CHAPTER 3

Binomial Distribution

3.1 DEFINITION

The *binomial distribution* can be defined, using the binomial expansion

$$(q+p)^n = \sum_{x=0}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{x=0}^{n} \frac{n!}{k!(n-k!)} p^k q^{n-k},$$

as the distribution of a random variable X for which

$$\Pr[X = x] = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, \ldots, n, \tag{3.1}$$

where $q + p = 1$, $p > 0$, $q > 0$, and n is a positive integer. Occasionally a more general form is used in which the variable X is transformed to $a + bX$, where a and b are real numbers with $b \neq 0$. When $n = 1$, the distribution is known as the *Bernoulli distribution*.

The characteristic function (cf) of the binomial distribution is $(1 - p + pe^{it})^n$, and the probability generating function (pgf) is

$$\begin{aligned} G(z) &= (1 - p + pz)^n = (q + pz)^n \\ &= \frac{{}_1F_0[-n; -; -pz/q]}{{}_1F_0[-n; -; -p/q]} \qquad (3.2) \\ &= {}_1F_0[-n; -; p(1-z)], \qquad 0 < p < 1. \qquad (3.3) \end{aligned}$$

The mean and variance are

$$\mu = np \qquad \text{and} \qquad \mu_2 = npq. \tag{3.4}$$

The distribution is a power series distribution (PSD) with finite support. From (3.2) it is a generalized hypergeometric probability distribution (GHPD) and from (3.3) it is a generalized hypergeometric factorial moment distribution. It is a member of the exponential family of distributions (when n is known), and it is an Ord and also a Katz distribution; for more details see Section 3.4.

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

3.2 HISTORICAL REMARKS AND GENESIS

If n independent trials are made and in each there is probability p that the outcome E will occur, then the number of trials in which E occurs can be represented by a rv X having the binomial distribution with parameters n, p. This situation occurs when a sample of fixed size n is taken from an *infinite* population where each element in the population has an equal and independent probability p of possession of a specified attribute. The situation also arises when a sample of fixed size n is taken from a *finite* population where each element in the population has an equal and independent probability p of having a specified attribute and elements are sampled independently and sequentially with replacement.

The binomial distribution is one of the oldest to have been the subject of study. The distribution was derived by James Bernoulli (in his treatise *Ars Conjectandi*, published in 1713), for the case $p = r/(r+s)$, where r and s are positive integers. Earlier Pascal had considered the case $p = \frac{1}{2}$. In his *Essay*, published posthumously in 1764, Bayes removed the rational restriction on p by considering the position relative to a randomly rolled ball of a second ball randomly rolled n times. The early history of the distribution is discussed, inter alia, by Boyer (1950), Stigler (1986), Edwards (1987), and Hald (1990).

A remarkable new derivation as the solution of the simple birth-and-emigration process was given by McKendrick (1914). The distribution may also be regarded as the stationary distribution for the Ehrenfest model (Feller, 1957). Haight (1957) has shown that the M/M/1 queue with balking gives rise to the distribution, provided that the arrival rate of the customers when there are n customers in the queue is $\lambda = (N-n)N^{-1}(n+1)^{-1}$ for $n < N$ and zero for $n \geq N$ (N is the maximum queue size).

3.3 MOMENTS

The moment generating function (mgf) is $(q + pe^t)^n$ and the cumulant generating function (cgf) is $n \ln(q + pe^t)$. The factorial cumulant generating function (fcgf) is $n \ln(1 + pt)$, whence $\kappa_{[r]} = n(r-1)!p^r$. The factorial moments can be obtained straightforwardly from the factorial moment generating function (fmgf), which is $(1 + pt)^n$. We have

$$\mu'_{[r]} = \frac{n!p^r}{(n-r)!};$$

that is,

$$\begin{aligned}
\mu'_{[1]} &= \mu = np, \\
\mu'_{[2]} &= n(n-1)p^2, \\
\mu'_{[3]} &= n(n-1)(n-2)p^3, \\
&\vdots
\end{aligned} \tag{3.5}$$

From $\mu'_r = \sum_{j=0}^{r} S(r, j)\mu'_{[j]}$ (see Section 1.2.7), it follows that the rth moment about zero is

$$\mu'_r = E[X^r] = \sum_{j=0}^{r} \frac{S(r, j)n!p^r}{(n-r)!}. \tag{3.6}$$

In particular

$$\mu'_1 = np,$$
$$\mu'_2 = np + n(n-1)p^2,$$
$$\mu'_3 = np + 3n(n-1)p^2 + n(n-1)(n-2)p^3,$$
$$\mu'_4 = np + 7n(n-1)p^2 + 6n(n-1)(n-2)p^3 + n(n-1)(n-2)(n-3)p^4.$$

Hence (or otherwise) the central moments can be obtained. The lower order central moments are

$$\begin{aligned} \mu_2 &= \sigma^2 = npq, \\ \mu_3 &= npq(q-p), \\ \mu_4 &= 3(npq)^2 + npq(1-6pq). \end{aligned} \tag{3.7}$$

The moment ratios $\sqrt{\beta_1}$ and β_2 are

$$\sqrt{\beta_1} = (q-p)(npq)^{-1/2}, \qquad \beta_2 = 3 + (1-6pq)(npq)^{-1}. \tag{3.8}$$

For a fixed value of p (and so of q) the (β_1, β_2) points fall on the straight line

$$\frac{\beta_2 - 3}{\beta_1} = \frac{1-6pq}{(q-p)^2} = 1 - \frac{2pq}{(q-p)^2}.$$

As $n \to \infty$, the points approach the limit (0, 3).

Note that the same straight line is obtained when p is replaced by q. The two distributions are mirror images of each other, so they have identical values of β_2 and the same absolute value of $\sqrt{\beta_1}$. The slope $(\beta_2 - 3)/\beta_1$ is always less than 1. The limit of the ratio as p approaches 0 or 1 is 1. For $p = q = 0.5$ the binomial distribution is symmetrical and $\beta_1 = 0$. For $n = 1$, the point (β_1, β_2) lies on the line $\beta_2 - \beta_1 - 1 = 0$. (Note that for *any* distribution $\beta_2 - \beta_1 - 1 \geq 0$.)

Romanovsky (1923) derived the following recursion formula for the central moments:

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right). \tag{3.9}$$

An analogous relation holds for moments about zero,

$$\mu'_{r+1} = pq\left[\left(\frac{n}{q}\right)\mu'_r + \frac{d\mu'_r}{dp}\right]. \tag{3.10}$$

Kendall (1943) used differentiation of the cf to derive the relationship

$$\mu_r = npq\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_j - p\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_{j+1}. \tag{3.11}$$

A simpler recursion formula holds for the cumulants

$$\kappa_{r+1} = pq\frac{\partial\kappa_r}{\partial p}, \qquad r \geq 1. \tag{3.12}$$

Formula (3.10) also holds for the incomplete moments, defined as

$$\mu'_{j,k} = \sum_{i=k}^{n} i^j \binom{n}{i} p^i q^{n-i}.$$

The mean deviation is

$$\nu_1 = E[|X - np|] = 2n\binom{n-1}{[np]}p^{[np]+1}q^{n-[np]}, \tag{3.13}$$

where [·] denotes the integer part [see Bertrand (1889), Frisch (1924), and Frame (1945)]. Diaconis and Zarbell (1991) have discussed the provenance and import of this formula and other equivalent formulas. They found that ν_1 is an increasing function of n but that ν_1/n is a decreasing function of n. Johnson's (1957) article led to a number of generalizations. Using Stirling's approximation for $n!$,

$$\nu_1 \approx \left(\frac{2npq}{\pi}\right)^{1/2}\left[1 + \frac{(np - [np])(nq - [nq])}{2npq} - \frac{1 - 2pq}{12npq}\right]; \tag{3.14}$$

this shows that the ratio of the mean deviation to the standard deviation approaches the limiting value $(2/\pi)^{1/2} \approx 0.798$ as $n \to \infty$.

Katti (1960) devised an ingenious method for obtaining the absolute moments of general order about m. All inverse moments of the binomial distribution [i.e., $E(X^{-r})$ with $r = 1, 2, \ldots$] are infinite because $\Pr[X = 0] > 0$. Inverse moments of the *positive binomial distribution* (formed by zero truncation) are discussed in Section 3.11.

Direct manipulation of the definition of the *inverse factorial moment*, as in Stancu (1968), yields

$$E[\{(X + r)^{(r)}\}^{-1}] = [(n + r)^{(r)}]^{-1}p^{-r}\left[1 - \sum_{y=0}^{r-1}\binom{n+r}{y}p^y q^{n+r-y}\right]. \tag{3.15}$$

Chao and Strawderman (1972) obtained a general result for $E[(X+a)^{-k}]$ in terms of a k-fold multiple integral of $t^{-1}E[t^{X+a-1}]$, and they applied this to the binomial distribution. A modification of their approach enabled Lepage (1978) to express the inverse ascending factorial moment

$$R_x(a,k) = E[\{(X+a)\cdots(X+a+k-1)\}^{-1}]$$

as a k-fold multiple integral of $E[t^{X+a-1}]$. Cressie et al. (1981) obtained $E[(X+a)^{-k}]$ both as a k-fold multiple integral of $E[e^{-t(X+a)}]$ and also from a single integral of $t^{k-1}E[e^{-t(X+a)}]$. Jones (1987) developed an analogous single-integral result for $R_x(a,k)$, namely,

$$R_x(a,k) = [\Gamma(k)]^{-1}\int_0^1 (1-t)^{k-1}E[t^{X+a-1}]\,dt,$$

and compared the different (though equivalent) expressions that are yielded by the different approaches for $R_x(a,k)$ for the binomial distribution.

3.4 PROPERTIES

The binomial distribution belongs to a number of families of distributions and hence possesses the properties of each of the families.

It is a distribution with finite support. As defined by (3.1), it consists of $n+1$ nonzero probabilities associated with the values $0, 1, 2, \ldots, n$ of the rv X. The ratio

$$\frac{\Pr[X=x+1]}{\Pr[X=x]} = \frac{(n-x)p}{(x+1)q}, \qquad x = 0, 1, \ldots, n-1, \tag{3.16}$$

shows that $\Pr[X=x]$ increases with x so long as $x < np - q = (n+1)p - 1$ and decreases with x if $x > np - q$. The distribution is therefore unimodal with the mode occurring at $x = [(n+1)p]$, where $[\cdot]$ denotes the integer part. If $(n+1)p$ is an integer, then there are joint modes at $x = np + p$ and $x = np - q$. When $p < (n+1)^{-1}$, the mode occurs at the origin.

The median is given by the minimum value of k for which

$$\sum_{j=0}^{k}\binom{n}{k}p^k q^{n-k} > \tfrac{1}{2}.$$

Kaas and Buhrman (1980) showed that

$$|\text{mean} - \text{median}| \le \max\{p, 1-p\}.$$

Hamza (1995) has sharpened this to

$$|\text{mean} - \text{median}| < \ln 2$$

when $p < 1 - \ln 2$ or $p > \ln 2$.

The distribution is a member of the exponential family of distributions with respect to $p/(1-p)$, since

$$\Pr[X = x] = \exp\left[x \ln\left(\frac{p}{1-p}\right) + \ln\binom{n}{x} + n\ln(1-p)\right].$$

Morris (1982, 1983) has shown that it is one of the six subclasses of the natural exponential family for which the variance is at most a quadratic function of the mean; he used this property to obtain unified results and to gain insight concerning limit laws. Unlike the other five subclasses, however, it is not infinitely divisible (no distribution with finite support can be infinitely divisible).

Because $\Pr[X = x]$ is of the form

$$\frac{b(x)\theta^x}{\eta(\theta)}, \qquad \theta > 0, \qquad x = 0, 1, \ldots, n,$$

where $\theta = p/(1-p)$ (Kosambi, 1949; Noack, 1950), the distribution belongs to the important family of PSDs (see Section 2.2). Patil has investigated these in depth [see, e.g., Patil (1986)]. He has shown that for the binomial distribution

$$\frac{\theta\eta^{(r+1)}(\theta)}{\eta^{(r)}(\theta)} = \mu - pr, \tag{3.17}$$

where $\eta(\theta) = \sum_x b(x)\theta^x$. Integral expressions for the tail probabilities of PSDs were obtained by Joshi (1974, 1975), who thereby demonstrated the duality between the binomial distribution and the beta distribution of the second kind. Indeed

$$\sum_{x=r}^{n}\binom{n}{x}p^x q^{n-x} = I_p(r, n-r+1) = \Pr\left[F \le \frac{\nu_2 p}{\nu_1 q}\right], \tag{3.18}$$

where F is a random variable that has an F distribution with parameters $\nu_1 = 2r$, $\nu_2 = 2(n-r+1)$; see Raiffa and Schlaifer (1961).

Berg (1974, 1983a) has explored the properties of the closely related family of factorial series distributions with

$$\Pr[X = x] = \frac{n^{(x)}c(x)}{x!h(n)},$$

to which the binomial distribution can be seen to belong by taking $c(x) = (p/q)^x$.

Expression (3.16) shows that the binomial distribution belongs to the discrete Pearson system (Katz, 1945, 1965; Ord, 1967b). Tripathi and Gurland (1977) have examined methods for selecting from those distributions having

$$\frac{\Pr[X = x+1]}{\Pr[X = x]} = \frac{A + Bx}{C + Dx + Ex^2}$$

a particular member such as the binomial.

Kemp (1968a,b) has shown that the binomial is a generalized hypergeometric distribution with pgf

$$G(z) = \frac{{}_1F_0[-n;\ ;pz/(p-1)]}{{}_1F_0[-n;\ ;p/(p-1)]}.$$

Moreover, because the ratio of successive factorial moments is $(n-r)p$, the distribution is also a generalized hypergeometric factorial moment distribution with pgf ${}_1F_0[-n;\ ;p(1-z)]$ (Kemp, 1968a; Kemp and Kemp, 1974). Its membership of those families enabled Kemp and Kemp to obtain differential equations and associated difference equations for the pgf and various mgf's, including the generating functions for the incomplete and the absolute moments.

The binomial has an increasing failure rate (Barlow and Proschan, 1965). The Mills ratio for a discrete distribution is defined as

$$\sum_{j\ge x} \Pr[X=j]/\Pr[X=x],$$

and therefore it is the reciprocal of the failure rate. Diaconis and Zarbell (1991) showed that the Mills ratio for the binomial distribution satisfies

$$\frac{x}{n} \le \frac{\sum_{j=x}^{n}\Pr[X=j]}{\Pr[X=x]} \le \frac{x(1-p)}{x-np},$$

provided that $x > np$. The binomial distribution is also a monotone likelihood-ratio distribution. The skewness of the distribution is positive if $p < 0.5$ and is negative if $p > 0.5$. The distribution is symmetrical iff $p = 0.5$.

Denoting $\Pr[X \le c]$ by $L_{n,c}(p)$, Uhlmann (1966) has shown that, for $n \ge 2$,

$$\begin{aligned}
L_{n,c}\left(\frac{c}{n-1}\right) &> \frac{1}{2} > L_{n,c}\left(\frac{c+1}{n+1}\right) \quad \text{for} \quad 0 \le c < \frac{n-1}{2},\\
L_{n,c}\left(\frac{c}{n-1}\right) &= \frac{1}{2} = L_{n,c}\left(\frac{c+1}{n+1}\right) \quad \text{for} \quad c = \frac{n-1}{2},\\
L_{n,c}\left(\frac{c+1}{n+1}\right) &> \frac{1}{2} > L_{n,c}\left(\frac{c}{n-1}\right) \quad \text{for} \quad \frac{n-1}{2} < c \le n.
\end{aligned} \tag{3.19}$$

The distribution of the *standardized binomial variable*

$$X' = \frac{X-np}{\sqrt{npq}}$$

tends to the unit-normal distribution as $n \to \infty$; that is, for any real numbers α, β (with $\alpha < \beta$)

$$\lim_{n\to\infty} \Pr[\alpha < X' < \beta] = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-u^2/2}\,du. \tag{3.20}$$

This result is known as the *De Moivre–Laplace theorem*. It forms a starting point for a number of approximations in the calculation of binomial probabilities; these will be discussed in Section 3.6.

In Section 3.1 we saw that the pgf of X is $(q + pz)^n$. If X_1, X_2 are independent rv's having binomial distributions with parameters n_1, p and n_2, p, respectively, then the pgf of $X_1 + X_2$ is $(q + pz)^{n_1}(q + pz)^{n_2} = (q + pt)^{n_1+n_2}$. Hence $X_1 + X_2$ has a binomial distribution with parameters $n_1 + n_2$, p. This property is also apparent on interpreting $X_1 + X_2$ as the number of occurrences of an outcome E having constant probability p in each of $n_1 + n_2$ independent trials.

The distribution of X_1, conditional on $X_1 + X_2 = k$, is

$$\Pr[X_1 = x|k] = \frac{\binom{n_1}{x} p^x q^{n_1-x} \binom{n_2}{k-x} p^{k-x} q^{n_2-k+x}}{\binom{n_1+n_2}{k} p^k q^{n_1+n_2-k}}$$

$$= \frac{\binom{n_1}{x}\binom{n_2}{k-x}}{\binom{n_1+n_2}{k}}, \tag{3.21}$$

where $\max(0, k - n_2) \le x \le \min(n_1, k)$. This is a *hypergeometric distribution*; see Chapter 6.

The distribution of the difference $X_1 - X_2$ is

$$\Pr[X_1 - X_2 = x] = \sum_{x_1} \binom{n_1}{x_1}\binom{n_2}{x_1 - x} p^{2x_1-x} q^{n_1+n_2-2x_1+x}, \tag{3.22}$$

where the summation is between the limits $\max(0, x) \le x_1 \le \min(n_1, n_2 + x)$.

When $p = q = 0.5$,

$$\Pr[X_1 - X_2 = x] = \binom{n_1+n_2}{n_2+x} 2^{-n_1-n_2}, \qquad -n_2 \le x \le n_1,$$

so that $X_1 - X_2$ has a binomial distribution of the more general form mentioned in Section 3.1.

From the De Moivre–Laplace theorem and the independence of X_1 and X_2, it follows that the distribution of the standardized difference

$$[X_1 - X_2 - p(n_1 - n_2)][pq(n_1 + n_2)]^{-1/2}$$

tends to the unit-normal distribution as $n_1 \to \infty$, $n_2 \to \infty$ (whatever the ratio n_1/n_2). A similar result also holds when X_1 and X_2 have binomial distributions with parameters n_1, p_1 and n_2, p_2 with $p_1 \ne p_2$; however, the conditional distribution of X_1, given $X_1 + X_2 = x$, is no longer hypergeometric. Its distribution has been studied by Stevens (1951) as well as by Hannan and Harkness (1963), who developed asymptotic normal approximations.

Springer (1979) has examined the distribution of products of discrete independent rv's; he used as an illustration the product of two binomial variables with parameters n_1, p_1 and n_2, p_2, where $n_1 = n_2 = 2$.

3.5 ORDER STATISTICS

As is the case for most discrete distributions, order statistics based on observed values of random variables with a common binomial distribution are not often used. Mention may be made, however, of discussions of binomial order statistics by Gupta (1965), Khatri (1962), and Siotani (1956); see also David (1981). Tables of the cumulative distribution of the smallest and largest order statistic and of the range [in random samples of sizes 1(1)20] are in Gupta (1960b), Siotani and Ozawa (1948), and Gupta and Panchapakesan (1974). These tables can be applied in selecting the largest binomial probability among a set of k, based on k independent series of trials. This problem has been considered by Somerville (1957) and by Sobel and Huyett (1957).

Gupta (1960b) and Gupta and Panchapakesan (1974) have tabulated the mean and variance of the smallest and largest order statistic. Balakrishnan (1986) has given general results for the moments of order statistics from discrete distributions, and he has discussed the use of his results in the case of the binomial distribution.

3.6 APPROXIMATIONS, BOUNDS, AND TRANSFORMATIONS

3.6.1 Approximations

The binomial distribution is of such importance in applied probability and statistics that it is frequently necessary to calculate probabilities based on this distribution. Although the calculation of sums of the form

$$\sum_x \binom{n}{x} p^x q^{n-x}$$

is straightforward, it can be tedious, especially when n and x are large and when there are a large number of terms in the summation. It is not surprising that a great deal of attention and ingenuity have been applied to constructing useful approximations for sums of this kind.

The *normal approximation* to the binomial distribution (based on the De Moivre–Laplace theorem)

$$\Pr[\alpha < (X - np)(npq)^{-1/2} < \beta] \approx \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-u^2/2}\, du = \Phi(\beta) - \Phi(\alpha) \tag{3.23}$$

has been mentioned in Section 3.4. This is a relatively crude approximation, but it can be useful when n is large. Numerical comparisons have been published in a number of textbooks (e.g., Hald, 1952).

A marked improvement is obtained by the use of a *continuity correction*. The following normal approximation is used widely on account of its simplicity:

$$\Pr[X \leq x] \approx \Phi\left(\frac{x + 0.5 - np}{(npq)^{1/2}}\right); \qquad (3.24)$$

its accuracy for various values of n and p was assessed by Raff (1956) and by Peizer and Pratt (1968), who used the absolute and the relative error, respectively. Various rules of thumb for its use have been recommended in various standard textbooks. Two such rules of thumb are

1. use when $np(1 - p) > 9$ and
2. use when $np > 9$ for $0 < p \leq 0.5 \leq q$.

Schader and Schmid (1989) carried out a numerical study of these two rules which showed that, judged by the absolute error, rule 1 guarantees increased accuracy at the cost of a larger minimum sample size. Their study also showed that for both rules the value of p strongly influences the error. For fixed n the maximum absolute error is minimized when $p = q = \frac{1}{2}$; it is reasonable to expect this since the normal distribution is symmetrical whereas the binomial distribution is symmetrical only when $p = \frac{1}{2}$. The maximum value that the absolute error can take (over all values of n and p) is $0.140(npq)^{-1/2}$; Schader and Schmid showed that under rule 1 it decreases from $0.0212(npq)^{-1/2}$ to $0.0007(npq)^{-1/2}$ as p increases from 0.01 to 0.5.

Decker and Fitzgibbon (1991) have given a table of inequalities of the form $n^c \geq k$, for different ranges of p and particular values of c and k, that yield specified degrees of error when (3.24) is employed.

Approximation (3.24) can be improved still further by replacing α and β on the right-hand side of (3.23) by

$$\frac{[\alpha\sqrt{npq} + np] - 0.5 - np}{\sqrt{npq}} \quad \text{and} \quad \frac{[\alpha\sqrt{npq} + np] + 0.5 - np}{\sqrt{npq}},$$

respectively, where $[\cdot]$ denotes the integer part. A very similar approximation was given by Laplace (1820).

For *individual* binomial probabilities, the normal approximation with continuity correction gives

$$\Pr[X = x] \approx (2\pi)^{-1/2} \int_{(x-0.5-np)/\sqrt{(npq)}}^{(x+0.5-np)/\sqrt{(npq)}} e^{-u^2/2}\, du. \qquad (3.25)$$

A nearly equivalent approximation is

$$\Pr[X = x] \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x - np)^2}{npq}\right]; \qquad (3.26)$$

see Prohorov (1953) concerning its accuracy.

Peizer and Pratt (1968) and Pratt (1968) developed a normal approximation formula for $\sum_{j=0}^{x}\binom{n}{j}p^j q^{n-j}$ in which the argument of $\Phi(\cdot)$ is

$$\frac{x+\frac{2}{3}-\left(n+\frac{1}{3}\right)p}{\left[\left(n+\frac{1}{6}\right)pq\right]^{1/2}}$$
$$\times\frac{1}{\delta_x}\left\{2\left[\left(x+\frac{1}{2}\right)\ln\left(\frac{x+\frac{1}{2}}{np}\right)+\left(n-x-\frac{1}{2}\right)\ln\left(\frac{n-x-\frac{1}{2}}{nq}\right)\right]\right\}^{1/2} \tag{3.27}$$

where $\delta_x=\left(x+\frac{1}{2}-np\right)/\sqrt{npq}$. This gives good results that are even better when the multiplier $x+\frac{2}{3}-\left(n+\frac{1}{3}\right)p$ is increased by

$$\tfrac{1}{50}\left[(x+1)^{-1}q-(n-x)^{-1}p+(n+1)^{-1}\left(q-\tfrac{1}{2}\right)\right].$$

With this adjustment, the error is less than 0.1% for $\min(x+1,n-x)\geq 2$.

Cressie (1978) suggested a slightly simpler formula, but it is not as accurate as Peizer and Pratt's improved formula, and the gain in simplicity is slight. Samiuddin and Mallick (1970) used the argument

$$\frac{\left(n-x-\frac{1}{2}\right)\left(x+\frac{1}{2}\right)}{n}\left[\ln\left(\frac{x+\frac{1}{2}}{np}\right)-\ln\left(\frac{n-x-\frac{1}{2}}{nq}\right)\right],$$

which has some points of similarity with Peizer and Pratt's formula. This approximation is considerably simpler but not as accurate.

Borges (1970) found that

$$Y=(pq)^{-1/6}\left(n+\tfrac{1}{3}\right)^{1/2}\int_p^y[t(1-t)]^{-1/3}\,dt, \tag{3.28}$$

where $y=\left(x+\frac{1}{6}\right)/\left(n+\frac{1}{3}\right)$, is approximately unit normally distributed; tables for the necessary beta integral are in Gebhardt (1971). This was compared numerically with other approximations by Gebhardt (1969). Another normal approximation is that of Ghosh (1980).

C. D. Kemp (1986) obtained an approximation for the *modal probability* based on Stirling's expansion (Section 1.1.2) for the factorials in the pmf,

$$\Pr[X=m]\approx\frac{e}{\sqrt{2\pi}}\left(\frac{a}{bc}\right)^{1/2}$$
$$\exp\left[n\ln\left(\frac{a(1-p)}{c}\right)+m\ln\left(\frac{cp}{b(1-p)}\right)+\frac{1}{12}\left(\frac{1}{a}-\frac{1}{b}-\frac{1}{c}\right)\right.$$
$$\left.-\frac{1}{360}\left(\frac{1}{a^3}-\frac{1}{b^3}-\frac{1}{c^3}\right)+\frac{1}{1260}\left(\frac{1}{a^5}-\frac{1}{b^5}-\frac{1}{c^5}\right)\right], \tag{3.29}$$

where m is the mode, $a=n+1$, $b=m+1$, and $c=n-m+1$. He reported that it gives at least eight-figure accuracy.

Littlewood (1969) made an exhaustive analysis of binomial sums. He obtained complicated asymptotic formulas for $\ln[\sum_{j=x}^{n}\binom{n}{j}p^j q^{n-j}]$ with uniform bounds of order $O(n^{-3/2})$ for each of the ranges

$$n\left(p+\frac{q}{24}\right) \le x \le n(1-n^{-1/5}p) \qquad \text{and} \qquad n(1-n^{-1/5}) \le x \le n$$

and also for

$$np \le x \le n\left(1-\tfrac{1}{2}q\right).$$

A number of approximations to binomial probabilities are based on the equation

$$\begin{aligned}\Pr[X \ge x] &= \sum_{j=x}^{n}\binom{n}{j}p^j q^{n-j} \\ &= B(x, n-x+1)^{-1}\int_0^p t^{x-1}(1-t)^{n-x}\,dt \\ &= I_p(x, n-x+1) \end{aligned} \qquad (3.30)$$

(this formula can be established by integration by parts). Approximation methods can be applied either to the integral or to the incomplete beta function ratio $I_p(x, n-x+1)$.

Bizley (1951) and Jowett (1963) pointed out that since there is an exact correspondence between sums of binomial probabilities and probability integrals for certain central F distributions (see Sections 3.4 and 3.8.3), approximations developed for the one distribution are applicable to the other, provided that the values of the parameters correspond appropriately.

The *Camp–Paulson* approximation (Johnson et al., 1995, Chapter 26) was developed with reference to the F distribution. When applied to the binomial distribution, it gives

$$\Pr[X \le x] \approx (2\pi)^{-1/2}\int_{-\infty}^{Y(3\sqrt{Z})^{-1}} e^{-u^2/2}\,du, \qquad (3.31)$$

where

$$\begin{aligned} Y &= \left[\frac{(n-x)p}{(x+1)q}\right]^{1/3}\left(9-\frac{1}{n-x}\right)-9+\frac{1}{(x+1)}, \\ Z &= \left[\frac{(n-x)p}{(x+1)q}\right]^{1/3}\left(\frac{1}{n-x}\right)+\frac{1}{(x+1)}. \end{aligned}$$

The maximum absolute error in this approximation cannot exceed $0.007(npq)^{-1/2}$.

A natural modification of the normal approximation that takes into account asymmetry is to use a Gram–Charlier expansion with one term in addition to the leading (normal) term. The maximum error is now $0.056(npq)^{-1/2}$. It varies with n and p in much the same way as for the normal approximations but is usually substantially smaller (about 50%). The relative advantage, however, depends on x as well as on n and p. For details, see Raff (1956).

If $n \to \infty$ and $p \to 0$ in such a way that $np = \theta$ remains finite and constant, then $\Pr[X = x] \to e^{-\theta}\theta^x/x!$; that is, the limiting form is the Poisson distribution (see Chapter 4). This is the basis for the *Poisson approximation* to the binomial distribution

$$\Pr[X \leq x] \approx e^{-np} \sum_{j=0}^{x} \frac{(np)^j}{j!}, \tag{3.32}$$

which has been used widely in inspection sampling. The maximum error is practically independent of n and approaches zero as p approaches zero. Anderson and Samuels (1967) showed that this gives an underestimate if $x \geq np$ and an overestimate if $x \leq np/(1 + n^{-1})$. Thus the Poisson approximation tends to overestimate tail probabilities at both tails of the distribution. The absolute error of approximation increases with x for $0 \leq x \leq (np + 0.5) - \sqrt{(np + 0.25)}$ and decreases with x for $(np + 0.5) + \sqrt{(np + 0.25)} \leq x \leq n$.

Rules of thumb for the use of (3.32) that have been recommended by various authors are summarized in Decker and Fitzgibbon (1991), together with their own findings regarding levels of accuracy. For practical work they advise the use of the normal approximation (3.24) when $n^{0.31}p \geq 0.47$; for $n^{0.31}p < 0.47$ they advise using the Poisson approximation (3.32).

Simons and Johnson (1971) were able to use a result due to Vervaat (1969) to show that, if $n \to \infty$ and $p \to 0$ with $np = \theta$, then

$$\sum_{j=0}^{\infty} \left| \binom{n}{j} p^j q^{n-j} - \frac{e^{-\theta}\theta^j}{j!} \right| h(j) \to 0 \tag{3.33}$$

for any $h(j)$ for which $\sum_{j=0}^{\infty} \theta^j (j!)^{-1} h(j)$ converges.

Ivchenko (1974) studied the ratio

$$\frac{\sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j}}{\sum_{j=0}^{x} [e^{-np}(np)^j/j!]}.$$

Hald (1967) and Steck (1973) constructed Poisson approximations to cumulative binomial probabilities by seeking solutions in θ to the equation

$$\sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j} = \sum_{j=0}^{x} \frac{e^{-\theta}\theta^j}{j!}.$$

Steck gave bounds for θ, while Hald obtained the approximation

$$\theta \approx \frac{(n - x/2)p}{1 - p/2}.$$

The accuracies of a number of Poisson approximations to the binomial distribution have been studied by Morice and Thionet (1969) and by Gebhardt (1969). Gebhardt used as an index of accuracy the maximum absolute difference between the approximate and the exact cdf's. Romanowska (1978) has made similar comparisons using the sum of the absolute differences between approximate and exact values.

The *Poisson Gram–Charlier approximation* for the cumulative distribution function is

$$\Pr[X \leq x] \approx \sum_{j=0}^{x} [P(j, np) + 0.5(j - np)\, \Delta P(j, np)], \tag{3.34}$$

where $P(j, np) = e^{-np}(np)^j/j!$ and the forward-difference operator Δ operates on j.

Kolmogorov's approximation is

$$\Pr[X \leq x] \approx \sum_{j=0}^{x} [P(j, np) - 0.5np^2\, \nabla^2 P(j, np)]. \tag{3.35}$$

The next term is $np^3 \nabla^3 P(j, np)/3$. It is a form of Gram–Charlier type B expansion. A detailed comparison of these two approximations is given in Dunin-Barkovsky and Smirnov (1955).

Galambos (1973) gave an interesting generalized Poisson approximation theorem. Let $S_x(n)$ denote the sum of the $\binom{n}{x}$ probabilities associated with different sets of x among n events $(E_1, \ldots, E_n)$. Then the conditions $S_1(n) \to a$ and $S_2(n) \to a^2/2$ as $n \to \infty$ are sufficient to ensure that the limiting distribution of the number of events that have occurred is Poissonian with parameter a.

Molenaar (1970a) provided a systematic review of the whole field of approximations among binomial, Poisson, and hypergeometric distributions. This is an important source of detailed information on relative accuracies of various kinds of approximation. For "quick work" his advice is to use

$$\Pr[X \leq x] \approx \begin{cases} \Phi(\{4x + 3\}^{1/2} q^{1/2} - \{4n - 4x - 1\}^{1/2} p^{1/2}) & \text{for } 0.05 < p < 0.93, \\ \Phi(\{2x + 1\}^{1/2} q^{1/2} - 2\{n - x\}^{1/2} p^{1/2}) & \text{for } p \geq 0.93, \end{cases}$$

or

$$\Pr[X \leq x] \approx \sum_{j=0}^{x} \frac{e^{-\lambda} \lambda^j}{j!} \tag{3.36}$$

with $\lambda = (2n - x)p/(2 - p)$ for p "small" (Molenaar suggests $p \leq 0.4$ for $n = 3$, $p \leq 0.3$ for $n = 30$, and $p \leq 0.2$ for $n = 300$).

Wetherill and Köllerström (1979) derived further interesting and useful inequalities among binomial, Poisson, and hypergeometric probabilities, with special reference to their use in the construction of acceptance sampling schemes.

3.6.2 Bounds

In the previous section we gave approximations to various binomial probabilities; in this section we examine bounds. Generally approximations are closer to the true values than bounds. Nevertheless, bounds provide one-sided approximations, and they often give useful limits to the magnitude of an approximation error.

Feller (1945) showed that, if $x \geq (n+1)p$, then

$$\Pr[X = x] \leq \Pr[X = m] \exp\left[-\frac{p[x-(n+1)p+1/2]^2}{2(n+1)pq} + \left[m - (n+1)p + \tfrac{1}{2}\right]^2 \right], \tag{3.37}$$

where m is the integer defined by $(n+1)p - 1 < m \leq (n+1)p$ and

$$\binom{n}{m}\left(\frac{m+1}{n+1}\right)^m \left(1 - \frac{m+1}{n+1}\right)^{n-m} \leq \binom{n}{m} p^m q^{n-m} \leq \binom{n}{m}\left(\frac{m}{n}\right)^m \left(1 - \frac{m}{n}\right)^{n-m}. \tag{3.38}$$

A number of formulas give bounds on the probability $\Pr[|X/n - p| \geq c]$, where c is some constant, that is, on the probability that the difference between the relative frequency X/n and its expected value p will have an absolute value greater than c; see Uspensky (1937), Lévy (1954), and Okamoto (1958). Kambo and Kotz (1966) and Krafft (1969) discussed these, sharpened Okamoto's bounds, and obtained the following improvement on Lévy's bound:

$$\Pr[|X/n - p| \geq c] < \sqrt{2}(c\sqrt{n})^{-1} \exp\left(-2nc^2 - \tfrac{4}{3}nc^4\right) \tag{3.39}$$

if $p, q \geq \max(4/n, 2c)$ and $n > 2$.

Both upper and lower bounds for $\Pr[X \geq x]$ were obtained by Bahadur (1960). Starting from the hypergeometric series representation

$$\Pr[X \geq x] = \binom{n}{x} p^x q^{n-x} \times q\, {}_2F_1[n+1, 1; x+1; p],$$

he obtained

$$\frac{q(x+1)}{x+1-(n+1)p}\left(1 + \frac{npq}{(x-np)^2}\right)^{-1} \leq \frac{\Pr[X \geq x]}{\binom{n}{x} p^x q^{n-x}} \leq \frac{q(x+1)}{x+1-(n+1)p}. \tag{3.40}$$

Slud (1977) developed further inequalities starting with the inequalities

$$\sum_{j=x+1}^{\infty} \frac{e^{-np}(np)^j}{j!} \le \Pr[X \ge x+1] \qquad \text{for } x \le \frac{n^2 p}{n+1}, \tag{3.41}$$

$$\sum_{j=x}^{\infty} \frac{e^{-np}(np)^j}{j!} \ge \max\left[\Pr[X \ge x],\ 1 - \Phi\left(\frac{x-np}{\sqrt{(npq)}}\right)\right] \qquad \text{for } x \ge (np+1), \tag{3.42}$$

and

$$\Pr[X \ge x] \ge \sum_{j=x}^{\infty} \frac{e^{-np}(np)^j}{j!} \ge 1 - \Phi\left(\frac{x-np}{\sqrt{(npq)}}\right) \qquad \text{for } x \le np. \tag{3.43}$$

The second inequality in (3.43) is valid for *all* x.

Prohorov (1953) quoted the following upper bound on the total error for the Poisson approximation to the binomial:

$$\sum_{j=0}^{\infty} \left| \binom{n}{j} p^j q^{n-j} - \frac{e^{-np}(np)^j}{j!} \right| \le \min\{2np^2, 3p\} \tag{3.44}$$

[see Sheu (1984) for a relatively simple proof]. Guzman's (1985) numerical studies suggest that in practice this bound is rather conservative.

We note the following inequalities for the ratio of a binomial to a Poisson probability when the two distributions have the same expected value:

$$e^{np}\left(1 - \frac{x}{n}\right)^x (1-p)^n \le \frac{\binom{n}{x} p^x q^{n-x}}{e^{-np}(np)^x/x!} \le e^{np}(1-p)^{n-x}. \tag{3.45}$$

Neuman's (1966) inequality is

$$\Pr[X \le np] > \frac{1}{2} + \frac{1+q}{3\sqrt{2\pi}}(npq)^{-1/2} - \frac{3q^2+12q+5}{48}(npq)^{-1} - \frac{1+q}{36\sqrt{2\pi}}(npq)^{-3/2}. \tag{3.46}$$

3.6.3 Transformations

Methods of transforming data to satisfy the requirements of the normal linear model generally seek to stabilize the variance, or to normalize the errors, or to remove interactions in order to make effects additive. Transformations are often used in the hope that they will at least partially fulfill more than one objective.

A widely used variance stabilization transformation for the binomial distribution is

$$u(X/n) = \arcsin\sqrt{\frac{X}{n}}. \tag{3.47}$$

Anscombe (1948) showed that replacing X/n by $\left(X + \frac{3}{8}\right) / \left(n + \frac{3}{4}\right)$ gives better variance stabilization; moreover, it produces a rv that is approximately normally distributed with expected value $\arcsin(\sqrt{p})$ and variance $1/(4n)$. Freeman and Tukey (1950) suggested the transformation

$$u\left(\frac{X}{n}\right) = \arcsin\sqrt{\frac{X}{n+1}} + \arcsin\sqrt{\frac{X+1}{n+1}}; \tag{3.48}$$

this leads to the same approximately normal distribution. Tables for applying this transformation were provided by Mosteller and Youtz (1961). For p close to 0.5, Bartlett (1947) suggested the transformation

$$u\left(\frac{X}{n}\right) = \ln\left(\frac{X}{n-X}\right). \tag{3.49}$$

3.7 COMPUTATION, TABLES, AND COMPUTER GENERATION

3.7.1 Computation and Tables

Recursive computation of binomial probabilities is straightforward. Since

$$\Pr[X = n] \gtrless \Pr[X = 0] \quad \text{according as} \quad p \gtrless 0.5,$$

forward recursion from $\Pr[X = 0]$ using

$$\Pr[X = x + 1] = \frac{(n - x)p}{(x + 1)q} \Pr[X = x]$$

is generally recommended for $p \leq 0.5$, and backward recursion from $\Pr[X = n]$ using

$$\Pr[X = x - 1] = \frac{xq}{(n - x + 1)p} \Pr[X = x]$$

for $p > 0.5$. Partial summation of the probabilities then gives the tail probabilities. When all the individual probabilities are required, computation with low overall rounding errors will result when an assumed value is taken for $\Pr[X = x_0]$ and both forward and backward recursion from x_0 are used; the resultant values must then be divided by their sum in order to give the true probabilities. Either the integer part of $n/2$ or an integer close to the mode of the distribution would be a sensible choice for x_0.

If only some of the probabilities are required, then recursion from the mode can be achieved using C. D. Kemp's (1986) very accurate approximation for the modal probability that was given in the previous section. This is the basis of his method for the computer generation of binomial rv's.

There are a number of tables giving values of individual probabilities and sums of these probabilities. Tables of the incomplete beta function ratio (Pearson, 1934) contain values to eight decimal places of

$$\Pr[X \geq k] = I_p(k, n-k+1)$$

for $p = 0.01(0.01)0.99$ and $\max(k, n-k+1) \leq 50$. Other tables are as follows:

Biometrika Tables for Statisticians (Pearson and Hartley, 1976)

Tables of the Binomial Probability Distribution (National Bureau of Standards, 1950)

Binomial Tables [Romig, 1953 (this supplements the tables of the National Bureau of Standards)].

Details concerning these and some other tables of binomial probabilities were given in the first edition of this book.

A method for computing $\Pr[X \geq x]$ (the binomial survival function) was devised by Bowerman and Scheuer (1990). It was designed to avoid underflow and overflow problems and is especially suitable for large n.

Stuart (1963) gave tables from which values of $[pq(n_1^{-1} + n_2^{-1})]^{1/2}$ can be obtained to four decimal places (this is the standard deviation of the difference between two independent binomial proportions with common parameter p).

Nomographs for calculating sums of binomial probabilities have been developed [see, e.g., Larson (1966)]. Such nomographs have also been constructed and labeled for the equivalent problem of calculating values of the incomplete beta function ratio (Hartley and Fitch, 1951).

3.7.2 Computer Generation

If large numbers of rv's are required from a binomial distribution with constant parameters, then the ease of computation of its probabilities coupled with the bounded support for the distribution makes nonspecific methods very attractive.

However, when successive calls to the generator are for random binomial variates with changing parameters, distribution-specific methods become important. A slow but very simple method is to simulate the flip of a biased coin n times and count the number of successes. When $p = 0.5$, it suffices to count the number of 1's in a random uniformly distributed computer word of n bits. For $p \neq 0.5$, the method requires n uniforms per generated binomial variate, making it very slow (recycling uniform random numbers in order to reduce the number required is not generally recommended). An ingenious improvement (the beta, or median, method) was devised by Relles (1972); see also Ahrens and Dieter (1974).

Devroye (1986) gives two interesting waiting-time methods based on the following features of the binomial distribution: First, let $G_1, G_2, \ldots$ be iid geometric rv's with parameter p, and let X be the smallest integer such that $\sum_{i=1}^{X+1} G_i > n$; then X is binomial with parameters n, p. Second, let $E_1, E_2, \ldots$ be iid exponential rv's, and let X be the smallest integer such that

$$\sum_{i=1}^{X+1} \frac{e_i}{n-i+1} > -\ln(1-p);$$

then X is binomial with parameters n, p. Both methods can be decidedly slow because of their requirement for very many uniform random numbers; however, as for the coin-flip method, their computer programs are very short.

The C. D. Kemp (1986) algorithm, based on inversion of the cdf by unstored search from the mode, competes favorably with the Ahrens and Dieter (1974) algorithm. Other unstored search programs are discussed in Kemp's paper.

Acceptance–rejection using a Poisson envelope was proposed by Fishman (1979). Kachitvichyanukul and Schmeiser's (1988) algorithm BTPE is a very fast, intricate composition–acceptance–rejection algorithm. Stadlober's (1991) algorithm is simpler, but not quite so fast; it uses the ratio of two uniforms.

3.8 ESTIMATION

3.8.1 Model Selection

The use of binomial probability paper in exploratory data analysis is described by Hoaglin and Tukey (1985). Binomially distributed data should produce a straight line with slope and intercept that can be interpreted in terms of estimates of the parameters.

Other graphical methods include the following:

1. a plot of the ratio of sample factorial cumulants $\tilde{\kappa}_{[r+1]}/\tilde{\kappa}_{[r]}$ against successive low values of r (Hinz and Gurland, 1967; see also Douglas, 1980);
2. a plot of xf_x/f_{x-1} against successive low values of x, where f_x is the observed frequency of x (Ord, 1967a), see Table 2.2; and
3. marking the position of $(\tilde{\kappa}_3/\tilde{\kappa}_2, \tilde{\kappa}_2/\tilde{\kappa}_1)$ on Ord's (1970) diagram of distributions, where $\tilde{\kappa}_1$, $\tilde{\kappa}_2$, and $\tilde{\kappa}_3$ are the first three sample cumulants.

Further graphical methods have been developed by Gart (1970) and Grimm (1970).

3.8.2 Point Estimation

Usually n is known. The method of moments, maximum likelihood, and minimum χ^2 estimators of p are then all equal to $\overline{x}/n$. This estimator is unbiased. Given k samples of size n, its variance is pq/nk, which is the Cramer–Rao lower bound for unbiased estimators of p; the estimator is in fact the minimum variance unbiased estimator (MVUE) of p. Its expected absolute error has been investigated by Blyth (1980).

An approximately median-unbiased estimator of p is

$$\frac{\overline{x}+\frac{1}{6}}{n+\frac{1}{3}}$$

(Crow, 1975; see also Birnbaum, 1964). A helpful expository account of estimation for p (including Bayesian estimation) is in Chew (1971). A useful summary of results is in Patel, Kapardia, and Owen (1976).

Estimation of certain functions of p (when n is known) has also been investigated. Sometimes an estimate of $\Pr[\alpha < X < \beta]$ is required. The MVUE of this polynomial function of p is

$$\sum_{\alpha<\xi<\beta} \binom{n}{\xi}\binom{n(N-1)}{T-\xi}\Bigg/\binom{nN}{T}$$

where N is the number of observations, $T=\sum_{j=1}^{N} x_j$, and ξ takes integer values. From this expression it can be seen that the MVUE of a probability $\Pr[X \in \omega]$, where ω is *any* subset of the integers $\{0, 1, \ldots, n\}$, has the same form with the range of summation $\alpha < \xi < \beta$ replaced by $\xi \in \omega$. Rutemiller (1967) studied the estimator of $\Pr[X = 0]$ in some detail, giving tables of its bias and variance. Pulskamp (1990) showed that the MVUE of $\Pr[X = x]$ is admissible under quadratic loss when $x = 0$ or n, but is inadmissible otherwise. He conjectured that the maximum likelihood estimator (MLE) is always admissible.

Another function of p for which estimators have been constructed is $\min(p, 1-p)$. A natural estimator to use, given a single observed value x, is $\min(x/n, 1-x/n)$. The moments of this statistic have been studied by Greenwood and Glasgow (1950) and the cumulative distribution by Sandelius (1952).

Cook, Kerridge, and Pryce (1974) have shown that, given a single observation x, $\psi(x)-\psi(n)$ is a useful estimator of $\ln(p)$; they also showed that

$$\left|E[\psi(x)-\psi(n)]-\ln(p)\right| < \frac{q^{n+1}p}{n+1},$$

where $\psi(y)$ is the derivative of $\ln\Gamma(y)$ (i.e., the psi function of Section 1.1.2). They also obtained an "almost unbiased" estimator of the entropy $p\ln(p)$ and used the estimator to construct an estimator of the entropy for a multinomial distribution.

Unbiased sequential estimation of $1/p$ has been studied by Gupta (1967), Sinha and Sinha (1975), and Sinha and Bose (1985). DeRouen and Mitchell (1974) constructed minimax estimators for linear functions of $p_1, p_2, \ldots, p_r$ corresponding to r different (independent) binomial variables.

Suppose now that $X_1, X_2, \ldots, X_k$ are independent binomial rv's and that X_j has parameters n_j, p, where $j = 1, 2, \ldots, k$. Then, given a sample of k

observations $x_1, \ldots, x_k$, comprising one from each of the k distributions, the maximum–likelihood estimator of p is the overall relative frequency

$$\hat{p} = \frac{\sum_{j=1}^{k} x_j}{\sum_{j=1}^{k} n_j}. \tag{3.50}$$

Moreover $\sum_{j=1}^{k} x_j$ is a sufficient statistic for p. Indeed, since $\sum_{j=1}^{k} X_j$ has a binomial distribution with parameters $\sum_{j=1}^{k} n_j$, p, the analysis is the same as for a single binomial distribution.

The above discussion assumes that $n_1, n_2, \ldots, n_k$ (or at least $\sum_{j=1}^{k} n_j$) are known. The problem of estimating the values of the n_j's was studied by Student (1919), Fisher (1941), Hoel (1947), and Binet (1953); for further historical details see Olkin, Petkau, and Zidek (1981).

Given a single observation of a rv X having a binomial distribution with parameters n, p, then, if p is known, a natural estimator for n is x/p. This is unbiased and has variance nq/p.

The equation for the MLE $\hat{n}$ of n when p is known is

$$\sum_{j=0}^{R-1} A_j(\hat{n} - j)^{-1} = -N \ln(1 - p); \tag{3.51}$$

A_j is the number of observations that exceed j and $R = \max(x_1, \ldots, x_N)$ (Haldane, 1941). When N is large,

$$\sqrt{N}\ \text{Var}(\hat{n}) \approx \left[\sum_{j=1}^{n} \left(\Pr[X = j] \sum_{i=0}^{j-1} (n - i)^{-2} \right) \right]^{-1}. \tag{3.52}$$

The consistency of this estimator was studied by Feldman and Fox (1968).

Dahiya (1981) has constructed a simple graphical method for obtaining the maximum-likelihood estimate of n, incorporating the integer restriction on n. In Dahiya (1986) he examined the estimation of m (an integer) when $p = \theta^m$ and θ is a known constant.

Suppose now that $X_1, X_2, \ldots, X_k$ are independent rv's all having the same binomial distribution with parameters n, p. Then equating the observed and expected first and second moments gives the moment estimators $\tilde{n}$ and $\tilde{p}$ of n and p as the solutions of $\overline{x} = \tilde{n}\tilde{p}$ and $s^2 = \tilde{n}\tilde{p}\tilde{q}$. Hence

$$\tilde{p} = 1 - \frac{s^2}{\overline{x}} \tag{3.53}$$

$$\tilde{n} = \frac{\overline{x}}{\tilde{p}}. \tag{3.54}$$

Note that, if $\overline{x} < s^2$, then $\tilde{n}$ is negative, suggesting that a binomial distribution is an inappropriate model.

Continuing to ignore the limitation that n must be an integer, the MLEs $\hat{n}$, $\hat{p}$ of n, and p satisfy the equations

$$\hat{n}\hat{p} = \overline{x}, \tag{3.55}$$

$$\sum_{j=0}^{R-1} A_j(\hat{n} - j)^{-1} = -N \ln\left(1 - \frac{\overline{x}}{\hat{n}}\right); \tag{3.56}$$

A_j is the number of observations that exceed j and $R = \max(x_1, \ldots, X_N)$. The similarity between (3.54) and (3.55) arises because the binomial distribution is a PSD; see Section 2.2. Unlike the method-of-moments equations, the maximum-likelihood equations require iteration for their solution. DeRiggi (1983) has proved that a maximum-likelihood solution exists iff the sample variance is less than $\overline{x}$ and that, if a solution exists, it is unique.

If N is large,

$$\text{Var}(\hat{n}) \approx \frac{n}{N}\left[\sum_{j=2}^{n}\left(\frac{p}{q}\right)^j \frac{(j-1)!(N-j)!}{j(N-1)!}\right]^{-1}, \tag{3.57}$$

and the asymptotic efficiency of $\tilde{n}$, relative to $\hat{n}$, is

$$\left[1 + 2\sum_{j=1}^{n-1}\left(\frac{p}{q}\right)^i \frac{j!(N-j-1)!}{(j+1)(N-2)!}\right]^{-1} \tag{3.58}$$

(Fisher, 1941).

Olkin, Petkau, and Zidek (1981) found theoretically and through a Monte Carlo study that, when both n and p are to be estimated, the method of moments and maximum-likelihood estimation both give rise to estimators that can be highly unstable; they suggested more stable alternatives based on (1) ridge stabilization and (2) jackknife stabilization. Blumenthal and Dahiya (1981) also recognized the instability of the MLE of n, both when p is unknown and when p is known; they too gave an alternative stabilized version of maximum-likelihood estimation.

In Carroll and Lombard's (1985) study of the estimation of population sizes for impala and waterbuck, these authors stabilized maximum-likelihood estimation of n by integrating out the nuisance parameter p using a beta distribution with parameters a and b. The need for a stabilized estimator of n has been discussed by Casella (1986).

The idea of minimizing the likelihood as a function of n, with p integrated out, can be interpreted in a Bayesian context. For a helpful description of the principles underlying Bayesian estimation for certain discrete distributions, including the binomial, see Irony (1992). Geisser (1984) has discussed and contrasted ways of choosing a prior distribution for binomial trials.

An early Bayesian treatment of the problem of estimating n is that of Draper and Guttman (1971). For p known they chose as a suitable prior distribution for n the rectangular distribution with pmf $1/k$, $1 \le n \le k$, k some large preselected integer. For p unknown, they again used a rectangular prior for n and, like Carroll and Lombard, adopted a beta prior for p, thus obtaining a marginal distribution for n of the form

$$p(n|x_1, \ldots, x_N) \propto \frac{(Nn - T + b - 1)!}{(Nn + a + b - 1)!} \prod_{j=1}^{N} \frac{n!}{(n - x_j)!},$$

where $\max(x_i) \le n \le k$, $T = \sum_{j=1}^{N} x_i$. Although this does not lead to tractable analytical results, numerical results are straightforward to obtain. Kahn (1987) has considered the tailweight of the marginal distribution for n after integrating out p when n is large; he has shown that the tailweight is determined solely by the prior density on n and p. This led him to recommend caution when adopting specific prior distributions. Hamedani and Walter (1990) have reviewed both Bayesian and non-Bayesian approaches to the estimation of n.

Empirical Bayes methods have been created by Walter and Hamedani (1987) for unknown p and by Hamedani and Walter (1990) for unknown n. In their 1990 paper they used an inversion formula and Poisson–Charlier polynomials to estimate the prior distribution of n; this can then be smoothed if it is thought necessary. Their methods are analogous to the Bayes–empirical Bayes approach of Deely and Lindley (1981). Barry (1990) has developed empirical Bayes methods, with smoothing, for the simultaneous estimation of the parameters p_i for many binomials in both one-way and two-way layouts.

Serbinowska (1996) has considered estimation of the number of changes in the parameter p in a stream of binomial observations.

Research concerning the estimation of the parameter n when p is known has been reviewed and extended by Zou and Wan (2003). Casella and Strawderman (1994) have investigated the simultaneous estimation of $n_1, n_2, \ldots$ for several binomial samples.

Kyriakoussis and Papadopoulos (1993) deal with the Bayesian estimation of p for the zero-truncated binomial distribution.

3.8.3 Confidence Intervals

The binomial distribution is a discrete distribution, and so it is not generally possible to construct a confidence interval for p with an exactly specified confidence coefficient using only a set of observations.

Let $x_1, x_2, \ldots, x_N$ be values of independent random binomial variables with exponent parameters $n_1, n_2, \ldots, n_N$ and common second parameter p. Then approximate $100(1 - \alpha)\%$ limits may be obtained by solving the following equations for p_L and p_U:

$$\sum_{j=T}^{n} \binom{n}{j} (p_L)^j (1 - p_L)^{n-j} = \frac{\alpha}{2}, \qquad (3.59)$$