JOURNAL OF MULTIVARIATE ANALYSIS 15, 183–200 (1984)

Trimmed Estimates in Simultaneous Estimation of Parameters in Exponential Families

MALAY GHOSH* AND DIPAK K. DEY

University of Florida and Texas Tech University

Communicated by A. Cohen

Let $X_1,..., X_p$ be p ($\geqslant 3$) independent random variables, where each X_i has a distribution belonging to the one-parameter exponential family of distributions. The problem is to estimate the unknown parameters simultaneously in the presence of extreme observations. C. Stein (*Ann. Statist.* **9** (1981), 1135–1151) proposed a method of estimating the mean vector of a multinormal distribution, based on order statistics corresponding to the $|X_i|$'s, which permitted improvement over the usual maximum likelihood estimator, for long-tailed empirical distribution functions. In this paper, the ideas of Stein are extended to the general discrete and absolutely continuous exponential families of distributions. Adaptive versions of the estimators are also discussed.

1. INTRODUCTION

Let $X_1,..., X_p$ be p ($\geqslant 3$) independent normal variables with respective means $\theta_1,..., \theta_p$ and common unit variance. For estimating $\theta = (\theta_1,..., \theta_p)$ under the loss $L(\theta, a) = \sum_{i=1}^{p} (\theta_i - a_i)^2$, the James–Stein estimator $\delta^0(X) = (1 - (p-2)/\sum_{i=1}^{p} X_i^2)X$ dominates $X = (X_1,..., X_p)$. Efron and Morris [6] noted that δ^0, while guaranteeing a reduction in the total risk of p of the maximum likelihood estimator (MLE) X, might do poorly in estimating those θ_i's with unusually large or small values. Accordingly, they proposed a compromise between $\delta^0(X)$ and X which consisted of using the components of δ^0 subject to a maximum deviation from the corresponding MLEs. An alternative compromise was proposed by Stein [15] based on order statistics corresponding to the $|X_i|$'s which permitted improvement over δ^0 for long-tailed empirical distribution functions. The resulting estimators could be viewed as trimmed versions of James–Stein estimators.

Received November 24, 1982; revised July 18, 1983.

AMS 1980 subject classifications: 62C15, 62F10, 62H99.

Key words and phrases: Exponential family, discrete, absolutely continuous, shrinkage estimators, trimmed estimators, adaptive estimators.

* Research supported by NSF Grant MCS-8202116.

0047-259X/84 $3.00

In this paper, we extend the ideas of Stein [15] to the general discrete and absolutely continuous exponential families of distributions. Section 2 is devoted to the development of a very general class of trimmed estimates for the natural parameter vector in the discrete exponential case. The important special cases involving the Poisson and negative binomial distributions are considered. In this process, trimmed versions of several estimators proposed earlier by Clevenson and Zidek [3], Peng [14], Hudson [12], Tsui [17, 18], Tsui and Press [19], Hwang [13], and Ghosh and Parsian [9] are developed. In Section 3, results analogous to those in Section 2 are obtained in the absolutely continuous case. Particular attention is paid to the special cases of normal and gamma distributions. In this process, trimmed versions of certain estimators proposed earlier by Baranchik [1], Strawderman [16], Efron and Morris [7], Berger [2], Ghosh and Parsian [8], and Ghosh *et al.* [11] are developed. Section 4 is devoted to the risk simulation study in the Poisson and gamma cases. In this section, certain additional adaptive trimmed estimators are also proposed in the Poisson and gamma cases, and their performance is studied through risk simulation. The proofs of some of the technical results are postponed until the Appendix.

2. Discrete Exponential Distribution

Let $X_1, ..., X_p$ be p independent discrete random variables, X_i having probability function (pf)

$$f_{\theta_i}(x_i) = P_{\theta_i}(X_i = x_i) = \pi(\theta_i)\, t_i(x_i)\, \theta_i^{x_i}, \qquad x_i = 0, 1, 2, ...,$$

where $\theta_i \in (0, \infty)$ is unknown, $i = 1, ..., p$. Then, the uniformly minimum variance unbiased estimator (UMVUE) of θ_i is given by $\delta_i^0(X_i) = t_i(X_i - 1)/t_i(X_i)$, where $t_i(-1)$ is defined as zero. Let $\delta^0(X) = (\delta_1^0(X_1), ..., \delta_p^0(X_p))$.

First assume the loss

$$L(\theta, a) = \sum_{i=1}^{p} (\theta_i - a_i)^2. \tag{2.1}$$

Write $\phi(X) = (\phi_1(X), ..., \phi_p(X))$ where it is assumed that

$$E_\theta[\phi_i^2(X)] < \infty \qquad \text{for all } i = 1, ..., p \text{ and all } \theta \in (0, \infty)^p. \tag{2.2}$$

Then, assuming the loss (2.1), the difference in the risk function of $\delta^0(X) + \phi(X)$ and $\delta^0(X)$ can be expressed as (cf. Hudson [12], Tsui [17], and Hwang [13])

$$R(\theta, \delta^0 + \phi) - R(\theta, \delta^0) = 2E_\theta u(X), \tag{2.3}$$

where

$$u(x) = \sum_{i=1}^{p} v_i(x_i)\, \Delta_i \phi_i(x) + \tfrac{1}{2} \sum_{i=1}^{p} \phi_i^2(x),$$
$$v_i(x_i) = \delta_i^0(x_i), \tag{2.4}$$
$$\Delta_i \phi_i(x) = \phi_i(x) - \phi_i(x - e_i),$$

and e_i is the ith p-dimensional vector whose ith coordinate is one and other coordinates are zero. Hence, the problem reduces to obtaining solutions to the difference inequality $u(x) < 0$.

Various solutions to the above difference inequality have been obtained by several authors (see, for example, Ghosh *et al.* [10] for further references). In this paper, we obtain yet another set of solutions to this difference inequality which leads to trimmed versions of the estimators obtained earlier.

We need to develop a few more notations before stating the first main result of this section. Define

$$h_i(x_i) = \sum_{k=1}^{x_i} v_i^{-1}(k), \qquad x_i = 1, 2, \ldots,$$
$$h_i(0) = 0. \tag{2.5}$$

Note that $h_i(x_i) \uparrow$ in x_i.

Suppose now that $x_{(1)} \leqslant \cdots \leqslant x_{(p)}$ are the ordered x_i's. Define

$$d_i(x_i) = h_i(x_i)\, h_i(x_i + 1) + b_0, \tag{2.6}$$

where b_0 ($\geqslant 0$) is a known constant. Let

$$D = D(x) = \sum_{i \ni x_i \leqslant x_{(l)}} d_i(x_i) + \sum_{i \ni x_i > x_{(l)}} d_i(x_{(l)}); \tag{2.7}$$

$$N(x) = \#\{i\colon 1 \leqslant x_i \leqslant x_{(l)}\}. \tag{2.8}$$

Define

$$\phi_i(X) = \begin{cases} -\dfrac{c(N(x)-2)^+}{D} h_i(x_i) & \text{if } \; x_i \leqslant x_{(l)}, \\ -\dfrac{c(N(x)-2)^+}{D} h_i(x_{(l)}) & \text{if } \; x_i > x_{(l)}, \end{cases} \tag{2.9}$$

where $0 < c < 2$ and $a^+ = \max(a, 0)$. We are now in a position to state the first main result of this section.

THEOREM 2.1. *Assume that $v_i(j)$ is nondecreasing in $j = 0, 1, 2, \ldots,$ for all $i = 1, \ldots, p$. Then defining $\phi_i(x)$ as in (2.9), $\phi(x) = (\phi_1(x), \ldots, \phi_p(x))$ provides a solution to $u(x) \leqslant 0$ for $l \geqslant 3$. Also, $P_\theta(u(X) < 0) > 0$ for every $\theta \in (0, \infty)^p$.*

Proof. We omit the proof of this theorem, because of its similarity to the proof of Theorem 2.3.

As a consequence of the above theorem, it follows that $(\delta_1^0(X_1) + \phi_1(X), \ldots, \delta_p^0(X_p) + \phi_p(X))$ dominates $(\delta_1^0(X_1), \ldots, \delta_p^0(X_p))$. This leads to a class of trimmed shrinkage estimators shrinking the UMVUE of θ towards zero. Also, the optimal choice of c is $c = 1$.

Remark. As two important applications of Theorem 2.1, consider the cases (i) $X_i \sim$ Poisson (θ_i), and (ii) $X_i \sim$ negative binomial (r_i, θ_i), i.e.,

$$P_{\theta_i}(X_i = x_i) = \binom{r_i + x_i - 1}{x_i} \theta_i^{x_i}(1 - \theta_i)^{r_i}, \tag{2.10}$$

$x_i = 0, 1, \ldots, r_i \geqslant 2, i = 1, \ldots, p$. In case (i), $\delta_i^0(x_i) = v_i(x_i) = x_i$ so that $h_i(0) = 0$ and $h_i(x_i) = \sum_{k=1}^{x_i} k^{-1}$. In case (ii), $v_i(x_i) = x_i/(x_i + r_i - 1)$ ($\uparrow$ in x_i) and $h_i(0) = 0$, $h_i(x_i) = \sum_{k=1}^{x_i} (r_i + k - 1)/k$ for $x_i = 1, 2, \ldots$. In both cases, it is easy to construct an appropriate class of trimmed shrinkage estimators dominating $\delta^0(X)$ by using (2.9).

In many instances, prior belief and other considerations dictate shrinking towards an arbitrarily specified point, not necessarily zero. Ghosh *et al.* [10] have investigated this possibility, and have constructed a class of estimators shrinking towards an arbitrarily specified point.

We shall now see how trimmed versions of the estimators proposed by Ghosh *et al.* [10] can be obtained in this context.

Suppose we want to shrink the estimate of θ_i towards a specified nonnegative integer λ_i. Define

$$d_i(x_i) = \begin{cases} b_0 + (h_i(x_i) - h_i(\lambda_i))(h_i(x_i + 1) - h_i(\lambda_i)) & \text{if } x_i \geqslant \lambda_i + 1, \\ (h_i(x_i) - h_i(\lambda_i))^2 + v_i^{-1}[\tfrac{3}{2}h_i(\lambda_i - 1) - h_i(1)]^+ & \text{for } x_i \leqslant \lambda_i, \end{cases} \tag{2.11}$$

where $b_0 \geqslant 0$. Let $x_{(1)} \leqslant \cdots \leqslant x_{(l)} \leqslant \cdots \leqslant x_{(p)}$ be the ordered x_i's. Let $D = \sum_{i \ni x_i \leqslant x_{(l)}} d_i(x_i) + \sum_{i \ni x_i > x_{(l)}} d_i(x_{(l)})$ and $N(x) = \#\{i : \lambda_i + 1 \leqslant x_i \leqslant x_{(l)}\}$. Define

$$\phi_i(x) = \begin{cases} -\dfrac{c(N(x) - 2)^+}{D} (h_i(x_i) - h_i(\lambda_i)) & \text{if} \quad x_i \leqslant x_{(l)}, \\ -\dfrac{c(N(x) - 2)^+}{D} (h_i(x_{(l)}) - h_i(\lambda_i)) & \text{if} \quad x_i > x_{(l)}, \end{cases} \tag{2.12}$$

where $0 < c < 2$. Then the following theorem is true.

THEOREM 2.2 *Let $v_i(j)$ be $\uparrow$ in j for all $i = 1,\dots, p$. Let $\phi_i(x)$, $i = 1,\dots, p$, be defined as in* (2.12). *Then* $(\delta_1^0(X_1) + \phi_1(X),\dots, \delta_p^0(X_p) + \phi_p(X))$ dominates $(\delta_1^0(X_1),\dots, \delta_p^0(X_p))$ *for all $l \geqslant 3$. The optimal choice for c is $c = 1$.*

The proof of this theorem is also omitted because of its similarity to the proof of Theorem 2.3. The Poisson and negative binomial examples follow as special cases by defining the $h_i(x_i)$'s appropriately.

Next, we consider certain trimmed estimators shrinking the usual estimator towards the minimum; untrimmed versions of such estimators are proposed in Ghosh *et al.* [10].

Assume that $v_1(x) = \cdots = v_p(x) = v(x)$ (say), and hence, $h_1(x) = \cdots = h_p(x) = h(x)$ (say). Such assumptions hold in the Poisson case or in the negative binomial case with $r_1 = \cdots = r_p$. Let $m = m(x) = x_{(1)}$ and $N(x) = \#\{i\colon x_i \geqslant m + 1\}$. Define

$$d(x_i) = \begin{cases} (h(x_i) - h(m))(h(x_i + 1) - h(m)) & \text{if } x_i \geqslant m + 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Let $D = \sum_{i \ni x_i \leqslant x_{(l)}} d(x_i) + \sum_{i \ni x_i > x_{(l)}} d(x_{(l)})$, and define

$$\phi_i(x) = \begin{cases} -\dfrac{c(N(x) - 2)^+}{D} (h(x_i) - h(m)) & \text{if } m + 1 \leqslant x_i \leqslant x_{(l)}, \\ -\dfrac{c(N(x) - 2)^+}{D} (h(x_{(l)}) - h(m)) & \text{if } x_i > x_{(l)}. \end{cases} \tag{2.14}$$

Then $(\delta_1^0(X_1) + \phi_1(X),\dots, \delta_p^0(X_p) + \phi_p(X))$ dominates $(\delta_1^0(X_1),\dots, \delta_p^0(X_p))$.

Next consider the more general loss

$$L(\theta, a) = \sum_{i=1}^{p} (\theta_i - a_i)^2/\theta_i^{m_i}, \qquad m_i \geqslant 0,\ i = 1,\dots, p. \tag{2.15}$$

Losses of this form with $m_1 = \cdots = m_p = 1$ were considered by Clevenson and Zidek [3] and Ghosh and Parsian [9] in the Poisson case, and by Tsui [17] when $m_1 = \cdots = m_p$. Hwang [13] considered the most general form of loss given in (2.15).

Suppose now that

$$E_\theta[\phi_i^2(X)] < \infty \qquad \text{for all } \theta \in [0, \infty)^p,\ i = 1,\dots, p, \tag{2.16}$$

$$\phi_i(x) = 0 \qquad \text{if } x_i < m_i,\ i = 1,\dots, p. \tag{2.17}$$

Under (2.16) and (2.17), one can write the risk difference (cf. Hwang [13])

$$R(\theta, \delta^0 + \phi) - R(\theta, \delta^0) = 2E_\theta u_0(X), \tag{2.18}$$

where

$$u_0(x) = \sum_{i=1}^{p} \nu_i(x_i)\,\Delta_i \psi_i(x) + \sum_{i=1}^{p} \omega_i(x_i)\,\psi_i^2(x), \tag{2.19}$$

$\nu_i(x_i) = t_i(x_i + m_i - 1)/t_i(x_i)$, $\omega_i(x) = \frac{1}{2}t_i(x_i + m_i)/t_i(x_i)$, $\psi_i(x) = \phi_i(x + m_i e_i)$, $i = 1,..., p$. Once again, our goal is to find solutions to the difference inequality $u_0(x) < 0$.

It is assumed that

$$\nu_i(x_i) = 0 \qquad \text{for } x_i < \alpha_i, \text{ and } \nu_i(\alpha_i) > 0. \tag{2.20}$$

Define now

$$h_i(x_i) = \sum_{k=\alpha_i}^{x_i} \nu_i^{-1}(k) \text{ if } x_i \geqslant \alpha_i \qquad \text{and} \qquad h_i(x_i) = 0 \text{ if } x_i < \alpha_i, \tag{2.21}$$

$i = 1,..., p$. Let

$$d_i(x_i) = \begin{cases} h_i(x_i) & \text{if } \quad m_i \geqslant 1, \\ h_i(x_i)\, h_i(x_i + 1) & \text{if } \quad m_i = 0. \end{cases} \tag{2.22}$$

Denoting by $x_{(1)} \leqslant \cdots \leqslant x_{(p)}$ the ordered x_i's, let

$$D = \sum_{i \ni x_i \leqslant x_{(l)}} d_i(x_i) + \sum_{i \ni x_i > x_{(l)}} d_i(x_{(l)}). \tag{2.23}$$

Also let

$$\beta_i = \begin{cases} 1 & \text{if } \quad m_i \geqslant 1, \\ 2 & \text{if } \quad m_i = 0, \end{cases}$$

$$\beta = \max_{1 \leqslant i \leqslant p} \beta_i, \tag{2.24}$$

$$N(x) = \#\{i\colon \alpha_i \leqslant x_i \leqslant x_{(l)}\}.$$

Define now

$$\psi_i(x) = \begin{cases} -\dfrac{c(N(x) - \beta)^+}{D}\, h_i(x_i) & \text{if } \quad x_i \leqslant x_{(l)}, \\[2ex] -\dfrac{c(N(x) - \beta)^+}{D}\, h_i(x_{(l)}) & \text{if } \quad x_i > x_{(l)}. \end{cases} \tag{2.25}$$

We are now in a position to state the final main result of this section.

THEOREM 2.3. *Assume that there exists a constant* L (>0) *independent of the* x_i's *such that*

$$\sum_{i \ni x_i \leqslant x_{(l)}} \omega_i(x_i)\, h_i^2(x_i) + \sum_{i \ni x_i > x_{(l)}} \omega_i(x_i)\, h_i^2(x_{(l)}) \leqslant LD. \tag{2.26}$$

Suppose also that $v_i(j) \uparrow$ *in* j *if* $m_i = 0$. *Then, assuming* (2.20), $\psi(x) = (\psi_1(x), \ldots, \psi_p(x))$ *provides a solution to* $u_0(x) < 0$ *with* $0 < c < L^{-1}$ *for* $l > \beta$. *The optimum choice of* c *is* $c = (2L)^{-1}$.

Proof. See Appendix.

As a consequence of the above theorem, it follows that $(\delta_1^0(X_1) + \phi_1(X), \ldots, \delta_p^0(X_p) + \phi_p(X))$ dominates $(\delta_1^0(X_1), \ldots, \delta_p^0(X_p))$, where $\phi_i(x) = \psi_i(x - m_i e_i)$, $i = 1, \ldots, p$. We now see an application of Theorem 2.3 in the Poisson example.

EXAMPLE 2.1. Let $X_1, \ldots, X_p$ be p independent random variables with $X_i \sim \text{Poisson}(\theta_i)$, $i = 1, \ldots, p$. Consider the loss (2.15). In this case $v_i(x_i) = x_i!/(x_i + m_i - 1)!$ and $\omega_i(x_i) = \frac{1}{2} x_i!/(x_i + m_i)!$.

Case I ($m_i = 0$). In this case $v_i(x_i) = x_i$ and $\omega_i(x_i) = \frac{1}{2}$. Thus, $v_i(j) \uparrow$ in j. Also, $h_i(0) = 0$ and $h_i(x_i) = \sum_{k=1}^{x_i} k^{-1}$ for $x_i \geqslant 1$. Hence, (2.20) holds with $\alpha_i = 1$. Thus,

$$\omega_i(x_i)\, h_i^2(x_i) \leqslant \tfrac{1}{2} h_i(x_i)\, h_i(x_i + 1) = \tfrac{1}{2} d_i(x_i). \tag{2.27}$$

Also, for those i with $x_i > x_{(l)}$,

$$\omega_i(x_i)\, h_i^2(x_{(l)}) < \tfrac{1}{2} h_i(x_{(l)})\, h_i(x_{(l)} + 1) = \tfrac{1}{2} d_i(x_{(l)}). \tag{2.28}$$

Case II ($m_i = 1$). In this case $v_i(x_i) = 1$ and $\omega_i(x_i) = \frac{1}{2}(x_i + 1)^{-1}$. Thus, $h_i(x_i) = x_i + 1$, so that

$$\omega_i(x_i)\, h_i^2(x_i) = \tfrac{1}{2}(x_i + 1) = \tfrac{1}{2} h_i(x_i) = \tfrac{1}{2} d_i(x_i). \tag{2.29}$$

Also, for those i with $x_i > x_{(l)}$,

$$\omega_i(x_i)\, h_i^2(x_{(l)}) = \tfrac{1}{2}(x_i + 1)^{-1}(x_{(l)} + 1)^2 \leqslant \tfrac{1}{2}(x_{(l)} + 1) = \tfrac{1}{2} d_i(x_{(l)}). \tag{2.30}$$

Case III ($m_i \geqslant 2$). In this case, $v_i(x_i) = (x_i + 1)^{-1} \cdots (x_i + m_i - 1)^{-1}$ and $\omega_i(x_i) = \frac{1}{2}(x_i + 1)^{-1} \cdots (x_i + m_i)^{-1}$. Now,

$$\begin{aligned} h_i(x_i) &= \sum_{j=0}^{x_i} (j+1) \cdots (j + m_i - 1) \\ &\leqslant (x_i + 1) \cdots (x_i + m_i - 1)(x_i + 1). \end{aligned} \tag{2.31}$$

Accordingly,

$$\omega_i(x_i)\, h_i^2(x_i) \leqslant \tfrac{1}{2}\{(x_i+1)/(x_i+m_i)\}\, h_i(x_i) \leqslant \tfrac{1}{2}h_i(x_i) = \tfrac{1}{2}d_i(x_i). \quad (2.32)$$

Also, for those i with $x_i > x_{(l)}$,

$$\begin{aligned}\omega_i(x_i)\, h_i^2(x_{(l)}) &\leqslant \tfrac{1}{2}(x_i+1)^{-1} \cdots (x_i+m_i)^{-1}(x_{(l)}+1) \\ &\quad \cdots (x_{(l)}+m_i-1)(x_{(l)}+1)\, h_i(x_{(l)}) \\ &\leqslant \tfrac{1}{2}h_i(x_{(l)}) = \tfrac{1}{2}d_i(x_{(l)}). \end{aligned} \quad (2.33)$$

Hence, combining (2.27)–(2.30) and (2.32)–(2.33), one finds that (2.26) holds with $L=\frac{1}{2}$. Also (2.20) holds with $\alpha_i = 1$ if $m_i = 0$ and $\alpha_i = 0$ if $m_i \geqslant 1$. Then, defining $\psi_i(x)$ as in (2.25) with $0 < c < 2$ and $h_i(x_i) = \sum_{j=0}^{x_i} (j+1) \cdots (j+m_i-1)$ if $m_i \geqslant 2$, $h_i(x_i) = x_i + 1$ if $m_i = 1$, and $h_i(x_i) = \sum_{j=1}^{x_i} j^{-1}$ for $x_i \geqslant 1$, $h_i(0) = 0$ if $m_i = 0$, one gets a solution to $u(x) < 0$. The optimal choice of c is $c = 1$. Now defining $\phi_i(x) = \psi_i(x - m_i e_i)$, $(\delta_1^0(X_1) + \phi_1(X),\dots, \delta_p^0(X_p) + \phi_p(X))$ dominates $(\delta_1^0(X_1),\dots, \delta_p^0(X_p))$ under the general loss (2.15).

In order to handle the negative binomial case under the general loss (2.15), we have to modify the definitions of the $d_i(x_i)$'s in (2.22). Also, it should be noted that in the Poisson case if $m_i \geqslant 1$ for each i, $\beta_i = 1$ for each i, and so $\beta = 1$. Then, one has the dominance over $\delta^0(X)$ for $l \geqslant 2$. In the negative binomial case, however, this dominance will hold only for $l \geqslant 3$. The situation is similar in the untrimmed situation as evidenced in Hwang [13].

It should also be noted that in the Poisson and negative binomial examples, we could have a more general class of estimators dominating $\delta^0(X)$. We have not aimed at a comprehensive list of solutions to $u(x) < 0$ in the theorems of this section, but have provided some simple estimators which seem to be potentially useful in practice.

3. Absolutely Continuous Case

Let $X_1,\dots, X_p$ be p independent random variables, X_i having pdf

$$f_{\theta_i}(x_i) = \pi_i(\theta_i)\, \rho_i(x_i) \exp(-\theta_i r_i(x_i)) \quad (3.1)$$

with respect to Lebesgue measure on (a_i, b_i), a_i and b_i being possibly infinite. It is assumed that $r_i(x_i)$ is absolutely continuous, and is strictly monotone. We want to estimate $\theta = (\theta_1,\dots, \theta_p)$ based on X. Assume the loss $L(\theta, a) = \sum_{i=1}^p (\theta_i - a_i)^2$.

For each $i = 1,\dots, p$, define $\partial s_i(x)/\partial x_i = \delta_i^0(x)\, r_i'(x_i)$, where $\delta^0(X) =$

$(\delta_1^0(X),..., \delta_p^0(X))$ is the estimator of θ to be improved upon. Let $q_i(x) = r_i'(x_i) \exp(s_i(x))/p_i(x_i)$. Consider the competing estimator $\delta(X) = (\delta_1(X),..., \delta_p(X))$ of θ, where $\delta_i(x) = \delta_i^0(x) - q_i(x)\,\phi_i(x)$, $i = 1,..., p$. We consider the special case where $q_i(x)$ is of the form $\eta(x)\,h_i(x_i)$. Then, assuming conditions (CI)–(CV) of Ghosh *et al.* [11], one gets

$$R(\theta, \delta) - R(\theta, \delta^0) = 2E_\theta \Delta(X), \tag{3.2}$$

where

$$\Delta(x) = \eta(x) \sum_{i=1}^{p} (h_i(x_i)/r_i'(x_i))\,\phi_i^{i(1)}(x) + \tfrac{1}{2}\eta^2(x) \sum_{i=1}^{p} h_i^2(x_i)\,\phi_i^2(x), \tag{3.3}$$

and $\phi_i^{i(1)}(x) = (\partial/\partial x_i)\,\phi_i(x)$, $i = 1,..., p$.

General solutions to the differential inequality $\Delta(x) < 0$ have been obtained earlier by Berger [2], Ghosh and Parsian [8], and Ghosh *et al.* [11]. In this section, we obtain yet another set of solutions to these inequalities which lead to trimmed shrinkage estimates.

With this end, first write $g_i'(x_i) = r_i'(x_i)/h_i(x_i)$. Let $b_i = |\,g_i(x_i) - \mu_i|^\beta$ for some $\beta > 0$, where the μ_i's are certain specified constants. Let $b_{(1)} \leqslant \cdots \leqslant b_{(p)}$ denote the ordered b_i's. Define

$$S = \sum_{i \ni b_i \leqslant b_{(l)}} b_i + (p - l)\, b_{(l)}. \tag{3.4}$$

Assume that $\eta(x) > 0$ for almost all x, $l > \beta$, and

$$\left\{ \sum_{i \ni b_i < b_{(l)}} h_i^2(x_i)(g_i(x_i) - \mu_i)^2 + b_{(l)}^{2/\beta} \sum_{i \ni b_i > b_{(l)}} h_i^2(x_i) \right\} \eta(x) \leqslant 2CS, \tag{3.5}$$

for some $C > 0$. Consider any $\tau(S)$ satisfying

$$\begin{aligned} &\text{(i)}\quad 0 < \tau(S) < C^{-1}(l - \beta),\\ &\text{(ii)}\quad U(S) = \tau(S)\, S^{(l-\beta)/\beta}/[C^{-1}(l-\beta) - \tau(S)] \uparrow \text{ (strictly) in } S. \end{aligned} \tag{3.6}$$

The existence of a $\tau(S)$ satisfying (i) and (ii) is trivially guaranteed. Now define

$$\phi_i(x) = \begin{cases} -\dfrac{\tau(S)}{S}(g_i(x_i) - \mu_i) & \text{if } b_i \leqslant b_{(l)},\\[2ex] -\dfrac{\tau(S)}{S} b_{(l)}^{1/\beta} \operatorname{sgn}(g_i(x_i) - \mu_i) & \text{if } b_i > b_{(l)}, \end{cases} \tag{3.7}$$

where sgn $u = 1$, 0, or -1 according as u >, =, or <0. The main result of this section is as follows.

THEOREM 3.1. *Assume that* (3.5) *holds. Then for any* $\tau(S)$ *satisfying* (3.6), $\phi(x)=(\phi_1(x),..., \phi_p(x))$ *with* $\phi_i(x)$ *defined in* (3.7) *provides a solution to the differential inequality* $\Delta(x)<0$, $\Delta(x)$ *being defined in* (3.3).

Proof. If $b_i > b_{(l)}$, $\phi_i^{i(1)}(x)=0$. If $b_i < b_{(l)}$, then for almost all x,

$$\begin{aligned}\phi_i^{i(1)}(x) &= \left(-\frac{\tau'(S)}{S}+\frac{\tau(S)}{S^2}\right)\frac{\partial S}{\partial x_i}(g_i(x_i)-\mu_i)\, g_i'(x_i)-\frac{\tau(S)}{S}\, g_i'(x_i)\\ &= \left(-\frac{\tau'(S)}{S}+\frac{\tau(S)}{S^2}\right)\beta\,|\, g_i(x_i)-\mu_i\,|^{\beta-1}\operatorname{sgn}(g_i(x_i)-\mu_i)\\ &\quad\times (g_i(x_i)-\mu_i)\, g_i'(x_i)-\frac{\tau(S)}{S}\, g_i'(x_i)\\ &=\beta\left(-\frac{\tau'(S)}{S}+\frac{\tau(S)}{S^2}\right)|\, g_i(x_i)-\mu_i\,|^{\beta}\, g_i'(x_i)-\frac{\tau(S)}{S}\, g_i'(x_i). \qquad (3.8)\end{aligned}$$

If $b_i = b_{(l)}$, then, for almost all x,

$$\begin{aligned}\phi_i^{i(1)}(x) &= \left(-\frac{\tau'(S)}{S}+\frac{\tau(S)}{S^2}\right)\beta(p-l+1)\,|\, g_i(x_i)-\mu_i|^{\beta-1}\\ &\quad\times \operatorname{sgn}(g_i(x_i)-\mu_i)\, g_i'(x_i)(g_i(x_i)-\mu_i)-\frac{\tau(S)}{S}\, g_i'(x_i)\\ &=\beta\left(-\frac{\tau'(S)}{S}+\frac{\tau(S)}{S^2}\right) b_{(l)}\, g_i'(x_i)(p-l+1)\\ &\quad-\frac{\tau(S)}{S}\, g_i'(x_i). \qquad (3.9)\end{aligned}$$

The rest of the proof follows the pattern of Ghosh *et al.* [11], and is omitted.

Next we consider applications of this theorem in the normal and gamma examples. First consider the normal case.

EXAMPLE 3.1. Let $X_1,..., X_p$ be p ($\geqslant 3$) independent random variables, $X_i \sim N(\theta_i, 1)$, $i=1,..., p$. In this case $r_i(x_i)=-x_i$ and $\delta_i^0(x_i)=x_i$. Hence, $s_i(x_i)=-x_i^2/2$, and so $q_i(x)=-1$ which is of the form $\eta(x)\, h_i(x_i)$ with $\eta(x)\equiv 1$, $h_i(x_i)=-1$. Now, $g_i'(x_i)=1$ so that $g_i(x_i)=x_i$. Hence, defining $b_i=(x_i-\mu_i)^2$, it follows that equality holds in (3.5) with $c=\frac{1}{2}$, S being defined in (2.4). Let $z_{(1)}\leqslant \cdots \leqslant z_{(l)} \leqslant \cdots \leqslant z_{(p)}$ denote the ordered $|x_i-\mu_i|$'s. With this notation,

$$S=\sum_{i \ni |x_i-\mu_i| \leqslant z_{(l)}} (x_i-\mu_i)^2+(p-l)\, z_{(l)}^2. \qquad (3.10)$$

Now for any $\tau(S)$ satisfying (3.7) with $\beta = 2$, define

$$\phi_i(x) = \begin{cases} -\dfrac{\tau(S)}{S}(x_i - \mu_i) & \text{if } |x_i - \mu_i| \leqslant z_{(l)}. \\ -\dfrac{\tau(S)}{S} z_{(l)} \operatorname{sgn}(x_i - \mu_i) & \text{if } |x_i - \mu_i| > z_{(l)}. \end{cases} \tag{3.11}$$

Then $(X_1 + \phi_1(X), \ldots, X_p + \phi_p(X))$ dominates X for $l \geqslant 3$.

The special case when $\tau(S) \equiv l - 2$ was considered in Stein [15] and further studied in Dey [4]. The $l = p$ case is considered in Efron and Morris [7] under the conditions (i) and (ii) on $\tau(S)$. Earlier, Baranchik [1] and Strawderman [16] considered the case when (i) holds with $l = p$ and $\tau(S) \uparrow$ (strictly) in S.

EXAMPLE 3.2. Let $X_1, \ldots, X_p$ be p independent gamma variables, X_i having pdf

$$f_{\theta_i}(x_i) = \exp(-\theta_i x_i)\, x_i^{\alpha_i - 1} \theta_i^{\alpha_i} / \Gamma(\alpha_i), \qquad x_i > 0, \theta_i > 0, \tag{3.12}$$

where $\alpha_i(> 2)$'s are known, but θ_i's are unknown. This example appears in Berger [2]. Assuming squared error loss, the best scale invariant estimator of θ is $\delta^0(X)$ with $\delta_i^0(X) \equiv \delta_i^0(X_i) = (\alpha_i - 2)/X_i$, $i = 1, \ldots, p$. Here, $r_i(x_i) = x_i$ so that $\partial s_i(x)/\partial x_i = (\alpha_i - 2)/x_i$. Thus, $s_i(x) = (\alpha_i - 2) \log x_i$. Hence, $q_i(x) = x_i^{\alpha_i - 2}/x_i^{\alpha_i - 1} = x_i^{-1}$ which is of the form $\eta(x)\, h_i(x_i)$ with $\eta(x) \equiv 1$ and $h_i(x_i) = x_i^{-1}$. Now, $g_i'(x_i) = x_i$ and taking $\mu_1 = \cdots = \mu_p = 0$ and $\beta = 1$, $b_i = g_i(x_i) = \frac{1}{2} x_i^2$. Denoting by $z_{(1)} \leqslant \cdots \leqslant z_{(p)}$ the ordered x_i's, $S = \frac{1}{2}[\sum_{i \ni x_i \leqslant z_{(l)}} x_i^2 + (p - l)\, z_{(l)}^2]$. Now from (3.5) it follows that

$$\left\{ \sum_{i \ni b_i \leqslant b_{(l)}} h_i^2(x_i)(g_i(x_i) - \mu_i)^2 + b_{(l)}^{2/\beta} \sum_{i \ni b_i > b_{(l)}} h_i^2(x_i) \right\} \eta(x)$$

$$= \tfrac{1}{4}\left[\sum_{i \ni x_i \leqslant z_{(l)}} x_i^2 + z_{(l)}^{(4)} \sum_{i \ni x_i > z_{(l)}} x_i^{-2} \right]$$

$$\leqslant \tfrac{1}{4}\left[\sum_{i \ni x_i \leqslant z_{(l)}} x_i^2 + (p - l)\, z_{(l)}^2 \right] = \tfrac{1}{2} S. \tag{3.13}$$

Hence, (3.5) holds with $C = \frac{1}{4}$. Hence, for $l \geqslant 2$, and any $\tau(S)$ satisfying (i) and (ii) with $C = \frac{1}{4}$ and $p = 1$, one defines

$$\phi_i(x) = \begin{cases} -\dfrac{\tau(S)}{S}\left(\dfrac{1}{2} x_i^2\right) & \text{if } x_i < z_{(l)}, \\ -\dfrac{\tau(S)}{S}\left(\dfrac{1}{2} z_{(l)}^2\right) & \text{if } x_i > z_{(l)}. \end{cases} \tag{3.14}$$

Then $(\delta_1^0(X_1) - q_1(X)\,\phi_1(X),..., \delta_p^0(X_p) - q_p(X)\,\phi_p(X))$ dominates $(\delta_1^0(X_1),..., \delta_p^0(X_p))$.

One can also obtain trimmed estimates improving on the usual ones when one considers estimation of the mean vector in the Hudson [12] subfamily of the general exponential family. The details are not pursued here.

4. Risk Simulations Study

In this section we will compute the risk of the shrinkage and the corresponding trimmed estimators for the simultaneous estimation of the Poisson means and the gamma scale parameters, using Monte Carlo simulation method.

4.1. *Numerical Studies for Poisson Case*

Suppose we want to estimate simultaneously p-Poisson means $(\theta_1,..., \theta_p)$ under squared norm loss as given in (2.1). Then the usual shrinkage estimator is given componentwise as

$$\hat{\theta}_i = X_i - \frac{(p - N_0 - 2)^+}{S}\, h_i(x_i), \tag{4.1}$$

where

$$N_0 = \#\{i: x_i = 0\},$$
$$h_i(x_i) = \sum_{k=1}^{x_i} k^{-1}, \tag{4.2}$$
$$a^+ = \max(a, 0),$$

and

$$S = \sum_{j=1}^{p} h_j(x_j)\, h_j(x_j + 1).$$

The corresponding trimmed version is given componentwise as

$$\hat{\theta}_i^{(l)} = \begin{cases} x_i - \dfrac{(l - N_0 - 2)^+}{D}\, h_i(x_i) & \text{if} \quad x_i \leqslant x_{(l)}, \\[2ex] x_i - \dfrac{(l - N_0 - 2)^+}{D}\, h_i(x_{(l)}) & \text{if} \quad x_i > x_{(l)}, \end{cases} \tag{4.3}$$

where

$$D = \sum_{x_j \leqslant x_{(l)}} h_j(x_j)\, h_j(x_j+1) + \sum h_j(x_{(l)})\, h_j(x_{(l)}+1) \tag{4.4}$$

and $h_i(x_i)$, N_0, and a^+ are defined in (4.2).

In the risk simulation study, first, 10 independent Poisson random variables are chosen. Second, 10 parameters θ_i are generated randomly within a certain range (a, b). Third, one observation of each of the 10 distributions with the 10 parameters obtained in the second step is generated. Then the estimates $\hat{\theta}$ and $\hat{\theta}^{(l)}$ ($l = 5, 6, 7, 8, 9$) are calculated. The third step is repeated 1000 times and the risks under squared norm loss for $\hat{\theta}$, $\hat{\theta}^{(l)}$, and MLE are calculated. The percentage of savings in using $\hat{\theta}$ and $\hat{\theta}^{(l)}$ as compared to the MLE are finally calculated.

In Table I, the percentage improvements over the MLE are computed when the estimators are trimmed at specified (lth) order statistics. The estimator $\hat{\theta}^{(l)}$ as defined in (4.3) stands for the shrinkage estimator which is trimmed at lth-order statistic ($l = 5, 6, 7, 8, 9$).

Table II is similar to Table I, except that one θ_i is chosen to be very large compared to other θ_i's so that the corresponding observation can be treated as possibly an outlier.

In Table III, an adaptive version of the trimmed estimator is considered. For corresponding results in the normal case see Dey and Berger [5]. The trimming point l as given in (4.3) is chosen (depending on the data) in the following way. Ten independent Poisson random variables with parameters θ_i are generated randomly, where θ_i's are generated uniformly within a certain range. The estimate of the expected improvement of $\hat{\theta}^{(l)}$ over the MLE is computed for $l = 3, \ldots, 10$, and the difference is then numerically maximized over l. These risks were calculated by simulation, repeating the calculations thousands of times. After finding the maximum, the risk of the

TABLE I

Shrinking at Specified Order Statistic

Range of the parameters θ_i	I	$I5$	$I6$	$I7$	$I8$	$I9$
(0, 4)	10.6	2.9	4.1	5.5	7.0	8.7
(4, 8)	1.6	0.2	0.3	0.6	0.9	1.3
(8, 12)	0.7	0.2	0.2	0.3	0.4	0.6
(12, 16)	0.5	0.2	0.2	0.3	0.4	0.5

$$I = \frac{R(\theta, X) - R(\theta, \hat{\theta})}{R(\theta, X)} \times 100, \qquad Il = \frac{R(\theta, X) - R(\theta, \hat{\theta}^{(l)})}{R(\theta, X)} \times 100, \qquad l = 5, \ldots, 9.$$

TABLE II

Shrinking at Specified Order Statistic in Presence of Outliers
$\theta_i \in (0, 4)$, $i = 1,\ldots, 9$, θ_{10} as Specified.

θ_{10}	I	$I5$	$I6$	$I7$	$I8$	$I9$
22.3147	4.5	2.7	3.2	3.7	4.2	4.7
20.2424	3.9	1.1	1.6	2.2	2.9	3.6
20.5531	0.7	0	0	0.3	0.3	0.8
21.1048	0.8	0	0	0	0	0.5

adaptive estimator, say, $\hat{\theta}^{(l^*)}$, is computed. Finally, the percentage improvement of $\hat{\theta}^{(l^*)}$ over MLE is calculated with that of $\hat{\theta}$.

In Tables I, II, and III, we observe, in general, that the percentage improvement decreases as the magnitude of the θ_i's increases. It is also observed that the improvements are always positive, which indicates that $\hat{\theta}^{(l)}$ and $\hat{\theta}^{(l^*)}$ are all minimax.

Tables I and II indicate that the percentage improvements are about the same if we use $\hat{\theta}$ or $\hat{\theta}^{(l)}$. Therefore for large p, truncation at a specified order statistic is quite desirable.

Table III indicates that the percentage improvement for the adaptive trimmed estimator is quite significant as compared to the usual shrinkage estimator. This indicates that the adaptive trimmed estimator is more suggestive.

4.2. *Numerical Studies for the Gamma Case*

Assume the X_i are gamma (α, θ_i) with known α (> 2), $i = 1,\ldots, p$. The problem is to estimate the parametric vector $\theta = (\theta_1,\ldots, \theta_p)$. Assuming squared error loss, the best scale invariant estimator of θ is $\delta^0(X)$ with

TABLE III

Percentage Improvement of $\hat{\theta}$ and $\hat{\theta}^{(l^*)}$

Range of the parameters θ_i	$I1$	$I2$
(0.4)	8.7	7.5
(4, 8)	2.5	2.4
(8, 12)	1.7	1.3
(12, 16)	0.8	0.8

$$I1 = \frac{R(\theta, X) - R(\theta, \hat{\theta})}{R(\theta, X)} \times 100, \qquad I2 = \frac{R(\theta, X) - R(\theta, \theta^{(l^*)})}{R(\theta, X)} \times 100.$$

$\delta_i^0(X) = (\alpha - 2)/X_i$, $i = 1,..., p$. From Berger [2] it follows that the improved estimator of θ, under squared error loss, is given componentwise as

$$\delta_i^J(X) = \frac{\alpha - 2}{X_i} + \frac{2(p-1) X_i^2}{\sum_{j=1}^p X_j^2} \tag{4.5}$$

and it can be easily shown that the risk of (4.5) is

$$R(\delta^J, \theta) = \frac{1}{\alpha - 2} \sum_{i=1}^p \theta_i^2 - E_\theta \left[\frac{4(p-1)^2}{\sum_{j=1}^p X_j^2} \right]. \tag{4.6}$$

Now from (3.14), it follows that the corresponding trimmed version of the estimator (4.5) can be defined componentwise as

$$\delta_i^l(X) = \frac{\alpha - 2}{X_i} + \frac{2(l-1) \min(X_i, Z_{(l)})}{\sum_{X_i \leqslant Z_{(l)}} X_i^2 + (p-l) Z_{(l)}^2}, \tag{4.7}$$

and the risk of (4.7) is given as

$$R(\delta^l, \theta) = \frac{1}{\alpha - 2} \sum_{i=1}^p \theta_i^2 - E_\theta \left[\frac{4(l-1)^2}{\sum_{X_i < Z_{(l)}} X_i^2 + (p-l) Z_{(l)}^2} \right]. \tag{4.8}$$

A very appealing possibility is to let the data select the trimming point l in the estimator (4.7). The obvious method of selection is to choose that $l \geqslant 3$ (call it l^*) which maximizes the unbiased estimate of the risk improvement (see (4.8))

$$(l-1)^2 \Big/ \left\{ \sum_{X_i \leqslant Z_{(l)}} X_i^2 + (p-l) Z_{(l)}^2 \right\}. \tag{4.9}$$

Theoretical analysis of this estimator is immensely difficult, due to the complicated dependence of l^* on X. Therefore we have selected the trimming point l adaptively and computed under squared error loss the risk of the adaptive trimmed estimator numerically. In Table IV, the percentage improvement of δ^{l^*} is calculated over δ^0 and δ^J for different α and p values for different ranges of the parameter values.

The parameters θ_i are chosen to be uniformly distributed within certain ranges. The ranges are so chosen that it is possible to check the performance of the estimators both when the parameters fall in a narrow range and when they fall in a wide range. The choice of uniform prior is an artifact, just to demonstrate the performance of the trimmed estimator. In fact, when the θ_i's are thought to have come from a possibly heavy-tailed prior distribution, it is expected (as in Dey and Berger [5]) that the trimmed estimator will perform much better. In all the cases, the percentage improvement of δ^{l^*} over δ^0 is

TABLE IV

Percentage Improvement in Risk under Squared error Loss for δ^{l^*} over δ^0 and δ^J

Range of the parameters θ_i	$\alpha=3$				$\alpha=20$			
	$p=10$		$p=20$		$p=10$		$p=20$	
	a	b	a	b	a	b	a	b
(0, 4)	35.09	24.44	36.63	30.96	5.89	3.13	6.32	4.82
(4, 8)	56.37	3.30	59.34	5.62	13.42	0.49	14.37	0.13
(8, 12)	59,95	0.91	63.35	3.14	14.74	0.64	15.81	0.17
(12, 16)	60.98	0.03	64.53	2.42	15.11	0.68	16.23	0.19
(0, 20)	55.09	4.52	57.18	12.91	11.87	0.98	12.35	0.15
(10, 30)	50.63	7.44	52.82	10.21	10.97	0.11	11.74	0

$$a=\frac{R(\theta,\delta^0)-R(\theta,\delta^{l^*})}{R(\theta,\delta^0)}\times 100, \qquad b=\frac{R(\theta,\delta^J)-R(\theta,\delta^{l^*})}{R(\theta,\delta^{JB})}\times 100$$

seen to be an increasing function of p, the number of independent gamma variates. For small α ($\alpha=3$), we observe that the precentage improvement of δ^{l^*} over δ^J increases as p increases. However, for large α ($\alpha=20$), the percentage improvement of δ^{l^*} over δ^J usually decreases as p increases. Table IV provides strong evidence that adaptive estimator δ^{l^*} performs much better than the "standard" estimator and does perform significantly better than the untrimmed estimator.

APPENDIX

Proof of Theorem 2.3. Using (2.26), it follows from (2.22) through (2.25) that

$$\sum_{i=1}^{p} \omega_i(x_i)\,\psi_i^2(x) \leqslant \frac{c^2((N(x)-\beta)^+)^2 L}{D}. \tag{A.1}$$

Next observe that $\Delta_i\psi_i(x)=0$ for $x_i \leqslant \alpha_i$ or $x_i > x_{(l)}$. For $\alpha_i \leqslant x_i \leqslant x_{(l)}$,

$$\Delta_i\psi_i(x) \leqslant c(N(x)-\beta)^+ \frac{h_i(x_i-1)\,\Delta_i d_i(x_i)-\Delta_i h_i(x_i)\,D_i}{DD_i}. \tag{A.2}$$

Next note that $\Delta_i d_i(x_i)=\Delta_i h_i(x_i)=\nu_i^{-1}(x_i)$ if $m_i \geqslant 1$ and $x_i \geqslant \alpha_i$. For $m_i=0$ and $x_i \geqslant \alpha_i$,

$$\Delta_i d_i(x_i) \leqslant 2h_i(x_i)\,\nu_i^{-1}(x_i). \tag{A.3}$$

Hence,

$$\sum_{i=1}^{p} v_i(x_i)\,\Delta_i\psi_i(x)$$

$$\leqslant c(N(x)-\beta)^+ \sum_{i\ni\alpha_i\leqslant x_i\leqslant x_{(l)}} \left[\frac{v_i(x_i)\,h_i(x_i-1)\,\Delta_i d_i(x_i)}{DD_i} - \frac{1}{D}\right]$$

$$\leqslant c(N(x)-\beta)^+ \left[\beta \sum_{i\ni\alpha_i\leqslant x_i\leqslant x_{(l)}} d_i(x_i-1)/(DD_i) - N(x)/D\right]$$

$$\leqslant -c((N(x)-\beta)^+)^2/D. \tag{A.4}$$

Combining (2.19), (A.1), and (A.4), it follows that

$$u_0(x) \leqslant -\frac{c((N(x)-\beta)^+)^2}{D}(1-Lc) \leqslant 0. \tag{A.5}$$

For any fixed L and $0 < c < L^{-1}$, since $c(1-Lc)$ is maximized at $c=(2L)^{-1}$, it follows from (A.5) that the optimal choice of c is $c=(2L)^{-1}$.

References

[1] Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645.

[2] Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* **8** 545–571.

[3] Clevenson, M., and Zidek, J. (1975). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.

[4] Dey, D. (1980). On the choice of coordinates in simultaneous estimation of normal means. Mimeograph Series No. 80-23, Department of Statistics, Purdue University.

[5] Dey, D., and Berger, J. (1983). On truncation of shrinkage estimators in simultaneous estimation of normal means. *J. Amer. Statist. Assoc.* **78** 865–869.

[6] Efron, B., and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.

[7] Efron, B, and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11–21.

[8] Ghosh, M., and Parsian, A. (1980). Admissible and minimax multiparameter estimation in exponential families. *J. Multivariate Anal.* **10** 551–564.

[9] Ghosh, M., and Parsian, A. (1981). Bayes minimax estimation of multiple Poisson parameters. *J. Multivariate Anal.* **11** 280–288.

[10] Ghosh, M., Hwang, J. T., and Tsui, K. W. (1983). Construction of improved estimators in multiparameter estimation for discrete exponential families. *Ann. Statist.* **11** 351–376.

[11] Ghosh, M., Hwang J. T., and Tsui, K. W. (1984). Construction of improved estimators in multiparameter estimation for continuous exponential families. *J. Multivariate Anal.* **14** 212–220.

[12] HUDSON, M. (1978). A natural identity for exponential families with application in multiparameter estimation. *Ann. Statist.* **6** 473–484.

[13] 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.

[14] PENG, J. C. M. (1975). *Simultaneous Estimation of the Parameters of Independent Poisson Distributions*. Technical Report No. 78, Department of Statistics, Stanford University.

[15] STEIN, C. (1981). Estimation of the parameters of a multivariate normal distribution. I. Estimation of the means. *Ann. Statist.* **9** 1135–1151.

[16] STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.

[17] TSUI, K. W. (1979). Multiparameter estimation of discrete exponential distributions. *Canad. J. Statist.* **7** 193–200.

[18] TSUI, K. W. (1981). Simultaneous estimation of several Poisson parameters under squared error loss. *Ann. Inst. Statist. Math.* **33** 215–223.

[19] 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.