CHAPTER 6

Hypergeometric Distributions

6.1 DEFINITION

The *classical hypergeometric distribution* is the distribution of the number of white balls in a sample of n balls drawn without replacement from a population of N balls, Np of which are white and $N - Np$ are black. The probability mass function (pmf) is

$$\Pr[X = x] = \binom{Np}{x}\binom{N - Np}{n - x} \Bigg/ \binom{N}{n} \tag{6.1}$$

$$= \binom{n}{x}\binom{N - n}{Np - x} \Bigg/ \binom{N}{Np}, \tag{6.2}$$

where $n \in \mathbb{Z}^+$, $N \in \mathbb{Z}^+$, $0 < p < 1$, and $\max(0, n - N + Np) \le x \le \min(n, Np)$. The probability generating function (pgf) is

$$G(z) = \frac{{}_2F_1[-n, -Np; N - Np - n + 1; z]}{{}_2F_1[-n, -Np; N - Np - n + 1; 1]}, \tag{6.3}$$

where

$${}_2F_1[\alpha, \beta; \gamma; z] = 1 + \frac{\alpha\beta}{\gamma} \cdot \frac{z}{1!} + \frac{\alpha(\alpha + 1)\beta(\beta + 1)}{\gamma(\gamma + 1)} \cdot \frac{z^2}{2!} + \cdots$$

is a Gaussian hypergeometric series (Section 1.1.6).

The parameters in

$$G(z) = \frac{{}_2F_1[a, b; c; z]}{{}_2F_1[a, b; c; 1]}$$

need not be restricted to $a = -n$, $b = -Np$, $c = N - Np - n + 1$. Eggenberger and Pólya (1923, 1928) studied a more general urn model leading to the pmf

$$\Pr[X = x] = \binom{n}{x}\binom{-n - (w + b)/c}{-x - w/c} \Bigg/ \binom{-(w + b)/c}{-w/c}$$

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz
Copyright © 2005 John Wiley & Sons, Inc.

and pgf

$$G(z) = \frac{{}_2F_1[-n, w/c; -n+1-b/c; z]}{{}_2F_1[-n, w/c; -n+1-b/c; 1]}, \tag{6.4}$$

where $w > 0$, $b > 0$, and n is a positive integer. Here it is possible for c to be either negative or positive. Taking $w/c = -Np$ leads to the classical hypergeometric distribution.

The case $w/c > 0$ arises from several models; the names *negative (inverse) hypergeometric distribution*, *hypergeometric waiting-time distribution*, and *beta–binomial distribution* all refer to the same mathematical distribution, as shown in Section 6.2.2.

The distributions mentioned so far have finite support. A distribution with infinite support, $x = 0, 1, 2, \ldots$, and pgf

$$G(z) = \frac{{}_2F_1[k, \ell; k+\ell+m; z]}{{}_2F_1[k, \ell; k+\ell+m; 1]}, \qquad k > 0, \qquad \ell > 0, \qquad m > 0, \tag{6.5}$$

arises as a beta mixture of negative binomial distributions. It is known as the *beta–negative binomial distribution* (*beta–Pascal distribution*) and also as the *generalized Waring distribution*; see Section 6.2.3.

The classical hypergeometric distribution, like the negative (inverse) hypergeometric (i.e., the beta–binomial) and the beta–negative binomial distributions, is a member of Ord's (1967a) difference-equation system of discrete distributions (see Sections 6.4 and 2.3.3). However, none of the three hypergeometric-type distributions is a power series distribution (PSD) (Section 2.2).

6.2 HISTORICAL REMARKS AND GENESIS

6.2.1 Classical Hypergeometric Distribution

The urn sampling problem giving rise to the classical hypergeometric distribution (Section 6.1) was first solved by De Moivre (1711, p. 236), when considering a generalization of a problem posed by Huygens. A multivariate version of the problem was solved by Simpson in 1740 [see Todhunter (1865, p. 206)], but little attention was given to the univariate distribution until Cournot (1843, pp. 43, 68, 69) applied it to matters concerning conscription, absent parliamentary representatives, and the selection of deputations and juries.

The reexpression of (6.1) as (6.2) shows that the distribution is unaltered when n and Np are interchanged. Note that the pmf of a hypergeometric-type distribution can always be manipulated into the form $\binom{a}{b}\binom{c}{d} \bigg/ \binom{a+c}{b+d}$.

The pgf (6.3) for the classical hypergeometric distribution can be restated as

$$G(z) = {}_2F_1[-n, -Np; -N; 1-z] \tag{6.6}$$

using a result from the theory of terminating Gaussian hypergeometric series. This shows that the distribution is both GHPD and GHFD; see Sections 2.4.1 and 2.4.2.

The characteristic function is

$$\frac{{}_2F_1[-n, -Np; N - Np - n + 1; e^{it}]}{{}_2F_1[-n, -Np; N - Np - n + 1; 1]}, \tag{6.7}$$

and the mean and variance are

$$\mu = E[X] = np \qquad \text{and} \qquad \mu_2 = \frac{np(1-p)(N-n)}{N-1}; \tag{6.8}$$

further moment formulas are given in Section 6.3.

The properties of the distribution were investigated in depth by Pearson (1895, 1899, 1924), who was interested in developing the system of continuous distributions that now bears his name via limiting forms of discrete distributions. Important further properties of the classical hypergeometric distribution were obtained by Romanovsky (1925).

6.2.2 Beta–Binomial Distribution, Negative (Inverse) Hypergeometric Distribution: Hypergeometric Waiting-Time Distribution

This distribution arises from a number of different models. We consider first the most widely used model.

The *beta–binomial* model gives the distribution as a mixture of binomial distributions, with the binomial parameter p having a beta distribution:

$$\Pr[X = x] = \int_0^1 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \times \frac{p^{\alpha-1}(1-p)^{\beta-1}\,dp}{B(\alpha,\beta)} \tag{6.9}$$

$$= \binom{n}{x}\binom{-\alpha-\beta-n}{-\alpha-x}\Big/\binom{-\alpha-\beta}{-\alpha}$$

$$= \binom{-\alpha}{x}\binom{-\beta}{n-x}\Big/\binom{-\alpha-\beta}{n}, \qquad n \in \mathbb{Z}^+,$$

$$0 \le \alpha, \qquad 0 \le \beta, \tag{6.10}$$

where $x = 0, 1, \ldots, n$. The pgf is

$$G(z) = \frac{{}_2F_1[-n, \alpha; -\beta - n + 1; z]}{{}_2F_1[-n, \alpha; -\beta - n + 1; 1]}$$

$$= {}_2F_1[-n, \alpha; \alpha + \beta; 1 - z] \tag{6.11}$$

(the distribution is therefore both GHPD and GHFD). The mean and variance are

$$\mu = \frac{n\alpha}{\alpha+\beta} \qquad \text{and} \qquad \mu_2 = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{6.12}$$

This model has been obtained and subsequently applied in many different fields by a number of research workers (see Section 6.9.2). When $\alpha = \beta = 1$, the outcome is the discrete rectangular distribution (see Section 6.10.1).

This derivation is closely related to Condorcet's *negative hypergeometric* model, which seems to have been derived for the first time by Condorcet in 1785; see Todhunter (1865, p. 383). Let A and B be two mutually exclusive events that have already occurred v and w times, respectively, in $v + w$ trials. Let $n = k + \ell$. Then the probability that in the next n trials events A and B will happen k and ℓ times, respectively (where k and ℓ are nonnegative integers), is

$$\begin{aligned}
\frac{(k+\ell)!}{k!\ell!} \int_0^1 x^{v+k}(1-x)^{w+\ell}\,dx \Big/ \int_0^1 x^v(1-x)^w\,dx \\
&= \frac{(k+\ell)!(v+k)!(w+\ell)!(v+w+1)!}{k!\ell!(v+k+w+\ell+1)!v!w!} \\
&= \binom{n}{k}\binom{-v-w-n-2}{-v-1-k} \Big/ \binom{-v-w-2}{-v-1} \\
&= \binom{-v-1}{k}\binom{-w-1}{n-k} \Big/ \binom{-v-w-2}{n}, \qquad k = 0, 1, \ldots, n;
\end{aligned} \tag{6.13}$$

see also Pearson (1907). The parameters here are $n = n$, $Np = -v - 1$, and $N = -v - w - 2$. Because N is negative, the name *negative hypergeometric distribution* is used. The pgf is

$$G(z) = \frac{{}_2F_1[-n, v+1; -w-n; z]}{{}_2F_1[-n, v+1; -w-n; 1]} \tag{6.14}$$

and the mean and variance are

$$\mu = \frac{n(v+1)}{v+w+2} \qquad \text{and} \qquad \mu_2 = \frac{n(v+1)(w+1)(v+w+n+2)}{(v+w+2)^2(v+w+3)}. \tag{6.15}$$

Higher moments are discussed in Section 6.3.

Condorcet's derivation assumes that sampling takes place from an infinite population. Prevost and Lhuilier in 1799 recognized that an equivalent expression is obtained when two samples are taken in succession from a finite population, without any replacements; see Todhunter (1865, p. 454).

The derivation of the *distribution of the number of exceedances* is mathematically similar to Condorcet's result; see Gumbel and von Schelling (1950) and Sarkadi (1957b). Gumbel (1958) credits this to Thomas (1948). Consider two independent random samples of sizes m and n drawn from a population in which a measured character has a continuous distribution. The number of exceedances X is defined as the number (out of n) of the observed values in the second sample that exceed the rth largest of the m values in the first sample. The probability that in n future trials there will be x values exceeding the rth largest value in m past trials is

$$\Pr[X = x] = \binom{m}{r} r \binom{n}{x} (m+n)^{-1} \Big/ \binom{m+n-1}{r+x-1}$$
$$= \binom{n}{x}\binom{-m-n-1}{-r-x} \Big/ \binom{-m-1}{-r}$$
$$= \binom{-r}{x}\binom{-m+r-1}{n-x} \Big/ \binom{-m-1}{n}, \tag{6.16}$$

where $x = 0, 1, \ldots, n$. The pgf is

$$G(z) = \frac{{}_2F_1[-n, r; -n-m+r; z]}{{}_2F_1[-n, r; -n-m+r; 1]}. \tag{6.17}$$

The parameters for the exceedance model are related to the negative hypergeometric model by $r = v + 1$, $m - r = w$.

Irwin (1954) pointed out that direct sampling for a sample of fixed size n from an urn with Np white balls and $N - Np$ black balls, as in Section 6.1, *but with replacement together with an additional similarly colored ball* after each ball is drawn also gives rise to the negative hypergeometric distribution. This is a particular case of Pólya-type sampling (see Section 6.2.4).

Suppose now that sampling without replacement (as described in Section 6.1) is continued until k white balls are obtained ($0 < k \le Np$). The distribution of the number of draws that are required is known as the *inverse hypergeometric distribution* or *hypergeometric waiting-time distribution*. (The model is analogous to the inverse binomial sampling model for the negative binomial distribution; however, the range of possible values for a negative hypergeometric distribution is finite because there is not an infinitude of black balls that might be drawn.) For this distribution

$$\Pr[X = x] = \binom{Np}{k-1}(Np-k+1)\binom{N-Np}{x-k}(N-x+1)^{-1} \Big/ \binom{N}{x-1}$$
$$= \binom{x-1}{x-k}\binom{N-x}{N-Np-x+k} \Big/ \binom{N}{N-Np}$$
$$= \binom{-k}{x-k}\binom{k-Np-1}{N-Np-x+k} \Big/ \binom{-Np-1}{N-Np}, \tag{6.18}$$

where $k \in \mathbb{Z}^+$, $N \in \mathbb{Z}^+$, $0 < p < 1$, and $x = k, k+1, \ldots, k+N-Np$, by manipulation of the factorials. Comparison with (6.13) shows that this is a negative hypergeometric distribution shifted k units away from the origin.

The pgf is

$$G(z) = z^k \frac{{}_2F_1[k, Np-n; k-N; z]}{{}_2F_1[k, Np-n; k-N; 1]} \tag{6.19}$$

and the mean and variance are

$$\mu = k + \frac{(N - Np)k}{Np + 1} = \frac{k(N+1)}{Np+1},$$
$$\mu_2 = \frac{k(N - Np)(N+1)(Np+1-k)}{(Np+1)^2(Np+2)}. \tag{6.20}$$

The term "inverse hypergeometric distribution" can refer either to the total number of draws, as above, or to the number of unsuccessful draws, as in Kemp and Kemp (1956a) and Sarkadi (1957a). Bol'shev (1964) related the inverse sampling model to a two-dimensional random walk. Guenther (1975) has written a helpful review paper concerning the negative (inverse) hypergeometric distribution.

6.2.3 Beta–Negative Binomial Distribution: Beta–Pascal Distribution, Generalized Waring Distribution

The *beta–negative binomial distribution* was obtained analogously to the beta–binomial distribution by Kemp and Kemp (1956a), who commented that it arises both as a beta mixture of negative binomial distributions with the pgf $(1-\lambda)^k/(1-\lambda z)^k$ and as an F distribution mixture of the negative binomial distribution with pgf $(1+P-Pz)^{-k}$. We have the pgf

$$G(z) = \int_0^1 \left(\frac{1-\lambda}{1-\lambda z}\right)^k \times \frac{\lambda^{\ell-1}(1-\lambda)^{m-1}\, d\lambda}{B(\ell, m)} \tag{6.21}$$
$$= \int_0^\infty (1+P-Pz)^{-k} \times \frac{P^{\ell-1}(1+P)^{-\ell-m}\, dP}{B(\ell, m)}$$
$$= \frac{{}_2F_1[k, \ell; k+\ell+m; z]}{{}_2F_1[k, \ell; k+\ell+m; 1]}, \qquad k \ge 0, \qquad \ell \ge 0, \qquad m \ge 0. \tag{6.22}$$

The probabilities are

$$\Pr[X = x] = \binom{-k}{x}\binom{m+k-1}{-\ell-x}\bigg/\binom{m-1}{-\ell}$$
$$= \binom{-\ell}{x}\binom{\ell+m-1}{-k-x}\bigg/\binom{m-1}{-k}, \qquad x = 0, 1, \ldots, \tag{6.23}$$

and the mean and variance are

$$\mu = \frac{k\ell}{m-1} \qquad \text{and} \qquad \mu_2 = \frac{k\ell(m+k-1)(m+\ell-1)}{(m-1)^2(m-2)}. \tag{6.24}$$

The moments exist only for $r < m$, however; see Section 6.3.

Another name that has sometimes been used for the distribution is *inverse Markov–Pólya distribution*. Kemp and Kemp (1956a) pointed out that this distribution also arises by inverse sampling from a Pólya urn *with additional replacements*; see the next section.

Unlike the classical hypergeometric and the negative hypergeometric distributions, the support of the beta–negative binomial distribution is infinite. Also, unlike those distributions, the beta–negative binomial is *not* a Kemp GHFD. It is, however, a Kemp GHPD.

The term "beta–Pascal" is often applied to a shifted form of the distribution (6.21) with support $k, k+1, \ldots$ [see, e.g., Raiffa and Schlaifer (1961, pp. 238, 270) and Dubey (1966a)]. Here k is necessarily an integer.

The unshifted distribution with pgf (6.22) and support $0, 1, \ldots$ has been studied in considerable detail by Irwin (1963, 1968, 1975a,b,c) and Xekalaki (1981, 1983a,b,c,d) under the name *generalized Waring distribution*. Irwin (1963) developed it from the following generalization of Waring's expansion (see Section 6.10.4):

$$\frac{(c-a-1)!}{(c-a+k-1)!} = \frac{(c-1)!}{(c+k-1)!}\left[1 + \frac{ak}{c+k} + \frac{a(a+1)k(k+1)}{(c+k)(c+k+1)2!} + \cdots\right]$$
$$= \frac{(c-1)!}{(c+k-1)!}\,{}_2F_1[a, k; c+k; 1].$$

Irwin's procedure of setting $\Pr[X = x]$ proportional to the $(x+1)$th term in this series is equivalent (as he realized) to adopting the pgf

$$G(z) = \frac{{}_2F_1[a, k; c+k; z]}{{}_2F_1[a, k; c+k; 1]}, \tag{6.25}$$

where $k = -n$, $a = -Np$, and $c + k = N - Np - n + 1$ would give the classical hypergeometric distribution. But note that Irwin's restrictions on the parameters are $c > a > 0$, $k > 0$. Irwin obtained the factorial moments

$$\mu'_{[r]} = \frac{(a+r-1)!(k+r-1)!(c-a-r-1)!}{(a-1)!(k-1)!(c-a-1)!}, \tag{6.26}$$

whence, provided that they exist,

$$\mu = \frac{ak}{c-a-1} \qquad \text{and} \qquad \mu_2 = \frac{ak(c-a+k-1)(c-1)}{(c-a-1)^2(c-a-2)}. \tag{6.27}$$

Further properties of the distribution, relationships to its Pearson-type continuous analogs, tail-length behavior, and parameter estimation are the subjects of Irwin (1968, 1975a,b,c). Xekalaki (1981) has written an anthology of results concerning urn models, mixture models, conditionality models, STER (Sums successively Truncated from the Expectation of the Reciprocal) models, and some related characterizations. In Xekalaki (1983a) she studied infinite divisibility, completeness, and regression properties of the distribution and in Xekalaki (1985) she showed that the distribution can be determined uniquely from a knowledge of certain conditional distributions and some appropriately chosen regression functions. Applications of the distribution are given in Section 9.3.

Special cases of the distribution are the Yule and Waring distributions; see Sections 10.3 and 10.4.

6.2.4 Pólya Distributions

The urn models described earlier in this chapter are all particular cases of the *Pólya urn model*. This was put forward by Eggenberger and Pólya (1923) as a model for contagious distributions, that is, for situations where the occurrence of an event has an aftereffect; see also Jordan (1927) and Eggenberger and Pólya (1928).

Suppose that a finite urn initially contains w white balls and b black balls and that balls are withdrawn one at a time, with immediate replacement, together with c balls of a similar color. Then the probability that x white balls are drawn in a sample of n withdrawals is

$$\begin{aligned}\Pr[X = x] &= \binom{n}{x}\frac{w(w+c)\dots[w+(x-1)c]b(b+c)\dots[b+(n-x-1)c]}{(w+b)(w+b+c)\dots[w+b+(n-1)c]}\\ &= \binom{n}{x}\frac{B(x+w/c, n-x+b/c)}{B(w/c, b/c)}\\ &= \binom{n}{x}\binom{-n-(w+b)/c}{-x-w/c}\Big/\binom{-(w+b)/c}{-w/c}\\ &= \binom{-w/c}{x}\binom{-b/c}{n-x}\Big/\binom{-(w+b)/c}{n}. \end{aligned} \tag{6.28}$$

Other ways of expressing the probabilities are discussed in Bosch (1963). The pgf is

$$G(z) = \frac{{}_2F_1[-n, w/c; -n+1-b/c; z]}{{}_2F_1[-n, w/c; -n+1-b/c; 1]} \tag{6.29}$$

and the mean and variance are

$$\mu = \frac{nw}{b+w} \qquad \text{and} \qquad \mu_2 = \frac{nwb(b+w+nc)}{(b+w)^2(b+w+c)}. \tag{6.30}$$

Pólya (1930) pointed out the following particular cases: If c is positive, then success and failure are both contagious; if $c = 0$, then events are independent (the classical binomial situation); while if c is negative, then each withdrawal creates a reversal of fortune. When c is negative such that w/c is a negative integer (e.g., $c = -1$), the outcome is the classical hypergeometric distribution, whereas when c is positive such that *w/c* is a positive integer (e.g., $c = +1$), the negative hypergeometric distribution is the result. Inverse sampling for a fixed number of white balls leads to the inverse (negative) hypergeometric when *w/c* is a negative integer; it gives a beta–negative binomial distribution when *w/c* is a positive integer.

6.2.5 Hypergeometric Distributions in General

In

$$\Pr[X = x] = \binom{a}{x}\binom{b}{n-x}\Big/\binom{a+b}{n} \tag{6.31}$$

it is clearly not essential that all the parameters n, a, b are positive; in fact, with certain restrictions, we can take any two of them negative and the remaining one positive and still obtain a valid pmf. The conditions under which (6.31) provides an honest distribution, with n, a, and b taking real values, were investigated by Davies (1933, 1934), Noack (1950), and Kemp and Kemp (1956a). Such distributions were termed generalized hypergeometric distributions by Kemp and Kemp, but the name is now used for a much wider class. *General hypergeometric distributions* have pgf's of the form

$$G(z) = \frac{{}_2F_1[-n, -a; b-n+1; z]}{{}_2F_1[-n, -a; b-n+1; 1]}; \tag{6.32}$$

they form a subset of Kemp's (1968a,b) wider class of GHPDs with pgf's of the form

$$G(z) = \frac{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda z]}{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda]};$$

see Section 2.4.1.

Kemp and Kemp distinguished four main types of distribution corresponding to (6.31), divided into subtypes as in Table 6.1. They adopted the following conventions:

(a) If $\Pr[X = r + 1] = 0$, then $\Pr[X = x] = 0$ for all $x \geq r + 1$.

(b) When $\alpha < 0$ and $\beta < 0$ with β an integer,

$$\frac{\alpha!}{(\alpha+\beta)!} = \frac{(-1)^\beta(-\alpha-\beta-1)!}{(-\alpha-1)!}.$$

They also imposed the restriction $\Pr[X = 0] \neq 0$ to ease the derivation of the moment and other properties of the distributions.

(i) The classical hypergeometric distribution belongs to type IA(i) or IA(ii) with n, a, and b all integers.

Table 6.1 Conditions for the Existence of Types I, II, III, and IV General Hypergeometric Distributions

Type	Conditions	Support
IA(i)	$n - b - 1 < 0$; n an integer; $0 \le n - 1 < a$	$x = 0, 1, \ldots, n$
IA(ii)	$n - b - 1 < 0$; a an integer; $0 \le a - 1 < n$	$x = 0, 1, \ldots, a$
IB	$n - b - 1 < 0$; $J < a < J + 1$; $J < n < J + 1$	$x = 0, 1, \ldots$
IIA	$a < 0 < n$; n an integer; $b < 0$; $b \neq -1$	$x = 0, 1, \ldots, n$
IIB	$a < 0 < a + b + 1$; $J < n < J + 1$; $J < n - b - 1 < J + 1$	$x = 0, 1, \ldots$
IIIA	$n < 0 < a$; a an integer; $b < n - a$; $b \neq n - a - 1$	$x = 0, 1, \ldots, a$
IIIB	$n < 0 < a + b + 1$; $J < a < J + 1$; $J < n - b - 1 < J + 1$	$x = 0, 1, \ldots$
IV	$a < 0$; $n < 0$; $0 < a + b + 1$	$x = 0, 1, \ldots$

Note: J is a nonnegative integer (the same for any one type of distribution).

(ii) The negative (inverse) hypergeometric distribution belongs to Type IIA or IIIA with n, a, and b again all integers.

(iii) The beta–negative binomial is Type IV.

(iv) A dualism exists between types IA(i) and IA(ii), between Types IIA and IIIA, and between Types IIB and IIIB (using the substitutions $a \leftrightarrow n$ and $a + b - n \leftrightarrow b$).

(v) No meaningful models have been found for Types IB, IIB, or IIIB.

Sarkadi (1957a) extended the class of distributions corresponding to (6.31) by including the cases $b = -1, b = n - a - 1$ that were excluded from types IIA and IIIA, respectively, by Kemp and Kemp. He pointed out that the sum of the probabilities over the ranges $0 \le x \le n$ and $0 \le x \le a$, respectively, is equal to unity in both cases, and so (6.31) defines a proper distribution.

By changing Kemp and Kemp's definition of $\alpha!/(\alpha + \beta)!$, Shimizu (1968) and Sibuya and Shimizu (1981) were able to include further distributions with support $[m_1, m_2]$, $[m_1, \infty)$, $[-m_2, -m_1]$, $(-\infty, -m_1]$, where m_1 and m_2 are positive integers. Their new types are distributions of the form $\pm X \pm k$, where X is a rv of one of the types in Table 6.1 and k is an integer. Table 6.2 gives the

Table 6.2 Relations between Types of Hypergeometric Distributions

Kemp and Kemp	Name	Support	Ord	Shimizu
IA(i), IA(ii)	Classical hypergeometric	Finite	I(a)	A1
IIA, IIIA	Negative (inverse) hypergeometric (beta–binomial)	Finite	I(b)	A2
IB	—	Infinite	I(e)	B1
IIB, IIIB	—	Infinite	I(e)	B2
IV	Beta–negative binomial	Infinite	VI	B3

broad relationships between Kemp and Kemp (1956a), Ord (1967a), and Shimizu (1968) hypergeometric-type distributions.

Review articles concerning general hypergeometric distributions are by Guenther (1983) and Sibuya (1983).

Kemp and Kemp's (1975) paper was concerned with models for general hypergeometric distributions. Besides urn models and models for contagion, it gave (1) models based on equilibrium stochastic processes, (2) STER models, (3) conditionality models, (4) weighting models, and (5) mixing models. It showed the following:

1. An equilibrium time-homogeneous stochastic process with birth and death rates λ_i and μ_i such that $\lambda_{i-1}/\mu_i = (a_1 + i - 1)/(b + i - 1)$ can yield a type IIA/IIIA or a type IV distribution by a suitable choice of parameters. Similarly
$$\frac{\lambda_{i-1}}{\mu_i} = \frac{(a_1 + i - 1)(a_2 + i - 1)}{(b + i - 1)i} \tag{6.33}$$
can lead to any one of type IA, IIA/IIIA, or IV by a suitable choice of parameters.
2. The STER distributions arise in connection with an inventory decision problem. If demand is a discrete rv with pgf $G(z) = \sum_{i \geq 0} p_i z^i$, then the corresponding STER distribution has probabilities that are Sums successively Truncated from the Expectation of the Reciprocal of the demand variable, giving the STER pgf
$$H(z) = (1 - z)^{-1}(1 - p_0)^{-1} \int_z^1 \frac{[G(z) - p_0]\, dz}{z}; \tag{6.34}$$
see Bissinger (1965) and Section 11.2.13. Kemp and Kemp found that, if a type IIA/IIIA demand distribution has the support $1, 2, \ldots, \min(n, a)$, it gives rise to a STER distribution that is also type IIA/IIIA; see also Kemp and Kemp (1969b).
3. Let X and Y be mutually independent discrete rv's. If X and Y are both binomial with parameters (n, p) and (m, p), then the conditional distribution of $X|(X + Y)$ is hypergeometric (type IA). If X and Y are both negative binomial with parameters (u, λ) and (v, λ), then $X|(X + Y)$ has a negative hypergeometric (type IIA/IIB) distribution (Kemp, 1968a). If
$$G_X(z) = \frac{{}_1F_1[-n; c; -\lambda z]}{{}_1F_1[-n; c; \lambda]} \qquad \text{and} \qquad G_Y(z) = \exp \lambda(z - 1),$$
then the distributions of $X|(X + Y)$ and $Y|(X + Y)$ are both type IA. Discrete Bessel distributions for X and Y can also lead to a type IA distribution for $X|(X + Y)$. Similarly, binomial distributions with parameters (n, p) and $(m, 1 - p)$ for X and Y lead to a type IA distribution for $X|(Y - X)$. Kemp and Kemp (1975) gave further models of this kind; see Kemp (1968a) for the general theory.

4. Weighting models give rise to distributions that have been modified by the method of ascertainment. When the weights (sampling chances) w_x are proportional to the value of the observation (i.e., to x), the distribution with pgf ${}_2F_1[a_1, a_2; b; z]/{}_2F_1[a_1, a_2; b; 1]$ is ascertained as the distribution with pgf

$$G(z) = z\frac{{}_2F_1[a_1+1, a_2+1; b+1; z]}{{}_2F_1[a_1+1, a_2+1; b+1; 1]}; \tag{6.35}$$

if w_x is proportional to $x!/(x-k)!$, then the same initial distribution is ascertained as the distribution with pgf

$$G(z) = z^k\,\frac{{}_2F_1[a_1+k, a_2+k; b+k; z]}{{}_2F_1[a_1+k, a_2+k; b+k; 1]} \tag{6.36}$$

(Kemp, 1968a).

5. Kemp and Kemp (1975) pointed out that a beta mixture of extended beta–binomial distributions can, under certain circumstances, give rise to a beta–binomial distribution. Two possibilities are as follows:

$$\begin{aligned} G_1(z) &= \int_0^1 {}_2F_1[-n, a; c; y(z-1)]\frac{y^{c-1}(1-y)^{d-c-1}\,dy}{B(c, d-c)} \\ &= {}_2F_1[-n, a; d; z-1]; \end{aligned} \tag{6.37}$$

$$\begin{aligned} G_2(z) &= \int_0^1 {}_2F_1[-n, d; b; y(z-1)]\frac{y^{c-1}(1-y)^{d-c-1}\,dy}{B(c, d-c)} \\ &= {}_2F_1[-n, c; b; z-1]. \end{aligned} \tag{6.38}$$

They also commented that a type IIA/IIIA distribution can be obtained as a gamma mixture of restricted Laplace–Haag distributions, and they pointed out that a type IV distribution can be derived as a mixture of Poisson distributions.

6.3 MOMENTS

The moment properties of the general hypergeometric distribution can be obtained from the factorial moments and indeed exist only when the factorial moments exist. The general form for the rth factorial moment (if it exists) for the distribution with pgf (6.31) is

$$\mu'_{[r]} = \frac{n!a!(a+b-r)!}{(n-r)!(a-r)!(a+b)!}, \tag{6.39}$$

and so

$${}_2F_1[-a, -n; -a-b; -t] \tag{6.40}$$

can be treated as the factorial moment generating function (fmgf) for the factorial moments.

Moments exist for the following:

Type IA(i)	Always (but are zero if $r > n$)
Type IA(ii)	Always (but are zero if $r > a$)
Type IB	When $r < a + b + 1$
Type IIA	Always (but are zero if $r > n$)
Type IIB	Never
Type IIIA	Always (but are zero if $r > a$)
Type IIIB	Never
Type IV	When $r < a + b + 1$

In other words,

$\mu'_{[r]}$ is finite for all r for types IA(i), IA(ii), IIA, and IIIA;
$\mu'_{[r]}$ is finite for $r < a + b + 1$ for types IB and IV; and
types IIB and IIIB have no moments.

Provided that the specified moment exists, it is straightforward (though tedious) to show via the factorial moments that

$$
\begin{aligned}
E[X] = \mu &= \frac{na}{a+b}, \\
\mathrm{Var}(X) = \mu_2 &= \frac{nab(a+b-n)}{(a+b)^2(a+b-1)}, \\
\mu_3 &= \frac{\mu_2(b-a)(a+b-2n)}{(a+b)(a+b-2)}, \\
\mu_4 &= \frac{\mu_2}{(a+b-2)(a+b-3)}\Bigg\{(a+b)(a+b+1-6n) \\
&\quad + 3ab(n-2) + 6n^2 + \frac{3abn(6-n)}{a+b} - \frac{18abn^2}{(a+b)^2}\Bigg\}.
\end{aligned}
\tag{6.41}
$$

The moment ratios are

$$
\sqrt{\beta_1} = \left[\frac{(a+b-1)}{abn(a+b-n)}\right]^{1/2} \frac{(b-a)(a+b-2n)}{(a+b-2)}, \tag{6.42}
$$

$$
\begin{aligned}
\beta_2 &= \frac{(a+b)^2(a+b-1)}{nab(a+b-n)(a+b-2)(a+b-3)} \\
&\quad \times \Bigg((a+b)(a+b+1-6n) + 3ab(n-2) \\
&\quad + 6n^2 + \frac{3abn(6-n)}{a+b} - \frac{18abn^2}{(a+b)^2}\Bigg) + 3.
\end{aligned}
\tag{6.43}
$$

Table 6.3 Comparison of Hypergeometric Types IA(ii), IIA, IIIA, and IV, Binomial, Poisson, and Negative Binomial Distributions

$\Pr[X = x]$

x	IA(ii)	Binomial	IIA	Poisson	IIIA	Negative Binomial	IV
0	0.076	0.107	0.137	0.135	0.123	0.162	0.197
1	0.265	0.269	0.266	0.271	0.265	0.269	0.267
2	0.348	0.302	0.270	0.271	0.284	0.247	0.220
3	0.222	0.201	0.184	0.180	0.195	0.164	0.144
4	0.075	0.088	0.093	0.090	0.093	0.089	0.083
5	0.013	0.026	0.036	0.036	0.032	0.042	0.044
6	0.001	0.006	0.011	0.012	0.007	0.017	0.023
7	0.000	0.001	0.003	0.003	0.001	0.007	0.011
8	0.000	0.000	0.000	0.001	0.000	0.002	0.005
9	—	0.000	0.000	0.000	—	0.001	0.003
10	—	0.000	0.000	0.000	—	0.000	0.001
11	—	—	—	0.000	—	0.000	0.001
$\geq$12	—	—	—	0.000	—	0.000	0.001

Note: For each distribution the mean is 2, $|n| = 10$, $|a/(a+b)| = 0.2$, and $|a+b| = 40$. A dash means that the probability is zero and 0.000 means that the probability is less than 0.0005.
Source: Adapted from Kemp and Kemp (1956a).

As $a \to \infty$, $b \to \infty$ such that $a/(a+b) = p$ (constant), the moment properties tend to those of the binomial distribution, provided that n is a positive integer. When n is negative and $a/(a+b)$ tends to λ, $\lambda < 0$, then the moment properties tend to those of the negative binomial distribution; see Table 6.3.

An alternative approach to the moment properties is via the differential equation for the mgf. The pgf is

$$G(z) = K\,{}_2F_1[-n, -a; b-n+1; z],$$

where K is a normalizing constant; this satisfies

$$\theta(\theta + b - n)G(z) = z(\theta - n)(\theta - a)G(z), \tag{6.44}$$

where θ is the differential operator $z\ d/dz$. The moment generating function (mgf) is $G(e^t)$; from the relationship between the θ and $D = d/dz$ operators (Section 1.1.4) it follows that $G(e^t)$ satisfies

$$D(D + b - n)G(e^t) = e^t(D - n)(D - a)G(e^t), \tag{6.45}$$

where D is the differential operator d/dt. The central mgf is $M(t) = e^{-\mu t}G(e^t)$, and this satisfies

$$(D + \mu)(D + \mu + b - n)M(t) = e^t(D + \mu - n)(D + \mu - a)M(t). \tag{6.46}$$

Identifying the coefficients of t^0, t^1, t^2, and t^3 in (6.46) gives expressions for the first four central moments that are equivalent to (6.41). Higher moments may be obtained similarly. This is essentially the method of Pearson (1899).

Lessing (1973) has shown that the uncorrected moments can be obtained from the following expression for the mgf:

$$G(e^t) = \frac{(a+b-n)!}{(a+b)!} \frac{\partial^n}{\partial y^n} \left[(1+ye^t)^a (1+y)^b\right]_{y=0}. \tag{6.47}$$

Janardan (1973b) commented that this result is a special case of an expression obtained by Janardan and Patil (1972).

The following finite difference relation holds among the central moments $\{\mu_j\}$:

$$(a+b)\mu_{r+1} = [(1+E)^r - E^r][\mu_2 + \alpha\mu_1 + \beta\mu_0], \tag{6.48}$$

where E is the displacement operator (i.e., $E^p[\mu_s] \equiv \mu_{s+p}$),

$$\alpha = -a + \frac{n(a-b)}{a+b}, \qquad \beta = \frac{nab(a+b-n)}{(a+b)^2},$$

and $\mu_0 = 1, \mu_1 = 0$ (Pearson, 1924).

The mean deviation is

$$\begin{aligned} \nu_1 &= E\left[\left|\frac{X-na}{a+b}\right|\right] \\ &= \frac{2m(b-n+m)}{a+b}\binom{a}{m}\binom{b}{n-m}\Big/\binom{a+b}{n}, \end{aligned} \tag{6.49}$$

where m is the greatest integer not exceeding $\mu + 1$ (Kamat, 1965).

Matuszewski (1962) and Chahine (1965) have studied the ascending factorial moments of the negative (inverse) hypergeometric distribution.

6.4 PROPERTIES

Let

$$f(x|n,a,b) = \Pr[X = x] = \binom{a}{x}\binom{b}{n-x}\Big/\binom{a+b}{n}, \tag{6.50}$$

$$F(x|n,a,b) = \sum_j \Pr[X = j] = \sum_j \binom{a}{j}\binom{b}{n-j}\Big/\binom{a+b}{n}, \tag{6.51}$$

where the range of summation for j is $\max(0, n-b) \le j \le x$ [or $0 \le j \le \min(x,n,a)$ when n and a are positive]. Then the following probability relationships hold:

$$f(x+1|n,a,b) = \frac{(a-x)(n-x)}{(x+1)(b-n+x+1)} f(x|n,a,b), \tag{6.52}$$

$$f(x|n, a+1, b-1) = \frac{(a+1)(b-n+x)}{(a+1-x)b} f(x|n, a, b), \tag{6.53}$$

$$f(x|n+1, a, b) = \frac{(b-n+x)(n+1)}{(n+1-x)(a+b-n)} f(x|n, a, b), \tag{6.54}$$

$$f(x|n, a, b+1) = \frac{(a+b-n+1)(b+1)}{(b-n+x+1)(a+b+1)} f(x|n, a, b). \tag{6.55}$$

Also

$$\begin{aligned} f(x|n, a, b) &= f(n-x|n, b, a) \\ &= f(a-x|a+b-n, a, b) \\ &= f(b-n+x|a+b-n, b, a); \end{aligned} \tag{6.56}$$

see Lieberman and Owen (1961). Furthermore

$$\begin{aligned} F(x|n, a, b) &= 1 - F(n-x-1|n, b, a) \\ &= F(b-n+x|a+b-n, b, a) \qquad (6.57) \\ &= 1 - F(a-x-1|a+b-n, a, b). \qquad (6.58) \end{aligned}$$

Raiffa and Schlaifer (1961) obtained relationships between the tails of the hypergeometric, beta–binomial (negative hypergeometric), and beta–negative binomial distributions. These authors used a different notation from Lieberman and Owen (1961). Let $F_h(\cdot)$ and $G_h(\cdot)$ denote the lower and upper tails of a classical hypergeometric distribution; then

$$\begin{aligned} &G_h(k|n, \ell+m-1, k+m-1) \\ &\quad = \sum_{x \geq k} \binom{k+m-1}{x}\binom{\ell+n-k}{n-x} \bigg/ \binom{\ell+m+n-1}{n} \\ &\quad = \sum_{x \leq n-k} \binom{\ell+n-k}{x}\binom{k+m-1}{n-x} \bigg/ \binom{\ell+m+n-1}{n} \\ &\quad = F_h(n-k|n, \ell+m-1, \ell+n-k). \end{aligned} \tag{6.59}$$

Furthermore, let

$$\begin{aligned} G_{\beta b}(k|m, \ell+m, n) &= \sum_{x \geq k} \binom{-m}{x}\binom{-\ell}{n-x} \bigg/ \binom{-\ell-m}{n} \\ &= \sum_{x \geq k} \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \times \frac{p^{m-1}(1-p)^{\ell-1}\,dp}{B(\ell, m)} \end{aligned} \tag{6.60}$$

be the upper tail of a beta–binomial (negative hypergeometric) distribution. Also let

$$
\begin{aligned}
&F_{\beta Pa}(n|m, \ell + m, k) \\
&\quad = \sum_{x \le n-k} \binom{-k}{x}\binom{k+m-1}{-\ell-x} \Big/ \binom{m-1}{-\ell} \\
&\quad = \sum_{x \le n-k} \int_0^1 \binom{k+x-1}{x}(1-\lambda)^k\lambda^x \times \frac{\lambda^{\ell-1}(1-\lambda)^{m-1}\,d\lambda}{B(\ell, m)}
\end{aligned}
\tag{6.61}
$$

denote the lower tail of a beta–negative binomial (beta–Pascal) distribution. Then from the relationship between the tails of a binomial and a negative binomial distribution (Section 5.6), Raiffa and Schlaifer proved that

$$G_{\beta b}(k|m, \ell + m, n) = F_{\beta Pa}(n|m, \ell + m, k)$$

and hence that

$$
\begin{aligned}
G_{\beta b}(k|m, \ell + m, n) &= F_{\beta Pa}(n|m, \ell + m, k) &(6.62)\\
&= G_h(k|n, \ell + m - 1, k + m - 1) &(6.63)\\
&= F_h(n - k - 1|n, \ell + m - 1, \ell + n - k) &(6.64)
\end{aligned}
$$

by a probabilistic argument; see Raiffa and Schlaifer (1961, pp. 238–239).

From (6.52), $\Pr[X = x + 1]$ is greater or less than $\Pr[X = x]$ according as

$$\frac{(a-x)(n-x)}{(x+1)(b-n+x+1)} \gtrless 1,$$

that is, according as

$$x \gtrless \frac{(n+1)(a+1)}{(a+b+2)} - 1. \tag{6.65}$$

Let $c = (n + 1)(a + 1)/(a + b + 2)$. Then $\Pr[X = x]$ increases with x, reaching a maximum at the greatest integer that does not exceed c, and then decreases. The mode of the distribution is therefore at $[c]$, where $[\cdot]$ denotes the integer part. If c is an integer, then there are two equal maxima at $c - 1$ and c. [Note that, if a and b are large, then the mode is very close to the mean, since $\mu = na/(a + b)$.]

The classical hypergeometric distribution is known to have a monotone likelihood ratio in x for known values of n and $a + b$ (Ferguson, 1967).

The following limiting results hold:

(i) The classical hypergeometric distribution tends to a Poisson distribution with mean μ as $n \to \infty$, $(a + b) \to \infty$ such that $na/(a + b) = \mu$, μ constant. Feller (1957), Nicholson (1956), Molenaar (1970a), and Lieberman and Owen (1961) have examined conditions under which it tends to a normal distribution.

(ii) As $a \to \infty$, $b \to \infty$ such that $a/(a+b) = p$, p constant, $0 < p < 1$, a type IA(i) distribution tends to a binomial distribution with parameters n, p.

(iii) If a is a positive integer, then as $n \to \infty$, $b \to \infty$ such that $n/b = p$, p constant, $0 < p < 1$, a type IA(ii) distribution tends to a binomial distribution with parameters a, p.

(iv) As a and b both tend to $-\infty$ in such a way that $a/(a+b) = p$, p constant, $0 < p < 1$, a type IIA distribution tends to a binomial distribution with parameters n, p.

(v) As $a \to \infty, b \to -\infty$ such that $a/b = -\lambda, \lambda$ constant, $0 < \lambda < 1$, a type IIIA distribution tends to a negative binomial distribution with pgf $(1-\lambda)^k/(1-\lambda z)^k$, where $k = -n$.

(vi) As $a \to -\infty, b \to \infty$ such that $a/b = -\lambda, \lambda$ constant, $0 < \lambda < 1$, a type IV distribution similarly tends to a negative binomial distribution.

(vii) From the duality relationship between type IIA and type IIIA, a type IIA distribution can also tend to a negative binomial and a type IIIA can tend to a positive binomial distribution.

A comparison between hypergeometric types IA(i), IIA, IIIA, and IV, binomial, Poisson, and negative binomial distributions [from Kemp and Kemp (1956a)] is presented in Table 6.3; see also Table 3.1.

6.5 APPROXIMATIONS AND BOUNDS

There is a considerable variety of approximations to the individual probabilities, and also to cumulative sums of probabilities, for the classical hypergeometric distribution. Many of these are based on the approximation of the hypergeometric distribution (6.1) by a binomial distribution with parameters n, p.

Sródka (1963) obtained some very good bounds on the probabilities:

$$\binom{n}{x}\left(\frac{Np-x}{N}\right)^x\left(\frac{N-Np-n+x}{N}\right)^{n-x}\left(1+\frac{6n^2-6n-1}{12N}\right) < \Pr[X=x] < \binom{n}{x}p^x(1-p)^{n-x}\left(1-\frac{n}{N}\right)^{-n}\left(1+\frac{6n^2+6n-1}{12N}\right)^{-1}. \tag{6.66}$$

For sufficiently large N these can be simplified to

$$\binom{n}{x}\left(\frac{Np-x}{N}\right)^x\left(\frac{N-Np-n+x}{N}\right)^{n-x} < \Pr[X=x] < \binom{n}{x}p^x(1-p)^{n-x}\left(1-\frac{n}{N}\right)^{-n}. \tag{6.67}$$

It is often adequate to use the simple binomial approximation

$$\Pr[X = x] \approx \binom{n}{x} p^x (1-p)^{n-x} \tag{6.68}$$

when $n < 0.1N$.

There is a marked improvement if n and p are replaced by n^* and p^*, where

$$p^* = \frac{(n-1) + (N-n)p}{N-1} \qquad \text{and} \qquad n^* = \frac{np}{p^*}, \tag{6.69}$$

that is, if $n^* p^*$ and $n^* p^*(1-p^*)$ are set equal to the theoretical mean and variance of the hypergeometric distribution (Sandiford, 1960).

Greater accuracy still may be obtained by using the following modification suggested by Ord (1968a):

$$\Pr[X = x] \approx \binom{n}{x} p^x (1-p)^{n-x} \left[1 + \frac{x(1-2p) + np^2 - (x-np)^2}{2Np(1-p)}\right]. \tag{6.70}$$

Burr (1973) showed that

$$\Pr[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \left[1 + \frac{x - (x-np)^2}{2Np} + O\left(\frac{1}{N^2 p^2}\right)\right], \tag{6.71}$$

and that for $n > Np$ a closer approximation is obtained by interchanging the roles of n and Np.

Ma (1982) independently derived an approximation for $n \le Np$ that is equivalent to Ord's approximation. He also showed that when $n > Np$ interchanging the roles of n and Np gives a better approximation.

Since, as already noted, the hypergeometric distribution is unchanged by interchanging n and Np, it is clear that a binomial with parameters Np, n/N has a claim equal to that of a binomial with parameters n, p as an approximating distribution for (6.1). In addition, the distribution of $n - x$ could be approximated by a binomial with parameters $N - Np$, n/N; similarly the distribution of $Np - x$ could be approximated by a binomial with parameters $N - n$, p. Brunk et al. (1968) compared these approximations. Their investigations support the opinion of Lieberman and Owen (1961) that it is best to use the binomial with smallest power parameter, that is, $\min(n, Np, N - Np, N - n)$.

The following binomial-type approximation for the cumulative probabilities was obtained by Wise (1954):

$$\sum_{j=0}^{x} \Pr[X = j] \approx \sum_{j=0}^{x} \binom{n}{x} w^j (1-w)^{n-j}, \tag{6.72}$$

where $w = \left(Np - \frac{1}{2}x\right) / \left(N - \frac{1}{2}n + \frac{1}{2}\right)$; that is, he showed that the distribution (6.1) is approximated by a binomial distribution with parameters n and

$$\left(Np - \tfrac{1}{2}x\right) / \left(N - \tfrac{1}{2}n - \tfrac{1}{2}\right).$$

A more complicated approximation of a similar type was constructed by Bennett (1965).

Molenaar (1970a) found that the use of

$$p^{\ddagger} = \frac{Np - x/2}{N - (n-1)/2} - \frac{n\left(x - np - \frac{1}{2}\right)}{6[N - (n-1)/2]^2} \qquad (6.73)$$

gives very accurate results even when $n/N > 0.1$.

Uhlmann (1966) made a systematic comparison between the hypergeometric distribution (with parameters n, Np, N, where $0 < p < 1$) and the binomial distribution (with parameters n, p). Denoting $\Pr[X \le c]$ for the two distributions by $L_{N,n,c}(p)$ and $L_{n,c}(p)$, respectively, he showed that in general

$$L_{N,n,c}(p) - L_{n,c}(p) \begin{cases} =0 & \text{for } p = 0, \\ >0 & \text{for } 0 < p \le c(n-1)^{-1}N(N+1)^{-1}, \\ <0 & \text{for } c(n-1)^{-1}N(N+1)^{-1} + (N+1)^{-1} \le p < 1, \\ =0 & \text{for } p = 1; \end{cases}$$

these results simplify when n is odd.

Subsequently, in a study of the relationship between hypergeometric and binomial pmf's, Ahrens (1987) used simple majorizing functions (upper bounds) for their ratio. He found that the ratio of the hypergeometric to the binomial pmf can always be kept below $\sqrt{2}$ by a suitable choice of approximating binomial.

If the binomial approximation to the hypergeometric distribution can itself be approximated by a Poisson or normal approximation (see Section 3.6.1), then there is a corresponding Poisson or normal approximation to the hypergeometric. Thus when p is small but n is large, the Poisson approximation

$$\Pr[X = x] \approx \frac{e^{-np}(np)^x}{x!} \qquad (6.74)$$

may be used. Burr (1973) sharpened this approximation.

The χ^2-test for association in a 2×2 contingency table uses a χ^2 approximation for the tail of a hypergeometric distribution; this is

$$\Pr[X \le x] \approx \Pr[\chi^2_{[1]} \ge T], \qquad (6.75)$$

where

$$T = \frac{(N-1)(x - np)^2}{p(1-p)n(N-n)}.$$

The relationship between the tails of a χ^2 distribution (Johnson et al., 1994, Chapter 17) and a Poisson distribution means that this is a Poisson-type approximation for the cumulative hypergeometric probabilities. It can be improved by the use of Yates' correction, giving "the usual $\frac{1}{2}$-corrected chi-statistic" of Ling and Pratt (1984). It is appropriate for n large, provided that p is not unduly small.

From the relationship between the χ^2 distribution with one degree of freedom and the normal distribution, we have, when p is not small and n is large,

$$\Pr[X \le x] \approx (2\pi)^{-1/2} \int_{-\infty}^{y} \exp\left(-\frac{u^2}{2}\right) du, \tag{6.76}$$

with

$$y = \frac{x - np + \frac{1}{2}}{[(N-n)np(1-p)/(N-1)]^{1/2}}.$$

Hemelrijk (1967) reported that, unless the tail probability is less than about 0.07 *and* $Np + n \le N/2$, some improvement is effected by replacing $(N-1)^{-1}$ by N^{-1} under the square-root sign. A more refined normal approximation was proposed by Feller (1957) and Nicholson (1956).

Pearson (1906) approximated hypergeometric distributions by (continuous) Pearson-type distributions. This work was continued by Davies (1933, 1934). The Pearson distributions that appeared most promising were type VI or III. Bol'shev (1964) also proposed an approximation of this kind that gives good results for $N \ge 25$.

Normal approximations would, however, seem to be the most successful. Ling and Pratt (1984) carried out an extensive empirical study of 12 normal and 3 binomial approximations for cumulative hypergeometric probabilities, including two relatively simple normal approximations put forward by Molenaar (1970a, 1973). Ling and Pratt considered that binomial approximations are not appropriate as competitors to normal approximations because of the computational problems with the tails. The four normal approximations that they found best originated from an unpublished paper; this was submitted to the *Journal of the American Statistical Association* by D. B. Peizer in 1968 but was never revised or resubmitted. Peizer's approximations are extremely good, but they are considerably more complicated than those above of Molenaar. See Ling and Pratt (1984) for details of the approximations that they studied.

Some new binomial approximations to the hypergeometric distribution have recently been obtained by López-Blázquez and Salamanca Miño (2000).

6.6 TABLES, COMPUTATION, AND COMPUTER GENERATION

An extensive set of tables of individual and cumulative probabilities for the classical hypergeometric distribution was prepared by Lieberman and Owen (1961). They give values for individual and cumulative probabilities to six decimal places for

$$N = 2(1)50(10)100, \qquad Np = 1(1)N-1, \qquad n = 1(1)Np.$$

Less extensive tables were published earlier by Chung and DeLury (1950). Graphs based on hypergeometric probabilities were given by Clark and Koopmans (1959).

Guenther (1983) thought that the best way to evaluate $\Pr[X \leq x]$ is from the Lieberman and Owen tables or by means of a packaged computer program. Computer algorithms have been provided by Freeman (1973) and Lund (1980). Lund's algorithm was improved by Shea (1989) and by Berger (1991).

Little serious attention seems to have been given to the use of Stirling's expansion for the computation of individual hypergeometric probabilities.

Computer generation of classical hypergeometric random variables has been discussed in detail by Kachitvichyanukul and Schmeiser (1985). When the parameters remain constant, the alias method of Walker (1977) and Kronmal and Peterson (1979) is a good choice. Kachitvichyanukul and Schmeiser gave an appropriate program with safeguards to avoid underflow.

The simplest of all algorithms needs a very fast uniform generator. It is based on a sequence of trials in which the probability of success depends on the number of previous successes; that is, it uses the model of finite sampling without replacement. The number of successes in a fixed number of trials is counted. Fishman (1973) and McGrath and Irving (1973) have given details.

Fishman's (1978) algorithm requires a search of the cdf. Kachitvichyanukul and Schmeiser (1985) suggest ways in which the speed of the method can be improved.

Devroye (1986) indicated in an exercise how hypergeometric rv's can be generated by rejection from a binomial envelope distribution.

For large-scale simulations with changing parameters, Kachitvichyanukul and Schmeiser's algorithm H2PE uses acceptance/rejection from an envelope consisting of a uniform with exponential tails. The execution time is bounded over the range of parameter values for which the algorithm is intended, that is, over the range $M - \max(0, n - N + Np) \geq 10$, where M is the mode of the distribution. Kachitvichyanukul and Schmeiser recommended inversion of the cumulative distribution function for other parameter values.

While the negative hypergeometric distribution could be generated by inverse sampling without replacement for a fixed number of successes, it would seem preferable to generate it using its beta–binomial model with the good extant beta and binomial generators.

Similarly a hypergeometric type IV distribution could be generated as a beta mixture of negative binomials; see Section 6.2.5.

6.7 ESTIMATION

Most papers on estimation for hypergeometric-type distributions have concentrated on particular distributions (e.g., beta–binomial distribution). Rodríguez-Avi et al. (2003) have recently studied a variety of estimation methods for distributions with pgf's of the general form

$$G(z) = \frac{{}_2F_1[\alpha, \beta; \lambda; z]}{{}_2F_1[\alpha, \beta; \lambda; 1]}.$$

These include (i) methods based on relations between moments and frequencies and the observed values; (ii) the minimum χ^2 procedure; and (iii) maximum likelihood. Two applications to real data are provided.

6.7.1 Classical Hypergeometric Distribution

In one of the most common situations, inspection sampling (see Section 6.9.1), there is a single observation of r defectives (successes!) in a sample of size n taken from a lot of size N. Both N and n are known, and the hypergeometric parameter Np denotes the number of defectives in a lot; an estimate of Np is required.

The maximum-likelihood estimator $\widehat{Np}$ is the integer maximizing

$$\binom{\widehat{Np}}{r}\binom{N-\widehat{Np}}{n-r} \tag{6.77}$$

for the observed value r. From the relationship between successive probabilities,

$$\Pr[X=r|n, Np+1, N] \gtrless \Pr[X=r|n, Np, N]$$

according as $Np \gtrless n^{-1}r(N+1)-1$. Hence $\widehat{Np}$ is the greatest integer not exceeding $r(N+1)/n$; if $r(N+1)/n$ is an integer, then $[r(N+1)/n]-1$ and $r(N+1)/n$ both maximize the likelihood. The variance of $r(N+1)/n$ is, from (6.8),

$$\frac{(N+1)^2(N-n)p(1-p)}{n(N-1)}.$$

Neyman confidence intervals for Np have been tabulated extensively, notably by Chung and DeLury (1950) and Owen (1962), and have been used widely. Steck and Zimmer (1968) outlined how these may be obtained; see also Guenther (1983). Steck and Zimmer also derived Bayes confidence intervals for Np based on various special cases of a Pólya prior distribution. They related these to Neyman confidence intervals. It would seem that Bayes confidence intervals are highly sensitive to choice of prior distribution.

A test of $Np = a$ against $Np = a_0$ is sometimes required. Guenther (1977) discussed hypothesis testing in this context, giving numerical examples.

In the simplest capture–recapture application (again see Section 6.9.1) we want to estimate the total size of a population N, with both $n_1 = Np$ (the number caught on the first occasion) and $n_2 = n$ (the number caught on the second occasion) known, given a single observation r. Here

$$\Pr[X=r|n_2, n_1, N+1] \gtrless \Pr[X=r|n_2, n_1, N] \tag{6.78}$$

according as $N \gtrless (n_1n_2/r)-1$. Hence the MLE $\hat{N}$ of N is the greatest integer not exceeding n_1n_2/r; if n_1n_2/r is an integer, then $(n_1n_2/r)-1$ and n_1n_2/r both maximize the likelihood.

The properties of the estimator $\hat{N}$ have been discussed in detail by Chapman (1951). Usually $n_1+n_2 \not> N$, in which case the moments of $\hat{N}$ are infinite. Because of the problems of bias and variability concerning $\hat{N}$, Chapman suggested instead the use of the estimator

$$N^* = \frac{(n_1+1)(n_2+1)}{r+1} - 1. \tag{6.79}$$

He found that

$$E[N^* - N] = \frac{(n_1+1)(n_2+1)(N-n_1)!(N-n_2)!}{(N+1)!(N-n_1-n_2-1)!}, \tag{6.80}$$

which is less than 1 when $N > 10^4$ and $n_1n_2/N > 9.2$. The variance of N^* is approximately

$$N^2(m^{-1} + 2m^{-2} + 6m^{-3}), \tag{6.81}$$

and its coefficient of variation is approximately $m^{-1/2}$, where $m = n_1n_2/N$. Chapman (1951, p. 148) concluded that "sample census programs in which the expected number of tagged members is much smaller than 10 may fail to give even the order of magnitude of the population correctly."

Robson and Regier (1964) discussed the choice of n_1 and n_2. Chapman (1948, 1951) showed how large-sample confidence intervals for N^* can be constructed; see also Seber (1982b).

The estimator

$$N^{**} = \frac{(n_1+2)(n_2+2)}{r+2} \tag{6.82}$$

has also been suggested for n_2 sufficiently large. Here

$$\begin{aligned} E(N^{**}) &\approx N(1 - m^{-1}), \\ \text{Var}(N^{**}) &\approx N^2(m^{-1} - m^{-2} - m^{-3}). \end{aligned} \tag{6.83}$$

In epidemiological studies the estimation of a target population size N is quite often achieved by merging two lists of sizes n_1 and n_2; see Section 9.1. Here it is usual to have $n_1 + n_2 > N$, in which case $\hat{N}$ is unbiased, and

$$s^2 = \frac{(n_1+1)(n_2+1)(n_1-r)(n_2-r)}{(r+1)^2(r+2)}$$

(where r is the number of items in common on the two lists) is an unbiased estimator of $\text{Var}(\hat{N})$ (Wittes, 1972).

6.7.2 Negative (Inverse) Hypergeometric Distribution: Beta–Binomial Distribution

The *beta–binomial distribution* is the most widely used of all the general hypergeometric distributions. It is particularly useful for regression situations involving binary data.

Consider the beta–binomial parameterization

$$\Pr[X = x] = \binom{-\alpha}{x}\binom{-\beta}{n-x}\Big/\binom{-\alpha-\beta}{n} \tag{6.84}$$

from Section 6.2.2, with $\alpha, \beta, n > 0$, n an integer. Moment and maximum-likelihood estimation procedures for the parameters α and β were devised by Skellam (1948) and Kemp and Kemp (1956b).

The moment estimators are obtained by setting

$$\overline{x} = \frac{\tilde{\alpha} n}{\tilde{\alpha} + \tilde{\beta}}, \qquad s^2 = \frac{n\tilde{\alpha}\tilde{\beta}(\tilde{\alpha} + \tilde{\beta} + n)}{(\tilde{\alpha} + \tilde{\beta})^2(\tilde{\alpha} + \tilde{\beta} + 1)}, \tag{6.85}$$

that is,

$$\tilde{\alpha} = \frac{(n - \overline{x} - s^2/\overline{x})\overline{x}}{(s^2/\overline{x} + \overline{x}/n - 1)n}, \qquad \tilde{\beta} = \frac{(n - \overline{x} - s^2/\overline{x})(n - \overline{x})}{(s^2/\overline{x} + \overline{x}/n - 1)n}. \tag{6.86}$$

Maximum-likelihood estimation is reminiscent of maximum-likelihood estimation for the negative binomial distribution. Let the observed frequencies be f_x, $x = 0, 1, \ldots, n$, and set

$$A_x = f_{x+1} + f_{x+2} + \cdots + f_n, \qquad B_x = f_0 + f_1 + \cdots + f_x;$$

then the total number of observations is $A_{-1} = B_n$.

The maximum-likelihood equations are

$$\begin{aligned} 0 = F &\equiv \sum_{x=0}^{n-1} \frac{A_x}{\hat{\alpha} + x} - \sum_{x=0}^{n-1} \frac{A_{-1}}{\hat{\alpha} + \hat{\beta} + x}, \\ 0 = G &\equiv \sum_{x=0}^{n-1} \frac{B_x}{\hat{\beta} + x} - \sum_{x=0}^{n-1} \frac{A_{-1}}{\hat{\alpha} + \hat{\beta} + x}. \end{aligned} \tag{6.87}$$

Iteration is required for their solution. Given initial estimates α_1 and β_1 (e.g., the moment estimates), corresponding values of F_1 and G_1 can be computed; better estimates, α_2 and β_2, can then be obtained by solving the simultaneous linear equations

$$\begin{aligned} F_1 &= (\alpha_2 - \alpha_1) \sum_{x=0}^{n-1} \frac{A_x}{(\alpha_1 + x)^2} - (\alpha_2 - \alpha_1 + \beta_2 - \beta_1) \sum_{x=0}^{n-1} \frac{A_{-1}}{(\alpha_1 + \beta_1 + x)^2}, \\ G_1 &= (\beta_2 - \beta_1) \sum_{x=0}^{n-1} \frac{B_x}{(\beta_1 + x)^2} - (\alpha_2 - \alpha_1 + \beta_2 - \beta_1) \sum_{x=0}^{n-1} \frac{A_{-1}}{(\alpha_1 + \beta_1 + x)^2}. \end{aligned} \tag{6.88}$$

The next cycle is then begun by calculating F_2 and G_2. Kemp and Kemp (1956b, p. 174) reported that "on average about five cycles were needed to stabilize the estimates to three decimal places."

Chatfield and Goodhardt (1970) put forward an estimation method based on the mean and zero frequency. For highly J-shaped distributions (e.g., distributions of numbers of items purchased) this method has good efficiency; however, it does require iteration.

Griffiths (1973) remarked that the MLEs can be obtained by the use of a computer algorithm to maximize the log-likelihood. Williams (1975), like Griffiths, considered it advantageous to reparameterize, taking $\pi = \alpha/(\alpha + \beta)$ (the mean of the beta distribution) and $\theta = 1/(\alpha + \beta)$ (a shape parameter). Williams was hopeful that convergence of a likelihood maximization algorithm would be more rapid with this parameterization.

Qu et al. (1990) pointed out that $\alpha > 0$, $\beta > 0$, that is, $\theta > 0$, gives the beta–binomial distribution, while $\alpha < 0$, $\beta < 0$, that is, $\theta < 0$, gives the hypergeometric distribution. When $\alpha \to \infty$, $\beta \to \infty$, that is, $\theta \to 0$, the distribution tends to the binomial. Hence maximum-likelihood estimation with this parameterization enables an appropriate distribution from within this group to be fitted to data without assuming one particular distribution. Qu et al. also proposed methods of testing $H_0 : \theta = 0$ using (1) a Wald statistic and (2) the likelihood ratio. He showed how the homogeneity of the parameters can be tested using the deviance.

Bowman, Kastenbaum, and Shenton (1992) have shown that the joint efficiency for the method of moments estimation of α and β is very high over much of the parameter space. Series were derived for the first four moments of the moment estimators; simulation approaches were used for validation.

Various methods of estimation for the beta–binomial and zero-truncated beta–binomial distributions are explored in Tripathi, Gupta, and Gurland (1994). They are compared with maximum likelihood on the basis of asymptotic relative efficiency. Examples of their use and recommendations are provided.

Moment and maximum-likelihood estimation for the case where all three parameters α, β, and n are unknown was discussed in outline by Kemp and Kemp (1956b).

6.7.3 Beta–Pascal Distribution

Dubey (1966a) studied estimation for the beta–Pascal distribution assuming that k is known.

Given the beta–Pascal distribution with parameterization

$$\Pr[X = y] = \int_0^1 \binom{y-1}{k-1} (1-\lambda)^k \lambda^{y-k} \times \frac{\lambda^{\ell-1}(1-\lambda)^{m-1}\, d\lambda}{B(\ell, m)}, \qquad y = k, k+1, \ldots, \tag{6.89}$$

the mean and variance are

$$\mu = k + \frac{k\ell}{m-1} = \frac{k(\ell+m-1)}{m-1}, \quad \text{provided that } m > 1,$$
$$\mu_2 = \frac{k\ell(k+m-1)(\ell+m-1)}{(m-1)^2(m-2)}, \quad \text{provided that } m > 2 \tag{6.90}$$

(note that the support is $k, k+1, \ldots$). Hence the moment estimates are

$$\tilde{m} = 2 + \frac{\overline{x}(\overline{x}-k)(k+1)}{(s^2k - \overline{x}^2 + k\overline{x})}, \qquad \tilde{\ell} = \frac{(\tilde{m}-1)(\overline{x}-k)}{k}. \tag{6.91}$$

Dubey also discussed maximum-likelihood estimation (assuming that k is known). There are close parallels with maximum-likelihood estimation for the beta–binomial distribution. Irwin (1975b) described briefly maximum-likelihood estimation procedures for the three-parameter generalized Waring distribution (i.e., k unknown).

6.8 CHARACTERIZATIONS

There are several characterizations for hypergeometric-type distributions.

Patil and Seshadri's (1964) very general result for discrete distributions has the following corollaries:

1. Iff the conditional distribution of X given $X + Y$ is hypergeometric with parameters a and b, then X and Y have binomial distributions with parameters of the form (a, θ) and (b, θ), respectively.
2. Iff the conditional distribution of X given $X + Y$ is negative hypergeometric with parameters α and β for all values of $X + Y$, then X and Y have negative binomial distributions with parameters of the form (α, θ) and (β, θ), respectively.

Further details are in Kagan, Linnik, and Rao (1973).

Consider now a family of $N + 1$ distributions indexed by $j = 0, 1, \ldots, N$, each supported on a subset of $\{0, 1, \ldots, n\}, n \le N$. Skibinsky (1970) showed that this is the hypergeometric family with parameters N, n, j iff for each θ, $0 \le \theta \le 1$, the mixture of the family with binomial (N, θ) mixing distribution is the binomial (n, θ) distribution. Skibinsky restated this characterization as follows: Let $h_0, h_1, \ldots, h_N$ denote $N + 1$ functions on $\{0, 1, \ldots, n\}$; then

$$h_j(i) = \binom{j}{i}\binom{N-j}{n-i}\Big/\binom{N}{n}, \tag{6.92}$$

$i = 0, 1, \ldots, n$, $j = 0, 1, \ldots, N$, iff the h_i are independent of θ and

$$\sum_{j=0}^{N} h_j b(j; N, \theta) = b(\cdot; n, \theta), \qquad 0 \le \theta \le 1, \tag{6.93}$$