The Annals of Statistics
1985, Vol. 13, No. 1, 282–294

ALL ADMISSIBLE LINEAR ESTIMATORS OF A MULTIVARIATE POISSON MEAN[1]

BY L. D. BROWN AND R. H. FARRELL

Cornell University

Admissible linear estimators $Mx + \gamma$ must be pointwise limits of Bayes estimators. Using properties of Bayes estimators preserved by taking limits, the structure of M and γ can be determined. Among M, γ with this structure, a necessary and sufficient condition for admissibility is obtained. This condition is applied to the case of linear (mixture) models. It is shown that only the most trivial such models admit linear estimators of full rank which are admissible or are even limits of Bayes estimators.

1. Introduction. In this paper p-dimensional discrete random variables X with

$$P(X = x) = c(\lambda)h(x)\lambda^{(x)} \tag{1.1}$$

are studied, λ a p-dimensional parameter vector with transpose $\lambda^t \in (0, \infty) \times \cdots \times (0, \infty)$ and $x \in \mathscr{X} = \{x: x^t \in [0, \infty) \times \cdots \times [0, \infty)$ has integer coordinates$\}$. Assume $h(x) > 0$ for all $x \in \mathscr{X}$. Let e_i be the unit vector with ith coordinate equal one, so that the notation $\lambda^{(x)}$ is defined by $\lambda^{(x)} = \prod_{i=1}^{p} (e_i^t\lambda)^{(e_i^t x)}$.

In another paper, Brown and Farrell (1985), when squared error is used to measure loss, it is shown that admissible estimators δ of the vector λ are pointwise limits of Bayes estimators. From this, together with some development of simple properties of Bayes estimators, it follows that

$$e_i^t\delta(x)e_j^t\delta(x + e_i) = e_j^t\delta(x)e_i^t\delta(x + e_j), \quad 1 \le i, \quad j \le p \tag{1.2}$$

for all $x \in \mathscr{X}$ and for all δ which are pointwise limits of Bayes estimators.

From the relation (1.2) the structure of estimators

$$\delta(x) = Mx + \gamma, \tag{1.3}$$

M a $p \times p$ matrix, γ a $p \times 1$ vector, can be deduced when δ is a pointwise limit of Bayes estimators. See Theorem 1.

When the components of X are independently distributed Poisson random variables (a special case of (1.1)), we obtain a necessary and sufficient condition that an estimator of the form (1.3) be admissible. See Theorem 2 for a statement. The result stated in Theorem 2 complements the result of Cohen (1966) in which all admissible linear estimators of a multivariate normal mean (known covariance) were determined. Linear estimators in the discrete context were discussed by N. L. Johnson (1957).

Received May 1983; revised August 1984.

[1] Research supported in part by the National Science Foundation Grant M.C.S. 8200031.

AMS 1980 *subject classifications*. Primary 62C07; secondary 62F10.

Key words and phrases. Estimation, multivariate Poisson parameter, decision theory, linear estimators, admissibility, linear models.

One motivation for the use of linear estimators is that they are the least squares estimators for linear models. Although many problems involving multivariate Poisson means are modeled by log linear models, there are also practical situations in which linear models are appropriate. One such situation is described in Example 1, below. The general model described in Example 1a occurs in applications to position emission tomography. See Shepp and Vardi (1982) or Vardi, Shepp, and Kaufman (1983).

Certain linear estimators other than least squares estimators also have a heuristic appeal, as discussed in Example 2.

EXAMPLE 1a (*A simple mixture problem*). Consider two radioactive materials with (unknown) decay rates per unit mass of ρ_1 and ρ_2 counts per unit time, respectively. Suppose the experimenter has three ingots of material. It can be determined that the ith unit $i = 1, 2, 3$, contains h_{ij} mass units of material $j = 1, 2$. The total number of counts, X_i, for each ingot is observed over time interval t_i. Thus the observed variables X_i are independent Poisson variables with $E(X_i) = t_i h_{i1}\rho_1 + t_i h_{i2}\rho_2$. There is no loss of generality in setting $t_i \equiv 1$, which we do in the general formulation below.

EXAMPLE 1b (*The general mixture problem*). The general version of Example 1 involves $X \in \mathbb{R}^p$ whose coordinates are independent Poisson variables with

$$\lambda = E(X) = H\rho \tag{1.4}$$

where $H(p \times m)$ has given nonnegative entries, and the parameter vector, $\rho(m \times 1)$, also has nonnegative entries. Assume that rank $H = m$ so that all coordinates of ρ are identifiable and hence estimable. We limit our later discussion to squared error for estimating ρ or for estimating λ. One might expect, since the problem is linear, that there should be reasonable linear estimators which are admissible, or at least limits of Bayes procedures. The obvious suggestion would be (generalized) least squares estimators. However, it follows from Theorem 1 that only in very special situations will there exist nontrivial linear estimators which are admissible or are even limits of Bayes procedures. This is discussed in more detail following the statement of Theorem 2.

EXAMPLE 2 (*Another linear estimator*). Even when λ is unrestricted (except for the condition $\lambda_i \geq 0$), heuristic considerations can motivate the use of linear estimators other than the simple $\delta(x) = ax + b$ $(a \geq 0,\ b \geq 0)$. Consider for example $\delta(x) = Mx$ with

$$M = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}.$$

Rewrite M as $M = (1/2)M_1 + (1/2)M_2$ with $M_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ and $M_2 = I$. Thus, $\delta = (1/2)\delta_1 + (1/2)\delta_2$ where δ_1 is the least square estimator corresponding to $\{\lambda\colon \lambda_1 = \lambda_2\}$ and of course $\delta_2(x) = x$. Both δ_1 and δ_2 are admissible linear estimators. This is taken by some to motivate their average, δ, as a plausible estimator. Such a situation is familiar in the normal means problem where δ is, in fact, admissible.

However, in the Poisson means problem δ is not admissible since it is not a limit of Bayes procedures. This can be easily deduced from Theorem 1.

In the sequel, ordinary squared error loss, $\|\delta(x) - \lambda\|^2$ is referred to as the Case 1 loss function. The Case 2 loss function is $\sum_{i=1}^{p} (e_i^t\lambda)^{-1}(e_i^t(\delta(x) - \lambda))^2$, used by Clevenson and Zidek, (1975). As is shown in Brown and Farrell, the analogue of (1.2) for the Case 2 loss function is

$$e_i^t\delta(x + e_i)e_j^t\delta(x + e_i + e_j) = e_j^t\delta(x + e_j)e_i^t\delta(x + e_i + e_j). \tag{1.5}$$

The necessary condition of Theorem 1 holds for both Case 1 and Case 2 loss functions and all distributions of the form (1.1). Theorem 2, the necessary and sufficient condition, is stated only for the Poisson problem in Case 1 and Case 2.

The proof of Theorem 2 in Case 1 uses known results about squared error due to Peng (1975). Admissibility proofs easier than Peng's are now available in Brown and Hwang (1982), and Ghosh, Hwang and Tsui (1983).

This work is partly an outgrowth of efforts to build a decision theory foundation for the Ph.D. thesis of Iain Johnstone (1981). The authors are indebted to Johnstone for many helpful conversations about the linear problem and the use of condition (1.2).

THEOREM 1. *Consider distributions of the form* (1.1) *with either the Case* 1 *or the Case* 2 *loss functions. Let the estimator* $\delta(x) = Mx + \gamma$ *be a pointwise limit of Bayes estimators,* M *a* $p \times p$ *matrix and* γ *a* $p \times 1$ *vector. Then*

(i) γ *and* M *have nonnegative entries.*

(ii) $M^2 = DM$ *where* D *is the diagonal matrix with* $d_{ii} = \sum_{j=1}^{p} m_{ji}$.

(iii) *If* M *is nonsingular then* $M = D$ *is a diagonal matrix.*

(iv) *If* $M = (m_{ij})$ *is singular there exists a permutation* σ *of* $1, \cdots, p$ *such that the matrix* $M' = (m_{\sigma(i)\sigma(j)})$ *has the block form*

$$M' = \begin{pmatrix} M_1 & 0 & \cdots & 0 & 0 & B_1 \\ 0 & M_2 & \cdots & 0 & 0 & B_2 \\ 0 & 0 & \cdots & M_s & 0 & B_s \\ 0 & 0 & \cdots & 0 & 0 & B_{s+1} \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}. \tag{1.6}$$

In (1.6), $s \ge 0$, $r_i > 0$ *for* $i = 1, \cdots s$, $r_{s+1} \ge 0$, *and* $r_{s+2} \ge 0$. *The matrices* M_i *are* $(r_i \times r_i)$ *rank one matrices of the form* $M_i = \sigma_i m_i(1, \cdots, 1)$ *where* $\sigma_i > 0$ *and* m_i *is an* $(r_i \times 1)$ *vector with positive entries satisfying* $(1, \cdots, 1)m_i = 1$. *The matrices* B_i, $i = 1, \cdots, s$, *satisfy* $B_i = m_i b_i^t$ *where* b_i *is* $r_{s+2} \times 1$ *and has nonnegative entries.* B_{s+1} *is an arbitrary* $(r_{s+1} \times r_{s+2})$ *matrix with nonnegative entries. We assume, without further loss of generality, that the row and column permutation leading to* (1.6) *has been chosen so that each of the last* r_{s+2} *columns of* M' *contains a positive entry.*

(v) *The following three conditions are equivalent*:

(α) *M is diagonalizable* (*i.e.*, *RMR^{-1} is diagonal for some nonsingular matrix R*).

(β) $r_{s+1} = 0$ *or* $B_{s+1} = 0$.

(γ) $m_{ii} = 0 \Rightarrow m_{ij} = 0, 1 \le j \le p$.

(vi) *A necessary condition on γ is*

$$M\gamma = D\gamma. \tag{1.7}$$

When the coordinates are permuted as in (1.6) *it follows from* (1.7) *that*

$$\gamma^t = (\beta_1 m_1^t, \cdots, \beta_s m_s^t, y^t, \mathbf{0}^t) \tag{1.8}$$

where $\beta_i \ge 0$, $i = 1, \cdots, s$, y is a $(r_{s+1} \times 1)$ *vector of nonnegative entries, and $\mathbf{0}^t$ is here the* $(1 \times r_{s+2})$ *zero vector.*

REMARK 1. The ith row of M being zero means the ith component of δ is $e_i^t\gamma$, a constant. The jth column being zero means e_j^tX is not used by δ in calculation of the estimate. Here is a peculiar consequence of Theorem 1. For $p = 2$ the matrix $M_1 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ corresponds to a limit of Bayes estimators while $M_2 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ does not. In fact, it follows from Theorem 2 that the estimator $\delta(x) = M_1 x$ is admissible for the Poisson problem. (Of course it is hard to see when an estimator like $M_1 x$ would be useful, which only reemphasizes the commonplace fact that admissibility alone does not guarantee the estimator is desirable).

THEOREM 2. *Let the components of X be independent Poisson variables. Assume that $1, \cdots, p$ has been permuted so that the estimator $\delta(x) = Mx + \gamma$ has matrix M of the block form* (1.6), *and that γ satisfies* (1.8). *In addition assume* (*as can be done by further permutation*) *that $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_s \ge 0$. A necessary and sufficient condition that δ be admissible for Case* 1 (*Case* 2, *respectively*) *is that*

$$1 \ge \sigma_1 \quad \text{and} \quad 1 > \sigma_3 \quad (1 > \sigma_2 \text{ respectively}), \tag{1.9}$$

and

$$\text{if} \quad \sigma_i = 1 \quad \text{then} \quad \beta_i = 0 \quad \text{and} \quad B_i = 0. \tag{1.10}$$

Section 2 consists of a proof of Theorem 1. Section 3 consists of a proof of Theorem 2.

REMARK 2. The results of Theorems 1 and 2 for Case 1 are also valid if the loss function is a positive definite quadratic form in $(\delta - \lambda)$, that is $(\delta - \lambda)^t \cdot Q(\delta - \lambda)$ where Q is symmetric positive definite. This can easily be checked from the proofs in Sections 2 and 3; and is also clear from the theorems themselves and general decision theoretic results noted in Bhattacharya (1966).

We now describe the application of Theorems 1 and 2 to the general mixture problem described in Example 1b, and to linear models in general.

Consider a linear model of the form

$$\lambda = H\rho + \rho_0 \tag{1.11}$$

$\rho \in R \subset \mathbb{R}^m$, where $H(p \times m)$ has rank m and int. $R \neq \varnothing$. This is a mild generalization of (1.4). Consider the Case 1 loss function for estimating λ. To avoid special cases we discuss only linear estimators $\delta(x) = Mx + \gamma$ in which the column space of $M(C(M))$ satisfies $C(M) = C(H)$. (If $C(M) \not\subset C(H)$ then the estimator can be improved for λ satisfying (1.11) by projecting it on $C(H) + \rho_0$. When $C(M) \subset_{\neq} C(H)$ the admissibility properties of δ follow from its properties for estimating λ subject to a submodal of (1.11) of the form $\lambda \doteq H^* \rho^* + \rho_0$ where $C(M) = C(H^*)$). We further limit consideration to the case where M is diagonalizable, for this class includes all the intuitively appealing linear estimators including the generalized least squares estimators.

It is then clear from Theorem 1 that δ can be a limit of Bayes estimators (and, hence, potentially admissible) only if $C(H) = C(M)$ is spanned by vectors $v_1, \cdots, v_m$ having nonnegative coordinates and satisfying

$$v_i^t v_j = 0, \quad i \neq j, \tag{1.12}$$

$\rho_0 \in C(H)$, and

$$R = \{\rho\colon \lambda \geq 0\} \tag{1.13}$$

which is a natural choice for R.

Under the above conditions, there is in fact a generalized least squares estimator which is a limit of Bayes estimators. Define V to be the diagonal matrix with entries $v_{ii} = \sum_{j=1}^m e_j^t v_i$. (Note that at most one of $e_j^t v_i$, $j = 1, \cdots, m$ is nonzero). Then the generalized least squares estimator appropriate for the covariance matrix V will be a limit of Bayes estimators. (This estimator can alternately be described as the projection on $C(H)$ in the (pseudo) inner product $(y, z) = y^t V^- z$. Note that the choice of the set $\{v_1, \cdots, v_m\}$ above is uniquely determined only up to multiplicative constants. Hence V is not uniquely determined. However, because of the orthogonality restriction (1.12) all allowable choices of $\{v_1, \cdots, v_m\}$ lead to the same generalized least squares estimator.)

This estimator will be admissible if and only if $m \leq 2$.

The general mixture problem of Example 1b is of the form (1.11) with $\rho_0 = 0$, and $R = \{\rho\colon \rho \geq 0\}$. The necessary (and sufficient) conditions (1.12) and (1.13) will be satisfied only in the very special case where each column of H is proportional to one of the orthogonal basis vectors $\{v_1, \cdots, v_m\}$. (If the condition (1.12) is satisfied but the columns of H are not themselves orthogonal then (1.13) will be violated in the mixture problem.) A mixture problem of this type is trivial. Thus to find limiting Bayes and admissible estimators for nontrivial mixture problems, one must look outside the class of linear estimators.

The same conclusions can be shown to also hold if the loss function is squared error for estimating the rate vector ρ in the mixture problem. For the proof combine Remark 2 with appropriate simple projection arguments.

La Motte (1982) describes in general terms the class of linear estimators which are admissible within the class of linear estimators. For estimation of a Poisson mean vector, many of these estimators are not admissible nor can they be limits of sequences of Bayes estimators since (1.2) is not satisfied. For example, in $p = 2$ dimensions with $\gamma = 0$, the estimators with matrices

$$M = \begin{pmatrix} 1/2 & 1/4 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} 3/7 & 2/7 \\ 1/7 & 3/7 \end{pmatrix}$$

are uniquely determined Bayes estimators within the class of linear estimators, but are, by Theorem 1, inadmissible in the class of all estimators. To see this, put mass ½ at $\binom{1}{1}$ and $\binom{1}{0}$ in the first case and $\binom{1}{0}$ and $\binom{2}{0}$ in the second case. For $p \geq 3$ dimensions, the linear estimator with $M = I_p$ is admissible within the class of linear estimators but as shown by Peng (1975) is an inadmissible estimator.

2. Proof of Theorem 1. Let $q_0 = 0$ and $q_k = \sum_{j=1}^{k} r_j$, $k = 1, 2, \cdots, s + 2$.

(i) Since $\delta(x) = Mx + \gamma$ is a pointwise limit of Bayes estimators and since the components of Bayes estimators are nonnegative, it follows that $e_i^t(Mx + \gamma) \geq 0$ for all i. Take $x = 0$ and find that the entries of γ are nonnegative. Take $x = ne_j$ and let $n \to \infty$, n an integer, to obtain $m_{ij} = \lim_{n\to\infty} n^{-1} e_i(Mne_j + \gamma) \geq 0$.

(ii) In the two cases the product rules (1.2) and (1.5) say, after substitution and cancellation, that, with $m_{ij} = e_i^t M e_j$,

$$\text{(2.1)} \qquad \begin{aligned} &\text{Case 1,} \quad e_i^t(Mx + \gamma)m_{ji} = m_{ij}e_j^t(Mx + \gamma) \\ &\text{Case 2,} \quad e_i^t(Mx + \gamma)m_{ji} + m_{ii}m_{ji} = e_j^t(Mx + \gamma) + m_{jj}m_{ij}. \end{aligned}$$

Substitute nx for x, n an integer, and let n go to infinity, to obtain in both cases that

$$\text{(2.2)} \qquad m_{ji}(e_i^t Mx) = m_{ij} e_j^t Mx.$$

Take $x = e_k$. Then (2.2) says that $m_{ji}m_{ik} = m_{ij}m_{jk}$. Sum on i to obtain

$$\text{(2.3)} \qquad M^2 = DM$$

where D is a diagonal matrix with jj entry the sum $\sum_{i=1}^{p} m_{ij}$ the jth column sum.

(iii) In particular if M is nonsingular, cancellation in (2.3) shows $M = D$, a diagonal matrix, which in the nonsingular case has nonzero diagonal entries, i.e., column sums.

(iv) From the substitutions $x = e_k$ and $x = e_i$ in (2.2), obtain

$$\text{(2.4)} \qquad m_{ji}m_{ik} = m_{ij}m_{jk} \quad \text{and} \quad m_{ii}m_{ji} = m_{ji}m_{ij}.$$

Then if $m_{ji} \neq 0$ it follows that $m_{ii} = m_{ij}$ and substitution into the first part of (2.4) yields

$$\text{(2.5)} \qquad \text{if} \quad m_{ji} \neq 0 \quad \text{then} \quad m_{ji}m_{ik} = m_{ii}m_{jk}, \quad \text{all } k.$$

This relationship says that if $m_{ji} \neq 0$ then either $m_{ii} = 0$ and the entire ith row of M is zero or $m_{ii} > 0$ and the ith and jth rows are proportional.

Permute the rows and columns of M by the same permutation, P, so that the matrix $M' = PMP^t$ has diagonal elements $m'_{ii} > 0$ for $1 \le i \le r$ and $m'_{ii} = 0$ for $r + 1 \le i \le p$. In this form we write

$$M' = \begin{pmatrix} A_1 & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{2.6}$$

where A_1 is $r \times r$. It will soon be clear that $r = q_s$. Similarly write

$$PDP^t = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}. \tag{2.7}$$

Apply (2.4) to M'. Sum over k to get

$$m'_{ji}c_i = m'_{ij}c_j \quad \text{with} \quad c_i = \textstyle\sum_k m'_{ik}. \tag{2.8}$$

Suppose $1 \le i \le r$, $r + 1 \le j \le p$ and $m'_{ji} > 0$. Then $c_i > 0$ since $m'_{ii} > 0$, yields $m'_{ij} > 0$. (2.5) applied to M' then yields $m'_{ji} = 0$ (a contradiction) since $m'_{jj} = 0$. Thus

$$B_{21} = 0. \tag{2.9}$$

We may further suppose P so chosen that rows $1, \cdots, r_1$ of A_1 are proportional rows $r_1 + 1, \cdots, r_1 + r_2$ are proportional, etc. Since $m_{ji} > 0$ implies rows i and j are proportional, $m_{ji} > 0$ implies that the indices i, j are in the same equivalence set, equivalence defined by nonzero proportionality of rows. Thus A_1 is in block form. We write

$$A_1 = \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & & 0 \\ 0 & 0 & \cdots & M_s \end{pmatrix} \tag{2.10}$$

in which M_i is a $r_i \times r_i$ matrix. Since the diagonal entries of M_i are nonzero, proportionality implies all entries of M_i are nonzero. Then it follows from (2.4) that all entries of a row of M_i are equal, thus

$$M_i = \sigma_i m_i(1, \cdots, 1) \tag{2.11}$$

where m_i is a $r_i \times 1$ vector with positive entries. Without loss of generality assume $(1, \cdots, 1)m_i = 1$.

Thus P can further be chosen so that

$$M' = \begin{bmatrix} M_1 & 0 & \cdots & 0 & 0 & B_1 \\ 0 & M_2 & \cdots & 0 & 0 & B_2 \\ 0 & 0 & \cdots & M_s & 0 & B_s \\ 0 & 0 & \cdots & 0 & 0 & B_{s+1} \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \tag{2.12}$$

where the last r_{s+2} rows of M' are $\mathbf{0}^t$ because of (2.5) since each of the last r_{s+2} columns of M' contains a nonzero element and $m'_{jj} = 0$ for $q_{s+1} + 1 \le j \le p$.

Expression (2.12) has the desired form, (1.6), except that it remains to establish the form of the matrices B_i. Use (2.7) applied to the problem after the permutation of coordinates described by P. Note that for $1 \le i \le s$ the column sum of M_i is $\sigma_i = \sigma_i(1, \cdots, 1)m_i > 0$. It follows that $M_i^2 = D_i M_i$ and

$$M_i B_i = D_i B_i, \quad 1 \le i \le s, \tag{2.13}$$

where $D_i = \sigma_i I_{r_i}$ with I_{r_i} the $r_i \times r_i$ identity matrix. Hence $\sigma_i m_i(1, \cdots, 1)B_i = \sigma_i B_i$, which shows that

$$B_i = m_i b_i^t, \quad 1 \le i \le s \tag{2.14}$$

for some $r_{s+2} \times 1$ vector b_i.

(v) Let $x = (x_{(1)}^t \cdots x_{(s)}^t x_{(s+1)}^t x_{(s+2)}^t)^t$ denote a $(p \times 1)$ vector partitioned to correspond to the matrix M'. M' is of rank s. It has nonzero eigenvalues σ_i, $1 \le i \le s$, and corresponding right eigenvectors formed by setting $x_{(i)} = m_i$ and $x_{(j)} = 0$, $j \neq i$. Suppose $B_{s+1} = 0$ (or $r_{s+1} = 0$), which is condition (β) in Theorem 1. Then $0 = M'x$ when

$$M_i x_{(i)} + B_i x_{(s+2)} = 0, \quad i = 1, \cdots, s. \tag{2.15}$$

This is a system of s independent equations in p unknowns (the entries of x). Therefore it has $p - s$ independent solutions. Consequently M' has $p - s$ independent eigenvectors for 0 in addition to the s eigenvectors for σ_j, $1 \le i \le s$. This proves M' is diagonalizable. Conversely, suppose $B_{s+1} \neq 0$. Then $0 = M'x$ only if (2.15) is satisfied and also

$$B_{s+1} x_{(s+1)} = 0. \tag{2.16}$$

Altogether this is a system of at least $s + 1$ independent equations and hence can have at most $p - s - 1$ independent solutions. It follows that M' does not possess a complete set of eigenvectors and hence cannot be diagonalizable. We have thus proved that (α) is equivalent to (β). It is easy to see that (β) is equivalent to (γ).

(vi) Substitute $x = 0$ in (2.1) and, in Case 2, use $m_{ji} m_{ii} = m_{ij} m_{ji}$ from (2.5) to get

$$(e_i^t \gamma) m_{ji} = (e_j^t \gamma) m_{ij}. \tag{2.17}$$

To prove (1.7) sum on the index j and reexpress the sum in matrix form. Now permute the coordinates and consider the problem with $M = M'$. With γ partitioned in the same manner as M', (1.7) yields

$$\sigma_i m_i (1^t \gamma_i) = \sigma_i \gamma_i, \tag{2.18}$$

and

$$0 = B_i \gamma_{s+2}, \quad i = 1, \cdots, s + 1. \tag{2.19}$$

It follows from (2.18) that $\gamma_i = \beta_i m_i$, $\beta_i \ge 0$, and from (2.19) that $\gamma_{s+2} = 0$ since for every column there is a B_i, $1 \le i \le s + 1$, with a nonzero entry in that column. These facts combined yield (1.8).

3. Proof of Theorem 2. The proofs for the Case 1 and Case 2 loss functions are parallel, with a number of small differences. We give the proof in full for

Case 1, and then sketch the modifications needed for Case 2. The next two paragraphs outline the argument for Case 1 and are followed by the detailed proof for that case.

The proof of admissibility and inadmissibility as stated requires examination of a number of cases. The argument to be presented below first establishes admissibility. In the case where $r_{s+1} = r_{s+2} = 0$ this is done by reducing the problem to looking at parameter vectors

$$\lambda_\theta^t = (\theta_1 m_1^t, \cdots, \theta_s m_s^t). \tag{3.1}$$

For parameters λ_θ the s-dimensional statistic $Z_1, \cdots, Z_s$ where

$$Z_i = \textstyle\sum_{j=q_{i-1}+1}^{q_i} (e_j^t x), \quad q_0 = 0, \tag{3.2}$$

is a sufficient statistic. In each case proof of admissibility reduces to multivariate statements about $(\sigma_1 Z_1 + \beta_1, \cdots, \sigma_s Z_s + \beta_s)$. It should be remembered in one dimension, if $0 \le \sigma < 1$ and $\beta > 0$ then $\sigma Z + \beta$ is the unique Bayes estimator resulting from a Gamma type prior. The case $r_{s+1} > 0$ and $r_{s+2} = 0$ corresponds to the multivariate situation of two estimators, one Bayes, one admissible, being joined together. These combinations are known to be admissible. The case $r_{s+2} > 0$ is reduced by a backward induction to the previous cases.

The admissibility arguments then follow. It is assumed that $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_s$. Except for the case $\sigma_1 = \sigma_2 = \sigma_3 = 1$ where appeal is made to results of Peng (1975), explicit construction of counterexamples is given.

Here is the proof of admissibility. Assume first that $r_{s+1} = r_{s+2} = 0$. Let δ be as in the theorem and suppose $R(\lambda, \hat\delta) \le R(\lambda, \delta)$ for all λ. Let Π be the projection onto the subspace of vectors $S^t = \{(a_1 m_1^t, \cdots, a_s m_s^t): a_i \in \mathbb{R}\}$. Then if $\lambda \in S$ $\| \Pi\delta - \lambda \| = \| \Pi(\delta - \lambda) \| \le \| \delta - \lambda \|$ with strict inequality if $\Pi\delta \ne \delta$. Note that $\lambda_\theta \in S$. We have

$$R(\lambda_\theta, \Pi\hat\delta) \le R(\lambda_\theta, \hat\delta) \le R(\lambda_\theta, \delta) \tag{3.3}$$

with strict inequality for some θ if $\Pi\hat\delta \ne \hat\delta$. Write $(\Pi\hat\delta)^t = (\hat d_1(x) m_1^t, \cdots, \hat d_s(x) m_s^t)$. By sufficiency the $\hat d_i$ may be assumed in the sequel to be functions of $Z_1, \cdots, Z_s$. Then

$$R(\lambda_\theta, \Pi\hat\delta) = \textstyle\sum_{i=1}^s \| m_i \|^2 E(\hat d_i(X) - \theta_i)^2 \tag{3.4}$$

and

$$R(\lambda_\theta, \delta) = \textstyle\sum_{i=1}^s \| m_i \|^2 E(d_i(X) - \theta_i)^2 \tag{3.5}$$

where $d_i(X) = \sigma_i Z_i + \beta_i$.

The random variables $Z_1, \cdots, Z_s$ are independent Poisson $\theta_1, \cdots, \theta_s$ random variables. For this problem with risk (3.4) the estimator in (3.5) is an admissible estimator of $\theta_1, \cdots, \theta_s$ if (and only if) $\sigma_3 > 1$ and $\beta_i = 0$ if $\sigma_i = 1$. This was shown in part by Peng (1975) and follows easily using the methods of Brown and Hwang (1982). Furthermore they show (3.3) implies $\Pi\hat\delta = \delta$, so that $\hat\delta = \delta$ and therefore δ is admissible.

Now consider the case $r_{s+1} > 0$ and $r_{s+2} = 0$. In this case the constant estimator, γ_{s+1} for the r_{s+1}-dimensional problem is Bayes and admissible while $Mx + \gamma$

restricted to the $q_s = \sum_{i=1}^{s} r_i$-dimensional problem was shown above to be admissible. Therefore $Mx + \gamma$ is admissible for the $q_{s+1} = q_s + r_{s+1}$ dimensional problem. (This is true even when $\gamma_{s+1} = 0$, in spite of the fact that this estimator is Bayes but not unique Bayes in the r_{s+1}-dimensional problem.)

Since the loss function is strictly convex if the estimator $\hat{\delta}$ differs from $\delta = Mx + \gamma$ at even a single lattice point and is as good as δ then $(\delta + \hat{\delta})/2$ has strictly better risk at all parameter values satisfying $\lambda_i > 0$, $i = 1, \cdots, p$. Thus if $\hat{\delta}$ is as good as δ (and $r_{s+2} = 0$) then $\hat{\delta} = \delta$, since δ is admissible.

We now proceed by induction on r_{s+2} to show that $\hat{\delta}$ as good as δ implies $\hat{\delta} = \delta$. Suppose $r_{s+2} = R$ and this hypothesis is true for $r_{s+2} \le R - 1$.

Assume $\hat{\delta}$ is as good as $\delta = MX + \gamma$. The last components of these vectors are $e_p^t\hat{\delta}$ and $e_p^t\delta = 0$. Let A be the least integer such that there exists x with $e_p^t x = A$ and $\hat{\delta}(x) \ne \delta(x)$. After cancellation of terms from both risk functions and division by $(e_p^t\lambda)^A = \xi^A$ let $\xi \to 0$. The result is

$$\begin{aligned}\sum_{x \ni e_p^t x = A} \|\, \delta(x) - \lambda \,\|^2 c(\lambda)h(x)\lambda^{(x - Ae_p)} \\ \ge \sum_{x \ni e_p^t x = A} \|\, \hat{\delta}(x) - \lambda \,\|^2 c(\lambda)h(x)\lambda^{(x - Ae_p)}.\end{aligned} \tag{3.6}$$

By construction $e_p^t\delta(x) = 0$ and $e_p^t\lambda = 0$. Let M^* denote the $(p-1) \times (p-1)$ matrix constructed by deleting the last row and column of M. Define $\lambda^* \in \mathbb{R}^{p-1}$ by $e_j^t\lambda^* = e_j^t\lambda + m_{jp}A$, $1 \le j \le p - 1$. Note that the value of r_{s+2} corresponding to M^* is $r_{s+2} = R - 1$. The expression on the left of (3.6) is the risk of the $p - 1$ dimensional estimator $\delta(x) = M^*x + \lambda^*$ with $x_i = x_i$, $i = 1, \cdots, p - 1$. The right side of (3.6) is

$$\begin{aligned}\sum_{x \ni e_p^t x = A} \sum_{j=1}^{p-1} (e_j^t(\hat{\delta}(x) - \lambda))^2 c(\lambda)h(x)\lambda^{(x - Ae_p)} \\ + \sum_{x \ni e_p^t x = A} (e_p^t\hat{\delta}(x))^2 c(\lambda)h(x)\lambda^{(x - Ae_p)}.\end{aligned} \tag{3.7}$$

By the induction hypothesis δ is uniquely admissible for the $p - 1$ dimensional problem. This implies $e_p^t\hat{\delta}(x) = 0$ for all x such that $e_p^t x = A$ and that $e_j^t\hat{\delta}(x) = e_j^t\delta(x)$, $1 \le j \le p - 1$, $e_p^t x = A$. This contradicts the choice of A, and completes the induction argument. We have thus proved the admissibility assertions of the theorem.

We now consider the asserted cases of inadmissibility. Suppose $\sigma_i \ge 1$ and either $B_i \ne 0$ or $\beta_i > 0$. For simplicity suppose $i = 1$. To avoid extra subscripts below let $m = m_1$, $\sigma = \sigma_1$, $Z = Z_1$, $\theta = \theta_1 = E(Z_1)$, $\beta = \beta_1 + \beta_1 X_{(S+2)}$ which is random if $B_1 \ne 0$, and λ^* be the $r_1 \times 1$ vector of the first r_1 parameter values. The first r_1 coordinates of the estimator will be modified keeping the remainder unchanged. In making risk comparison only the first r_1 terms of the sum in the risk function need be examined. We omit the other terms in the following. The relevant parts of the estimator are

$$e_j^t\delta(x) = e_j^t((\sigma Z + \beta)m - \varepsilon w), \quad 1 \le j \le r, \quad \varepsilon > 0, \tag{3.8}$$

with w a $r_1 \times 1$ vector to be specified. Consider a family of norms of vectors w defined by

$$\|\, w \,\|_\alpha^2 = \sum_{j=1}^{r_1} (e_j^t\lambda^*)^{-\alpha}(e_j^t w)^2,$$

so that in Case 1, $\alpha = 0$, and in Case 2, $\alpha = 1$. Think of the weighted sum of squares of the first r_1 coordinates of δ as a norm, and obtain for the relevant part of the risk of (3.8)

$$\begin{aligned} &E\| (\sigma Z + \beta)m - \varepsilon w - \lambda^* \|_\alpha^2 \\ &\quad = \sigma^2\theta \| m \|_\alpha^2 + E\| ((\sigma - 1)\theta + \beta)m + (\theta m - \lambda^*) - \varepsilon w \|_\alpha^2. \end{aligned} \tag{3.9}$$

In Case 1 with the choice $w = \mathbf{1}$ use the observation that $(w, m)_0 = 1$ and $(w, \theta m - \lambda^*)_0 = 0$. In Case 2 with the choice $w = m$, $(\lambda^*, w)_1 = 1$. Thus the Case 1 risk (3.9) becomes

$$\sigma^2\theta \| m \|_0^2 + E \| (\sigma\theta + \beta) - \lambda^* \|_0^2 - 2\varepsilon\, E((\sigma - 1)\theta + \beta) + \varepsilon^2 \| w \|_0^2. \tag{3.10}$$

The Case 2 risk (3.9) becomes

$$\sigma^2\theta \| m \|_1^2 + E \| (\sigma\theta + \beta) - \lambda^* \|_1^2 - (2\varepsilon\, E(\sigma\theta + \beta) - \varepsilon^2) \| m \|_1^2 + 2\varepsilon. \tag{3.11}$$

In Case 1, since $\sigma \geq 1$ and $E\beta > 0$, the sum of the last terms is negative for $\varepsilon > 0$ and small. In Case 2,

$$1 = \textstyle\sum_{j=1}^{r} (e_j^t m) \leq \| m \|_1 (\sum_{j=1}^{n} e_j^t \lambda)^{1/2} = \| m \|_1 \theta^{1/2} \tag{3.12}$$

so that $\theta^{-1} \leq \| m \|_1^2$. Thus in Case 2, for $\varepsilon > 0$ near zero, the last two terms of (3.11) are bounded by

$$(2\varepsilon\, E(\sigma\theta + \beta) - \varepsilon^2) \| m \|_1^2 - 2\varepsilon \geq (2\varepsilon(\sigma - 1)\theta + E\beta - \varepsilon^2)\theta^{-1} > 0. \tag{3.13}$$

In both cases a strict improvement is obtained.

In the case $\sigma > 1$ and $B_1 = 0$, $\beta = 0$, the first r_1 coordinates of $(\sigma m - \varepsilon w)Z$ contribute to the risk

$$\| \sigma m - \varepsilon w \|_\alpha^2 \theta + \| (\sigma m - \varepsilon w)\theta - \lambda \|_\alpha^2. \tag{3.14}$$

In Case 1, using $\| \mathbf{1} \|_0^2 = r_1$ and $(m, \mathbf{1})_0 = 1$, with $w = \mathbf{1}$, the difference of the risk functions for the cases $\varepsilon = 0$ and $\varepsilon > 0$ is

$$(2\varepsilon\sigma - r_1\varepsilon^2)\theta + (2\varepsilon(\sigma - 1) - \varepsilon^2 r_1)\theta^2 > 0 \tag{3.15}$$

for $\varepsilon < 2(\sigma - 1)/r_1$. In Case 2 with $w = m$ and $\varepsilon = 0$, the risk function is

$$(\sigma^2\theta + \sigma^2\theta^2) \| m \|_1^2 - 2\sigma\theta + \theta. \tag{3.16}$$

This risk is minimized as a function of σ by the choice

$$\sigma = 1/(1 + \theta) \| m \|_1^2 \leq \theta/(1 + \theta) \leq 1.$$

Thus in Case 2, if $\sigma > 1$, then strict improvement results.

Finally, suppose $\sigma_1 = \sigma_2 = \sigma_3 = 1$, $B_1 = B_2 = B_3 = 0$, and $\beta_1 = \beta_2 = \beta_3 = 0$. The estimator δ^* to be constructed differs from δ only on its first $q_3 = r_1 + r_2 + r_3$ coordinates and the difference of the risks depends only on these coordinates of X. Recall that Z_1, Z_2 and Z_3 are independent Poisson random variables with means θ_1, θ_2, θ_3. Peng (1975) has shown the existence of an estimator which dominates (Z_1, Z_2, Z_3) as an estimator of $(\theta_1, \theta_2, \theta_3)$. See also Ghosh, Hwang and Tsui (1983). Write the coordinates of this improved estimator as

$$(1 - \alpha_i(Z))Z_i. \tag{3.17}$$

Thus

$$
\begin{aligned}
0 &\le E_\theta(\textstyle\sum_{i=1}^{3} [(Z_i - \theta_i)^2 - ((1 - \alpha_i(Z))Z_i - \theta_i)^2]) \\
&= E_\theta(\textstyle\sum_{i=1}^{3} \alpha_i(Z)Z_i(2(Z_i - \theta_i) - \alpha_i(Z)Z_i)) \\
&= E_\theta(\textstyle\sum_{i=1}^{3} (Z_i^2(2\alpha_i(Z) - \alpha_i^2(Z)) - 2Z_i\theta_i\alpha_i(Z)))
\end{aligned} \tag{3.18}
$$

with strict inequality for some θ.

Define a modified estimator δ^* by $e_j^t\delta^* = ((e_j^t m_i) - \varepsilon\alpha_i(Z))Z_j$, $q_{i-1} + 1 \le j \le q_i$, $i = 1, 2, 3$, and otherwise $e_j^t\delta^* = e_j^t\delta$. Then

$$
\begin{aligned}
&R(\lambda, \delta) - R(\lambda, \delta^*) \\
&\quad = \varepsilon\, E_\theta(\textstyle\sum_{i=1}^{3} \alpha_i(Z)Z_i(2(Z_i - \theta_i) - r_i\varepsilon\alpha_i(Z)Z_i)) \\
&\quad = \varepsilon\, E_\theta(\textstyle\sum_{i=1}^{3} [Z_i^2(2\alpha_i(Z) - r_i\varepsilon\alpha_i^2(Z)) - 2\theta_i Z_i\alpha_i(Z)]).
\end{aligned} \tag{3.19}
$$

It follows from (3.15) that (3.16) is nonnegative and positive for some θ provided $r_j\varepsilon < 1$, $1 \le j \le 3$. That completes the proof for Case 1.

As previously remarked, the proof of Case 2 is parallel to the preceding proof. It begins with the case $r_{s+1} = r_{s+2} = 0$ by letting Π be the projection onto S in the inner product $\langle u, v\rangle = u^tC^{-1}v$ where C is the $(p \times p)$ diagonal matrix with diagonal elements $(m_1^t, \cdots, m_s^t)$. Note that if $\lambda = \lambda_\theta \in S$ and $(\Pi\hat{\delta})^t = (\hat{d}_1 m_1^t, \cdots, \hat{d}_s m_s^t)$ then

$$
\begin{aligned}
R(\lambda_\theta, \delta) \ge R(\lambda_\theta, \hat{\delta}) &= \textstyle\sum_{i=1}^{p} \lambda_i^{-1}(\hat{\delta}_i - \lambda_i)^2 \ge \sum_{i=1}^{p} \lambda_i^{-1}((\Pi\hat{\delta})_i - \lambda_i)^2 \\
&= \textstyle\sum_{i=1}^{s} (1/\theta_i) \sum_{j=1}^{r_i} (1/m_{ij})(\hat{d}_i m_{ij} - \theta_i m_{ij})^2 \\
&= \textstyle\sum_{i=1}^{s} (1/\theta_i)(\hat{d}_i - \theta_i)^2 = R(\lambda_\theta, \Pi\hat{\delta}).
\end{aligned} \tag{3.20}
$$

This expression is parallel to (3.3), (3.4). When $\sigma_2 < 1$ and $\beta_1 = 0$ if $\sigma_1 = 1$ the estimator with $d_i(Z) = \sigma_i Z + \beta$, $i = 1, \cdots, s$, is an admissible estimator of $\theta_1, \cdots, \theta_s$ under the loss in (3.20). It follows, as in Case 1, that δ is admissible under these conditions.

The argument for $r_{s+1} > 0$, $r_{s+2} = 0$ is as before.

Suppose $r_{s+2} \ge 1$. Proceed by induction as before. Suppose there is a least integer A such that $\hat{\delta}(x) \ne \delta(x)$ for some x with $e_p^t x = A$. Proceed as in (3.6) to get

$$
\begin{aligned}
&\textstyle\sum_{x\ni e_p^t x = A} [\sum_{i=1}^{p} \lambda_i^{-1}(\delta_i(x) - \lambda_i)^2]c(\lambda)h(x)\lambda^{(x-Ae_p)} \\
&\quad \ge \textstyle\sum_{x\ni e_p^t x = A} [\sum_{i=1}^{p} \lambda_i^{-1}(\hat{\delta}_i(x) - \lambda_i)^2]c(\lambda)h(x)\lambda^{(x-Ae_p)}.
\end{aligned} \tag{3.21}
$$

Now let $\lambda_k \to 0$, keeping the other coordinates of λ as fixed, positive numbers. The left side of (3.21) stays bounded as the right side approaches ∞; a contradiction, unless $\hat{\delta}_p(x) = \delta_p(x) = 0$ for all $x \ni e_p^t x = A$. Thus $\hat{\delta}_p(x) = \delta_p(x) = 0$ for all x. It now follows from the uniqueness assertion in the induction hypothesis that $\hat{\delta}(x) = \delta(x)$ for all x.

The first two parts of the inadmissibility proof are exactly as in Case 1.

In the third part of the inadmissibility proof $\sigma_1 = \sigma_2 = 1$, $B_1 = B_2 = 0$ and $\beta_1 = \beta_2 = 0$. Now, when Z_1, Z_2 are independent Poisson random variables with means θ_1, θ_2, Clevenson and Zidek (1975) have shown the existence of an estimator $((1 - \alpha_1(Z))Z_1, (1 - \alpha_2(Z))Z_2)$ which dominates $\delta(Z) = Z$ under Case 2 loss. Let

$$\delta_j^* = (1 - \alpha_i(Z))e_j^t m_i Z_i, \quad q_{i-1} + 1 \leq j \leq q_i, \quad i = 1, 2.$$

Otherwise $\delta_j^* = \delta_j$. (Note that $e_j^t m_i$ occurs in a different place here than in the estimator used in (3.19) for Case 1, and $\varepsilon = 1$.) Then,

$$\begin{aligned} &R(\lambda, \delta) - R(\lambda, \delta^*) \\ &\geq E_\theta(\textstyle\sum_{i=1}^2 \theta_i^{-1}[Z_i^2(2\alpha_i(Z) - \alpha_i^2(Z)) - 2\theta_i Z_i \alpha_i(Z)]) \qquad (3.22) \\ &\geq 0 \end{aligned}$$

with strict inequality for some θ, since the middle expression in (3.22) is the difference in risks for the problem of Clevenson and Zidek. This completes the proof.

REFERENCES

BHATTACHARYA, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. *Ann. Math. Statist.* **37** 1819–1824.

BROWN, L. D. and FARRELL, R. H. (1985). Complete class theorems for estimation of multivariate Poisson means and related problems. To appear in *Ann. Statist.* in June.

BROWN, L. D. and HWANG, J. I. (1982). A unified admissibility proof. *Statistical Decision Theory and Related Topics III*, Vol. 1. *Academic Press*, New York.

CLEVENSON, M. and ZIDEK, J. (1975). Simultaneous estimation of the mean 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.

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

JOHNSON, N. L. (1957). Uniqueness of a result in the theory of accident proneness. *Biometrika* **44** 530–531.

JOHNSTONE, I. M. (1981). Admissible estimation of Poisson means, birth and death processes, and discrete Dirichlet problems. Ph.D. thesis, Cornell University.

LA MOTTE, L. R. (1982). Admissibility in linear estimation. *Ann. Statist.* **10** 245–255.

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

SHEPP, L. A. and VARDI, Y. (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Medical Imaging.* Vol. M1, 113–122.

VARDI, Y., SHEPP, L. A., and KAUFMAN, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.*, to appear.

HEBREW UNIVERSITY
MT. SCOPUS CAMPUS
91905 JERUSALEM
ISRAEL

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853