QÜESTIIÓ, Vol 17, 3 pp. 301-326, 1993

UNBIASED ESTIMATORS OF MULTIVARIATE DISCRETE DISTRIBUTIONS AND CHI-SQUARE GOODNESS-OF-FIT TEST

NIKULIN, M.S.* and VOINOV, V.G.†

We consider the problem of estimation of the value of a real-valued function $u(\theta)$, $\theta = (\vartheta_1, \ldots, \vartheta_K)^T$, on the basis of a sample from non-truncated or truncated multivariate Modified Power Series Distributions. Using the general theory of estimation and the results of Patil(1965) and Patel(1978) we give the tables of MVUE*'s for functions of parameter θ of trinomial, multinomial, negative multinomial and left-truncated modified power series distributions. We have applied the properties of* MVUE*'s and the results from the theory of MVU estimation to construct a goodness-of-fit chi-squared test for multivariate modified power series distributions.*

Key words: Multivariate modified power series distributions, sufficient statistic, MVUE, Rao-Kolmogorov-Blackwell method, chi-squared goodness-of-fit test, minimum chi-squared estimator, maximum likelihood estimator, Chernoff-Lehmann theorem, BAN estimator.

*Nikulin, M.S. The Steklov Mathematical Institute in St.Petersburg and The University Bordeaux-2.

†Voinov, V.G. Physical-Technical Institute of Academy of Sciences of Kazakhstan, Alma-Ata.

–Article presentat al Seventh International Conference on Multivariate Analysis, setembre 1992.

–Acceptat l'abril de 1993.

INTRODUCTION

The problem of obtaining the MVUE's of the probability densities and related functions for various families of distributions (for which a complete sufficient statistic, Z, exists) is of great interest for many reasons, as noted, for example, by Halmos (1946), Rao (1949), Kolmogorov (1950), Lehmann (1951), Joshi and Park (1974), Hoeffding (1983,1984) etc. The main one is the general problem of estimation E $u(X)$, where $u(x)$ is an arbitrary given function. This problem really consists in obtaining the MVUE $\hat{p}(x;\theta)$ of the density $p(x;\theta)$ and integrating with respect to x the product of this estimator and $u(x)$, i.e.,

$$\hat{\mathrm{E}}u(X) = \int u(x)\hat{p}(x;\theta)\,\mathrm{d}x.$$

Neyman and Scott (1960) were the first to consider the problem in the particular case of a normal distibution with unknown mean and variance. In this way one can obtain for example MVUE's for characteristic functions, moment generating functions etc., and derive MVUE's for moments by differentiating.

Another problem is to estimate the probabilty $\mathcal{P}(X \in A)$. It is also tantamount to obtaining the MVUE $\hat{p}(x;\theta)$ and integrating this estimator over the set A, i.e.,

$$\hat{\mathcal{P}}(X \in A) = \int_A \hat{p}(x;\theta)\,\mathrm{d}x.$$

There are many ways of obtaining the MVUE's of probability densities $p(x;\theta)$. One of them consists in obtaining the solution $\varphi(x|Z) = \hat{p}(x;\theta)$ of the unbiasedness integral equation

$$\int \varphi(x|z)q(z;\theta)\,\mathrm{d}z = p(x;\theta).$$

where $q(z|\theta)$ is the density of the sufficient statistic Z. This approach has been used, for example, by Ghurye and Olkin (1969) to obtain the MVUE's for multivariate normal probability families, see also Lumelsky and Sapoznikov (1969). Another way consists in Rao-Kolmogorov-Blackwellizing an arbitrary unbiased estimator with respect to the minimal sufficient statistics Z, which were used by Patel (1965) and Patel (1978) for the multivariate modified power series distributions. The general theory for the representation of MVUE's has been developed in depth, see, for example, Stein (1950), Bahadur (1957). Some useful properties of multivariate power series distributions and general results related with the theory of MVU estimation can be found in Ghosh, Sinha and Sinha (1977) and "*Encyclopedia of Statistical Sciences*", edited by S.Kotz and N.L.Johnson, see also Johnson and Kotz (1969).

We shall no longer discuss this or any related problems any more, since they have been considered in details in the monograph Voinov and Nikulin (1993), where many examples, applications and tables of MVUE's are given.

Being the most often used a chi-square goodness-of-fit test is based on the limiting theorem established by Carl Pearson in 1900. The theorem states that a quadratic form of Pearson

$$X^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i}$$

of vector $\nu = (\nu_1, \ldots, \nu_r)^T$ following the multinomial distribution $M_r(n,p)$, $p = (p_1, \ldots, p_r)^T$, tends as $n \Rightarrow \infty$ to the limiting chi-square distribution with $r-1$ degrees of freedom.

Later on R. Fisher showed that if multinomial probabilities $p_i = p_i(\theta)$ depend on unknown parameter $\theta = (\vartheta_1, \ldots, \vartheta_k)^T$, which can be replaced by *the minimum chi-squared estimator* $\tilde{\theta}_n$ or an estimator asymptotically equivalent to it which is a root of the equation

$$\sum_{i=1}^{r} \frac{\nu_i}{np_i(\theta)} \cdot \frac{\partial p_i(\theta)}{\partial \vartheta_j} = 0, \qquad j = 1, \ldots, k, \quad (k \le r-2),$$

then the Pearson's statistic

$$X^2(\tilde{\theta}_n) = \sum_{i=1}^{r} \frac{\left[\nu_i - np_i(\tilde{\theta}_n)\right]^2}{np_i(\tilde{\theta}_n)}$$

has in the limite a chi-square distribution with $r-k-1$ degrees of freedom. This fact allows to construct a reasonable statistical test for testing the hypothesis that outcome probabilities belong to a given parametric family. Untill 50s one assumed that Pearson's test is applicable in more complicated situations, where the conditions of Fisher apparenltly do not hold. In particular this case appears when testing the hypothesis that a continuous probability distribution belongs to a given parametric family. Untill now one may found many examples of incorrect recommendations on applications of Pearson's test in statistical guide books (see, for example, monographs of H.Cramer, A.Hald, M.Kendall, W.Feller). One assumed that a violation of conditions of Pearson's and Fisher's theorems can not essentially distort a statistical inference. However, in 1954 Chernoff and Lehmann showed that a formal application of Pearson's test when one uses *the maximum likelihood estimator* $\hat{\theta}_n$ leeds to a substantial difference of a limiting distribution of the statistic $X^2(\hat{\theta}_n)$ from the chi-square distribution: the statistic $X^2(\hat{\theta}_n)$ is distributed in the limite as $n \Rightarrow \infty$ like

$$\xi_1^2 + \cdots + \xi_{r-k-1}^2 + \mu_1 \xi_{r-k}^2 + \cdots + \mu_k \xi_{k-1}^2,$$

where $\xi_1, \ldots, \xi_{k-1}$ are idependent standard normally distributed random variables, the numbers $\mu_1, \ldots, \mu_k$ lie between 0 and 1, and generally speaking, depend on the unknown value of the parameter θ. For this reason Nikulin (1973), Dzjaparidze and Nikulin (1974) (see, also Rao and Robson (1974), Moore and Spruill (1975)) proposed to consider two modifications Y^2 and W^2 of the statistic of Pearson, depending on the method of estimation of the parameter θ. The statistic Y^2 is applied if one uses the maximum likelihood estimator $\hat{\theta}_n$ or an estimator asymptotically equivalent to it (*BAN estimator*) and the limit distribution of Y^2 is a chi-square with $r-1$ degrees of freedom. The statistic W^2 is different from the statistic Y^2 and allows us to apply a chi-square test when we use any *$\sqrt{n}$−consistent estimator* θ_n^* of θ and the limiting distributions of W^2 is a chi-square with $r-k-1$ degrees of freedom. More details one can find, for example, in the paper of Nikulin (1991), where is pointed out that one may construct a chi-square-type test by using the best minimum variance unbiased estimators of the unknown probabilities. This paper generalizes the results of Bolshev and Mirvaliev (1978) and Voinov and Nikulin (1993) for a wide class of multivariate discrete distributions, and with the paper of Nikulin and Greenwood (1990) gives some general presentation about the theory and practice of the chi-square test.

1. NON-TRUNCATED MULTIVARIATE MODIFIED POWER SERIES DISTRIBUTIONS

Following Patil (1965) and Patel (1978) consider first the problem of minimum variance unbiased estimation of non-truncated multivariate modified power series distributions.

Let

$$\mathcal{A}^{(i)} = \{a_1^{(i)}, a_2^{(i)}, \ldots\} \qquad i = 1, 2, \ldots, n,$$

be arbitrary subsets of k−vectors, $a_j^{(i)} = (a_{j1}^{(i)}, \ldots, a_{jk}^{(i)})^T \in \mathbb{R}^k$.

A sum

$$\mathcal{A}_n = \sum_{i=1}^{n} \mathcal{A}^{(i)}$$

is the set of all k−vectors of the form

$$\sum_{i=1}^{n} a^{(i)},$$

where $a^{(i)} \in \mathcal{A}^{(i)}$. If $\mathcal{A}^{(i)} = \mathcal{A},\ i = 1, 2, \ldots, n$, then

$$\mathcal{A}_n = \sum_{i=1}^{n} \mathcal{A}$$

is denoted by $n[\mathcal{A}]$. A set $\{a\}$ denotes *the singleton*, the set of only one vector. A set T of a $k-$dimensional space is said *to be stable with respect to a vector* $r = (r_1, r_2, \ldots, r_k)^T$ if $x = (x_1, x_2, \ldots, x_k)^T \in T$ implies $x + r \in T$. Equivalently, $T + \{r\} \in T$.

A subset T of a $k-$dimensional space is said to be *the index set of the function*

$$f(\theta) = \sum a(x) \vartheta_1^{x_1} \cdots \vartheta_k^{x_k}, \theta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_k)^T \in \Theta \in \mathbb{R}^k,\ \vartheta_i \geq 0,$$

where $a(x) \neq 0,\ x \in T,$ so that $f(\theta)$ is finite and differentiable.

This index set is denoted by

$$T = W[f(\theta)]. \tag{1}$$

A real valued function $u(\theta), \theta \in \Theta$, is called *MVU estimable* if it possesses the MVU estimator based on a sample

$$\mathbb{X} = (X_1, X_2, \ldots, X_n)^T \qquad X_i = (X_{i1}, \ldots, X_{ik}),$$

of size n drawn from an appropriate distribution.

Let T be a set of $k-$fold cartesian product of the set I, where I is the set of non-negative integers, i.e.

$$T = \{(x_1, \ldots, x_k)^T : x_i \in I, \quad i = 1, 2, \ldots, k\} \subseteq I \times I \times \cdots \times I,$$

and let

$$f(\theta) = \sum a(x)(g(\vartheta_1))^{x_1} \cdots (g(\vartheta_k))^{x_k}, \quad \theta = (\vartheta_1, \ldots, \vartheta_k)^T \in \Theta \subseteq \mathbb{R}^k, \quad \vartheta_i \geq 0,$$

where the summation extends over the set $T = W[f(\theta)]$ and where $a(x)$ is positive, Θ is the *$k-$dimensional parameter space*, which is the region of convergence of the power series of $f(\theta)$. Let $f(\theta),\ g(\vartheta_1), \ldots, g(\vartheta_k)$ be positive, finite and differentiable. The parameter space Θ is the *region of convergence* of the series $f(\theta),\ \theta \in \Theta$.

A $k-$dimensional random vector $X = (X_1, X_2, \ldots, X_k)^T$ with the density function

$$\mathcal{P}\{X = x; \vartheta\} = \begin{cases} \dfrac{a(x)\,(g(\vartheta_1))^{x_1} \cdots (g(\vartheta_k))^{x_k}}{f(\theta)}, & x \in T, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

may be defined to have *multivariate modified power series distribution* (MMPSD) with *range* T and the *series function* $f(\theta)$, $\theta \in \Theta$.

When $g(\vartheta_i) = \exp(\vartheta_i)$, $i = 1, 2, \ldots, k$, (2) can be rewritten as

$$\mathcal{P}\{X = x; \theta\} = \begin{cases} \dfrac{a(x) \exp\left(\sum_{i=1}^{k} \vartheta_i x_i\right)}{f(\theta)}, & x \in T, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

i.e. MMPSD belongs to a *class of discrete multivariate exponential type distributions.* This class includes the multivariate power series distributions (MPSD) considered by Patil (1965) when functions $g(\vartheta_i)$, $i = 1, \ldots, k$; are invertible. On the other hand the class (3) is larger than the class of MPSD, since functions $g(\vartheta_i)$ may include another parameters. For example, *the generalized negative multinomial distribution*

$$\mathcal{P}\{X = x; \theta\} = \frac{m\left(m + \sum_{i=1}^{k} \beta_i x_i - 1\right)! \prod_{i=1}^{k} \left[\vartheta_i (1 - \vartheta_i)^{\beta_i - 1}\right]^{x_i}}{x_1! \cdots x_k! \left(m + \sum_{i=1}^{k} x_i(\beta_i - 1)\right)!} \tag{4}$$

can not be included in the class of MPSD ($0 < \vartheta_i < 1$; $\beta_i \geq 1$; $\vartheta_i \beta_i < 1$).

Let $\mathbb{X} = (X, \ldots, X_n)^T$ be a random sample from (2), $X_i = (X_{i1}, \ldots, X_{ik})^T$. Then the statistic

$$Z_n = (Z_1, Z_2, \ldots, Z_k)^T = \sum_{i=1}^{n} X_i \tag{5}$$

with components

$$Z_j = \sum_{i=1}^{n} X_{ij} \quad (j = 1, 2, \ldots, k)$$

is the *complete sufficient statistic for* θ. The density function of Z_n is

$$\mathcal{P}\{Z_n = z; \theta\} = \begin{cases} \dfrac{b(z, n) \left[g(\vartheta_1)\right]^{z_1} \cdots \left[g(\vartheta_k)\right]^{z_k}}{\left[f(\theta)\right]^n}, & z \in n[T], \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where $z = (z_1, \ldots, z_k)^T$ and

$$b(z, n) = \sum \prod_{i=1}^{n} a(x_{i1}, \ldots, x_{ik})$$

and summation extends over all values of x_{ij} such that

$$\sum_{i=1}^{n} x_{ij} = z_j \quad (j = 1, 2, \ldots, k).$$

Theorem 1 (PATEL (1978))

A necessary and sufficient condition for the parametric function

$$u(\theta) = \prod_{i=1}^{k} (g(\vartheta_i))^{r_i}$$

to be MVU estimable is that $n[T]$ is stable with respect to the vector $r = (r_1, \ldots, r_k)^T$ for some positive integer n. Whenever it exists the MVU $h(Z_n)$ of the parametric function $u(\theta)$ is

$$h(Z_n) = \frac{b(Z_n - r, n)}{b(Z_n, n)}, \quad Z_n \in n[T] + \{r\}. \tag{7}$$

Corollary 1

A parametric function

$$u(\theta) = \prod_{i=1}^{k} (g(\vartheta_i))^{r_i}$$

is not MVU estimable if T is finite.

Example 1

Let $\mathbb{X} = (X_1, \ldots, X_n)^T$ be a random sample from *multinomial distribution* $M_k(m, \theta)$ with parameters m and $\theta = (\vartheta_1, \ldots, \vartheta_k)^T$:

$$\mathcal{P}\{X_i = x; \theta\} = \frac{m! \vartheta_1^{x_1} \cdots \vartheta_k^{x_k}}{x_1! \cdots x_k! \left(m - \sum_{i=1}^{k} x_i\right)! \left(1 + \sum_{i=1}^{k} \vartheta_i\right)^m}, \tag{8}$$

$$0 \le \sum_{i=1}^{k} x_i \le m, \quad 0 \le x_i \le m, \quad \vartheta_i > 0, \quad i = 1, 2, \ldots, k. \text{ If } k = 2,$$

then we have a random sample from binomial distribution, if $k = 3$, then we have a sample from trinomial distribution.

Since T is finite a parametric function

$$u(\theta) = \prod_{i=1}^{k} \vartheta_i^{r_i}$$

is not MVU estimable for the multinomial distribution (7), i.e.

$$\mathrm{E}\ h(Z_n) \not\equiv u(\theta),\ \theta \in \Theta.$$

Corollary 2

A parametric function

$$u(\theta) = \prod_{i=1}^{k} \vartheta_i^{r_i}$$

is not MVU estimable if r_i is fractional or negative for some i, $i = 1, 2, \dots, k$. For example, if X is binomial $B(n, p)$ random variable, then the ratio $p/(1-p)$ is not MVU. (When p is the proportion of males in species, it would mean that the sex ratio is not MVU estimable). It follows from the next results of Kolmogorov (1950) and Hoeffding (1984), related with the problem of estimation of the value of a real-valued function $u(\theta)$, $\theta = (\vartheta_1, \dots, \vartheta_k)^T$, on the basis of a sample $\mathbb{X}$ from the multinomial distribution $M_k(m, \theta)$. According these results the MVUE exists if and only if $u(\theta)$ is a polynomial of degree at most m.

Let

$$u(\theta) = u\left(g(\vartheta_1), \dots, g(\vartheta_k)\right)$$

be a function of $g(\vartheta_1), \dots, g(\vartheta_k)$ and let $u(\theta)[f(\theta)]^n$ admit expansion in the power series of $g(\vartheta_i)$, i.e.

$$u(\theta)[f(\theta)]^n = \sum c(z, n)[g(\vartheta_1)]^{z_1} \cdots [g(\vartheta_k)]^{z_k}, \tag{9}$$

$z = (z_1, \dots, z_k)^T$, z_i are non-negative integer, $z_i \in I$.

Then it holds the

Theorem 2 (PATEL (1978))

A necessary and sufficient condition for $u(\theta), \theta \in \Theta$, to be MVU estimable function on the basis of a random sample $\mathbb{X} = (X_1, \dots, X_n)^T$ of size n from MMPSD (2) is

$$W[u(\theta)(f(\theta))^n] \subseteq W[(f(\theta))^n], \tag{10}$$

where $W[\cdot]$ is an index set (see(1)). Whenever it exists, the MVUE $h(Z_n)$ of $u(\theta)$ is given by

$$
(11) \qquad h(Z_n) = \begin{cases} \dfrac{c(Z_n, n)}{b(Z_n, n)}, & Z_n \in W[u(\theta)[f(\theta)^n], \\ 0, & \text{otherwise.} \end{cases}
$$

Corollary 1 (Patil (1965))

The MVUE $\hat{\sigma}^2$ of the variance

$$
\sigma^2 = \text{Var } h(Z_n)
$$

reduces, if exists, to

$$
(12) \qquad \hat{\sigma}^2 = \widehat{\text{Var}} h(Z_n) = h(Z_n)\left(h(Z_n) - h(Z_n - r)\right),
$$

where $Z_n = X_1 + \cdots + X_n$ is the complete sufficient statistic for θ.

Corollary 2

The density function $\mathcal{P}\{X = x; \theta\}$ of MMPSD is always estimable and the MVUE $\hat{\mathcal{P}}\{X = x; \theta\} = h(x; Z_n)$ is defined by the expression

$$
\begin{aligned}
\hat{\mathcal{P}}\{X = x; \theta\} &= h(x; Z_n) = \mathcal{P}\{X = x | Z_n\} = \\
(13) \qquad &= \begin{cases} \dfrac{a(x) b(Z_n - x, n-1)}{b(Z_n, n)}, & Z_n \in (n-1)[T] + \{x\}, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

The corollary immediately follows from (9) and the condition of unbiasedness.

On can verify that the conditional distribution of X given Z_n is the distribution (5) if and only if X is an element of a random sample $\mathbb{X}$ from (2). It means that MMPSD is characterized through its conditional distribution given Z_n (see, for example, J.Ghosh, B.K.Sihna,B.K.Sihna, 1977).

Example 2

For generalized negative multinomial distribution defined by (4) we have

$$
a(x) = \frac{m\left(m + \sum_{i=1}^{k} \beta_i x_i - 1\right)!}{x_1! \cdots x_k! \left(m + \sum_{i=1}^{k} x_i(\beta_i - 1)\right)!}
$$

and

$$b(Z_n, n) = \frac{mn\left(mn + \sum_{i=1}^{k}\beta_i Z_i - 1\right)!}{Z_1! \cdots Z_k! \left(mn + \sum_{i=1}^{k} Z_i(\beta_i - 1)\right)!}$$

Then the MVUE $h(x; Z_n)$ of $\mathcal{P}\{X = x; \theta\}$ in accordance with (12) is

$$h(x; Z_n) = \hat{\mathcal{P}}(X = x|Z_n) = \prod_{i=1}^{k} \binom{Z_i}{x_i} \frac{m(mn - m)}{mn} \times$$

$$\times \frac{\left(m + \sum_{i=1}^{k}\beta_i x_i - 1\right)! \left(mn - m + \sum_{i=1}^{k}\beta_i(Z_i - x_i) - 1\right)!}{\left(mn - m + \sum_{i=1}^{k}\beta_i(Z_i - x_i) - \sum_{i=1}^{k}(Z_i - x_i)\right)!} \times$$

$$\times \frac{\left(mn + \sum_{i=1}^{k} Z_i(\beta_i - 1)\right)!}{\left(m + \sum_{i=1}^{k} x_i(\beta_i - 1)\right)! \left(mn + \sum_{i=1}^{k}\beta_i Z_i - 1\right)!}.$$

Example 3

Let $\mathbb{X} = (X_1, \ldots, X_n)$ be a sample, a $k-$dimensional random vector X_i has the probability distribution (2). Then the complete sufficient statistic

$$Z_n = \sum_{i=1}^{n} X_i$$

for a parameter θ has the density function

$$g_n(z; \theta) = \mathcal{P}\{Z_n = z; \theta\}$$

defined by (5).

Since random vectors X_i and $\mathcal{U}$,

$$\mathcal{U} = Z_n - X_1 = \sum_{i=2}^{n} X_i,$$

are independent and the Jacobian of the transformation from

$$(X_1, \mathcal{U}) \quad \text{to} \quad (X_1, Z_n)$$

is equal to 1, we may write down the density function $f(x_1, z; \theta)$ of X_1 and Z_n as

$$f(x_1, z; \theta) = \mathcal{P}\{X = x; \theta\} g_{n-1}(z - x_1; \theta).$$

Then the density of conditional distribution of X_1 given $Z_n = z$ is

$$f(x_1|z) = \frac{\mathcal{P}\{X = x; \theta\} g_{n-1}(z - x_1, \theta)}{g_n(z; \theta)} \tag{14}$$

Substituting Z_n instead of z in (13) we get evidently the MVUE (12) of $\mathcal{P}\{X = x; \theta\}$. Performing the same manipulations we may obtain the joint conditional density of $X_1, X_2, \ldots, X_m$ $(m \leq n)$ given $Z_n = z$:

$$f(x_1, \ldots, x_m|z) = \frac{\mathcal{P}\{X_1 = x_1; \theta\} \cdots \mathcal{P}\{X_m = x_m; \theta\} g_{n-m}\left(z - \sum_{i=1}^{m} x_i; \theta\right)}{g_n(z; \theta)} \tag{15}$$

Substituting Z_n instead of z in (14) we get the MVUE for

$$\mathcal{P}\{X_1 = x_1, \ldots, X_m = x_m; \theta\}.$$

Let, for example, $\mathcal{P}\{X = x; \theta\}$ be the density of the multinomial $M_k(m, \theta)$ distribution given by (7), i.e.

$$a(x) = \frac{m!}{x_1! \cdots x_k! \left(m - \sum_{i=1}^{k} x_i\right)!} = \binom{m}{x_1, \ldots, x_k} \tag{16}$$

and

$$g_n(z; \theta) = \frac{b(z, n) \vartheta_1^{z_1} \cdots \vartheta_k^{z_k}}{[f(\theta)]^n}, \tag{17}$$

where

$$b(z, n) = \binom{mn}{z_1, \ldots, z_k}, \tag{18}$$

$$f(\theta) = \sum a(x) \vartheta_1^{x_1} \cdots \vartheta_k^{x_k}, \tag{19}$$

and the summation extends over all non-negative integers x_i such that

$$0 \leq \sum_{i=1}^{k} x_i \leq m.$$

Substituting (15),(16) and (17) into (14) we get the MVUE

$$\hat{\mathcal{P}}\{X_1 = x_1, \ldots, X_j = x_j; \theta\} = \mathcal{P}\{X_1 = x_1, \ldots, X_j = x_j | Z_n\}$$

of $\mathcal{P}\{X_1 = x_1, \ldots, X_j = x_j; \theta\}$, $j \leq n$:

$$\hat{\mathcal{P}}\{X_1 = x_1, \ldots, X_j = x_j; \theta\} = \frac{a(x_1) \cdots a(x_j) b(Z_n - x_1 - \cdots - x_j, n)}{b(Z_n, n)} =$$

$$= \frac{\prod_{i=1}^{j} \binom{m}{x_{i1}, \ldots, x_{ik}} \cdot \binom{m(n-j)}{Z_1 - x_{11} - \cdots - x_{j1}, \ldots, Z_k - x_{1k} - \cdots - x_{jk}}}{\binom{mn}{Z_1, \ldots, Z_k}}$$

2. THE TABLE OF MVUE'S FOR FUNCTIONS $u(\theta)$ OF PARAMETER θ OF THE MULTINOMIAL DISTRIBUTION

Let $\mathbb{X} = (X_1, \ldots, X_n)^T$ be a sample from multinomial $M_k(m, \theta)$ distribution with parameters m and $\theta = (\vartheta_1, \ldots, \vartheta_k)^T$, i.e. the density function of $X_i = (X_{i1}, \ldots, X_{ik})^T$ is given by (7). When a parameter m is known then

$$Z_n = \sum_{i=1}^{n} X_i$$

is the complete and sufficient statistic for a parameter θ and

$$\mathcal{P}\{Z_n = z; \theta\} = \frac{(mn)! \vartheta_1^{z_1} \cdots \vartheta_k^{z_k}}{z_1! \cdots z_k! \left(mn - \sum_{i=1}^{k} z_i\right)! \left(1 + \sum_{i=1}^{k} \vartheta_i\right)^{mn}},$$

$$0 \leq \sum_{i=1}^{k} z_i \leq mn, \quad z_i = 0, 1, \ldots, mn \ (i = 1, 2, \ldots, k).$$

N	$u(\theta)$	MVUE
1	$\mathcal{P}\{X = x;\theta\}$ [12], [20]	$\hat{\mathcal{P}} = \dfrac{\binom{Z_1}{x_1}\cdots\binom{Z_k}{x_k}\binom{mn-\sum_{i=1}^{k} Z_i}{m-\sum_{i=1}^{k} x_i}}{\binom{mn}{m}}$
2	Var $\hat{\mathcal{P}}$ [20]	$\dfrac{\binom{Z_1}{x_1}\cdots\binom{Z_k}{x_k}\binom{mn+\sum_{i=1}^{k} Z_i}{m+\sum_{i=1}^{k} x_i}^2}{\binom{mn}{m}^2} - \dfrac{\binom{Z_1}{x_1}\cdots\binom{Z_k}{x_k}\binom{mn-\sum_{i=1}^{k} Z_i}{m-\sum_{i=1}^{k} x_i}}{\binom{mn}{m}} \times \dfrac{\binom{Z_1-x_1}{x_1}\cdots\binom{Z_k-x_k}{x_k}\binom{m(n-1)-\sum_{i=1}^{k} Z_i+\sum_{i=1}^{k} x_i}{m-\sum_{i=1}^{k} x_i}}{\binom{m(n-1)}{m}}$
3	$\mathcal{P}\{X_1 = x_1, \dots, X_\ell = x_\ell;\theta)\}$ $\ell \le n$	$\hat{\mathcal{P}} = \dfrac{\prod_{j=1}^{\ell}\binom{m}{x_{j1},\dots,x_{jk}}}{\binom{mn}{Z_1,\dots,Z_k}} \times \binom{m(n-\ell)}{Z_1-x_{11}-\dots-x_{\ell 1},\dots,Z_k-x_{1k}-\dots-x_{\ell k}}$, where $\binom{m}{x_1,\dots,x_k} = \dfrac{m!}{x_1!\cdots x_k!\left(m-\sum_{i=1}^{k} x_i\right)!}$
4	$\mathrm{E}\,X = \dfrac{m\theta}{1+\sum_{i=1}^{k}\vartheta_i}$	$\dfrac{1}{n}Z_n$

Remark (TRINOMIAL DISTRIBUTION)

As it was remarked by Patil (1965) the multivariate distribution given by (7) with the series function

$$f(\theta) = (1 + \vartheta_1 + \vartheta_2 + \cdots + \vartheta_k)^m,$$

can be expressed in its traditional form:

$$\mathcal{P}\{X = x; \theta\} = \frac{m! p_1! p_2! \cdots p_k!}{x_1! x_2! \cdots x_k! \left(m - \sum_{i=1} x_i\right)!} \cdot \left(1 - \sum_{i=1}^{k} p_i\right)^{m - \sum_{i=1}^{k} x_i},$$

where

$$p = (p_1, \ldots, p_k)^T, \quad p_i = \frac{\vartheta_i}{1 + \sum_{i=1}^{k} \vartheta_i}, \quad 0 < \sum_{i=1}^{k} p_i < 1,\ 0 \le \sum_{i=1}^{k} x_i \le m.$$

Let suppose that $k = 3$. In this case we have $X = (X_1, X_2, X_3)^T$ and

$$\mathcal{P}\{X = x; \theta\} = \begin{pmatrix} & m & \\ x_1 & & x_2 \end{pmatrix} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{m - x_1 - x_2},$$

$0 \le x_i \le m;\ x_1 + x_2 \le m;\ x_i \in I$, where

$$\begin{pmatrix} & m & \\ x_1 & & x_2 \end{pmatrix} = \frac{m!}{x_1! x_2! (m - x_1 - x_2)!}.$$

If $\mathbb{X} = (X_1, \ldots, X_n)^T$ is a sample from the trinomial distribution, then the sufficient statistic for p is $Z_n = X_1 + \cdots + X_n$ and

$$\mathcal{P}\{Z_1 = z_1, Z_2 = z_2; \theta\} = \begin{pmatrix} mn \\ z_1, z_2 \end{pmatrix} p_1^{z_1} p_2^{z_2} (1 - p_1 - p_2)^{mn - z_1 - z_2},$$

$0 \le z_i \le mn;\ 0 \le z_1 + z_2 \le mn.$

THE TABLE OF MVUE'S FOR FUNCTIONS $u(p)$ OF PARAMETER p OF THE TRINOMIAL DISTRIBUTION.

N	$u(p)$	MVUE
1	$\mathcal{P}\{X = x; p\}$	$\hat{\mathcal{P}} = \dfrac{\binom{m}{x_1, x_2}\binom{mn-m}{Z_1 - x_1, Z_2 - x_2}}{\binom{mn}{Z_1, Z_2}}$
2	$\mathcal{P}^2\{X = x; p\}$	$\hat{\mathcal{P}}^2 = \dfrac{\binom{m}{x_1, x_2}^2\binom{mn-2m}{Z_1 - 2x_1, Z_2 - 2x_2}}{\binom{mn}{Z_1, Z_2}}$
3	$p_i, \quad i = 1,2$	$\hat{p}_i = \dfrac{1}{mn} \cdot Z_i$
4	$p_1 p_2$	$\dfrac{Z_1 Z_2}{mn(mn-1)}$
5	$\dfrac{1}{p_i}, \quad i = 1,2$	$\dfrac{mn+1}{Z_i+1}$
6	$\dfrac{p_1^i p_2^j}{(1-p_1-p_2)^r}$ $i, j, r \in \mathbb{Z}$	$\dfrac{\binom{mn+r-i-j}{Z_1 - i, Z_2 - j}}{\binom{mn}{Z_1, Z_2}}$ $mn + r - i - j > 0$

3. THE TABLE OF MVUE'S FOR FUNCTIONS $u(\theta)$ OF PARAMETER θ OF THE NEGATIVE MULTINOMIAL DISTRIBUTION

Let $\mathbb{X} = (X_1, \ldots, X_n)^T$ be a sample from a *negative multinomial* $NM_k(m, \theta)$ distribution with parameters m and $\theta = (\vartheta_1, \ldots, \vartheta_k)^T$:

$$\mathcal{P}\{X_i = x; \theta\} = \frac{\Gamma\left(m + \sum_{i=1}^{k} x_i\right)}{\Gamma(m)\prod_{i=1}^{k} x_i} \cdot \vartheta_1^{x_1} \cdots \vartheta_k^{x_k} \left(1 - \sum_{i=1}^{k} \vartheta_i\right)^m,$$

$$X_i = (X_{i1}, \ldots, X_{ik})^T,\ x = (x_1, \ldots, x_k)^T,\ x_i = 0, 1, 2, \ldots;\ \vartheta_i > 0,\ 0 < \sum_{i=0}^{k} \vartheta_i < 1.$$

When a parameter m is known, then

$$Z_n = \sum_{i=1}^{n} X_i$$

is the complete and sufficient statistic for a parameter θ, and

$$\mathcal{P}\{Z_n = z; \theta\} = \frac{\Gamma\left(mn + \sum_{i=1}^{k} z_i\right)}{\Gamma(mn)\prod_{i=1}^{k} z_i!} \cdot \vartheta_1^{z_1} \cdots \vartheta_k^{z_k} \left(1 - \sum_{i=1}^{k} \vartheta_i\right)^{mn},$$

$_i = 0, 1, 2, \ldots;\ i = 1, 2, \ldots, k.$

N	$u(\theta)$	MVUE
1	$\mathcal{P}\{X = x;\theta\}$ [23], [20]	$\hat{\mathcal{P}} = \begin{cases} \dfrac{m\binom{mn-1}{m}\prod_{i=1}^{k}\binom{Z_i}{x_i}}{\left(m+\sum_{i=1}^{k}x_i\right)\cdot\binom{mn+\sum_{i=1}^{k}Z_i-1}{m+\sum_{i=1}^{k}x_i}}, & Z_i \geq x_i \\ 0, & \text{otherwise.} \end{cases}$
2	Var $\hat{\mathcal{P}}$ [20]	$\dfrac{m^2\binom{mn-1}{m}^2\prod_{i=1}^{k}\binom{Z_i}{x_i}^2}{\left(m+\sum_{i=1}^{k}x_i\right)^2\binom{mn+\sum_{i=1}^{k}Z_i-1}{m+\sum_{i=1}^{k}x_i}} - \dfrac{m^2\binom{mn-1}{m}\prod_{i=1}^{k}\binom{Z_i}{x_i}}{\left(m+\sum_{i=1}^{k}x_i\right)^2\binom{mn+\sum_{i=1}^{k}Z_i-1}{m+\sum_{i=1}^{k}x_i}} \times \dfrac{\binom{m(n-1)-1}{m}\prod_{i=1}^{k}\binom{Z_i-x_i}{x_i}}{\binom{m(n-1)+\sum_{i=1}^{k}Z_i-\sum_{i=1}^{k}x_i-1}{m+\sum_{i=1}^{k}x_i}}$

N	$u(\theta)$	MVUE
3	$\prod_{i=1}^{k} \vartheta_i^{r_i}$, $r = 0, 1, 2, \dots$ [23], [20]	$\hat{u} = \dfrac{\left(\sum_{i=1}^{k} Z_i - \sum_{i=1}^{k} r_i + mn - 1\right)!}{\left(\sum_{i=1}^{k} Z_i + mn - 1\right)!} \prod_{i=1}^{k} \dfrac{Z_i!}{(Z_i - r_i)!}$, $Z_i \geq r_i, \quad i = 1, 2, \dots, k$
4	Var $\hat{u}$ [20]	$\dfrac{\left[\left(\sum_{i=1}^{k} Z_i - \sum_{i=1}^{k} r_i + mn - 1\right)!\right]^2}{\left[\left(\sum_{i=1}^{k} Z_i + mn - 1\right)!\right]^2} \prod_{i=1}^{k} \dfrac{(Z_i!)^2}{[(Z_i - r_i)!]^2} - \dfrac{\left(\sum_{i=1}^{k} Z_i - 2\sum_{i=1}^{k} r_i + mn - 1\right)!}{\left(\sum_{i=1}^{k} Z_i + mn - 1\right)!} \prod_{i=1}^{k} \dfrac{Z_i!}{(Z_i - 2r_i)!}$
5	E X	$\dfrac{1}{n} \cdot Z_n$
6	θ	$\dfrac{Z_n}{\sum_{i=1}^{k} Z_i + mn - 1}$
7	$\mathcal{P}(X_1 = x_1, \dots, X_\ell = x_\ell; \theta)$ $\ell \leq n$	$\dfrac{\Gamma(mn) \prod_{i=1}^{k} Z_i \Gamma\left[m(n-\ell) + \sum_{i=1}^{k} (Z_i - x_{1i} - \dots - x_{\ell i})\right]}{[\Gamma(m)]^\ell \, \Gamma[m(n-\ell)] \prod_{i=1}^{k} (Z_i - x_{1i} - \dots - x_{\ell i})!} \times \dfrac{\prod_{i=1}^{k} \Gamma\left(m + \sum_{j=1}^{\ell}\sum_{i=1}^{k} x_{ji}\right)}{\Gamma\left(mn + \sum_{i=1}^{k} Z_i\right) \prod_{j=1}^{\ell}\prod_{i=1}^{k} x_{ji}!}$

4. LEFT TRUNCATED MULTIVARIATE MODIFIED POWER SERIES DISTRIBUTION

Let $\alpha \in \mathcal{T} = \{x = (x_1, \dots, x_k)^T : x_i \in I,\ i = 1, 2, \dots, k\} \subseteq I * I * \dots * I, I$ being the set of non-negative integers, and let

$$\mathcal{S}_{\alpha_j} = \{x : x \in I, x \geq \alpha_j\},\ j = 1, 2, \dots, k,$$

and $\mathcal{S} = \{(x_1, \dots, x_k) : x_j \in \mathcal{S}_{\alpha_j};\ j = 1, 2, \dots, k\} \subseteq \mathcal{S}_{\alpha_1} \times \mathcal{S}_{\alpha_2} \times \dots \times \mathcal{S}_{\alpha_k}$.

Let also

$$f(\alpha, \theta) = \sum a(x)(g(\vartheta_1))^{x_1} \cdots (g(\vartheta_k))^{x_k}, \tag{20}$$

where the summation extends over the set $\mathcal{S}$, $a(x) > 0$, and $\theta \in \Theta$ $k-$dimensional parametric space. Let $f(\alpha, \theta), g(\vartheta_1), \dots, g(\vartheta_k)$ be positive, finite and differentiable functions.

A $k-$dimensional random vector $X = (X_1, X_2, \dots, X_k)^T$ with density function

$$\mathcal{P}\{X = x; \alpha, \theta\} = \begin{cases} \dfrac{a(x)(g(\vartheta_1))^{x_1} \cdots (g(\vartheta_k))^{x_k}}{f(\alpha, \theta)}, & x \in \mathcal{S}, \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

may be defined to have *left truncated multivariate modified power series distribution* (LTMMPSD) with range $W[f(\alpha, \theta)]$:

$$W[f(\alpha, \theta)] = \{x = (x_1, \dots, x_k)^T : x_j \geq \alpha_j,\ j = 1, 2, \dots, k\},$$

and a series function $f(\alpha, \theta)$ defined in (18).

Let $\mathbb{X} = (X_1, \dots, X_n)^T$ be a random sample of size n from LTMMPSD (19), $X_j = (X_{j1}, X_{j2}, \dots, X_{jk})^T$. Denote

$$Z = (Z_1, \dots, Z_k)^T \quad \text{and} \quad Y = (Y_1, \dots, Y_k)^T,$$

where

$$Z_i = \sum_{j=1}^{n} X_{ji}, \qquad Y_i = \min(X_{1i}, X_{2i}, \dots, X_{ni}), \quad i = 1, 2, \dots, k.$$

If $\alpha = (\alpha_1, \dots, \alpha_k)^T$ and $\theta = (\vartheta_1, \dots, \vartheta_k)^T$ are both unknown, then

$$(Y^T, Z^T)^T = (Y_1, \dots, Y_k; Z_1, \dots, Z_k)^T$$

is the complete sufficient statistic for (α, θ).

Theorem 3 (PATEL (1979))

If $\mathbb{X} = (X_1, \ldots, X_n)^T$ be a random sample of size n from LTMMPSD (19) then $(Y^T, Z^T)^T$ has the density

$$\mathcal{P}\{Y = y, Z = z; \alpha, \theta\} = \begin{cases} \dfrac{b(y,z,n)\,[g(\vartheta_1)]^{z_1} \cdots [g(\vartheta_k)]^{z_k}}{[f(\alpha,\theta)]^n}, & z \in W\{[f(\alpha,\vartheta)]^n\}, \\ & y \in W[f(\alpha,\theta)], \\ 0, \text{ otherwise}, \end{cases}$$

(22)

where $W[f(\alpha, \theta)] = \mathcal{S}$ is the index set of $f(\alpha, \theta)$,

$$\begin{aligned} b(y,z,n) &= c(y,z,n) - c(y+1,z,n), \\ c(y,z,n) &= \sum \prod_{j=1}^{n} a(x_{j1}, \ldots, x_{jk}) \end{aligned}$$

and summation extents over all n–tuples $(x_{1i}, x_{2i}, \ldots, x_{ni})$ for $i = 1, 2, \ldots, k$ of integers $x_{ji} > y_i$ with

$$\sum_{j=1}^{n} x_{ji} = z_i, \quad i = 1, 2, \ldots, k.$$

Denote $\eta = (\eta_1, \ldots, \eta_k)^T$, where $\eta_i = g(\vartheta_i)$, $i = 1, 2, \ldots, k$.

Theorem 4 (PATEL (1979))

A real valued parametric function $u(\alpha, \eta)$, α And η being unknown, of the LTMMPSD (19) is MVU estimable iff for every $\alpha \in \mathcal{T}$, $u(\alpha, \eta)[f(\alpha, \theta)]^n$ admits a power series expansion in $\eta_1, \eta_2, \ldots, \eta_k$, i.e.

$$u(\alpha,\eta)[f(\alpha,\theta)]^n = \sum d(\alpha,z,n)\eta_1^{z_1} \cdots eta_k^{z_k}, z \in W\{u(\alpha,\eta)[f(\alpha,\theta)]^n\}$$

and

$$W\{u(\alpha,\eta)[f(\alpha,\theta)]\} \subseteq W\{[f(\alpha,\theta)]^n\}.$$

The MVU estimator $h(Y, Z, n)$ of $u(\alpha, \eta)$ if exists, is given by the expression

$$h(Y,Z,n) = \begin{cases} \dfrac{d(Y,Z,n) - d(Y+1,Z,n)}{b(Y,Z,n)}, & Z \in W\{u(Y,\eta)[f(\alpha,\vartheta)]^n\}, \\ 0, & \text{otherwise.} \end{cases}$$

(23)

Remark 1

The parametric function

$$u(\theta) = \prod_{i=1}^{k} (g(\vartheta_i))^{r_i}$$

is not MVU estimable if $\mathcal{T}$ is finite or if r_i is fractional or non-negative for some i.

Corollary 1

A necessary and suffficient condition for the parametric function

(24) $$u(\theta) = \prod_{i=1}^{k} (g(\vartheta_i))^{r_i},$$

$r_1, r_2, \ldots, r_k$ being non-negative integers, to be MVU estimable is that

$$W\{[f(\alpha, \theta)]^n\} + \{r\} \subseteq W\{[f(\alpha, \theta)]^n\}.$$

Whenever it exists the MVUE $h(Y, Z, n)$ of $u(\theta)$ is given by the expression

(25) $$h(Y, Z, n) = \begin{cases} \dfrac{b(Y, Z - r, n)}{b(Y, Z, n)}, & Z \in W\{[f(\alpha, \vartheta)]^n\} + \{r\}, \\ 0, & \text{otherwise.} \end{cases}$$

In particular if $r = 1 = (1, 1, \ldots, 1)^T$, then from (22) and (23) it follows that the MVUE for

(26) $$u(\theta) = \prod_{i=1}^{k} g(\vartheta_i)$$

is given by formula

(27) $$h(Y, Z, n) = \begin{cases} \dfrac{b(Y, Z - 1, n)}{b(Y, Z, n)}, & Z \in W\{[f(\alpha, \vartheta)]^n + \{1\}, \\ 0, & \text{otherwise.} \end{cases}$$

One can verify that the MVUE of the variance Var $h(Y, Z - 1, n)$ of the statistic (25) is given by formula

(28) $$\widehat{\text{Var}}\, h(Y, Z - 1, n) = h(Y, Z, n)[h(Y, Z, n) - h(Y, Z - 1, n)].$$

Example 5

Let $u(\theta) = [g(\vartheta_i)]^{r_i}$. In this case, according to (23), the MVUE for $u(\theta)$ is

$$\hat{\eta}(Z_i - r_i) = \frac{b(Y, Z - r, n)}{b(Y, Z, n)}, \tag{29}$$

where $r = (0, \ldots, 0, r_i, 0, \ldots, 0)^T$ is the vector with component r_i at i–th place and zero elsewhere. The MVUE of the variance Var $\hat{\eta}(Z_i - r_i)$ of the statistic (27) is

$$\widehat{\text{Var}}\ \hat{\eta}(Z_i - r_i) = \hat{\eta}(Z_i - r_i) - \hat{\eta}(Z_i - 2r_i). \tag{30}$$

In particular, if $r_i = 1$, then we obtain that the MVUE for $g(\vartheta_i)$ is the statistic $\hat{\eta}(Z_i - 1)$ and the MVUE of its variance is

$$\widehat{\text{Var}}\ \hat{\eta}(Z_i - 1) = \hat{\eta}(Z_i - 1) - \hat{\eta}(Z_i - 2). \tag{31}$$

Corollary 2

The MVU estimator $\hat{\alpha}(Y, Z, n)$ of a parametric function

$$u(\alpha) = \prod_{i=1}^{k} \alpha_i^{r_i} \tag{32}$$

for non-negative integers $r_i (i = 1, 2, \ldots, k)$ always exists and is

$$\hat{\alpha}(Y, Z, n) = Y_1^{r_1} \cdots Y_k^{r_k} - \frac{c(Y + 1, Z, n)}{b(Y, Z, n)} e(Y) \tag{33}$$

where

$$e(Y) = \Sigma^* \prod_{i=1}^{k} \binom{r_i}{j_i} Y_i^{j_i}, \tag{34}$$

Σ^* denotes summation over all non-negative $j_1, \ldots, j_k$ except $j_i = r_i$, $i = 1, 2, \ldots, k$. From (30)-(32) it follows that the MVUE for

$$u(\alpha) = \alpha_i^{r_i}$$

is given by formula:

$$\hat{\alpha}(Y, Z, n) = Y_i^{r_i} - \frac{c(Y + 1, Z, n)}{b(Y, Z, n)} \left[(1 + Y_i)^{r_i} - Y_i^{r_i}\right], \tag{35}$$

and the MVUE of its variance is

$$\widehat{\text{Var}}\ \hat{\alpha}(Y, Z, n) = \frac{c(Y + 1, Z, n)}{b(Y, Z, n)} \left[1 + \frac{c(Y + 1, Z, n)}{b(Y, Z, n)}\right] \left[(1 + Y_i)^{r_i} - Y_i^{r_i}\right] \tag{36}$$

From (33) and (34) we can find that the MVUE $\hat{\alpha}_i$ for α_i is

$$\hat{\alpha}_i = Y_i - \frac{c(Y+1, Z, n)}{b(Y, Z, n)} \tag{37}$$

and the MVUE of its variance is

$$\widehat{\text{Var}}\ \hat{\alpha}_i = \frac{c(Y+1, Z, n)}{b(Y, Z, n)} \left[1 + \frac{c(Y+1, Z, n)}{b(Y, Z, n)}\right] \tag{38}$$

By the same way we can find the MVUE for the function

$$u(\alpha) = \prod_{i=1}^{k} \alpha_i^r = \left(\prod_{i=1}^{k} \alpha_i\right)^r, \tag{39}$$

obtained from (30) in the case $r_1 = r_2 = \cdots = r_k = r$. From (31) we find its MVUE

$$\hat{\alpha}(Y, Z, n) = \left[1 + \frac{c(Y+1, Z, n)}{b(Y, Z, n)}\right] \mathcal{A}^r - \frac{c(Y+1, Z, n)}{b(Y, Z, n)} \mathcal{B}^r, \tag{40}$$

where

$$\mathcal{A} = \prod_{i=1}^{k} Y_i \quad \text{and} \quad \mathcal{B} = \prod_{i=1}^{k} (1 + Y_i).$$

The MVUE of the variance of the statistic (37) is

$$\widehat{\text{Var}}\ \hat{\alpha}(Y, Z, n) = \frac{c(Y+1, Z, n)}{b(Y, Z, n)} \left[1 + \frac{c(Y+1, Z, n)}{b(Y, Z, n)}\right] (\mathcal{B}^r - \mathcal{A}^r)^2 \tag{41}$$

Now we can obtain easily the MVUE's for $u(\alpha) = \alpha_1 \alpha_2 \cdots \alpha_k$ and for its variance, for which we need to put $r = 1$ in (37) and (38).

Corollary 3

In general $\mathcal{P}\{X = x; \alpha, \theta\}$ defined in (19) is not MVU estimable for an arbitrary point $\alpha \in \mathcal{T}$. Nevertheless, if $\mathcal{T}$ is finite then $\mathcal{P}\{X = \alpha + r; \alpha, \theta\}$ is MVU estimable and its estimator is given by
$\hat{\mathcal{P}}\{X = \alpha + r; \alpha, \theta\} =$

$$\begin{cases} \dfrac{a(Y+r)c(Y+r, Z-Y-r, n-1) - a(Y+r+1)c(Y+r+1, Z-Y-r, n-1)}{b(Y, Z, n)}, \\ \qquad\qquad \text{if } Z_i > n(Y_i+1) + r_i \text{ for every } i = 1, 2, \ldots, k, \\ 0, \text{otherwise.} \end{cases}$$

5. A CHI-SQUARE GOODNESS-OF-FIT TEST FOR MULTIVARIATE DISCRETE DISTRIBUTIONS

Let $\mathbb{X} = (X_1, \ldots, X_n)^T$ be a sample, $X_i \in \mathbb{R}^k$. We consider the problem of constructing a chi-square criteria for testing the composite hypothesis H_0 according to which the law of X_i belongs to the family of multivariate discrete distributions, given by (2) or (19). We denote Z_n the complete and sufficient statistic for θ. Let Ω denote the sample space of the vector X_i given Z_n. We pertition Ω into $r(r > 1)$ mutually disjoint subsets $\Omega_1, \ldots, \Omega_r$. Let $\nu = (\nu_1, \ldots, \nu_r)^T$ the frequency vector arising from groupping the data $X_1, \ldots, X_n$ over the sets $\Omega_1, \ldots, \Omega_r$, $p = (p_1, \ldots, p_r)^T$, where

$$p_i = \mathcal{P}\{X_1 \in \Omega_i | Z_n\},$$

and let

$$V = \frac{1}{\sqrt{n}}(\nu - np), \; \Sigma = \mathrm{E}\ \{VV^T | Z_n\}.$$

The rank of the covariance matrix Σ is $r-1$. Let Σ^- be a general invers matrix of Σ. Using the results of Voinov and Nikulin (see, for example, (1989,1990) we obtain the next

Theorem 5

If the size n of the sample tends to the infinity then under H_0 a statistic $Y_n^2 = V^T \Sigma^- V$ is distributed asymptoticaly as χ^2_{r-1}:

$$\lim_{n\to\infty} \mathcal{P}\{Y_n \le x | \mathrm{H}_0\} = \mathcal{P}\{\chi^2_{r-1} \le x\}.$$

ACKNOWLEDGMENT

This is a revised version of our talk given in 1992 in The University of Barcelona, Steklov Mathematical Institute in St. Petersburg and The University Paris-5. We thank Professors C.Cuadras, C.Huber, S.Kotz, T.Smith for their advice and encouragement, and the referee for his careful reading of the submitted work.

REFERENCES

[1] **Chernoff H., Lehmann E.L.** (1954). "The use of maximum likelihood estimates in χ^2 tests for goodness of fit". *Ann. Math. Statist.*, **25**, 579–586.

[2] **Dzhaparidze K., Nikulin M.** (1974). "On a modification of the standard statistic of Pearson". *Theory of probability and its applications*, **19.2**, 851–853.

[3] **Ghosh J.K., Sinha B.K., Sinha B.K.** (1977). "Multivariate Power Series Distributions and Neyman's Properties for Multinomials". *Journal of Multivariate Analysis*, **7**, 397–408.

[4] **Ghurye F.G., Olkin I.** (1969). "Unbiased estimation of some multivariate probability densities and related functions". *Ann. Math. Statist.*, **40**, #1, 1261–1271.

[5] **Halmos P.R.** (1946). "The Theory of Unbiased estimation". *Ann. Math., Statist.*, **17**, 34–43.

[6] **Hoeffding W.** (1983). *Unbiased Range Preserving Estimators.* In: A Festschrift for Erich I. Lehmann. California, Wadsworth.

[7] **Hoeffding W.** (1984). "Range Preserving Unbiased Estimators in the Multinomial Case". *JASA*, **79**, #387, 712–714.

[8] **Johnson N.L., Kotz S.** (1969). *Distributions in Statistics: Discrete Distributions.* New York: Wiley.

[9] **Joshi S.W., Park C.J.** (1974). "Minimum variance unbiased estimation for truncated power series distrubutions". *Sankhyā*, **A 36**, 305–314.

[10] **Kolmogorov A.N.** (1950). "Unbiased estimates" (in russian). Izv. Acad. Nauk SSSR, *Ser. Mat.*, **14**, 303–326 (Amer. Math. Soc. Translation, 1953, #98).

[11] **Lehmann E.L.** (1951). "A General Concept of Unbiasedness". *Ann. Math. Statist.*, **22**, 587–592.

[12] **Lumelsky Ya., Sapoznikov P.N.** (1969). "Unbiased estimators of density functions". *Theory of probability and its applications*, **14.2**, 372–380.

[13] **Moore D.S., Spruill M.C.** (1975). "Unified large-sample theory of general chi-squared statistics for test of fit". *Ann. Statist.*, **3.3**, 599–616.

[14] **Neyman J., Scott E.L.** (1960). "Correction for the bias introduced by a transformation of variables". *Ann. Math. Statist.*, **31**, 643–655.

[15] **Nikulin M.** (1973). "Chi-square test for continuous distributions with shift and scale parameters". *Theory of Probability and its Applications*, **18**, 559–568.

[16] **Nikulin M.** (1973). "On a chi-square test for continuous distributions". *Theory of Probability and its Applications*, **18**, 638–639.

[17] **Nikulin M.S., Voinov V.G.** (1989). "A chi-square goodness-of-fit test for exponential distributions of the first order". *Springer-Verlag Lecture Notes in Mathematics*, **1412**, 248–267.

[18] **Nikulin M.S.** (1991). "Some recent results on chi-squared tests". *Queen's Papers in Pure and Applied Mathematics*, **86**, Queen's University, Kingston, Ontario, Canada, 74p.

[19] **Nikulin M.S., Greenwood P.E.** (1990). "A guide to chi-squared testing". *Technical Report*, **94**, Departement of Statistics, University of British Columbia, Vancouver, Canada, 199p.

[20] **Patel S.R.** (1978). "Minimum variance unbiased estimation of multivariate modified power series distribution". *Metrica*, **25**, 155–161.

[21] **Patel S.R.** (1979). "Minimum variance unbiased estimation of left truncated multivariate modified power series distribution". *Metrica*, **26**, 87–94.

[22] **Patil G.P.** (1965). *On multivariate generalized power series distribution and its application to the multinomial and negative multinomial.* In: Classical and Contagious discrete distributions, Calcutta, 183–194.

[23] **Rao K.C., Robson D.S.** (1974). "A chi square statistic for goodness-of-fit test within exponential family". *Commun. Statists.*, **3.12**, 1139–1153.

[24] **Rao C.R.** (1949). "Sufficient statistics and minimum variance estimates". *Proc. Camb. Phil. Soc.*, **45**, 213–218.

[25] **Sibuya M., Yoshimura I., Shimizu R.** (1964). "Negative multinomial distribution". *Ann. Inst. Statist. Math.*, **26**, 409–424.

[26] **Voinov V.G., Nikulin M.S.** (1993). *Unbiased Estimators and their applications, v. 1: Univariate Case.* Kluwer Academic Publisher, Holland (translation in English from Russian).

[27] **Voinov V.G., Nikulin M.S.** (1990). *A chi-squared goodness-of-fit tests for one- and multidimentional distributions.* In: Studies in Mathematical Statistics, 9. (Zapiski Nauchnyckh Seminarov LOMI im. V.A.Steklova AN SSSR, 184), 62–79. The English thranslation of this paper will appear in *Journal of Soviet Mathematics* in 1993.