Chapter 3

Common Families of Distributions

"How do all these unusuals strike you, Watson?"
"Their cumulative effect is certainly considerable, and yet each of them is quite possible in itself."

Sherlock Holmes and Dr. Watson
The Adventure of the Abbey Grange

3.1 Introduction

Statistical distributions are used to model populations; as such, we usually deal with a *family* of distributions rather than a single distribution. This family is indexed by one or more parameters, which allow us to vary certain characteristics of the distribution while staying with one functional form. For example, we may specify that the normal distribution is a reasonable choice to model a particular population, but we cannot precisely specify the mean. Then, we deal with a parametric family, normal distributions with mean μ, where μ is an unspecified parameter, $-\infty < \mu < \infty$.

In this chapter we catalog many of the more common statistical distributions, some of which we have previously encountered. For each distribution we will give its mean and variance and many other useful or descriptive measures that may aid understanding. We will also indicate some typical applications of these distributions and some interesting and useful interrelationships. Some of these facts are summarized in tables at the end of the book. This chapter is by no means comprehensive in its coverage of statistical distributions. That task has been accomplished by Johnson and Kotz (1969–1972) in their multiple-volume work *Distributions in Statistics* and in the updated volumes by Johnson, Kotz, and Balakrishnan (1994, 1995) and Johnson, Kotz, and Kemp (1992).

3.2 Discrete Distributions

A random variable X is said to have a discrete distribution if the range of X, the sample space, is countable. In most situations, the random variable has integer-valued outcomes.

Discrete Uniform Distribution

A random variable X has a *discrete uniform* $(1, N)$ *distribution* if

$$P(X = x|N) = \frac{1}{N}, \quad x = 1, 2, \ldots, N, \tag{3.2.1}$$

where N is a specified integer. This distribution puts equal mass on each of the outcomes $1, 2, \ldots, N$.

A note on notation: When we are dealing with parametric distributions, as will almost always be the case, the distribution is dependent on values of the parameters. In order to emphasize this fact and to keep track of the parameters, we write them in the pmf preceded by a "|" (given). This convention will also be used with cdfs, pdfs, expectations, and other places where it might be necessary to keep track of the parameters. When there is no possibility of confusion, the parameters may be omitted in order not to clutter up notation too much.

To calculate the mean and variance of X, recall the identities (provable by induction)

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2} \quad \text{and} \quad \sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}.$$

We then have

$$\mathrm{E}X = \sum_{x=1}^{N} xP(X = x|N) = \sum_{x=1}^{N} x\frac{1}{N} = \frac{N+1}{2}$$

and

$$\mathrm{E}X^2 = \sum_{x=1}^{N} x^2\frac{1}{N} = \frac{(N+1)(2N+1)}{6},$$

and so

$$\begin{aligned} \operatorname{Var} X &= \mathrm{E}X^2 - (\mathrm{E}X)^2 \\ &= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{(N+1)(N-1)}{12}. \end{aligned}$$

This distribution can be generalized so that the sample space is any range of integers, $N_0, N_0 + 1, \ldots, N_1$, with pmf $P(X = x|N_0, N_1) = 1/(N_1 - N_0 + 1)$.

Hypergeometric Distribution

The hypergeometric distribution has many applications in finite population sampling and is best understood through the classic example of the urn model.

Suppose we have a large urn filled with N balls that are identical in every way except that M are red and $N-M$ are green. We reach in, blindfolded, and select K balls at random (the K balls are taken all at once, a case of sampling without replacement). What is the probability that exactly x of the balls are red?

The total number of samples of size K that can be drawn from the N balls is $\binom{N}{K}$, as was discussed in Section 1.2.3. It is required that x of the balls be red, and this can be accomplished in $\binom{M}{x}$ ways, leaving $\binom{N-M}{K-x}$ ways of filling out the sample with $K-x$ green balls. Thus, if we let X denote the number of red balls in a sample of size K, then X has a *hypergeometric distribution* given by

$$P(X=x|N,M,K)=\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}},\quad x=0,1,\ldots,K. \tag{3.2.2}$$

Note that there is, implicit in (3.2.2), an additional assumption on the range of X. Binomial coefficients of the form $\binom{n}{r}$ have been defined only if $n \geq r$, and so the range of X is additionally restricted by the pair of inequalities

$$M \geq x \quad \text{and} \quad N-M \geq K-x,$$

which can be combined as

$$M-(N-K) \leq x \leq M.$$

In many cases K is small compared to M and N, so the range $0 \leq x \leq K$ will be contained in the above range and, hence, will be appropriate. The formula for the hypergeometric probability function is usually quite difficult to deal with. In fact, it is not even trivial to verify that

$$\sum_{x=0}^{K} P(X=x)=\sum_{x=0}^{K}\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}=1.$$

The hypergeometric distribution illustrates the fact that, statistically, dealing with finite populations (finite N) is a difficult task.

The mean of the hypergeometric distribution is given by

$$\mathrm{E}X=\sum_{x=0}^{K} x\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}=\sum_{x=1}^{K} x\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}. \qquad \text{(summand is 0 at } x=0\text{)}$$

To evaluate this expression, we use the identities (already encountered in Section 2.3)

$$x\binom{M}{x}=M\binom{M-1}{x-1},$$

$$\binom{N}{K}=\frac{N}{K}\binom{N-1}{K-1},$$

and obtain

$$\mathrm{E}X = \sum_{x=1}^{K} \frac{M\binom{M-1}{x-1}\binom{N-M}{K-x}}{\frac{N}{K}\binom{N-1}{K-1}} = \frac{KM}{N}\sum_{x=1}^{K}\frac{\binom{M-1}{x-1}\binom{N-M}{K-x}}{\binom{N-1}{K-1}}.$$

We now can recognize the second sum above as the sum of the probabilities for another hypergeometric distribution based on parameter values $N-1$, $M-1$, and $K-1$. This can be seen clearly by defining $y = x-1$ and writing

$$\begin{aligned}\sum_{x=1}^{K}\frac{\binom{M-1}{x-1}\binom{N-M}{K-x}}{\binom{N-1}{K-1}} &= \sum_{y=0}^{K-1}\frac{\binom{M-1}{y}\binom{(N-1)-(M-1)}{K-1-y}}{\binom{N-1}{K-1}}\\ &= \sum_{y=0}^{K-1} P(Y=y|N-1, M-1, K-1) = 1,\end{aligned}$$

where Y is a hypergeometric random variable with parameters $N-1$, $M-1$, and $K-1$. Therefore, for the hypergeometric distribution,

$$\mathrm{E}X = \frac{KM}{N}.$$

A similar, but more lengthy, calculation will establish that

$$\operatorname{Var} X = \frac{KM}{N}\left(\frac{(N-M)(N-K)}{N(N-1)}\right).$$

Note the manipulations used here to calculate $\mathrm{E}X$. The sum was transformed to another hypergeometric distribution with different parameter values and, by recognizing this fact, we were able to sum the series.

Example 3.2.1 (Acceptance sampling) The hypergeometric distribution has application in acceptance sampling, as this example will illustrate. Suppose a retailer buys goods in lots and each item can be either acceptable or defective. Let

$$\begin{aligned} N &= \#\text{ of items in a lot,}\\ M &= \#\text{ of defectives in a lot.}\end{aligned}$$

Then we can calculate the probability that a sample of size K contains x defectives. To be specific, suppose that a lot of 25 machine parts is delivered, where a part is considered acceptable only if it passes tolerance. We sample 10 parts and find that none are defective (all are within tolerance). What is the probability of this event if there are 6 defectives in the lot of 25? Applying the hypergeometric distribution with $N = 25, M = 6, K = 10$, we have

$$P(X=0) = \frac{\binom{6}{0}\binom{19}{10}}{\binom{25}{10}} = .028,$$

showing that our observed event is quite unlikely if there are 6 (or more!) defectives in the lot. ||

Binomial Distribution

The binomial distribution, one of the more useful discrete distributions, is based on the idea of a *Bernoulli trial.* A Bernoulli trial (named for James Bernoulli, one of the founding fathers of probability theory) is an experiment with two, and only two, possible outcomes. A random variable X has a *Bernoulli*(p) *distribution* if

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases} \qquad 0 \le p \le 1. \tag{3.2.3}$$

The value $X = 1$ is often termed a "success" and p is referred to as the success probability. The value $X = 0$ is termed a "failure." The mean and variance of a Bernoulli(p) random variable are easily seen to be

$$\begin{aligned} \mathrm{E}X &= 1p + 0(1-p) = p, \\ \operatorname{Var} X &= (1-p)^2 p + (0-p)^2(1-p) = p(1-p). \end{aligned}$$

Many experiments can be modeled as a sequence of Bernoulli trials, the simplest being the repeated tossing of a coin; p = probability of a head, $X = 1$ if the coin shows heads. Other examples include gambling games (for example, in roulette let $X = 1$ if red occurs, so p = probability of red), election polls ($X = 1$ if candidate A gets a vote), and incidence of a disease (p = probability that a random person gets infected).

If n identical Bernoulli trials are performed, define the events

$$A_i = \{X = 1 \text{ on the } i\text{th trial}\}, \quad i = 1, 2, \ldots, n.$$

If we assume that the events $A_1, \ldots, A_n$ are a collection of independent events (as is the case in coin tossing), it is then easy to derive the distribution of the total number of successes in n trials. Define a random variable Y by

$$Y = \text{ total number of successes in } n \text{ trials.}$$

The event $\{Y = y\}$ will occur only if, out of the events $A_1, \ldots, A_n$, exactly y of them occur, and necessarily $n - y$ of them do not occur. One particular outcome (one particular ordering of occurrences and nonoccurrences) of the n Bernoulli trials might be $A_1 \cap A_2 \cap A_3^c \cap \cdots \cap A_{n-1} \cap A_n^c$. This has probability of occurrence

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3^c \cap \cdots \cap A_{n-1} \cap A_n^c) &= pp(1-p) \cdot \cdots \cdot p(1-p) \\ &= p^y(1-p)^{n-y}, \end{aligned}$$

where we have used the independence of the A_is in this calculation. Notice that the calculation is not dependent on *which* set of y A_is occurs, only that *some* set of y occurs. Furthermore, the event $\{Y = y\}$ will occur no matter which set of y A_is occurs. Putting this all together, we see that a particular sequence of n trials with exactly y successes has probability $p^y(1-p)^{n-y}$ of occurring. Since there are $\binom{n}{y}$

such sequences (the number of orderings of y 1s and $n-y$ 0s), we have

$$P(Y=y|n,p)=\binom{n}{y}p^y(1-p)^{n-y},\quad y=0,1,2,\ldots,n,$$

and Y is called a *binomial(n,p) random variable.*

The random variable Y can be alternatively, and equivalently, defined in the following way: In a sequence of n identical, independent Bernoulli trials, each with success probability p, define the random variables $X_1,\ldots,X_n$ by

$$X_i=\begin{cases}1 & \text{with probability } p\\ 0 & \text{with probability } 1-p.\end{cases}$$

The random variable

$$Y=\sum_{i=1}^n X_i$$

has the binomial(n,p) distribution.

The fact that $\sum_{y=0}^n P(Y=y)=1$ follows from the following general theorem.

Theorem 3.2.2 (Binomial Theorem) *For any real numbers x and y and integer $n\geq 0$,*

$$(x+y)^n=\sum_{i=0}^n\binom{n}{i}x^iy^{n-i}. \tag{3.2.4}$$

Proof: Write

$$(x+y)^n=(x+y)(x+y)\cdot\cdots\cdot(x+y),$$

and consider how the right-hand side would be calculated. From each factor $(x+y)$ we choose either an x or y, and multiply together the n choices. For each $i=0,1,\ldots,n$, the number of such terms in which x appears exactly i times is $\binom{n}{i}$. Therefore, this term is of the form $\binom{n}{i}x^iy^{n-i}$ and the result follows. □

If we take $x=p$ and $y=1-p$ in (3.2.4), we get

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

and we see that each term in the sum is a binomial probability. As another special case, take $x=y=1$ in Theorem 3.2.2 and get the identity

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

The mean and variance of the binomial distribution have already been derived in Examples 2.2.3 and 2.3.5, so we will not repeat the derivations here. For completeness, we state them. If $X \sim \text{binomial}(n, p)$, then

$$\mathrm{E}X = np, \quad \operatorname{Var} X = np(1-p).$$

The mgf of the binomial distribution was calculated in Example 2.3.9. It is

$$M_X(t) = [pe^t + (1-p)]^n.$$

Example 3.2.3 (Dice probabilities) Suppose we are interested in finding the probability of obtaining at least one 6 in four rolls of a fair die. This experiment can be modeled as a sequence of four Bernoulli trials with success probability $p = \frac{1}{6} = P(\text{die shows } 6)$. Define the random variable X by

$$X = \text{total number of 6s in four rolls.}$$

Then $X \sim \text{binomial}(4, \frac{1}{6})$ and

$$\begin{aligned}
P(\text{at least one } 6) &= P(X > 0) = 1 - P(X = 0) \\
&= 1 - \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 \\
&= 1 - \left(\frac{5}{6}\right)^4 \\
&= .518.
\end{aligned}$$

Now we consider another game; throw a pair of dice 24 times and ask for the probability of at least one double 6. This, again, can be modeled by the binomial distribution with success probability p, where

$$p = P(\text{roll a double } 6) = \frac{1}{36}.$$

So, if Y = number of double 6s in 24 rolls, $Y \sim \text{binomial}(24, \frac{1}{36})$ and

$$\begin{aligned}
P(\text{at least one double } 6) &= P(Y > 0) \\
&= 1 - P(Y = 0) \\
&= 1 - \binom{24}{0}\left(\frac{1}{36}\right)^0\left(\frac{35}{36}\right)^{24} \\
&= 1 - \left(\frac{35}{36}\right)^{24} \\
&= .491.
\end{aligned}$$

This is the calculation originally done in the eighteenth century by Pascal at the request of the gambler de Meré, who thought both events had the same probability. (He began to believe he was wrong when he started losing money on the second bet.) ||

Poisson Distribution

The Poisson distribution is a widely applied discrete distribution and can serve as a model for a number of different types of experiments. For example, if we are modeling a phenomenon in which we are waiting for an occurrence (such as waiting for a bus, waiting for customers to arrive in a bank), the number of occurrences in a given time interval can sometimes be modeled by the Poisson distribution. One of the basic assumptions on which the Poisson distribution is built is that, for small time intervals, the probability of an arrival is proportional to the length of waiting time. This makes it a reasonable model for situations like those indicated above. For example, it makes sense to assume that the longer we wait, the more likely it is that a customer will enter the bank. See the Miscellanea section for a more formal treatment of this.

Another area of application is in spatial distributions, where, for example, the Poisson may be used to model the distribution of bomb hits in an area or the distribution of fish in a lake.

The Poisson distribution has a single parameter λ, sometimes called the intensity parameter. A random variable X, taking values in the nonnegative integers, has a *Poisson*(λ) *distribution* if

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots. \tag{3.2.5}$$

To see that $\sum_{x=0}^{\infty} P(X = x|\lambda) = 1$, recall the Taylor series expansion of e^y,

$$e^y = \sum_{i=0}^{\infty} \frac{y^i}{i!}.$$

Thus,

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

The mean of X is easily seen to be

$$\begin{aligned}
\mathrm{E}X &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} && \text{(substitute } y = x - 1\text{)} \\
&= \lambda
\end{aligned}$$

A similar calculation will show that

$$\operatorname{Var} X = \lambda,$$

and so the parameter λ is both the mean and the variance of the Poisson distribution.

The mgf can also be obtained by a straightforward calculation, again following from the Taylor series of e^y. We have

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

(See Exercise 2.33 and Example 2.3.13.)

Example 3.2.4 (Waiting time) As an example of a waiting-for-occurrence application, consider a telephone operator who, on the average, handles five calls every 3 minutes. What is the probability that there will be no calls in the next minute? At least two calls?

If we let X = number of calls in a minute, then X has a Poisson distribution with $EX = \lambda = \frac{5}{3}$. So

$$\begin{aligned} P(\text{no calls in the next minute}) &= P(X = 0) \\ &= \frac{e^{-5/3}\left(\frac{5}{3}\right)^0}{0!} \\ &= e^{-5/3} = .189; \\ P(\text{at least two calls in the next minute}) &= P(X \ge 2) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - .189 - \frac{e^{-5/3}\left(\frac{5}{3}\right)^1}{1!} \\ &= .496. \end{aligned}$$

||

Calculation of Poisson probabilities can be done rapidly by noting the following recursion relation:

$$P(X = x) = \frac{\lambda}{x} P(X = x - 1), \quad x = 1, 2, \ldots. \tag{3.2.6}$$

This relation is easily proved by writing out the pmf of the Poisson. Similar relations hold for other discrete distributions. For example, if $Y \sim \text{binomial}(n, p)$, then

$$P(Y = y) = \frac{(n - y + 1)}{y} \frac{p}{1 - p} P(Y = y - 1). \tag{3.2.7}$$

The recursion relations (3.2.6) and (3.2.7) can be used to establish the Poisson approximation to the binomial, which we have already seen in Section 2.3, where the approximation was justified using mgfs. Set $\lambda = np$ and, if p is small, we can write

$$\frac{n - y + 1}{y} \frac{p}{1 - p} = \frac{np - p(y - 1)}{y - py} \approx \frac{\lambda}{y}$$

since, for small p, the terms $p(y-1)$ and py can be ignored. Therefore, to this level of approximation, (3.2.7) becomes

$$P(Y=y)=\frac{\lambda}{y}P(Y=y-1), \tag{3.2.8}$$

which is the Poisson recursion relation. To complete the approximation, we need only establish that $P(X=0)\approx P(Y=0)$, since all other probabilities will follow from (3.2.8). Now

$$P(Y=0)=(1-p)^n=\left(1-\frac{np}{n}\right)^n=\left(1-\frac{\lambda}{n}\right)^n$$

upon setting $np=\lambda$. Recall from Section 2.3 that for fixed λ, $\lim_{n\to\infty}(1-(\lambda/n))^n=e^{-\lambda}$, so for large n we have the approximation

$$P(Y=0)=\left(1-\frac{\lambda}{n}\right)^n\approx e^{-\lambda}=P(X=0),$$

completing the Poisson approximation to the binomial.

The approximation is valid when n is large and p is small, which is exactly when it is most useful, freeing us from calculation of binomial coefficients and powers for large n.

Example 3.2.5 (Poisson approximation) A typesetter, on the average, makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?

If we assume that setting a word is a Bernoulli trial with success probability $p=\frac{1}{500}$ (notice that we are labeling an error as a "success") and that the trials are independent, then X = number of errors in five pages (1500 words) is binomial$\left(1500,\frac{1}{500}\right)$. Thus

$$\begin{aligned}P(\text{no more than two errors})&=P(X\le 2)\\&=\sum_{x=0}^{2}\binom{1500}{x}\left(\frac{1}{500}\right)^x\left(\frac{499}{500}\right)^{1500-x}\\&=.4230,\end{aligned}$$

which is a fairly cumbersome calculation. If we use the Poisson approximation with $\lambda=1500\left(\frac{1}{500}\right)=3$, we have

$$P(X\le 2)\approx e^{-3}\left(1+3+\frac{3^2}{2}\right)=.4232. \qquad \|$$

Negative Binomial Distribution

The binomial distribution counts the number of successes in a fixed number of Bernoulli trials. Suppose that, instead, we count the number of Bernoulli trials required to get a fixed number of successes. This latter formulation leads to the negative binomial distribution.

In a sequence of independent Bernoulli(p) trials, let the random variable X denote the trial at which the rth success occurs, where r is a fixed integer. Then

$$P(X = x|r,p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r,\ r+1, \ldots, \tag{3.2.9}$$

and we say that X has a *negative binomial*(r,p) *distribution.*

The derivation of (3.2.9) follows quickly from the binomial distribution. The event $\{X = x\}$ can occur only if there are exactly $r-1$ successes in the first $x-1$ trials, and a success on the xth trial. The probability of $r-1$ successes in $x-1$ trials is the binomial probability $\binom{x-1}{r-1} p^{r-1}(1-p)^{x-r}$, and with probability p there is a success on the xth trial. Multiplying these probabilities gives (3.2.9).

The negative binomial distribution is sometimes defined in terms of the random variable Y = number of failures before the rth success. This formulation is statistically equivalent to the one given above in terms of X = trial at which the rth success occurs, since $Y = X - r$. Using the relationship between Y and X, the alternative form of the negative binomial distribution is

$$P(Y = y) = \binom{r+y-1}{y} p^r (1-p)^y, \quad y = 0, 1, \ldots. \tag{3.2.10}$$

Unless otherwise noted, when we refer to the negative binomial(r,p) distribution we will use this pmf.

The negative binomial distribution gets its name from the relationship

$$\binom{r+y-1}{y} = (-1)^y \binom{-r}{y} = (-1)^y \frac{(-r)(-r-1)(-r-2)\cdot\cdots\cdot(-r-y+1)}{(y)(y-1)(y-2)\cdot\cdots\cdot(2)(1)},$$

which is, in fact, the defining equation for binomial coefficients with negative integers (see Feller 1968 for a complete treatment). Substituting into (3.2.10) yields

$$P(Y = y) = (-1)^y \binom{-r}{y} p^r (1-p)^y,$$

which bears a striking resemblance to the binomial distribution.

The fact that $\sum_{y=0}^{\infty} P(Y = y) = 1$ is not easy to verify but follows from an extension of the Binomial Theorem, an extension that includes negative exponents. We will not pursue this further here. An excellent exposition on binomial coefficients can be found in Feller (1968).

The mean and variance of Y can be calculated using techniques similar to those used for the binomial distribution:

$$
\begin{aligned}
\mathrm{E}Y &= \sum_{y=0}^{\infty} y \binom{r+y-1}{y} p^r (1-p)^y \\
&= \sum_{y=1}^{\infty} \frac{(r+y-1)!}{(y-1)!(r-1)!} p^r (1-p)^y \\
&= \sum_{y=1}^{\infty} r \binom{r+y-1}{y-1} p^r (1-p)^y.
\end{aligned}
$$

Now write $z = y - 1$, and the sum becomes

$$
\begin{aligned}
\mathrm{E}Y &= \sum_{z=0}^{\infty} r \binom{r+z}{z} p^r (1-p)^{z+1} \\
&= r\frac{(1-p)}{p} \sum_{z=0}^{\infty} \binom{(r+1)+z-1}{z} p^{r+1}(1-p)^z \qquad \begin{pmatrix}\text{summand is negative} \\ \text{binomial pmf}\end{pmatrix} \\
&= r\frac{(1-p)}{p}.
\end{aligned}
$$

Since the sum is over all values of a negative binomial$(r+1, p)$ distribution, it equals 1. A similar calculation will show

$$\mathrm{Var}\, Y = \frac{r(1-p)}{p^2}.$$

There is an interesting, and sometimes useful, reparameterization of the negative binomial distribution in terms of its mean. If we define the parameter $\mu = r(1-p)/p$, then $\mathrm{E}Y = \mu$ and a little algebra will show that

$$\mathrm{Var}\, Y = \mu + \frac{1}{r}\mu^2.$$

The variance is a quadratic function of the mean. This relationship can be useful in both data analysis and theoretical considerations (Morris 1982).

The negative binomial family of distributions includes the Poisson distribution as a limiting case. If $r \to \infty$ and $p \to 1$ such that $r(1-p) \to \lambda, 0 < \lambda < \infty$, then

$$
\begin{aligned}
\mathrm{E}Y &= \frac{r(1-p)}{p} \to \lambda, \\
\mathrm{Var}\, Y &= \frac{r(1-p)}{p^2} \to \lambda,
\end{aligned}
$$

which agree with the Poisson mean and variance. To demonstrate that the negative binomial$(r, p) \to$ Poisson(λ), we can show that all of the probabilities converge. The fact that the mgfs converge leads us to expect this (see Exercise 3.15). ∥

Example 3.2.6 (Inverse binomial sampling) A technique known as inverse binomial sampling is useful in sampling biological populations. If the proportion of individuals possessing a certain characteristic is p and we sample until we see r such individuals, then the number of individuals sampled is a negative binomial random variable.

For example, suppose that in a population of fruit flies we are interested in the proportion having vestigial wings and decide to sample until we have found 100 such flies. The probability that we will have to examine at least N flies is (using (3.2.9))

$$\begin{aligned} P(X \geq N) &= \sum_{x=N}^{\infty} \binom{x-1}{99} p^{100}(1-p)^{x-100} \\ &= 1 - \sum_{x=100}^{N-1} \binom{x-1}{99} p^{100}(1-p)^{x-100}. \end{aligned}$$

For given p and N, we can evaluate this expression to determine how many fruit flies we are likely to look at. (Although the evaluation is cumbersome, the use of a recursion relation will speed things up.) ||

Example 3.2.6 shows that the negative binomial distribution can, like the Poisson, be used to model phenomena in which we are waiting for an occurrence. In the negative binomial case we are waiting for a specified number of successes.

Geometric Distribution

The geometric distribution is the simplest of the waiting time distributions and is a special case of the negative binomial distribution. If we set $r = 1$ in (3.2.9) we have

$$P(X = x|p) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots,$$

which defines the pmf of a *geometric random variable* X with success probability p. X can be interpreted as the trial at which the first success occurs, so we are "waiting for a success." The fact that $\sum_{x=1}^{\infty} P(X = x) = 1$ follows from properties of the geometric series. For any number a with $|a| < 1$,

$$\sum_{x=1}^{\infty} a^{x-1} = \frac{1}{1-a},$$

which we have already encountered in Example 1.5.4.

The mean and variance of X can be calculated by using the negative binomial formulas and by writing $X = Y + 1$ to obtain

$$\mathrm{E}X = \mathrm{E}Y + 1 = \frac{1}{p} \quad \text{and} \quad \operatorname{Var} X = \frac{1-p}{p^2}.$$

The geometric distribution has an interesting property, known as the "memoryless" property. For integers $s > t$, it is the case that

$$P(X > s|X > t) = P(X > s - t); \tag{3.2.11}$$

that is, the geometric distribution "forgets" what has occurred. The probability of getting an additional $s-t$ failures, having already observed t failures, is the same as the probability of observing $s-t$ failures at the start of the sequence. In other words, the probability of getting a run of failures depends only on the length of the run, not on its position.

To establish (3.2.11), we first note that for any integer n,

$$\begin{aligned} P(X>n) &= P(\text{no successes in } n \text{ trials}) \\ &= (1-p)^n, \end{aligned} \tag{3.2.12}$$

and hence

$$\begin{aligned} P(X>s|X>t) &= \frac{P(X>s \text{ and } X>t)}{P(X>t)} \\ &= \frac{P(X>s)}{P(X>t)} \\ &= (1-p)^{s-t} \\ &= P(X>s-t). \end{aligned}$$

Example 3.2.7 (Failure times) The geometric distribution is sometimes used to model "lifetimes" or "time until failure" of components. For example, if the probability is .001 that a light bulb will fail on any given day, then the probability that it will last at least 30 days is

$$P(X>30) = \sum_{x=31}^{\infty} .001(1-.001)^{x-1} = (.999)^{30} = .970. \qquad \|$$

The memoryless property of the geometric distribution describes a very special "lack of aging" property. It indicates that the geometric distribution is not applicable to modeling lifetimes for which the probability of failure is expected to increase with time. There are other distributions used to model various types of aging; see, for example, Barlow and Proschan (1975).

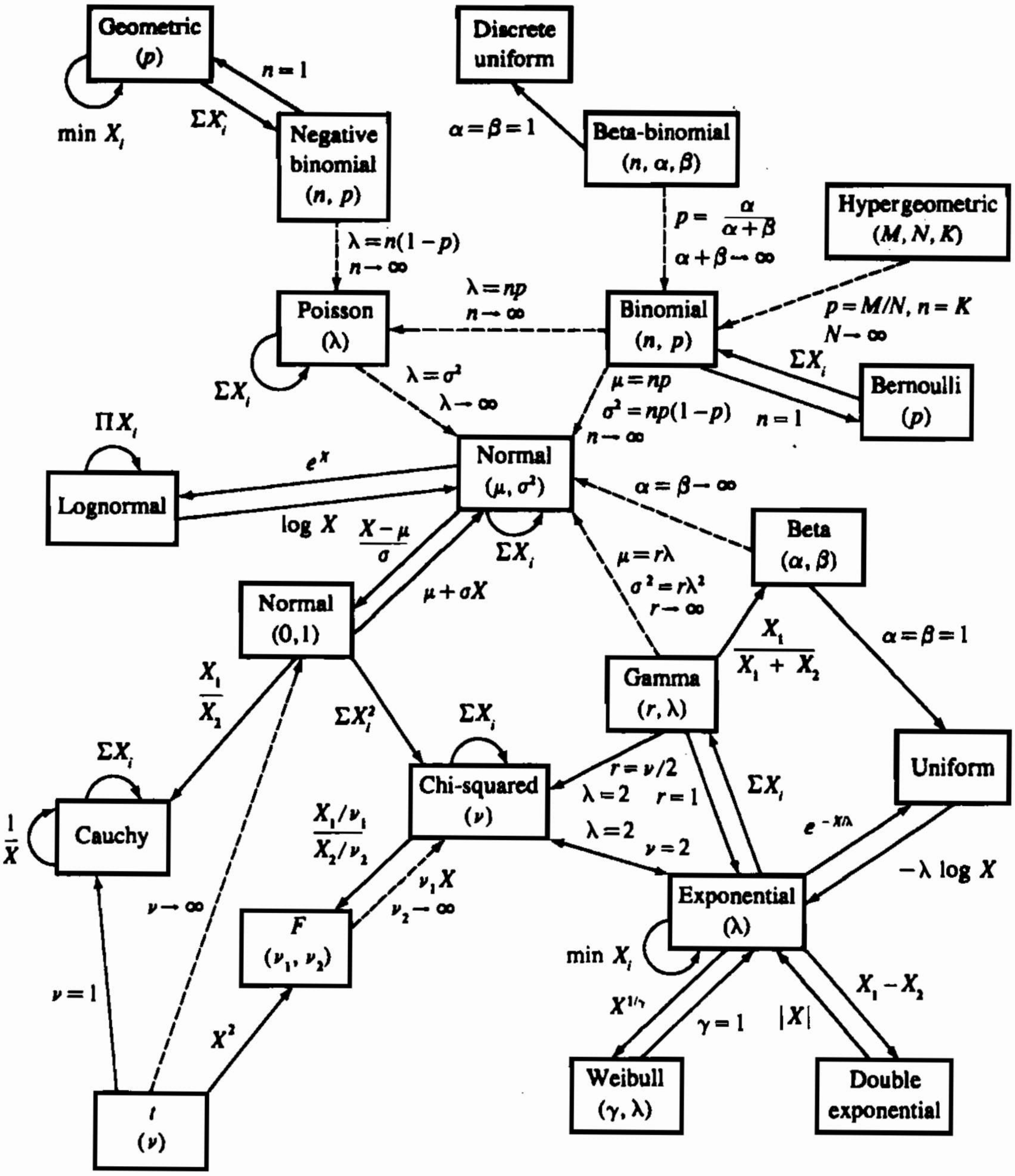
3.3 Continuous Distributions

In this section we will discuss some of the more common families of continuous distributions, those with well-known names. The distributions mentioned here by no means constitute all of the distributions used in statistics. Indeed, as was seen in Section 1.6, any nonnegative, integrable function can be transformed into a pdf.

Uniform Distribution

The continuous *uniform distribution* is defined by spreading mass uniformly over an interval $[a,b]$. Its pdf is given by

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise.} \end{cases} \tag{3.3.1}$$



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).