CHAPTER 5

Negative Binomial Distribution

5.1 DEFINITION

Many different models give rise to the negative binomial distribution, and consequently there is a variety of definitions in the literature. The two main dichotomies are (a) between parameterizations and (b) between points of support.

Formally, the negative binomial distribution can be defined in terms of the expansion of the negative binomial expression $(Q - P)^{-k}$, where $Q = 1 + P$, $P > 0$, and k is positive real; the $(x + 1)$th term in the expansion yields $\Pr[X = x]$. This is analogous to the definition of the binomial distribution in terms of the binomial expression $(\pi + \omega)^n$, where $\omega = 1 - \pi$, $0 < \pi < 1$, and n is a positive integer.

Thus the *negative binomial distribution* with parameters k, P is the distribution of the rv X for which

$$\Pr[X = x] = \binom{k + x - 1}{k - 1} \left(\frac{P}{Q}\right)^x \left(1 - \frac{P}{Q}\right)^k, \qquad x = 0, 1, 2, \ldots, \tag{5.1}$$

where $Q = 1 + P$, $P > 0$, and $k > 0$. Unlike the binomial distribution, here there is a nonzero probability for X to take any specified nonnegative integer value, as in the case of the Poisson distribution.

The probability generating function (pgf) is

$$G(z) = (1 + P - Pz)^{-k} \tag{5.2}$$

$$= {}_1F_0[k; \; ; P(z - 1)] \tag{5.3}$$

$$= \frac{{}_1F_0[k; \; ; Pz/(1 + P)]}{{}_1F_0[k; \; ; P/(1 + P)]} \tag{5.4}$$

and the characteristic function is $(1 + P - Pe^{it})^{-k}$. The mean and variance are

$$\mu = kP \qquad \text{and} \qquad \mu_2 = kP(1 + P). \tag{5.5}$$

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

This parameterization (but with the symbol p instead of P) is the one introduced by Fisher (1941).

Other early writers adopted different parameterizations. Jeffreys (1941) had $b = P/(1+P)$, $\rho = kP$, giving the pgf $[(1-bz)/(1-b)]^{\rho-\rho/b}$, and $\mu = \rho$, $\mu_2 = \rho/(1-b)$. Anscombe (1950) used the form $\alpha = k$, $\lambda = kP$, giving the pgf $(1+\lambda/\alpha - \lambda z/\alpha)^{-\alpha}$, and $\mu = \lambda$, $\mu_2 = \lambda(1+\lambda/\alpha)$.

Evans (1953) took $a = P$, $m = kP$, giving the pgf $(1+a-az)^{-m/a}$, and $\mu = m$, $\mu_2 = m(1+a)$. This parameterization has been popular in the ecological literature. Some writers, for instance Patil et al. (1984), called this the Pólya–Eggenberger distribution, as it arises as a limiting form of Eggenberger and Pólya's (1923) urn model distribution. Other authors, notably Johnson and Kotz (1977) and Berg (1988b), called the (nonlimiting) urn model distribution the Pólya–Eggenberger distribution; see Section 6.2.4. For both distributions "Pólya–Eggenberger" is quite often abbreviated to "Pólya."

A further parameterization that has gained wide favor is $p = 1/(1+P)$, that is, $q = P/(1+P)$, and $k = k$, giving

$$G(z) = \left(\frac{1-q}{1-qz}\right)^k, \tag{5.6}$$

$$\Pr[X=x] = \binom{k+x-1}{k-1} q^x (1-q)^k, \qquad x = 0, 1, 2, \ldots, \tag{5.7}$$

and $\mu = kq/(1-q)$, $\mu_2 = kq/(1-q)^2$. Sometimes $\lambda = P/(1+P)$ is used to avoid confusion with the binomial parameter q.

Clearly k need not be an integer. When k is an integer, the distribution is sometimes called the *Pascal distribution* (Pascal, 1679). The name "Pascal distribution" is, however, more often applied to the distribution shifted k units from the origin, that is, with support $k, k+1, \ldots$; this is also called the binomial waiting-time distribution.

The geometric is the special case $k = 1$ of the negative binomial distribution.

Kemp (1967a) summarized four commonly encountered formulations of pgf's for the negative binomial and geometric distributions as follows:

Formulation	**Negative Binomial**	**Geometric**	**Conditions**
1	$p^k(1-qz)^{-k}$	$p(1-qz)^{-1}$	$p+q=1$, $0<p<1$
2	$p^k z^k (1-qz)^{-k}$	$pz(1-qz)^{-1}$	
3	$(Q-Pz)^{-k}$	$(Q-Pz)^{-1}$	$Q = 1+P$, $P>0$ ($Q = 1/p$, $P = q/p$, i.e., $p = 1/Q$, $q = P/Q$)
4	$z^k(Q-Pz)^{-k}$	$z(Q-Pz)^{-1}$	

In cases 1 and 3, k is positive real, the support is $0, 1, 2, \ldots$, and the distribution is a power series distribution (PSD). For cases 2 and 4, k is necessarily a positive integer; the distribution has support $k, k+1, k+2, \ldots$, and the distribution is a generalized power series distribution (GPSD) (see Section 2.1).

Case 1 shows that the distribution is a generalized hypergeometric probability distribution [with argument parameter q; cf. (5.4)], while case 3 shows that it is a generalized hypergeometric factorial moment distribution [with argument parameter P; cf. (5.3)]. Other families to which the negative binomial belongs are the exponential (provided k is fixed), Katz, Willmot, and Ord families.

5.2 GEOMETRIC DISTRIBUTION

In the special case $k = 1$ the pmf is

$$\Pr[X = j] = Q^{-1}\left(\frac{P}{Q}\right)^j = pq^j, \qquad j = 0, 1, 2, \ldots. \tag{5.8}$$

These values are in geometric progression, so this distribution is called a *geometric distribution*; sometimes it is called a *Furry distribution* (Furry, 1937).

Its properties can be obtained from those of the negative binomial as the special case $k = 1$.

The geometric distribution possesses a property similar to the "nonaging" (or "Markovian") property of the exponential distribution (Johnson et al. 1995, Chapter 18). This is

$$\Pr[X = x + j | X \geq j] = \frac{Q^{-1}(P/Q)^{x+j}}{(P/Q)^j} = Q^{-1}\left(\frac{P}{Q}\right)^x = \Pr[X = x]. \tag{5.9}$$

This property characterizes the geometric distribution (among all distributions restricted to the nonnegative integers), just as the corresponding property characterizes the exponential distribution. The distribution is commonly said to be a discrete analog of the exponential distribution. It is a special case of the grouped exponential distribution; see Spinelli (2001).

The geometric distribution may be extended to cover the case of a variable taking values $\theta_0, \theta_0 + \delta, \theta_0 + 2\delta, \ldots \quad (\delta > 0)$. Then, in place of (5.8), we have

$$\Pr[X = \theta_0 + j\delta] = Q^{-1}\left(\frac{P}{Q}\right)^j. \tag{5.10}$$

The characterization summarized in (5.9) also applies to this distribution with X replaced by $\theta_0 + X\delta$ and j replaced by $\theta_0 + j\delta$.

Other characterizations are described in Section 5.9.1.

Another special property of the geometric distribution is that, if a mixture of negative binomial distributions [as in (5.1)] is formed by supposing k to have the geometric distribution

$$\Pr[k = j] = (Q')^{-1}\left(\frac{P'}{Q'}\right)^{j-1}, \qquad j = 1, 2, \ldots, \tag{5.11}$$

then the resultant mixture distribution is also a geometric distribution of the form (5.8) with Q replaced by $QQ' - P'$.

The geometric, like the negative binomial distribution, is infinitely divisible; see Section 1.2.10 for a definition of infinite divisibility.

The Shannon entropy (first-order entropy) of the geometric distribution is $P \log_2 P - Q \log_2 Q$. The second-order entropy is $\log_2(1 + 2P)$.

Margolin and Winokur (1967) obtained formulas for the moments of the order statistics for the geometric distribution and tabulated values of the mean and variance to two decimal places; see also Kabe (1969). Steutel and Thiemann's (1989a) expressions for the order statistics were derived using the independence of the integer and fractional parts of exponentially distributed rv's. The computation of the order statistics from the geometric distribution has also been studied by Adatia (1991). Adatia also obtained an explicit formula for the expected value of the product of two such order statistics.

Order statistics from a continuous distribution form a Markov chain, but this is not in general true for discrete distributions. At first it was thought that exceptionally the Markov property holds for the geometric distribution (Gupta and Gupta, 1981); later this was disproved by Nagaraja (1982) and Arnold et al. (1984). Nagaraja's (1990) lucid survey article on order statistics from discrete distributions documents a number of characterizations of the geometric distribution based on its order statistics; see Section 5.9.1.

Estimation of the parameter of the geometric distribution is particularly straightforward. Because it is a PSD, the first-moment equation is also the maximum-likelihood equation. Hence

$$\hat{P} = \overline{x}. \tag{5.12}$$

A moment-type estimator for the geometric distribution with either or both tails truncated was obtained by Kapadia and Thomasson (1975), who compared its efficiency with that of the maximum-likelihood estimator (MLE). Estimation for the geometric distribution with unknown P and unknown location parameter was studied by Klotz (1970) (maximum-likelihood estimation), Iwase (1986) (minimum-variance unbiased estimation), and Yanagimoto (1988) (conditional maximum-likelihood estimation). Vit (1974) examined tests for homogeneity.

If $X_1, X_2, \ldots, X_k$ are rv's each with the geometric distribution with pmf

$$\Pr[X = x] = Q^{-1}\left(\frac{P}{Q}\right)^x, \qquad x = 0, 1, 2, \ldots,$$

then $\sum_{i=1}^{k} X_i$ is a negative binomial variable with parameters k and P; see Section 5.5. Using this fact, Clemans (1959) constructed charts from which confidence intervals for P, given $k^{-1}\sum_{i=1}^{k} x_i$, can be read off.

Applications of the geometric distribution include runs of one plant species with respect to another in transects through plant populations (Pielou, 1962, 1963), a ticket control problem (Jagers, 1973), a surveillance system for congenital malformations (Chen, 1978), and estimation of animal abundance (Seber, 1982b). Mann et al. (1974) looked at applications in reliability theory.

The distribution is used in Markov chain models, for example, in meteorological models of weather cycles and precipitation amounts (Gabriel and Neumann, 1962). Many other applications in queueing theory and applied stochastic models were discussed by Taylor and Karlin (1998) and Bhat (2002). Daniels (1961) investigated the representation of a discrete distribution as a mixture of geometric distributions and applied this to busy-period distributions in equilibrium queueing systems. Sandland (1974) put forward a building-society-membership scheme and a length-of-tenure scheme as models for the truncated geometric distribution with support $0, 1, \ldots, n-1$.

5.3 HISTORICAL REMARKS AND GENESIS OF NEGATIVE BINOMIAL DISTRIBUTION

Special forms of the negative binomial distribution were discussed by Pascal (1679). A derivation as the distribution of the number of tosses of a coin necessary to achieve a fixed number of heads was published by Montmort (1713) in his solution of the problem of points; see Todhunter (1865, p. 97). A very clear interpretation of the pmf as a density function was given by Galloway (1839, pp. 37–38) in his discussion of the problem of points. Let X be the rv representing the number of independent trials necessary to obtain k occurrences of an event that has a constant probability of occurring at each trial. Then

$$\Pr[X = k + j] = \binom{k+j-1}{k-1} p^k (1-p)^j, \qquad j = 1, 2, \ldots; \tag{5.13}$$

that is, X has a negative binomial distribution (case 2 in Kemp's list in the previous section).

Meyer (1879, p. 204) obtained the pmf as the probability of exactly j male births in a birth sequence containing a fixed number of female births; he assumed a known constant probability of a male birth. He also gave the cdf in a form that we now recognize as the upper tail of an F distribution (equivalent to an incomplete beta function; see Section 5.6).

Student (1907) found empirically that certain hemocytometer data could be fitted well by a negative binomial distribution. Whittaker (1914) continued this approach. Unfortunately she did not realize that the Poisson distribution is a limiting form for both the binomial and the negative binomial distributions (see Section 5.12.1), and she aroused considerable controversy concerning the relative merits of the Poisson and the negative binomial distributions.

Greenwood and Yule (1920) derived the following relationship between the Poisson and negative binomial distributions. Suppose that we have a mixture of Poisson distributions such that the expected values θ of the Poisson distributions vary according to a gamma distribution with pdf

$$f(\theta) = [\beta^\alpha \Gamma(\alpha)]^{-1} \theta^{\alpha-1} \exp\left(-\frac{\theta}{\beta}\right), \qquad \theta > 0, \qquad \alpha > 0, \qquad \beta > 0.$$

Then

$$\Pr[X = x] = [\beta^\alpha \Gamma(\alpha)]^{-1} \int_0^\infty \theta^{\alpha-1} e^{-\theta/\beta} (\theta^x e^{-\theta}/x!)\, d\theta$$
$$= \binom{\alpha + x - 1}{\alpha - 1} \left(\frac{\beta}{\beta + 1}\right)^x \left(\frac{1}{\beta + 1}\right)^\alpha. \qquad (5.14)$$

So X has a negative binomial distribution with parameters α and β. This type of model was used to represent "accident proneness" by Greenwood and Yule. The parameter θ represents the expected number of accidents for an individual. This is assumed to vary from individual to individual.

Another important derivation is that of Lüders (1934); see also Quenouille (1949). Here the negative binomial arises as the distribution of the sum of N independent random variables each having the same logarithmic distribution (Chapter 7), where N has a Poisson distribution. Thyrion (1960) called this an *Arfwedson process*. Boswell and Patil (1970) termed it a Poisson sum (Poisson–stopped sum) of logarithmic rv's. Let $Y = X_1 + X_2 + \cdots + X_N$, where the X_i are iid logarithmic rv's with pgf $\ln(1 - \theta z)/\ln(1 - \theta)$. Assume also that N is a Poisson rv (with parameter λ) which is independent of the X_i. Then the pgf of Y is

$$\exp\left[\lambda\left(\frac{\ln(1 - \theta z)}{\ln(1 - \theta)} - 1\right)\right] = \left(\frac{1 - \theta}{1 - \theta z}\right)^{-\lambda/\ln(1-\theta)}; \qquad (5.15)$$

see Section 7.1.2.

The negative binomial as a limiting form for Pólya and Eggenberger's urn model was mentioned in Section 5.1. Consider a random sample of n balls from an urn containing Np white balls and $N(1 - p)$ black balls. Suppose that after each draw the drawn ball is replaced together with $c = N\beta$ others *of the same color*. Let the number of white balls in the sample be X. Then

$$\Pr[X = x] = \binom{n}{x} \left(\frac{p}{\beta}\right)^{[x]} \left(\frac{q}{\beta}\right)^{[n-x]} \Big/ \left(\frac{1}{\beta}\right)^{[n]}, \qquad (5.16)$$

where $a^{[x]} = a(a + 1) \cdots (a + x - 1)$; see Section 6.2.4. The limiting form as $n \to \infty$, $p \to 0$, $\beta \to 0$ such that $np \to \eta k$, $n\beta \to \eta$ is negative binomial with pmf (5.1), where $P = \eta$; see Eggenberger and Pólya (1923, 1928).

This limiting form of (5.16) has been called a "Pólya" distribution by, for instance, Gnedenko (1961), Arley and Buch (1950), and Hald (1952). On the other hand, Bosch (1963) called (5.16) a "Pólya" distribution.

Patil and Joshi (1968) called the negative binomial a "Pólya–Eggenberger" and the distribution (5.16) simply a "Pólya" distribution. Proofs of the limiting form appear in Bosch (1963), Lundberg (1940), Feller (1968), and Boswell and Patil (1970). Thompson (1954) showed that a negative binomial distribution can also be obtained (approximately) from a modified form of Neyman's contagious distribution model (Section 9.6).

Distribution (5.16) arises also as a beta mixture of binomial distributions (Skellam, 1948); see Sections 6.2.2 and 8.3.4. Boswell and Patil (1970) derived the negative binomial as a limiting form of this mixture of binomials.

Feller (1957, p. 253) pointed out that the negative binomial can be regarded as a convolution of a fixed number of geometric distributions; here, as for the inverse sampling model (5.13), the exponent k is necessarily an integer. Maritz (1952) considered how the negative binomial could arise from the addition of a set of correlated Poisson rv's. Kemp (1968a) showed that weighting a negative binomial with parameters k and $P = q/p$ using the weight function (sampling chance) $a_x = \alpha^x$ gives another negative binomial but with parameters k and $\alpha q/(1 - \alpha q)$.

Bhattacharya (1966) obtained the negative binomial by mixing his confluent hypergeometric distributions with a "generalized exponential" distribution; the pgf of the outcome is

$$\int_0^\infty \frac{{}_1F_1(a; b; \theta z)}{{}_1F_1(a; b; \theta)} \times \frac{c^a (c+1)^{b-a} \theta^{b-1} e^{-(c+1)\theta}\ {}_1F_1(a; b; \theta)\, d\theta}{\Gamma(b)} = \frac{[1 - 1/(c+1)]^a}{[1 - z/(c+1)]^a}. \tag{5.17}$$

Bhattacharya showed that the generalized exponential mixing distribution is unique by virtue of the uniqueness of the Mellin transform (Section 1.1.10), and he applied his results to the theory of accident proneness in the case where $a = 1$ and the number of accidents sustained by an individual has a sub-Poisson distribution (Section 4.12.4).

The result of mixing negative binomials with constant exponent parameter k using a beta distribution with parameters c and $k - c$, where $k > c > 0$, is another negative binomial distribution with exponent parameter c; we have

$$\int_0^1 [(1 + P\theta - P\theta z)^{-k}] \frac{\theta^{c-1}(1-\theta)^{k-c-1}\, d\theta}{B(c, k-c)} = (1 + P - Pz)^{-c}. \tag{5.18}$$

A mixture of Katti (1966) type H_2 distributions (Section 6.11) using a particular beta distribution can also yield a negative binomial distribution:

$$\begin{aligned}\int_0^1 {}_2F_1[k, a; b; P\theta(z-1)] \frac{\theta^{b-1}(1-\theta)^{a-b-1}\, d\theta}{B(b, a-b)} &= {}_3F_2[k, a, b; b, a; P(z-1)] \\ &= (1 + P - Pz)^{-k}. \end{aligned}\tag{5.19}$$

Also a gamma mixture of Poisson $\bigwedge$ beta distributions (Section 8.3.3) gives rise to a negative binomial:

$$\begin{aligned}\int_0^\infty {}_1F_1[a; a+b; P\theta(z-1)]\frac{e^{-\theta}\theta^{a+b-1}\,d\theta}{\Gamma(a+b)} \\ = {}_2F_1[a, a+b; a+b; P(z-1)] \\ = (1+P-Pz)^{-a}. \qquad (5.20)\end{aligned}$$

These results are special cases of those in Section 8.3.6; see also Kemp (1968a).

The negative binomial arises also from several well-known stochastic processes. The time-homogeneous birth-and-immigration process with zero initial population was first obtained by McKendrick (1914); the equivalence of the distributions arising from this process, from Greenwood and Yule's model as a gamma mixture of Poisson distributions, and from Lüders and Quenouille's Poisson–stopped sum of logarithmic distributions model was discussed by Irwin (1941). The nonhomogeneous process with zero initial population known as the *Pólya process* was developed by Lundberg (1940) in the context of risk theory. Other stochastic processes that lead to the negative binomial include the simple birth process with nonzero initial population size (Yule, 1925; Furry, 1937), Kendall's (1948) nonhomogeneous birth-and-death process with zero death rate, and the simple birth-death-and-immigration process with zero initial population of Kendall (1949).

The geometric distribution is the equilibrium distribution of queue length for the M/M/1 queue, while the negative binomial is the equilibrium queue length distribution for the M/M/1 queue with a particular form of balking; see Haight (1957) and also Bhat (2002). The negative binomial can also be obtained as the equilibrium solution for a particular type of Markov chain known as a Foster process (Foster, 1952).

5.4 MOMENTS

From the pgf $(1+P-Pz)^{-k}$, it follows that the factorial moment generation function (fmgf) is $(1-Pt)^{-k}$, and so

$$\mu'_{[r]} = \frac{(k+r-1)!}{(k-1)!}P^r, \qquad r = 1, 2, \ldots. \qquad (5.21)$$

Also the factorial cumulant generating function (fcgf) is $-k\ln(1-Pt)$, whence

$$\kappa_{[r]} = k(r-1)!P^r, \qquad r = 1, 2, \ldots. \qquad (5.22)$$

The relationship with the binomial pgf is readily apparent—replacing N by $-k$ and π by $-P$ in the well-known formulas for the moment properties of the binomial distribution gives the corresponding formulas for the negative binomial distribution. In particular

$$\begin{aligned}
\mu_r &= kPQ\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_j + P\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_{j+1},\\
\mu\ &= \kappa_1 = kP = \frac{kq}{p},\\
\mu_2 &= \kappa_2 = kP(1+P) = \frac{kq}{p^2},\\
\mu_3 &= \kappa_3 = kP(1+P)(1+2P) = \frac{kq(1+q)}{p^3},\\
\mu_4 &= 3k^2P^2(1+P)^2 + kP(1+P)(1+6P+6P^2)\\
&= \frac{3k^2q^2}{p^4} + \frac{kq(p^2+6q)}{p^4},
\end{aligned} \tag{5.23}$$

and

$$\begin{aligned}
\sqrt{\beta_1} &= \frac{1+2P}{[kP(1+P)]^{1/2}} = \frac{1+q}{\sqrt{kq}},\\
\beta_2 &= 3 + \frac{1+6P+6P^2}{kP(1+P)} = 3 + \frac{p^2+6q}{kq},
\end{aligned} \tag{5.24}$$

where $p = 1/(1+P) = Q^{-1}$ and $q = P/(1+P) = PQ^{-1}$.

The alternative notation, with the pgf in the form $p^k(1-qz)^{-k}$, shows that the mgf is $p^k(1-qe^t)^{-k}$, the central mgf is $e^{-kqt/p}p^k(1-qe^t)^{-k}$, whence

$$\mu_{r+1} = q\frac{\partial\mu_r}{\partial q} + \frac{rkq}{p^2}\mu_{r-1},$$

and the cumulant generating function (cgf) is $k\ln p - k\ln(1-qe^t)$.

Because this is a PSD with series parameter q, the cumulants satisfy

$$\kappa_{r+1} = q\frac{\partial\kappa_r}{\partial q}, \qquad r = 1, 2, \ldots. \tag{5.25}$$

The distribution is overdispersed (variance greater than the mean), with an index of dispersion equal to $p^{-1} = 1+P$. The coefficient of variation is $(kq)^{-1/2} = [(1+P)/(kP)]^{1/2}$.

The factorial moments of negative order are

$$\mu_{-k} = E\left[\frac{X!}{(X+k)!}\right] = \int_0^1\int_0^{t_{k-1}}\cdots\int_0^{t_1}\left(\frac{1-q}{1-qz}\right)^k ds\,dt_1\cdots dt_{k-1};$$

in particular

$$\mu_{-1} = E\left[\frac{1}{X+1}\right] = \frac{(1-q)^k}{q(1-k)}[1-(1-q)^{1-k}]$$

(Balakrishnan and Nevzorov, 2003).

The mean deviation is

$$\nu_1 = \frac{2m(k+m-1)!P^m}{m!(k-1)!Q^{m+k-1}} = \frac{2m(k+m-1)!p^{k-1}q^m}{m!(k-1)!}, \tag{5.26}$$

where $m = [kP] + 1$ (that is, m is the smallest integer greater than the mean μ); see Kamat (1965).

5.5 PROPERTIES

From the relationship

$$\frac{\Pr[X = x+1]}{\Pr[X = x]} = \frac{(k+x)P}{(x+1)Q}, \tag{5.27}$$

it can be seen that

$$\Pr[X = x+1] < \Pr[X = x] \quad \text{if } x > kP - Q$$

and that

$$\Pr[X = x] \geq \Pr[X = x-1] \quad \text{if } x \leq kP - P. \tag{5.28}$$

So when $(k-1)P$ is not an integer, there is a single mode at $[(k-1)P]$, where $[\cdot]$ denotes the integer part. When $(k-1)P$ is an integer, then there are two equal modes at $X = (k-1)P$ and $X = kP - Q$. If $kP < Q$, the mode is at $X = 0$.

Van de Ven and Weber (1993) have obtained bounds for the median of the negative binomial distribution which are valid for all parameter values. Their definition of the median is $\inf\left\{x : \Pr[X \leq x] \geq \frac{1}{2}\right\}$. Göb (1994) commented that long-standing inequalities for the percentage points of the binomial cdf provide bounds for the binomial median. He then used the relationship between the binomial and the negative binomial cdf's to obtain bounds for the negative binomial median.

For fixed values of x and k the probabilities increase monotonically with P; for fixed x and P they increase monotonically with k.

When $k < 1$, we have $p_x p_{x+2}/p_{x+1}^2 > 1$ (where $p_x = \Pr[X = x]$), and therefore the probabilities are logconvex; when $k > 1$, we have $p_x p_{x+2}/p_{x+1}^2 < 1$ and so now the probabilities are logconcave. Although the probabilities satisfy the logconvexity condition that is a sufficient condition for infinite divisibility only when $k < 1$, nevertheless the distribution is a Poisson–stopped sum of logarithmic rv's and so is infinitely divisible for all values of k.

The logconvexity/logconcavity properties imply that the distribution has a decreasing hazard (failure) rate for $k < 1$ and an increasing hazard rate for $k > 1$. For $k = 1$ the failure rate is constant. This is the no-memory (Markovian) property of the geometric distribution; see Sections 5.2 and 5.9.1.

If X_1 and X_2 are independent variables each having a negative binomial distribution with the same series parameter q but with possibly different power parameters k_1 and k_2, then $X_1 + X_2$ also has a negative binomial distribution; its pgf is

$$(1 + P - Pz)^{-k_1-k_2} = p^{k_1+k_2}(1 - qz)^{-k_1-k_2}. \tag{5.29}$$

As k tends to infinity and P to zero, with kP remaining fixed ($kP = \theta$), the right-hand side of (5.1) tends to the value $e^{-\theta}\theta^k/k!$, corresponding to a Poisson distribution with expected value θ.

Young (1970) gave formulas for the moments of the order statistics for the negative binomial distribution and tabulated $E[X_{(r)}]$ for samples of size $n = 2(1)8$ to two decimal places. He showed that when $p = Q^{-1}$ is close to unity there is a good gamma approximation, enabling Gupta's (1960a) tables for gamma order statistics to be used.

Pessin (1961, 1965) noted that, as $Q \to \infty$ with k constant, the standardized negative binomial tends to a gamma distribution.

5.6 APPROXIMATIONS AND TRANSFORMATIONS

The sum of a number of negative binomial terms can be expressed in terms of an incomplete beta function ratio and hence as a sum of binomial terms. We have

$$\begin{aligned}
\sum_{j=r}^{\infty} \Pr[X = j] &= \frac{(k + r - 1)!p^k q^r}{(k - 1)!r!}\left(1 + \frac{(k + r)q}{(r + 1)} + \cdots\right) \\
&= \frac{(k + r - 1)!p^k q^r}{(k - 1)!r!}\, {}_2F_1[1, k + r; r + 1; q] \\
&= \frac{(k + r - 1)!q^r}{(k - 1)!r!}\, {}_2F_1[r, 1 - k; r + 1; q] \\
&= \frac{B_q(r, k)}{B(r, k)} = I_q(r, k).
\end{aligned} \tag{5.30}$$

Therefore

$$\begin{aligned}
\Pr[X \le r] = 1 - I_q(r + 1, k) &= I_p(k, r + 1) \\
&= \Pr[Y \ge k],
\end{aligned} \tag{5.31}$$

where Y is a binomial rv with pgf $(q + pz)^{k+r}$. This formula has been rediscovered on many occasions. Patil (1963a) gives a list of references; see also Morris (1963). Approximations for binomial distributions (already discussed in Section 3.6.1) can thereby be applied to negative binomial distributions.

Negative binomial approximations to the negative hypergeometric probabilities have been obtained by López-Blázquez and Salamanca-Miño (2001).

Bartko (1966) studied five different approximations for cumulative negative binomial probabilities. Their accuracy is similar to that of approximations for the binomial distribution. The two most useful approximations in Bartko's opinion are as follows:

1. A corrected (Gram–Charlier) Poisson approximation

$$\Pr[X \le x] = e^{-kP} \sum_{j=0}^{x} \frac{(kP)^j}{j!} - \frac{(x - kP)}{2(1 + P)} e^{-kP} \frac{(kP)^x}{x!}. \tag{5.32}$$

2. The Camp–Paulson approximation (see Johnson et al., 1995, Chapter 26)

$$\Pr[X \le x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K} e^{-u^2/2}\, du,$$

where

$$K = \frac{\left[\dfrac{9x + 8}{x + 1} - \dfrac{(9k - 1)\{kP/(x + 1)\}^{1/3}}{k}\right]}{3\left[\dfrac{\{kP/(x + 1)\}^{2/3}}{k} - \dfrac{1}{x + 1}\right]^{1/2}}. \tag{5.33}$$

Of these (5.33) is remarkably accurate, but it is much more complicated than (5.32).

Peizer and Pratt (1968) and Pratt (1968) have obtained extremely accurate normal approximations of the form

$$\Pr[X \le r] = \int_{-\infty}^{z_i} (2\pi)^{-1/2} e^{-x^2/2}\, dx,$$

where

$$\begin{aligned} z_i &= d_i \left(\frac{1 + pg(a) + qg(b)}{(r + k + 1/6)pq}\right)^{1/2}, \qquad i = 1, 2, \\ a &= \frac{r + 0.5}{(r + k)q}, \qquad b = \frac{k - 0.5}{(r + k)p}, \qquad g(u) = \frac{1 + u}{1 - u} + \frac{2u \ln u}{(1 - u)^2}, \end{aligned} \tag{5.34}$$

and

$$d_1 = \left(r + \frac{2}{3}\right) p - \left(s - \frac{1}{3}\right) q, \qquad d_2 = d_1 + 0.02 \left(\frac{p}{r + 1} - \frac{q}{k} + \frac{p - 0.5}{r + k + 1}\right).$$

Peizer and Pratt provided a table giving evidence concerning the remarkable accuracy of the two approximations.

Guenther's (1972) approximation is based on the incomplete gamma function; it is

$$\Pr[X \le r] \approx \Pr\left[\chi^2_{2kq} \le (2r+1)p\right], \tag{5.35}$$

where χ^2_{2kq} is a chi-squared variable; see Johnson et al. (1994, Chapter 17). It enables tables of the incomplete gamma function to be used. Best and Gipps (1974) presented evidence that

$$\Pr[X \le r] \approx \Pr\left[\chi^2_{8kq/(q+1)^2} \le \frac{[4r+2+4kq/(1+q)]p}{1+q}\right] \tag{5.36}$$

provides a considerable improvement over (5.35).

A transformation that approximately normalizes and approximately equalizes the variance is useful. The formulas $E[X] = kP$, $\text{Var}(X) = kP(1+P)$ suggest the transformation

$$Y_1 = \sqrt{k}\sinh^{-1}\sqrt{\frac{X}{k}}, \tag{5.37}$$

with Y_1 approximately distributed as a standard normal variable.

More detailed investigations by Anscombe (1948) indicated that the transformation

$$Y_2 = \sqrt{k-0.5}\sinh^{-1}\sqrt{\frac{X+\frac{3}{8}}{k-\frac{3}{4}}} \tag{5.38}$$

is preferable; see also Laubscher (1961).

5.7 COMPUTATION AND TABLES

Computation of individual negative binomial probabilities can be reduced to the calculation of the corresponding binomial probabilities by the use of the relationship between the tails of the binomial and negative binomial distributions; see the previous section. For low values of r the probabilities can also be computed by recursion from $\Pr[X = 0]$ using

$$\Pr[X = r+1] = \frac{(k+r)q}{r+1}\Pr[X = r]. \tag{5.39}$$

For higher values of r Stirling's expansion (Section 1.1.2) can be used for the gamma functions in the expression

$$\Pr[X = r+1] = \frac{\Gamma(k+r)p^k q^r}{\Gamma(r+1)\Gamma(k)};$$

this gives

$$\ln \Pr[X = r] \approx (k-1)\ln\left(\frac{(k+r)p}{k}\right) + (r+0.5)\ln\left(\frac{(k+r)q}{r}\right) - 0.5\ln\left(\frac{2\pi kq}{p^2}\right) - \frac{1}{12k} - \frac{k}{12r(k+r)}. \qquad (5.40)$$

Cumulative negative binomial probabilities can be computed from cumulative binomial probabilities or by summation of individual probabilities; alternatively they can be approximated using appropriate formulas from the previous section.

However, for fractional values of k, and for convenience in looking up sequences of values, direct tables can be useful. Williamson and Bretherton (1963) provided comprehensive six-decimal tables of $\Pr[X = r]$.

Grimm (1962) gave values of individual probabilities and of the cumulative distribution function to five decimal places; see also Brown (1965). Taguti (1952) gave minimum values of r for which

$$\sum_{j=0}^{r} (j!)^{-1} h(h+d)\cdots[h+(j-1)d](1+d)^{-(h/d)-j} \geq \alpha$$

for $\alpha = 0.95, 0.99$; these are (approximate) percentage points of negative binomial distributions with $k = h/d$, $P = d$.

Computer generation of rv's from a geometric distribution is very straightforward. One method is to exploit the waiting-time property. Consider a stream of uniform rv's. Then geometric rv's can be generated by counting the number of uniforms needed to obtain a uniform less than p (the number of failures needed to obtain the first success). Devroye (1986, p. 498) considers that "for $p \geq \frac{1}{3}$ the method is probably difficult to beat in any programming environment."

A second way to generate a geometric rv G is by analytic inversion of the cdf. Let U be a uniform rv. Then $G = [\ln(U)/\ln(1-p)]$ (where $[\cdot]$ denotes the integer part). If a stream of exponential rv's is available, then discretizing the exponential (E) gives $G = [-E/\ln(1-p)]$. Devroye notes (in an exercise) that there may be an accuracy problem for low values of p and that one way that this may be overcome is via the expansion

$$\ln(1-p) = \frac{2}{c}\left(1 + \frac{1}{3c^2} + \frac{1}{5c^4} + \cdots\right),$$

where $c = 1 - 2/p$ (c is negative).

The negative binomial with an integer parameter $k = N$ can be generated as the sum of N geometric rv's. Except for low values of N (say $N = 2, 3, 4$), this method cannot be advocated as it requires many uniforms for a single output negative binomial rv. This argument applies a fortiori to the use of the sum of a Poisson number of logarithmic rv's.

The method generally recommended for generating negative binomial rv's with changing parameters is to generate Poisson rv's with random parameters drawn from a gamma distribution [see, e.g., algorithm NB3 in Fishman (1978)]. For fixed parameters the use of a fast general method, such as indexed table look-up, alias, or frequency table, is recommended.

Three simple stochastic models that can be used to generate *correlated* negative binomial rv's have been described by Sim and Lee (1989). Two of their methods are based on the autoregressive scheme of the first-order Markovian process. The third uses the Poisson process from a first-order autoregressive gamma sequence.

5.8 ESTIMATION

5.8.1 Model Selection

Early graphical methods for identifying whether or not a negative binomial model is appropriate for a particular type of data were based on ratios of factorial moments (Ottestad, 1939), or probability-ratio cumulants (Gurland, 1965), or ratios of factorial cumulants (Hinz and Gurland, 1967). Ord's method of plotting $u_r = rf_r/f_{r-1}$ against r (where f_r is an observed frequency) gives an upward-sloping straight line, $u_r \approx (k + r - 1)p$; see Ord (1967a, 1972) and Tripathi and Gurland (1979). Grimm's (1970) method and the methods of Hoaglin, Mosteller, and Tukey (1985) can also be used; see Section 4.7.1.

5.8.2 *P* Unknown

Consider the total number of trials $k + X$ needed to obtain k successes when the probability of a success is p (the inverse sampling model). Then the minimum variance unbiased estimator (MVUE) of $p = (1 + P)^{-1}$, based on a single observation x of X, is

$$p^\circ = \frac{k-1}{k+x-1}, \tag{5.41}$$

and the Cramér–Rao lower bound on its variance is

$$\mathrm{Var}(p^\circ) \geq \frac{p^2 q}{k}. \tag{5.42}$$

Best (1974) stated that

$$\mathrm{Var}(p^\circ) = p^2 \sum_{r=1}^{\infty} \binom{k+r-1}{r}^{-1} q^r; \tag{5.43}$$

Mikulski and Smith (1976) showed that

$$\mathrm{Var}(p^\circ) \leq \frac{p^2 q}{k - p + 2}. \tag{5.44}$$

These bounds on the variance of p° were sharpened by Ray and Sahai (1978) and Sahai and Buhrman (1979).

Given a sample of observations from a negative binomial distribution with power parameter k, consideration of the pgf in the form $G(z) = p^k(1 - qz)^{-k}$ shows that the distribution is a PSD, and hence the maximum-likelihood equation for q is the first-moment equation $\overline{x} = k\hat{q}/(1 - \hat{q})$; thus

$$\hat{q} = \frac{\overline{x}}{k + \overline{x}}. \tag{5.45}$$

Roy and Mitra (1957) showed that the uniformly minimum variance unbiased estimator (UMVUE) of $P = q/p$ is $\tilde{\theta}/(1 - \tilde{\theta})$, where

$$\tilde{\theta} = \frac{\sum_x x f_x}{\sum_x (k + x) f_x - 1}, \tag{5.46}$$

and the f_x are the observed frequencies. The UMVUE of μ is $\sum x f_x/(k \sum f_x)$ and the UMVUE of μ_2 is $\sum x f_x \sum (x + 1) f_x/[n(n + 1)]$ (Guttman, 1958).

Irony (1992) commented that the steps needed to make Bayesian inferences about q parallel those needed for Bayesian inferences about the binomial parameter p.

Maynard and Chow (1972) constructed an approximate Pitman-type "close" estimator of P for small sample sizes. Scheaffer (1976) has studied methods for obtaining confidence intervals for $p = 1 - q$. Gerrard and Cook (1972) and Binns (1975) considered sequential estimation of the mean $kq/(1 - q)$ when k is known.

5.8.3 Both Parameters Unknown

Consider now the situation where both parameters are unknown. Because of the variability of the sample variance of the negative binomial distribution, samples with the sample variance less than the sample mean ($s^2 < \overline{x}$) will occasionally be encountered, even when a negative binomial model is appropriate. However, when this occurs, the appropriateness of the model should be examined (see, e.g., Clark and Perry, 1989).

Method of Maximum Likelihood The maximum-likelihood estimators satisfy the equations

$$\hat{k}\hat{P} = \overline{x}, \tag{5.47}$$

$$\ln(1 + \hat{P}) = \sum_{j=1}^{\infty} \left((\hat{k} + j - 1)^{-1} \sum_{i=j}^{\infty} f_j \right), \tag{5.48}$$

where f_j is an observed frequency; see Fisher (1941), Bliss and Fisher (1953), and Wise (1946). Iteration is required for the solution of these equations. It is important to realize that iteration may be very slow if the initial estimates are poor. Rapid explicit methods that can provide good initial estimates have therefore been studied extensively. Ross and Preece (1985) have advocated the use of the maximum-likelihood program (MLP) of Ross (1980).

Method of Moments The simplest way to estimate the parameters is by the method of moments, that is, by equating the sample mean $\overline{x}$ and sample variance s^2 to the corresponding population values.

Thus, if $x_1, x_2, \ldots, x_n$ are n observed values (supposed independent), we calculate the solutions $\tilde{k}$, $\tilde{P}$ of the equations

$$\tilde{k}\tilde{P} = \overline{x} \qquad \text{and} \qquad \tilde{k}\tilde{P}(1 + \tilde{P}) = s^2;$$

this gives

$$\tilde{P} = \frac{s^2}{\overline{x}} - 1, \qquad \tilde{k} = \frac{\overline{x}^2}{s^2 - \overline{x}}. \tag{5.49}$$

Bowman and Shenton (1965, 1966) obtained asymptotic formulas for the variances, covariances, and biases of the moment and maximum-likelihood estimators.

Method of Mean and Zero Frequency In place of (5.48) an equation obtained by equating the observed and expected numbers of zero values may be used. This equation is

$$f_0 = (1 + P^\dagger)^{-k^\dagger}, \tag{5.50}$$

where f_0 is the number of zero values. Combining this equation with the equation $k^\dagger P^\dagger = \overline{x}$ gives

$$\frac{P^\dagger}{\ln(1 + P^\dagger)} = -\frac{\overline{x}}{\ln f_0}. \tag{5.51}$$

Other Methods Gurland (1965) and Gurland and Tripathi (1975) put forward a method based on the solution of linear equations involving functions of the moments and/or frequencies; see also Katti and Gurland (1962a). Gurland (1965) and Hinz and Gurland (1967) concluded that estimators based on the factorial cumulants and a certain function of the zero frequency have good efficiency relative to maximum likelihood.

Pieters et al. (1977) made small-sample comparisons of various methods using simulation. Willson, Folks, and Young (1984) extended this work by considering not only the bias but also the standard deviation and the mean square error of the method of moments and maximum-likelihood estimators and by comparing these to a proposed multistage estimation procedure. In her comment on their work, Bowman (1984) pointed out the riskiness in depending on small samples when estimating k and questioned the choice of $n = 5$ as the initial sample size for the multistage procedure.