BULLETIN 13 (1), 1970

ON THE DISTRIBUTION OF SUM OF INDEPENDENT POSITIVE BINOMIAL VARIABLES

BY
J. C. AHUJA

1. **Introduction.** Let $X_1, X_2, \ldots, X_n$ be n independent and identically distributed random variables having the positive binomial probability function

$$f(x;p) = \binom{N}{x} p^x(1-p)^{N-x}/1-(1-p)^N, \quad x \in T \tag{1}$$

where $0<p<1$, and $T=\{1, 2, \ldots, N\}$. Define their sum as $Y=X_1+X_2+\cdots+X_n$. The distribution of the random variable Y has been obtained by Malik [2] using the inversion formula for characteristic functions. It appears that his result needs some correction. The purpose of this note is to give an alternative derivation of the distribution of Y by applying one of the results, established by Patil [3], for the generalized power series distribution. The distribution function of Y is also found in an explicit form in terms of a linear combination of the incomplete beta functions.

2. **Distribution of sum.** If we take $p=\theta/1+\theta$, the probability function (1) may be written as

$$f(x;\theta) = \binom{N}{x} \theta^x/g(\theta), \quad x \in T \tag{2}$$

where $g(\theta)=(1+\theta)^N-1$. The binomial expansion of $[g(\theta)]^n$ in powers of θ gives

$$\begin{aligned} [g(\theta)]^n &= [(1+\theta)^N-1]^n \\ &= \sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r} (1+\theta)^{Nr} \\ &= \sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r} \sum_{y=0}^{Nr} \binom{Nr}{y} \theta^y \end{aligned}$$

which, after changing the order of summation, becomes

$$[g(\theta)]^n = \sum_{y=0}^{Nn} \left[\sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r}\binom{Nr}{y}\right] \theta^y \tag{3}$$

where the terms in the second summation are zero for $r<y/N$. Using the binomial coefficient identity (12.17) given by Feller [1, p. 65], it can be easily verified that

$$\sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r}\binom{Nr}{y} = 0$$

https://doi.org/10.4153/CMB-1970-035-7 Published online by Cambridge University Press

for $y=0, 1, \ldots, n-1$, so that (3) reduces to

$$[g(\theta)]^n = \sum_{y=n}^{Nn} \left[\sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r}\binom{Nr}{y}\right] \theta^y. \tag{4}$$

We now recall that the probability function (2) is a special case of the generalized power series distribution as defined by Patil [3] with range T and the series function $g(\theta)=(1+\theta)^N-1$. So the result (7) of Patil [3] and the series expansion (4) provide us the distribution of Y as

$$h(y) = \sum_{r=1}^{n} (-1)^{n-r} \binom{n}{r}\binom{Nr}{y} \theta^y/[(1+\theta)^N-1]^n \tag{5}$$

for $y=n, n+1, \ldots, Nn$, since the term in the summation is zero for $r=0$. Taking $\theta=p/1-p$ in (5), we get the distribution of Y in the form

$$h(y) = \sum_{r=1}^{n} (-1)^{n-r} \binom{n}{r}\binom{Nr}{y} p^y(1-p)^{Nn-y}/[1-(1-p)^N]^n \tag{6}$$

for $y=n, n+1, \ldots, Nn$. Further, it may be easily seen that the distribution function of Y is obtained as

$$\begin{aligned} F(y) &= 1-\sum_{x=y+1}^{Nn} \left\{\sum_{r=1}^{n} (-1)^{n-r} \binom{n}{r}\binom{Nr}{x} p^x(1-p)^{Nn-x}/[1-(1-p)^N]^n\right\} \\ &= 1-[1-(1-p)^N]^{-n} \sum_{r=1}^{n} (-1)^{n-r} \binom{n}{r} (1-p)^{N(n-r)} I_p(y+1, Nr-y) \end{aligned} \tag{7}$$

where $I_p(y+1, Nr-y)$ is the incomplete beta function tabulated by Pearson [4].

It may be remarked that the distribution of Y obtained by Malik [2, p. 335] should read as

$$f(y) = \sum_{r=0}^{n-1} b_r \binom{N(n-r)}{y} p^y q^{Nn-y} \tag{8}$$

for $y=n, n+1, \ldots, Nn$, where $p+q=1$, and $b_r=\binom{n}{r}(-1)^r/(1-q^N)^n$, which shows that (6) and (8) are identical.

References

1. W. Feller, *An introduction to probability theory and its applications*, Wiley, New York, **1** (third edition), 1968.

2. H. J. Malik, *Distribution of the sum of truncated binomial variates*. Canad. Math. Bull. **12** (1969), 334–336.

3. G. P. Patil, *Minimum variance unbiased estimation and certain problems of additive number theory*, Ann. Math. Statist. **34** (1963), 1050–1056.

4. K. Pearson, *Tables of the incomplete beta function*, Cambridge Univ. Press, London 1934.

Portland State University,
Portland, Oregon