



Construction of Improved Estimators in Multiparameter Estimation for Discrete Exponential Families

Author(s): Malay Ghosh, Jiunn Tzon Hwang and Kam-Wah Tsui

Source: *The Annals of Statistics*, Vol. 11, No. 2 (Jun., 1983), pp. 351–367

Published by: [Institute of Mathematical Statistics](#)

Stable URL: <http://www.jstor.org/stable/2240551>

Accessed: 28/01/2015 13:58

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *The Annals of Statistics*.

<http://www.jstor.org>

CONSTRUCTION OF IMPROVED ESTIMATORS IN MULTIPARAMETER ESTIMATION FOR DISCRETE EXPONENTIAL FAMILIES¹

BY MALAY GHOSH², JIUNN TZON HWANG³ AND KAM-WAH TSUI

*Iowa State University and University of Florida; Cornell University; and
University of Wisconsin-Madison*

This paper extends and unifies the theory of simultaneous estimation for the discrete exponential family. We discuss construction of estimators which theoretically dominate the uniformly minimum variance unbiased estimator (UMVUE) under a weighted squared error loss function, and show by means of computer simulation results that new simultaneous Poisson means estimators perform more favorably than those previously proposed. Our improved estimators shift the UMVUE towards a possibly nonzero point or a data-based point.

1. Introduction. Stein (1956) obtained the surprising result that for estimating p independent normal means simultaneously, the sample mean was inadmissible under squared error loss when $p \geq 3$. An explicit estimator dominating the sample mean was introduced by James and Stein (1961). Later, Stein (1973) introduced an integration by parts technique to transform the problem of finding improved estimators of the sample mean into that of solving certain differential inequalities. This technique has been used in the problem of multiparameter estimation for the continuous exponential family (Hudson, 1978), as well as in the multivariate normal mean estimation problem. The important role of differential inequalities in the study of inadmissibility was also emphasized by Brown (1979).

Various improved multivariate normal mean estimators have been derived under different assumptions about the covariance structure. The covariance matrix may be assumed to be known or unknown, or perhaps diagonal. The loss function used in much of the research is quadratic. (Brown (1966) discussed a wider class of loss functions for the simultaneous estimation problem.) Berger et al. (1977), Berger and Haff (1981), and other references cited therein describe the results more fully.

There has also been considerable interest in simultaneous estimation problems for nonnormal distributions, particularly in estimating the parameters $\theta_1, \dots, \theta_p$ of p independent Poisson distributions, with one observation available from each of the p Poisson populations. (This one observational setting can always be obtained by a sufficiency reduction.)

Let

$$(1.1) \quad L_m(\theta, \delta) = \sum_{i=1}^p (\theta_i - \delta_i)^2 / \theta_i^{m_i}$$

where $\delta = (\delta_1, \dots, \delta_p)$ is an estimator of $(\theta_1, \dots, \theta_p)$ and $m = (m_1, \dots, m_p)$ is a vector of p nonnegative integers. Though the maximum likelihood estimator (MLE), which is also the uniformly minimum variance unbiased estimator (UMVUE), is admissible under L_m

Received June 1982; revised October 1982.

¹ Research for this article was carried out independently by the three authors, whose collaboration was invited by the editor. The order of authors' names is alphabetical, and does not indicate their relative contributions to the paper.

² Research supported by the NSF Grant Number MCS-8005485.

³ Research supported by the NSF Grant Number MCS-8003568.

AMS 1980 subject classification. 62C15, 62F10, 62H99.

Key words and phrases. Difference inequality, differential inequality, exponential family, improved estimator, UMVUE, adaptive estimators, Poisson, negative binomial, simultaneous estimation.

TABLE 1
Catalog of some recent results in simultaneous estimation of the parameters of p independent distributions belonging to one-parameter exponential families (discrete or continuous)

Authors	Values of m_i in loss function (1.1)	Distributions	Results†
Hudson (1974)	0	Cont. Exp.	A ($p \geq 4$)
Clevenson and Zidek (1975)	1	Poisson	A ($p \geq 2$)
Peng (1975)	0	Poisson	A ($p \geq 3$)
Hudson (1978)	0	Disc. Exp.	A ($p \geq 4$)
Hudson (1978)	0	Cont. Exp.	A, C ($p \geq 4$)
Tsui and Press (1977 and 1982)	$m^* > 0$	Poisson	A ($p \geq 2$)
Tsui (1978)	possibly different positive integers	Poisson	A ($p \geq 2$)
Tsui (1979)	0	Disc. Exp.	B, C ($p \geq 4$)
Tsui (1981a)	0	Poisson	B, C ($p \geq 3$)
Berger (1980)	integer m^*	Cont. Exp.	A, D ($p \geq 2$)
Hwang (1979)	possibly different nonnegative integers	Disc. Exp.	A ($p \geq 3$)
Hudson and Tsui (1981)	0	Poisson	B ($p \geq 3$)
Ghosh and Parsian (1980)	integer m^*	Cont. Exp.	A ($p \geq 2$)
Ghosh, Hwang and Tsui††	possibly different nonnegative integers	Disc. Exp.	B, C ($p \geq 3$)

† A—Improved estimator shifts the components of the UMVUE towards zero.
B—Improved estimator shifts the components of the UMVUE towards some fixed point.
C—Improved estimator shifts the components of the UMVUE towards some data-based point.
D—Improved estimator shifts the components of the UMVUE possibly towards infinity.
††—This paper.

for $p = 1$ (see, e.g., Hodges and Lehmann, 1951), it is shown to be inadmissible for large p . The case when $m_i = 1, i = 1, \dots, p$, was considered by Clevenson and Zidek (1975) (see also Ghosh and Parsian, 1981), who obtained a class of estimators which are uniformly better than the MLE provided that $p \geq 2$. In the squared error loss case (i.e., $m_i = 0, i = 1, \dots, p$), Peng (1975) showed that the MLE is inadmissible when $p \geq 3$, rather than when $p \geq 2$. Tsui and Press (1977) considered L_{m^*} , the special case of (1.1) in which all the m_i 's are equal to a constant, m^* , and provided estimators that dominate the MLE under such loss functions for all positive integers m^* , as long as $p \geq 2$. Tsui and Press (1982) further developed their results and performed computer simulations, which showed that their Poisson parameters estimators can in some situations improve on the risk of the MLE by over 30%. Tsui (1978) showed that the MLE is inadmissible under the most general loss function (1.1) for $p \geq 2$ if all the m_i 's are positive.

More generally, Hudson (1978) showed that under squared error loss, the UMVUE is inadmissible in simultaneously estimating the natural parameters of p independent distributions belonging to a one-parameter discrete exponential family, provided $p \geq 4$. This family of distributions includes the Poisson distribution and the negative binomial distribution as special cases. Some of Hudson's results were generalized by Tsui (1979a) and Hwang (1982). Almost all of the inadmissibility results for the Poisson case, and more generally, for the one-parameter discrete exponential family case, were obtained by first deriving some difference inequalities using Stein's (1973) idea, and then solving these inequalities by guesswork. The same approach was used to obtain inadmissibility results for the continuous exponential family. The solutions of these inequalities were used to construct the improved estimators. Table 1 summarizes some recent results. In Table 1, not all the papers listed discussed improving on the UMVUE. In many problems, such as those considered in Berger (1980), the UMVUE is inadmissible even in the one dimensional case. This, however, does not occur in the Poisson and negative binomial estimation

problems considered in this paper. The one dimensional estimators are all admissible under L_m .

The desirability of finding general methods of solving these types of inequalities was evident; Berger's (1980) suggestion of some general methods for solving differential inequalities for the continuous exponential family was therefore a welcome advancement. He constructed, for example, improved estimators for the scale parameters of several independent gamma distributions when L_m is the loss function, where m^* is an arbitrary integer. Some of his results were extended by Ghosh and Parsian (1980). Hwang (1979) adapted Berger's method and was able to provide some general methods of solving difference inequalities for the discrete exponential family. However, the solutions of these differential/difference inequalities mainly lead to improved estimators which shift the usual one towards the origin.

A question which is frequently raised is whether there are improved estimators for the discrete or continuous exponential family that shift the usual one towards a point other than the origin. This question has been answered affirmatively for the multivariate normal mean problem under quadratic loss; see, for example, Lindley's discussion in Stein (1962), and Stein (1981). Similarly, improved estimators which shift the usual one towards a prefixed point or towards a data-based point have been suggested by Tsui (1981a) and Hudson and Tsui (1981) for the Poisson distribution, by Hudson (1978) for the continuous exponential family, and by Tsui (1979) for the discrete exponential family. The loss function in these cases was the squared error loss function. It is natural to conjecture that the general methods of solving inequalities developed by Hwang (1979) and Berger (1980) can be modified in order to yield improved estimators shifting the usual one towards a point other than the origin. This is desirable in view of the simulation results of Tsui (1981a), which indicate that in many cases, shrinking towards a data-based point yields more reduction in risk than shrinking towards the origin. In this paper, we demonstrate that Hwang's (1979) method can be modified in such a way that most previous discrete exponential simultaneous estimation results can be encompassed and the desirable extensions outlined above can be accommodated.

Sections 3 and 4 are devoted to the general theory of simultaneous estimation for the one-parameter exponential family. In particular, Section 3 focuses on improved estimators which shrink the UMVUE toward a prefixed point; adaptive estimators (those which shrink towards a point determined by the data) are derived in Section 4. The most important special case is that of estimating several Poisson parameters. Improved estimators have been applied to oil well discovery (Clevenson and Zidek, 1975), crime rate estimation (Rolph, Chaiken and Houchens, 1981), and error rate estimation in audit sampling (Matsumura and Tsui, 1982). Most of these improved estimators are similar to the ones proposed in this paper. Because of the importance of the improved Poisson means estimators, their theoretical properties and simulated performances are discussed in detail in Section 2.

2. Representative estimators and simulation results. In order to provide concreteness to the general theory developed in later sections, we first introduce and discuss some simple improved estimators in the simultaneous estimation problem for the Poisson distributions. We also report the results of a computer simulation study designed to test the performances of several Poisson means estimators.

It will be convenient to first define our notation. Let $X = (X_1, \dots, X_p)$, where the X_i 's are p independent Poisson random variables with means $\theta_1, \dots, \theta_p$, respectively. If there is no ambiguity, X_i will also be used to denote the realization of X_i as well. The UMVUE of $\theta = (\theta_1, \dots, \theta_p)$ is $\delta^0(X) = X$ based on X . The improved estimator $\delta^* = (\delta_1^*, \dots, \delta_p^*)$ is defined as $\delta^* = X + \phi(X)$, where $\phi(X) = (\phi_1(X), \dots, \phi_p(X))$ and the ϕ_i 's satisfy the following regularity conditions:

$$(2.1) \quad \phi_i(X) = 0 \quad \text{if} \quad X_i < m_i$$

and

$$(2.2) \quad E_{\theta} \phi_i^2(X) < \infty.$$

Let e_i be a p -dimensional vector whose i th coordinate is one and whose other coordinates are zero. Let $\Delta_i F(X)$ be the finite difference $F(X) - F(X - e_i)$. For any integer a , let $1/a^{(-1)} = a + 1$, $a^{(0)} = 1$, and $a^{(n)} = a(a-1) \cdots (a-n+1)$ for any positive integer n . From Tsui and Press (1982, Lemmas 1-4), the difference in risks of δ^* and δ^0 under the loss (1.1) can be written as

$$(2.3) \quad R(\theta, \delta^*) - R(\theta, \delta^0) = 2E_{\theta} \mathcal{D}(X),$$

where

$$(2.4) \quad \mathcal{D}(X) = \sum_{i=1}^p \left\{ \frac{\Delta_i \phi_i(X + m_i e_i)}{(X_i + m_i - 1)^{(m_i-1)}} + \frac{\phi_i^2(X + m_i e_i)}{2(X_i + m_i)^{(m_i)}} \right\}.$$

It is clear from (2.3) and (2.4) that an estimator δ^* which has a correction term ϕ that satisfies (2.1), (2.2) and $\mathcal{D}(X) \leq 0$ for all X , with strict inequality for some X , dominates $\delta^0(X) = X$. Two special cases of (1.1), and hence (2.4), are of special interest. One is the squared error loss case, L_0 (i.e., $m_i = 0$ for all i), and the other is the normalized squared error loss case, L_1 ($m_i = 1$ for all i). For these two cases, $\mathcal{D}(X) \leq 0$ becomes, respectively

$$(2.5) \quad \mathcal{D}(X) = \sum_{i=1}^p \{X_i \Delta_i \phi_i(X) + \frac{1}{2} \phi_i^2(X)\} \leq 0,$$

$$(2.6) \quad \mathcal{D}(X) = \sum_{i=1}^p \{\Delta_i \phi_i(X + e_i) + \frac{1}{2} \phi_i^2(X + e_i)/(X_i + 1)\} \leq 0.$$

In the remainder of this paper, $\sum_{j=a}^b S_j$ is defined to be zero if $b < a$.

We now consider a Poisson means estimator which shifts the UMVUE towards a fixed point $\lambda = (\lambda_1, \dots, \lambda_p)$, where the λ_i 's are nonnegative integers. Intuitively, shrinkage estimators tend to do very well only at points (or subspaces) to which they shrink. Thus if θ is thought to be away from the origin, an estimator shrinking only to the origin will not offer much improvement. Therefore, in order to realize substantial improvement, one should use prior information to guess θ and use an improved estimator that shrinks the usual estimator toward the prior guess of θ . The following example provides a pertinent estimator when the prior guess of θ is λ .

EXAMPLE 2.1 (Under L_0). Let $(C)_+ = \max\{C, 0\}$ for any real number C . Define

$$(2.7) \quad N(X) = \# \{i: X_i > \lambda_i\}$$

and

$$(2.8) \quad h(j) = \sum_{n=1}^j 1/n, \quad j = 1, 2, \dots, \text{zero otherwise.}$$

Furthermore, let $D(X) = \sum_{i=1}^p d_i(X_i)$, where

$$(2.9) \quad d_i(X_i) = \begin{cases} \{h(X_i) - h(\lambda_i)\}^2 + \frac{1}{2} \{3h(\lambda_i) - 2\}_+ & \text{if } X_i < \lambda_i \\ \{h(X_i) - h(\lambda_i)\} \{h(X_i + 1) - h(\lambda_i)\} & \text{if } X_i \geq \lambda_i. \end{cases}$$

Let $\delta^1(X) = X + \phi(X)$ be an estimator of θ with i th component

$$(2.10) \quad \delta_i^1(X) = X_i - \{N(X) - 2\}_+ \{h(X_i) - h(\lambda_i)\} / D(X), \quad i = 1, \dots, p.$$

Then, it follows from Theorem 3.1 in Section 3 that (2.5) becomes

$$(2.11) \quad \mathcal{D}(X) \leq -\frac{1}{2} \{N(X) - 2\}_+^2 / D(X).$$

Thus, δ^1 dominates $\delta^0(X) = X$ under L_0 if $p \geq 3$, and the reduction in risk is expected to be

sizable if $D(X)$ is likely to be small and $(N(X) - 2)_+$ is likely to be large. This will occur if λ is close to the true θ , and the components of λ are not large.

Estimator δ^1 given above is similar to some of the estimators proposed by Hudson and Tsui (1981, Theorems 1 and 2). However, $p \geq 3$ is required here instead of $p \geq 4$.

A better understanding of the structure of the Poisson means estimators under L_0 can be gained by examining the James-Stein estimator δ^{JS} and the Lindley estimator δ^L in the simultaneous normal means estimation problem. To this end, let Y_1, \dots, Y_p be independent, normally distributed, and have variance 1 with means μ_1, \dots, μ_p , respectively. Let $Y = (Y_1, \dots, Y_p)$ and $\bar{Y} = \sum_{j=1}^p Y_j/p$. The forms of δ^{JS} and δ^L , respectively, are

$$(2.12) \quad \delta_i^{JS}(Y) = Y_i - (p-2)(Y_i - \eta_i)/\sum_{j=1}^p (Y_j - \eta_j)^2,$$

$$(2.13) \quad \delta_i^L(Y) = Y_i - (p-3)(Y_i - \bar{Y})/\sum_{j=1}^p (Y_j - \bar{Y})^2, \quad i = 1, \dots, p$$

where $\eta = (\eta_1, \dots, \eta_p)$ is a prechosen point, often taken to be the origin when the James-Stein estimator is quoted. Now observe that $h(X_i)$ given in (2.8) is close to $\ln X_i$ when X_i is large and is close to $\sqrt{X_i}$ when X_i is small. Moreover, both the log transform and the square-root transform are commonly used for the Poisson data with the transformed data treated as approximately normally distributed. Thus, the correction term of δ^1 given in (2.10) is similar to that of δ^{JS} given in (2.12). With this observation, one might speculate that most, if not all, the normal theory results can be adapted to yield similar estimators for the Poisson means. In particular, an estimator δ^M similar to the Lindley estimator given in (2.13) should have a correction term of the following form

$$(2.14) \quad \phi_i^M(X) = -C(X)\{h(X_i) - \bar{h}\}/\sum_{j=1}^p \{h(X_j) - \bar{h}\}^2, \quad i = 1, \dots, p,$$

where $\bar{h}(X) = \sum_{j=1}^p h(X_j)/p$ and $C(X)$ is some appropriate function similar to $(N(X) - 2)_+$ given in (2.10). One guess for $C(X)$ is $[\#\{i: h(X_i) > \bar{h}(X)\} - 2]_+$. Since $h(X_j) \simeq \ln X_j$, we have $\bar{h}(X) \simeq \ln(\prod_{j=1}^p X_j)^{1/p}$ and δ^M therefore shifts the UMVUE, X_i , towards the “geometric mean” of the X_i ’s. Although δ^M is intuitively appealing, its dominance over the UMVUE δ^0 , is difficult to prove with the tools developed in Section 4, or with any other known technique.

Hudson (1981) attempted to apply the normal theory results by using the log transform and approximating finite differences in (2.5) by differentials, resulting in the differential inequality $\mathcal{D}^*(X) \leq 0$ derived in the normal means case. He then proposed estimators whose correction terms satisfy $\mathcal{D}^*(X) \leq 0$, but it is not known if these estimators dominate δ^0 . Adaptive estimators which do theoretically dominate the usual one under squared error loss are described in the examples below.

EXAMPLE 2.2. For random variables Y_1, \dots, Y_p , let $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(p)}$ be the order statistics. Define $N(X) = \#\{i: X_i > X_{(1)}\}$; and $H_i(X) = h(X_i) - h(X_{(1)})$, $i = 1, \dots, p$, where $h(\cdot)$ is as given in (2.8). Let $D(X) = \sum_{i=1}^p H_i(X)H_i(X + e_i)$. Consider the estimator δ^2 with the i th component

$$(2.15) \quad \delta_i^2(X) = X_i + \phi_i(X) \quad \text{and} \quad \phi_i(X) = -\{N(X) - 2\}_+ H_i(X)/D, \quad i = 1, \dots, p.$$

This δ^2 shifts the UMVUE, δ^0 , towards the minimum of the X_i ’s. Its correction term ϕ satisfies (2.11) for $\mathcal{D}(X)$ given in (2.5), and hence dominates δ^0 under L_0 if $p \geq 4$. This new estimator is similar to the one proposed by Tsui (1981a), but, according to our simulation results, has greater reduction in risk over δ^0 than that of Tsui. Notice that the ϕ_i ’s given in (2.15) can be obtained by replacing the λ_i ’s in (2.10) with $X_{(1)}$. Thus, $X_{(1)}$ can be viewed as a conservative estimate of the θ_i ’s, towards which the usual estimator should be corrected. Other order statistics can be used to construct adaptive estimators in the same manner. One such example is described below.

EXAMPLE 2.3. Define the median $\text{med}(X)$ of $X = (X_1, \dots, X_p)$ to be the smallest integer c such that $\#\{i: X_i \leq c\} \geq p/2$. The estimator δ^3 , whose i th component is given by

TABLE 2
Percentage reductions in risk under L_0 for estimators* $\delta^1, \delta^2, \delta^3, \delta^z$, and δ^M over δ^0 .

Range of the Parameters θ_i	$p = 6$				$p = 10$				
	δ^1	δ^2	δ^3	δ^z	δ^1	δ^2	δ^3	δ^z	δ^M
(0, 4)	7.1	8.8	5.5	7.0	15.6	9.7	12.0	8.3	28.5
(0, 8)	8.5	12.8	3.8	1.3	9.3	13.7	6.3	2.0	36.9
(8, 12)	5.4	15.0	2.3	.7	5.2	15.8	3.6	.8	42.7
(12, 16)	4.1	15.9	1.6	.3	3.7	17.4	2.4	.5	43.4
** (0, 12)	—	—	—	—	5.9	2.8	3.8	1.5	9.6
** (4, 16)	—	—	—	—	5.7	6.5	3.0	0.6	19.0

* δ^0 is the UMVUE, δ^1 shifts δ^0 towards $\lambda = (\lambda_1, \dots, \lambda_p)$ where $\lambda_i =$ integer part of θ_i , δ^2 shifts δ^0 towards the minimum of the Poisson observations, δ^3 shifts δ^0 towards the median of the Poisson observations, δ^z shifts δ^0 towards the origin (i.e., δ^1 with $\lambda_i = 0$ for all i), δ^M shifts δ^0 towards the “geometric mean” of the Poisson observations.

** simulation results for this range were only done for $p = 10$.

(2.7) through (2.10) except that the λ_i 's are replaced by $\text{med}(X)$, shifts the UMVUE toward the median of the X_i 's. That is,

$$\delta_i^3(X) = X_i - \{N(X) - 2\}_+ \{h(X_i) - h(\text{Med}(X))\} / D(X), \quad i = 1, \dots, p.$$

This adaptive estimator dominates the UMVUE under L_0 if $p \geq 6$. Although this estimator is intuitively appealing, the simulation results below show that its performance is not as good as that of δ^2 given in Example 2.2 when the Poisson parameters are large but close to each other. For purposes of comparison, we also recorded the performance of δ^z , an estimator which is the same as δ^1 except that all the λ_i 's are equal to zero. This estimator is similar to Peng's estimator.

Table 2 summarizes our computer simulation results for $p = 6$ and $p = 10$ when the loss function is L_0 ; the entries are the percentages of reduction in risk, $[R(\theta, \delta^0) - R(\theta, \delta)] \cdot 100\% / R(\theta, \delta^0)$, of the estimator δ over the UMVUE, δ^0 , where δ is $\delta^1, \delta^2, \delta^3, \delta^z$, or δ^M . Since $R(\theta, \delta^0) = \sum_{i=1}^p \theta_i$, the absolute reduction in risk can be very large for, say, δ^2 . The simulation procedure is as described in Tsui and Press (1982, pages 97–98), with the exception that the third step is repeated 1000 times instead of 2000 times. The θ_i 's are p random samples from a uniform distribution over a fixed interval. The simulation results indicate that δ^2 , the estimator that shifts the UMVUE towards the minimum of the Poisson observations, is good in almost all the situations when the parameters are close to one another. Its percentage of reduction in risk over the UMVUE increases as the parameters θ_i become larger but close to one another. Estimators δ^1 and δ^3 , which are described in Examples 2.1 and 2.3, respectively, perform remarkably well when the θ_i 's are small and p is large. This favorable performance of δ^1 is expected, since it adjusts the UMVUE to a point $\lambda = (\lambda_1, \dots, \lambda_p)$ close to the true θ_i 's. In this simulation study, λ_i was chosen to the integer part of θ_i . This is, of course, not realistic in practical situations (because λ_i depends on θ_i), but we desired to determine how well δ^1 could perform in most favorable situations. When the θ_i 's are large but close to one another, the performances of both δ^1 and δ^3 are somewhat disappointing. The percentages of reduction in risk for both δ^1 and δ^3 over δ^0 decrease as the θ_i 's become larger. Looking closely at (2.9) and (2.11) for δ^1 , and the corresponding expressions for δ^3 , we discover that this is because the expression $D(X)$ is dominated by the terms $\frac{1}{2} \{3h(\lambda_i) - 2\}_+$ in the case of δ^1 , and by the terms $\frac{1}{2} \{3h(\text{med}(X)) - 2\}_+$ in the case of δ^3 . These terms are likely to be large when the θ_i 's are large, and hence the reductions in risk are likely to be small. Furthermore, the $N(X)$ terms for δ^1 or δ^3 are likely to be smaller than $N(X)$ for δ^2 . These reasons partly explain why the risk reduction percentage of δ^2 is bigger than those of δ^1 or δ^3 when the θ_i 's are large. For comparison, simulation results for relatively larger ranges of θ_i 's were performed. Esti-

TABLE 3
Percentage reductions in risk under L_1 for the estimators* δ^4 , δ^5 over δ^0 .

Range of the Parameters θ_i	$p = 6$		$p = 10$	
	δ^4	δ^5	δ^4	δ^5
(0, 4)	25.2	26.0	26.3	31.4
(4, 8)	11.4	22.6	11.3	23.7
(8, 12)	5.9	17.8	6.9	20.4
(12, 16)	4.9	17.0	5.2	18.6
** (0, 12)	—	—	12.4	13.3
** (4, 16)	—	—	7.3	13.9

* δ^4 shrinks δ^0 towards the origin, δ^5 shrinks δ^0 towards the minimum of the Poisson observations.

** Simulation results for this range were only done for $p = 10$.

mators δ^1 , δ^2 and δ^3 all have similar performances which are better than δ^2 's. Although δ^M , given in (2.14), is not known to dominate δ^0 , its performance was tested out of curiosity. The simulation results in Table 2 show that, rather surprisingly, δ^M 's performance is impressive both when the θ_i 's are close to one another as well as when they are relatively far apart.

We next consider the Poisson means estimators which dominate the UMVUE under normalized squared error, $L_1(m_i = 1$ for all i in (1.1)). Motivation for using such a loss function has been discussed by Clevenson and Zidek (1975) and Tsui and Press (1982).

EXAMPLE 2.4. Consider the estimator δ^4 with correction term

$$\theta_i(X) = -(p-1)X_i/(\sum_{j=1}^p X_j + p-1), \quad i = 1, \dots, p.$$

This estimator, proposed by Clevenson and Zidek (1975), dominates $\delta^0(X) = X$ under L_1 . In fact, $\mathcal{D}(X)$ given in (2.6) becomes

$$(2.16) \quad \mathcal{D}(X) \leq -\frac{1}{2} (p-1)^2 / (\sum_{j=1}^p X_j + p).$$

Therefore, by Jensen's inequality, the reduction in risk, $2E_\theta \mathcal{D}(X)$, of δ^4 over δ^0 , is at least $(p-1)^2 / (\sum_{j=1}^p \theta_j + p)$. Large savings are thus expected when the sum $\sum_{j=1}^p \theta_j$ is close to zero. This is intuitively clear because δ^4 shrinks δ^0 towards the origin. What if the θ_i 's are large? As above, it should be advantageous to shrink δ^0 toward some prior guess of θ . In Section 3, we show that this is true under the general loss (1.1). Moreover, the following adaptive estimator is proved, in Section 4, to dominate δ^0 under L_1 .

EXAMPLE 2.5. Recall $X_{(1)} = \min_{i=1}^p(X_i)$. Let $N(X) = \#\{i: X_i > X_{(1)}\}$. Define $g_i(X) = X_i - X_{(1)}$, $D(X) = \sum_{i=1}^p g_i(X)$ and

$$(2.17) \quad \psi_i(X) = \frac{-[N(X) - 1]_+ g_i(X)}{D(X)}, \quad i = 1, \dots, p.$$

Consider the estimator $\delta^5(X)$ with correction term $\phi(X)$, whose i th component is $\phi_i(X) = \psi_i(X - e_i)$, where $\psi_i(X)$ is as given in (2.17). This estimator dominates $\delta^0(X) = X$ under L_1 , and $\mathcal{D}(X)$ given in (2.6) satisfies the inequality

$$(2.18) \quad \mathcal{D}(X) \leq -\frac{1}{2} \{N(X) - 1\}_+^2 / D(X),$$

indicating that $\mathcal{D}(X)$ can be sizable if all the θ_i 's are large, as long as they are close to one another. This is not so for δ^4 , whose corresponding $\mathcal{D}(X)$ is likely to be small when the θ_i 's are large. We thus expect that δ^5 will perform better than δ^4 when the θ_i 's are large, especially if the θ_i 's are close to one another. Our simulation results reported in Table 3 show that the performance of δ^5 is outstanding.

The use of the loss function (1.1) with positive m_i 's implies that a heavy penalty is imposed for overestimating small parameters θ_i . Since uniform domination requires that $R(\theta, \delta) \leq R(\theta, \delta^0)$ be true for all θ , it is known to be very difficult to shift the UMVUE, δ^0 , upward. This is the motivation for the construction of δ^5 as well as estimators proposed in Sections 3 and 4 when such a loss is considered.

The general theory for the discrete exponential family under loss (1.1) will be developed in the next two sections.

3. Adjusting the UMVUE towards a fixed point. Suppose $X_i, i = 1, \dots, p$ are p independent random variables having discrete density of the form

$$(3.1) \quad f(X_i | \theta_i) = \xi_i(\theta_i) t_i(X_i) \theta_i^{X_i}, \quad X_i = 0, 1, \dots,$$

where θ_i is a positive unknown parameter varying in a certain known interval. The problem of interest is estimating $\theta = (\theta_1, \dots, \theta_p)$ based on the vector of observations $X = (X_1, \dots, X_p)$. The UMVUE of θ_i is $\delta_i^0(X_i) = t_i(X_i - 1)/t_i(X_i)$, where we define $t_i(y) = 0$ if $y < 0$. To improve upon the estimator $\delta^0(X) = (\delta_1^0(X), \dots, \delta_p^0(X))$, one writes the competitor $\delta^*(X)$ as $\delta^*(X) + \phi(X)$, where $\phi(X) = (\phi_1(X), \dots, \phi_p(X))$. Our goal in Sections 3 and 4 is to find various $\phi(X)$ such that δ^* dominates δ^0 under the loss L_m given by (1.1).

As in Section 2, it is assumed that the correction term $\phi(X)$ satisfies regularity conditions (2.1) and (2.2). Under the loss (1.1), the difference in risk of δ^* and δ^0 equals (cf. Hwang, 1979, pages 20–22, or Tsui and Press, 1982)

$$(3.2) \quad R(\theta, \delta^*) - R(\theta, \delta^0) = 2E_\theta \mathcal{D}(X)$$

where

$$(3.3) \quad \mathcal{D}(X) = \sum_{i=1}^p \left\{ \frac{t_i(X_i + m_i - 1)}{t_i(X_i)} \Delta_i \phi_i(X + m_i e_i) + \frac{t_i(X_i + m_i)}{2t_i(X_i)} \phi_i^2(X + m_i e_i) \right\}.$$

Examples of (3.3) for Poisson X_i 's are given in (2.4) for the loss function (1.1), in (2.5) for L_0 , and in (2.6) for L_1 .

As mentioned before, in order to find an estimator δ^* that improves upon δ^0 , it suffices to find ϕ such that $\mathcal{D}(X) \leq 0$, with

$$(3.4) \quad E_\theta \mathcal{D}(X) < 0 \quad \text{for some } \theta.$$

The inequality $\mathcal{D}(X) \leq 0$ can be rewritten as

$$(3.5) \quad \mathcal{D}(X) = \sum_{i=1}^p \{v_i(X_i) \Delta_i \psi_i(X) + w_i(X) \psi_i^2(X)\} \leq 0$$

with $v_i(X_i) = t_i(X_i + m_i - 1)/t_i(X_i)$, $w_i(X) = t_i(X_i + m_i)/\{2t_i(X_i)\}$, and $\psi_i(X) = \phi_i(X + m_i e_i)$. The solutions to (3.5) that are provided in this paper apply to general v_i and w_i functions. Our solutions are obtained under:

ASSUMPTION AI. For each $i = 1, \dots, p$, there exists some α_i for which $v_i(X_i) > 0$ whenever $X_i \geq \alpha_i$.

In the Poisson problems, this assumption is satisfied for $\alpha_i = 1$ under L_0 , and for $\alpha_i = 0$ under L_1 .

Our solutions to (3.5), similar to those in Hwang (1982), have the form

$$(3.6) \quad \psi_i(X) = \frac{-C(X)H_i(X_i)}{D(X)}, \quad \text{where } D(X) = \sum_{j=1}^p d_j(X_j).$$

The functions H_i , $C(X) \geq 0$ and $d_i(X_i) \geq 0$ are specified in Assumptions AII through AIV below. We first explain the essence of these functions and then state the conditions on them more precisely. The function $C(X)$ is essentially a positive constant, and is often taken to be a function of the number of X_i 's that are larger than prespecified or data-based values. The function H_i is typically of the form

$$(3.7) \quad H_i(X_i) = h_i(X_i) - h_i(\lambda_i), \quad h_i(X_i) = \sum_{j=\lambda_i}^{X_i} v_i^{-1}(j),$$

where λ_i is an integer chosen according to some prior information about θ_i (cf. Example 2.1). The function $d_i(X_i)$ will be, for some nonnegative constants β_i and b_i , similar to $H_i^{\beta_i}(X_i) + b_i$. Nontrivial solutions are provided only for $p > \max_{1 \leq j \leq p} \beta_j$. It is reasonable to have some restriction on p , since in many cases, estimators can be improved only when p is large.

The precise assumptions on H_i and d_i are as follows:

ASSUMPTION AII. For an integer $\lambda_i \geq \alpha_i - 1$, $H_i(X_i) = H_i(X_i | \lambda_i)$ is an arbitrary nondecreasing function such that $\Delta_i H_i(X_i) \geq v_i^{-1}(X_i)$ if $X_i > \lambda_i$ and $H_i(\lambda_i) = 0$.

ASSUMPTION AIIL. Let $d_i(X_i) = d_i(X_i | \lambda_i)$ be a nonnegative function satisfying

$$(3.8) \quad v_i(X_i)H_i(X_i - 1)\Delta_i d_i(X_i) \leq \beta_i \min(d_i(X_i - 1), d_i(X_i))$$

for all $X_i \neq \lambda_i$ and some nonnegative constant β_i . (Note that (3.8) is automatically satisfied if $X_i = \lambda_i + 1$.)

ASSUMPTION AIV. There exists a finite constant K so that

$$(3.9) \quad \sum_{i=1}^p w_i(X) H_i^2(X_i) \leq KD(X).$$

We now provide some solutions to the difference inequality (3.5). Define $\beta = \max_{1 \leq i \leq p} \beta_i$.

THEOREM 3.1. Under Assumptions AI through AIV, $\psi(X) = (\psi_1(X), \dots, \psi_p(X))$ given in (3.6) is a solution to (3.5) where for all X , $C(X)$ satisfies

$$(3.10) \quad H_i(X_i - 1)\Delta_i C(X) \geq 0$$

and

$$(3.11) \quad 0 \leq C(X) \leq K^{-1}(N(X) - \beta)_+.$$

Furthermore,

$$(3.12) \quad \mathcal{D}(X) \leq -C(X)\{N(X) - \beta - KC(X)\}_+/D(X),$$

with strict inequality for those X satisfying the following condition:

$$(3.13) \quad C(X)H_i(X_i - 1)\Delta_i d_i(X_i) > 0 \quad \text{for at least two } i\text{'s}.$$

PROOF. Clearly,

$$\Delta_i \psi_i(X) = \frac{-C(X)H_i(X_i)}{D} + \frac{C(X - e_i)H_i(X_i - 1)}{D(X - e_i)} \leq -C(X)\Delta_i \{H_i(X_i)/D(X)\},$$

where the last inequality follows from (3.10). Now, by defining $D_i = D(X - e_i)$, direct calculation shows that

$$-\Delta_i \{H_i(X_i)/D(X)\} = \frac{-\Delta_i H_i(X_i)}{D} + \frac{H_i(X_i - 1)\Delta_i D}{DD_i}.$$

Let $D' = \sum_{j=1}^p \min\{d_j(X_j - 1), d_j(X_j)\}$. Since $X_i > \lambda_i$ implies $X_i \geq \alpha_i$, it follows that

$$(3.14) \quad \begin{aligned} \sum v_i(X_i)\Delta_i \psi_i(X) &\leq \frac{C(X)}{D} \sum \left[-v_i(X_i)\Delta_i H_i(X_i) + \frac{\{v_i(X_i)H_i(X_i - 1)\Delta_i d_i(X_i)\}_+}{D_i} \right] \\ &\leq \frac{C(X)}{D} [-N(X) + \sum \{v_i(X_i)H_i(X_i - 1)\Delta_i d_i(X_i)\}_+/D']. \end{aligned}$$

In the last transition, the inequality is strict for those X satisfying (3.13). By (3.8), the upper bound in (3.14) is, in turn, bounded above by $C(X)(\beta - N(X))/D$. From Assumption AIV, it follows that $\sum \psi_i^2(X)w_i(X) \leq KC^2(X)/D$. These last two statements then imply that

$$\mathcal{D}(X) \leq -C(X)\{N(X) - \beta - KC(X)\}/D.$$

By condition (3.11), $C(X)\{N(X) - \beta - KC(X)\} = C(X)\{N(X) - \beta - KC(X)\}_+$, and hence (3.12) is established.

One remaining question is how to choose each d_i so that it satisfies (3.8) and has a simple form. The functions $d_i(X_i)$ proposed in this paper are non-increasing for $X_i \leq \lambda_i$ and nondecreasing for $X_i \geq \lambda_i + 1$. Thus, for $X_i \geq \lambda_i + 1$, condition (3.8) reduces to

$$(3.15) \quad v_i(X_i)H_i(X_i - 1)\Delta_i d_i(X_i) \leq \beta_i d_i(X_i - 1),$$

which is the inequality considered in Hwang (1982). Simple choices of d_i 's satisfying (3.15) as given in Hwang (1982) will be used here for $X_i \geq \lambda_i + 1$. Therefore, in the following corollaries, we focus on the choice of d_i only for $X_i < \lambda_i$. Recall that $\sum_{i=a}^b S_i$ is defined to be zero whenever $a > b$.

COROLLARY 3.1. *Let*

$$h_i(X_i) = \sum_{j=\alpha_i}^{X_i} v_i^{-1}(j)$$

and $H_i(X_i) = h_i(X_i) - h_i(\lambda_i)$. Define, for $X_i < \lambda_i$,

$$d_i(X_i) = |H_i(X_i)|^{\beta_i} + a_i,$$

where β_i is an arbitrary number greater or equal to one and

$$a_i = \max_{\alpha_i \leq X_i \leq \lambda_i - 1} \{ |H_i(X_i - 1)|^{\beta_i} - |H_i(X_i)|^{\beta_i} \}_+.$$

Then H_i satisfies Assumption AII and $d_i(X_i)$ satisfies AIII for $X_i < \lambda_i$. In particular, if $\beta_i = 2$ and $v_i(X_i)$ is nondecreasing, the same conclusion holds for $a_i = v_i^{-1}(\alpha_i)\{\frac{3}{2} h_i(\lambda_i) - h_i(\alpha_i)\}_+$.

PROOF. See the Appendix.

COROLLARY 3.2. *Let $H_i(X_i) = \sum_{j=\lambda_i+1}^{X_i} v_i^{-1}(j)$, or $-\mu_i$ according as $X_i \geq$, or $< \mu_i$, where μ_i is a nonnegative constant. Assume for some nonnegative constant b_i , $d_i(X_i) = b_i$ for $X_i < \lambda_i$. Then H_i satisfies AII and $d_i(X_i)$ satisfies AIII for $X_i < \lambda_i$.*

PROOF. Assumption AII is trivially satisfied. AIII is satisfied for $X_i < \lambda_i$, since $\Delta_i d_i(X_i) = 0$ for $X_i < \lambda_i$.

The following examples illustrate the application of Theorem 3.1 in solving difference inequalities in order to develop improved estimators. Example 3.1 slightly generalizes Example 2.1 in Section 2.

EXAMPLE 3.1 (Under L_0). Let X_1, \dots, X_p be independent Poisson random variables with means $\theta_1, \dots, \theta_p$, respectively. To improve upon $\delta^0(X) = X$ under L_0 , it is desired to solve (2.5). Let $H_i(X_i) = h(X_i) - h(\lambda_i)$ where $h(\cdot)$ is as given in (2.8). Let $\lambda_i \geq \alpha_i - 1 = 0$ be an integer and let $d_i(X_i)$ be as given in (2.9). Assumption AIV is clearly satisfied with $K = \frac{1}{2}$. Condition (3.8) is satisfied when $X_i < \lambda_i$, by Corollary 3.1, and by Corollary 2.2.1 in Hwang (1979) when $X_i \geq \lambda_i + 1$. The rest of the assumptions, AI and AII, are trivially satisfied. It follows that $\phi^0(X) = (\phi_1^0(X), \dots, \phi_p^0(X))$ with,

$$\phi^0(X) = -c\{N(X) - 2\}_+ H_i(X_i)/D(X)$$

is a solution to (2.5) for $0 \leq c \leq 2$ and $N(X)$ as in (2.7). For $0 < c \leq 2$ and $p > 2$, (2.1) and (2.2) are clearly satisfied and, by Theorem 3.1, (3.4) is satisfied. (When $c = 2$, there are infinitely many X satisfying (3.13).) Therefore, $X + \phi^0(X)$ dominates X for $0 < c \leq 2$ and $p > 2$. The optimum choice of c , in the sense of minimizing the right hand side of (3.12), is $c = 1$.

EXAMPLE 3.2 (Continuation of Example 3.1). Corollary 3.2 also provides alternative solutions. Let $H_i(X_i) = h(X_i) - h(\lambda_i)$ or $-\mu_i$, according as $X_i \geq \lambda_i$ or $X_i < \lambda_i$, where μ_i is some nonnegative constant. Furthermore, define $d_i(X_i)$ as $H_i(X_i)H_i(X_i + 1)$ or μ_i^2 , depending on whether $X_i \geq \lambda_i$ or $X_i < \lambda_i$. Again, $X + \phi^0(X)$ dominates X for $0 < c \leq 2$ and $p > 2$. These estimators are similar to those proposed in Tsui (1979a). The recommended choice for c is 1.

In Example 3.3 below, Theorem 3.1 and Corollaries 3.1 and 3.2 can be applied to solve (2.6), derived under L_1 , nontrivially. However, because some of the solutions fail (2.1), not every solution corresponds to an estimator dominating X . This forces us to consider H_i and d_i only of the form in Corollary 3.2 with $\mu_i = 0$. In general, this difficulty is encountered when L_m with some positive m_i is used.

EXAMPLE 3.3 (Continuation of Example 3.1 under L_1). To apply Theorem 3.1 and Corollary 3.2 to solve (2.6), let $\mu_i = 0$, $H_i(X_i) = (X_i - \lambda_i)_+$ and $d_i(X_i) = H_i(X_i)$. With such choices and $\beta_i = 1$, Assumption AIII is satisfied for $X_i > \lambda_i$ by direct calculation, and is satisfied for $X_i < \lambda_i$ by Corollary 3.2. Condition AIV is clearly satisfied with $K = 1/2$. Assumptions AI and AII also hold for $\alpha_i = 0$. Let $\psi(X)$ be as in (3.6) with $C(X)$ nondecreasing in each coordinate (which implies (3.10)) and $0 \leq C(X) \leq 2\{N(X) - 1\}_+$, with $N(X)$ as in (2.7). Furthermore, assume $C(X) \neq 0$. It then follows that for $p > 1$, X is dominated by $X + \phi(X)$, where the i th component of $\phi(X)$ is $\psi_i(X - e_i)$.

The following result provides solutions to the difference inequality (3.5) under the most general loss L_m , given by (1.1).

THEOREM 3.2. Let X_1, \dots, X_p be as in Example 3.1. For some nonnegative integers λ_i , define $H_i(X_i) = \{h_i(X_i) - h_i(\lambda_i)\}_+$, with h_i as specified below according to the value of m_i .

(i) If $m_i = 0$, define $\alpha_i = 1$, $h_i(X_i) = \sum_{j=1}^{X_i} j^{-1}$, and $d_i(X_i) = H_i(X_i + 1)H_i(X_i)$.

(ii) If $m_i > 0$, define $\alpha_i = 0$, $h_i(X_i) = m_i^{-1}(X_i + m_i)^{(m_i)}$, and $d_i(X_i) = m_i^{-1}H_i(X_i)$.

Let ψ_i be as in (3.6), with $C(X) \neq 0$ nondecreasing in each coordinate and satisfying $0 \leq C(X) \leq 2\{N(X) - \max_{1 \leq j \leq p}(\alpha_j + 1)\}_+$, and let $N(X)$ be as in (2.7). If $p > \max_{1 \leq j \leq p}(\alpha_j + 1)$, then $\delta^*(X)$, with i th component $X_i + \psi_i(X - m_i e_i)$, dominates the usual estimator, $\delta^0(X) = X$.

The proof of Theorem 3.2 is similar to that of Example 3.2 and is omitted. The recommended choice of $C(X)$ is $\{N(X) - \max_{1 \leq j \leq p}(\alpha_j + 1)\}_+$.

All the above improved estimators shrink the UMVUE towards a point with integer coordinates. More general improved estimators shrinking towards a point with noninteger coordinates can be constructed by modifying Theorems 3.2 and 3.1. For details, see Tsui (1981b).

Theorem 3.1 will next be applied to the situation in which X_i , $i = 1, \dots, p$, are independent negative binomial random variables measuring the number of successes before the r_i th failure. More precisely,

$$P_{\theta_i}(X_i = x_i) = \binom{r_i + x_i - 1}{r_i - 1} \theta_i^{x_i} (1 - \theta_i)^{r_i}, \quad x_i = 0, 1, \dots$$

The UMVUE of the success probability vector, $\theta = (\theta_1, \dots, \theta_p)$, is $\delta^0(X) = (\delta_1^0(X), \dots, \delta_p^0(X))$, where $\delta_i^0(X_i) = X_i / (r_i + X_i - 1)$. We will consider only the squared error loss, L_0 ; a negative binomial example under the more general loss L_m can be found in Tsui (1981b). Estimators dominating $\delta^0(X)$ for $p \geq 3$, have i th component

$$(3.16) \quad \delta_i^0(X) - c\{N(X) - 2\}_+ H_i(X_i)/D,$$

where $D = D(X) = \sum_{j=1}^p d_j(X_j)$, $N(X)$ is as in (2.7), and $0 < c \leq 2$. The functions H_i and d_i are specified below, and two choices of these functions are given. In both of these cases

below, $h_i(X_i) = \sum_{j=1}^{X_i} (r_i + j - 1)/j$.

$$(i) \quad \begin{aligned} H_i(X_i) &= h_i(X_i) - h_i(\lambda_i) \text{ and} \\ d_i(X_i) &= \begin{cases} H_i^2(X_i) + b_i H_i(X_i) & \text{if } X_i \geq \lambda_i \\ H_i^2(X_i) + a_i & \text{if } X_i < \lambda_i, \end{cases} \end{aligned}$$

where $a_i = r_i \{3h_i(\lambda_i)/2 - 1\}_+$ and $b_i = (r_i + \lambda_i + 1)/(\lambda_i + 2)$.

$$(ii) \quad H_i(X_i) = \begin{cases} h_i(X_i) - h_i(\lambda_i) & \text{if } X_i \geq \lambda_i \\ -\mu_i & \text{if } X_i < \lambda_i \end{cases}$$

for some nonnegative μ_i and

$$d_i(X_i) = \begin{cases} H_i^2(X_i) + b_i H_i(X_i) & \text{if } X_i \geq \lambda_i \\ \mu_i^2 & \text{if } X_i < \lambda_i, \end{cases}$$

where b_i is as given in (i).

The negative binomial result is summarized in the following theorem, whose proof is again similar to the one in Example 3.2 and is omitted.

THEOREM 3.3. *For the negative binomial problem described above, if the functions H_i and d_i are as in either (i) or (ii), then the estimator given componentwise in (3.16) dominates the UMVUE $\delta^0(X)$ under L_0 , provided $p > 2$.*

The recommended choice for c is 1.

4. Adaptive estimators for the discrete exponential family. This section develops improved estimators that shrink the UMVUE towards the n th order statistic of the observations. Shrinking the UMVUE towards a point which depends on the observations, rather than towards a constant point (as in Section 3) creates some technical problems in proving domination. However, the difficulties can be surmounted with the aid of the following lemmas. Their proofs are direct and are omitted.

LEMMA 4.1. *Let $y = (y_1, \dots, y_p)$ be a sequence of integers. If $y_i \neq y_{(n)}$, then $(y - e_i)_{(n)} = y_{(n)}$.*

LEMMA 4.2. *Let y be as in Lemma 4.1. If $y_i = y_{(n)}$, then $y_{(n)} - 1 \leq (y - e_i)_{(n)}$.*

Other solutions to the difference inequality (3.5) are provided in Theorem 4.1 below. This theorem can be used to generate improved estimators which shrink the UMVUE towards $X_{(n)}$. The solutions proposed, similar to (3.6) are of the form

$$(4.1) \quad \psi_i(X) = -C(X)H_i(X)/D(X), \text{ where } D(X) = \sum_{j=1}^p d_j(X).$$

We assume that the functions H_i and d_i satisfy Assumptions BI through BIV below, which are modifications of Assumptions AII through AIV.

ASSUMPTION BI. $H_i(X)$ and $d_i(X) \geq 0$ depend on X only through X_i and $X_{(n)}$.

ASSUMPTION BII. $\Delta_i H_i(X) \geq 0$ for all X , $\Delta_i H_i(X) \geq v_i^{-1}(X_i)$ for $X_i > X_{(n)}$, and $H_i(X) = 0$ whenever $X_i = X_{(n)}$.

ASSUMPTION BIII. For $X_i \neq X_{(n)}$,

$$(4.2) \quad |v_i(X_i)H_i(X - e_i)\Delta_i d_i(X)| \leq \beta_i \min(d_i(X - e_i), d_i(X)).$$

(Inequality (4.2) is clearly satisfied if $X_i = X_{(n)} + 1$, since $H_i(X - e_i) = 0$ by BI, BII, and Lemma 4.1.)

ASSUMPTION BIV. $\sum_{i=1}^p w_i(X)H_i^2(X)/\sum_{j=1}^p d_j(X) \leq K$, for some positive constant K .

Define

$$(4.3) \quad \beta = \max_{1 \leq i \leq p} \beta_i \quad \text{and} \quad N(X) = \#\{i: X_i > X_{(n)}\}.$$

THEOREM 4.1. *Suppose Assumption AI, and Assumptions BI through BIV hold, and that $C(X)$ satisfies*

$$(4.4) \quad 0 \leq C(X) \leq K^{-1}(N(X) - \beta)_+$$

and

$$(4.5) \quad H_i(X - e_i) \Delta_i C(X) \geq 0.$$

Then ψ , given componentwise in (4.1), is a solution to (3.5) for all X with $X_{(n)} \geq \max_{1 \leq j \leq p} \alpha_j - 1$. Furthermore, for such X ,

$$(4.6) \quad \mathcal{D}(X) \leq -C(X) \{N(X) - \beta - KC(X)\}_+ / D,$$

with strict inequality for those X satisfying

$$(4.7) \quad (X_i - X_{(n)})C(X)H_i(X - e_i)\Delta_i d_i(X) \neq 0$$

for at least two i 's.

The proof of Theorem 4.1 is similar to that of Theorem 3.1, and is provided in the Appendix. Because of Lemma 4.1, $X_{(n)}$ is essentially constant (like λ_i in Theorem 3.1) with respect to the operator Δ_i , except when $X_{(n)} = X_i$. In that case, Lemma 4.2 applies.

Corollaries 4.1 and 4.2 below, which provide choices of d_i satisfying (4.2), parallel Corollaries 3.1 and 3.2, and can be similarly proved. Corollary 4.1 concentrates only on the case $\beta_i = 2$.

COROLLARY 4.1. *Suppose that $v_i(X_i)$ is nondecreasing in X_i . Let $h_i(X_i) = \sum_{j=\alpha_i}^{X_i} v_i^{-1}(j)$ and $H_i(X_i) = h_i(X_i) - h_i(X_{(n)})$. Define $d_i(X) = H_i^2(X) + a_i$, where*

$$a_i = v_i^{-1}(\alpha_i) \{\frac{3}{2}h_i(X_{(n)}) - h_i(\alpha_i)\}_+.$$

Then d_i and H_i satisfy (4.2) for $X_i < X_{(n)}$.

COROLLARY 4.2. *Let $H_i(X) = h_i(X_i) - h_i(X_{(n)})$ or $-\mu_i$, according as $X_i \geq$ or $< X_{(n)}$, where μ_i is an arbitrary nonnegative constant and h_i is as defined in Corollary 4.1. For $X_i < X_{(n)}$, define $d_i(X) = b_i$ for some nonnegative constant b_i . Then d_i and H_i satisfy (4.2) for all X with $X_i < X_{(n)}$.*

The following examples illustrate the application of Theorem 4.1.

EXAMPLE 4.1 (continuation of Example 3.1). To apply Theorem 4.1 to the difference inequality (2.5), let $h(X_i)$ be as in (2.8), $H_i(X) = h(X_i) - h(X_{(n)})$, and

$$d_i(X) = \begin{cases} H_i(X)H_i(X + e_i) & \text{for } X_i \geq X_{(n)} \\ H_i^2(X) + a_i(X) & \text{for } X_i < X_{(n)}, \end{cases}$$

where $a_i(X) = \{\frac{3}{2}h(X_{(n)}) - 1\}_+$. Take

$$\phi_i(X) = -c\{N(X) - 2\}_+ H_i(X) / \sum_{j=1}^p d_j(X),$$

where $N(X)$ is as in (4.3). Direct calculation and Corollary 4.1 show that (4.2) is satisfied with $\beta_i = 2$. Assumption BIV is also satisfied with $K = \frac{1}{2}$. The remaining assumptions in Theorem 4.1 are clearly satisfied. This implies that ϕ is a solution to (2.5) when $p > n + 2$ and $0 < c \leq 2$. Therefore, $X + \phi(X)$ dominates X under L_0 when $p > n + 2$ and $0 < c \leq 2$. The special case where $n = 1$ and $c = 1$ was given in Example 2.2.

REMARK. Another choice of H_i and d_i is

$$H_i(X) = \begin{cases} h_i(X_i) - h_i(X_{(n)}) & X_i \geq X_{(n)}, \\ -\mu_i & X_i < X_{(n)}, \end{cases}$$

for some constant $\mu_i \geq 0$, and

$$d_i(X) = \begin{cases} H_i(X)H_i(X + e_i) & \text{if } X_i \geq X_{(n)}, \\ \mu_i^2 & \text{if } X_i < X_{(n)}. \end{cases}$$

The corresponding estimator, $X + \phi(X)$, again dominates X when $p > n + 2$ and $0 < c \leq 2$.

Theorem 4.2 below provides some adaptive improved Poisson means estimators under the general loss function L_m .

THEOREM 4.2. *Let X_1, \dots, X_p be as in Example 3.1. The functions H_i and d_i are defined according to the value of m_i . Let h_i be as in (i) and (ii) of Theorem 3.2. In both cases below, $H_i(X) = \{h_i(X_i) - h_i(X_{(n)})\}_+$. (i) $m_i = 0$: $\alpha_i = 1$ and $d_i(X) = H_i(X + e_i)H_i(X)$, (ii) $m_i > 0$: $\alpha_i = 0$ and $d_i(X) = m_i^{-1}H_i(X)$.*

Let $\psi_i(X) = -c(X)H_i(X)/\sum_{j=1}^p d_j(X)$, where $c(X) \neq 0$ is nondecreasing in X_i for $X_i > X_{(n)}$ and satisfies

$$(4.8) \quad 0 \leq c(X) \leq 2\{N(X) - \max_{1 \leq j \leq p} (\alpha_j + 1)\}_+,$$

and let $N(X)$ be as in (4.3). If $p > n + \max_{1 \leq j \leq p} \alpha_j + 1$, then $\delta^*(X)$, with i th coordinate $X_i + \psi_i(X - m_i e_i)$, dominates the usual estimator, $\delta^0(X) = X$.

The proof of Theorem 4.2 is similar to that of Theorem 3.2 and Example 3.2 and is omitted. A recommended choice of $c(X)$ is $\{N(X) - \max_{1 \leq j \leq p} (\alpha_j + 1)\}_+$, which satisfies the assumptions of Theorem 4.2.

When all the X_i 's belong to the negative binomial family, a theorem similar to Theorem 4.2 can be developed by another application of Theorem 4.1. The details, however, will not be given here. The result in the squared error loss case can be found in Ghosh and Hwang (1981).

APPENDIX

PROOF OF COROLLARY 3.1. Clearly, H_i satisfies AII. It is also obvious that d_i satisfies AIII for $X_i < \alpha_i$, since $H_i(X_i)$ is a constant for $X_i < \alpha_i$, and so is d_i . In what follows, we show that d_i satisfies AIII for $\alpha_i \leq X_i < \lambda_i$. Since $\beta_i \leq 1$, $d_i(X_i)$ is nonincreasing for $X_i < \lambda_i$. Assumption AIII is therefore equivalent to

$$(A.1) \quad v_i(X_i)H_i(X_i - 1)\Delta_i d_i(X_i) \leq \beta_i d_i(X_i).$$

By the Mean Value Theorem and the fact that $|H_i(X_i)| \leq |H_i(X_i - 1)|$ when $X_i < \lambda_i$,

$$(A.2) \quad \begin{aligned} v_i(X_i)H_i(X_i - 1)\Delta_i d_i(X_i) &\leq v_i(X_i)|H_i(X_i - 1)|\beta_i|H_i(X_i - 1)|^{\beta_i - 1}v_i^{-1}(X_i) \\ &= \beta_i|H_i(X_i - 1)|^{\beta_i}. \end{aligned}$$

Now the upper bound in the above expression is

$$\beta_i|H_i(X_i)|^{\beta_i} + \beta_i|H_i(X_i - 1)|^{\beta_i} - \beta_i|H_i(X_i)|^{\beta_i} \leq \beta_i|H_i(X_i)|^{\beta_i} + \beta_i a_i = \beta_i d_i(X_i).$$

Hence, the first part of this corollary is proved.

When $\beta_i = 2$, consider again only those X_i 's such that $d_i \leq X_i \leq \lambda_i$. Instead of using the Mean Value Theorem in deriving (A.2), we directly calculate the difference $\Delta_i d_i(X_i)$, and obtain

$$v_i(X_i)H_i(X_i - 1)\Delta_i d_i(X_i) = H_i(X_i - 1)(H_i(X_i) + H_i(X_i - 1)).$$

To show that the above expression is bounded above by $2d_i(X_i)$, it is sufficient to show that

$$(A.3) \quad H_i(X_i - 1)(H_i(X_i) + H_i(X_i - 1)) - 2H_i^2(X_i) \leq 2a_i.$$

The left hand side of (A.3) is

$$\{3h_i(\lambda_i) - 2h_i(X_i) - h_i(X_i - 1)\}v_i^{-1}(X_i) \leq \{3h_i(\lambda_i) - 2h_i(\alpha_i)\}v_i^{-1}(X_i).$$

This upper bound is, in turn, bounded by $2a_i$ (due to the monotonicity assumption of $v_i(\cdot)$), which establishes (A.3).

PROOF OF THEOREM 4.1. We first give two lemmas which are needed for the proof of Theorem 4.1. The proof of Lemma A.1 is straightforward and is omitted.

LEMMA A.1. *Let y be as in Lemma 4.1 and $y_i \neq y_{(n)}$. Then $(y + e_i)_{(n)} = y_{(n)}$.*

LEMMA A.2. *Assumption BII implies that $H_i(X) \leq 0$ if $X_i < X_{(n)}$.*

PROOF OF LEMMA A.2. If $X_i = X_{(n)}$, then $H_i(X) = 0$ by Assumption BII. If $X_i = X_{(n)} - 1$, then Assumption BII implies that $H_i(X + e_i) - H_i(X) \geq 0$. From Lemma A.1, it follows that $(X + e_i)_{(n)} = X_{(n)} = X_{(n)} = X_i + 1$. The first statement in the proof implies that $H_i(X + e_i) = 0$, and hence $H_i(X) \leq 0$.

For $X_i = X_{(n)} - 2$, again we have $(X + e_i)_{(n)} = X_{(n)} = (X_i + 1) + 1$. By the result stated in the last paragraph, $H_i(X + e_i) \leq 0$. Repeating the above procedure, Lemma A.2 is established.

To prove Theorem 4.1, note that

$$(A.4) \quad \Delta_i \psi_i(X) = \frac{-C(X)H_i(X)}{D(X)} + \frac{C(X - e_i)H_i(X - e_i)}{D(X - e_i)}.$$

For those X with $X_i = X_{(n)}$, $H_i(X) = 0$ by assumption. By Lemma 4.2, $X_i - 1 = X_{(n)} - 1 \leq (X - e_i)_{(n)}$. It then follows from Lemma A.2 that $H_i(X - e_i) \leq 0$. Moreover, $w_i(X)\phi_i^2(X) = 0$, which implies that

$$(A.5) \quad \mathcal{D}(X) \leq \sum_{\{i: X_i \neq X_{(n)}\}} v_i(X_i) \Delta_i \phi_i(X) + w_i(X) \phi_i^2(X).$$

Therefore, we need consider only those X with $X_i \neq X_{(n)}$ below. By Lemma 4.1, Δ_i treats $X_{(n)}$ as a constant. It follows from (A.4) and (4.5) that

$$\Delta_i \psi_i(X) \leq -C(X) \Delta_i \{H_i(X)/D(X)\}.$$

Now

$$(A.6) \quad -\Delta_i \{H_i(X)/D(X)\} = \frac{-\Delta_i H_i(X)}{D(X)} + \frac{H_i(X - e_i) \Delta_i D(X)}{D(X) D(X - e_i)}.$$

Let $D' = \sum_{j=1}^p \min\{d_j(X - e_j), d_j(X)\}$. It then follows that

$$(A.7) \quad \begin{aligned} \sum v_i(X_i) \Delta_i \psi_i(X) &\leq \frac{C(X)}{D} \sum_{\{i: X_i \neq X_{(n)}\}} \left\{ -v_i(X_i) \Delta_i H_i(X) + \frac{|v_i(X_i) H_i(X_i - e_i) \Delta_i d_i(X)|}{D_i} \right\} \\ &\leq \frac{C(X)}{D} \{-N(X) + \sum_{\{i: X_i \neq X_{(n)}\}} |v_i(X_i) H_i(X - e_i) \Delta_i d_i(X)| / D'\}. \end{aligned}$$

In the last transition, the inequality is strict for those X satisfying (4.7). This is straightforward if one considers three cases: (i) $\Delta_i d_i(X) > 0$ for both i , (ii) $\Delta_i d_i(X) < 0$ for both i , and (iii) $\Delta_i d_i(X)$ is positive for one i and negative for the other.

By (4.2), the upper bound in (A.7) is, in turn, bounded by $C(X)(\beta - N(X))/D$. From

Assumption BIV, $\sum \psi_i^2(X) w_i(X) \leq KC^2(X)/D$. Together, these imply that

$$\mathcal{D}(X) \leq -C(X)\{N(X) - \beta - KC(X)\}/D.$$

By (4.4), $C(X)(N(X) - \beta - KC(X)) = C(X)(N(X) - \beta - KC(X))_+$, which establishes (4.6).

Acknowledgment. Thanks are due to Mr. Youngjo Lee who carried out the computer simulations, and to the Editor and the Associate Editor for their constructive suggestions regarding the overall presentation of the paper. Thanks are also due to Ella Mae Matsumura for her detailed editorial comments.

REFERENCES

- BERGER, J. O. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of Gamma scale parameters. *Ann. Statist.* **8** 545–571.
- BERGER, J., BOCK, M. E., BROWN, L. D., CASELLA, G., and GLESER, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771.
- BERGER, J. and HAFF, L. R. (1981). A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. Mimeograph Series #81-7, Purdue University.
- BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.
- BROWN, L. D. (1979). A heuristic method for determining admissibility of estimators—with applications. *Ann. Statist.* **7** 960–994.
- CLEVENSON, M. L., and ZIDEK, J. V. (1975). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.
- GHOSH, M. and HWANG, J. T. (1981). Shrinking estimators towards arbitrary points in discrete and absolutely continuous exponential family. Cornell Statistical Center Technical Report. Ser. 82-19.
- GHOSH, M., and PARSIAN, A. (1980). Admissible and minimax multiparameter estimation in exponential families. *J. Multivariate Anal.* **10** 551–564.
- GHOSH, M., and PARSIAN, A. (1981). Bayes minimax estimation of multiple Poisson parameters. *J. Multivariate Anal.* **11** 280–288.
- HODGES, J. L., JR. and LEHMAN, E. L. (1951). Some applications of the Cramer-Rao inequality. *Proc. 2nd Berk. Symp. Math. Statist. Probab.* 13–20.
- HUDSON, H. M. (1974). Empirical Bayes estimation. Technical Report No. 58, Department of Statistics, Stanford University.
- HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.
- HUDSON, H. M. (1981). Adaptive estimation for simultaneous estimation of Poisson means. Unpublished manuscript.
- HUDSON, H. M. and TSUI, K. W. (1981). Simultaneous Poisson estimators for a priori hypotheses about means. *J. Amer. Statist. Assoc.* **76** 182–187.
- HWANG, J. T. (1979). Improving upon inadmissible estimators in discrete exponential families. Technical Report No. 79-14, Department of Statistics, Purdue University.
- 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.
- JAMES, W., and STEIN, C. (1961). Estimation with quadratic loss. *Proc. 4th Berk. Symp. Math. Statist. Probab.* **1** 361–379.
- MATSUMURA, E. M. and TSUI, K. W. (1982). Stein-type Poisson estimators in audit sampling. *J. Accounting Research* **20** 162–170.
- PENG, J. C. M. (1975). Simultaneous estimation of the parameters of independent Poisson distributions. Technical Report No. 78, Department of Statistics, Stanford University.
- ROLPH, J. E., CHAIKEN, J. M. and HOUCHESS, R. L. (1981). Methods for estimating crime rates of individuals. The Rand Corporation, R-2730-NIJ.
- STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berk. Symp. Math. Statist. Probab.* **1** 197–206.
- STEIN, C. (1973). Estimation of the mean of a multivariate distribution. *Proc. Prague Symp. Asymp. Statist.* 345–381.
- STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.
- TSUI, K. W. (1978). Simultaneous estimation of several Poisson parameters under generalized

- k -normalized squared error loss. Technical Report No. 78-39, Institute of Applied Math. and Statist., University of British Columbia.
- TSUI, K. W. (1979a). Multiparameter estimation of discrete exponential distributions. *Can. J. Statist.* **7** 193–200.
- TSUI, K. W. (1979b). Estimation of Poisson means under weighted squared error loss. *Can. J. Statist.* **7** 201–204.
- TSUI, K. W. (1981a). Simultaneous estimation of several Poisson parameters under squared error loss. *Ann. Inst. Statist. Math.* **33** 214–223.
- TSUI, K. W. (1981b). Difference inequalities and inadmissibility of the usual estimator in multiparameter estimator of discrete exponential families. Technical Report No. 653, Department of Statistics, University of Wisconsin-Madison.
- TSUI, K. W. and PRESS, S. J. (1977). Simultaneous estimation of several Poisson parameters under k -normalized squared error loss. Working Paper No. 456, University of British Columbia.
- 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.

MALAY GHOSH
DEPARTMENT OF STATISTICS
507NSB
UNIVERSITY OF FLORIDA
GAINESVILLE, FL 32611

J. T. HWANG
DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

KAM-WAH TSUI
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706