CONVOLUTION OF INDEPENDENT LEFT-TRUNCATED NEGATIVE BINOMIAL VARIABLES AND LIMITING DISTRIBUTIONS

J. C. AHUJA AND E. A. ENNEKING

(Received Jan. 13, 1972)

1. Introduction

Let X_i $(i=1, 2, \cdots, n)$ be n independent and identically distributed random variables having the left-truncated negative binomial distribution

$$(1) \qquad f_c(x; k, \theta)=\binom{x+k-1}{x}\theta^x/g_c(k, \theta) \qquad x \in T$$

where $0<\theta<1$, $k>0$, $g_c(k, \theta)=(1-\theta)^{-k}-\sum_{x=0}^{c}\binom{x+k-1}{x}\theta^k$, and $T=\{c+1, c+2, \cdots\}$. Define their sum as $Z=\sum_{i=1}^{n} X_i$. The distribution of Z for $c=0$ has been recently obtained by the author [1] which we call the associated Lah distribution. This paper derives the distribution of Z for the general case in terms of the generalized Lah numbers which we name the generalized Lah distribution. A recurrence relation is provided for the generalized Lah numbers and is utilized to obtain a recurrence formula for the probability function (pf) of Z. The distribution function (df) of Z is found in an explicit form in terms of a linear combination of the incomplete beta functions. It is shown further that, under certain limiting conditions, the generalized Lah distribution approaches the generalized Stirling distribution of the first kind and the generalized Stirling distribution of the second kind obtained by the author [2] and Tate and Goen [6] respectively.

2. Definitions, notation and preliminary results

The Lah numbers $L(z, n)$ (see Riordan [5], p. 44) with arguments z and n are given by the relation

$$(2) \qquad L(z, n)=(-1)^n(z!/n!)\binom{z-1}{n-1}$$

where $L(z, n)=0$ for $n>z$. Since the sign of $L(z, n)$ is the same as that of $(-1)^n$, we may write (2) in absolute value as

$$(3) \qquad |L(z, n)| = (z!/n!)\binom{z-1}{n-1} .$$

We introduce the associated Lah numbers $B_k(z, n)$ and the generalized Lah numbers $L_{c,k}(z, n)$ by

$$(4) \qquad B_k(z, n) = (z!/n!) \sum_{r=1}^{n} (-1)^{n-r}\binom{n}{r}\binom{z+kr-1}{z}$$

where $B_k(z, n)=0$ for $n>z$, and

$$(5) \qquad L_{c,k}(z, n) = (z!/n!) \sum (-1)^{n-r_1} \frac{n!}{r_1!r_2!\cdots r_{c+2}!} \times \prod_{j=0}^{c} \left[\binom{j+k-1}{j}\right]^{r_{j+2}} \binom{z-\sum_{j=0}^{c} jr_{j+2}+kr_1-1}{kr_1-1}$$

for integral $c \geqq 0$, and $z=n(c+1), n(c+1)+1, \cdots$, where the summation extends over all $r_j \geqq 0$ such that $\sum_{j=1}^{c+2} r_j = n$. As a consequence of these definitions, writing $L_{0,k}(z, n)$ as an iterated sum over r_1 and r_2, we have the following:

Property 1. $L_{0,k}(z, n) = B_k(z, n)$.

Property 2. $L_{0,1}(z, n) = B_1(z, n) = |L(z, n)|$.

DEFINITION 1. A random variable Z is said to have the associated Lah distribution (ALD) with parameters n, k, and θ if its pf is given by

$$(6) \qquad p(z; n, k, \theta) = n!B_k(z, n)\theta^z/z![g(k, \theta)]^n$$

for $z=n, n+1, \cdots$, where $0<\theta<1$, and $g(k, \theta)=(1-\theta)^{-k}-1$.

DEFINITION 2. A random variable Z is said to have the generalized Lah distribution (GLD) with parameters c, n, k, and θ if its pf is given by

$$(7) \qquad p_c(z; n, k, \theta) = n!L_{c,k}(z, n)\theta^z/z![g_c(k, \theta)]^n$$

for $z=n(c+1), n(c+1)+1, \cdots$, where $0<\theta<1$ and $L_{c,k}(z, n)$ and $g_c(k, \theta)$ are as defined above. Note that the ALD is a special case of the GLD for $c=0$.

Recently, the author [2] has obtained the exact distribution of the sum of n independent random variables having the logarithmic series distribution truncated on the left at 'c' called the generalized Stirling distribution of the first kind, while Tate and Goen [6] have provided the n-fold convolution of the Poisson distribution truncated on the left at 'c' which we call the generalized Stirling distribution of the second kind. These generalized Stirling distributions are of the following form:

DEFINITION 3. The generalized Stirling distribution of the first kind (GSDFK) with parameters c, n, and θ has the pf

$$u_c(z;n,\theta)=n!d_c(z,n)\theta^z/z![g_c(\theta)]^n \tag{8}$$

for $z=n(c+1), n(c+1)+1,\cdots$, where $0<\theta<1$, $g_c(\theta)=\sum_{x=c+1}^{\infty}\theta^x/x$, and $d_c(z,n)$ are the generalized Stirling numbers of the first kind defined in [2].

DEFINITION 4. The generalized Stirling distribution of the second kind (GSDSK) with parameters c, n, and θ has the pf

$$v_c(z;n,\theta)=n!D_c(z,n)\theta^z/z![h_c(\theta)]^n \tag{9}$$

for $z=n(c+1), n(c+1)+1,\cdots$, where $0<\theta<\infty$, $h_c(\theta)=\sum_{x=c+1}^{\infty}\theta^x/x!$, and $D_c(z,n)$ are the generalized Stirling numbers of the second kind defined in [6].

We now extend the definition of the generalized power series distribution with two parameters given in the dictionary of discrete distributions by Patil and Joshi [3].

DEFINITION 5. A random variable Z is said to have the generalized power series distribution (GPSD) with $m+1$ parameters $c_1, c_2, \cdots, c_m$, and θ if its pf is given by

$$q(z;c_1,c_2,\cdots,c_m,\theta)=\frac{a_z(c_1,c_2,\cdots,c_m)\theta^z}{g(c_1,c_2,\cdots,c_m,\theta)} \qquad z\in S \tag{10}$$

where $a_z(c_1,c_2,\cdots,c_m)>0$, $\theta\geqq 0$, S is a countable subset of the set of real numbers without any limit point, and the series function $g(c_1, c_2, \cdots, c_m, \theta)=\sum a_z(c_1,c_2,\cdots,c_m)\theta^z$, the summation extending over S, is positive, finite and differentiable for all admissible values of the parameters.

It may be noted that the ALD, the GSDFK, and the GSDSK are all special cases of the GPSD with three parameters, while the GLD is a special case of the GPSD with four parameters.

3. Derivation of the GLD

The characteristic function of the sum Z of n independent random variables having the distribution (1) is obtainable as

$$\phi_z(t)=[g_c(k,\theta)]^{-n}\left[(1-\theta e^{it})^{-k}-\sum_{x=0}^{c}\binom{x+k-1}{x}(\theta e^{it})^x\right]^n. \tag{11}$$

Taking the multinomial expansion of the second factor on the right side of (11) and employing the inversion formula for characteristic functions, we obtain the pf of Z in the form

$$(12)\qquad p_c(z; n, k, \theta)=[g_c(k, \theta)]^{-n}\sum(-1)^{n-r_1}\frac{n!}{r_1!r_2!\cdots r_{c+2}!}\theta^{\sum_{j=0}^{c} jr_{j+2}} \times\prod_{j=0}^{c}\left[\binom{j+k-1}{j}\right]^{r_{j+2}}(1/2\pi)\int_{-\pi}^{\pi}(1-\theta e^{it})^{-kr_1} \times\exp\left[-it\left(z-\sum_{j=0}^{c} jr_{j+2}\right)\right]dt$$

where the summation extends over $r_j\geqq 0$ such that $\sum_{j=1}^{c+2} r_j=n$. But we observe that

$$(13)\qquad (1/2\pi)\int_{-\pi}^{\pi}(1-\theta e^{it})^{-kr_1}\exp\left[-it\left(z-\sum_{j=0}^{c} jr_{j+2}\right)\right]dt =\binom{z-\sum_{j=0}^{c} jr_{j+2}+kr_1-1}{kr_1-1}\theta^{z-\sum_{j=0}^{c} jr_{j+2}}\ .$$

Therefore, the equation (12) together with (13) and the definition of the generalized Lah numbers $L_{c,k}(z, n)$ provide us the following:

THEOREM 1. *Let X_i $(i=1, 2, \cdots, n)$ be n independent and identically distributed random variables having the left-truncated negative binomial distribution* (1), *and let Z denote their sum. Then Z has the GLD given by* (7).

4. Recurrence for $L_{c,k}(z, n)$ and some properties of the GLD

It is easy to see that the GLD (7) which is a special case of the GPSD (10) has the series function

$$(14)\qquad [g_c(k, \theta)]^n=\sum_{z=n(c+1)}^{\infty} n!L_{c,k}(z, n)\theta^z/z!\ .$$

Differentiating both sides of (14) with respect to θ, then multiplying both sides by $(1-\theta)$ and using the fact that

$$r\binom{k+r-1}{r}-(r-1)\binom{k+r-2}{r-1}=k\binom{k+r-2}{r-1}$$

for $r=1, 2, \cdots, c$, we obtain

$$(15)\qquad n[g_c(k, \theta)]^{n-1}\left[kg_c(k, \theta)+(k+c)\binom{k+c-1}{c}\theta^c\right] =(1-\theta)\sum n!L_{c,k}(z, n)\theta^{z-1}/(z-1)!$$

which, using (14), becomes

$$nk \sum n!L_{c,k}(z, n)\theta^z/z! + n(k+c)\binom{k+c-1}{c}\theta^c \times \sum (n-1)!L_{c,k}(z, n-1)\theta^z/z! = (1-\theta) \sum n!L_{c,k}(z, n)\theta^{z-1}/(z-1)! \ . \tag{16}$$

Equating the coefficients of θ^z in (16), we arrive at the following:

Property 3. The generalized Lah numbers $L_{c,k}(z, n)$ satisfy the recurrence formula

$$L_{c,k}(z+1, n) = (z+nk)L_{c,k}(z, n) + n(k+c)\binom{k+c-1}{c} \times [z!/(z-c)!]L_{c,k}(z-c, n-1) \ . \tag{17}$$

It can now be verified that the recurrence relation (17), after some algebra, enables us to obtain the following:

Property 4. The pf $p_c(z; n, k, \theta)$ satisfies the recurrence formula

$$p_c(z+1; n, k, \theta) = [\theta/(z+1)]\Big[(z+nk)p_c(z; n, k, \theta) + n(k+c) \times \binom{k+c-1}{c}\theta^c(g_c(k, \theta))^{-1}p_c(z-c; n-1, k, \theta)\Big] \ . \tag{18}$$

Property 5. The mean and variance of the pf $p_c(z; n, k, \theta)$ are given by

$$\begin{aligned} &\mathrm{E}\,(Z) = n\theta(k+M)/(1-\theta) \\ &\mathrm{V}\,(Z) = n\theta[k+M(c+1)-\theta M(k+c+M)]/(1-\theta)^2 \end{aligned} \tag{19}$$

where

$$M = (k+c)\binom{k+c-1}{c}\theta^c/g_c(k, \theta) \ .$$

Property 6. The df $F_c(z; n, k, \theta)$ of the pf $p_c(z; n, k, \theta)$ is obtainable as

$$\begin{aligned} F_c(z; n, k, \theta) &= 1 - \sum_{x=z+1}^{\infty} p_c(x; n, k, \theta) \\ &= 1 - [g_c(k, \theta)]^{-n} \sum (-1)^{n-r} \frac{n!}{r_1!r_2!\cdots r_{c+2}!} \\ &\quad \times \prod_{j=0}^{c} \left[\binom{j+k-1}{j}\right]^{r_{j+2}} \theta^{\sum_{j=0}^{c} jr_{j+2}} (1-\theta)^{-kr_1} \\ &\quad \times I_\theta\Big(z - \sum_{j=0}^{c} jr_{j+2} + 1, kr_1\Big) \end{aligned} \tag{20}$$

where the summation extends over $r_j \geqq 0$ such that $\sum_{j=0}^{c+2} r_j = n$, and $I_\theta(m, n)$ is the incomplete beta function tabulated by Pearson [4].

It may be remarked that the df of the ALD obtained by the author [1] follows from (20) for $c=0$.

5. Limiting distributions

Let us now consider the limit of the characteristic function of the GLD given by (11) when $k \rightarrow 0$. We find that

$$\lim_{k\rightarrow 0} \phi_z(t) = \lim_{k\rightarrow 0} [g_c(k, \theta)]^{-n}\left[(1-\theta e^{it})^{-k} - \sum_{x=0}^{c} \binom{x+k-1}{k} (\theta e^{it})^x\right]^n$$

$$= [g_c(\theta)]^{-n}\left[-\log(1-\theta e^{it}) - \sum_{x=0}^{c} (\theta e^{it})^x / x\right]^n$$

which, in fact, is the characteristic function of the ASDFK with parameters c, n, and θ.

Similarly, it can be shown that, when $k \rightarrow \infty$, and $\theta \rightarrow 0$ such that $k\theta \rightarrow \lambda$, $0 < \lambda < \infty$, the characteristic function of the GLD approaches that of the ASDSK with parameters c, n, and λ. Using the uniqueness and continuity theorems for characteristic functions, we have proved the following:

THEOREM 2. *As $k \rightarrow 0$, the GLD with parameters c, n, k, and θ tends to the ASDFK with parameters c, n, and θ.*

THEOREM 3. *As $k \rightarrow \infty$ and $\theta \rightarrow 0$ such that $k\theta \rightarrow \lambda$, $0 < \lambda < \infty$, the GLD with parameters c, n, k, and θ tends to the ASDSK with parameters c, n, and λ.*

PORTLAND STATE UNIVERSITY

REFERENCES

[1] Ahuja, J. C. (1971). Distribution of the sum of independent decapitated negative binomial variables, *Ann. Math. Statist.*, **42**, 383-384.

[2] Ahuja, J. C. (1971). Exact distribution of the sum of independent left-truncated logarithmic series variables, *Ann. Math. Statist.*, **42**, 846 (Abstract).

[3] Patil, G. P. and Joshi, S. W. (1968). *A Dictionary and Bibliography of Discrete Distributions*, Oliver and Boyd, Edinburgh.

[4] Pearson, K. (1934). *Tables of Incomplete Beta Function*, Cambridge Univ. Press, London.

[5] Riordan, John (1958). *An Introduction to Combinatorial Analysis*, Wiley, New York.

[6] Tate, R. F. and Goen, R. L. (1958). Minimum variance unbiased estimation for the truncated Poisson distribution, *Ann. Math. Statist.*, **29**, 755-765.