Chapter 3

Some Special Distributions

3.1 The Binomial and Related Distributions

In Chapter 1 we introduced the *uniform distribution* and the *hypergeometric distribution*. In this chapter we discuss some other important distributions of random variables frequently used in statistics. We begin with the binomial and related distributions.

A **Bernoulli experiment** is a random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, for instance, success or failure (e.g., female or male, life or death, nondefective or defective). A sequence of **Bernoulli trials** occurs when a Bernoulli experiment is performed several independent times so that the probability of success, say p, remains the same from trial to trial. That is, in such a sequence, we let p denote the probability of success on each trial.

Let X be a random variable associated with a Bernoulli trial by defining it as follows:

$$X(\text{success}) = 1 \quad \text{and } X(\text{failure}) = 0.$$

That is, the two outcomes, success and failure, are denoted by one and zero, respectively. The pmf of X can be written as

$$p(x) = p^x(1-p)^{1-x}, \quad x = 0, 1, \tag{3.1.1}$$

and we say that X has a *Bernoulli distribution*. The expected value of X is

$$\mu = E(X) = \sum_{x=0}^{1} xp^x(1-p)^{1-x} = (0)(1-p) + (1)(p) = p,$$

and the variance of X is

$$\begin{aligned}\sigma^2 = \operatorname{var}(X) &= \sum_{x=0}^{1}(x-p)^2p^x(1-p)^{1-x} \\ &= p^2(1-p) + (1-p)^2p = p(1-p).\end{aligned}$$

It follows that the standard deviation of X is $\sigma = \sqrt{p(1-p)}$.

In a sequence of n Bernoulli trials, we shall let X_i denote the Bernoulli random variable associated with the ith trial. An observed sequence of n Bernoulli trials will then be an n-tuple of zeros and ones. In such a sequence of Bernoulli trials, we are often interested in the total number of successes and not in the order of their occurrence. If we let the random variable X equal the number of observed successes in n Bernoulli trials, the possible values of X are $0, 1, 2, \ldots, n$. If x successes occur, where $x = 0, 1, 2, \ldots, n$, then $n - x$ failures occur. The number of ways of selecting the x positions for the x successes in the n trials is

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Since the trials are independent and the probabilities of success and failure on each trial are, respectively, p and $1 - p$, the probability of each of these ways is $p^x(1-p)^{n-x}$. Thus the pmf of X, say $p(x)$, is the sum of the probabilities of these $\binom{n}{x}$ mutually exclusive events; that is,

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

Recall, if n is a positive integer, that

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} b^x a^{n-x}.$$

Thus it is clear that $p(x) \geq 0$ and that

$$\begin{aligned} \sum_x p(x) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \\ &= [(1-p)+p]^n = 1. \end{aligned}$$

Therefore, $p(x)$ satisfies the conditions of being a pmf of a random variable X of the discrete type. A random variable X that has a pmf of the form of $p(x)$ is said to have a **binomial distribution**, and any such $p(x)$ is called a **binomial pmf**. A binomial distribution will be denoted by the symbol $b(n, p)$. The constants n and p are called the **parameters** of the binomial distribution. Thus, if we say that X is $b(5, \frac{1}{3})$, we mean that X has the binomial pmf

$$p(x) = \begin{cases} \binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x} & x = 0, 1, \ldots, 5 \\ 0 & \text{elsewhere.} \end{cases} \tag{3.1.2}$$

The mgf of a binomial distribution is easily obtained as follows,

$$\begin{aligned} M(t) &= \sum_x e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= [(1-p) + pe^t]^n \end{aligned}$$

for all real values of t. The mean μ and the variance σ^2 of X may be computed from $M(t)$. Since

$$M'(t) = n[(1-p) + pe^t]^{n-1}(pe^t)$$

and

$$M''(t) = n[(1-p+pe^t]^{n-1}(pe^t) + n(n-1)[(1-p)+pe^t]^{n-2}(pe^t)^2,$$

if follows that

$$\mu = M'(0) = np$$

and

$$\sigma^2 = M''(0) - \mu^2 = np + n(n-1)p^2 - (np)^2 = np(1-p).$$

Example 3.1.1. Let X be the number of heads (successes) in $n = 7$ independent tosses of an unbiased coin. The pmf of X is

$$p(x) = \begin{cases} \binom{7}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{7-x} & x = 0, 1, 2, \ldots, 7 \\ 0 & \text{elsewhere.} \end{cases}$$

Then X has the mgf

$$M(t) = (\tfrac{1}{2} + \tfrac{1}{2}e^t)^7,$$

has mean $\mu = np = \frac{7}{2}$, and has variance $\sigma^2 = np(1-p) = \frac{7}{4}$. Furthermore, we have

$$P(0 \le X \le 1) = \sum_{x=0}^{1} p(x) = \frac{1}{128} + \frac{7}{128} = \frac{8}{128}$$

and

$$P(X = 5) = p(5) = \frac{7!}{5!2!}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^2 = \frac{21}{128}. \quad \blacksquare$$

Most computer packages have commands which obtain the binomial probabilities. To give the R (Ihaka and Gentleman, 1996) or S-PLUS (S-PLUS, 2000) commands, suppose X has a $b(n,p)$ distribution. Then the command **dbinom(k,n,p)** returns $P(X = k)$, while the command **pbinom(k,n,p)** returns the cumulative probability $P(X \le k)$.

Example 3.1.2. If the mgf of a random variable X is

$$M(t) = (\tfrac{2}{3} + \tfrac{1}{3}e^t)^5,$$

then X has a binomial distribution with $n = 5$ and $p = \frac{1}{3}$; that is, the pmf of X is

$$p(x) = \begin{cases} \binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x} & x = 0, 1, 2, \ldots, 5 \\ 0 & \text{elsewhere.} \end{cases}$$

Here $\mu = np = \frac{5}{3}$ and $\sigma^2 = np(1-p) = \frac{10}{9}$. $\blacksquare$

Example 3.1.3. If Y is $b(n, \frac{1}{3})$, then $P(Y \geq 1) = 1 - P(Y = 0) = 1 - (\frac{2}{3})^n$. Suppose that we wish to find the smallest value of n that yields $P(Y \geq 1) > 0.80$. We have $1 - (\frac{2}{3})^n > 0.80$ and $0.20 > (\frac{2}{3})^n$. Either by inspection or by use of logarithms, we see that $n = 4$ is the solution. That is, the probability of at least one success throughout $n = 4$ independent repetitions of a random experiment with probability of success $p = \frac{1}{3}$ is greater than 0.80. ■

Example 3.1.4. Let the random variable Y be equal to the number of successes throughout n independent repetitions of a random experiment with probability p of success. That is, Y is $b(n, p)$. The ratio Y/n is called the relative frequency of success. Recall expression (1.10.3), the second version of Chebyshev's inequality (Theorem 1.10.3). Applying this result, we have for all $\epsilon > 0$ that

$$P\left(\left|\frac{Y}{n} - p\right| \geq \epsilon\right) \leq \frac{\text{Var}(Y/n)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}.$$

Now, for every fixed $\epsilon > 0$, the right-hand member of the preceding inequality is close to zero for sufficiently large n. That is,

$$\lim_{n\to\infty} P\left(\left|\frac{Y}{n} - p\right| \geq \epsilon\right) = 0$$

and

$$\lim_{n\to\infty} P\left(\left|\frac{Y}{n} - p\right| < \epsilon\right) = 1.$$

Since this is true for every fixed $\epsilon > 0$, we see, in a certain sense, that the relative frequency of success is for large values of n, close to the probability of p of success. This result is one form of the *Weak Law of Large Numbers*. It was alluded to in the initial discussion of probability in Chapter 1 and will be considered again, along with related concepts, in Chapter 4. ■

Example 3.1.5. Let the independent random variables X_1, X_2, X_3 have the same cdf $F(x)$. Let Y be the middle value of X_1, X_2, X_3. To determine the cdf of Y, say $F_Y(y) = P(Y \leq y)$, we note that $Y \leq y$ if and only if at least two of the random variables X_1, X_2, X_3 are less than or equal to y. Let us say that the ith "trial" is a success if $X_i \leq y$, $i = 1, 2, 3$; here each "trial" has the probability of success $F(y)$. In this terminology, $F_Y(y) = P(Y \leq y)$ is then the probability of at least two successes in three independent trials. Thus

$$F_Y(y) = \binom{3}{2}[F(y)]^2[1 - F(y)] + [F(y)]^3.$$

If $F(x)$ is a continuous cdf so that the pdf of X is $F'(x) = f(x)$, then the pdf of Y is

$$f_Y(y) = F'_Y(y) = 6[F(y)][1 - F(y)]f(y). \quad ■$$

Example 3.1.6. Consider a sequence of independent repetitions of a random experiment with constant probability p of success. Let the random variable Y denote the total number of failures in this sequence before the rth success, that is,

$Y + r$ is equal to the number of trials necessary to produce exactly r successes. Here r is a fixed positive integer. To determine the pmf of Y, let y be an element of $\{y : y = 0, 1, 2, \ldots\}$. Then, by the multiplication rule of probabilities, $P(Y = y) = g(y)$ is equal to the product of the probability

$$\binom{y+r-1}{r-1} p^{r-1}(1-p)^y$$

of obtaining exactly $r - 1$ successes in the first $y + r - 1$ trials and the probability p of a success on the $(y + r)$th trial. Thus the pmf of Y is

$$p_Y(y) = \begin{cases} \binom{y+r-1}{r-1} p^r(1-p)^y & y = 0, 1, 2, \ldots \\ 0 & \text{elsewhere.} \end{cases} \tag{3.1.3}$$

A distribution with a pmf of the form $p_Y(y)$ is called a **negative binomial distribution**; and any such $p_Y(y)$ is called a negative binomial pmf. The distribution derives its name from the fact that $p_Y(y)$ is a general term in the expansion of $p^r[1 - (1 - p)]^{-r}$. It is left as an exercise to show that the mgf of this distribution is $M(t) = p^r[1 - (1 - p)e^t]^{-r}$, for $t < -\ln(1 - p)$. If $r = 1$, then Y has the pmf

$$p_Y(y) = p(1-p)^y, \quad y = 0, 1, 2, \ldots, \tag{3.1.4}$$

zero elsewhere, and the mgf $M(t) = p[1 - (1 - p)e^t]^{-1}$. In this special case, $r = 1$, we say that Y has a **geometric distribution** of the form. ■

Suppose we have several independent binomial distributions with the same probability of success. Then it makes sense that the sum of these random variables is binomial, as shown in the following theorem. Note that the mgf technique gives a quick and easy proof.

Theorem 3.1.1. *Let $X_1, X_2, \ldots, X_m$ be independent random variables such that X_i has binomial $b(n_i, p)$ distribution, for $i = 1, 2, \ldots, m$. Let $Y = \sum_{i=1}^m X_i$. Then Y has a binomial $b(\sum_{i=1}^m n_i, p)$ distribution.*

Proof: Using independence of the X_is and the mgf of X_i, we obtain the mgf of Y as follows:

$$\begin{aligned} M_Y(t) &= E\left[\exp\{\sum_{i=1}^m tX_i\}\right] = E\left[\prod_{i=1}^m \exp\{tX_i\}\right] \\ &= \prod_{i=1}^m E\left[\exp\{tX_i\}\right] = \prod_{i=1}^m (1-p+pe^t)^{n_i} = (1-p+pe^t)^{\sum_{i=1}^m n_i}. \end{aligned}$$

Hence, Y has a binomial $b(\sum_{i=1}^m n_i, p)$ distribution. ■

The binomial distribution is generalized to the multinomial distribution as follows. Let a random experiment be repeated n independent times. On each repetition, the experiment results in but one of k mutually exclusive and exhaustive ways, say $C_1, C_2, \ldots, C_k$. Let p_i be the probability that the outcome is an element of C_i

and let p_i remain constant throughout the n independent repetitions, $i = 1, 2, \ldots, k$. Define the random variable X_i to be equal to the number of outcomes that are elements of C_i, $i = 1, 2, \ldots, k-1$. Furthermore, let $x_1, x_2, \ldots, x_{k-1}$ be nonnegative integers so that $x_1 + x_2 + \cdots + x_{k-1} \le n$. Then the probability that exactly x_1 terminations of the experiment are in $C_1, \ldots,$ exactly x_{k-1} terminations are in C_{k-1}, and hence exactly $n - (x_1 + \cdots + x_{k-1})$ terminations are in C_k is

$$\frac{n!}{x_1! \cdots x_{k-1}! x_k!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} p_k^{x_k},$$

where x_k is merely an abbreviation for $n - (x_1 + \cdots + x_{k-1})$. This is the **multinomial pmf** of $k-1$ random variables $X_1, X_2, \ldots, X_{k-1}$ of the discrete type. To see that this is correct, note that the number of distinguishable arrangements of $x_1\ C_1$s, $x_2\ C_2$s, $\ldots, x_k\ C_k$s is

$$\binom{n}{x_1}\binom{n-x_1}{x_2}\cdots\binom{n-x_1-\cdots-x_{k-2}}{x_{k-1}} = \frac{n!}{x_1! x_2! \cdots x_k!}$$

and the probability of each of these distinguishable arrangements is

$$p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

Hence the product of these two latter expressions gives the correct probability, which is an agreement with the formula for the multinomial pmf.

When $k = 3$, we often let $X = X_1$ and $Y = X_2$; then $n - X - Y = X_3$. We say that X and Y have a **trinomial distribution**. The joint pmf of X and Y is

$$p(x, y) = \frac{n!}{x! y! (n-x-y)!} p_1^x p_2^y p_3^{n-x-y},$$

where x and y are nonnegative integers with $x+y \le n$, and p_1, p_2, and p_3 are positive proper fractions with $p_1 + p_2 + p_3 = 1$; and let $p(x, y) = 0$ elsewhere. Accordingly, $p(x, y)$ satisfies the conditions of being a joint pmf of two random variables X and Y of the discrete type; that is, $p(x, y)$ is nonnegative and its sum over all points (x, y) at which $p(x, y)$ is positive is equal to $(p_1 + p_2 + p_3)^n = 1$.

If n is a positive integer and a_1, a_2, a_3 are fixed constants, we have

$$\begin{aligned}
&\sum_{x=0}^{n}\sum_{y=0}^{n-x} \frac{n!}{x! y! (n-x-y)!} a_1^x a_2^y a_3^{n-x-y} \\
&= \sum_{x=0}^{n} \frac{n! a_1^x}{x!(n-x)!} \sum_{y=0}^{n-x} \frac{(n-x)!}{y!(n-x-y)!} a_2^y a_3^{n-x-y} \\
&= \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} a_1^x (a_2 + a_3)^{n-x} \\
&= (a_1 + a_2 + a_3)^n. \qquad (3.1.5)
\end{aligned}$$

Consequently, the mgf of a trinomial distribution, in accordance with Equation (3.1.5), is given by

$$\begin{aligned} M(t_1,t_2) &= \sum_{x=0}^{n}\sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!}(p_1e^{t_1})^x(p_2e^{t_2})^y p_3^{n-x-y} \\ &= (p_1e^{t_1}+p_2e^{t_2}+p_3)^n, \end{aligned}$$

for all real values of t_1 and t_2. The moment-generating functions of the marginal distributions of X and Y are, respectively,

$$M(t_1,0) = (p_1e^{t_1}+p_2+p_3)^n = [(1-p_1)+p_1e^{t_1}]^n$$

and

$$M(0,t_2) = (p_1+p_2e^{t_2}+p_3)^n = [(1-p_2)+p_2e^{t_2}]^n.$$

We see immediately, from Theorem 2.5.5 that X and Y are dependent random variables. In addition, X is $b(n,p_1)$ and Y is $b(n,p_2)$. Accordingly, the means and variances of X and Y are, respectively, $\mu_1 = np_1$, $\mu_2 = np_2$, $\sigma_1^2 = np_1(1-p_1)$, and $\sigma_2^2 = np_2(1-p_2)$.

Consider next the conditional pmf of Y, given $X = x$. We have

$$p_{2|1}(y|x) = \begin{cases} \frac{(n-x)!}{y!(n-x-y)!}\left(\frac{p_2}{1-p_1}\right)^y\left(\frac{p_3}{1-p_1}\right)^{n-x-y} & y = 0,1,\ldots,n-x \\ 0 & \text{elsewhere.} \end{cases}$$

Thus the conditional distribution of Y, given $X = x$, is $b[n-x, p_2/(1-p_1)]$. Hence the conditional mean of Y, given $X = x$, is the linear function

$$E(Y|x) = (n-x)\left(\frac{p_2}{1-p_1}\right).$$

Also, the conditional distribution of X, given $Y = y$, is $b[n-y, p_1/(1-p_2)]$ and thus

$$E(X|y) = (n-y)\left(\frac{p_1}{1-p_2}\right).$$

Now recall from Example 2.4.2 that the square of the correlation coefficient ρ^2 is equal to the product of $-p_2/(1-p_1)$ and $-p_1/(1-p_2)$, the coefficients of x and y in the respective conditional means. Since both of these coefficients are negative (and thus ρ is negative), we have

$$\rho = -\sqrt{\frac{p_1p_2}{(1-p_1)(1-p_2)}}.$$

In general, the mgf of a multinomial distribution is given by

$$M(t_1,\ldots,t_{k-1}) = (p_1e^{t_1}+\cdots+p_{k-1}e^{t_{k-1}}+p_k)^n$$

for all real values of $t_1, t_2, \ldots, t_{k-1}$. Thus each one-variable marginal pmf is binomial, each two-variable marginal pmf is trinomial, and so on.