COMMUN. STATIST.-THEOR. METH., 15(3), 873-891 (1986)

ON SOME DISTRIBUTIONS ARISING FROM CERTAIN GENERALIZED SAMPLING SCHEMES.

John Panaretos

University of Patras
Patras, Greece
and
University of Crete, Greece

Evdokia Xekalaki

Department of Mathematics
University of Crete
Iraklio, Crete, Greece

Key Words and Phrases: distribution of order k; hypergeometric; negative hypergeometric; logarithmic series; generalized Waring distribution; binomial; Poisson; negative binomial; Polya and inverse Polya distribution.

1. INTRODUCTION

Most random phenomena appear to be analogous, in certain basic aspects, to a suitable random sampling scheme. Therefore, urn models have become very popular stochastic devices translating realistic problems into mathematical ones. Almost all of the basic discrete distributions of probability theory can be obtained from considerations of sampling-from-an-urn problems while others can be thought of as limiting cases of the former. Characteristic is the case of the binomial and the negative binomial distributions and that of the Poisson and the logarithmic series distributions. In all the cases a series of drawings is made from an urn filled with balls of two or more colours, a sampled ball of a particular colour being considered as a success. A comprehensive account of such urn models is given by Johnson and Kotz (1977).

Quite early, the notion of success was generalized to refer to a run of k balls of one particular colour where $k \geq 1$. Fréchet (1943), for example, examined the problem of determining the distribution of the number of runs of k like outcomes in a series of n Bernoulli trials as well as that of the number of trials needed to the first success run. Also Feller (1970) discusses the asymptotic behaviour of these two distributions. Since then, several authors have considered problems of this or similar nature leading to generalized forms of the basic probability distributions. The papers by Philippou (1983), Philippou et al. (1983), Hirano (1984), Panaretos and Xekalaki (1984), Xekalaki et al. (1984) and the references there in cover a substantial amount of work in this direction. The distributions obtained through models associated with runs of k like elements are usually called distributions of order k. So, in the above mentioned literature the binomial, the geometric, the negative binomial, the Poisson, the compound Poisson, the beta-geometric the beta-negative binomial and the logarithmic series distributions of order k have been defined. Of these, the binomial, geometric and negative binomial distributions of order k have arisen in sampling, straight or inverse, with replacement while the rest have been obtained through limiting or mixing operations.

In this paper an urn containing balls of two colous is again used, but sampling, straight or inverse, is done without replacement or with additional replacements. So, in sections 2,3 and 5 the hypergeometric, negative hypergeometric and generalized Waring distributions of order k are defined and studied. Means, variances, models leading to these distributions as well as limiting cases are examined.

The relationship of each of these distributions to distributions of order k already existing in the literature is also examined. Moreover, in section 4 a logarithmic series distribution of order k is introduced.

2. THE HYPERGEOMETRIC DISTRIBUTION OF ORDER k

Consider an urn that contains a white and b black balls. Assume that n balls are drawn, one at a time and without replacement. Let X denote the number of occurrences of a run of k white balls.

Theorem 2.1: The probability function (p.f.) of the random variable (r.v.) X defined as above is given by

$$P(X=x) = \sum_{m=0}^{k-1} \sum \binom{\Sigma x_i + x}{x_1,\ldots,x_k,x} \frac{b^{(\Sigma x_i)} a^{(n-\Sigma x_i)}}{(a+b)^{(n)}} \qquad x=0,1,2,\ldots,\left[\frac{n}{k}\right] \tag{2.1}$$

where the symbol $\sum$ denotes summation over all the non-negative values of $x_1,x_2,\ldots,x_k$ subject to the condition $x_1+2x_2+\ldots kx_k = n-m-kx$ and $a^{(r)}=a(a-1)\ldots(a-r+1)$, $r=0,1,\ldots$, ($a^{(0)}=1$ and $\Sigma ix_i=x_1+2x_2+\ldots+kx_k$).

Proof: Let W denote the outcome {white ball} and B the outcome {black ball} in a single trial. Then, a typical outcome of the event {X=x} can be represented by the arrangement

$$Z_1 Z_2 \ldots Z_{\Sigma x_i + x} \underbrace{WW\ldots W}_{m} \qquad (0 \le m \le k-1)$$

where x_1 of the Z's are of the form B x_2 of the Z's are of the form WB,...; x_k of the Z's are of the form $\underbrace{W\ldots W}_{k-1}$ B and x of the Z's are of the form $\underbrace{WW\ldots W}_{k}$ so that $\Sigma ix_i+kx+m=n$, $0\le m\le k-1$.

The number of all possible such outcomes is $\binom{\Sigma x_i + x}{x_1,x_2,\ldots,x_k,x}$. These are mutually exclusive and each of them has probability

$$P(Z_1 Z_2 \ldots Z_{\Sigma x_i + x} \underbrace{WW\ldots W}_{m}) = \frac{b^{(\Sigma x_i)} a^{(kx+m+\Sigma(i-1)x_i)}}{(a+b)^{(\Sigma ix_i+kx+m)}}$$

Therefore,

$$P(X=x) = \sum_{m=0}^{k-1} \sum \binom{\Sigma x_i + x}{x_1, \ldots, x_k, x} \frac{b^{(\Sigma x_i)} a^{(kx+m+\Sigma(i-1)x_i)}}{(a+b)^{(\Sigma i x_i + kx + m)}} \qquad x=0,1,\ldots,\left[\frac{n}{k}\right]$$

with $\sum$ as defined before. This expression is equivalent to (2.1) and hence the theorem has been established.

Note that, when k=1, (2.1) reduces to

$$P(X=x) = \binom{n}{x} \frac{b^{(n-x)} a^{(x)}}{(a+b)^{(n)}} = \binom{a}{x} \binom{b}{n-x} \Big/ \binom{a+b}{n}, \quad x=0,1,\ldots,n$$

which is the p.f. of the usual hypergeometric distribution with parameters a,b and n. Therefore, we can regard the distribution defined by (2.1) as a generalization of the hypergeometric distribution and consider the following definition.

Definition 2.1: A non-negative integer-valued r.v. X with support in $\{0,1,\ldots,\left[\frac{n}{k}\right]\}$, (n>0, k a positive integer) will be said to have the hypergeometric distribution of order k if its p.f. is given by (2.1).

Hirano (1984) defined the binomial distribution of order k with p.f.

$$P(N_{k,n,p}=x) = \sum_{m=0}^{k-1} \sum \binom{\Sigma x_i + x}{x_1, \ldots, x_k, x} p^n \left(\frac{q}{p}\right)^{\Sigma x_i}, \tag{2.2}$$

$$x=0,1,\ldots,\left[\frac{n}{k}\right], \quad 0<p<1, \ q=1-p.$$

in the context of an urn model similar in nature to the one considered in this section only each of the n sampled balls was returned to the urn before the next ball was drawn. The theorem that follows shows that the hypergeometric distribution of order k tends to the binomial distribution of order k if the numbers of black and white balls are increased while their proportion is kept constant.

Theorem 2.2: Consider a r.v. X whose p.f. is given by (2.1). Let $\lim_H$ stand for limit as $a\to+\infty$, $b\to+\infty$ so that $a/(a+b)\to p$. Then

$$\lim_H P(X=x) = P(N_{k,n,p}=x).$$

Proof:

$$\lim_{H} P(X=x)=\lim_{H}\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\frac{b^{(\Sigma x_i)}a^{(n-\Sigma x_i)}}{(a+b)^{(n)}}$$

$$=\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\lim_{H}\left\{\left(\frac{a}{a+b}\right)^{n-\Sigma x_i}\left(\frac{b}{a+b}\right)^{\Sigma x_i}\right\}$$

which leads to (2.2) and hence establishes the result. The case k=1 leads to the well known result concerning the binomial limiting case of the hypergeometric distribution.

3. THE NEGATIVE HYPERGEOMETRIC DISTRIBUTION OF ORDER k.

Consider again an urn containing a white and b black balls.

In the previous section the distribution of the number X of occurrences of a run of k white balls in a sample of n balls taken without replacement was derived. If sampling were done with replacement, the resulting distribution would, of course, be the binomial distribution of order k as defined by (2.2). Suppose now that each sampled ball is returned to the urn along with one additional ball of the same colour before the next ball is drawn and let X be the number of runs of k white balls in a sample of size n. As shown by the following theorem, the resulting distribution of X is a negative hypergeometric type of distribution.

Theorem 3.1: The p.f. of the r.v. X defined as above is given by

$$P(X=x)=\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\frac{b_{(\Sigma x_i)}\,a_{(n-\Sigma x_i)}}{(a+b)_{(n)}}\qquad x=0,1,2,\ldots,[\tfrac{n}{k}] \tag{3.1}$$

where the inner summation $\sum$ extends over all the non-negative values of $x_1,\ldots,x_k$ such that $\Sigma i x_i = n-m-kx$ and $a_{(r)}=a(a+1)\ldots(a+r-1)$, $r=1,2,\ldots,(a_{(0)}=1)$.

Proof: The event $\{X=x\}$ is the union of $\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}$ mutually exclusive outcomes each of which is represented by

$Z_1Z_2\ldots Z_{\Sigma x_i+x}\underbrace{WW\ldots W}_{m}$, $0\le m\le k-1$ where again x_i of the Z's are of the form $\underbrace{WW\ldots W}_{i-1}B$, $i=1,2,\ldots,k$ and x of Z's are of the form $\underbrace{WW\ldots W}_{k}$ so that $\Sigma ix_i+m+kx=n$. (The outcomes W and B are defined as in section 2). Each of these outcomes occurs with a probability given by

$$P(Z_1Z_2\ldots Z_{\Sigma x_i+x}\underbrace{WW\ldots W}_{m})=\frac{b_{(\Sigma x_i)}a_{(m+kx+\Sigma(i-1)x_i)}}{(a+b)_{(\Sigma ix_i+m+kx)}}$$

for fixed m, $0\le m\le k-1$. Hence,

$$P(X=x)=\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\frac{b_{(\Sigma x_i)}a_{(m+kx+\Sigma(i-1)x_i)}}{(a+b)_{(\Sigma ix_i+m+kx)}}\qquad x=0,1,\ldots,\left[\frac{n}{k}\right]$$

which by the definition of $\sum$, is equivalent to (3.1). Hence the result.

For k=1, (3.1) takes the form

$$P(X=x)=\binom{n}{x}\frac{b_{(n-x)}a_{(x)}}{(a+b)_{(n)}}=\binom{-a}{x}\binom{-b}{n-x}\Big/\binom{-a-b}{n},\quad x=0,1,\ldots,n.$$

which represents the p.f. of the negative hypergeometric distribution with parameters a,b and n. Therefore, (3.1) defines a generalized negative hypergeometric distribution.

<u>Definition 3.1</u>: A non-negative, integer-valued r.v. X with support in {0,1,...,n} is said to have the negative hypergeometric distribution of order k with parameters a,b and n if its p.f. is given by (3.1).

<u>Theorem 3.2</u>: Let X be a r.v. whose distribution, conditional on another r.v. p with support (0,1), is the binomial distribution of order k with parameters n and p and p.f. given by (2.2). Let the distribution of p be the beta with parameters a and b and probability density function (p.d.f.)

$$f(p)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,p^{a-1}\,(1-p)^{b-1},\quad 0<p<1,\ a>0,\ b>0 \qquad (3.2)$$

Then the distribution of X is the negative hypergeometric of order k with parameters a,b and n as defined by (3.1).

<u>Proof</u>: The proof is straightforward.

The result of this theorem establishes a relationship between the negative hypergeometric distribution of order k and the binomial distribution of order k similar to that existing for the case k=1, namely that the negative hypergeometric of order k is a mixture on p of the bimonial distribution of the same order when p has a beta distribution. This relationship is used in what follows to derive the probability generating function (p.g.f) and the moments of the negative hypergeometric distribution of order k.

Let X and p be defined as in Theorem 3.2 and let $G_X(s)$ denote the p.g.f. of X. Then

$$G_X(s) = \int_0^1 G_{X|p}(s).f(p)dp \tag{3.3}$$

where, $G_{X|p}(s)$ is the p.g.f. of the binomial distribution of order k given by

$$G_{X|p}(s)=1+(s-1)\sum_{i=1}^{[n/k]} s^{i-1} \sum_{j=ki}^{n} \sum_{\Sigma rx_r=j-ki} \binom{\Sigma x_i+i-1}{x_1,\dots,x_k,i-1} p^j\left(\frac{q}{p}\right)^{\Sigma x_i} \tag{3.4}$$

(see Hirano (1984)). Then, using (3.2), (3.3) and (3.4) we obtain

$$G_X(s)=1+(s-1)\sum_{i=1}^{[n/k]} s^{i-1} \sum_{j=ki}^{n} \sum_{\Sigma rx_r=j-ki} \binom{\Sigma x_i+i-1}{x_1,\dots,x_k,i-1} \frac{b^{(\Sigma x_i)}a^{(j-\Sigma x_i)}}{(a+b)_{(j)}}$$

Also, since

$$E(X^r|p)=\sum_{i=1}^{[n/k]} \{i^r-(i-1)^r\} \sum_{j=ki}^{n} \sum_{\Sigma rx_r=j-ki} \binom{\Sigma x_i+i-1}{x_1,\dots,x_k,i-1} p^j\left(\frac{q}{p}\right)^{\Sigma x_i} \qquad r=1,2,\dots$$

(see Hirano (1984)) we have for the moments of X

$$E(X^r)=\sum_{i=1}^{[n/k]} \{i^r-(i-1)^r\} \sum_{j=ki}^{n} \sum_{\Sigma rx_r=j-ki} \binom{\Sigma x_i+i-1}{x_1,\dots,x_k,i-1} \frac{b^{(\Sigma x_i)}a^{(j-\Sigma x_i)}}{(a+b)_{(j)}}$$

It will now be shown that the binomial distribution of order k is a limiting case of the negative hypergeometric distribution of order k.

Theorem 3.3: Let X be a r.v. whose p.f. is given by (3.1) and let $N_{k,n,p}$ be another r.v. having the binomial distribution of order k as defined by (2.2). Then, if $a\to+\infty$, $b\to+\infty$ so that $a/(a+b)\to p$, $0<p<1$, the distribution of X tends to the distribution of $N_{k,n,p}$.

Proof: Let $\lim_H$ denote limit as $a\to+\infty$, $b\to+\infty$ so that $a/(a+b)\to p$. Then

$$\lim_H P(X=x)=\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\lim_H\frac{b_{(\Sigma x_i)}a_{(n-\Sigma x_i)}}{(a+b)_{(n)}}$$

$$=\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\lim_H\left\{\left(\frac{b}{a+b}\right)^{\Sigma x_i}\left(\frac{a}{a+b}\right)^{n-\Sigma x_i}\right\}$$

$$=P(N_{k,n,p}=x),\quad x=0,1,\ldots,\left[\frac{n}{k}\right].$$

The above result indicates that, just like in the case k=1, sampling from an urn with an additional replacement becomes equivalent to sampling with replacement in the usual sense when the initial numbers of black and white balls are increased at a constant proportion p.

Note: If each ball that is drawn at random is replaced together with $c(c\geq1)$ balls of the same colour, the resulting p.f. of X will be

$$P(X=x)=\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\frac{b(b+c)\ldots(b+(\Sigma x_i-1)c)a(a+c)\ldots(a+(n-\Sigma x_i-1)c)}{(a+b)(a+b+c)\ldots(a+b+(n-1)c)}$$

$$x=0,1,\ldots\left[\frac{n}{k}\right]$$

i.e.

$$P(X=x)=\sum_{m=0}^{k-1}\sum\binom{\Sigma x_i+x}{x_1,\ldots,x_k,x}\frac{\left(\frac{b}{c}\right)_{(\Sigma x_i)}\left(\frac{a}{c}\right)_{(n-\Sigma x_i)}}{\left(\frac{a+b}{c}\right)_{(n)}}\quad x=0,1,\ldots,\left[\frac{n}{k}\right]\qquad(3.5)$$

The distribution given in (3.5) is a generalization of the Polya distribution (which is obtained for k=1) and hence we can give the following definition.

Definition 3.2: A non-negative, integer-valued r.v. X taking values in $\{0,1,\ldots,[n/k]\}$, where k is a positive integer, is said to have the Polya distribution of order k with parameters a,b,c and n if its p.f. is given by (3.5).

Note that the distribution in (3.5) reduces to the distribution in (3.1), (2.2) or (2.1) according as c equals 1,0 or -1.

4. A LOGARITHMIC SERIES DISTRIBUTION OF ORDER k.

In this section a logarithmic series distribution is defined as limiting form of a distribution arising in the context of an inverse sampling scheme.

Consider an urn containing balls numbered from 0 through to k, a balls bearing a zero and $b_i=b$ balls bearing number i, i=1,2,...,k. Suppose that successive drawings are made with replacement till m zeroes are obtained. Let X denote the sum of the numbers drawn before the m-th zero.

Theorem 4.1: Let X be a.r.v. defined as above and let $P_m(X=x)$ denote its p.f. Then

$$\lim_{m\to 0} P_m(X=x|X>0)=\left(\ln\left(\frac{a}{bk+a}\right)\right)\sum\frac{(\Sigma x_i-1)!}{x_1!\ldots x_k!}\left(\frac{b}{bk+a}\right)^{\Sigma x_i}$$

$$x=1,2,\ldots \qquad (4.1)$$

where $\sum$ denotes summation over all non-negative values of $x_1,x_2,\ldots,x_k$ so that $\Sigma ix_i=x$.

Proof: Let X_i denote the number of balls bearing number i, i=1,2,...,k drawn before the m-th zero-ball. Then

$P_m(X=x)=\sum P(X_1=x_1,\ldots,X_K=x_k)$ where $\sum$ extends over all non-negative values of $x_1,\ldots,x_k$ so that $\Sigma ix_i=x$ and

$$P(X_1=x_1,\ldots,X_k=x_k)=\binom{\Sigma x_i+m-1}{x_1,\ldots,x_k,m-1}\left(\frac{a}{bk+a}\right)^m\left(\frac{b}{bk+a}\right)^{\Sigma x_i}$$

$$x_i=0,1,\ldots\ ;\quad i=1,2,\ldots,k$$

i.e.,

$$P_m(X=x)=\sum\binom{\Sigma x_i+m-1}{x_1,\ldots,x_k,m-1}\left(\frac{a}{bk+a}\right)^m\left(\frac{b}{bk+a}\right)^{\Sigma x_i} \qquad x=0,1,2,\ldots \tag{4.2}$$

The probability distribution given by (4.2) was defined by Steyn (1956) and shown to arise as a gamma mixture of the Poisson distribution of order k by Philippou (1983) who refers to it as the compound Poisson distribution of order k.

Obviously,

$$P_m(X=x|X>0)=\sum\binom{\Sigma x_i+m-1}{x_1,\ldots,x_k,m-1}\left(\frac{b}{bk+a}\right)^{\Sigma x_i}\frac{a^m}{(bk+a)^m-a^m} \qquad x=1,2,\ldots$$

Therefore,

$$\lim_{m\to 0}P_m(X=x|X>0)=\sum\frac{1}{x_1!\ldots x_k!}\left(\frac{b}{bk+a}\right)^{\Sigma x_i}\lim_{m\to 0}\frac{a^m m_{(\Sigma x_i)}}{(bk+a)^m-a^m}.$$

But

$$\lim_{m\to 0}\frac{a^m m_{(\Sigma x_i)}}{(bk+a)^m-a^m}=\lim_{m\to 0}\frac{m_{(\Sigma x_i)}}{\left(\frac{bk+a}{a}\right)^m-1}=\lim_{m\to 0}\frac{(m+1)_{(\Sigma x_i-1)}}{\left(\frac{bk+a}{a}\right)^m\ln\frac{bk+a}{a}}=\frac{(\Sigma x_i-1)!}{-\ln(a/(bk+a))}$$

This implies that (4.3) is true and hence the theorem has been established.

For k=1, (4.1) reduces to the ordinary logarithmic series distribution with parameter b/a and p.f.

$$P(X=x)=\frac{1}{\ln(1+\frac{b}{a})}\,\frac{(b/a)^x}{x}, \quad x=1,2,\ldots$$

So we may give the following definition.

Definition 4.1: A positive, integer-valued r.v. X will be said to have the logarithmic series distribution of order k with parameter $\theta>0$ if its p.f. is given by

$$P(X=x)=\sum\frac{(\Sigma x_i-1)!}{\ln(\theta k+1)}\,\frac{(\theta/(\theta k+1))^{\Sigma x_i}}{x_1!\ldots x_k!} \qquad x=1,2,\ldots \tag{4.3}$$

Note that the logarithmic series distribution of order k defined by (4.3) differs from Aki et al.'s (1983) logarithmic series distribution of order k with p.g.f.

$$G(s) = \ln \frac{1-s+qp^k s^{k+1}}{1-ps} \Big/ k\ln p \tag{4.4}$$

This is natural since (4.3) is the limiting case of the negative binomial distribution of order k given by (4.2) which is different from that considered by Aki et al. (1983).

Let G(s) denote the p.g.f. of X defined as above. Then

$$G(s)=\sum_{x=1}^{\infty} s^x \sum \frac{(\Sigma x_i-1)!}{\ln(\theta k+1)} \frac{(\theta/(\theta k+1))^{\Sigma x_i}}{x_1!\ldots x_k!}$$

$$= \frac{1}{\ln(\theta k+1)} \sum_{x=0}^{\infty} \sum_{\Sigma i x_i = x+1} \binom{\Sigma x_i - 1}{x_1,\ldots,x_k} \prod_{i=1}^{k} \left(\frac{\theta s^i}{1+\theta k}\right)^{x_i}$$

Letting $x_i = r_i$, $i=1,2,\ldots,k$ and $x = r + \sum_{i=1}^{k} (i-1)x_i$ the above expression becomes

$$G(s) = \frac{1}{\ln(\theta k+1)} \sum_{r=1}^{\infty} \sum_{\Sigma r_i = r} \binom{r}{r_1,\ldots,r_k} \prod_{i=1}^{k} \left(\frac{\theta s^i}{1+\theta k}\right)^{r_i}$$

$$= \frac{1}{\ln(\theta k+1)} \sum_{r=1}^{\infty} \frac{1}{r} \left(\frac{\theta}{1+\theta k}\right)^r \sum_{\Sigma r_i = r} \binom{r}{r_1,\ldots,r_k} s^{r_1} s^{2r_2} \ldots s^{kr_k}$$

$$= \frac{1}{\ln(\theta k+1)} \sum_{r=1}^{\infty} \left(\frac{\theta}{\theta k+1}\right)^r \frac{\left(\sum_{i=1}^{k} s^i\right)^r}{r}$$

i.e.

$$G(s) = -\frac{\ln\left(1 - \frac{\theta}{1+\theta k} \sum_{i=1}^{k} s^i\right)}{\ln(1+\theta k)} \tag{4.5}$$

Then the moments of X can be derived using (4.3). For example we have for the mean

$$E(X) = \theta \sum_{i=1}^{k} i \Big/ \ln(1+\theta k) = \theta k(k+1) / (2\ln(1+\theta k)).$$

The form of (4.5) suggests the possibility of representing the logarithmic series distribution of order k as a random sum of discrete uniform r.v.'s as indicated by the following proposition.

Proposition 4.1: Let $Y_1, Y_2, \ldots$ be a sequence of independent, positive, integer-valued r.v.s distributed uniformly in $\{1,2,\ldots,k\}$. Let Z be another r.v. independent of $Y_1, Y_2, \ldots$ whose distribution is the logarithmic series distribution with parameter $\theta k/(1+\theta k)$. Then the distribution of the r.v. $X = Y_1 + Y_2 + \ldots + Y_Z$ is the logarithmic series distribution of order k as defined by (4.3) or (4.5).

The following theorem will now be shown.

Theorem 4.2: Let $Y_1, Y_2, \ldots$ be a sequence of independently and identically distributed r.v.'s with p.f.

$$P(Y=y) = \frac{(\theta k/(\theta k+1))^y}{y \ln(\theta k+1)}, \qquad y=1,2,\ldots$$

and let Z be another r.v. distributed independently of $Y_1, Y_2, \ldots$ according to a Poisson distribution with parameter $\lambda \ln(\theta k+1)$. Then, the distribution of the r.v. $X = Y_1 + Y_2 + \ldots + Y_Z$ is the negative binomial distribution of order k defined by (4.2) for $b/a = \theta$.

Proof: The result is an immediate consequence of the fact that

$$G_X(s) = G_Z(G_Y(s)) = (1+\theta(k-\sum_{i=1}^{k} s^i))^{-\lambda},$$

Note: The Poisson distribution of order k has a p.g.f. of the form

$$\exp\{\lambda(\sum_{i=1}^{k} s^i - k)\}, \quad \lambda>0$$

(Philippou et.al. (1983)) and can be regarded as the distribution of a Poisson sum of discrete uniform r.v.'s on $\{1,2,\ldots,k\}$ as noted by Xekalaki et al (1984)). Further, the negative binomial distribution of order k defined by (4.2) has a p.g.f. of the form $\{1+b(k-\sum_{i=1}^{k} s^i)/a\}^{-m}$ (Philippou,(1983))

which suggests that it can arise as a negative binomial distribution with parameters m and $p=\frac{a}{a+bk}$ generalized by a uniform distribution on $\{1,\ldots,k\}$, $i=1,2,\ldots$. Combining these facts with the results of Proposition 4.1 and Theorem 4.2, one is led to the following equivalent genesis schemes of the negative binomial distribution of order k as defined by (4.2):

Negative binomial $(m, p = \frac{a}{a+bk}) \vee$ uniform $\{1,\ldots,k\}$

Poisson $(\lambda) \wedge$ gamma $(m, \frac{a}{bk}) \vee$ uniform $\{1,\ldots,k\}$

Poisson of order k (λ) $\wedge$ gamma $(m,\frac{a}{b})$

Poisson (λk) $\vee$ uniform $\{1,\ldots,k\}$ $\wedge$ gamma $(m,\frac{a}{b})$

Poisson $(\lambda \ln(1+\theta k))$ $\vee$ log. series of order k (θ)

Poisson $(\lambda \ln(1+\theta))$ $\vee$ log. series $(\frac{\theta k}{1+\theta k})$ $\vee$ uniform $\{1,\ldots,k\}$

(Here $\wedge$ denotes mixing (compounding) and $\vee$ denotes generalization)

5. THE GENERALIZED WARING DISTRIBUTION OF ORDER k

Let us now consider sampling from an urn filled with a white and b black balls according to the following scheme. A ball is drawn at random, its colour is noted and the ball is replaced along with one additional ball of the same colour before the next ball is drawn. Let W and B denote the outcomes {white ball}, and {black ball} in one drawing respectively, and define S_k to be the event $\{\underbrace{WW\ldots W}_{k}\}$ in k consecutive drawings. Let Y be the number of drawings to the m-th occurrence of S_k.

Theorem 5.1: Let Y be a r.v. defined as above. Then its probability distribution is given by

$$P(Y=y) = \sum \binom{\Sigma r_i+m-1}{r_1,\ldots,r_k,m-1} \frac{b_{(\Sigma r_i)}\, a_{(y-\Sigma r_i)}}{(a+b)_{(y)}} \qquad y=km,km+1,\ldots \tag{5.1}$$

where $\sum$ denotes summation over all non-negative values of $r_1,\ldots,r_k$ subject to $\Sigma ir_i+km = y$.

Proof: The event $\{Y=y\}$ occurs if any one of the $\binom{\Sigma r_i+m-1}{r_1,\ldots,r_k,m-1}$ mutually exclusive events of the form $Z_1Z_2\ldots Z_{\Sigma r_i+m-1}S_k$ occurs, where r_i of the Z's are of the form $\underbrace{WW\ldots W}_{i-1}B$, $i=1,2,\ldots,k$ and m-1 of the Z's are S_k's $(m\geq 1)$ so that $\Sigma ir_i+km=y$. The probability of any such event is

$$P(Z_1Z_2\ldots Z_{\Sigma r_i+m-1}S_k) = \frac{b_{(\Sigma r_i)}\, a_{(\Sigma(i-1)r_i+km)}}{(a+b)_{(\Sigma ir_i+km)}} \qquad \begin{array}{l} r_i=0,1,\ldots \\ i=1,2,\ldots k. \end{array}$$

Therefore, the probability $p_{r_1,\ldots,r_k}$ of observing r_1 B's, r_2 WB's,...,r_k $\underbrace{WW...W}_{k-1}B$'s and m S_k's is given by

$$p_{r_1,\ldots,r_k} = \binom{\Sigma r_i+m-1}{r_1,\ldots,r_k,m-1} \frac{b_{(\Sigma r_i)}\, a_{(\Sigma(i-1)r_i+km)}}{(a+b)_{(\Sigma i r_i+km)}} \qquad \begin{array}{l} r_i=0,1,\ldots \\ i=1,2,\ldots,k \end{array}$$

Then the event {Y=y}, y=km,km+1,... occurs with a probability equal to the sum of all possible values of $p_{r_1,\ldots,r_k}$ of the above relationship for all the non-negative values of $r_1,\ldots,r_k$ satisfying the condition $\Sigma i r_i+km=y$. Hence the result.

Consider now the r.v. X=Y-km. Obviously,

$$P(X=x) = \sum \binom{\Sigma r_i+m-1}{r_1,\ldots,r_k,m-1} \frac{b_{(\Sigma r_i)}\, a_{(x+km-\Sigma r_i)}}{(a+b)_{(x+km)}} \qquad x=0,1,2,\ldots \tag{5.2}$$

where $\sum$ denotes summation over all the non-negative values of $r_1,r_2,\ldots,r_k$ such that $\Sigma i r_i=x$. For k=1, (5.2) represents the distribution of the number of black balls drawn before the mth white ball which is known in the literature as the generalized Waring distribution (Irwin (1963), Xekalaki (1981)), i.e.

$$P(X=x) = \frac{a_{(m)}}{(a+b)_{(m)}} \frac{b_{(x)} m_{(x)}}{(a+b+m)_{(x)}} \frac{1}{x!}, \quad x=0,1,2,\ldots \tag{5.3}$$

(For information concerning the structure and applications of this distribution see Irwin (1975), Xekalaki (1981), (1983a,b,c), (1984a,b) and Xekalaki and Panaretos (1983)).

Hence (5.2) provides a generalization of (5.3) in the context of distributions of order k and therefore it leads to the following definition.

Definition 5.1: A non-negative integer-valued r.v. X will be said to have the generalized Waring distribution of order k with parameters a,b and m if its p.f. is given by (5.2).

Theorem 5.2: Let X be a non-negative, integer-valued r.v. and let p be a continuous r.v. with support in (0,1). Assume that conditional on p the distribution of X is the negative binomial distribution of order k with p.f.

$$P(X=x|p) = \sum \binom{\Sigma x_i+m-1}{x_1,\ldots,x_k,m-1} p^{x+km} (q/p)^{\Sigma x_i} \quad \begin{array}{l} x=0,1,2,\ldots \\ q=1-p \\ 0<p<1 \end{array} \tag{5.4}$$

Here $\sum$ extends over all the non-negative values of $x_1,\ldots,x_k$ for which $\Sigma ix_i=x$. Let the distribution of p be the beta of type I with p.d.f. given by (3.2). Then, the unconditional distribution of X is the generalized Waring distribution of order k with parameters a,b and m as defined by (5.2).

Proof: The proof is straightforward.

The result of this theorem is a generalization of the chance mechanism that gives rise to the ordinary generalized Waring distribution as a beta mixture on p of the ordinary negative binomial distribution as defined by (5.4) for k=1 (see e.g. Xekalaki (1981)).

The relationship between the negative binomial and generalized Waring distributions of order k can be used to derive the mean and variance of the latter as shown by the following corollary.

Corollary 5.1: Let X be a r.v. whose p.f. is given by (5.2).Then

$$E(X) = m\left\{\frac{(a+b-k-1)_{(k+1)}}{(a-k)_{(k)}(b-1)} - \frac{a+b-1}{b-1} - k\right\} \tag{5.5}$$

$$V(X)= \frac{m}{b-1}\left\{\frac{(a+b-2k-2)_{(2k+2)}}{(b-2)\ (a-2k)_{(2k)}} - \frac{(2k+1)(a+b-k-1)_{(k+1)}}{(a-k)_{(k)}} - \frac{a(a+b-1)}{b-2}\right\} \tag{5.6}$$

Proof: From Theorem 5.2 it follows that the distribution of X can be regarded as a mixture on p of the distribution of X|p when X|p has a negative binomial distribution of order k with parameters m and p as defined by (5.4) and p has a beta distribution of the first type with p.d.f. given by (3.2). Therefore,

$E(X)=E(E(X|p))=E(m(1-p^k)/qp^k)$ and

$V(X)=E(V(X|p))=m\ E\left(1-(2k+1)qp^k-p^{2k+1}\right)/q^2p^{2k}$

which leads to (5.5) and (5.6).

For k=1 (5.5) and (5.6) reduce to the mean and variance of the ordinary genaralized Waring distribution.

Theorem 5.3: Let X be a r.v. having the Poisson distribution with parameter $-mk\ln p; m>0, k>0, 0<p<1$ and let $Y_1, Y_2, \ldots$ be a sequence of mutually independent r.v.'s that are identically distributed independently of X according to a logarithmic series distribution of order k with parameter p, $0<p<1$ and p.g.f. given by (4.4). Assume that p has a beta distribution of the first kind with p.d.f. given by (3.2). Then, the r.v. $Z=Y_1+Y_2+\ldots+Y_X$ has the generalized Waring distribution of order k with parameters a, b and m.

Proof: The result follows immediately by Theorem 5.1 if one notes that the r.v. $Z|p$ has the negative binomial distribution of order k with parameters m and p and p.f. given by (5.4)

The case k=1 leads to Xekalaki's (1981) derivation of the ordinary generalized Waring distribution as a beta mixture on p of the Poisson ($-m\ln p$) V logarithmic series (p) distribution.

The following theorem can be easily shown.

Theorem 5.4: Let X be a non-negative, integer-valued r.v. whose distribution is the generalized Waring of order k with parameters a, b and m and p.f. given by (5.2). Then as $a\to+\infty$, $b\to+\infty$ so that $b/(a+b)<+\infty$ the probability distribution of X tends to that of the negative binomial distribution of order k with parameters m and $\frac{b}{a+b}$.

So, the urn scheme considered in this section can give rise to the negative binomial distribution of order k if the initial numbers of black and white balls are increased at a fixed proportion.

Theorem 5.5: Let X be defined as in Theorem 5.4. Then if $a\to+\infty$, $b\to+\infty$ $m\to+\infty$ so that $b/(a+b)\to 0$ and $mb/(a+b)<+\infty$, the probability distribution of X tends to that of the Poisson distribution of order k with parameter $mb/(a+b)$.

Proof: Let lim denote limit as $a\to+\infty$, $b\to+\infty$, $m\to+\infty$ so that $b/(a+b)\to 0$ and $mb/(a+b)\to+\infty$ (H′). Then, from (5.2)

$$\lim_{H'} P(X=x) = \sum \lim_{H'} \frac{m^{(\Sigma x_i)} b^{(\Sigma x_i)} a^{(x+km-\Sigma x_i)}}{(a+b)_{(x+km)} x_1!x_2!\ldots x_k!}$$

$$= \lim_{H'}\left(\frac{a}{a+b}\right)^{km} \sum \left(\frac{bm}{a+b}\right)^{\Sigma x_i} \lim_{H'}\left(\frac{a}{a+b}\right)^{x-\Sigma x_i} \Big/ \prod_{i=1}^{k} x_i! \, .$$

$$\lim_{H'}\left(\frac{a}{a+b}\right)^{km} = \lim_{H'} e^{km\ln\frac{a}{a+b}} \qquad -km\frac{b}{a} \le km\ln\frac{a}{a+b} \le -km\frac{b}{a+b}$$

This implies that

$$-\lim_{H'} km\frac{b}{a+b} \lim_{H'}\frac{a+b}{a} \le \lim_{H'} km\ln\frac{a}{a+b} \le \lim_{H'} km\frac{b}{a+b}$$

i.e. $$\lim_{H'} km\ln\frac{a}{a+b} = -k\frac{mb}{a+b}$$

Therefore

$$\lim_{H'} P(X=x) = e^{-kmb/(a+b)} \sum \left(\frac{bm}{a+b}\right)^{\Sigma x_i} \Big/ \prod_{i=1}^{k} x_i!$$

which shows that under H′ the distribution of X tends to the Poisson distribution of order k with parameter $mb/(a+b)$.

The sampling scheme, considered in this section can be slightly modified so as to give rise to a more general form of distribution. In fact, if each sampled ball is returned to the urn along with c ($c\ge 1$) balls of the same colour, the p.f. of the r.v. X where X+km is the number of drawings to the mth occurence of a run of k white balls is given by

$$P(X=x) = \sum \binom{\Sigma x_i+m-1}{x_1,\ldots,x_k,m-1} \frac{(b/c)^{(\Sigma x_i)}(a/c)^{(x+km-\Sigma x_i)}}{((a+b)/c)^{(x+km)}} \qquad (5.7)$$

$$x=0,1,2,\ldots$$

Here $\sum$ denotes again summation subject to the constraint $\Sigma i x_i = x$.

For k=1, (5.7) represents the p.f. of the inverse Polya distribution. Hence the following definition can be given.

Definition 5.2: A non-negative, integer-valued r.v. X taking values in {0,1,2,...} is said to have the inverse Polya distribution of order k with parameters a,b,m and c if its p.f. is given by (5.7).

BIBLIOGRAPHY

Feller, W. (1970). An introduction to probability theory and its applications. Vol. 1 (Revised printing of the third edition). Willey, New York.

Fréchet, M. (1943). Les probabilités associeés à un système d' événements compatibles et dépendants. 2. Cas particulieurs et applications. Actualités Scientifiques et Industrielles, no 942, Hermann:Paris.

Hirano, K. (1984). Some properties of the distributions of order k. Research memorandum no. 280, the Institute of Statistical Mathematics. Tokyo, Japan.

Irwin, J.O. (1963). The place of mathematics in medical and biological statistics. J.Roy.Stat.Soc., A 126, 1-44.

Irwin, J.O. (1975). The generalized Waring distribution. J.Roy. Stat.Soc., A 138, 18-31 (Part I), 204-227 (Part II), 374-384 (Part III).

Johnson, N.L. and Kotz, S. (1977). Urn models and their application. Wiley, New York.

Panaretos, J. and Xekalaki, E. (1984). On generalized binomial distributions and their relation to generalized Poisson distributions. Ann.Inst.Statist.Math., Part A Vol.38, (1986) (to appear).

Philippou, A.N. (1983). The Poisson and compound Poisson distributions of order k and some of their properties. Zap. Nauchn.Sem. Leningrad Otdel.Mat. Inst.-Steklov (LOMI), 130, 175-180 (in Russian)

Philippou, A.N., Georghiou, and Philippou, G.N. (1983). A generalized geometric distribution and some of its properties. Statistics and Probability Letters, 1, 171-175.

Steyn, H.S. (1956). On the univariable series $F(t) \equiv F(a;b_1b_2...,b_\kappa;c;t,t^2,...,t^\kappa)$ and its applications in probability theory. Proc.Kon.Ned.Akad.V.Watensch, Ser.A, 59, 190-197.

Xekalaki, E. (1981). Chance mechanisms for the univariate generalized Waring distribution and related characterizations. Statistical Distributions in Sientific Work, 4, (Models, Structures and Characterizations eds. C. Taillie, G.P. Patil and B. Baldessari), D. Reidel, Holland, 157-171.

Xekalaki, E. (1983a). The univariate generalized Waring distribution in relation to accident theory:Proneness, spells or contagion? Biometrics, 39, 887-895.

Xekalaki, E. (1983b). A property of the Yule distribution and its application. Commun.Statist.(Theory and Methods), 12,1181-1189.

Xekalaki, E. (1983c). Hazard functions and life distributions in discrete time.Commun.Statist.(Theory and Methods),12, 2503-2509.

Xekalaki, E. (1984a). The bivariate generalized Waring distribution and its application to accident theory. J.Roy. Statist. Soc. Series A, 147(3). 488-498.

Xekalaki, E. (1984b). Linear regression and the Yule distribution. J. of Econometrics, 24, 397-403.

Xekalaki, E. and Panaretos, J. (1983). Identifiability of compound Poisson distributions.Scand. Actuarial J., 39-45.

Xekalaki, E., Panaretos, J. and Philippou, A. (1984). On some mixtures of distributions of order k. (Submitted).

Received May, 1985; Revised December, 1985.

Refereed by S. A. Patil, Tennessee Technological University, Cookeville, TN.