The Annals of Statistics
1988, Vol. 16, No. 1, 278–291

SIMULTANEOUS ESTIMATION OF POISSON MEANS UNDER ENTROPY LOSS

BY MALAY GHOSH[1] AND MING-CHUNG YANG

University of Florida and University of Iowa

This paper characterizes admissible linear estimators of multiple Poisson parameters under entropy loss. Estimators dominating some of the standard estimators are given. Further, hierarchical Bayes estimators are generated and conditions under which they dominate the standard estimators are also given. Monte Carlo simulations are undertaken to indicate the extent of the risk dominance.

1. Introduction. Let $X_1, \dots, X_p$ be p independent Poisson variables with respective means $\theta_1, \dots, \theta_p$. We write $\mathbf{X} = (X_1, \dots, X_p)$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$. It is assumed that $\boldsymbol{\theta} \in (0, \infty)^p$. Consider the problem of estimating $\boldsymbol{\theta}$ when the loss is of the form

$$L_0(\boldsymbol{\theta}, \mathbf{a}) = \sum_{i=1}^{p} \theta_i^{-m_i}(\theta_i - a_i)^2, \tag{1.1}$$

where m_i's are known constants. For $p = 1$, admissibility of the usual (MLE, UMVUE, etc.) estimator $\mathbf{X}$ of $\boldsymbol{\theta}$ follows from more general results of Karlin (1958) or Brown and Hwang (1982) for every m_1 [other proofs are available in Girshick and Savage (1951) or Hodges and Lehmann (1951)]. Estimation of $\boldsymbol{\theta}$ in higher dimensions has received considerable attention in recent years beginning with the pioneering work of Clevenson and Zidek (1975). Clevenson and Zidek showed that when $m_1 = \cdots = m_p = 1$, $\mathbf{X}$ was an inadmissible estimator of $\boldsymbol{\theta}$ for $p \geq 2$. Peng (1975) considered the case $m_1 = \cdots = m_p = 0$ and proved the inadmissibility of $\mathbf{X}$ for $p \geq 3$ and its admissibility for $p = 2$. In more recent publications, estimation of $\boldsymbol{\theta}$ is considered for general $m_1, \dots, m_p$ [see Ghosh, Hwang and Tsui (1983) for a unified treatment].

The present paper considers instead the loss as the entropy distance (or the Kullback–Leibler information number) between two distributions of p independent Poisson variables. Accordingly, the loss is defined as

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{a}) &= E_{\boldsymbol{\theta}}\left[\log\left\{\left(\prod_{i=1}^{p} \exp(-\theta_i)\theta_i^{X_i}/X_i!\right)\Big/\left(\prod_{i=1}^{p} \exp(-a_i)a_i^{X_i}/X_i!\right)\right\}\right] \\ &= \sum_{i=1}^{p} (a_i - \theta_i - \theta_i \log(a_i/\theta_i)) = \sum_{i=1}^{p} \theta_i(a_i/\theta_i - \log(a_i/\theta_i) - 1). \end{aligned} \tag{1.2}$$

Received January 1986; revised May 1987.

[1]Research partially supported by NSF Grants DMS-82-18091 and DMS-86-00666. Also partially supported by the Research Development Award of the Division of Sponsored Research, University of Florida.

AMS 1980 *subject classifications*. 62C15, 62F10, 62C10.

Key words and phrases. Poisson means, simultaneous estimation, entropy loss, linear estimators, admissibility, hierarchical Bayes, empirical Bayes.

Note that the entropy loss can differ from the conventional loss measured in terms of the distance between the parameter and its estimate.

The entropy loss was first introduced by James and Stein (1961) for estimation of the multinormal variance–covariance matrix. Later, the same loss was considered in Brown (1968), Haff (1977, 1979, 1980, 1982) and Dey and Srinivasan (1985) for estimating either the multinormal variance–covariance matrix or its inverse. Dey, Ghosh and Srinivasan (1987) considered the entropy loss for simultaneous estimation of p independent gamma scale parameters or their reciprocals, while Ighodaro and Santner (1982) and Ighodaro, Santner and Brown (1982) considered entropy loss for simultaneous estimation of independent binomial and multinomial proportions.

Note that when $(a_i - \theta_i)\theta_i^{-1}$ is near zero, retaining only the first two terms in the Taylor expansion,

$$\log(a_i/\theta_i) = \log\left(1 + (a_i - \theta_i)\theta_i^{-1}\right) \doteq (a_i - \theta_i)\theta_i^{-1} - \tfrac{1}{2}(a_i - \theta_i)^2\theta_i^{-2}.$$

Hence, from (1.2) it follows that if $(a_i - \theta_i)\theta_i^{-1}$ is close to zero for all $i = 1, \dots, p$, then $L(\boldsymbol{\theta}, \mathbf{a}) = \frac{1}{2}\sum_{i=1}^{p}(\theta_i - a_i)^2/\theta_i$, which is a constant multiple of the loss given in (1.1) with $m_1 = \cdots = m_p = 1$. However, the loss (1.1) (with $m_1 = \cdots = m_p = 1$) and (1.2) may differ considerably when the preceding does not hold.

Note that the usual MLE estimator $\mathbf{X}$ of $\boldsymbol{\theta}$ is not very appropriate in this case since under the loss (1.2), $\mathbf{X}$ has infinite risk for all $\boldsymbol{\theta}$. One way to generate meaningful estimators is to use certain priors (proper or improper) and obtain the resulting generalized Bayes estimators of $\boldsymbol{\theta}$ under the loss (1.2). For any such prior, if the resulting posterior pdf $f(\boldsymbol{\theta}|\mathbf{x})$ is proper, the generalized Bayes estimator of $\boldsymbol{\theta}$ under (1.2) is obtained by minimizing $\sum_{i=1}^{p}\int(a_i - \theta_i \log a_i)f(\boldsymbol{\theta}|\mathbf{x})\,d\theta_1 \cdots d\theta_p$ with respect to $\mathbf{a}$ which gives the solution $a_i \equiv a_i(\mathbf{x}) = E(\theta_i|\mathbf{x})$, $i = 1, \dots, p$. Thus, the generalized Bayes estimator of $\boldsymbol{\theta}$ under (1.2) is the same as its generalized Bayes estimator under squared error loss and is given by $E(\boldsymbol{\theta}|\mathbf{X})$. Thus, use of independent gamma (α_i, k_i) priors

$$(1.3) \qquad g_{\boldsymbol{\alpha},\mathbf{k}}(\boldsymbol{\theta}) = \prod_{i=1}^{p}\left\{\exp(-\alpha_i\theta_i)\theta_i^{k_i-1}\alpha_i^{k_i}/\Gamma(k_i)\right\},$$

$\alpha_i > 0$, $k_i > 0$ for all $i = 1, \dots, p$, leads to the proper Bayes estimators

$$(1.4) \qquad \mathbf{e}_B(\mathbf{X}) = \left((1 + \alpha_1)^{-1}(X_1 + k_1), \dots, (1 + \alpha_p)^{-1}(X_p + k_p)\right)$$

of $\boldsymbol{\theta}$, and these estimators have finite Bayes risks (expected loss over the joint distribution of the parameters and the samples). The class of priors described in (1.3) can be extended to include improper priors as well, where some or all of the α_i's and k_i's are allowed to be zeros. The resulting generalized Bayes estimators are obtained as appropriate pointwise limits of the estimators derived in (1.4).

The preceding estimators are members of a more general class of estimators of the form $\mathbf{XC} + \mathbf{b}$, where $\mathbf{C}$ is a diagonal matrix of constants and $\mathbf{b}$ is a vector of constants. In Section 2 of this paper, we characterize the admissible subclass within this general class of estimators. A fortiori, this will characterize the admissibility of generalized Bayes estimators under independent gamma priors or their limits. Cohen (1966) characterized admissible linear estimators of the multivariate normal mean, while Brown and Farrell (1985) characterized admissible linear estimators of Poisson means under the loss (1.1) with $m_1 = \cdots = m_p = 1$ or 0.

Following Ghosh (1983), in Section 3 of this paper, we obtain certain generalized Bayes estimators of $\boldsymbol{\theta}$ using hierarchical priors. Such estimators have also an interesting empirical Bayes interpretation.

In Section 4, using Monte Carlo simulations, we compare the risk performances of $\mathbf{X} + \mathbf{1}$ and the estimators dominating them. In the preceding, $\mathbf{1}$ is a p-component row vector with all elements equal to 1.

In Section 5, we have made some concluding remarks about how the results of Section 2 can be generalized when $\mathbf{C}$ is not necessarily diagonal. The proof of the technical Lemma 4 is deferred to the Appendix.

2. The admissibility results. Let the parameter space be $\Theta = (0, \infty)^p$. Write $\mathbf{C} = \operatorname{diag}(c_1, \dots, c_p)$ and $\mathbf{b} = (b_1, \dots, b_p)$, where each $b_i \geq 0$. Note that since $P(X_i = 0) > 0$, if one allows some $b_i < 0$, the corresponding $c_i X_i + b_i$ can take negative values with positive probability for which the loss given in (1.2) will be undefined. The main theorem of this section characterizes the admissibility of the estimators $\mathbf{XC} + \mathbf{b}$ under the loss (1.2).

THEOREM 1. *Under the loss* (1.2), $\mathbf{XC} + \mathbf{b}$ *is an admissible estimator of* $\boldsymbol{\theta}$ *if and only if* (i) $b_i > 0$ *and* $0 \leq c_i \leq 1$ *for all* $i \in S = \{1, \dots, p\}$ *and* (ii) $0 < \sum_{\{i \in S:\, c_i = 1\}} b_i \leq 1$ *holds.*

The preceding characterization theorem is a consequence of Lemmas 1–5. For all these lemmas, the loss is the one given in (1.2).

LEMMA 1. $\mathbf{XC} + \mathbf{b}$ *is an inadmissible estimator of* $\boldsymbol{\theta}$ *if there exists at least one* $c_i > 1$ *with the corresponding* $b_i > 0$.

PROOF. Suppose $c_l > 1$ and $b_l > 0$. Then, writing once again $\boldsymbol{\delta}^0(\mathbf{X}) = \mathbf{XC} + \mathbf{b}$, but $\boldsymbol{\delta}^*(\mathbf{X})$ as an estimator with all but its lth element equal to the corresponding element of $\boldsymbol{\delta}^0(\mathbf{X})$ and the lth element of $\boldsymbol{\delta}^*(\mathbf{X})$ equal to $X_l + b_l c_l^{-1}$, it follows that

$$
\begin{aligned}
R(\boldsymbol{\theta}, \boldsymbol{\delta}^0) - R(\boldsymbol{\theta}, \boldsymbol{\delta}^*) &= E_{\theta_l}\big[L(\theta_l, c_l X_l + b_l) - L\big(\theta_l, X_l + b_l c_l^{-1}\big)\big] \\
&= \big(c_l\theta_l + b_l\big) - \big(\theta_l + b_l c_l^{-1}\big) \\
&\quad - \theta_l E_{\theta_l}\log\big[\big(c_l X_l + b_l\big)/\big(X_l + b_l c_l^{-1}\big)\big] \qquad (2.1)\\
&= \big(c_l - 1 - \log c_l\big)\theta_l + b_l\big(1 - c_l^{-1}\big) \\
&> \big(c_l - 1 - \log c_l\big)\theta_l \geq 0,
\end{aligned}
$$

so that $\boldsymbol{\delta}^*$ dominates $\boldsymbol{\delta}^0$. □

LEMMA 2. $\mathbf{XC} + \mathbf{b}$ *is an admissible estimator of* $\boldsymbol{\theta}$ *if* $0 \leq c_l < 1$ *and* $b_l > 0$ *for every* $l \in S$.

PROOF. Consider the joint prior distribution which is the product of *independent* gamma $((1-c_l)/c_l, b_l/c_l)$ priors if $c_l > 0$ and priors degenerate at b_l if $c_l = 0$. $\mathbf{XC} + \mathbf{b}$ is a proper Bayes estimator of $\boldsymbol{\theta}$ under this prior with finite Bayes risk, and hence, it is admissible. □

LEMMA 3. *$\mathbf{XC} + \mathbf{b}$ is an inadmissible estimator of $\boldsymbol{\theta}$ if $b_l = 0$ for some $l \in S$.*

PROOF. This is an immediate consequence of the fact that if $b_l = 0$ for some $l \in S$, $R(\boldsymbol{\theta}, \mathbf{XC}) \geq R(\theta_l, X_l c_l) = +\infty$, while there exist estimators, for example, positive constants with finite risk for all $\boldsymbol{\theta} \in (0, \infty)^p$. □

LEMMA 4. *For every $p \geq 1$, $\boldsymbol{\delta}^0(\mathbf{X}) = (X_1 + b_1, \dots, X_p + b_p)$ is an admissible estimator of $\boldsymbol{\theta}$ under the loss* (1.2) *if $b_i > 0$ for all $i \in S$ and $\sum_{i \in S} b_i \leq 1$.*

The proof of Lemma 4 is technical and is deferred to the Appendix. Like many other proofs of admissibility, our proof employs Blyth's (1951) technique, but instead of using a sequence of conjugate gamma priors to approximate the improper prior with respect to which $\boldsymbol{\delta}^0(\mathbf{X})$ is generalized Bayes, our proof employs a different sequence of priors closely akin to Brown and Hwang (1982). However, in spite of the similarity of the technique of proof with the one in Brown and Hwang, the calculations are different because the present entropy loss differs from the squared error loss of Brown and Hwang.

The final lemma of this section proves the inadmissibility of $\boldsymbol{\delta}^0(\mathbf{X})$ defined in Lemma 4 when $b_i > 0$ for all $i \in S$ and $\sum_{i \in S} b_i > 1$.

LEMMA 5. *Suppose $\boldsymbol{\delta}^*(\mathbf{X}) = \boldsymbol{\delta}^0(\mathbf{X}) + \boldsymbol{\phi}(\mathbf{X})$ when $\boldsymbol{\phi}(\mathbf{X}) = (\phi_1(\mathbf{X}) \cdots \phi_p(\mathbf{X}))$ with*

$$\phi_i(\mathbf{X}) = -\frac{c(\mathbf{X})(X_i + b_i)}{\sum_{j=1}^p (X_j + d)}, \qquad 1 \leq i \leq p,$$

where $d > 0$ and $c(\mathbf{X})$ is nondecreasing in each coordinate. We assume that

(i) $\sum_{j=1}^p b_j = B > 1$;
(ii) $0 < c(\mathbf{X}) < [2(B-1)pd]/[pd + 1 + 2(B-1)] = G$ *(say).*

Then $\boldsymbol{\delta}^*(\mathbf{X})$ dominates $\boldsymbol{\delta}^0(\mathbf{X})$ under the loss (1.2).

PROOF. First note that using (1.1) of Hwang (1982),

$$\begin{aligned} R(\boldsymbol{\theta}, \boldsymbol{\delta}^*) - R(\boldsymbol{\theta}, \boldsymbol{\delta}^0) &= \sum_{i=1}^p E_{\boldsymbol{\theta}}\left[\phi_i(\mathbf{X}) - \theta_i \log\left(1 + \frac{\phi_i(\mathbf{X})}{X_i + b_i}\right)\right] \\ &= \sum_{i=1}^p E_{\boldsymbol{\theta}}\left[\phi_i(\mathbf{X}) - X_i \log\left(1 + \frac{\phi_i(\mathbf{X} - \mathbf{e}_i)}{X_i + b_i - 1}\right) I_{[X_i \geq 1]}\right] \qquad (2.2) \\ &\leq E_{\boldsymbol{\theta}}[U(\mathbf{X})], \end{aligned}$$

where

$$(2.3)\qquad U(\mathbf{X}) = \sum_{i=1}^{p}\left[\phi_i(\mathbf{X}) - X_i\log\left(1 + \frac{\phi_i(\mathbf{X}-\mathbf{e}_i)}{X_i + b_i - 1}\right)I_{[X_i\geq 1]}\right],$$

$\mathbf{e}_i$ being the ith unit vector. Next, observe that writing $T = \sum_{j=1}^{p} X_j$ and using assumption (ii),

$$(2.4)\qquad \left|\frac{\phi_i(\mathbf{X}-\mathbf{e}_i)}{X_i + b_i - 1}\right| I_{[X_i\geq 1]} \leq \frac{c(\mathbf{X}-\mathbf{e}_i)}{T + pd - 1} I_{[X_i\geq 1]} < \frac{G}{pd} < 1.$$

Next, using (2.4) and Lemma 3.1 of Dey, Ghosh and Srinivasan (1987), it follows that

$$\log\left(1 + \frac{\phi_i(\mathbf{X}-\mathbf{e}_i)}{X_i + b_i - 1}\right)I_{[X_i\geq 1]}$$

$$(2.5)\qquad \geq \left\{\frac{\phi_i(\mathbf{X}-\mathbf{e}_i)}{X_i + b_i - 1} - \frac{3 - G(pd)^{-1}}{6\left(1 - G(pd)^{-1}\right)}\frac{\phi_i^2(\mathbf{X}-\mathbf{e}_i)}{(X_i + b_i - 1)^2}\right\}I_{[X_i\geq 1]}$$

$$\geq \left\{-\frac{c(\mathbf{X}-\mathbf{e}_i)}{T + pd - 1} - \frac{1}{2}pd(pd - G)^{-1}\frac{c^2(\mathbf{X}-\mathbf{e}_i)}{(T + pd - 1)^2}\right\}I_{[X_i\geq 1]}.$$

Using the fact that $c(\mathbf{X})$ is monotone in its arguments, it follows from (2.3) and (2.5) that

$$U(\mathbf{X}) \leq -\frac{c(\mathbf{X})(T + B)}{T + pd} + \left\{\frac{c(\mathbf{X})T}{T + pd - 1} + \frac{pd}{2(pd - G)}\frac{Tc^2(\mathbf{X})}{(T + pd - 1)^2}\right\}I_{[T\geq 1]}$$

$$\leq \left[-\frac{c(\mathbf{X})(T + B)}{T + pd} + \frac{c(X)\mathbf{T}}{T + pd - 1} + \frac{pd}{2(pd - G)}\frac{Tc^2(\mathbf{X})}{(T + pd - 1)^2}\right]I_{[T\geq 1]}$$

$$(2.6)\qquad < \frac{-c(\mathbf{X})T}{(T + pd)(T + pd - 1)}\left[(B - 1) - \frac{pd + 1}{2(pd - G)}c(\mathbf{X})\right]I_{[T\geq 1]}$$

$$= -\frac{c(\mathbf{X})T(pd + 1)}{(T + pd)(T + pd - 1)}(G - c(\mathbf{X}))I_{[T\geq 1]}$$

$$< 0,$$

where assumption (ii) is used in the last step. The lemma follows now from (2.2) and (2.6). □

Proof of Theorem 1. We are now in a position to prove Theorem 1. From Lemmas 1 and 3, it follows that $\boldsymbol{\delta}^0(\mathbf{X}) = \mathbf{XC} + \mathbf{b}$ is inadmissible when any element of $\mathbf{b} = 0$ or a diagonal element of $\mathbf{C}$, say c_l, exceeds 1 with corresponding $b_l > 0$. Further, if some of the $c_j = 1$ and sum of the corresponding $b_j > 1$, then using Lemma 5, one can prove the inadmissibility of $\boldsymbol{\delta}^0(\mathbf{X})$. Hence, suppose that $\boldsymbol{\delta}^0(\mathbf{X})$ has components of the type $c_jX_j + b_j$, where each j in S belongs to one of the two sets

$$S_1 = \{j: 0 \le c_j < 1,\ b_j > 0\}$$

and

$$S_2 = \left\{j: c_j = 1,\ b_j > 0 \text{ and } \sum_{j:\ c_j = 1} b_j \le 1\right\}.$$

If either S_1 or S_2 is empty, then appealing to Lemmas 2 and 4, one proves the admissibility of $\boldsymbol{\delta}^0(\mathbf{X})$.

Suppose now neither S_1 nor S_2 is empty. Let $\boldsymbol{\delta}(\mathbf{X}) = (\delta_1(\mathbf{X}), \dots, \delta_p(\mathbf{X}))$ dominate $\boldsymbol{\delta}^0(\mathbf{X})$. Writing the component losses of $L(\boldsymbol{\theta}, \mathbf{a})$ as $L(\theta_j, a_j)$, one gets

$$E_{\boldsymbol{\theta}}\left[\sum_{j=1}^{p} L\big(\theta_j, \delta_j(\mathbf{X})\big)\right] \le E_{\boldsymbol{\theta}}\left[\sum_{j=1}^{p} L\big(\theta_j, \delta_j^0(\mathbf{X})\big)\right], \tag{2.7}$$

for all $\boldsymbol{\theta} \in (0, \infty)^p = \Theta$ with strict inequality for some $\boldsymbol{\theta} \in \Theta$. Integrating both sides of (2.7) with respect to pdf's corresponding to the priors $\xi_j(\theta_j)$ described in Lemma 2 for $j \in S_1$, one gets

$$\begin{aligned}
&\int \sum_{j \in S_1} E_{\boldsymbol{\theta}}\big[L\big(\theta_j, \delta_j(\mathbf{X})\big)\big] \prod_{j \in S_1} d\xi_j(\theta_j) \\
&\qquad + \int \sum_{j \in S_2} E_{\boldsymbol{\theta}}\big[L\big(\theta_j, \delta_j(\mathbf{X})\big)\big] \prod_{j \in S_1} d\xi_j(\theta_j) \\
&\le \int \sum_{j \in S_1} E_{\boldsymbol{\theta}}\big[L\big(\theta_j, \delta_j^0(\mathbf{X})\big)\big] \prod_{j \in S_1} d\xi_j(\theta_j) \\
&\qquad + \int \sum_{j \in S_2} E_{\boldsymbol{\theta}}\big[L\big(\theta_j, \delta_j^0(\mathbf{X})\big)\big] \prod_{j \in S_1} d\xi_j(\theta_j).
\end{aligned} \tag{2.8}$$

But, since $\delta_j^0(\mathbf{X})$ is the Bayes estimator of θ_j with finite Bayes risk for every $j \in S_1$, the first term in the rhs of (2.8) is less than or equal to the first term in the lhs of (2.8). Hence,

$$\begin{aligned}
&\int \sum_{j \in S_2} E_{\boldsymbol{\theta}}\big[L\big(\theta_j, \delta_j(\mathbf{X})\big)\big] \prod_{j \in S_1} d\xi_j(\theta_j) \\
&\qquad \le \int \sum_{j \in S_2} E_{\boldsymbol{\theta}}\big[L\big(\theta_j, \delta_j^0(\mathbf{X})\big)\big] \prod_{j \in S_1} d\xi_j(\theta_j).
\end{aligned} \tag{2.9}$$

Let $\boldsymbol{\theta}_+$ denote the vector with its elements equal to θ_j ($j \in S_2$). Both the lhs and rhs of (2.9) depend on $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}_+$. Also, the rhs of (2.9) involves only the elements of $\mathbf{X}_+$, the vector with its elements X_j ($j \in S_2$). The lhs of (2.9) involves the elements of $\mathbf{X}$, where the elements of $\mathbf{X}_+$ are independent Poisson with parameters equal to the corresponding elements of $\boldsymbol{\theta}_+$, while the remaining elements have certain negative binomial distributions depending on the b_j's and c_j's or are Poisson with parameters b_j ($j \in S_1$). Using convexity of the loss and the sufficiency of $\mathbf{X}_+$ for $\boldsymbol{\theta}_+$, it follows that there exists $\boldsymbol{\delta}^1(\mathbf{X}_+)$ such that

$$\begin{aligned}
&\int E_{\boldsymbol{\theta}_+}\left[\sum_{j\in S_2} L\left(\theta_j, \delta_j^1(\mathbf{X}_+)\right)\right] \prod_{j\in S_1} d\xi_j(\theta_j) \\
(2.10)\qquad &\le \int \sum_{j\in S_2} E_{\boldsymbol{\theta}}\left[L\left(\theta_j, \delta_j(\mathbf{X})\right)\right] \prod_{j\in S_1} d\xi_j(\theta_j) \\
&\le \int E_{\boldsymbol{\theta}_+}\left[\sum_{j\in S_2} L\left(\theta_j, \delta_j^0(\mathbf{X}_+)\right)\right] \prod_{j\in S_1} d\xi_j(\theta_j).
\end{aligned}$$

But using Lemma 4, $\boldsymbol{\delta}^0(\mathbf{X}_+)$ is an admissible estimator of $\boldsymbol{\theta}_+$ within the class ot all estimators depending only on $\mathbf{X}_+$. Thus, using convexity of the loss, one must have $\boldsymbol{\delta}^1(\mathbf{X}_+) = \boldsymbol{\delta}^0(\mathbf{X}_+)$ a.e., because otherwise $\frac{1}{2}(\boldsymbol{\delta}^1(\mathbf{X}_+) + \boldsymbol{\delta}^0(\mathbf{X}_+))$ dominates $\boldsymbol{\delta}^0(\mathbf{X}_+)$. Hence, one must have equality in (2.10) and, accordingly, in (2.9). Hence, from (2.8), (2.9) and the unique Bayesness of $\delta_j^0(\mathbf{X})$ with respect to the prior ξ_j, one gets

$$(2.11)\qquad E_{\boldsymbol{\theta}}\left[\sum_{j\in S_1} L\left(\theta_j, \delta_j(\mathbf{X})\right)\right] = E_{\boldsymbol{\theta}}\left[\sum_{j\in S_1} L\left(\theta_j, \delta_j^0(\mathbf{X})\right)\right].$$

Hence, from (2.8) and (2.11), one must have

$$(2.12)\qquad E_{\boldsymbol{\theta}}\left[\sum_{j\in S_2} L\left(\theta_j, \delta_j(\mathbf{X})\right)\right] \le E_{\boldsymbol{\theta}}\left[\sum_{j\in S_2} L\left(\theta_j, \delta_j^0(\mathbf{X})\right)\right],$$

with strict inequality for some $\boldsymbol{\theta} \in \Theta$. Arguing as before with the sufficiency of $\mathbf{X}_+$ for $\boldsymbol{\theta}_+$, convexity of the loss and using Lemma 4, one finds that (2.12) leads to a contradiction. Thus, $\boldsymbol{\delta}^0(\mathbf{X})$ is admissible when all $j \in S_1 \cup S_2$. The proof of Theorem 1 is now complete. □

3. Hierarchical Bayes estimators. In this section, we develop certain Bayes estimators using hierarchical priors of Ghosh (1983). According to this formulation, conditional on $U = u$ $(0 < u < 1)$, θ_i's have independent gamma $(u/(1-u), k_i)$ prior distributions, while U has a beta (m, n) prior distribution. Then, using the results of Ghosh (1983), the Bayes estimator of $\boldsymbol{\theta}$ is given by $\boldsymbol{\delta}^H(\mathbf{X})$, where

$$\begin{aligned}
\delta_j^H(\mathbf{X}) &= \frac{n+T}{k+m+n+T}(X_j + k_j) \\
(3.1)\qquad &= \left(1 - \frac{k+m}{k+m+n+T}\right)(X_j + k_j),
\end{aligned}$$

for $j = 1, \dots, p$, where $k = \sum_{j=1}^p k_j$ and $T = \sum_{j=1}^p X_j$. These estimators have an interesting empirical Bayes interpretation. Suppose we consider only the first stage prior as described previously. Then the Bayes estimator of $\boldsymbol{\theta}$ is $(1-u)$ $(\mathbf{X} + \mathbf{k})$, where $\mathbf{k} = (k_1, \dots, k_p)$. In an empirical Bayes framework, u is unknown and is estimated from the marginal distribution of the X_i's. It is easy to find that marginally X_i's are independently distributed negative binomial random variables with probability functions

$$(3.2)\qquad P(X_i = x_i) = \frac{\Gamma(x_i + k_i)}{x_i!\,\Gamma(k_i)} u^{k_i}(1-u)^{x_i},$$

$x_i = 0, 1, \dots, k_i$ $(i = 1, \dots, p)$. Then, marginally $T = \sum_{j=1}^{p} X_j$ is sufficient for u, having a negative binomial distribution of the form

$$P(T = t) = \frac{\Gamma(t+k)}{t!\Gamma(k)} u^k (1-u)^t, \qquad t = 0, 1, \dots, \tag{3.3}$$

where $k = \sum_{i=1}^{p} k_i$. Estimating u from the marginal distribution of T, one gets $(1 - \hat{u}(T))(\mathbf{X} + \mathbf{k})$ as an empirical Bayes estimator of $\boldsymbol{\theta}$. The estimator given in (3.1) is one such estimator with $\hat{u}(T) = (k + m)/(k + m + n + T)$. If one identifies k_i with b_i in the definition of $\phi_i(\mathbf{X})$ in Lemma 5 and takes $c(\mathbf{X}) = k + m$ and $pd = p + m + n$, then, using Lemma 5, the estimator $\boldsymbol{\delta}^H(\mathbf{X})$ dominates the generalized Bayes estimator of $\boldsymbol{\theta}$ provided $k = \sum_{i=1}^{p} k_i > 1$ and $k + m < 2(k-1)(k+m+n)/(k+m+n+1+2(k-1))$. This is equivalent to $n/(k + m + n + 1) > (k + m)/(2(k - 1))$. This is impossible to achieve for every given m, n and p, for example, when $m > k - 2$. However, if $m < k - 2$ $(k \geq 3)$ and n is very large compared to $k + m + 1$, it is possible to construct a proper Bayes estimator which dominates $\mathbf{X} + \mathbf{k}$. Thus, for $k \geq 3$, it is possible to construct proper Bayes estimators using hierarchical priors which dominate $\mathbf{X} + \mathbf{k}$. One merit of constructing such estimators is that they are typically admissible.

4. Monte Carlo simulations. It is pointed out in the introduction that estimators $\mathbf{X} + \mathbf{b}$ of $\boldsymbol{\theta}$ are generalized Bayes with respect to improper (limiting gamma) priors. One such estimator is $\mathbf{X} + \mathbf{1}$. In some sense, this is a plausible estimator of $\boldsymbol{\theta}$, since it prevents its components from being too small near the origin. Note also, that $\mathbf{X} + \mathbf{1}$ has finite risk for all $\boldsymbol{\theta} \in (0, \infty)^p$. However, it is shown in Section 2 that $\mathbf{X} + \mathbf{1}$ is an inadmissible estimator of $\boldsymbol{\theta}$ for $p \geq 2$ under entropy loss. In this section, using Monte Carlo simulations, we show the risk improvement of certain estimators given in Lemma 5 dominating $\mathbf{X} + \mathbf{1}$.

With this end, for different p, first take certain values of $\theta_1, \dots, \theta_p$ and compute the risks of $\boldsymbol{\delta}^0(\mathbf{X}) = \mathbf{X} + \mathbf{1}$ and

$$\boldsymbol{\delta}^*(\mathbf{X}) = \mathbf{X} - \frac{2pd(p-1)}{(pd + 2p - 1)\sum_{j=1}^{p}(X_j + d)}(\mathbf{X} + \mathbf{1})$$

under the loss (1.2) for $d = 1, 2, 3, 5, 10$ and $p = 2, 5, 10$. Now, $X_1, \dots, X_p$ are generated by the IMSL routine GGPON such that X_i has a Poisson distribution with parameter θ_i, $i = 1, \dots, p$. Then, the losses $L(\boldsymbol{\theta}, \boldsymbol{\delta}^0)$ and $L(\boldsymbol{\theta}, \boldsymbol{\delta}^*)$ are calculated. This procedure is repeated 1000 times to compute the simulated versions of $R(\boldsymbol{\theta}, \boldsymbol{\delta}^0)$ and $R(\boldsymbol{\theta}, \boldsymbol{\delta}^*)$ as the averages of $L(\boldsymbol{\theta}, \boldsymbol{\delta}^0)$ and $L(\boldsymbol{\theta}, \boldsymbol{\delta}^*)$, respectively. From the preceding, the percentage risk reduction is computed as $100[R(\boldsymbol{\theta}, \boldsymbol{\delta}^0) - R(\boldsymbol{\theta}, \boldsymbol{\delta}^*)]/R(\boldsymbol{\theta}, \boldsymbol{\delta}^0)$. The preceding procedure is repeated 10 times by keeping the θ_i's fixed. The averages and standard deviations of the 10 percentage risk reductions are reported in Tables 1–3 corresponding to $p = 2$, 5 and 10, respectively.

TABLE 1
($p = 2$) *Percentage of risk reduction for* δ^* *over* δ^0.
(*Estimated standard deviations are in parentheses.*)

d / $(\theta_1, \ldots, \theta_p)$	1	2	3	5	10
(0.5, 0.5)	44.52 (0.26)	40.20 (0.17)	34.51 (0.11)	25.99 (0.09)	15.94 (0.04)
(0.5, 2.0)	23.57 (0.23)	27.29 (0.15)	26.39 (0.14)	22.04 (0.14)	14.88 (0.07)
(0.5, 3.5)	14.07 (0.28)	18.25 (0.23)	18.65 (0.20)	17.20 (0.15)	12.77 (0.13)
(1.0, 3.0)	15.36 (0.38)	20.23 (0.19)	20.42 (0.18)	18.63 (0.18)	14.26 (0.07)
(2.0, 2.0)	14.93 (0.23)	20.41 (0.28)	20.71 (0.18)	19.33 (0.18)	14.45 (0.11)
(2.0, 4.0)	10.40 (0.29)	12.75 (0.27)	14.47 (0.19)	14.69 (0.21)	11.57 (0.12)
(3.0, 4.0)	8.48 (0.33)	10.33 (0.31)	12.10 (0.26)	12.75 (0.21)	10.80 (0.14)
(4.0, 6.0)	5.02 (0.30)	7.27 (0.26)	8.16 (0.27)	9.12 (0.33)	8.22 (0.23)
(4.0, 8.0)	4.60 (0.32)	6.06 (0.24)	6.27 (0.18)	7.82 (0.30)	7.26 (0.17)
(5.0, 7.0)	4.35 (0.21)	5.70 (0.28)	6.46 (0.14)	7.28 (0.16)	7.33 (0.09)
(6.0, 6.0)	4.45 (0.15)	5.69 (0.36)	6.46 (0.20)	7.37 (0.15)	7.50 (0.21)
(7.0, 8.0)	3.33 (0.20)	4.29 (0.27)	4.70 (0.21)	5.67 (0.22)	5.94 (0.23)
(8.0, 8.0)	3.37 (0.14)	3.89 (0.26)	4.49 (0.25)	4.92 (0.21)	5.68 (0.21)
(8.0, 10.0)	2.79 (0.19)	3.50 (0.27)	4.00 (0.27)	4.64 (0.08)	5.38 (0.23)
(8.0, 12.0)	2.48 (0.13)	3.07 (0.27)	3.52 (0.29)	3.83 (0.24)	4.53 (0.21)
(9.0, 11.0)	2.29 (0.15)	3.25 (0.18)	4.02 (0.23)	4.10 (0.25)	4.55 (0.25)
(10.0, 12.0)	2.19 (0.16)	2.78 (0.14)	3.28 (0.30)	3.75 (0.28)	4.15 (0.24)

TABLE 2
($p = 5$) *Percentage of risk reduction for* δ^* *over* δ^0.
(*Estimated standard deviations are in parentheses.*)

d / $(\theta_1, \ldots, \theta_p)$	1	2	3	5	10
(1.0, 1.0, 1.0, 1.0, 1.0)	48.97 (0.20)	50.34 (0.11)	47.45 (0.09)	39.32 (0.09)	26.43 (0.04)
(3.0, 3.0, 3.0, 3.0, 3.0)	19.32 (0.15)	22.10 (0.33)	22.50 (0.21)	22.43 (0.17)	18.15 (0.14)
(0.5, 1.0, 2.0, 3.0, 4.0)	25.90 (0.20)	29.51 (0.22)	29.51 (0.24)	27.54 (0.08)	20.64 (0.13)
(0.8, 1.6, 2.4, 3.2, 4.0)	23.48 (0.17)	27.15 (0.28)	27.50 (0.14)	26.68 (0.14)	20.02 (0.09)
(5.0, 5.0, 5.0, 5.0, 5.0)	11.34 (0.12)	13.08 (0.15)	13.46 (0.18)	14.29 (0.28)	13.25 (0.21)
(7.0, 7.0, 7.0, 7.0, 7.0)	8.15 (0.13)	9.07 (0.30)	9.86 (0.24)	10.50 (0.16)	10.04 (0.20)
(4.0, 5.0, 6.0, 7.0, 8.0)	9.67 (0.11)	10.95 (0.25)	11.81 (0.25)	12.42 (0.22)	11.41 (0.17)
(4.8, 5.6, 6.4, 7.2, 8.0)	8.93 (0.25)	10.28 (0.26)	11.14 (0.25)	11.10 (0.28)	10.91 (0.18)
(9.0, 9.0, 9.0, 9.0, 9.0)	6.52 (0.13)	7.45 (0.21)	7.66 (0.22)	8.26 (0.09)	8.10 (0.19)
(11.0, 11.0, 11.0, 11.0, 11.0)	5.48 (0.12)	6.11 (0.20)	6.62 (0.22)	6.03 (0.22)	6.84 (0.21)
(8.0, 9.0, 10.0, 11.0, 12.0)	5.78 (0.10)	6.92 (0.29)	7.00 (0.32)	7.40 (0.19)	7.27 (0.17)
(8.8, 9.6, 10.4, 11.2, 12.0)	5.47 (0.09)	5.91 (0.19)	6.47 (0.18)	6.86 (0.19)	6.02 (0.15)

The preceding simulation results suggest that for a given p and d, the maximum risk reduction takes place when all the θ_i's are near zeros. This is quite intuitive since $\delta^*(\mathbf{X})$ shrinks $\delta^0(\mathbf{X})$ toward zero. Also, when the θ_i values are equally spaced within certain intervals [for example, (2.0, 4.0) for $p = 2$, (0.8, 1.6, 2.4, 3.2, 4.0) for $p = 5$ and (0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2, 3.6, 4.0) for $p = 10$ all involve spaced θ_i valued within (0.00, 4.00)], the risk improvement seems to be more significant when the dimension increases.

TABLE 3
($p = 10$) *Percentage of risk reduction for* δ^* *over* δ^0.
(*Estimated standard deviations are in parentheses.*)

d / $(\theta_1, \ldots, \theta_p)$	1	2	3	5	10
(1.0, 1.0, 1.0, 1.0, 1.0, 3.0, 3.0, 3.0, 3.0, 3.0)	32.33 (0.17)	35.37 (0.08)	35.13 (0.15)	32.25 (0.15)	24.33 (0.07)
(2.0, 2.0, 2.0, 2.0, 2.0, 4.0, 4.0, 4.0, 4.0, 4.0)	23.09 (0.19)	25.87 (0.17)	25.93 (0.19)	24.93 (0.16)	20.34 (0.13)
(0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2, 3.6, 4.0)	29.33 (0.15)	32.05 (0.14)	32.25 (0.24)	29.60 (0.14)	22.81 (0.08)
(5.0, 5.0, 5.0, 5.0, 5.0, 7.0, 7.0, 7.0, 7.0, 7.0)	11.57 (0.15)	13.24 (0.30)	13.51 (0.24)	13.79 (0.17)	12.88 (0.16)
(6.0, 6.0, 6.0, 6.0, 6.0, 8.0, 8.0, 8.0, 8.0, 8.0)	10.13 (0.13)	11.29 (0.22)	11.68 (0.08)	12.06 (0.16)	11.54 (0.12)
(4.4, 4.8, 5.2, 5.6, 6.0, 6.4, 6.8, 7.2, 7.6, 8.0)	11.27 (0.13)	12.80 (0.16)	13.22 (0.26)	13.04 (0.16)	12.78 (0.11)

5. Concluding remarks. For simultaneous estimation of means of p independent Poisson variables under the entropy loss given in (1.2), so far we have considered estimators of the form $\mathbf{XC} + \mathbf{b}$ when $\mathbf{C}$ is diagonal. Theorem 1, given in Section 2, provides necessary and sufficient conditions for admissibility of such estimators. Utilizing our Theorem 1 and Theorem 1 of Brown and Farrell (1985), one can extend the admissibility results for an arbitrary $\mathbf{C}$.

With this end, first note that following the line of argument of Brown and Farrell (1983) [see also Brown (1981)], one can show that a necessary condition for an estimator δ of θ to be admissible is that δ is the pointwise limit of a sequence of Bayes estimators. Since Bayes estimators under the squared error and entropy losses are the same, such estimators retain the same property as given in (1.2) of Brown and Farrell (1985). Now minor modifications of the proofs of Theorems 1 and 2 of Brown and Farrell (1985) lead to the following theorem which characterizes admissible estimators of the form $\mathbf{XC} + \mathbf{b}$, where $\mathbf{C}$ is not necessarily diagonal. The proof of the theorem is omitted for brevity.

THEOREM 2. *Under the loss* (1.2):

(i) *If* $\mathbf{C}$ *is nonsingular, then* $\delta(\mathbf{X}) = \mathbf{X}C + \mathbf{b}$ *is admissible if and only if* $\mathbf{C} = \mathbf{D}$ *is a diagonal matrix and satisfies the conditions* (i)–(ii) *of Theorem* 1.

(ii) *If* $\mathbf{C}$ *is singular, then* $\delta(\mathbf{X}) = \mathbf{XC} + \mathbf{b}$ *is admissible if and only if there exists a permutation* σ *of* $1, \ldots, p$ *for* $\mathbf{C}^t$ *and* $\mathbf{b}^t$ (*where* $\mathbf{C}^t$ *and* $\mathbf{b}^t$ *denote, respectively, the transpose of* $\mathbf{C}$ *and* $\mathbf{b}$) *such that the permuted matrix* $(\mathbf{C}^t)' = (\mathbf{C}^t_{\sigma(i)\sigma(j)})$ *has the block form*

$$(\mathbf{C}^t)' = \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \ddots & & & \\ \mathbf{0} & & \cdots & \mathbf{M}_s & \cdots & \mathbf{0} \\ \mathbf{0} & & \cdots & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

and the permuted vector $(\mathbf{b}^t)' = (\beta_1 \mathbf{m}_1, \beta_2 \mathbf{m}_2, \dots, \beta_s \mathbf{m}_s, \mathbf{y})^t$. *In the preceding* $\mathbf{M}_i$ *are* $(r_i \times r_i)$ *with the rank* 1 *and* $\mathbf{M}_i = \sigma_i \mathbf{m}_i^t (1, \dots, 1)$, $i = 1, 2, \dots, s$, *where* $\mathbf{m}_i^t$ *is an* $(r_i \times 1)$ *vector with positive entries satisfying* $(1, \dots, 1)\mathbf{m}_i^t = 1$, $1 \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_s > 0$, $\mathbf{y}$ *is a* $(r_{s+1} \times 1)$ *vector of positive entries*, $\Sigma_{i=1}^{s+1} r_i = p$, $\beta_i > 0$, $i = 1, \dots, s$, *and if there exists an* i *such that* $\sigma_i = 1$ *and* $\sigma_{i+1} < 1$ (*with* $\sigma_{s+1} = 0$), *then* $\Sigma_{k=1}^{i} \beta_k \leq 1$.

APPENDIX

PROOF OF LEMMA 4. Under the prior $\pi(\boldsymbol{\theta}) = (\prod_{j=1}^{p} \theta_j^{b_j - 1}) g(\boldsymbol{\theta})$, where $b_j > 0$ for all j, and the loss (1.2), the Bayes estimator of $\boldsymbol{\theta}$ is $\boldsymbol{\delta}^g(\mathbf{X}) = (\delta_1^g(\mathbf{X}), \dots, \delta_p^g(\mathbf{X}))$, where

$$(\text{A.1}) \qquad \begin{aligned} \delta_i^g(\mathbf{x}) = & \left[\int_0^\infty \cdots \int_0^\infty \theta_i \exp\left(-\sum_1^p \theta_j \right) \prod_{j=1}^{p} \theta_j^{x_j + b_j - 1} g(\boldsymbol{\theta})\, d\theta_1 \cdots d\theta_p \right] \\ & \div \left[\int_0^\infty \cdots \int_0^\infty \exp\left(-\sum_1^p \theta_j \right) \prod_{j=1}^{p} \theta_j^{x_j + b_j - 1} g(\boldsymbol{\theta})\, d\theta_1 \cdots d\theta_p \right]. \end{aligned}$$

Define $I_{\mathbf{x}}(g)$ as the denominator given in (A.1). Then, integration by parts leads to

$$(\text{A.2}) \qquad \delta_i^g(\mathbf{x}) = x_i + b_i + \left[I_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g) / I_{\mathbf{x}}(g) \right],$$

where $\nabla_i g = \partial g / \partial \theta_i$ and $\mathbf{e}_i$ is the ith unit vector. Take $g(\boldsymbol{\theta}) = 1$ and $g_n(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) h_n^2(\boldsymbol{\theta})$, where $h_n^2(\boldsymbol{\theta})$ will be defined later. Then $I_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g) = 0$ and one gets from (A.2),

$$(\text{A.3}) \qquad g_i^g(\mathbf{x}) - \delta_i^{g_n}(\mathbf{x}) = -\frac{I_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g_n)}{I_{\mathbf{x}}(g_n)}.$$

Hence, writing $r(\pi, \boldsymbol{\delta})$ as the Bayes risk of an estimator $\boldsymbol{\delta}$ of $\boldsymbol{\theta}$ with respect to the prior π, and writing $\pi_n(\boldsymbol{\theta}) = \prod_{j=1}^{p} \theta_j^{b_j - 1} g_n(\theta)$, one gets

$$(\text{A.4}) \quad r(\pi_n, \boldsymbol{\delta}^g) - r(\pi_n, \boldsymbol{\delta}^{g_n}) = \sum_{i=1}^{p} E\left[\delta_i^g(\mathbf{X}) - \delta_i^{g_n}(\mathbf{X}) - \theta_i \log \frac{\delta_i^g(\mathbf{X})}{\delta_i^{g_n}(\mathbf{X})} \right],$$

where E denotes expectation over the joint distribution of $\boldsymbol{\theta}$ and $\mathbf{X}$. Note that

$$(\text{A.5}) \qquad \begin{aligned} E\left[\delta_i^g(\mathbf{X}) - \delta_i^{g_n}(\mathbf{X}) \right] &= -\sum_{x_1 = 0}^{\infty} \cdots \sum_{x_p = 0}^{\infty} \int_0^\infty \cdots \int_0^\infty \exp\left(-\sum_1^p \theta_j \right) \\ & \qquad \times \left(\frac{\prod_{j=1}^{p} \theta_j^{x_j + b_j - 1}}{\prod_{j=1}^{p} x_j!} \right) \theta \frac{\partial g_n}{\partial \theta_i}\, d\theta_1 \cdots d\theta_p \\ &= -\int_0^\infty \cdots \int_0^\infty \left(\prod_{j=1}^{p} \theta_j^{b_j - 1} \right) \theta_i \frac{\partial g_n}{\partial \theta_i}\, d\theta_1 \cdots d\theta_p \end{aligned}$$

and

$$-E\left[\theta_i \log\frac{\delta_i^g(\mathbf{X})}{\delta_i^{g_n}(\mathbf{X})}\right] = -E\left[\theta_i \log\frac{X_i+b_i}{X_i+b_i+\left\{I_{\mathbf{X}+\mathbf{e}_i}(\nabla_i g_n)/I_{\mathbf{X}}(g_n)\right\}}\right]$$

$$\text{(A.6)}\qquad \begin{aligned} &= E\left[\theta_i \log\left(\frac{X_i+b_i}{X_i+b_i+\left\{I_{\mathbf{X}+\mathbf{e}_i}(\nabla_i g_n)/I_{\mathbf{X}}(g_n)\right\}}\right)^{-1}\right] \\ &\le E\left[\theta_i \frac{I_{\mathbf{X}+\mathbf{e}_i}(\nabla_i g_n)}{(X_i+b_i)I_{\mathbf{X}}(g_n)}\right] \\ &= \sum_{x_1=0}^{\infty}\cdots\sum_{x_p=0}^{\infty}\frac{1}{\prod_{j=1}^p x_j!}\frac{I_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g_n)I_{\mathbf{x}+\mathbf{e}_i}(g_n)}{(x_i+b_i)I_{\mathbf{x}}(g_n)}. \end{aligned}$$

Next, noting that $I_{\mathbf{x}+\mathbf{e}_i}(g_n) = (x_i+b_i)I_{\mathbf{x}}(g_n) + I_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g_n)$, one gets from (A.6),

$$\text{(A.7)}\qquad \begin{aligned} &-E\left[\theta_i \log\frac{\delta_i^g(\mathbf{X})}{\delta_i^{g_n}(\mathbf{X})}\right] \\ &\le \sum_{x_1=0}^{\infty}\cdots\sum_{x_p=0}^{\infty}\frac{1}{\prod_{j=1}^p x_j!}\left[I_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g_n) + \frac{I^2_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g_n)}{(x_i+b_i)I_{\mathbf{x}}(g_n)}\right] \\ &= \int_0^\infty\cdots\int_0^\infty\left(\prod_{j=1}^p \theta_j^{b_j-1}\right)\theta_i\frac{\partial g_n}{\partial\theta_i}\,d\theta_1\cdots d\theta_p \\ &\quad+\sum_{x_1=0}^{\infty}\cdots\sum_{x_p=0}^{\infty}\frac{1}{\prod_{j=1}^p x_j!}\frac{I^2_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g_n)}{(x_i+b_i)I_{\mathbf{x}}(g_n)}. \end{aligned}$$

Combining (A.4), (A.5) and (A.7), one gets

$$\text{(A.8)}\qquad \begin{aligned} &r(\pi_n,\delta^g) - r(\pi_n,\delta^{g_n}) \\ &\le \sum_{i=1}^p\sum_{x_1=0}^{\infty}\cdots\sum_{x_p=0}^{\infty}\frac{1}{\left(\prod_{j=1}^p x_j!\right)(x_i+b_i)}\frac{I^2_{\mathbf{x}+\mathbf{e}_i}(\nabla_i g_n)}{I_{\mathbf{x}}(g_n)} \\ &\le \sum_{i=1}^p\sum_{x_1=0}^{\infty}\cdots\sum_{x_p=0}^{\infty}\frac{1}{\left(\prod_{j=1}^p x_j!\right)(x_i+b_i)} \\ &\quad\times\int_0^\infty\cdots\int_0^\infty e^{-\Sigma_1^p\theta_j}\theta_i^2\left(\prod_{j=1}^p\theta_j^{x_j+b_j-1}\right)\left(\frac{\partial h_n}{\partial\theta_i}\right)^2 d\theta_1\cdots d\theta_p \\ &\le \sum_{i=1}^p\left(1+b_i^{-1}\right)\sum_{x_i=0}^{\infty}\left((x_i+1)!\right)^{-1} \\ &\quad\times\int_0^\infty\cdots\int_0^\infty e^{-\theta_i}\theta_i^{x_i+1}\theta_i\left(\prod_{j=1}^p\theta_j^{b_j-1}\right)\left(\frac{\partial h_n}{\partial\theta_i}\right)^2 d\theta_1\cdots d\theta_p \\ &\le \left(1+\max_{1\le i\le p} b_i^{-1}\right)\sum_{i=1}^p\int_0^\infty\cdots\int_0^\infty\theta_i\left(\prod_{j=1}^p\theta_j^{b_j-1}\right)\left(\frac{\partial h_n}{\partial\theta_i}\right)^2 d\theta_1\cdots d\theta_p. \end{aligned}$$

Now take

$$\text{(A.9)}\qquad \begin{aligned} h_n(\boldsymbol{\theta}) &= 1, \quad \text{if } 0 \le \mu = \sum_{i=1}^{p} \theta_i \le 1,\\ &= 1 - \frac{\log \mu}{\log n}, \quad \text{if } 1 < \mu < n,\\ &= 0, \quad \text{otherwise}. \end{aligned}$$

Then, from (A.8) it follows that

$$\text{(A.10)}\qquad \begin{aligned} &r(\pi_n, \delta^g) - r(\pi_n, \delta^{g_n})\\ &\le \Bigl(1 + \max_{1 \le i \le p} b_i^{-1}\Bigr)(\log n)^{-2} \int \cdots \int_{1<\mu<n} \mu^{-1}\Biggl(\prod_{j=1}^{p} \theta_j^{b_j - 1}\Biggr) d\theta_1 \cdots d\theta_p. \end{aligned}$$

Using the transformation $\mu = \Sigma_1^p \phi_i$, $\theta_i = \theta_i/\mu$, $i = 1, \dots, p-1$, it follows that the integral in the rhs of (A.10) equals

$$\text{(A.11)}\qquad \begin{aligned} &\int \cdots \int_{\phi_j > 0, \Sigma_1^{p-1}\phi_j \le 1} \int_1^n \mu^{-1}\Bigl(\mu^{\Sigma_1^p b_j - p + p - 1}\Bigr)\\ &\qquad \times \prod_1^{p-1} \phi_j^{b_j - 1}\Biggl(1 - \sum_1^{p-1} \phi_j\Biggr)^{b_p - 1} d\mu\, d\phi_1 \cdots d\phi_{p-1}\\ &\le \text{const.} \int_1^n \mu^{\Sigma_1^p b_j - 2}\, d\mu\\ &\le \text{const.}, \quad \text{if } \sum_1^p b_j < 1,\\ &\le \text{const.}\,(\log n), \quad \text{if } \sum_1^p b_j = 1. \end{aligned}$$

It follows from (A.10) and (A.11) that $r(\pi_n, \delta^g) - r(\pi_n, \delta^{g_n}) \to 0$ and $n \to \infty$ when $\Sigma_{j=1}^p b_j \le 1$, and using Blyth's theorem [see, for example, Brown and Hwang (1982)], the proof Lemma 4 is complete. □

Acknowledgment. Thanks are due the referee for his constructive comments which led to substantial improvement over the original version of the paper.

REFERENCES

BLYTH, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22** 22–42.

BROWN, L. D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.* **39** 29–48.

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300.

BROWN, L. D. and FARRELL, R. H. (1983). Elementary proof of a necessary and sufficient condition for admissibility in discrete multivariate problems. Unpublished.

BROWN, L. D. and FARRELL, R. H. (1985). All admissible linear estimators of a multivariate Poisson mean. *Ann. Statist.* **13** 282–294.

BROWN, L. D. and HWANG, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1** 205–230. Academic, New York.

CLEVENSON, M. L. and ZIDEK, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.

COHEN, A. (1966). All admissible linear estimates of the mean vector. *Ann. Math. Statist.* **37** 458–463.

DEY, D. K., GHOSH, M. and SRINIVASAN, C. (1987). Simultaneous estimation of parameters under entropy loss. *J. Statist. Plann. Inference* **15** 347–363.

DEY, D. K. and SRINIVASAN, C. (1985). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* **13** 1581–1591.

DEY, D. K. and SRINIVASAN, C. (1986). On estimation of Poisson logits. *Comm. Statist. A—Theory Methods* **15** 2087–2097.

GHOSH, M. (1983). Estimates of multiple Poisson means: Bayes and empirical Bayes. *Statist. Decisions* **1** 183–195.

GHOSH, M., HWANG, J. T. and TSUI, K. W. (1983). Construction of improved estimators in multiparameter estimation for discrete exponential families (with discussion). *Ann. Statist.* **11** 351–376.

GIRSHICK, M. A. and SAVAGE, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. *Proc. Second Berkeley Symp. Math. Statist. Probab.* **1** 53–73. Univ. California Press.

HAFF, L. R. (1977). Minimax estimators for a multinormal precision matrix. *J. Multivariate Anal.* **7** 374–385.

HAFF, L. R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.* **7** 1264–1276.

HAFF, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8** 586–597.

HAFF, L. R. (1982). Identities for the Wishart distribution with applications to regression. *Sankhyā Ser. B* **44** 245–258.

HODGES, J. L., JR. and LEHMANN, E. L. (1951). Some applications of the Cramér–Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Probab.* **1** 13–22. Univ. California Press.

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.

IGHODARO, A. O. (1980). Ridge and James–Stein methods for contingency tables. Ph.D. thesis, School of Operations Research/Industrial Engineering, Cornell Univ.

IGHODARO, A. O. and SANTNER, T. (1982). Ridge type estimators of multinomial cell probabilities. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **2** 31–53. Academic, New York.

IGHODARO, A., SANTNER, T. and BROWN, L. D. (1982). Admissibility and complete class results for the multinomial estimation problem with entropy and squared error loss. *J. Multivariate Anal.* **12** 469–479.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press.

KARLIN, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29** 406–436.

PENG, J. C. M. (1975). Simultaneous estimation of the parameters of independent Poisson distributions. Technical Report 78, Dept. Statistics, Stanford Univ.

DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

GRADUATE INSTITUTE OF STATISTICS
NATIONAL CENTRAL UNIVERSITY
CHUNG LI, TAIWAN
REPUBLIC OF CHINA