THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

ON NEYMAN'S CONJECTURE: A CHARACTERIZATION OF THE MULTINOMIALS

Bikas Kumar Sinha and Thomas M. Gerig

Department of Statistics
North Carolina State University
Raleigh, North Carolina 27650

Institute of Statistics Mimeo Series # 1611

September 1982

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

ON NEYMAN'S CONJECTURE: A CHARACTERIZATION OF THE MULTINOMIALS

Bikas Kumar Sinha[1] and Thomas M. Gerig

Department of Statistics
North Carolina State University
Raleigh, North Carolina 27650

0. ABSTRACT

We provide a complete solution to a problem posed in Neyman (1965) and reformulated in Ghosh, Sinha and Sinha (1977) regarding a characterization of (positive and negative) multinomial distributions based, among other things, on the properties of regression in power series distributions.

AMS 1970 Subject Classification: Primary 62H05

Key words and phrases: Multivariate power series distributions, positive and negative multinomials, linear regression.

[1]On leave from the Indian Statistical Institute (Calcutta), India.

1. INTRODUCTION

There is no denying the fact that over the past two decades there has been an increasing interest in characterization of well-known discrete as well as continuous distributions and in characterization problems in general. Some excellent references, e.g., Kagan, Linnik and Rao (1972), Kotz (1974), Patil (edited-1963) and Patil, Kotz and Ord (edited-1975/ edited-1981), lead one to wonder about the vastness of the literature and diverse research interests on this topic. Our concern in this article is, with multivariate power series distributions (Khatri (1959)) and, more specifically, with a conjecture set forth by the late Professor J. Neyman in 1965 regarding a characterization of positive and negative multinomial distributions. We settle his conjecture completely after properly formulating it in this section below.

The key reference to this paper is Ghosh, Sinha and Sinha (1977) - hereafter abbreviated as GSS - wherein this particular problem has been studied and partially solved. Our work may be regarded as a supplement to that of GSS and the two together settle the conjecture. The paper by Sinha and Sinha (1976) accounts for the first attempt to attack the problem, (Another not-so-related paper, but of independent interest, is Sinha and Gerig (1982).)

To start with, suppose $(X_1, X_2, \ldots, X_k)$ follow a k-variate power series distribution with the joint pmf

$$f(x_1, x_2, \ldots, x_k) = a_{x_1 x_2 \ldots x_k} \theta_1^{x_1} \theta_2^{x_2} \ldots \theta_k^{x_k} / \Psi(\underline{\theta}) \qquad (1.1)$$

where $a_{x_1x_2\ldots x_k} \geq o$; $x_i \in I$(the set of non-negative integers), $1 \leq i \leq k$; $\underline{\theta} = (\theta_1,\theta_2,\ldots,\theta_k) \in H_k = \{(\theta_1,\ldots,\theta_k)|\theta_i > o,\ 1 \leq i \leq k;$ $\Psi(\theta) = \Sigma_{i_1i_2\ldots i_k}\ a_{i_1i_2\ldots i_k}\ \theta_1^{i_1}\ \theta_2^{i_2} \ldots \theta_k^{i_k} < \infty\}$.

A conjecture of Neyman (1965), reformulated in GSS, runs as follows:

Within the class of power series distributions, the multinomials are characterized by the following properties:

Q1 *The regression of* X_i *on the remaining variables is a linear function of the sum of the remaining variables.*

Q2 *The distribution of* $X_1 + X_2 + \ldots + X_k$ *is of the power series type.*

In GSS, the conjecture has been *settled* in the *affirmative* under the *conditions* $a_{oo\ldots o} > o$ and $a_{1o\ldots o} + \ldots + a_{oo\ldots o1} > o$. Without these conditions, however, the claim above (*as such*) turns out to be false. See Counter-Example 1 in Section 4 of the present paper.

Consider the following statements:

S1X The conditional distribution of X_i given $\{X_j = x_j(1\leq j\leq k,\ j \neq i)\}$ is *non-degenerate* for at least one set of values of the x_j's and the regression is linear and moreover, depends on the x_j's only through $\sum_{j(\neq i)} x_j$.* Further, $\sum_{j(\neq i)} x_j$ assumes at least three distinct values $(1\leq i\leq k)$.**

S2X The distribution of $X = X_1 + X_2 + \ldots + X_k$ is of the power series type.

---

*We note that in view of the Proposition in GSS (pp. 399), this conditional distribution also depends on the x_j's *only through* $\sum_{j(\neq i)} x_j$

**Throughout, we will assume this in order that the linearity of regression carries non-trivial sense. Without this condition, again, Theorem 1.1 is false. See Counter-Example 2 in Section 4.

In this article, we state and prove what we believe to be the correct form of Neyman's conjecture regarding the multinomials. In particular, we prove the following:

Theorem 1.1 Whenever S1X and S2X obtain, there exist integers $r \geq 1$, $t_1, t_2, \ldots, t_k \geq 0$ *and a set of r.v.'s* $(Z_1, Z_2, \ldots, Z_k)$ *having a joint (positive or negative) multinomial distribution such that the representation* $X_i = rZ_i + t_i$, $1 \leq i \leq k$, *holds with probability one.*

2. NEYMAN'S PROPERTIES FOR MULTINOMIALS

First of all we intend to develop certain basic results pertaining to this investigation. Let t_i = min(values of X_i) . Transform X_i to $Y_i = X_i - t_i$, $1 \leq i \leq k$. It is easy to observe that $(Y_1, Y_2, \ldots, Y_k)$ follow a k-variate power series distribution with the joint pmf given by, say,

$$h(y_1, y_2, \ldots, y_k) = b_{y_1 y_2 \ldots y_k} \theta_1^{y_1} \theta_2^{y_2} \ldots \theta_k^{y_k} / \xi(\underline{\theta}) \; ; \; y_1, \ldots, y_k \geq 0 \qquad (2.1)$$

Further, the statements (S1X,S2X) are <u>equivalent</u> to the analogous statements (S1Y,S2Y) concerning the Y_i's (and are obtained from S1X and S2X respectively by changing X_i to Y_i , x_j to y_j , $1 \leq i \neq j \leq k$) . Let us write $Y = Y_1 + Y_2 + \ldots + Y_k = Y_i + Y_i^*$ and $y = y_1 + y_2 + \ldots + y_k = y_i + y_i^*$, $1 \leq i \leq k$. Then Y_i^* has at least three distinct values and $y_i \geq 0$. Further, for any i, $b_{y_1 y_2 \ldots y_k} > 0$ for some $(y_1, y_2, \ldots, y_k)$ with $y_i = 0$ and $y_i^* = y - y_i \geq 0$. Our first result is concerned with the conditional distribution of Y_i given $Y_i^* = y_i^*$.

Proof. Take $i = 1$ for notational simplicity. In view of the Proposition in GSS (pp. 399), we note that the joint marginal distribution of $(Y_2,\ldots,Y_k)$ is a power series distribution with the pmf

$$\begin{aligned} g(y_2,\ldots,y_k) &= \sum_{y_1} h(y_1,\ldots,y_k) \\ &= c(y_2,\ldots,y_k)\, \theta_2^{*y_2}\theta_3^{*y_3}\ldots\theta_k^{*y_k}/\xi^*(\theta) \end{aligned} \tag{2.2}$$

where $\theta_i^* = \theta_i B(\theta_1)$, $2 \le i \le k$, $\xi^*(\theta) = \xi(\theta)/A(\theta_1)$; $A(\theta_1), B(\theta_1) > o$. Again, according to S1Y, the conditional distribution of Y_1 given $\{Y_i = y_i | 2 \le i \le k\}$ is (trivially) a power series distribution <u>depending only on</u> $y_2 + y_3 + \ldots + y_k$. Hence, the identity

$$b_{y_1 y_2 \ldots y_k}\, \theta_1^{y_1}\theta_2^{y_2}\ldots\theta_k^{y_k}/\xi(\theta) = \{c(y_2,\ldots,y_k)\, \theta_2^{*y_2}\ldots\theta_k^{*y_k}/\xi^*(\underline{\theta})\} \cdot \left\{ \frac{b_{y_1 y_2\ldots y_k}}{c(y_2,\ldots,y_k)} \cdot \frac{\theta_1^{y_1}}{A(\theta_1)(B(\theta_1))^{y_2+y_3+\ldots+y_k}} \right\}$$

leads to

$$b_{y_1 y_2\ldots y_k} = c(y_2,\ldots y_k) \cdot d_{y_1}(y_2+\ldots+y_k) \quad \text{(say)} \tag{2.3}$$

Finally, the numerator of

$$P[Y_1 = i | Y_1^* = y_1^*] = \frac{P[Y_1 = i \cap Y_1^* = y_1^*]}{P[Y_1^* = y_1^*]}$$

given by

$$\sum_{\substack{y_2,\ldots,y_k \\ y_2+\ldots+y_k = y_1^*}} b_{iy_2\ldots y_k}\,\theta_1^{\,i}\theta_2^{\,y_2}\ldots\theta_k^{\,y_k}/\xi(\theta)$$

simplifies, in view of (2.3), to

$$d_{i(y_1^*)}\,\theta_1^{\,i}\left[\sum_{\substack{y_2,\ldots,y_k \\ y_2+y_3+\ldots+y_k = y_1^*}} c(y_2,\ldots,y_k)\,\theta_2^{\,y_2}\ldots\theta_k^{\,y_k}/\xi(\theta)\right]$$

$$= \{d_{i(y_1^*)}\,\theta_1^{\,i}/A(\theta_1)(B(\theta_1))^{y_1^*}\}\ .\ \ P(Y_1^* = y_1^*)\quad \text{(using (2.2))}$$

Hence the Lemma.

The next result is interesting in itself and has been of fundamental importance in developing the proof of Theorem 1.1.

Theorem 2.1 (Under the set-up of Lemma 2.1). Whenever S1Y *and* S2Y *obtain,* $P[Y_i = o \mid Y_i^* = y_i^*] > o$ *for every value* y_i^* of Y_i^* .

The proof is given in the Appendix.

We are now in a position to look to the problem more closely. As a matter of fact, our next lemma, together with Theorem 5.1 in GSS, provides a complete proof of Theorem 1.1 stated in Section 1.

Lemma 2.2 *Whenever* S1Y *and* S2Y *obtain, we necessarily have*

(i) $b_{oo\ldots o} > o$,

(ii) *if* r *is the least positive value of* Y, *then*

$b_{ro\ldots o} > o,\ b_{oro\ldots o} > o,\ \ldots,\ b_{oo\ldots or} > o$ *while*

$b_{y_1y_2\ldots y_k} = o$ *for any other* $(y_1,\ldots,y_k)$ *satisfying*

$y_1 + y_2 + \ldots + y_k = r$,

(iii) *if* s *is any other value of* Y, $r|s$.

Proof. Let us write the power series distribution of Y as

$$P(Y=t) = u(t)v^{-1}(\underline{\theta})\lambda^{t}(\underline{\theta}) \tag{2.4}$$

where $t \geq o$, $u(t) \geq o$, $v(\underline{\theta}) = \sum_{t} u(t)\,\lambda^{t}(\underline{\theta})$ and $\lambda(\underline{\theta})$ is a function of the θ_i's . Clearly, $v(\underline{\theta}) < \infty$ for $\underline{\theta} \in H$.

Referring to (2.1) and (2.3), we then have

$$u(t)v^{-1}(\underline{\theta})\xi(\underline{\theta})\lambda^{t}(\underline{\theta}) = \sum_{i=0}^{t} d_{i(t-i)}\,\theta_1^{\,i}\,\eta(\theta_2,\ldots,\theta_k|t-i) \tag{2.5}$$

where

$$\eta(\theta_2,\ldots,\theta_k|t-i) = \sum_{\substack{y_2,\ldots,y_k \\ y_2+\ldots+y_k = t-i}} c(y_2,\ldots,y_k)\theta_2^{y_2}\ldots\theta_k^{y_k} \tag{2.6}$$

is a homogeneous polynomial of degree $(t-i)$ in $(\theta_2,\ldots,\theta_k)$. Note incidentally that we have taken the representation

$$P(Y_1^{*} = t-i) = \eta(\theta_2,\ldots,\theta_k|t-i)/\xi^{*}(\underline{\theta})B^{t-i}(\theta_1) \tag{2.7}$$

Next observe that a consequence of Y_1^{*} having at least three distinct values is that the total $Y = Y_1 + Y_1^{*}$ must also have at least three distinct values (since, according to Theorem 2.1, the value o of Y_1 must necessarily combine with each value of Y_1^{*}) . Let (t_1,t_2,t_3) be <u>any</u> triplet of values of Y satisfying $o \leq t_1 < t_2 < t_3$. Denoting by $P_t(\theta)$ the homogeneous polynomial of degree t (in the θ_i's) in the right hand side of (2.5), we now deduce, using (2.5),

$$P_{t_2}^{t_3-t_1}(\underline{\theta}) = A(t_1,t_2,t_3)\, P_{t_1}^{t_3-t_2}(\underline{\theta})\; P_{t_3}^{t_2-t_1}(\underline{\theta}) \text{ identically in } \underline{\theta} \qquad (2.8)$$

where $A(t_1,t_2,t_3) = \{u(t_2)\}^{t_3-t_1}/\{u(t_1)\}^{t_3-t_2}\{u(t_3)\}^{t_2-t_1}$ is a function only of t_1,t_2,t_3.

We now take up demonstration of the results one by one. Let r be the <u>least positive</u> value of Y and let s_i be the least positive value of Y_i^* in combination with $Y_i = o$, $1 \leq i \leq k$. Then we have a series of implications shown below concerning the values of Y_i's or of (Y_i, Y_i^*)'s.

$$\begin{bmatrix} \underline{Y_1} & \underline{Y_1^*} \\ o & s_1(\geq r) \\ r_1 & r_2 \end{bmatrix} \xrightarrow[\text{(if } r_1 < r)]{\text{(by Th. 2.1)}} \begin{bmatrix} \underline{Y_1} & \underline{Y_1^*} \\ o & r_2 \end{bmatrix} \xrightarrow[\text{least positive value)}]{\text{(Since } r \text{ is the}} r_2 = r, r_1 = o$$

$$(r_1 + r_2 = r)$$

$$\Longrightarrow \begin{bmatrix} \underline{Y_1} & \underline{Y_1^*} \\ o & r \end{bmatrix} \xrightarrow{\text{(for } k \geq 3} \begin{bmatrix} \underline{Y_1} & \underline{Y_2} & \underline{Y_3} & \cdots & \underline{Y_k} \\ o & r_2 & r_3 & \cdots & r_k \end{bmatrix} \xrightarrow{\text{(if } r_2 < r)}$$

$$r_2 + r_3 + \ldots + r_k = r$$

$$\begin{bmatrix} \underline{Y_2} & \underline{Y_2^*} \\ r_2 & r_2^* \end{bmatrix} \text{ which in combination with } \begin{bmatrix} \underline{Y_2} & \underline{Y_2^*} \\ o & s_2(\geq r) \end{bmatrix} \xrightarrow{\text{(by Th. 2.1)}}$$

$$(r_2 + r_2^* = r)$$

$$\begin{bmatrix} \underline{Y_2} & \underline{Y_2^*} \\ o & r_2^* \end{bmatrix} \xrightarrow[\text{least positive value)}]{\text{(since } r \text{ is the}} r_2^* = r,\ r_2 = o \text{ and so on.}$$

Thus, $P[Y = r]$ arises essentially due to all or a (non-null) subset of the following combinations of values y_i's of the Y_i's: $\{(r,o,\ldots,o), (o,r,o,\ldots,o), \ldots (o,o,\ldots,o,r)\}$. Specifically, let $(o,o,\ldots,o,r)$ be a contributing term in $P(Y=r)$. Then

$$\begin{bmatrix} \underline{Y_k} & \underline{Y_k^*} \\ r & o \\ o & s_k(\geq r) \end{bmatrix} \xrightarrow{\text{(by Th. 2.1)}} \begin{bmatrix} \underline{Y_k} & \underline{Y_k^*} \\ o & o \end{bmatrix}$$, thereby establishing (i).

If possible, let $(r,o,\ldots,o)$ be missing in the effective subset determining $P(Y=r)$. Then $P_r(\underline{\theta})$ does <u>not</u> involve any term involving θ_1. If now $s(>r)$ be any other value of Y, we can set $t_1 = o$, $t_2 = r$, $t_3 = s$ in (2.8) and claim

$$P_r^s(\underline{\theta}) = A(r,s)\, P_s^r(\underline{\theta}) \quad \text{identically in } \underline{\theta} \tag{2.9}$$

This shows that $P_s(\underline{\theta})$ also must <u>not</u> involve any term involving θ_1 for any s whatsoever. But this is equivalent to $P[Y_1=o] = 1$ which contradicts SlY. Hence, $P_r(\underline{\theta})$ is, as a matter of fact, an <u>irreducible homogeneous polynomial</u> of degree r in all of $(\theta_1,\ldots,\theta_k)$ and has the form, say, $\sum_{i=1}^{k} a_i\theta_i^r$ with $a_i > o$ for every i. If, for any s, $(r,s) = 1$, i.e., r and s are relatively prime to each other, (2.9) becomes impossible unless $r = 1$. [This can be seen as follows. The polynomial $P_s^r(\underline{\theta})$ contains a term of the form $\theta_1^{s\,r_1}\theta_2^{s\,r_2}$ for $o < r_1, r_2 < r$, $r_1 + r_2 = r$. Hence, there exists a decomposition of s as $s = s_1 + s_2$, $o < s_1, s_2 < s$, such that $rs_1 = sr_1$ and

$rs_2 = sr_2$. Taking $r_1 = r - 1$, this means that $r|s$ which is a contradiction unless $r = 1$] . If, again, no two positive values of Y are relatively prime to each other, we can set, for some $s > r$, $(r,s) = h > 1$, $r = ph$, $s = qh$, $(p,q) = 1$. Then (2.9) reads as $P_r^q(\underline{\theta}) = P_s^p(\underline{\theta})$ identically in $\underline{\theta}$. The polynomial $P_s^p(\underline{\theta})$ contains a term of the form $\theta_1^{sp_1}\theta_2^{sp_2}$ for $o < p_1, p_2 < p$, $p_1 + p_2 = p$. Hence, there exist integers q_1, q_2, $o < q_1, q_2 < q$, $q_1 + q_2 = q$ such that $rq_1 = sp_1$ and $rq_2 = sp_2$, i.e., $pq_1 = qp_1$ and $pq_2 = qp_2$. Once again, taking $p_1 = p - 1$, we get that $p|q$ so that $p = 1$ necessarily. Hence, in any case, we must have, for any $s > r$, $r|s$ as also $b_{ro\ldots o} > o$, $b_{oro\ldots o} > o$, ..., $b_{oo\ldots or} > o$. This proves the lemma.

<u>Remark 1</u> It is now enough to define Z_i by $X_i = t_i + rZ_i$, $1 \leq i \leq k$ and note that the conditions for applicability of Theorem 5.1 of GSS on the joint distribution of $(Z_1, Z_2, \ldots Z_k)$ have all been established. Thus, Theorem 1.1 gets through, thereby establishing Neyman's conjecture.

3. APPENDIX

<u>Proof of Theorem 2.1</u> We will use various notations already established through (2.1)-(2.8). We proceed through the following steps.

<u>Step I</u> Certainly, $P[Y_i = o | Y_i^* = y_i^*] > o$ for some value y_i^* of Y_i^*. Also, if <u>at least two</u> of the conditional distributions (of Y_i given Y_i^*) are <u>degenerate</u>, then linearity of regression would demand all such distributions to be degenerate - thereby contradicting S1Y. Consequently, only one of them at the most can be degenerate. Again, Y_i^* assumes at least three distinct values. Therefore, $Y = Y_i + Y_i^*$ will also assume at least three distinct values except under the following situation:

Y_i^*	Y_i	Y
r	s-r, t-r	s,t
s	o, t-s	s,t
t	o	t

Calculations yield (for some a_i, b_i, a_i^*, b_i^*, c_i^*, d_i^*)

$$E(Y_i|Y_i^* = r) = \frac{a_i + b_i\theta_i^{t-s}}{a_i^* + b_i^*\theta_i^{t-s}} ,$$

$$E(Y_i|Y_i^* = s) = \theta_i^{t-s}/c_i^* + d_i^*\theta_i^{t-j}$$

$E(Y_i|Y_i^* = t) = o$ and linearity is violated.

Step II We will now treat the cases when Y has at least three distinct values, say, r, s, t in the order $o \leq r < s < t$. Without any loss of generality, $Y_i = o$ may be made to correspond to one of $Y_i^* = r, s, t$. No matter to which it corresponds, the important point is whether the underlying conditional distribution is degenerate at o or not. We will settle both the cases.

Step III Suppose one of the conditional distributions is non-degenerate with $Y_i = o$ having a positive conditional probability while, if possible, there is another with $Y_i = o$ having zero probability in the conditional distribution. Then, a direct analysis, similar to that in Step I, would result in non-linearity of regression. Hence, in such situations, $P[Y_i = o|Y_i^* = y_i^*] > o$ becomes a necessity for every value y_i^* of Y_i^*.

Step IV Now suppose that the only degenerate conditional distribution concentrates on $Y_i = o$. We will argue that this violates S2Y whenever Y_i^* is not an extreme value of Y. For notational simplicity we take $i = 1$ and $P[Y_1=o|Y_1^*=s] = 1$ with $P[Y=r] > o$, $P[Y=t] > o$ and $r < s < t$. Referring to (2.8), we must have $P_s^{t-r}(\underline{\theta}) = A(r,s,t) P_r^{t-s}(\underline{\theta})\, P_t^{s-r}(\underline{\theta})$ identically in $\underline{\theta}$ where $P_r(\underline{\theta})$, etc., are to be obtained from (2.5). Now $P[Y_1=o|Y_1^*=s] = 1$ implies, from (2.5), that

$P_s(\underline{\theta})$ involves the term $\eta(\theta_2,\ldots,\theta_k|s)$ which is a homogeneous polynomial of degree s in $(\theta_2,\theta_3,\ldots,\theta_k)$. Taking limit on both sides of (1) as $\theta_1 \to o$, the left hand side tends to a positive quantity (more specifically, $\to d_{o(s)}\eta(\theta_2,\ldots,\theta_k|s))$. Hence, none of $P_r(\underline{\theta})$ and $P_t(\underline{\theta})$ can vanish as $\theta_1 \to o$. Clearly, this is the case when $P[Y_1=o|Y_1^*=r] > o$, $P[Y_1=o|Y_1^*=t] > o$. Certainly, this must be true for all values of Y .

Step V The highly non-trivial case yet to be settled is: $P[Y_1=o|Y_1^*=r] = 1$ where r is the minimum value of Y (or the maximum value of Y, in case Y is finite with probability one). This is the only degenerate distribution and no other includes the value o for Y_1 . Consider the following

Table I

Y_1^*	Y_1	Conditional probability	Regression
r	o	1	o
s_o	s_1	$a_1\theta_1^{s_1}/\{a_1\theta_1^{s_1}+a_2\theta_1^{s_2}+\ldots\}$	$\dfrac{a_1s_1 + a_2s_2\theta_1^{s_2-s_1} + \ldots}{a_1 + a_2\theta_1^{s_2-s_1} + \ldots}$
	s_2	$a_2\theta_1^{s_2}/\{a_1\theta_1^{s_1}+a_2\theta_1^{s_2}+\ldots\}$	
	$\vdots$		
t_o	t_1	$b_1\theta_1^{t_1}/\{b_1\theta_1^{t_1}+b_2\theta_2^{t_2}+\ldots\}$	$\dfrac{b_1t_1 + b_2t_2\theta_1^{t_2-t_1} + \ldots}{b_1 + b_2\theta_1^{t_2-t_1} + \ldots}$
	t_2	$b_2\theta_1^{t_2}/\{b_1\theta_1^{t_1}+b_2\theta_1^{t_2}+\ldots\}$	
	$\vdots$	$\vdots$	

Linearity of regression means

$$\frac{a_1 s_1 + a_2 s_2 \theta_1^{s_2 - s_1} + \dots}{a_1 + a_2 \theta_1^{s_2 - s_1} + \dots} = \left(\frac{s_o - r}{t_o - r}\right) \cdot \frac{b_1 t_1 + b_2 t_2 \theta_1^{t_2 - t_1} + \dots}{b_1 + b_2 \theta_1^{t_2 - t_1} + \dots} \qquad (1)$$

identically in θ_1. Hence,

$$s_1 = \frac{s_o - r}{t_o - r}\ t_1 \qquad (2)$$

and further,

$$s_2 = \frac{s_o - r}{t_o - r}\ t_1$$

$$s_1 = \frac{s_o - r}{t_o - r}\ t_2 \qquad (3)$$

provided $s_2 - s_1 \neq t_2 - t_1$. But (2) and (3) contradict each other. Hence, we must have, among other relations (involving the values in the conditional distributions displayed in the above table),

$$s_1 = \frac{s_o - r}{t_o - r}\, t_1 \ , \ s_2 - s_1 = t_2 - t_1 \qquad (4)$$

[Note that all values of Y_1^* must lie on the <u>same</u> side of r as, otherwise, the regressions (positive quantities) cannot be collinear with o in the middle. This justifies the same sign of $(s_o - r)$ and $(t_o - r)$]. We will now apply (4) to get into a contradiction. Once again we refer to the above table and this time we wish to examine S2Y. We take $r < s_o < t_o$ and write $s = s_o + s_1 < t = t_o + t_1$. We now refer to the identity in Step IV involving P_r, P_s, P_t and set $\theta_i = \theta_i^o$,

$2 \le i \le k$ so that it reads as an identity in θ_1. Arrange P_s in increasing order of powers of θ_1 and let $\theta_1^{\,i}$ be the first member. Similarly, let $\theta_1^{\,j}$ be the first member in the expression for P_t. Note that $1 \le i \le s - r - 1$ and $1 \le j \le t - r - 1$ (as, otherwise, we will end up with situations discussed in our earlier steps). But now $i(t-r) = j(s-r)$ is a necessity and this is not satisfied whenever $(s-r, t-r) = 1$, i.e., they are relatively prime to each other. So, we are left with the situation $s = r + ph$, $t = r + qh$, $h > 1$, $(p,q) = 1$ and, then, for some ε, $1 \le \varepsilon < h$, we have $i = \varepsilon p$ and $j = \varepsilon q$. Hence, we end up with the following set-up (verifying a part of (4) and utilizing the other part, e.g., $s_2 - s_1 = t_2 - t_1$):

Table II

Y_i^*	Y_i	Y
r	o	r
$r + (h-\varepsilon)p$	εp	$r + ph$
	$\varepsilon p + x$	$r + x + ph$
	$\vdots$	$\vdots$
$r + (h-\varepsilon)q$	εq	$r + qh$
	$\varepsilon q + x$	$r + qh + x$
	$\vdots$	$\vdots$

If in Table I, there are only two values under each of the conditional distributions (non-degenerate), we get, besides (4), $s_2 = \frac{s_o - r}{t_o - r}\, t_2$ and this applied on the latter table, means that

$\varepsilon p + x = \frac{p}{q}(\varepsilon q + x)$ must hold. But this is not true since $p \neq q$. Again, on the other hand, if (in Table I) there are values s_3 and t_3 but $(s_3 - s_1) \neq 2(s_2 - s_1)$, $(t_3 - t_1) \neq 2(t_2 - t_1)$, then also we arrive at a contradiction (to linear regression and/or power series distribution) in the same manner. Hence, in Table II, all successive values must necessarily have an equal increment of x in both the (non-degenerate) distributions. Suppose the last terms are $\varepsilon p + nx$ and $\varepsilon q + nx$ respectively. Then, once again, we must have $\varepsilon p + nx = \frac{p}{q}(\varepsilon q + nx)$ implying $p = q$ which is not the case.

Like this, under any such situation, one can come up with a contradiction to S1Y and/or S2Y. We will stop here.

4. COUNTER-EXAMPLES

1. That Q1 and Q2 alone do not lead to the multinomials can easily be seen through counter-examples such as

X_1	X_2	...	X_k	$a_{x_1 x_2 \ldots x_k}$
o	o	...	o	1
1	1	...	1	1
2		...	2	1
⋮	⋮		⋮	⋮
n	n	...	n	1

2. If the conditioning variable assumes only two values, linear regression becomes a trivial property of any set of conditional distributions. Then the stated claim is not valid as the following example illustrates:

X_1	X_2	. . .	X_k	$a_{x_1 x_2 \dots x_k}$
o	o	. . .	o	1
2	o	. . .	o	1
o	2	. . .	o	1
⋮	⋮		⋮	⋮
o	o	. . .	2	1

5. CONCLUDING REMARKS

As is well-known, the most appropriate prior for the multinomial parameters θ_i's is the Dirichlet prior given by

$$\Pi(\theta_1,\theta_2,\dots,\theta_k) = \frac{\Gamma(\alpha_1+\alpha_2+\dots+\alpha_k+\alpha_{k+1})}{\Gamma(\alpha_1)\Gamma(\alpha_2)\dots\Gamma(\alpha_{k+1})} \prod \theta_i^{\alpha_i-1}\,(1-\sum_1^k \theta_i)^{\alpha_{k+1}-1},$$

$$o < \alpha_i,\ \theta_i \geq o,\ \sum_1^k \theta_i < 1 .$$

Various characterizations of such distributions are available in the literature. This distribution also possesses <u>all</u> the properties of the multinomials (vide Neyman (1965), Sinha and Sinha (1976)). It would be natural to investigate a similar characterization of these distributions through the properties of regression. We will undertake this investigation in a separate communication.

6. ACKNOWLEDGEMENT

We thank Professor John Bishir for drawing our attention to the reference [7].

7. REFERENCES

[1] Ghosh, J. K., Sinha, B. K. and Sinha, B. K. (1977): Multivariate power series distributions and Neyman's properties for multinomials. Jour. Multivariate Analysis 7(3), 397-408.

[2] Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1972): Characterization problems in mathematical statistics (in Russian). Nauka, Moscow (English edition, 1972, John Wiley, New York).

[3] Khatri, C. G. (1959): On certain properties of power-series distributions. Biometrika, 46, 486-490.

[4] Kotz, S. (1974): Characterizations of statistical distributions. A supplement to recent surveys. International Statistical Review 42, 39-65.

[5] Neyman, J. (1965): Certain chance mechanisms involving discrete distributions in Classical and Contagious Discrete Distributions (Ed. G. P. Patil). Calcutta Statistical Publishing Soc., 4-14.

[6] Classical and Contagious Discrete Distributions. Ed. G. P. Patil. Calcutta Statistical Publishing Soc. (1965).

[7] Statistical Distributions in Scientific Work: Volume 3 - Characterizations and applications (1975). Ed. Patil, G. P., Kotz, S. and Ord, J. K. Volume 4 - Models, structures and characterizations. Ed. Taillie, C., Patil, G. P. and Baldessari, B. A. (1981).

[8] Sinha, Bimal Kumar and Sinha, Bikas Kumar (1976): On a characterization of the dispersion matrix based on the properties of regression. Comm. Statist. A 5(13), 1215-1224.

[9] Sinha, Bikas Kumar and Thomas M. Gerig (1982): On a further characterization of dispersion matrices based on the properties of regression. Submitted to the Journal of Statistical Planning and Inference.