Ann. Inst. Statist. Math.
38 (1986), Part A, 45-56

# MULTIPARAMETER ESTIMATION FOR SOME MULTIVARIATE DISCRETE DISTRIBUTIONS WITH POSSIBLY DEPENDENT COMPONENTS

KAM-WAH TSUI

(Received June 4, 1984)


## Summary

In multiparameter estimation for multivariate discrete distributions with infinite support, inadmissibility problems in situations where the multivariate probability distribution function is *not* a product of the one-dimensional marginal probability distribution functions have previously been unexplored. This paper examines the inadmissibility problem in some of these situations. Special attention is given to estimating the mean of a negative multinomial distribution. In estimating the mean vector, certain Clevenson-Zidek type estimators are shown to be uniformly better than the usual estimator under a large class of generally scaled squared loss functions. Some of the results are generalized to other multivariate discrete distributions and to situations where several independent negative multinomial distributions are considered.


## 1. Introduction

There has been considerable interest in the problem of multiparameter estimation for one-parameter discrete exponential families. Particular interest has been given to simultaneous estimation of the means of several Poisson random variables. Various improved Poisson means estimators have been shown to dominate the maximum likelihood estimator (MLE) uniformly under quadratic type losses. These results can be found in Peng [7], Clevenson and Zidek [2], Tsui [9], Hwang [5], Tsui and Press [11], and Ghosh, Hwang and Tsui [3], among others. The last paper also provides a review of the research in the general area of simultaneous estimation for one-parameter exponential families, including the Poisson distributions and the negative binomial distribu-

Key words and phrases: Dependent, difference inequality, dominate, estimator, loss function, multivariate discrete distribution, negative multinomial.

tions as special cases.

In all of the previous research, the underlying marginal distributions are assumed to be independent of one another. This paper examines the simultaneous estimation problem in some situations where the underlying marginal distributions are dependent.

Let $X=(X_1,\cdots,X_n)$, $n\geqq 2$, be a random vector which has a negative multinomial distribution (also called a multivariate negative binomial distribution), $\mathrm{NM}(k,p)$, with probability distribution function equal to

$$P(X_1=x_1,\cdots,X_n=x_n\,|\,k,p)=\frac{\Gamma(k+Z)}{\Gamma(k)\prod_{i=1}^{n}x_i!}q^k\prod_{i=1}^{n}p_i^{x_i}, \tag{1.1}$$

where $k>0$, $Z=\sum_{i=1}^{n}x_i$, $p=(p_1,\cdots,p_n)$, $p_i>0$, $x_i=0,1,2,\cdots$, for $i=1,\cdots,n$, and $q=1-\sum_{i=1}^{n}p_i>0$. Note that the $X_i$'s are positively correlated and the mean of $X_i$ is $\theta_i=kp_i/q$.

Bates and Neyman [1] provide an application of the negative multinomial distribution in models representing accident proneness; other applications are described in Neyman [6]. Sibuya, Yoshimura and Shimizu [8] provide a comprehensive review of the properties of the negative multinomial distribution. In particular, many situations where such a distribution arises are described. For example, suppose that, given $\alpha>0$, $X_1,\cdots,X_n$ are independent and have Poisson distributions with means $\alpha\lambda_1,\cdots,\alpha\lambda_n$, respectively; suppose further that $\alpha$ has a gamma distribution. Then the unconditional joint distribution of the $X_i$'s (after integrating out $\alpha$) is a negative multinomial distribution. Thus, the negative multinomial distribution can be viewed as an $n$-variate "dependent Poisson distribution". Note that in (1.1), if $\theta_i=kp_i/q$, $i=1,\cdots,n$, are kept fixed while letting $k\rightarrow\infty$, then the random variables $X_1,\cdots,X_n$ become independent and have Poisson distributions with means $\theta_1,\cdots,\theta_n$, respectively. Therefore, a negative multinomial distribution $\mathrm{NM}(k,p)$, with a large value of $k$, is close to the independent Poisson distribution case.

Suppose that $X=(X_1,\cdots,X_n)$ has a negative multinomial distribution $\mathrm{NM}(k,p)$, given in (1.1). One problem of interest is to estimate simultaneously the means $\theta_1,\cdots,\theta_n$ of $X_1,\cdots,X_n$, respectively. The maximum likelihood estimation (MLE) of $\theta=(\theta_1,\cdots,\theta_n)$ is $\hat{\theta}^{\circ}(X)=X$. For this case of *dependent* $X_i$'s, it will be shown that $\hat{\theta}^{\circ}(X)$ is uniformly dominated by a class of simultaneous means estimators $\hat{\theta}(X)=(\hat{\theta}_1(X),\cdots,\hat{\theta}_n(X))$ under a large class of loss functions of the form

$$L_K(\theta,\hat{\theta}(X))=\sum_{i=1}^{n}K(\theta_i)(\theta_i-\hat{\theta}_i(X))^2/\theta_i\,, \tag{1.2}$$

where $K(\cdot)$ is a nonincreasing function and $K(\cdot)>0$.

Some of the improved estimators $\hat{\theta}$ have the form

$$\hat{\theta}(X)=\left(1-\frac{c}{Z+b}\right)X \tag{1.3}$$

where $b\geqq(n-1)$, $0\leqq c\leqq 2(n-1)$, and $Z=\sum_{i=1}^{n} X_i$. The estimators given by (1.3) are proposed in Clevenson and Zidek [2] for the independent Poisson distributions case. A proof of a result more general than this is provided in Section 2. The proof, which is similar to that of Theorem 2.1 of Clevenson and Zidek [2], does not use the difference inequality method commonly used in simultaneous estimation problems for discrete distributions.

What if the simultaneous means estimation problem mentioned above involves several independent negative multinomial distributions? Using a difference inequality to be developed, we show in Theorem 3 that the usual estimator of the means is dominated by a large class of the estimators under the normalized squared error loss $L_1$, a special case of (1.2) with $K(\cdot)\equiv 1$. Tsui's [10] result concerning several independent negative binomial distributions is a special case of the new result. When only one negative multinomial distribution is considered, this difference inequality proof provides an alternative proof of the result in Section 2 under $L_1$. The difference inequality proof, however, cannot establish the dominance result under the general loss (1.2). Further generalization of the results and concluding remarks are given in Section 4.

## 2. One negative multinomial case

Suppose $X=(X_1,\cdots,X_n)$ has the negative multinomial distribution NM$(k,p)$, given in (1.1). Theorem 1 shows that in this case of dependent $X_i$'s, the usual estimator, $\hat{\theta}^{\circ}(X)=X$, of the mean vector $E(X)=(\theta_1,\cdots,\theta_n)$, is dominated uniformly by a class of estimators under the general loss function given in (1.2). The proof of Theorem 1 makes use of several facts which are stated as lemmas below. Recall that $\theta_i=kp_iq^{-1}$, and hence the sum $\theta.\equiv\sum_{i=1}^{n}\theta_i=k(1-q)/q$ and $\theta_i\theta.^{-1}=p_i\left(\sum_{j=1}^{n}p_j\right)^{-1}=p_i(1-q)^{-1}$.

LEMMA 1. *Suppose $X=(X_1,\cdots,X_n)$ has the probability distribution function given by* (1.1). *Then*

(1) *conditional on $Z\equiv\sum_{i=1}^{n}X_i$, $X=(X_1,\cdots,X_n)$ has a multinomial distribution and*

(2.1) $$E_{X|Z}(X_i)=Z\theta_i\theta_\cdot^{-1}, \qquad \text{and}$$

(2.2) $$E_{X|Z}(X_i^2)=Z\theta_i\theta_\cdot^{-1}(1-\theta_i\theta_\cdot^{-1})+Z^2\theta_i^2\theta_\cdot^{-2},$$

*where* $E_{X|Z}$ *denotes expectation with respect to the conditional distribution of* $X$ *given* $Z$.

(2) *The sum* $Z=\sum_{i=1}^{n} X_i$ *has a negative binomial distribution with probability distribution function equal to*

(2.3) $$P(Z=z)=\frac{\Gamma(z+k)}{z!\Gamma(k)}q^k(1-q)^z, \qquad z=0, 1, \cdots$$

(2.4) $$\text{and} \qquad E_Z(Z)=k(1-q)/q=\theta_\cdot,$$

*where* $E_Z$ *stands for expectation with respect to the distribution of* $Z$.

LEMMA 2. *Let* $Z$ *be a random variable. Suppose* $h_1(\cdot)$ *and* $h_2(\cdot)$ *are real-valued functions. Then, if* $h_1(z)$ *is nondecreasing and* $h_2(z)$ *is nonincreasing in* $z$, *then*

(2.5) $$E[h_1(Z)h_2(Z)]\leqq E[h_1(Z)]\,E[h_2(Z)],$$

*provided that all the expected values above are finite. Moreover, if* $h_1(z)$ *is strictly nondecreasing,* $h_2(z)$ *is strictly nonincreasing and* $Z$ *is not a degenerate random variable, then the inequality in* (2.5) *is strict.*

LEMMA 3. *Suppose* $K(\cdot)$ *is a nonincreasing real-valued function such that* $K(\cdot)>0$. *Suppose* $\theta_i>0$, $i=1,\cdots,n$. *Then*

(2.6) $$\frac{1}{n}\sum_{i=1}^{n}\theta_i K(\theta_i)\leqq\left[\frac{1}{n}\sum_{i=1}^{n}\theta_i\right]\left[\frac{1}{n}\sum_{i=1}^{n}K(\theta_i)\right].$$

The proof of Lemma 3 follows directly from Lemma 2.

Let $\hat\theta(X)=(\hat\theta_1(X),\cdots,\hat\theta_n(X))$ be an estimator of $\theta$. Denote by $R_K(\theta, \hat\theta)=EL_K(\theta,\hat\theta(X))$ the risk function of $\hat\theta$, where $L_K(\theta,\hat\theta)$ is given in (1.2). For the special case $K(\cdot)\equiv 1$, denote the corresponding risk function as $R_1(\theta,\hat\theta)$. The notation is used throughout this paper.

THEOREM 1. *Suppose* $X=(X_1,\cdots,X_n)$ *has the negative multinomial distribution* NM$(k, p)$, *given in* (1.1). *Let* $\hat\theta^*(X)$ *be an estimator of the mean vector* $\theta=(\theta_1,\cdots,\theta_n)$ *of* $X$ *defined by*

(2.7) $$\hat\theta^*(X)=[1-\phi(Z)/(Z+b)]X,$$

*where* $Z=\sum_{i=1}^{n}X_i$, $b>0$, *and the real-valued function* $\phi(\cdot)$ *satisfies*

(2.8) $$0 \leqq \phi(z) \leqq \min\{2b, 2(n-1)\}\ ,$$

*and*

(2.9) $\phi(\cdot)$ *is nondecreasing and* $\phi(\cdot) \not\equiv 0$ .

*Then, under the loss function given in* (1.2), *the difference in risk functions of* $\hat{\theta}^*(X)$ *and* $\hat{\theta}^\circ(X) = X$, *respectively, is*

(2.10) $$\Delta \equiv R_K(\theta, \hat{\theta}^*) - R_K(\theta, \hat{\theta}^\circ) \leqq 0\ , \qquad \textit{for all } \theta$$

*with strict inequality for some* $\theta$. *That is,* $\hat{\theta}^*(X)$ *dominates* $\hat{\theta}^\circ(X)$ *uniformly under the loss function* (1.2).

PROOF. For notational convenience, let $K_0 = \sum_{i=1}^n K(\theta_i)$ and $K_1 = \sum_{i=1}^n \theta_i \cdot K(\theta_i)$. The difference in risk functions of $\hat{\theta}^*$ and $\hat{\theta}^\circ$ is

(2.11)
$$\begin{aligned}\Delta &= \mathrm{E}_Z\, \mathrm{E}_{X|Z} \Big\{ \sum_{i=1}^n \theta_i^{-1} K(\theta_i)[-2(X_i - \theta_i)(Z+b)^{-1} \\ &\qquad \times \phi(Z) X_i + \phi^2(Z)(Z+b)^{-2} X_i^2] \Big\} \\ &= \mathrm{E}_Z\, \mathrm{E}_{X|Z} \Big\{ \sum_{i=1}^n \theta_i^{-1} K(\theta_i)[2\theta_i (Z+b)^{-1} \phi(Z) X_i \\ &\qquad + ((Z+b)^{-2}\phi^2(Z) - 2(Z+b)^{-1}\phi(Z)) X_i^2] \Big\}\ .\end{aligned}$$

By Lemma 1, (2.1) and (2.2), (2.11) becomes

(2.12)
$$\begin{aligned}&\theta_\cdot^{-1}\, \mathrm{E}_Z \{2Z(Z+b)^{-1}\phi(Z)K_1 + Z[(Z+b)^{-2}\phi^2(Z) - 2(Z+b)^{-1}\phi(Z)] \\ &\qquad \times [K_0 - K_1\theta_\cdot^{-1} + ZK_1\theta_\cdot^{-1}]\} \\ &= \theta_\cdot^{-1}\, \mathrm{E}_Z \{2Z(Z+b)^{-1}\phi(Z)K_1(1 - Z\theta_\cdot^{-1}) + Z^2(Z+b)^{-2}\phi^2(Z)K_1\theta_\cdot^{-1} \\ &\qquad + Z((Z+b)^{-2}\phi^2(Z) - 2(Z+b)^{-1}\phi(Z))[K_0 - K_1\theta_\cdot^{-1}]\}\ .\end{aligned}$$

Since $h_1(Z) = Z(Z+b)^{-1}\phi(Z)$ is nondecreasing in $Z$ by (2.8) and (2.9), and $h_2(Z) = (1 - Z\theta_\cdot^{-1})$ is nonincreasing in $Z$, (2.5) of Lemma 2 and the fact that $\mathrm{E}_Z(1 - Z\theta_\cdot^{-1}) = 0$ imply that the first summand in the braces in (2.12) has expectation less than or equal to zero. To complete the proof, it is necessary to show that the expectation of the sum of the remaining two summands in (2.12) is less than or equal to zero. The expectation of these two summands is

(2.13)
$$\begin{aligned}&\mathrm{E}_Z \{\theta_\cdot^{-1} Z(Z+b)^{-2}\phi(Z)[Z\phi(Z)K_1 + \phi(Z)(K_0\theta_\cdot - K_1) \\ &\qquad - 2(Z+b)(K_0\theta_\cdot - K_1)]\} \\ &= \mathrm{E}_Z \{\theta_\cdot^{-1} Z(Z+b)^{-2}\phi(Z)[(K_0\theta_\cdot - K_1)(\phi(Z) - 2b) \\ &\qquad + Z(\phi(Z)K_1 - 2(K_0\theta_\cdot - K_1))]\}\ .\end{aligned}$$

By (2.6) of Lemma 3, $nK_1 \leqq K_0\theta_\cdot$. Moreover, $\phi(z) - 2b \leqq 0$ and $\phi(z) \geqq 0$ by (2.8), and hence (2.13) is less than or equal to

$$\mathrm{E}_Z\{\theta_\cdot^{-1}Z(Z+b)^{-2}\phi(Z)K_1[(n-1)(\phi(Z)-2b)+Z(\phi(Z)-2(n-1))]\}\leqq 0\ ,$$

again by (2.8). Therefore, (2.10) holds.

*Remark.* Recall that the case of independent Poisson random variables $X_i$ is a limiting case of $X=(X_1,\cdots,X_n)$ with a negative multinomial distribution. Hence, Theorem 3.1 of Clevenson and Zidek [2] can be viewed as a special case of Theorem 1.

The case of simultaneously estimating the means of several independent negative multinomial distributions is considered in the next section. Theorem 2 provides a result similar to that in Theorem 1, but for this multivariate negative multinomial case. The proof of Theorem 1 can be modified to prove Theorem 2, but the resulting class of dominating estimators is small. A large class of estimators dominating the usual estimator can be obtained if the loss function used is the special case of (1.1) with $K(\cdot)\equiv 1$. The proof of the latter result uses the difference inequality method. The result and the proof will be discussed further in the next section.

## 3. Multivariate negative multinomial case

In this section, the following setting is assumed:

For $j=1,\cdots,m$, $\boldsymbol{X}_j=(X_{j1},\cdots,X_{jn_j})$ has a negative multinomial distribution $\mathrm{NM}(k_j,\boldsymbol{p}_j)$, where $k_j>0$, $\boldsymbol{p}_j=(p_{j1},\cdots,p_{jn_j})$, $p_{ji}>0$, and $q_j=1-\sum_{i=1}^{n_j}p_{ji}>0$. Furthermore, the $\boldsymbol{X}_j$, $j=1,\cdots,m$, are mutually independent. The random vector $\boldsymbol{X}=(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_m)$ is said to have a multivariate negative multinomial distribution. Let $\boldsymbol{\theta}_j=(\theta_{j1},\cdots,\theta_{jn_j})$ be the mean of $\boldsymbol{X}_j$ and let $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_m)$. This setting will be referred to as a multivariate negative multinomial setting.

The problem is to estimate $\boldsymbol{\theta}$ based on $\boldsymbol{X}$. The corresponding loss function (1.1) in the above setting becomes

$$L_K(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})=\sum_{j=1}^{m}\sum_{i=1}^{n_j}K(\theta_{ji})(\theta_{ji}-\hat{\theta}_{ji})^2/\theta_{ji}\ , \tag{3.1}$$

where $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ is an estimator of $\boldsymbol{\theta}$. Denote

$$Z_j=\sum_{i=1}^{n_j}X_{ji}\ ,\qquad Z=\sum_{j=1}^{m}Z_j\ ,\qquad Z_{(ji)}=Z_j-X_{ji}\ , \tag{3.2}$$

$$\boldsymbol{\theta}_j=k_j\boldsymbol{p}_j/q_j\ ,\qquad \theta_{j\cdot}=\sum_{i=1}^{n}\theta_{ji}=k_j(1-q_j)/q_j\ , \tag{3.3}$$

$$K_{j0}=\sum_{i=1}^{n_j}K(\theta_{ji})\ ,\qquad K_{j1}=\sum_{i=1}^{n_j}\theta_{ji}K(\theta_{ji})\ ,\qquad n_*=\mathrm{Min}\,\{n_j\}\ . \tag{3.4}$$

Theorem 2 below provides a class of estimators $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ of $\boldsymbol{\theta}$ uniformly dominating $\hat{\boldsymbol{\theta}}^{\circ}(\boldsymbol{X})=\boldsymbol{X}$ under loss (3.1) for the case $n_*\geqq 2$.

THEOREM 2. *Let $\boldsymbol{X}=(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_m)$ be as described in the multivariate negative multinomial setting. Suppose $n_*\geqq 2$. Then the estimator $\hat{\boldsymbol{\theta}}^*(\boldsymbol{X})$ given in* (2.7) *uniformly dominates $\hat{\boldsymbol{\theta}}^{\circ}(\boldsymbol{X})=\boldsymbol{X}$ under loss* (3.1) *if the function $\phi(Z)$ in* (2.7) *satisfies conditions* (2.8) *and* (2.9) *except that the value of $n$ in* (2.8) *must be replaced by $n_*$.*

PROOF. Given $\boldsymbol{Z}=(Z_1,\cdots,Z_m)$, the conditional expectations of $X_{ji}$ and $X_{ji}^2$ have similar expressions as in Lemma 1, (2.1) and (2.2). Proceeding in the proof of Theorem 1, the difference in risk functions, $\Delta$, of $\hat{\boldsymbol{\theta}}^*(\boldsymbol{X})$ and $\hat{\boldsymbol{\theta}}^{\circ}(\boldsymbol{X})$ can be shown to be less than or equal to an expression similar to that of (2.13), namely,

$$(3.3)\quad \Delta\leqq\sum_{j=1}^{m} \mathrm{E}_Z\{\theta_{j\cdot}^{-1}Z_j(Z+b)^{-2}\phi(Z)[(K_{j0}\theta_{j\cdot}-K_{j1})(\phi(Z)-2b) + Z_j(\phi(Z)K_{j1}-2(K_{j0}\theta_{j\cdot}-K_{j1}))-2(Z-Z_j)(K_{j0}\theta_{j\cdot}-K_{j1})]\}\ .$$

Now, (2.6) of Lemma 3 implies $n_jK_{j1}\leqq K_{j0}\theta_{j\cdot}$. The assumed conditions on $\phi(Z)$ then imply that (3.3) cannot exceed zero and the proof is complete.

For the case of several independent negative binomial distributions, that is, for $n_*=1$, the proof of Theorem 2 does not seem to be modifiable to produce estimators $\hat{\theta}^*(X)$ dominating $\hat{\theta}^{\circ}(X)$ under loss function (3.1). However, if the loss function is $L_1(\theta,\hat{\theta})$, the special case of (3.1) (or (1.1)) with $K(\cdot)\equiv 1$, it can be shown by using the difference inequality method that $\hat{\theta}^{\circ}(X)=X$ is dominated by a class of estimators even for $n_*=1$. The result is summarized in Theorem 3. For simplicity, denote $n^*=\operatorname{Max}_{j=1}^{m}\{n_j\}$ and $n.=\sum_{j=1}^{m} n_j$.

THEOREM 3. *Let $\boldsymbol{X}=(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_m)$ be as described in the multivariate negative multinomial setting. Then, to estimate the mean $\boldsymbol{\theta}$ of $\boldsymbol{X}$ under loss function $L_1(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})$, the estimators given in* (2.7) *dominate the estimator $\hat{\theta}^{\circ}(X)=X$, provided that the function $\phi(Z)$ in* (2.7) *satisfies* (2.9),

$$(3.4)\qquad b>n^*-1\ ,\qquad 0\leqq\phi(Z)\leqq 2\operatorname{Min}\{(n.-1),b\}\ ,$$

*and*

$$(3.5)\qquad (Z+b)^{-1}\phi(Z)\ \text{is nonincreasing in}\ Z\ .$$

With the convention that $\sum_{t=1}^{a} n_t=0$ for $a\leqq 0$, the $\left[i+\sum_{t=1}^{j-1} n_t\right]$th coordinate of an $n.$-vector of real values is called the $(ji)$th coordinate of the vector for simplicity. Thus, the $(ji)$th coordinate of $X$ is $X_{ji}$. Let $e_{ji}$ be the $n.$-vector whose $(ji)$th coordinate is one and whose other coordinates are zero. The following lemmas are useful in proving Theorem 3.

LEMMA 4. *Suppose $X=(X_1,\cdots,X_m)$ is as described in the multivariate negative multinomial setting. Suppose $g(X)$ is a real-valued function such that*

$$\mathrm{E}\, g(X)<\infty\,, \quad \textit{and} \quad g(X)=0 \qquad \textit{if } X_{ji}\leqq 0\,. \tag{3.6}$$

*Then*

$$\mathrm{E}\,\{p_{ji}g(X)\}=\mathrm{E}\,\{X_{ji}g(X-e_{ji})/(Z_j+k_j-1)\}\,, \tag{3.7}$$

$$\mathrm{E}\,\{g(X)/p_{ji}\}=\mathrm{E}\,\{(Z_j+k_j)g(X-e_{ji})/(X_{ji}+1)\}\,, \tag{3.8}$$

*and*

$$\mathrm{E}\,\{g(X)/\theta_{ji}\}=\frac{1}{k_j}\,\mathrm{E}\left\{\left(\frac{Z_j+k_j}{X_{ji}+1}g(X+e_{ji})\right.\right.\\ \left.\left.-\sum_{t\neq i}^{n_j}\frac{g(X+e_{ji}-e_{jt})}{X_{ji}+1}X_{jt}-g(X)\right)\right\}\,. \tag{3.9}$$

PROOF. The proofs of (3.7) and (3.8) are similar to the one in Hwang ([4], p. 19–20), where several independent negative binomial distributions were considered. The proof of (3.9) follows from (3.7) and (3.8) and the observation that $\theta_{ji}=k_jp_{ji}/q_j$ and $q_j=1-\sum_{i=1}^{n_j}p_{ji}$.

LEMMA 5. *Suppose $X=(X,\cdots,X_m)$ is as described in the multivariate multinomial setting. Let $\hat{\theta}(X)=X+f(X)$ be an estimator of $\theta$, where $f(X)$ is an $n.$-vector of real-valued functions $f_{ji}(X)$, $j=1,\cdots,m$, $i=1,\cdots,n_j$ such that $f_{ji}(X)$ satisfies condition* (3.6) *in Lemma* 4. *Then under the loss function $L_1(\theta,\hat{\theta})$, the difference in risk functions of $\hat{\boldsymbol{\theta}}(X)$ and $\hat{\boldsymbol{\theta}}^\circ(X)=X$ is*

$$\Delta\equiv R(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})-R(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}^\circ)=\sum_{j=1}^{m}\mathrm{E}\,[D_{j1}(X)+D_{j2}(X;k_j)] \tag{3.10}$$

*where*

$$D_{j1}(X)=\sum_{i=1}^{n_j}f^2(X+e_{ji})/(X_{ji}+1)+2\sum_{i=1}^{n_j}[f_{ji}(X+e_{ji})-f_{ji}(X)]\,, \tag{3.11}$$

$$k_jD_{j2}(X;k_j)=\sum_{i=1}^{n_j}\left[Z_j\frac{f_{ji}^2(X+e_{ji})}{X_{ji}+1}-\sum_{t\neq i}^{n_j}X_{jt}\frac{f_{ji}^2(X+e_{ji}-e_{jt})}{X_{ji}+1}-f_{ji}(X)\right]\\ +2\sum_{i=1}^{n_j}\left[Z_jf_{ji}(X+e_{ji})-\sum_{t\neq i}^{n_j}X_{jt}f_{ji}(X+e_{ji}-e_{jt})\right. \tag{3.12}$$

$$-X_{ji}f_{ji}(X)\Big] .$$

PROOF. $\varDelta=\mathrm{E}\left\{\sum_{j=1}^{m}\sum_{i=1}^{n_j}[f_{ji}^2(X)/\theta_{ji}+2(f_{ji}(X)/\theta_{ji}]\right\}$. Applying identity (3.9) of Lemma 5 consecutively with $g(X)=f_{ji}^2(X)$ and $g(X)=X_{ji}f_{ji}(X)$ yields (3.10) after some algebraic manipulations.

From (3.10), a sufficient condition for an estimator $\hat{\theta}(X)=X+f(X)$ to dominate $\hat{\theta}^\circ(X)=X$ under loss function $L_1(\theta,\hat{\theta})$ is that the functions $f_{ji}$ satisfy the difference inequality:

$$\sum_{j=1}^{m}[D_{j1}(X)+D_{j2}(X;k_j)]\leqq 0\ , \qquad \text{for all } X, \tag{3.13}$$

with strict inequality with positive probability.

PROOF OF THEOREM 3. Define $f_{ji}(X)=-\phi(Z)X_{ji}/(Z+b)$ if $X_{ji}\geqq 0$, zero otherwise. We show that these $f_{ji}$'s satisfy the difference inequality (3.13). First, it is straightforward to show that

$$\sum_{j=1}^{m}D_{j1}(X)=\frac{\phi^2(Z+1)(Z+n.)}{(Z+b+1)^2}-2\phi(Z)\left[\frac{Z+n.}{Z+b+1}-\frac{Z}{Z+b}\right] . \tag{3.14}$$

By the same argument as in (3.2) of Tsui and Press ([11], p. 95), (3.14) is less than or equal to zero if $\phi(Z)$ satisfies (2.9) and (3.4). To complete the proof, we show that for each $j$, $D_{j2}(X;k_j)\leqq 0$ for all $X$. Now, from (3.12) and (3.2),

$$\begin{aligned}k_jD_{j2}(X;k_j)=&\frac{\phi^2(Z+1)Z_j(Z_j+n_j)}{(Z+b+1)^2}-\frac{\phi(Z)\sum_{i=1}^{n_j}[Z_{(ji)}(X_{ji}+1)+X_{ji}^2]}{(Z+b)^2}\\&-\frac{2\phi(Z+1)(Z_j+n_j)Z_j}{(Z+b+1)}+\frac{2\phi(Z)(Z_j+n_j-1)Z_j}{(Z+b)} .\end{aligned} \tag{3.15}$$

Using the identity $\sum_{i=1}^{n_j}[Z_{(ji)}(X_{ji}+1)+X_{ji}^2]\equiv Z_j(Z_j+n_j-1)$, (3.15) becomes

$$\frac{Z_j\phi(Z+1)(Z_j+n_j)}{(Z+b+1)}\left[\frac{\phi(Z+1)}{Z+b+1}-2\right]-\frac{Z_j\phi(Z)(Z_j+n_j-1)}{(Z+b)}\left[\frac{\phi(Z)}{Z+b}-2\right] . \tag{3.16}$$

By (3.4), $b>n^*-1$, hence $(Z_j+n_j-1)(Z+b)^{-1}$ is nondecreasing in $Z_j$. Therefore, $\phi(Z)(Z_j+n_j-1)(Z+b)^{-1}$ is nondecreasing in $Z_j$. By (3.4) again, the expressions in both sets of brackets are nonpositive and (3.16) becomes

$$\frac{Z_j\phi(Z+1)(Z_j+n_j)}{(Z+b+1)}\left[\frac{\phi(Z+1)}{Z+b+1}-\frac{\phi(Z)}{Z+b}\right]\leqq 0\ , \qquad \text{by (3.5)} .$$

Thus, $D_{j2}(X; k_j) \leqq 0$ for all $X$. This completes the proof.

*Remark.* If the $X_j$'s are independent binomial random variables, that is, $n_* = 1$, Theorem 1 of Tsui [10] is a special case of Theorem 3 with $b > m-1$.

## 4. Further Generalizations and concluding remarks

Theorem 1 states that in estimating the mean vector $\theta = (\theta_1, \cdots \theta_n)$ of a negative multinomial distribution based on the observed $X = (X_1, \cdots, X_n)$, $n \geqq 2$, the estimator $\hat{\theta}^*(X)$ given in (2.7) is better than the MLE $\hat{\theta}^\circ(X) = X$ under the fairly general loss function (1.1). The proof of the result depends mainly on the properties (2.1) and (2.2) of the conditional distribution of $X$ given $Z = \sum_{i=1}^{n} X_i$. The marginal distribution of $Z$ does not appear to play an important role. It is therefore natural to expect that the result of Theorem 1 remains true for other multivariate discrete distributions as long as properties (2.1) and (2.2) continue to hold.

Recall that the negative multinomial distribution can be obtained as a mixture of independent Poisson distributions with the gamma distribution as mixing distribution. More precisely, the probability distribution function $X$ can be written in the form

$$P(X_1 = x_1, \cdots, X_n = x_n) = E_{F(\alpha)} \left\{ \prod_{i=1}^{n} P(X_i = x_i \mid \lambda_i \alpha) \right\} , \tag{4.1}$$

$x_i = 0, 1, \cdots,$ for $i = 1, \cdots, n$, where $P(X_i = x_i \mid \lambda_i \alpha)$ is the Poisson probability distribution function with mean $\lambda_i \alpha$ and the expectation is taken with respect to the distribution $F(\alpha)$ of $\alpha$, the gamma distribution in this case. For other mixing distributions $F(\cdot)$ of $\alpha$, it is not difficult to show that the resulting multivariate discrete distribution in (4.1) has properties (2.1) and (2.2). Moreover, the components of $X$ are dependent.

Theorem 1 can be further generalized in another direction, namely, to an even more general loss function than (1.1). Note that in the proof of Theorem 1, the transition from (2.13) to (2.14) made use of the inequality (2.6). Suppose the loss function used is (1.1), except that $K(\theta_i)$ in (1.1) is replaced by $K_i(\theta_i)$, $i = 1, \cdots, n$, where the $K_i(\cdot)$ are possibly different non-increasing functions and $K_i(\cdot) > 0$ for all $i$. Denote this loss function by $L^*$. Suppose further that for some $\beta > 1$, the functions $K_i(\cdot)$ satisfy a similar inequality as (2.6), namely,

$$\beta \sum_{i=1}^{n} \theta_i K_i(\theta_i) \leqq \left[ \sum_{i=1}^{n} \theta_i \right] \left[ \sum_{i=1}^{n} K_i(\theta_i) \right] , \qquad \text{for all } \theta_i > 0 . \tag{4.2}$$

Then, using the same argument as in the proof of Theorem 1, we have the following generalization of Theorem 1.

THEOREM 4. *Let $n \geqq 2$. Suppose the probability distribution function of $X=(X_1, \cdots, X_n)$ is given by* (4.1), *which includes the negative multinomial distribution as a special case. Suppose further that the loss function $L^*$ is used and that inequality* (4.2) *holds. Let $\phi(\cdot)$ be a real-valued function satisfying conditions* (2.8) *and* (2.9) *except that $n$ in* (2.8) *must be replaced by $\beta$ given in* (4.2). *Then the estimator $\hat{\theta}^*(X)$ of the mean $\theta$ of $X$, given in* (2.7), *is uniformly better than the naive estimator $\hat{\theta}^\circ(X)=X$.*

An example of (4.2) is $K_i(\theta_i)=\gamma_i K(\theta_i)$ for $\gamma_i>0$, $i=1, \cdots, n$, and $K(\theta_i)$ is as described in Lemma 3. In this example, $\beta=\gamma./\max\{\gamma_j\}>1$, where $\gamma. \equiv \sum_{j=1}^{n} \gamma_j$. To see this, observe that

$$\sum_{i=1}^{n} \theta_i K_i(\theta_i)=\gamma.\left[\sum_{i=1}^{n} \theta_i K(\theta_i)\gamma_i/\gamma.\right] . \tag{4.3}$$

By Lemma 1, (4.3) cannot exceed

$$\begin{aligned}\gamma.\left[\sum_{i=1}^{n} \theta_i\gamma_i/\gamma.\right]\left[\sum_{i=1}^{n} K(\theta_i)\gamma_i/\gamma.\right] &= \gamma.^{-1}\left[\sum_{i=1}^{n} \theta_i\gamma_i\right]\left[\sum_{i=1}^{n} K_i(\theta_i)\right] \\ &\leqq \gamma.^{-1}[\operatorname{Max}\{\gamma_j\}]\left(\sum_{i=1}^{n} \theta_i\right)\left(\sum_{i=1}^{n} K_i(\theta_i)\right) .\end{aligned}$$

Inequality (4.2) then follows.

In summary, this paper examined the previously unexplored problem of inadmissibility of the usual estimator of the mean of a multivariate discrete distribution with marginal distributions *not necessarily independent* of one another. Special attention was given to the negative multinomial distribution. It was shown that the Clevenson-Zidek type estimators given in (2.7) uniformly dominate the moment estimator in two major settings. The first setting assumes a general loss function, $L_K$, given in (1.1), and involves one negative multinomial distribution. More generally, it involves the multivariate discrete distribution given in (4.1) and a loss function $L^*$, which is more general than $L_K$. The second setting assumes a more specific loss function, $L_1$, which is a special case of $L_K$ with $K(\cdot)\equiv 1$, and involves several independent negative multinomial distributions. The superiority of the Clevenson-Zidek type estimators over the usual one may hence be considered robust both with respect to a large class of loss functions and various distributional assumptions.

UNIVERSITY OF WISCONSIN-MADISON

## REFERENCES

[1] Bates, G. F. and Neyman, J. (1952). Contributions to the theory of accident proneness, I, 'an optimistic model of the correlation between light and severe accidents', and II, 'True or false contagion', *University of California Publications in Statistics*, 1, 215-275.

[2] Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws, *J. Amer. Statist. Ass.*, 70, 698-705.

[3] Ghosh, M., Hwang, J. T. and Tsui, K. W. (1983). Construction of improved estimators in multiparameter estimation for exponential families (with Discussion), *Ann. Statist.*, 11, 351-367.

[4] Hwang, J. T. (1979). Improving upon inadmissible estimators in discrete exponential families, *Technical Report*, No. 79-14, Purdue University, Department of Statistics.

[5] Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases, *Ann. Statist.*, 10, 857-867.

[6] Neyman, J. (1963). Certain chance mechanisms involving discrete distributions, *Proceedings of the International Symposium on Classical and Contagious Discrete Distributions* (ed. G. T. Patil, McGill University, Montréal), Statistical Publishing Society, 4-14.

[7] Peng, J. C. M. (1975). Simultaneous estimation of the parameters of independent Poisson distributions, *Technical Report*, No. 78, Department of Statistics, Stanford University.

[8] Sibuya, M., Yoshimura, I. and Shimizu, R. (1964). Negative multinomial distribution, *Ann. Inst. Statist. Math.*, 16, 409-426.

[9] Tsui, K. W. (1981). Simultaneous estimation of several Poisson parameters under squared error loss, *Ann. Inst. Statist. Math.*, 33, 214-223.

[10] Tsui, K. W. (1984). Robustness of Clevenson-Zidek type estimators, *J. Amer. Statist. Ass.*, 79, 152-157.

[11] Tsui, K. W. and Press, S. J. (1982). Simultaneous estimation of several Poisson parameters under $k$-normalized squared error loss, *Ann. Statist.*, 10, 93-100.