$$P(S \text{ on } 2 \mid S \text{ on } 1) = \frac{399{,}999}{499{,}999} = .80000$$

and

$$P(S \text{ on } 10 \mid S \text{ on first } 9) = \frac{399{,}991}{499{,}991} = .799996 \approx .80000$$

These calculations suggest that while the trials are not exactly independent, the conditional probabilities differ so slightly from one another that for practical purposes the trials can be regarded as independent with constant $P(S) = .8$. Thus, to a very good approximation, the experiment is binomial with $n = 10$ and $p = .8$. ■

We will use the following rule of thumb in deciding whether a "without-replacement" experiment can be treated as a binomial experiment.

Rule

> Suppose each trial of an experiment can result in S or F, but the sampling is without replacement from a population of size N. If the sample size (number of trials) n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

By "analyzed," we mean that probabilities based on the binomial experiment assumptions will be quite close to the actual "without-replacement" probabilities, which are typically more difficult to calculate. In Example 3.28, $n/N = 5/50 = .1 > .05$, so the binomial experiment is not a good approximation, but in Example 3.29, $n/N = 10/500{,}000 < .05$.

The Binomial Random Variable and Distribution

In most binomial experiments it is the total number of S's, rather than knowledge of exactly which trials yielded S's, that is of interest.

Definition

> Given a binomial experiment consisting of n trials, the **binomial random variable** X associated with this experiment is defined as
>
> X = the number of S's among the n trials

Suppose, for example, that $n = 3$. Then there are eight possible outcomes for the experiment:

SSS, SSF, SFS, SFF, FSS, FSF, FFS, FFF

From the definition of X, $X(SSF) = 2$, $X(SFF) = 1$, and so on. Possible values for X in an n-trial experiment are $x = 0, 1, 2, \dots, n$. We will often write $X \sim \text{Bin}(n, p)$ to indicate that X is a binomial rv based on n trials with success probability p.

Notation

Because the pmf of a binomial rv X depends on the two parameters n and p, we denote the pmf by $b(x; n, p)$.

To gain insight into $b(x; n, p)$ for general n, consider the case $n = 4$ for which each outcome, its probability, and corresponding x value are listed in Table 3.1. For example,

$$\begin{aligned} P(SSFS) &= P(S) \cdot P(S) \cdot P(F) \cdot P(S) && \text{(independent trials)} \\ &= p \cdot p \cdot (1-p) \cdot p && \text{(constant } P(S)\text{)} \\ &= p^3 \cdot (1-p) \end{aligned}$$

Table 3.1 Outcomes and probabilities for a binomial experiment with four trials

Outcome	x	*Probability*	*Outcome*	x	*Probability*
SSSS	4	p^4	*FSSS*	3	$p^3(1-p)$
SSSF	3	$p^3(1-p)$	*FSSF*	2	$p^2(1-p)^2$
SSFS	3	$p^3(1-p)$	*FSFS*	2	$p^2(1-p)^2$
SSFF	2	$p^2(1-p)^2$	*FSFF*	1	$p(1-p)^3$
SFSS	3	$p^3(1-p)$	*FFSS*	2	$p^2(1-p)^2$
SFSF	2	$p^2(1-p)^2$	*FFSF*	1	$p(1-p)^3$
SFFS	2	$p^2(1-p)^2$	*FFFS*	1	$p(1-p)^3$
SFFF	1	$p(1-p)^3$	*FFFF*	0	$(1-p)^4$

In this special case, we wish $b(x; 4, p)$ for $x = 0, 1, 2, 3,$ and 4. For $b(3; 4, p)$, we identify which of the 16 outcomes yield an x value of 3 and add up the probabilities associated with each such outcome:

$$b(3; 4, p) = P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF) = 4p^3(1-p)$$

Since the number of outcomes with $x = 3$ is four and each of the four has probability $p^3(1 - p)$ (the order of S's and F's is not important, but only the number of S's),

$$b(3; 4, p) = \begin{Bmatrix}\text{number of outcomes} \\ \text{with } X = 3\end{Bmatrix} \cdot \begin{Bmatrix}\text{probability of any particular} \\ \text{outcome with } X = 3\end{Bmatrix}$$

Similarly, $b(2; 4, p) = 6p^2(1 - p)^2$, which is also the product of the number of outcomes with $X = 2$ and the probability of any such outcome.

In general,

$$b(x; n, p) = \begin{Bmatrix}\text{number of sequences of} \\ \text{length } n \text{ consisting of } x\ S\text{'s}\end{Bmatrix} \cdot \begin{Bmatrix}\text{probability of any} \\ \text{particular such sequence}\end{Bmatrix}$$

Since the ordering of S's and F's is not important, the second factor above is $p^x(1 - p)^{n-x}$ (e.g., the first x trials resulting in S and the last $n - x$ resulting in F). The first factor is the number of ways of choosing x of the n trials to be S's—that is, the number of combinations of size x that can be constructed from n distinct objects (trials here).

Theorem

$$b(x;\, n,\, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Example 3.30

Each of six randomly selected cola drinkers is given a glass containing cola S and one containing cola F. The glasses are identical in appearance except for a code on the bottom to identify the cola. Suppose there is actually no tendency among cola drinkers to prefer one cola to the other. Then $p = P(\text{a selected individual prefers } S) = .5$, so with $X =$ the number among the six who prefer S, $X \sim \text{Bin}(6, .5)$.

Thus,

$$P(X = 3) = b(3;\, 6,\, .5) = \binom{6}{3}(.5^3(.5)^3 = 20\,(.5)^6 = .313$$

The probability that at least three prefer S is

$$P(3 \le X) = \sum_{x=3}^{6} b(x;\, 6,\, .5) = \sum_{x=3}^{6} \binom{6}{x}(.5)^x(.5)^{6-x} = .656$$

while the probability that at most one prefers S is

$$P(X \le 1) = \sum_{x=0}^{1} b(x;\, 6,\, .5) = .109$$

■

Using Binomial Tables

Even for a relatively small value of n, the computation of binomial probabilities can be tedious. Appendix Table A.1 tabulates the cdf $F(x) = P(X \le x)$ for $n = 5, 10, 15, 20, 25$ in combination with selected values of p. Various other probabilities can then be calculated using the proposition on cdf's from Section 3.2.

Notation

For $X \sim \text{Bin}(n, p)$, the cdf will be denoted by

$$P(X \le x) = B(x;\, n,\, p) = \sum_{y=0}^{x} b(y;\, n,\, p), \qquad x = 0, 1, \ldots, n$$

Example 3.31

Suppose that 20% of all copies of a particular textbook fail a certain binding strength test. Let X denote the number among 15 randomly selected copies that fail the test. Then X has a binomial distribution with $n = 15$ and $p = .2$.

1. The probability that at most 8 fail the test is

$$P(X \le 8) = \sum_{y=0}^{8} b(y;\, 15,\, .2) = B(8;\, 15,\, .2)$$

which is the entry in the $x = 8$ row and the $p = .2$ column of the $n = 15$ binomial table. From Appendix Table A.1, the probability is $B(8; 15, .2) = .999$.

2. The probability that exactly 8 fail is

$$P(X = 8) = P(X \le 8) - P(X \le 7) = B(8; 15, .2) - B(7; 15, .2)$$

which is the difference between two consecutive entries in the $p = .2$ column. The result is $.999 - .996 = .003$.

3. The probability that at least 8 fail is

$$\begin{aligned} P(X \ge 8) &= 1 - P(X \le 7) = 1 - B(7; 15, .2) \\ &= 1 - \begin{pmatrix} \text{entry in } x = 7 \\ \text{row of } p = .2 \text{ column} \end{pmatrix} \\ &= 1 - .996 = .004 \end{aligned}$$

4. Finally, the probability that between 4 and 7, inclusive, fail is

$$\begin{aligned} P(4 \le X \le 7) &= P(X = 4, 5, 6, \text{ or } 7) = P(X \le 7) - P(X \le 3) \\ &= B(7; 15, .2) - B(3; 15, .2) = .996 - .648 = .348 \end{aligned}$$

Notice that this latter probability is the difference between entries in the $x = 7$ and $x = 3$ rows, *not* the $x = 7$ and $x = 4$ rows. ■

Example 3.32 An electronics manufacturer claims that at most 10% of its power supply units need service during the warranty period. To investigate this claim, technicians at a testing laboratory purchase 20 units and subject each one to accelerated testing to simulate use during the warranty period. Let p denote the probability that a power supply unit needs repair during the period (the proportion of all such units that need repair). The laboratory technicians must decide whether the data resulting from the experiment supports the claim that $p \le .10$. Let X denote the number among the 20 sampled that need repair, so $X \sim \text{Bin}(20, p)$. Consider the decision rule

> Reject the claim that $p \le .10$ in favor of the conclusion that $p > .10$ if $x \ge 5$ (where x is the observed value of X) and consider the claim plausible if $x \le 4$.

The probability that the claim is rejected when $p = .10$ (an incorrect conclusion) is

$$P(X \ge 5 \text{ when } p = .10) = 1 - B(4; 20, .1) = 1 - .957 = .043$$

The probability that the claim is not rejected when $p = .20$ (a different type of incorrect conclusion) is

$$P(X \le 4 \text{ when } p = .2) = B(4; 20, .2) = .630$$

The first probability is rather small, but the second is intolerably large. When $p = .20$, so that the manufacturer has grossly understated the percentage of units that need service, and the stated decision rule is used, 63% of all samples will result in the manufacturer's claim being judged plausible!

One might think that the probability of this second type of erroneous conclusion could be made smaller by changing the cutoff value 5 in the decision rule to something else. However, while replacing 5 by a smaller number would yield a probability smaller than .630, the other probability would then increase. The only way to make both "error probabilities" small is to base the decision rule on an experiment involving many more units. ■

Note that a table entry of 0 signifies only that a probability is 0 to three significant digits, for all entries in the table are actually positive. Much more extensive tables of binomial probabilities have been published. In Chapter 4, we will present a method for obtaining quick and accurate approximations to binomial probabilities when n is large.

The Mean and Variance of X

For $n = 1$, the binomial distribution becomes the Bernoulli distribution. From Example 3.17 the mean value of a Bernoulli variable is $\mu = p$, so the expected number of S's on any single trial is p. Since a binomial experiment consists of n trials, intuition suggests that for $X \sim \text{Bin}(n, p)$, $E(X) = np$, the product of the number of trials and the probability of success on a single trial. The expression for $V(X)$ is not so intuitive.

Proposition

> If $X \sim \text{Bin}(n, p)$, then $E(X) = np$, $V(X) = np(1 - p) = npq$, and $\sigma_X = \sqrt{npq}$ (where $q = 1 - p$).

Thus, calculating the mean and variance of a binomial rv does not necessitate evaluating summations. The proof of the result for $E(X)$ is sketched in Exercise 58.

Example 3.33 If 75% of all purchases at a certain store are made with a credit card and X is the number among ten randomly selected purchases made with a credit card, then $X \sim \text{Bin}(10, .75)$. Thus $E(X) = np = (10)(.75) = 7.5$, $V(X) = npq = 10(.75)(.25) = 1.875$, and $\sigma = \sqrt{1.875}$. Again, even though X can take on only integer values, $E(X)$ need not be an integer. If we perform a large number of independent binomial experiments, each with $n = 10$ trials and $p = .75$, then the average number of S's per experiment will be close to 7.5. ■

Exercises Section 3.4* (44–61)

44. Compute the following binomial probabilities directly from the formula for $b(x; n, p)$.

a. $b(3; 8, .6)$

b. $b(5; 8, .6)$

c. $P(3 \le X \le 5)$, when $n = 8$ and $p = .6$

d. $P(1 \le X)$, when $n = 12$ and $p = .1$

45. Use Appendix Table A.1 to obtain the following probabilities:

*"Between a and b inclusive" is equivalent to $(a \le X \le b)$.

Let X denote the number of passengers on a randomly selected trip. Obtain the probability mass function of X.

61. Refer to Chebyshev's inequality given in Exercise 43. Calculate $P(|X - \mu| \geq k\sigma)$ for $k = 2$ and $k = 3$ when $X \sim \text{Bin}(20, .5)$ and compare to the corresponding upper bound. Repeat for $X \sim \text{Bin}(20, .75)$.

3.5 Hypergeometric and Negative Binomial Distributions

The hypergeometric and negative binomial distributions are both closely related to the binomial distribution. Whereas the binomial distribution is the approximate probability model for sampling without replacement from a finite dichotomous (S–F) population, the hypergeometric distribution is the exact probability model for the number of S's in the sample. The binomial rv X is the number of S's when the number n of trials is fixed, whereas the negative binomial distribution arises from fixing the number of S's and letting the number of trials be random.

The Hypergeometric Distribution

The assumptions leading to the hypergeometric distribution are as follows:

1. The population or set to be sampled consists of N individuals, objects, or elements (a *finite* population).
2. Each individual can be characterized as a success (S) or a failure (F), and there are M successes in the population.
3. A sample of n individuals is drawn in such a way that each subset of size n is equally likely to be chosen.

The random variable of interest is $X =$ the number of S's in the sample. The probability distribution of X depends on the parameters n, M, and N, so we wish to obtain $P(X = x) = h(x; n, M, N)$.

Example 3.34 An undergraduate library has 20 copies of a certain introductory economics text, of which 8 are first printings and 12 are second printings (containing corrections of some minor errors that appeared in the first printing). The course instructor has requested that 5 copies be put on 2-hour reserve. If the copies are selected in a completely random fashion, so that every subset of size 5 has the same probability of being selected, what is the probability that x ($x = 0, 1, 2, 3, 4$, or 5) of those selected are second printings?

In this example, the population size is $N = 20$, the sample size is $n = 5$, and the number of S's (second printing $= S$) and F's in the population are $M = 12$ and $N - M = 8$, respectively. Consider the value $x = 2$. Because all outcomes (each one consisting of five particular books) are equally likely,

$$P(X = 2) = h(2; 5, 12, 20) = \frac{\text{Number of outcomes having } X = 2}{\text{Number of possible outcomes}}$$

The number of possible outcomes in the experiment is the number of ways of selecting 5 from the 20 objects without regard to order—that is, $\binom{20}{5}$. To count the number of outcomes having $X = 2$, note that there are $\binom{12}{2}$ ways of selecting 2 of the second printings, and for each such way there are $\binom{8}{3}$ ways of selecting the 3 first printings to fill out the sample. The product rule from Chapter 2 then gives $\binom{12}{2}\binom{8}{3}$ as the number of outcomes with $X = 2$, so

$$h(2;\,5,\,12,\,20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = \frac{77}{323} = .238$$

■

In general, if the sample size n is smaller than the number of successes in the population (M), then the largest possible X value is n. However, if $M < n$ (e.g., a sample size of 25 and only 15 successes in the population), then X can be at most M. Similarly, whenever the number of population failures ($N - M$) exceeds the sample size, the smallest possible X value is 0 (since all sampled individuals might then be failures). However, if $N - M < n$, the smallest possible X value is $n - (N - M)$. Summarizing, the possible values of X satisfy the restriction $\max(0,\, n - (N - M)) \le x \le \min(n,\, M)$. An argument parallel to that of the previous example gives the pmf of X.

Proposition

If X is the number of S's in a completely random sample of size n drawn from a population consisting of M S's and $(N - M)$ F's, then the probability distribution of X, called the **hypergeometric distribution,** is given by

$$P(X = x) = h(x;\,n,\,M,\,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad (3.15)$$

for x an integer satisfying $\max(0,\, n - N + M) \le x \le \min(n,\, M)$.

In Example 3.34, $n = 5$, $M = 12$, and $N = 20$, so $h(x;\,5,\,12,\,20)$ for $x = 0, 1, 2, 3, 4, 5$ can be obtained by substituting these numbers into Equation (3.15).

Example 3.35

Five individuals from an animal population thought to be near extinction in a certain region have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix in, a random sample of 10 of these animals is selected. Let X = the number of tagged animals in the second sample. If there are actually 25 animals of this type in the region, what is the probability that (a) $X = 2$? (b) $X \le 2$?

Example (Example 3 continued)

The parameter values are $n = 10$, $M = 5$ (5 tagged animals in the population), and $N = 25$, so

$$h(x;\, 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}} \qquad x = 0, 1, 2, 3, 4, 5$$

For (a),

$$P(X = 2) = h(2;\, 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = .385$$

For (b),

$$\begin{aligned} P(X \le 2) = P(X = 0, 1, \text{ or } 2) &= \sum_{x=0}^{2} h(x;\, 10, 5, 25) \\ &= .057 + .257 + .385 = .699 \end{aligned}$$

Comprehensive tables of the hypergeometric distribution are available, but because the distribution has three parameters, these tables require much more space than tables for the binomial distribution.

The Mean and Variance of X

As in the binomial case, there are simple expressions for $E(X)$ and $V(X)$ for hypergeometric rv's.

Proposition

> The mean and variance of the hypergeometric rv X having pmf $h(x;\, n, M, N)$ are
>
> $$E(X) = n \cdot \frac{M}{N}, \qquad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

The ratio M/N is the proportion of S's in the population. If we replace M/N by p in $E(X)$ and $V(X)$, we get

$$\begin{aligned} E(X) &= np \\ V(X) &= \left(\frac{N-n}{N-1}\right) \cdot np(1-p) \end{aligned} \qquad (3.16)$$

Expression (3.16) shows that the means of the binomial and hypergeometric rv's are equal, while the variances of the two rv's differ by the factor $(N - n)/(N - 1)$, often called the **finite population correction factor.** This factor is less than 1, so the hypergeometric variable has smaller variance than does the binomial rv. The correction factor can be written $(1 - n/N)/(1 - 1/N)$, which is approximately 1 when n is small relative to N.

Example 3.36 *(Example 3.35 continued)*

In the animal-tagging example, $n = 10$, $M = 5$, and $N = 25$, so $p = \frac{5}{25} = .2$ and

$$E(X) = 10(.2) = 2$$

$$V(X) = \frac{15}{24}(10)(.2)(.8) = (.625)(1.6) = 1$$

If the sampling was carried out with replacement, $V(X) = 1.6$. Suppose the population size N is not actually known, so the value x is observed and we wish to estimate N. It is reasonable to equate the observed sample proportion of S's, x/n, with the population proportion, M/N, giving the estimate

$$\hat{N} = \frac{M \cdot n}{x}$$

If $M = 100$, $n = 40$, and $x = 16$, then $\hat{N} = 250$. ■

Approximating Hypergeometric Probabilities

Our general rule of thumb in Section 3.4 stated that if sampling was without replacement but n/N was at most .05, then the binomial distribution could be used to compute approximate probabilities involving the number of S's in the sample. A more precise statement is as follows: Let the population size, N, and number of population S's, M, get large with the ratio M/N remaining fixed at p. Then $h(x; n, M, N)$ approaches $b(x; n, p)$, so for n/N small, the two are approximately equal provided that p is not too near either 0 or 1. This is the rationale for our rule of thumb.

The Negative Binomial Distribution

The negative binomial rv and distribution are based on an experiment satisfying the following conditions:

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in either a success (S) or a failure (F).
3. The probability of success is constant from trial to trial, so $P(S \text{ on trial } i) = p$ for $i = 1, 2, 3 \ldots$.
4. The experiment continues (trials are performed) until a total of r successes have been observed, where r is a specified positive integer.

The random variable of interest is $X =$ the number of failures that precede the rth success; X is called a **negative binomial random variable** because, in contrast to the binomial rv, the number of successes is fixed and the number of trials is random.

Possible values of X are $0, 1, 2, \ldots$. Let $nb(x; r, p)$ denote the pmf of X. The event $\{X = x\}$ is equivalent to $\{r - 1$ S's in the first $(x + r - 1)$ trials and an S on the $(x + r)$th trial} (e.g., if $r = 5$ and $x = 10$, then there must be four S's in the first 14 trials and the trial 15 must be an S). Since trials are independent,

$$\begin{aligned} nb(x;\ r,\ p) &= P(X = x) \\ &= P(r - 1\ S\text{'s on the first } x + r - 1 \text{ trials}) \cdot P(S) \end{aligned} \tag{3.17}$$

The first probability on the far right of Expression (3.17) is the binomial probability

$$\binom{x + r - 1}{r - 1} p^{r-1}(1 - p)^x, \qquad \text{while } P(S) = p$$

Proposition

> The pmf of the negative binomial rv X with parameters r = number of S's and $p = P(S)$ is
>
> $$nb(x;\ r,\ p) = \binom{x + r - 1}{r - 1} p^r(1 - p)^x, \qquad x = 0, 1, 2, \ldots$$

Example 3.37

A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let $p = P$(a randomly selected couple agrees to participate). If $p = .2$, what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with S = {agrees to participate}, what is the probability that 10 F's occur before the fifth S? Substituting $r = 5$, $p = .2$, and $x = 10$ into $nb(x;\ r,\ p)$ gives

$$nb(10;\ 5,\ .2) = \binom{14}{4}(.2)^5(.8)^{10} = .034$$

The probability that at most 10 F's are observed (at most 15 couples are asked) is

$$P(X \le 10) = \sum_{x=0}^{10} nb(x;\ 5,\ .2) = (.2)^5 \sum_{x=0}^{10} \binom{x + 4}{4} (.8)^x = .164$$

■

In some sources, the negative binomial rv is taken to be the number of trials $X + r$ rather than the number of failures.

In the special case $r = 1$, the pmf is

$$nb(x;\ 1,\ p) = (1 - p)^x p, \qquad x = 0, 1, 2, \ldots \tag{3.18}$$

In Example 3.10, we derived the pmf for the number of trials necessary to obtain the first S, and the pmf there is similar to Expression (3.18). Both X = number of F's and Y = number of trials ($= 1 + X$) are referred to in the literature as **geometric random variables,** and the pmf (3.18) is called the **geometric distribution.**

In Example 3.18, the expected number of trials until the first S was shown to be $1/p$, so that the expected number of F's until the first S is $(1/p) - 1 = (1 - p)/p$. Intuitively, we would expect to see $r \cdot (1 - p)/p$ F's before the rth S, and this is indeed $E(X)$. There is also a simple formula for $V(X)$.

Proposition

> If X is a negative binomial rv with pmf $nb(x; r, p)$, then
>
> $$E(X) = \frac{r(1-p)}{p}, \qquad V(X) = \frac{r(1-p)}{p^2}$$

Finally, by expanding the binomial coefficient in front of $p^r(1-p)^x$ and doing some cancellation, it can be seen that $nb(x; r, p)$ is well defined even when r is not an integer. This *generalized negative binomial distribution* has been found to fit observed data quite well in a wide variety of applications.

Exercises Section 3.5 (62–72)

62. A shipment of 15 concrete cylinders has been received by a contractor, 5 for a small project and the other 10 for a larger project. Suppose that 6 of the 15 have a crushing strength below the specified minimum. If the 5 for the smaller project are randomly selected from the 15 and X = the number among the 5 that have a below minimum crushing strength, then X has a hypergeometric distribution with parameters $n = 5$, $M = 6$, and $N = 15$. Compute the following:

a. $P(X = 2)$

b. $P(X \le 2)$

c. $P(X \ge 2)$

d. $E(X)$ and $V(X)$

63. Each of 12 refrigerators of a certain type has been returned to a distributor because of the presence of a high-pitched oscillating noise when the refrigerator is running. Suppose that 4 of these 12 have defective compressors and the other 8 have less serious problems. If they are examined in random order, let X = the number among the first 6 examined that have a defective compressor. Compute the following:

a. $P(X = 1)$

b. $P(X \ge 4)$

c. $P(1 \le X \le 3)$

64. A tennis coach has a basket containing 25 balls; 15 of these are Penn balls and the other 10 are Wilsons. Each of four players randomly selects 3 balls for a match.

a. What is the probability that exactly 8 of the balls selected are Penns?

b. What are the mean value and standard deviation of the number of Penn balls selected?

c. What is the probability that the number of Penn balls selected is more than 1 SD away from its mean value?

d. What are the mean value and standard deviation of the number of Penn balls left in the basket?

65. A geologist has collected 10 specimens of basaltic rock and 10 specimens of granite. If the geologist instructs a laboratory assistant to randomly select 15 of the specimens for analysis, what is the pmf of the number of basalt specimens selected for analysis? What is the probability that all specimens of one of the two types of rock are selected for analysis?

66. A personnel director interviewing 11 senior engineers for four job openings has scheduled six interviews for the first day and five for the second day of interviewing. Assume that the candidates are interviewed in random order.

a. What is the probability that x of the top four candidates are interviewed on the first day?

b. How many of the top four candidates can be expected to be interviewed on the first day?

67. Twenty pairs of individuals playing in a bridge tournament have been seeded 1, . . . , 20. In the first part of the tournament, the 20 are randomly divided into 10 east–west pairs and 10 north–south pairs.

a. What is the probability that x of the top 10 pairs end up playing east–west?

b. What is the probability that all of the top five pairs end up playing the same direction?

c. If there are $2n$ pairs, what is the pmf of X = the number among the top n pairs who end up playing east–west, and what are $E(X)$ and $V(X)$?

68. A second-stage smog alert has been called in a certain area of Los Angeles County in which there are 50 industrial firms. An inspector will visit 10 randomly selected firms to check for violations of regulations.

a. If 15 of the firms are actually violating at least one regulation, what is the pmf of the number of firms visited by the inspector that are in violation of at least one regulation?

b. If there are 500 firms in the area, of which 150 are in violation, approximate the pmf of part (a) by a simpler pmf.

c. For X = the number among the 10 visited that are in violation, compute $E(X)$ and $V(X)$ both for the exact pmf and the approximating pmf in part (b).

69. Suppose that $p = P(\text{male birth}) = .5$. A couple wishes to have exactly two female children in their family. They will have children until this condition is fulfilled.

a. What is the probability that the family has x male children?

b. What is the probability that the family has four children?

c. What is the probability that the family has at most four children?

d. How many male children would you expect this family to have? How many children would you expect this family to have?

70. A family decides to have children until it has three children of the same sex. Assuming $P(B) = P(G) = .5$, what is the pmf of X = the number of children in the family?

71. Three brothers and their wives decide to have children until each family has two female children. What is the pmf of X = the total number of male children born to the brothers? What is $E(X)$, and how does it compare to the expected number of male children born to each brother?

72. Individual A has a red die and B has a green die (both fair). If they each roll until they obtain five "doubles" (1–1, . . . , 6–6), what is the pmf of X = the total number of times a die is rolled? What are $E(X)$ and $V(X)$?

3.6 The Poisson Probability Distribution

The binomial, hypergeometric, and negative binomial distributions were all derived by starting with an experiment consisting of trials or draws and applying the laws of probability to various outcomes of the experiment. There is no simple experiment on which the Poisson distribution is based, though we will shortly describe how it can be obtained by certain limiting operations.

Definition

> An rv X is said to have a **Poisson distribution** if the pmf of X is
>
> $$p(x;\ \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$
>
> for some $\lambda > 0$.

The value of λ is frequently a rate per unit time or per unit area. The letter e in $p(x;\ \lambda)$ represents the base of the natural logarithm system, and its numerical value is approximately 2.71828. Because λ must be positive, $p(x;\ \lambda) > 0$ for all possible

x values. The fact that $\Sigma_{x=0}^{\infty} p(x;\ \lambda) = 1$ is a consequence of the Maclaurin infinite series expansion of e^{λ}, which appears in most calculus texts:

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \tag{3.19}$$

If the two extreme terms in Expression (3.19) are multiplied by $e^{-\lambda}$ and then $e^{-\lambda}$ is placed inside the summation, the result is

$$1 = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!}$$

which shows that $p(x;\ \lambda)$ fulfills the second condition necessary for specifying a pmf.

Example 3.38 Let X denote the number of flaws on the surface of a randomly selected boiler of a certain type. Suppose X has a Poisson distribution with $\lambda = 5$. Then the probability that a randomly selected boiler has exactly two flaws is

$$P(X = 2) = \frac{e^{-5}(5)^2}{2!} = .084$$

The probability that a boiler contains at most two flaws is

$$P(X \le 2) = \sum_{x=0}^{2} \frac{e^{-5}(5)^x}{x!} = e^{-5}\left(1 + 5 + \frac{25}{2}\right) = .125$$

■

The Poisson Distribution as a Limit

The rationale for using the Poisson distribution in many situations is provided by the following proposition.

Proposition

> Suppose that in the binomial pmf $b(x;\ n,\ p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np remains fixed at value $\lambda > 0$. Then $b(x;\ n,\ p) \to p(x;\ \lambda)$.

According to this proposition, *in any binomial experiment in which n is large and p is small, $b(x;\ n,\ p) \approx p(x;\ \lambda)$ where $\lambda = np$.* As a rule of thumb, this approximation can safely be applied if $n \ge 100$, $p \le .01$, and $np \le 20$.

Example 3.39 If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is .005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly 1 page with errors? At most three pages with errors?

With S denoting a page containing at least one error and F an error-free page, the number X of pages containing at least one error is a binomial rv with $n = 400$ and $p = .005$, so $np = 2$. We wish

$$P(X = 1) = b(1;\ 400,\ .005) \approx p(1;\ 2) = \frac{e^{-2}(2)^1}{1!} = .271$$

Similarly,

$$P(X \leq 3) \approx \sum_{x=0}^{3} p(x;\ 2) = \sum_{x=0}^{3} e^{-2}\frac{2^x}{x!} = .135 + .271 + .271 + .180$$
$$= .857$$

Appendix Table A.2 exhibits the cdf $F(x;\ \lambda)$ for $\lambda = .1, .2, \ldots, 1, 2, \ldots, 10$, 15, and 20. For example, if $\lambda = 2$, then $P(X \leq 3) = F(3;\ 2) = .857$ as in Example 3.39, while $P(X = 3) = F(3;\ 2) - F(2;\ 2) = .180$.

The Mean and Variance of X

Since $b(x;\ n, p) \rightarrow p(x;\ \lambda)$ as $n \rightarrow \infty$, $p \rightarrow 0$, $np = \lambda$, the mean and variance of a binomial variable should approach those of a Poisson variable. These limits are $np = \lambda$ and $np(1 - p) \rightarrow \lambda$.

Proposition

If X has a Poisson distribution with parameter λ, then $E(X) = V(X) = \lambda$.

These results can also be derived directly from the definitions of mean and variance.

Example 3.40 *(Example 3.38 continued)*

Both the expected number of flaws and the variance of the number of flaws equal five, and $\sigma_X = \sqrt{\lambda} = \sqrt{5} = 2.24$.

The Poisson Process

A very important application of the Poisson distribution arises in connection with the occurrence of events of a particular type over time. As an example, suppose that starting from a time point that we label $t = 0$, we are interested in counting the number of radioactive pulses recorded by a Geiger counter. We make the following assumptions about the way in which pulses occur:

1. There exists a parameter $\alpha > 0$ such that for any short time interval of length Δt, the probability that exactly one pulse is received is $\alpha \cdot \Delta t + o(\Delta t)$.*
2. The probability of more than one pulse being received during Δt is $o(\Delta t)$ [which, along with Assumption 1, implies that the probability of no pulses during Δt is $1 - \alpha \cdot \Delta t - o(\Delta t)$].
3. The number of pulses received during the time interval Δt is independent of the number received prior to this time interval.

Informally, Assumption 1 says that, for a short interval of time, the probability of receiving a single pulse is approximately proportional to the length of the time interval, where α is the constant of proportionality. Now let $P_k(t)$ denote the

* A quantity is $o(\Delta t)$ (read "little o of delta t") if as Δt approaches 0, so does $o(\Delta t)/\Delta t$. That is, $o(\Delta t)$ is even more negligible than Δt itself. The quantity $(\Delta t)^2$ has this property, but $\sin(\Delta t)$ does not.

Propos

Examp

Exercis

73. Let X
and us
probab
a. P(
b. P(
c. P(

probability that k pulses will be received by the counter during any particular time interval of length t.

Proposition

> $P_k(t) = e^{-\alpha t} \cdot (\alpha t)^k/k!$, so that the number of pulses during a time interval of length t is a Poisson rv with parameter $\lambda = \alpha t$. The expected number of pulses during any such time interval is then αt, so the expected number during a unit interval of time is α.

Example 3.41 Suppose pulses arrive at the counter at an average rate of six per minute, so that $\alpha = 6$. To find the probability that in a .5-min interval at least one pulse is received, note that the number of pulses in such an interval has a Poisson distribution with parameter $\alpha t = 6(.5) = 3$ (.5 min is used because α is expressed as a rate per minute). Then with X = the number of pulses received in the 30-sec interval,

$$P(1 \le X) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = .950$$

■

If in Assumptions 1–3 we replace "pulse" by "event," then the number of events occurring during a fixed time interval of length t has a Poisson distribution with parameter αt. Any process that has this distribution is called a **Poisson process,** and α is called the *rate of the process.* Other examples of situations giving rise to a Poisson process include monitoring the status of a computer system over time, with breakdowns constituting the events of interest, recording the number of accidents in an industrial facility over time, answering calls at a telephone switchboard, and observing the number of cosmic-ray showers from a particular observatory over time.

Instead of observing events over time, consider observing events of some type that occur in a two- or three-dimensional region. For example, we might select on a map a certain region R of a forest, go to that region, and count the number of trees. Each tree would represent an event occurring at a particular point in space. Under assumptions similar to 1–3, it can be shown that the number of events occurring in a region R has a Poisson distribution with parameter $\alpha \cdot a(R)$, where $a(R)$ is the area of R. The quantity α is the expected number of events per unit area or volume.

Exercises Section 3.6 (73–87)

73. Let X have a Poisson distribution with parameter $\lambda = 5$ and use Appendix Table A.2 to compute the following probabilities.

a. $P(X \le 8)$

b. $P(X = 8)$

c. $P(9 \le X)$

d. $P(5 \le X \le 8)$

e. $P(5 < X < 8)$

74. Suppose the number X of tornadoes observed in a particular region during a 1-year period has a Poisson distribution with $\lambda = 8$.

a. Compute $P(X \le 5)$.