

A CHARACTERIZATION OF THE NEGATIVE MULTINOMIAL DISTRIBUTION

J. PANARETOS*

Trinity College, Dublin

1. INTRODUCTION

The negative multinomial distribution (n.m.d.) arises in situations where we go on making trials until exactly n_0 occurrences of the 0th outcome have been noted and we require the joint probability of n_i occurrences of the i th outcome ($i = 1, 2, \dots, s$) noted before the n_0 th occurrence of the 0th outcome. Clearly it is a generalization of the negative binomial

*Present Address: Department of Statistics, University of Missouri, Columbia, Missouri 65211 USA.

distribution and just as the latter can be deduced from a number of different models so can the n.m.d. (see Sibuya *et al.*, 1964).

In the present note we use the conditional distribution of $\underline{Y}|\underline{X}$ of two n -dimensional random vectors \underline{X} , \underline{Y} and a relation between \underline{Y} and $\underline{X} - \underline{Y}$ to characterize the n.m.d. The relation we use is an extension to the multivariate case of a condition known in the literature as the Rao-Rubin condition (Rao and Rubin, 1964).

In Section 2 we quote a theorem by Shanbhag (1977) and its multivariate extension derived by Panaretos (1977). A similar result concerning truncated distributions is also stated. In Section 3 we state and prove the characterization of the n.m.d. Subsequently, characterizations of truncated forms of n.m.d.'s are mentioned. Finally, in Section 4 an application of the characterization is suggested and a comparison is made to another characterization derived by Janardan (1974).

2. SHANBHAG'S RESULT AND ITS MULTIVARIATE EXTENSION

Shanbhag (1977) arrived at a general characterization of discrete distributions using the conditional distribution of two random variables. His result can be stated as follows.

Theorem 1. Let $\{(a_n, b_n) : n = 0, 1, \dots\}$ be a sequence of real vectors with $a_n > 0$ for every $n \geq 0$ and $b_0, b_1 > 0, b_n \geq 0$ for $n \geq 2$. Denote by $\{c_n\}$ the convolution of $\{a_n\}$ and $\{b_n\}$. Let (X, Y) be a random vector of non-negative integer-valued components such that $P(X = n) = P_n, n \geq 0$ with $P_0 < 1$ and whenever $P_n > 0$ we have

$$P(Y = r | X = n) = \frac{a_r b_{n-r}}{c_n} \quad r = 0, 1, \dots, n. \quad (1)$$

$$\text{Then} \quad P(Y = r) = P(Y = r | X = Y), \quad r = 0, 1, \dots \quad (2)$$

if and only if (iff)

$$\frac{P_n}{c_n} = \frac{P_0}{c_0} \theta^n \quad n = 1, 2, \dots, \text{ for some } \theta > 0. \quad (3)$$

Condition (2) is known in the literature as the Rao-Rubin condition. It was first used by Rao and Rubin (1964) to show that if the distribution of $\underline{Y}|\underline{X}$ is binomial, (2) is necessary

and sufficient for X to be a Poisson random variable (r.v.). (It is clear that Rao and Rubin's result is a corollary of Theorem 1.)

Panaretos (1977) extended Shanbhag's result to the multivariate case in the following way.

Theorem 2. Let $\{(a_{\underline{n}}, b_{\underline{n}}) : \underline{n} = (n_1, \dots, n_s), n_i = 0, 1, \dots; i = 1, 2, \dots, s; s = 1, 2, \dots\}$ be a sequence of real vectors such that $a_{\underline{n}} > 0, b_{\underline{n}} \geq 0$ for every $n_i \geq 0, i = 1, 2, \dots, s$ with $b_{\underline{0}} > 0, b_{0, \dots, 0, 1} > 0$ and some $b_{0, 0, \dots, 0, 1, n_s} > 0$
 \vdots
 some $b_{0, 0, \dots, 1, n_{s-1}, n_s} > 0,$
 \vdots
 some $b_{1, n_2, n_3, \dots, n_s} > 0.$

Define $\{c_{\underline{n}}\}$ to be the convolution of $\{a_{\underline{n}}\}$ and $\{b_{\underline{n}}\}$ given by $c_{\underline{n}} = \sum_{\underline{r}=\underline{0}}^{\underline{n}} a_{\underline{r}} b_{\underline{n}-\underline{r}}$ where $a_{\underline{r}} = a_{r_1, \dots, r_s}$ and $\sum_{\underline{r}=\underline{0}}^{\underline{n}}$ denoting $\sum_{r_1=0}^{n_1} \sum_{r_2=0}^{n_2} \dots \sum_{r_s=0}^{n_s}$. Consider a random vector (X, Y) where

$X = (X_1, \dots, X_s), Y = (Y_1, \dots, Y_s)$ with $X_i, Y_i, i = 1, 2, \dots, s$ non-negative integer-valued r.v.'s such that $P(X = \underline{n}) = P_{\underline{n}},$ i.e., $P(X_1 = n_1, \dots, X_s = n_s) = P_{n_1, \dots, n_s}$ with $P_{n_1, \dots, n_s} > 0$ for some n_i and for every $i = 1, 2, \dots, s$ and whenever $P_{\underline{n}} > 0$

$$P(Y = \underline{r} | X = \underline{n}) = \frac{a_{\underline{r}} b_{\underline{n}-\underline{r}}}{c_{\underline{n}}}, \quad r_i = 0, 1, \dots, n_i; i = 1, 2, \dots, s. \tag{4}$$

Also define $X^{(j)} = (X_1, \dots, X_j), Y^{(j)} = (Y_1, \dots, Y_j), j = 2, 3, \dots, s$ and let $X^{(j)} > Y^{(j)}$ denote that $(X_k = Y_k, k = 1, 2, \dots, j-1, \text{ and } X_j > Y_j).$ Then

$$P(Y = \underline{r}) = P(Y = \underline{r} | X = Y) = P(Y = \underline{r} | X^{(j)} > Y^{(j)}), \quad j = 2, 3, \dots, s \tag{5}$$

$$\text{iff } \frac{P_n}{c_n} = \frac{P_0}{c_0} \prod_{i=1}^s \theta_i^{n_i} \text{ for some } \theta_1, \dots, \theta_s > 0. \quad (6)$$

Also if (6) is true then \underline{Y} and $\underline{X} - \underline{Y}$ are independent.

In an attempt to extend Rao and Rubin's result, Talwaker (1970) used the following relation as the multivariate analogue of the Rao-Rubin condition

$$P(\underline{Y} = \underline{r}) = P(\underline{Y} = \underline{r} | \underline{X} = \underline{Y}) = P(\underline{Y} = \underline{r} | \underline{X} > \underline{Y}) \quad (7)$$

Clearly (7) is more restrictive than (5). Patil and Ratnaparkhi (1977) replaced condition (7) with the linear regression of \underline{Y} on \underline{X} and characterized the double binomial and double inverse hypergeometric distributions as the distribution of $\underline{Y} | \underline{X}$.

If one uses the techniques employed by Panaretos (1979) to generalize Shanbhag's result so as to characterize truncated distributions, it is possible to derive the following result characterizing truncated multivariate distributions.

Theorem 3. Consider the following changes in the conditions of Theorem 2. Suppose that $a_n > 0$ for $n_1 \geq k_1$, $n_2 \geq k_2, \dots, n_s \geq k_s$ where k_1, k_2, \dots, k_s are non-negative integers. (Observe that in this case c_n is positive for all $n \geq k$.) Assume also this time that P_n is truncated at $k - \underline{1}$, i.e., that $P(\underline{X} \geq \underline{k}) = 1$, $P(X_i > k_i) > 0$; $i = 1, 2, \dots, s$ and whenever $P_n > 0$

$$P(\underline{Y} = \underline{r} | \underline{X} = \underline{n}) = \frac{a_r b_{n-r}}{c_n}, \quad r_i = 0, 1, \dots, n_i; \quad n_i = k_i, k_i + 1, \dots; \quad i = 1, 2, \dots, s. \quad (8)$$

Then

$$P(\underline{Y} = \underline{r} | \underline{Y} \geq \underline{k}) = P(\underline{Y} = \underline{r} | \underline{X} = \underline{Y}) = P(\underline{Y} = \underline{r} | X^{(j)} > Y^{(j)}, Y_j \geq k_j), \quad j = 2, 3, \dots, s \quad (9)$$

iff

$$\frac{P_{\underline{n}}}{c_{\underline{n}}} = \frac{P_{\underline{k}}}{c_{\underline{k}}} \prod_{i=1}^s c_i^{n_i - k_i} \text{ for some } \theta_1, \theta_2, \dots, \theta_s > 0 \quad (10)$$

$$n_i = k_i, k_i + 1, \dots; i = 1, 2, \dots, s.$$

Evidently, for $\underline{k} = \underline{0}$, Theorem 3 reduces to Theorem 2.

3. CHARACTERIZATION OF THE NEGATIVE MULTINOMIAL DISTRIBUTION

It was mentioned in the Introduction that the conditional distribution of \underline{Y} on \underline{X} will be used to characterize the n.m.d. as the distribution of \underline{X} . The form of the distribution of $\underline{Y}|\underline{X}$ required for this purpose is the negative inverse hypergeometric (n.i.h.d.) which arises again as a model in inverse sampling without replacement from a finite population. (For this and other models see Sibuya *et al.*, 1964).

Theorem 4. (Characterization of the negative multinomial distribution.) Suppose that

$$P(\underline{Y} = \underline{r} | \underline{X} = \underline{n}) = \frac{B(m+r_1+\dots+r_s, \rho+(n_1-r_1)+\dots+(n_s-r_s))}{B(m, \rho)} \prod_{i=1}^s \binom{n_i}{r_i} \quad (11)$$

$$r_i = 0, 1, \dots, n_i, m > 0, \rho > 0, i = 1, 2, \dots, s$$

(multivariate inverse hypergeometric with parameters m, ρ). Then, condition (5) holds iff

$$P(\underline{X} = \underline{n}) = \frac{\Gamma(m+\rho+n_1+\dots+n_s)}{\Gamma(m+\rho)} p_0^{m+\rho} \prod_{i=1}^s \frac{p_i^{n_i}}{n_i!} \quad (12)$$

$$n_i = 0, 1, \dots; 0 < p_i < 1, \sum_{i=1}^s p_i < 1, i = 1, 2, \dots, s; p_0 = 1 - \sum_{i=1}^s p_i$$

(negative multinomial with parameters $m + \rho, p_1, \dots, p_s$).

Proof. Let us consider the following sequences

$$\tilde{a}_r = \frac{\Gamma(m+r_1+\dots+r_s)}{\Gamma(m) \prod_{i=1}^s r_i!}, \quad \tilde{b}_n = \frac{\Gamma(\rho+n_1+\dots+n_s)}{\Gamma(\rho) \prod_{i=1}^s n_i!} \quad (13)$$

$$r_i = 0, 1, \dots, n_i; \quad n_i = 0, 1, \dots$$

The convolution $\{c_n\}$, $n_i = 0, 1, \dots$ of these sequences is

$$c_n = \frac{\Gamma(m+\rho+n_1+\dots+n_s)}{\Gamma(m+\rho) \prod_{i=1}^s n_i!} \sum_{r=0}^n \frac{B(m+r_1+\dots+r_s, \rho+(n_1-r_1)+\dots+(n_s-r_s))}{B(m, \rho)} \times \prod_{i=1}^s \binom{n_i}{r_i}$$

i.e.

$$c_n = \frac{\Gamma(m+\rho+n_1+\dots+n_s)}{\Gamma(m+\rho) \prod_{i=1}^s n_i!}, \quad n_i = 1, 2, \dots; \quad i = 1, 2, \dots, s. \quad (14)$$

It can be checked that the conditional distribution (11) can be expressed in the form $\frac{a_r b_{n-r}}{c_n}$ with $\tilde{a}_r, \tilde{b}_n, \tilde{c}_n$ given by

(13) and (14). Hence from Theorem 2 we have that condition (5) is equivalent to

$$\frac{P_n}{c_n} = \frac{P_0}{c_0} \prod_{i=1}^s \theta_i^{n_i} \quad \text{for some } \theta_i > 0, \quad i = 1, 2, \dots, s,$$

i.e., to

$$P_n = P_0 \frac{\Gamma(m+\rho+n_1+\dots+n_s)}{\Gamma(m+\rho)} \prod_{i=1}^s \frac{\theta_i^{n_i}}{n_i!} \quad (15)$$

Since $\sum_n P_n = 1$, it follows that

$$p_0^{-1} = \frac{\left(\sum_{i=1}^s \theta_i \right)^n}{p_0^{m+\rho}} \sum_{n=0}^{\infty} \frac{\Gamma(m+\rho+n_1+\dots+n_s)}{\Gamma(m+\rho)} p_0^{m+\rho} \prod_{i=1}^s \frac{\theta_i^{n_i}}{n_i!}$$

where $n = \sum_{i=1}^s n_i$, i.e.,

$$p_0^{-1} = \frac{\left(\sum_{i=1}^s \theta_i \right)^n}{p_0^{m+\rho}} \tag{16}$$

Substituting (16) in (15) gives the required result with

$$p_i = \theta_i / \sum_{j=1}^s \theta_j.$$

A similar result in the bivariate case has been proved by Patil and Ratnaparkhi (1975) but with the additional condition that $\partial^{r+\ell} G(t_1, t_2) / \partial t_1^r \partial t_2^\ell$ exists, for r, ℓ positive integers, and $G(t_1, t_2)$ the probability generating function of (X_1, X_2) .

Theorem 5. (Characterization of the truncated negative multinomial distribution). Assume that the conditional distribution of Y on X is m.i.h. as in (11). Then, condition (9) holds iff

$P_{\underline{n}} = P(X = \underline{n})$ is a n.m.d. truncated at $\underline{k} - \underline{1}$, i.e.,

$$P_{\underline{n}} = K \frac{\Gamma(m+\rho+n_1+n_2+\dots+n_s)}{\Gamma(m+\rho)} p_0^{m+\rho} \prod_{i=1}^s \frac{p_i^{n_i}}{n_i!},$$

$n_i = k_i, k_i+1, \dots; 0 < p_i < 1; i = 1, 2, \dots, s; \sum p_i < 1, p_0 = 1 - \sum_{i=1}^s p_i$

and K is the normalizing constant.

Proof. The proof follows from Theorem 3 if one considers as $a_{\underline{n}}$ and $b_{\underline{n}}$ the sequences given by (13).

Theorem 6. (Characterization of the convolution of a negative multinomial distribution with a truncated negative multinomial distribution.) Suppose that the distribution of $\underline{Y}|\underline{X}$ is m.i.h. truncated at $\underline{k} - \underline{1}$, i.e.,

$$P(\underline{Y} = \underline{r} | \underline{X} = \underline{n}) = R \frac{B(m+r_1+r_2+\dots+r_s, \rho+(n_1-r_1)+\dots+(n_s-r_s))}{B(m, \rho)} \times \prod_{i=1}^s \binom{n_i}{r_i}$$

$r_i = k_i, k_i+1, \dots, n_i$; $m > 0, \rho > 0$; $i = 1, 2, \dots, s$; R the normalizing constant. Then, condition (9) holds iff $P_{\underline{n}} = P(\underline{X} = \underline{n})$ is the convolution of a n.m.d. (m, p_1, \dots, p_s) truncated at $\underline{k} - \underline{1}$ with a n.m.d. (ρ, p_1, \dots, p_s) .

Proof. The proof follows again from Theorem 3 if we consider $a_{\underline{n}}$ to have the form of a n.m.d. (m, p_1, \dots, p_s) truncated at $\underline{k} - \underline{1}$ and $b_{\underline{n}}$ to have the form of a n.m.d. $(\rho, p_1, p_2, \dots, p_s)$.

4. AN APPLICATION

The characterization of the n.m.d. derived in the previous section can be of some importance in practice where conditions are satisfied for a m.i.h.d. to be the distribution of the conditional random variable $\underline{Y}|\underline{X} = \underline{n}$. In this case we might be able, because of the characterization, to deduce a n.m.d. for \underline{X} . One may argue that such a form for the distribution of $\underline{Y}|\underline{X} = \underline{n}$ may not be feasible in practice. However, it is a distribution used in connection with pollen analysis. Janardan (1973) for instance, assumed that counts of various kinds of pollen grains found at a given depth in sediment follow independent binomial distributions with constant proportion p . He then allowed p to vary from depth to depth, according to a beta distribution. Averaging over all depths in this manner he obtained the m.i.h.d. as the joint distribution for counts of various kinds of pollen grains. Therefore, in a problem of pollen analysis with a m.i.h.d. as a survival mechanism the results of Theorem 4 may indicate that the counts \underline{X} of the different pollen species. Similarly, it might indicate that \underline{X} has a distribution which is definitely not the negative multinomial.

Remark. Janardan (1974) has shown that if \underline{Y} , $\underline{X} - \underline{Y}$ are independent random vectors then each of them follows a negative multinomial distribution iff the conditional distribution of $\underline{Y}|\underline{X}$ is multivariate inverse hypergeometric. The result of Theorem 4 extends this result by making use of condition (5) which is less restrictive than independence between \underline{Y} and $\underline{X} - \underline{Y}$. It may, also, be observed that the "if" part of Janardan's result remains valid if condition (5) replaces the assumption of independence between \underline{Y} and $\underline{X} - \underline{Y}$.

REFERENCES

- Janardan, K. G. (1973). Chance mechanisms for multivariate hypergeometric models. *Sankhyā, Series A*, 35, 465-478.
- Janardan, K. G. (1974). A characterization of multinomial and negative multinomial distributions. *Skandinavisk Actuarietidskrift*, 57, 58-62.
- Panaretos, J. (1979). On characterizing some discrete distributions using an extension of the Rao-Rubin theorem (submitted for publication).
- Panaretos, J. (1977). A characterization of a general class of multivariate discrete distributions. In *Analytic Function Methods in Probability Theory, Colloquia Mathematica Societatis Janos Bolyai. No. 21 (1980)*, B. Gyires, ed. North Holland. Pages 243-252.
- Patil, G. P. and Ratnaparkhi, M. V. (1977). Certain characterizations with linearity or regression in additive damage models. *Journal of Multivariate Analysis*, 7, 598-601.
- Patil, G. P. and Ratnaparkhi, M. V. (1975). Problems of damaged random variables and related characterizations. In *Statistical Distributions in Scientific Work, Vol. 3*, G. P. Patil, S. Kotz, and J. K. Ord, eds. Reidel, Dordrecht-Holland, Pages 255-270.
- Rao, C. R. and Rubin H. (1964). On a characterization of the Poisson distribution. *Sankhyā, Series A*, 26, 295-298.
- Shanbhag, D. N. (1977). An extension of the Rao-Rubin characterization of the Poisson distribution. *Journal of Applied Probability*, 14, 640-646.
- Sibuya, M., Yoshimura, I., and Shimizu, R. (1964). Negative multinomial distribution. *Annals of the Institute of Statistical Mathematics*, 16, 409-426.
- Talwaker, S. (1970). A characterization of the double Poisson distribution. *Sankhyā, Series A*, 32, 265-270.

[Received June 1980]