receives no calls on 4.98% of the days and more than five calls on 8.39% of the days. **It is important to understand that \( X \) has only been modeled as a Poisson random variable, and other models could also be reasonable.**

**Example 6.16.** Lengths of an electronic tape contain, on average, one defect per 100 ft. If we need a tape of 50 ft., what is the probability that it will be defect-free?

Let \( X \) denote the number of defects per 50 ft. of this tape. We can think of lengths of the tape as a window of time, although not in a literal sense. If we assume that the defective rate is homogeneous over the length of the tape, then we can model \( X \) as \( X \sim \text{Poi}(\frac{1}{2}) \). That is, if 100 ft. contain one defect on average, then 50 ft. of tape should contain half a defect on average. This can be made rigorous by using the concept of a homogeneous Poisson process.

Therefore,

\[
P(X = 0) = e^{-\frac{1}{2}} = .6065.
\]
Example 6.17 (Events over an Area). Suppose a 14 inch circular pizza has been baked with 20 pieces of barbecued chicken. At a party, you were served a $4 \times 4 \times 2$ (in inches) triangular slice. What is the probability that you got at least one piece of chicken?

The area of a circle of radius 7 is $\pi \times 7^2 = 153.94$. The area of a triangular slice of lengths 4, 4, and 2 inches on a side is $\sqrt{s(s-a)(s-b)(s-c)} = \sqrt{5 \times 1 \times 1 \times 3} = \sqrt{15} = 3.87$, where $a, b, c$ are the lengths of the three sides and $s = (a + b + c)/2$. Therefore, we model $X$, the number of pieces of chicken in the triangular slice, as $X \sim \text{Poi}(\lambda)$, where $\lambda = 20 \times 3.87/153.94 = .503$. Using the Poisson pmf,

$$P(X \geq 1) = 1 - e^{-.503} = .395.$$  

Example 6.18 (A Hierarchical Model with a Poisson Base). Suppose a chick lays a $\text{Poi}(\lambda)$ number of eggs in some specified period of time, say a month. Each egg has a probability $p$ of actually developing. We want to find the distribution of the number of eggs that actually develop during that period of time.

Let $X \sim \text{Poi}(\lambda)$ denote the number of eggs the chick lays and $Y$ the number of eggs that develop. For example,

$$P(Y = 0) = \sum_{x=0}^{\infty} P(Y = 0|X = x)P(X = x) = \sum_{x=0}^{\infty} (1 - p)^x \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda(1 - p))^x}{x!} = e^{-\lambda} e^{\lambda(1 - p)} = e^{-p\lambda}.$$  

In general,

$$P(Y = y) = \sum_{x=y}^{\infty} \binom{x}{y} p^y (1 - p)^{x-y} \frac{e^{-\lambda} \lambda^x}{x!} = \frac{(p/(1 - p))^y}{y!} e^{-\lambda} \sum_{x=y}^{\infty} \frac{1}{(x-y)!} (1 - p)^x \lambda^x = \frac{(p/(1 - p))^y}{y!} e^{-\lambda} (\lambda(1 - p))^y \sum_{n=0}^{\infty} \frac{(\lambda(1 - p))^n}{n!} = \frac{(\lambda p)^y}{y!} e^{-\lambda} e^{\lambda(1 - p)} = \frac{e^{-\lambda p} (\lambda p)^y}{y!},$$

on writing $n = x - y$ in the summation, so we recognize by inspection that $Y \sim \text{Poi}(\lambda p)$. What is interesting here is that the distribution of $Y$ still remains Poisson under assumptions that seem to be very realistic physically.

Example 6.19 (Meteor Showers). Between the months of May and October, you can see a shooting star at the rate of about one per 20 minutes. If you sit on your patio
for one hour each evening, how many days would it be before you see ten or more
shooting stars on the same day?

This example combines the Poisson and the geometric distributions in an inter-
esting way. Let \( p \) be the probability of seeing ten or more shooting stars in one day. If we let \( N \) denote the number of shooting stars observed in one day and model \( N \) as \( N \sim \text{Poi}(\lambda) \), with \( \lambda = 3 \) (since one hour is equal to three 20 minute intervals), then

\[
p = P(N \geq 10) = \sum_{x=10}^{\infty} \frac{e^{-3}3^x}{x!} = .0011.
\]

Now, if we let \( X \) denote the number of days that you have to watch the sky until you see this shower of ten or more shooting stars, then \( X \sim \text{Geo}(p) \) and therefore \( E(X) = \frac{1}{p} = 909.1 \), which is about 30 months. You are observing for six months each year because there are six months between May and October (inclusive). So, you can expect that if you observe for about five years, you will see a shower of ten or more shooting stars on some evening.

**Example 6.20 (Poisson Forest).** It is a common assumption in forestry and ecology that the number of plants in a part of a forest is distributed according to a Poisson distribution with mean proportional to the area of the part of the forest.

Suppose on average there are ten trees per 100 square ft. in a forest. An ento-
mologist is interested in estimating an insect population in a forest of size 10,000
square ft. The insects are found in the trees, and it is believed that there are 100
of them per tree. The entomologist will cover a 900 square ft. area and count the
insects on all trees in that area. What are the chances that the entomologist will
discover more than 9200 insects in this area?

Suppose \( X \) is the number of trees in the 900 square ft. area the entomologist
covers, and let \( Y \) be the number of insects the entomologist discovers. We assume
that \( X \sim \text{Poi}(\lambda) \), with \( \lambda = 90 \). Then, because there are 100 insects per tree,

\[
P(Y > 9200) = P(X > 92) = \sum_{x=93}^{\infty} \frac{e^{-90}(90)^x}{x!} \approx .3898.
\]

The .3898 value was found by direct summation on a computer. A more realistic
model will assume the number of insects per tree is a random variable rather than
being constantly equal to 100. However, finding an answer to the question would
then be much harder.

**Example 6.21 (Gamma-Ray Bursts).** Gamma-ray bursts are thought to be the most
intense electromagnetic events observed in the sky, and they typically last a few
seconds. While they are on, their intense brightness covers up any other gamma-ray
source in the sky. They occur at the rate of about one episode per day. It was initially
thought that they were events within the Milky Way galaxy, but most astronomers
now believe that is not true or not entirely true.
The 2000th gamma-ray burst since 1991 was detected at the end of 1997 at NASA’s Compton Gamma Ray Observatory. Are these data compatible with a model of a Poisson-distributed number of bursts with a rate of one per day?

Using a model of homogeneously distributed events, the number of bursts in a seven-year period is \( \textit{Poi}(\lambda) \) with  
\[
\lambda = 7 \times 365 \times 1 = 2555.
\]
The observed number of bursts is 2000, less than the expected number of bursts. But is it so much less that the postulated model is in question? To assess this, we calculate \( P(X \leq 2000) \), the probability that we could observe an observation as deviant from the expected one as we did just by chance. Statisticians call such a deviation probability a \textit{p-value}. The \( p \)-value then equals

\[
P(\textit{X} \leq 2000) = \sum_{x=0}^{2000} \frac{e^{-2555}(2555)^x}{x!}.
\]

Due to the large values of \( \lambda \) and the range of the summation, directly summing this is not recommended. But the sum can be approximated by using various other indirect means, including a theorem known as the \textit{central limit theorem}, which we will later discuss in detail. The approximate \( p \)-value can be seen to be extremely small, virtually zero. So, the chance of such a deviant observation, if the Poisson model at the rate of one burst per day was correct, is very small. One would doubt the model in such a case. The bursts may not occur at a homogeneous rate of one per day.

\section*{6.6.1 Mean Absolute Deviation and the Mode}

Similar to the binomial case, a closed-form formula is available for the mean absolute deviation \( E[|X - \lambda|] \) of a Poisson distribution; we can also characterize the mode; i.e., the value with the largest probability. Again, see Diaconis and Zabell (1991) for these results.

\textbf{Theorem 6.8 (Mean Absolute Deviation and Mode).} Let \( \textit{X} \sim \textit{Poi}(\lambda) \). Then:

(a) \( E[|X - \lambda|] = 2\lambda P(X = [\lambda]) \).

(b) A Poisson distribution is unimodal and \( P(X = k) \leq P(X = [\lambda]) \) \( \forall k \geq 0 \).

\textit{Proof.} Part (a) requires nontrivial calculations; see Diaconis and Zabell (1991). Part (b), however, is easy to prove. Consider the ratio

\[
\frac{P(X = k + 1)}{P(X = k)} = \frac{\lambda}{k + 1},
\]

and note that this is \( \geq 1 \) if and only if \( k + 1 \leq [\lambda] \), which proves that \( [\lambda] \) is always a mode. If \( \lambda \) is an integer, then \( \lambda \) and \( \lambda - 1 \) will both be modes; that is, there would be two modes. If \( \lambda \) is not an integer, then \( [\lambda] \) is the unique mode.
6.7 Poisson Approximation to Binomial

A binomial random variable is the sum of \( n \) indicator variables. When the expectation of these indicator variables, namely \( p \), is small, and the number of summands \( n \) is large, the Poisson distribution provides a good approximation to the binomial. The Poisson distribution can also sometimes serve as a good approximation when the indicators are independent but have different expectations \( p_i \), or when the indicator variables have some weak dependence. We will start with the Poisson approximation to the binomial when \( n \) is large and \( p \) is small.

**Theorem 6.9.** Let \( X_n \sim \text{Bin}(n, p_n), n \geq 1 \). Suppose \( np_n \to \lambda, 0 < \lambda < \infty \), as \( n \to \infty \). Let \( Y \sim \text{Poi}(\lambda) \). Then, for any given \( k, 0 \leq k < \infty \),

\[
P(X_n = k) \to P(Y = k)
\]
as \( n \to \infty \).

**Proof.** For ease of explanation, let us first consider the case \( k = 0 \). We have

\[
P(X_n = 0) = (1 - p)^n = \left(1 - \frac{np}{n}\right)^n \sim \left(1 - \frac{\lambda}{n}\right)^n \sim e^{-\lambda}.
\]
Note that we did not actually prove the claimed fact that \((1 - \frac{np}{n})^n \sim (1 - \frac{\lambda}{n})^n\), but it is true and is not hard to prove.

Now consider \(k = 1\). We have

\[
P(X_n = 1) = np(1 - p)^{n-1} = (np)(1 - p)^n \frac{1}{1 - p} \sim \lambda(e^{-\lambda})(1) = \lambda e^{-\lambda}.
\]

The same technique works for any \(k\). Indeed, for a general \(k\),

\[
P(X_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

\[
= \frac{1}{k!} [n(n-1) \cdots (n-k+1)] p^k (1 - p)^n \left[ \frac{1}{(1 - p)^k} \right]
\]

\[
= \frac{1}{k!} n^k \left[ \frac{n-1}{n} \cdots \frac{n-k+1}{n} \right] p^k (1 - p)^n \left[ \frac{1}{(1 - p)^k} \right]
\]

\[
= \frac{1}{k!} (np)^k \left[ \frac{n-1}{n} \cdots \frac{n-k+1}{n} \right] (1 - p)^n \left[ \frac{1}{(1 - p)^k} \right]
\]

\[
\sim \frac{1}{k!} (\lambda)^k [1] e^{-\lambda} [1] = \frac{e^{-\lambda} \lambda^k}{k!},
\]

which is what the theorem says.

In fact, the convergence is not just pointwise for each fixed \(k\) but is *uniform* in \(k\). This will follow from the following more general theorem, which we state for reference (see Le Cam, 1960; Barbour and Hall, 1984; Steele, 1994)

**Theorem 6.10 (Le Cam, Barbour and Hall, Steele).** Let \(X_n = B_1 + B_2 + \cdots + B_n\), where \(B_i\) are independent Bernoulli variables with parameters \(p_i = p_{i,n}\). Let \(Y_n \sim \text{Poi}(\lambda)\), where \(\lambda = \lambda_n = \sum_{i=1}^{n} p_i\). Then,

\[
\sum_{k=0}^{\infty} |P(X_n = k) - P(Y_n = k)| \leq 2 \frac{1 - e^{-\lambda}}{\lambda} \sum_{i=1}^{n} p_i^2.
\]

Here are some more examples of the Poisson approximation to the binomial.

**Example 6.22 (Lotteries).** Consider a weekly lottery in which three numbers out of 25 are selected at random and a person holding exactly those three numbers is the winner of the lottery. Suppose the person plays for \(n\) weeks, for large \(n\). What is the probability that he will win the lottery at least once? At least twice?

Let \(X\) be the number of weeks that the player wins. Then, assuming the weekly lotteries are independent, \(X \sim \text{Bin}(n, p)\), where \(p = \frac{1}{25} = \frac{1}{2300} = 0.00043\). Since \(p\) is small and \(n\) is supposed to be large, \(X \approx \text{Poi}(\lambda), \lambda = np = 0.00043n\). Therefore,
6.7 Poisson Approximation to Binomial

\[ P(X \geq 1) = 1 - P(X = 0) \approx 1 - e^{-0.00043n} \]

and

\[ P(X \geq 2) = 1 - P(X = 0) - P(X = 1) \approx 1 - e^{-0.00043n} - 0.00043ne^{-0.00043n} \]

\[ = 1 - (1 + .00043n)e^{-0.00043n}. \]

We can compute these for various \( n \). If the player plays for five years,

\[ 1 - e^{-0.00043n} = 1 - e^{-0.00043 \times 5 \times 52} = .106 \]

and

\[ 1 - (1 + .00043n)e^{-0.00043n} = .006. \]

If he plays for ten years,

\[ 1 - e^{-0.00043n} = 1 - e^{-0.00043 \times 10 \times 52} = .200 \]

and

\[ 1 - (1 + .00043n)e^{-0.00043n} = .022. \]

We can see that the chances of any luck are at best moderate even after pro-
longed tries.

**Example 6.23 (An Insurance Example).** Suppose 5000 clients are each insured for one million dollars against fire damage in a coastal property. Each residence has a 1 in 10,000 chance of being damaged by fire in a 12 month period. How likely is it that the insurance company has to pay out as much as 3 million dollars in fire damage claims in one year? Four million dollars?

If \( X \) is the number of claims made during a year, then \( X \sim Bin(n, p) \) with \( n = 5000 \) and \( p = 1/10,000 \). We assume that no one makes more than one claim and the clients are independent. Then we can approximate the distribution of \( X \) by \( Poi(np) = Poi(.5) \). We need

\[ P(X \geq 3) = 1 - P(X \leq 2) \approx 1 - (1 + .5 + .5^2 / 2)e^{-.5} = .014 \]

and

\[ P(X \geq 4) = 1 - P(X \leq 3) \approx 1 - (1 + .5 + .5^2 / 2 + .5^3 / 6)e^{-.5} = .002. \]

These two calculations are done above by using the Poisson approximation, namely \( e^{-5 \cdot \frac{k^k}{k!}} \), for \( P(X = k) \). The insurance company is quite safe being prepared for 3 million dollars in payout and very safe being prepared for 4 million dollars.
6.8 * Miscellaneous Poisson Approximations

A binomial random variable is the sum of independent and identically distributed Bernoulli variables. Poisson approximations are also often accurate when the individual Bernoulli variables are independent but have small and different parameters \( p_i \) or when the Bernoulli variables have a weak dependence. A rule of thumb is that if the individual \( p_i \)'s are small and their sum is moderate, then a \( \text{Poi}(\sum p_i) \) approximation should be accurate. There are many rigorous theorems in this direction. There are the first-generation Poisson approximation theorems and the more modern Poisson approximation theorems, that go by the name of the Stein-Chen method. The Stein-Chen method is now regarded as the principal tool for approximating the distribution of sums of weakly dependent Bernoulli variables, with associated bounds on the error of the approximation. The two original papers are Stein (1972) and Chen (1975). More recent sources with modern applications in a wide variety of fields are Barbour et al. (1992) and Diaconis and Holmes (2004).

We will first work out a formal Poisson approximation in some examples below.

**Example 6.24 (Poisson Approximation in the Birthday Problem).** In the birthday problem, \( n \) unrelated people gather around and we want to know if there is at least one pair of individuals with the same birthday. Defining \( I_{i,j} \) as the indicator of the event that individuals \( i \) and \( j \) have the same birthday, we have

\[
X = \text{number of different pairs of people who share a common birthday} = \sum_{1 \leq i < j \leq n} I_{i,j}.
\]

Each \( I_{i,j} \sim \text{Ber}(p) \), where \( p = 1/365 \). Note, however, that the \( I_{i,j} \) are definitely not independent. Now, the expected value of \( X \) is \( \lambda = \binom{n}{2}/365 \). This is moderate (\( > .5 \)) if \( n \geq 20 \). So, a Poisson approximation may be accurate when \( n \) is about 20 or more.

If we use a Poisson approximation when \( n = 23 \), we get

\[
P(X > 0) \approx 1 - e^{-\binom{23}{2}/365} = 1 - e^{-693151} = .500002,
\]

which is almost exactly equal to the true value of the probability that there will be a pair of people with the same birthday in a group of 23 people; this was previously discussed in Chapter 2.

**Example 6.25 (Three People with the Same Birthday).** Consider again a group of \( n \) unrelated people, and ask what the chances are that we can find three people in the group with the same birthday. We proceed as in the preceding example. Define \( I_{i,j,k} \) as the indicator of the event that individuals \( i, j, k \) have the same birthday. Then, \( I_{i,j,k} \sim \text{Ber}(p) \), \( p = 1/(365)^2 \). Let

\[
X = \sum_{1 \leq i < j < k \leq n} I_{i,j,k}
\]
The expected value of $X$ is $\lambda = \binom{n}{3}/365^2$. We want to approximate $P(X \geq 1)$. If we use the Poisson approximation $X \sim \text{Poi}(\binom{n}{3}/365^2)$, we get with $n = 84$

$$P(X \geq 1) \approx 1 - e^{-\left(\frac{84}{3}\right)/365^2} = 1 - e^{-0.715211} = .5109.$$  

In fact, $n = 84$ is truly the first $n$ for which the probability that we can find three people with the same birthday exceeds .5. We again see the effectiveness of the Poisson approximation in approximating sums of dependent Bernoulli variables. Note how much harder it is to find three people with the same birthday than it was to find two!

A reasonably simple first-generation theorem on the validity of the Poisson approximation for suitable sums of (not necessarily independent) Bernoulli variables can be described by using the so-called binomial moments of the sum. We will first define the term binomial moment.

**Definition 6.2.** Let $X$ be a nonnegative integer-valued random variable with a finite $j$th moment for a given $j \geq 1$. The $j$th binomial moment of $X$ is defined as $M_j = E\left[\binom{X}{j}\right] = \sum_{x=j}^{\infty} \binom{x}{j} P(X = x)$.

**Remark.** Note that the binomial moments and the factorial moments are related as $M_j = E[X(X-1)\cdots(X-j+1)]!$, thus the $j$th binomial moment is finite if and only if the $j$th factorial moment is finite, which is true if and only if the $j$th moment is finite.

We give an example.

**Example 6.26 (Factorial Moments of Poisson).** Let $X \sim \text{Poi}(\lambda)$. Let $n \geq 1$. Then,

$$E[X(X-1)\cdots(X-n+1)] = \sum_{x=0}^{\infty} x(x-1)\cdots(x-n+1) \frac{e^{-\lambda}\lambda^x}{x!}$$

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

$$= e^{-\lambda} \sum_{x=n}^{\infty} \frac{\lambda^x}{(x-n)!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x+n}}{x!} = e^{-\lambda} \lambda^n \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \lambda^n \times e^\lambda = \lambda^n,$$

a remarkably pretty result.

In some problems, typically of a combinatorial nature, careful counting lets one use the binomial moments and establish the validity of a Poisson approximation.
Here is a theorem that shows how to do it; see Galambos and Simonelli (1996) for a proof of it.

**Theorem 6.11 (Basic General Poisson Approximation Theorem).** Let $X_n = B_1 + B_2 + \cdots + B_n$, where $B_i$ are Bernoulli random variables. Let $M_k = M_{k,n}$ be the $k$th binomial moment of $X_n$. If there exists $0 < \lambda < \infty$ such that, for every fixed $k$, $M_k \to \frac{\lambda^k}{k!}$ as $n \to \infty$, then $P(X_n = j) \to \frac{e^{-\lambda} \frac{\lambda^j}{j!}}{j}$ for any $j$ as $n \to \infty$.

Here is an application of this theorem.

**Example 6.27 (The Committee Problem).** From $n$ people, $N = N(n)$ committees are formed, each committee of a fixed size $m$. We let $N, n \to \infty$, holding $m$ fixed. The Bernoulli variable $B_i = B_{i,n}$ is the indicator of the event that the $i$th person is not included in any committee. The purpose of this example is to derive a Poisson approximation for $X$, the total number of people who are not included in any committee.

Under the usual assumptions of independence and also the assumption of random selection, the binomial moment $M_k$ can be shown to be

$$M_k = \binom{n}{k} \left[ \frac{(n-k)}{m} \right]^{N}.$$

Stirling’s approximation now shows that $M_k \sim \frac{n^k}{k!} e^{-kN \left( \frac{m}{n} + O(n^{-2}) \right)}$ as $n \to \infty$. One now sees, on inspection, that if $N, n$ are related as $N = \frac{n \log n}{m} - n \log \lambda + o(n^{-1})$ for some $0 < \lambda < \infty$, then $M_k \to \frac{\lambda^m}{k!}$, so from the basic general Poisson approximation theorem above, the number of people who are left out of all committees converges to $Poi(\lambda^m)$.

### 6.9 Benford’s Law

Benford’s law asserts that if we pick numbers at random from a statistical data set or mathematical tables such as a table of logarithms or a table of physical constants, then the leading digit tends to be 1 with much greater frequency than the 11.1% one would expect if the distribution was just discrete uniform on $\{1, 2, \ldots, 9\}$. The law was first asserted by the astronomer Simon Newcomb (1881). However, the distribution has come to be known as the Benford distribution, attributable to a publication by Frank Benford (1938). The distribution was later found to give quite reasonable fits to various kinds of data, such as the first digit in randomly picked home addresses, daily returns of stocks, leading digits in geological variables, baseball statistics, half-lives of radioactive particles, etc.
The Benford distribution is a distribution on \{1, 2, \ldots, 9\} with the pmf

\[
p(x) = \frac{\log(x + 1) - \log x}{\log 10}, \quad x = 1, 2, \ldots, 9.
\]

Note that clearly \(p(x) \geq 0\) for any \(x\) and \(\sum_{x=1}^{9} p(x) = \frac{1}{\log 10} \left[ \log 2 - \log 1 + \log 3 - \log 2 + \cdots + \log 10 - \log 9 \right] = 1\). Therefore, it is a valid pmf. The numerical values of the probabilities are as follows:

<table>
<thead>
<tr>
<th>(x)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p(x))</td>
<td>.301</td>
<td>.176</td>
<td>.125</td>
<td>.097</td>
<td>.079</td>
<td>.067</td>
<td>.058</td>
<td>.051</td>
<td>.046</td>
</tr>
</tbody>
</table>

The moments of the distribution are easily calculated. In particular, the mean and the variance equal 3.44 and 6.0565. The distribution is right skewed, and the coefficient of skewness equals .796.

Contemporary literature on the Benford distribution includes Hill (1996), Diaconis (1977), and Diaconis and Freedman (1979). Benford’s distribution is a very simple distribution from a purely mathematical point of view. Its appeal lies in its ability to give mysteriously good fits to the leading digit for diverse types of empirical data. A recent discovery is that the fits are better when data from apparently unrelated sources are combined. In other words, if a set of variables have their individual distributions and then those distributions are mixed, then the leading digit in the mixed distribution would often approximately follow the Benford law.

### 6.10 Distribution of Sums and Differences

Sums of random variables arise very naturally in practical applications. For example, the revenue over a year is the sum of the monthly revenues; the time taken to finish a test with ten problems is the sum of the times taken to finish the individual problems, etc. Likewise, the difference of two intrinsically similar random variables is also a natural quantity to study; e.g., the number of crimes of some specific kind committed last year and the number committed this year. It is also interesting to look at the absolute difference of similar random variables in addition to the difference itself.

Sometimes we can reasonably assume that the various random variables being added are independent. Thus, the following general question is an important one. Suppose \(X_1, X_2, \ldots, X_k\) are \(k\) independent random variables and that we know the distributions of the individual \(X_i\). What is the distribution of the sum \(X_1 + X_2 + \cdots + X_k\)?

In general, this is a very difficult question. Interestingly, if the individual \(X_i\) have one of the distinguished distributions we have discussed in this chapter, then their sum is also often a distribution of that same type. For example, sums of independent Poisson random variables would be Poisson also. This loyalty to types is a very useful fact, and we present a theorem in this regard below.
Theorem 6.12. \(\text{(a) Suppose } X_1, X_2, \ldots, X_k \text{ are } k \text{ independent binomial random variables with } X_i \sim \text{Bin}(n_i, p). \text{ Then } X_1 + X_2 + \ldots + X_k \sim \text{Bin}(n_1 + n_2 + \cdots + n_k, p).\)

\(\text{(b) Suppose } X_1, X_2, \ldots, X_k \text{ are } k \text{ independent negative binomial random variables with } X_i \sim \text{NB}(r_i, p). \text{ Then } X_1 + X_2 + \ldots + X_k \sim \text{NB}(r_1 + r_2 + \cdots + r_k, p).\)

\(\text{(c) Suppose } X_1, X_2, \ldots, X_k \text{ are } k \text{ independent Poisson random variables with } X_i \sim \text{Poi}(\lambda_i). \text{ Then } X_1 + X_2 + \ldots + X_k \sim \text{Poi}(\lambda_1 + \lambda_2 + \cdots \lambda_k).\)

**Proof.** Each of the three parts can be proved by various means. One possibility is to attack the problem directly. Alternatively, the results can also be proved by using generating functions or mgfs. It is useful to see a proof using both methods, so we do this for the Poisson case. The proof for the other two cases is exactly the same and will be omitted.

First, note that it is enough to consider only the case \(k = 2\) because then the general case follows by induction. We denote \(X_1, X_2\) as \(X, Y\) for notational simplicity. Then,

\[
P(X + Y = z) = \sum_{x=0}^{z} P(X = x, Y = z - x) = \sum_{x=0}^{z} P(X = x)P(Y = z - x)
\]

\[
= \sum_{x=0}^{z} \frac{e^{-\lambda_1} \lambda_1^x}{x!} \frac{e^{-\lambda_2} \lambda_2^{z-x}}{(z-x)!}
\]

\[
= e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_1^z}{z!} \sum_{x=0}^{z} \frac{(\lambda_1/\lambda_2)^x}{x!(z-x)!}
\]

\[
= e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_2^z}{z!} \sum_{x=0}^{z} \left(\frac{z}{x}\right) \frac{(\lambda_1/\lambda_2)^x}{x!}
\]

\[
= e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_2^z}{z!} \left(1 + \frac{\lambda_1}{\lambda_2}\right)^z = e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_1 + \lambda_2}{z!},
\]

as was required to prove.

The second method uses the formula for the mgf of a Poisson distribution. Since \(X\) and \(Y\) are both Poisson and they are independent, the mgf of \(X + Y\) is

\[
\psi_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}]E[e^{tY}] = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)}
\]

\[= e^{(\lambda_1 + \lambda_2)(e^t-1)},\]

which agrees with the mgf of the \(\text{Poi}(\lambda_1 + \lambda_2)\) distribution, and therefore, by the distribution-determining property of mgfs, the distribution of \(X + Y\) must be \(\text{Poi}(\lambda_1 + \lambda_2)\).
The calculation used in each of these two methods of proof is useful, and it is important to be familiar with each method.

**Example 6.28.** Suppose \( X \sim \text{Poi}(1), Y \sim \text{Poi}(5), \) and \( Z \sim \text{Poi}(10), \) and suppose \( X, Y, Z \) are independent. We want to find \( P(X + Y + Z \geq 20). \)

By the previous theorem, \( X + Y + Z \sim \text{Poi}(16), \) and therefore

\[
P(X + Y + Z \geq 20) = 1 - P(X + Y + Z \leq 19) = 1 - \sum_{x=0}^{19} \frac{e^{-16}16^x}{x!}
\]

\[
= 1 - .8122 = .1878.
\]

In the absence of the result that \( X + Y + Z \sim \text{Poi}(16), \) computing this probability would call for enumeration of all the ways that \( X + Y + Z \) could be 19 or smaller and adding up those probabilities. Clearly, it would be a much more laborious calculation.

### 6.10.1 Distribution of Differences

We now turn to differences of random variables of the same type; e.g., the difference of two independent Poisson random variables. Obviously, it cannot be Poisson, because it can take negative values. Similarly, the difference of two binomial random variables cannot be binomial because it will take negative values. Indeed, the distribution of differences is not nearly as nice or neat as the distribution of sums of variables of the same type. We present the Poisson case below; the binomial case is a chapter exercise.

**Theorem 6.13 (Difference of Independent Poissons).** Let \( X \) and \( Y \) be independent random variables, \( X \sim \text{Poi}(\lambda_1), Y \sim \text{Poi}(\lambda_2). \) Then,

\[
P(X - Y = z) = e^{-(\lambda_1 + \lambda_2)} \left( \frac{\lambda_1}{\lambda_2} \right)^{z/2} I_z(2\sqrt{\lambda_1\lambda_2}) \text{ if } z \geq 0,
\]

where \( I_n(x) \) is the modified Bessel function of order \( n \) with the power series expansion

\[
I_n(x) = \left( \frac{x}{2} \right)^n \sum_{k=0}^{\infty} \frac{x^{2k}}{4^k k!(n + k)!}.
\]

**Proof.** The formula for \( z \leq 0 \) follows from the formula for \( z \geq 0 \) by switching the roles of \( X \) and \( Y. \) So we only consider the case \( z \geq 0. \)
By the independence of $X$ and $Y$,

$$
P(X - Y = z) = \sum_{y=0}^{\infty} P(X = y + z, Y = y) = \sum_{y=0}^{\infty} P(X = y + z) P(Y = y)
$$

$$
= \sum_{y=0}^{\infty} e^{-\lambda_1} \frac{\lambda_1^{y+z}}{(y+z)!} e^{-\lambda_2} \frac{\lambda_2^y}{y!}
$$

$$
= e^{-(\lambda_1 + \lambda_2)} \sum_{y=0}^{\infty} \frac{(\lambda_1 \lambda_2)^y}{(y+z)! y!}
$$

$$
= e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_1^{z/2}}{(\lambda_1 \lambda_2)^{z/2}} I_z(2\sqrt{\lambda_1 \lambda_2})
$$

$$
= e^{-(\lambda_1 + \lambda_2)} \left( \frac{\lambda_1}{\lambda_2} \right)^{z/2} I_z(2\sqrt{\lambda_1 \lambda_2}),
$$

as was claimed.

Of course, directly, $E(X - Y) = \lambda_1 - \lambda_2$ and $\text{Var}(X - Y) = \lambda_1 + \lambda_2$.

### 6.11 * Discrete Does Not Mean Integer-Valued*

Although the named discrete distributions are all on integers, discrete random variables need not be integer-valued. Indeed, according to the definition we have provided in this text, a random variable taking values in any countable set is discrete. One naturally thinks of the rationals as a natural countable set after the set of integers. In particular, the rationals in the unit interval $[0,1]$ also form a countably infinite set. There are many ways to write reasonable distributions on rationals in the unit interval. We give one example.

**Example 6.29 (Distribution on Rationals in the Open Unit Interval).** Let $X$ and $Y$ be independent random variables, each distributed as geometric with parameter $p$, and let $R = \frac{X}{X+Y}$. Then clearly $R$ takes only rational values, and $P(0 < R < 1) = 1$ for any $p$.

The pmf can be investigated as follows. Let $r = \frac{m}{n}$, $0 < m < n$ be a rational in its irreducible form; i.e., $m$ and $n$ have no common factors. Then,

$$
P(R = r) = P(X = mk, X + Y = nk \text{ for some } k \geq 1)
$$

$$
= P(X = mk, Y = (n - m)k \text{ for some } k \geq 1)
$$

$$
= \sum_{k=1}^{\infty} P(X = mk) P(Y = (n - m)k)
$$
for all \( n \geq 2 \) and all \( m < n \) such that \( m \) is relatively prime to \( n \). Note that all such \( m \)’s, for a given \( n \), result in the same pmf value, as is seen in the formula above. In the special case \( p = \frac{1}{2} \), the pmf becomes

\[
P \left( R = \frac{m}{n} \right) = \frac{1}{2n-1}, n \geq 2, (m, n) = 1,
\]

where the \( (m, n) = 1 \) notation means that they are relatively prime. These probabilities would have to add to one. The number of \( m \)’s relatively prime to a given \( n \) is the so-called Euler totient function \( \phi(n) \). It is interesting that this example therefore shows the number-theoretic identity

\[
\sum_{n=2}^{\infty} \frac{\phi(n)}{2^n - 1} = 1.
\]

6.12 Synopsis

(a) If \( X \sim Bin(n, p) \) and \( q = 1 - p \), then

\[
P(X = x) = \binom{n}{x} p^x q^{n-x}, 0 \leq x \leq n; E(X) = np; \text{Var}(X) = npq.
\]

The mgf of \( X \) equals \( \psi(t) = (pe^t + q)^n \) at any \( t \). The integer part of \( np + p \) is always a mode of \( X \).

(b) If \( X \sim Geo(p) \) and \( q = 1 - p \), then

\[
P(X = x) = pq^{x-1}, x \geq 1; E(X) = \frac{1}{p}; \text{Var}(X) = \frac{q}{p^2}.
\]
More generally, if $X \sim NB(r, p), r \geq 1$, then

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x \geq r; E(X) = \frac{r}{p}; \text{Var}(X) = \frac{rq}{p^2}.$$

The mgf and the generating function of $X$ equal $\psi(t) = \left( \frac{pe^t}{1-qp} \right)^r, t < \log \left( \frac{1}{q} \right),$ and $G(s) = \left( \frac{ps}{1-qs} \right)^r, s < \frac{1}{q}$.

(c) If $X \sim Hypergeo(n, D, N), p = \frac{D}{N}$, and $q = 1 - p$, then

$$P(X = x) = \binom{D}{x} \binom{N-D}{n-x} \binom{N}{n}, n - N + D \leq x \leq D; E(X) = np; \text{Var}(X) = npq \left( \frac{N-n}{N-1} \right).$$

(d) The geometric distribution satisfies the lack of memory property

$$P(X > m + n | X > n) = P(X > m)$$

for any $m, n \geq 1$.

(e) If $X \sim Poi(\lambda)$, then

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, x \geq 0; E(X) = \text{Var}(X) = \lambda.$$

The integer part of $\lambda$ is always a mode of $X$. The mgf of $X$ equals $\psi(t) = e^{\lambda (e^t - 1)}$.

(f) If $X = X_n \sim Bin(n, p), n \to \infty, p = p_n \to 0$, and $np \to \lambda$ for some $\lambda, 0 < \lambda < \infty$, then, for any fixed $k$, $P(X_n = k) \to \frac{e^{-\lambda} \lambda^k}{k!}$ as $n \to \infty$.

(g) Suppose $X_1, X_2, \ldots, X_k$ are $k$ independent binomial random variables with $X_i \sim Bin(n_i, p)$. Then $X_1 + X_2 + \cdots + X_k \sim Bin(n_1 + n_2 + \cdots + n_k, p)$.

(h) Suppose $X_1, X_2, \ldots, X_k$ are $k$ independent negative binomial random variables, with $X_i \sim NB(r_i, p)$. Then $X_1 + X_2 + \cdots + X_k \sim NB(r_1 + r_2 + \cdots + r_k, p)$.

(i) Suppose $X_1, X_2, \ldots, X_k$ are $k$ independent Poisson random variables with $X_i \sim Poi(\lambda_i)$. Then $X_1 + X_2 + \cdots + X_k \sim Poi(\lambda_1 + \lambda_2 + \cdots + \lambda_k)$. 
6.13 Exercises

Exercise 6.1. Suppose a fair coin is tossed \( n \) times. Find the probability that exactly half of the tosses result in heads when \( n = 10, 30, 50 \). Where does the probability seem to converge as \( n \) becomes large?

Exercise 6.2. Suppose one coin with probability .4 for heads, one with probability .6 for heads, and one that is a fair coin are each tossed once. Find the pmf of the total number of heads obtained. Is it a binomial distribution?

Exercise 6.3. Suppose the IRS audits 5% of those having an annual income exceeding 200,000 dollars. What is the probability that at least one in a group of 15 such individuals will be audited? What is the expected number that will be audited? What is the most likely number of people who will be audited?

Exercise 6.4. Suppose that each day the price of a stock moves up 12.5 cents with probability 1/3 and moves down 12.5 cents with probability 2/3. If the movements of the stock from one day to another are independent, what is the probability that after ten days the stock has its original price?

Exercise 6.5. * (Pepy’s Problem). Find the probability that at least \( n \) sixes are obtained when \( 6n \) fair dice are rolled. Write a formula for it, and compute it for \( 1 \leq n \leq 5 \). Do you see a pattern in the values?

Exercise 6.6. In repeated rolling of a fair die, find the minimum number of rolls necessary in order for the probability of at least one six to be

(a) \( \geq .5 \).

(b) \( \geq .9 \).

Exercise 6.7. * In repeated rolling of a fair die, find the minimum number of rolls necessary in order for the probability of at least \( k \) sixes to be \( \geq .9 \) when \( k = 2, 3 \).

Exercise 6.8 (System Reliability). A communication system consists of \( n \) components, each of which will independently function with probability \( p \). The total system will be able to operate effectively if at least half of its components function. For what values of \( p \) is a five-component system more likely to operate effectively than a three-component system?

Exercise 6.9. * (A Waiting Time Problem). Tim, Jack, and John are going to have coffee at the local coffee shop. They will each toss a fair coin, and if one comes out as the “odd man,” then he pays for all three. They keep tossing until an odd man is found. What is the probability that a decision will be reached within two rounds of tosses?

Can you generalize this with \( n \) people, a general coin with probability \( p \) of heads, and the question being what the probability is that a decision will be reached within \( k \) rounds?
Exercise 6.10. A certain firm is looking for five qualified engineers to add to its staff. If from past experience it is known that only 20% of engineers applying for a position with this firm are judged to be qualified, what is the probability that the firm will interview exactly 40 applicants to fill the five positions? At least 40 applicants to fill the five positions?

Exercise 6.11. *(Distribution of Maximum).* Suppose \( n \) numbers are drawn at random from \( \{1, 2, \cdots, N\} \). What is the probability that the largest number drawn is a specified number \( k \) if sampling is (a) with replacement; (b) without replacement?

Exercise 6.12. *(Poisson Approximation).* One hundred people will each toss a fair coin 200 times. Approximate the probability that at least 10 of the 100 people would have obtained exactly 100 heads and 100 tails.

Exercise 6.13. *(A Design Problem).* You are allowed to choose a number \( n \) and then toss a fair coin \( n \) times. You will get a prize if you can get either seven or nine heads. What is your best choice of the number \( n \)?

Exercise 6.14. *(A Novel Way to Give a Test).* A student takes a five-answer multiple-choice oral test. His grade is determined by the number of questions required in order for him to get five correct answers. A grade of A is given if he requires only five questions; a grade of B is given if he requires six or seven questions; a grade of C is given if he requires eight or nine questions; and he fails otherwise.

Suppose the student guesses independently at random on each question. What is his most likely grade?

Exercise 6.15. A binomial random variable has mean 14 and variance 4.2. Find the probability that it is strictly larger than 10.

Exercise 6.16 *(Distribution of Sum).* The demand for the daily newspaper in a vending stall is distributed as \( \text{Bin}(20, .75) \) on weekdays and \( \text{Bin}(50, .75) \) on the weekend. Assuming that all days are independent, what is the distribution of the weekly demand?

Exercise 6.17 *(Distribution of Difference).* The demand for the daily newspaper on a Monday in a vending stall is distributed as \( \text{Bin}(20, .75) \) and that on a Sunday as \( \text{Bin}(50, .75) \). Find the probability that at least 20 more newspapers are sold on a Sunday than on a Monday at this stall.

Exercise 6.18. *(Distribution of Difference).* The number of earthquakes per year in Los Angeles of magnitude greater than 4 has a mean of .5 and that in Manila, Phillipines has a mean of 1. Find the pmf of the absolute difference between the number of earthquakes of magnitude greater than 4 in the two cities and approximately calculate the mean of the absolute difference.

Exercise 6.19. * Suppose \( X \sim \text{Poi}(\lambda), Y \sim \text{Bin}(n, \frac{\lambda}{n}) \), and that \( X \) and \( Y \) are independent. Derive a formula for \( P(X = Y) \).
Exercise 6.20. * The most likely value of a binomial random variable is 50, and the probability that it takes the value $n$ is .357. What is its variance?

Exercise 6.21. *(An Interesting Property of Binomial Distributions). Suppose $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(n - 1, p)$. Let $a_n = \max_k P(X = k), b_n = \max_k P(Y = k)$. Show that $b_n \leq a_n$, for any $p$.

Exercise 6.22. Suppose a fair coin is tossed repeatedly. Find the probability that three heads will be obtained before four tails.

Generalize to $r$ heads and $s$ tails.

Exercise 6.23. Twelve vegetable cans, all of the same size, have lost their labels. It is known that five contain tomatoes and the rest contain beets. What is the probability that in a random sample of four cans all contain beets?

Exercise 6.24. *(Domination of the Minority). In a small town in Alaska, there are 60 Republicans and 40 Democrats. Ten are selected at random for a council. What is the probability that there will be more Democrats on the council than Republicans?

Exercise 6.25 (Negative Hypergeometric Distribution). In a small town in Alaska, there are 60 Republicans and 40 Democrats. The mayor wants to form a council, selecting residents at random until five Democrats have been chosen. Find the distribution and the expected value of the number of Republicans in the council and hence the distribution and the expected value of the size of the council.

Exercise 6.26. The number of customers $X$ that enter a store during a one-hour interval has a Poisson distribution. Experience has shown that $P(X = 0) = .1111$. Find the probability that during a randomly chosen one-hour interval, more than five customers enter the store.

Exercise 6.27 (A Skewness and Kurtosis Calculation). Suppose $X$ and $Y$ are independent Poisson, that $X$ has skewness .5, and that $X + Y$ has skewness $\frac{1}{3}$. What are the skewness and kurtosis of $Y$?

Exercise 6.28. *(A Pretty Question). Suppose $X$ is a Poisson-distributed random variable. Can three different values of $X$ have equal probabilities?

Exercise 6.29. Suppose $X$ has a Poisson distribution such that $P(X = k) = P(X = k + 1)$ for some fixed integer $k$. Find the mean of $X$.

Exercise 6.30 (A P-Value Calculation). It is estimated that the risk of going into a coma with surgical anesthesia is 6 in 100,000. In the movie Coma, two patients out of ten go into a coma during surgery. Calculate the p-value for these data.

Exercise 6.31. *(Couples Wishing Large families). Suppose a couple want to have children until they have two children of each sex. What are the mean and the variance of the total number of children they will have?
Exercise 6.32 (Capture-Recapture). Suppose there are 10,000 fish in a pond. One hundred were captured, marked, and released. Then 1000 were recaptured. What is the probability that the recapture will contain more than 15 marked fish? Also do a Poisson approximation.

Exercise 6.33. Suppose $X$ has a hypergeometric distribution. Is it possible for $E(X)$ to be equal to $\text{Var}(X)$?

Exercise 6.34. * Suppose $X \sim \text{Poi}(\lambda)$. Find an expression for the probability that $X$ takes an even value.

Exercise 6.35. * (A Distribution on Rationals). Suppose $X$ and $Y$ are independent Poisson random variables with means $\lambda$ and $\mu$, respectively. Let $R = \frac{X}{X+Y} I_{X>0}$. Find the distribution of $R$; i.e., $P(R = \frac{m}{n})$, $(m, n) = 1$, and also $P(R = 0)$. What can you say about the expected value of $R$ if $\lambda = \mu$?

Exercise 6.36 (Poisson Approximation). Assume that each of 2000 individuals living near a nuclear power plant is exposed to particles of a certain kind of radiation at the rate of one per week. Suppose that each hit by a particle is harmless with probability $1 - 10^{-5}$ and produces a tumor with probability $10^{-5}$. Find the approximate distribution of:

(a) the total number of tumors produced in the whole population over a one-year period by this kind of radiation;
(b) the total number of individuals acquiring at least one tumor over a year from this radiation.

Exercise 6.37. * (Poisson Approximation). Twenty couples are seated at a rectangular table, husbands on one side and wives on the other, in a random order. Using a Poisson approximation, find the probability that:

(a) exactly two husbands are seated directly across from their wives;
(b) at least three are;
(c) at most three are.

Exercise 6.38 (Poisson Approximation). There are five coins on a desk, with probabilities .05, .1, .05, .01, and .04 for heads. Using a Poisson approximation, find the probability of obtaining at least one head when the five coins are each tossed once. Is the number of heads obtained binomially distributed in this problem?

Exercise 6.39 (Poisson Approximation). Typically, about 6% of guests with a confirmed reservation at a hotel with 1100 rooms do not show up. During a convention, the hotel is already completely booked. How many additional reservations can the hotel grant and be 99% sure that the number of guests with a confirmed reservation who will be denied a room is at most two?

Exercise 6.40 (Use Your Computer). Simulate the birthday problem to find the first person with the same birthday as yours. Perform the simulation 500 times. How many people did it take to find the first match? Was it typically about the same as the theoretical expected value?
Exercise 6.41 (Use Your Computer). Simulate the capture-recapture experiment with $N = 5000$ fish, a first catch of size $D = 500$, and a second catch of size $n = 250$. Perform the simulation 500 times. About how many marked fish did you find in the second catch? Did you ever see a second catch without any marked fish?

Exercise 6.42. Let $X \sim Bin(n, p)$. Prove that $P(X \text{ is even}) = \frac{1}{2} + \frac{(1-2p)^n}{2}$. Hence, show that $P(X \text{ is even})$ is larger than $\frac{1}{2}$ for any $n$ if $p < \frac{1}{2}$ but is larger than $\frac{1}{2}$ for only even values of $n$ if $p > \frac{1}{2}$.

References

Fundamentals of Probability: A First Course
DasGupta, A.
2010, XVI, 450 p., Hardcover