Complete Class Theorems for Estimation of Multivariate Poisson Means and Related Problems
Author(s): L. D. Brown and R. H. Farrell
Source: *The Annals of Statistics*, Vol. 13, No. 2 (Jun., 1985), pp. 706-726
Published by: Institute of Mathematical Statistics
Stable URL: http://www.jstor.org/stable/2241205
Accessed: 25/03/2010 15:27

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=ims.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

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

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

http://www.jstor.org

The Annals of Statistics
1985, Vol. 13, No. 2, 706–726

COMPLETE CLASS THEOREMS FOR ESTIMATION OF MULTIVARIATE POISSON MEANS AND RELATED PROBLEMS[1]

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

Cornell University

Basic decision theory for discrete random variables of the multivariate geometric (power series) type is developed. Some properties of Bayes estimators that carry over in the limit to admissible estimators are obtained. A stepwise generalized Bayes representation of admissible estimators is developed with estimation of the mean of a multivariate Poisson random variable in mind. The development carries over to estimation of the mean of a multivariate negative Binomial random variable. Due to the natural boundary of the parameter space there is an interesting pathology illustrated to some extent by the examples given. Examples include one to show that admissible estimators with somewhere infinite risk do exist in two or more dimensions.

1. Introduction. A complete class theorem for estimates of a one-dimensional Poisson parameter, λ, was worked out by both authors some years ago. One author viewed the result as decomposing the set of risk functions into certain faces which are convex and which correspond to the largest integer n such that the estimator $\delta(x) = 0$, for $0 \le x \le n$. Details are given in Section 5. This viewpoint does not easily generalize to higher dimensions. To the other author was suggested the possibility of a stepwise generalized Bayes representation somewhat analogous to that in Brown (1981) for problems with finite sample spaces. This type of representation does lend itself to generalization to higher dimensions. Section 4 contains a description of this stepwise Bayes representation for multidimensional parameters, which is formally described in Theorem 4.1. The representation in Section 4 contains two features not present in Brown (1981): at each step the estimators need not be (conditionally) Bayes but instead may be only (conditionally) generalized Bayes; and the sequence of subsets of the parameter space appearing in the stepwise representation is no longer necessarily finite, but may be well ordered as a transfinite sequence.

The results in this paper apply to geometric type discrete probabilities as defined in (2.1), of which the Poisson and negative Binomial probabilities are special cases. Except for examples in Section 7 and the discussion of the negative Binomial in Section 8, no further reference is made to particular parametric families.

Basic theory about convergence of sequences of Bayes estimators is briefly outlined in Section 2 and is developed more fully in Brown and Farrell (1983). Here it is shown that admissible estimators have continuous risk where the risk is finite and that admissible estimators are pointwise limits of sequences of

Received August 1983; revised December 1984.
[1] Research supported in part by National Science Foundation Grant D.M.S. 8200031.
AMS 1980 *subject classifications.* Primary 62C07; secondary 62F10.
Key words and phrases. Estimation, multivariate Poisson parameter, decision theory.

estimators which are Bayes relative to discrete measures supported on finite sets of points. Notations used throughout are established in Section 2.

Bayes estimators for the loss functions used have some monotonicity and convexity properties which carry over in the limit to the limit estimators. Also, an important property, called in this paper the product rule, carries over to the limit estimator. These results are developed in Section 3. They are applied in Brown and Farrell (1985) to resolve the question of which $p \times p$ matrices M give rise to estimators $Mx + \gamma$ which are limits of Bayes estimators. That paper also settles the question of which of these estimators are admissible.

The stepwise decomposition of the sample space, discussed in Section 4, uses the results of Sections 2 and 3. The results here are similar to Brown (1981) but the arguments are very different. The reader can get a better feel for the stepwise result if he or she tries to apply the theory of Section 4 to the examples of Section 7.

The complete class described in Section 4 includes many inadmissible estimators. Section 6 presents some very partial results concerning admissibility and dominance of estimators in the complete class. These results depend on a description of the set of zeros of the estimators in question. Some of the results are applied in the examples in Section 7.

A theorem showing that a broad subclass of admissible estimators of the mean of a multidimensional Poisson variable could be represented as generalized Bayes was used in Johnstone (1981, 1984, 1985). Johnstone's Proposition 8.1 appears as Corollary 6.3 of our paper.

2. Notations, preliminary decision theory. Let Q denote the integers and Q_+ the nonnegative integers. Let e_i denote the ith coordinate vector and the superscript t denote transpose. For λ, $\lambda_1 \in \mathbb{R}^p$ define the multivariate orderings $\lambda < \lambda_1$ ($\lambda \le \lambda$, respectively) to mean $e_i^t\lambda < e_i^t\lambda_1$ ($e_i^t\lambda \le e_i^t\lambda_1$), $i = 1, \cdots, p$. The sample space is $X = Q_+^p$. We write Λ for the parameter space. For the multivariate Poisson measures $\Lambda = (0, \infty)^p$ and for the multivariate negative binomial $\Lambda = (0, 1)^p$. The decision theory depends on the lattice of x values being unbounded but needs only that $\Lambda \subset (0, \infty)^p$ contains the intervals $\{\lambda \in (0, \infty)^p: \mathbf{0} < \lambda \le \lambda_1\}$ whenever $\lambda_1 \in \Lambda$. ($\mathbf{0}$ denotes the zero vector.) Define $\hat{\Lambda} \subset [0, \infty)^p$ by

$$\hat{\Lambda} = \{\lambda \in [0, \infty)^p: \exists \lambda_1 \in \Lambda, \lambda \le \lambda_1\},$$

and let $\bar{\Lambda}$ denote the closure of Λ (and of $\hat{\Lambda}$). In the multivariate Poisson case $\hat{\Lambda} = \bar{\Lambda}$, but not in the multivariate negative binomial case.

A nonrandomized estimator is a function $\delta: X \to \bar{\Lambda}$. Throughout we use only nonrandomized estimators except in Lemma 2.3 where randomized estimators are introduced in order to prove that the nonrandomized ones are a complete class. The discrete probability density functions are of the geometric form

$$p_\lambda(x) = c(\lambda)h(x)\lambda^{(x)}, \quad \lambda^{(x)} = \prod_{i=1}^{p} (e_i^t\lambda)^{(e_i^t x)}, \quad x \in X, \quad \lambda \in \hat{\Lambda}, \tag{2.1}$$

where $1 = \sum_{x \ge 0} c(\lambda)h(x)\lambda^{(x)}$. While not entirely necessary it is convenient to assume that $h(x) > 0$, $x \in X$, and we do so throughout. In estimation of the parameter vector λ the loss (L) and risk (R) are defined by

$$(2.2)\qquad \begin{aligned} L(\lambda, \delta(x)) &= \textstyle\sum_{i=1}^{p} (e_i^t\lambda)^{\alpha_i}(e_i^t(\delta(x) - \lambda))^2, \\ R(\lambda, \delta) &= \textstyle\sum_{x\geq 0} c(\lambda)h(x)\lambda^{(x)}L(\lambda, \delta(x)), \end{aligned}$$

where $\alpha_1, \cdots, \alpha_p$ are specified constants.

Note that p_λ and $R(\lambda, \cdot)$ are defined on $\hat{\Lambda}$ but admissibility is defined relative to the parameter space Λ—that is, δ is admissible (relative to Λ) if $R(\lambda, \delta') \leq R(\lambda, \delta)$ $\lambda \in \Lambda$, implies $R(\lambda, \delta') = R(\lambda, \delta)$, $\lambda \in \Lambda$. This is done only to avoid having to treat certain pathologies as special cases in the proof for the general characterization of Theorem 4.1. See Remark 2.2 and Example 7.5.

Some general theory is presented but applications are dependent on the specification of the parameters $\alpha_1, \cdots, \alpha_p$ in the loss function. In particular, the character of the stepwise reduction of the sample space is dependent on the choice of these parameters and in order to simplify the discussion specific constructions are limited to two special cases:

Case 1: $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$;
Case 2: $\alpha_1 = \alpha_2 = \cdots = \alpha_p = -1$.

The Case 1 loss function is used in Peng (1975) and in Ghosh, Hwang, Tsui (1983); the Case 2 loss is used by Clevenson and Zidek (1975), Johnstone (1981), and Tsui (1982). See Section 8 on the estimation of a negative binomial mean for a different loss function.

In general, the notation of (2.2) is simplified by writing

$$(2.3)\qquad R(\lambda, \delta) = \textstyle\sum_{i=1}^{p} \sum_{x\geq 0} c(\lambda)h(x)\lambda^{(x+\alpha_i e_i)}(e_i^t(\delta(x) - \lambda))^2.$$

In the sequel we will say a parameter set Λ is monotone (to the lower left) provided $\lambda \in \Lambda$ and $\mathbf{0} < \lambda' \leq \lambda$ implies $\lambda' \in \Lambda$. (The complementary concept of monotonicity (to the upper right) is defined and used in Section 4.) The finite risk set of an estimator is defined by $\Lambda^* = \Lambda^*(\delta) = \{\lambda \in \hat{\Lambda}: R(\lambda, \delta) < \infty\}$. Let $(\Lambda^*)^\circ$ denote the relative interior of Λ^* considered as a subset of $[0, \infty)^p$.

LEMMA 2.1. *Consider any procedure, δ. Then Λ^* is monotone. If $\lambda_0 \in (\Lambda^*)^\circ$ then $R(\cdot, \delta)$ is continuous at λ_0. If λ_0 is a boundary point of Λ^* then* $\lim_{\lambda\to\lambda_0, \lambda\leq\lambda_0} R(\lambda, \delta) = R(\lambda_0, \delta)$.

PROOF. For any δ, $R(\lambda, \delta) < \infty$ if and only if

$$(2.4)\qquad H(\lambda, \delta) = \textstyle\sum_{i=1}^{p} \sum_{x\geq 0} h(x)\lambda^{(x+\alpha_i e_i)}(e_i^t\delta(x))^2 < \infty$$

by Cauchy-Schwartz and the fact that $R(\lambda, \mathbf{0}) < \infty$, for all $\lambda \in \Lambda$. $H(\cdot, \delta)$ is monotone nondecreasing. Hence $\Lambda^* = \{\lambda: H(\lambda, \delta) < \infty\}$ is monotone. If $R(\lambda_0, \delta) = \infty$ then $\lim_{\lambda\to\lambda_0} R(\lambda, \delta) = R(\lambda_0, \delta) = \infty$ by Fatou's Lemma. The remaining continuity assertions follow from the dominated convergence theorem and (2.4). □

REMARK 2.2. If $R(\lambda, \delta) < \infty$ for all $\lambda \in \Lambda$ then $R(\cdot, \delta)$ is continuous on $\Lambda^* = \hat{\Lambda}$. It then follows that δ is admissible (relative to Λ) if and only if it is admissible relative to $\hat{\Lambda}$. This result holds more generally if the closure, $((\Lambda^*)^\circ)^-$, of $(\Lambda^*)^\circ$ satisfies

$$((\Lambda^*)^\circ)^- \supset \Lambda^*. \tag{2.5}$$

It can be shown that if δ is admissible then $(\Lambda^*)^\circ \neq \emptyset$. However it may be that $\hat{\Lambda} - \Lambda^* \neq \emptyset$. See Example 7.4. Furthermore, Example 7.5 shows that (2.5) is not always valid. When (2.5) fails it may be that δ is admissible relative to $\hat{\Lambda}$ but not admissible (relative to Λ). Estimators failing to satisfy (2.5) are of no apparent practical interest. For this reason, and for simplicity of presentation, our results are formulated only to characterize admissibility (relative to Λ).

LEMMA 2.3. *The nonrandomized estimators form a complete class.*

PROOF. Let $\rho(\cdot \mid x)$ denote a randomized estimator. If ρ is admissible then $R(\lambda_0, \rho) < \infty$ for some $\lambda_0 \in \Lambda$. (Otherwise $\mathbf{0}$ would be a better estimator.) Note that $p_{\lambda_0}(x) > 0$ for all $x \in X$. Jensen's inequality now yields that $\delta(x) = \int a\rho(da \mid x)$ is well defined and $R(\lambda, \delta) \leq R(\lambda, \rho)$, $\lambda \in \Lambda$, with strict inequality for λ_0 (in fact for all $\lambda \in \Lambda^* \cap \Lambda$) unless $\rho(\{\delta(x)\} \mid x) = 1$, $x \in X$,—that is, unless ρ is actually nonrandomized. □

The complete class results of Sections 4–6 are based on a standard decision theoretic result. For the sake of completeness we now state this result.

THEOREM 2.4. *Suppose δ is admissible. Then there is a sequence of Bayes procedures $\{\delta_n\}$ relative to priors having compact support in Λ such that*

$$\delta_n(x) \longrightarrow \delta(x) \quad \forall x \in X. \tag{2.6}$$

PROOF. (2.6) can be deduced from results in Wald (1949). A more elegant method would be to use the Stein-LeCam theorem (see Farrell (1968)) as in Brown (1971, Theorem 3.1). An elementary proof is given in Brown and Farrell (1983). □

3. The structure of the Bayes rules. If ν is a σ-finite Borel measure on $\bar{\Lambda}$, we will work with the integrals $\pi_\nu(x) = \pi(x)$ defined by

$$\pi(x) = \int c(\lambda)\lambda^{(x)}\, \nu(d\lambda). \tag{3.1}$$

Note that π is defined (but possibly infinite) for all $x \in Q^p_+$. In terms of the functions $\pi(x)$, the generalized Bayes estimator δ determined by ν is given (see (2.2) for the loss function) by the ratios

$$\pi(x + (\alpha_i + 1)e_i)/\pi(x + \alpha_i e_i) = e_i^t \delta(x). \tag{3.2}$$

If ν is a finite measure with compact support in Λ then $R^p = \{x\colon 0 < \pi(x) < \infty\}$, so that the corresponding Bayes estimator is unique and well defined by (3.2).

The product rule, so called, depends on integral values of the parameters $\alpha_1, \cdots, \alpha_p$. We state results for the two cases of special interest.

LEMMA 3.1. *Let δ be well defined by* (3.2). *If $\alpha_1 = \cdots = \alpha_p = 0$, then with $e_{i_0} = 0$, (Case* 1)

(3.3) $$\prod_{j=1}^{k} e_{i_j}^t \delta(x + e_{i_0} + \cdots + e_{i_{j-1}}) = \pi(x + e_{i_1} + \cdots + e_{i_k})/\pi(x).$$

If $\alpha_1 = \cdots = \alpha_p = -1$ (Case 2)

(3.4) $$\prod_{j=1}^{k} e_{i_j}^t \delta(x + e_{i_1} + \cdots + e_{i_j}) = \pi(x + e_{i_1} + \cdots + e_{i_k})/\pi(x).$$

PROOF. Immediate after substitution in (3.2). □

THEOREM 3.2. *Let δ be admissible and be a pointwise limit of the Bayes estimators δ_n relative to ν_n for which* (3.2) *is well defined. If*

$$\pi_n(x) = \int c(\lambda)\lambda^{(x)} \nu_n(d\lambda)$$

then in Case 1,

$$\lim_{n\to\infty} \pi_n(x + e_{i_1} + \cdots + e_{i_k})/\pi_n(x) = \prod_{j=1}^{k} e_{i_j}^t \delta(x + e_{i_0} + \cdots + e_{i_{j-1}}),$$

and in Case 2,

$$\lim_{n\to\infty} \pi_n(x + e_{i_1} + \cdots + e_{i_k})/\pi_n(x) = \prod_{j=1}^{k} e_{i_j}^t \delta(x + e_{i_1} + \cdots + e_{i_j}).$$

PROOF. Immediate from Lemma 3.1. □

This yields the product rule for admissible estimators:

COROLLARY 3.3. *Let δ be admissible. Let $e_{i_1} + \cdots + e_{i_k} = e_{j_1} + \cdots + e_{j_k}$. Then in Case* 1,

(3.5) $$\prod_{q=1}^{k} e_{i_q}^t \delta(x + e_{i_0} + \cdots + e_{i_{q-1}}) = \prod_{q=1}^{k} e_{j_q}^t \delta(x + e_{j_0} + \cdots + e_{j_{q-1}}),$$

and in Case 2,

(3.6) $$\prod_{q=1}^{k} e_{i_q}^t \delta(x + e_{i_1} + \cdots + e_{i_q}) = \prod_{q=1}^{k} e_{j_q}^t \delta(x + e_{j_1} + \cdots + e_{j_q}).$$

PROOF. Immediate from Theorem 2.4 and Theorem 3.2. □

4. Proof of the pointwise decomposition of the sample space. The basic idea to be followed here was introduced by Brown (1981). We seek to construct a decreasing well-ordered sequence of subsets $X_1 \supset X_2 \cdots$ indexed by ordinals such that $\cap_\beta X_\beta = \emptyset$. It can be seen that indexing by integers may not be enough, and that limit ordinals may occur; see Example 7.2. Each set X_β will have a monotonicity and a convexity property to be stated below.

At step $\beta + 1$ it is to be shown in Case 1 that the values of $\delta(x)$ of the admissible estimator δ for $x \in X_\beta - X_{\beta+1}$ are given by an expression somewhat like a generalized Bayes rule, i.e., as a ratio of two integrals. A similar representation also holds for Case 2. This is stated formally in Theorem 4.1.

Let $C \subset Q^p$. We say C is *monotone* (to the upper right) if $x \in C$ and $y \geq x$, $y \in Q^p$, implies $y \in C$. We say $C \subset Q^p$ is *convex* if $C = Q^p \cap$ (convex hull C). C is *bounded* (below) if there is an $x \in Q^p$ such that $y \in C$ implies $y \geq x$.

In the following, C denotes a convex, monotone, bounded subset of Q^p. It is shown in Lemma 4.2 that C possesses a finite set of minimal points. ($y \in C$ is minimal if $x \leq y$, $x \neq y$ implies $x \notin C$.) Denote this set by $\mathscr{M}(C)$. A family of finite nonnegative measures $\{\omega_y : y \in \mathscr{M}(C)\}$ on $\bar{\Lambda}$ is called nontrivial if $\omega_y \neq 0$ for some $y \in \mathscr{M}(C)$. It is called compatible if for any two points $y_1, y_2 \in \mathscr{M}(C)$, and $x \geq y_i$, $i = 1, 2$, the two measures in (4.1) are equal.

$$\lambda^{(x-y_1)}\omega_{y_1}(d\lambda) = \lambda^{(x-y_2)}\omega_{y_2}(d\lambda). \tag{4.1}$$

For a set of compatible measures on $\bar{\Lambda}$ indexed by $\mathscr{M}(C)$ let $C^+ = \{x \in C: \exists y \in \mathscr{M}(C), y \leq x, \omega_y \neq 0\}$, and let $C^0 = C - C^+$. If $x \in C^+$ define

$$\pi(x) = \int \lambda^{(x-y)}\,\omega_y(d\lambda) \quad \text{for} \quad y \leq x, \qquad \omega_y \neq 0. \tag{4.2}$$

If $x \in C^0$ define $\pi(x) = 0$. π is well defined because of (4.1). Examples of nontrivial compatible sets of measures other than the obvious $\omega_y(d\lambda) = \lambda^{(y)}\omega(d\lambda)$ for a measure ω on Λ are given in Section 7.

A compatible family $\{\omega_y : y \in \mathscr{M}(C)\}$ is said to satisfy *condition L* if there exists a sequence ν_n of finite measures with compact support in Λ such that

$$\omega_{ny}(d\lambda) = \lambda^{(y)}c(\lambda)\nu_n(d\lambda) \tag{4.3}$$

satisfies $\omega_{ny} \to \omega_y$ weakly, $y \in \mathscr{M}(C)$.

Example 7.6 displays a compatible family which does not satisfy condition L. Many related examples exist. However they all appear to be related to procedures which are rather artificial. Thus, the important feature of the stepwise description in the following main theorem is the appearance of compatible families and the resulting integral representations, (4.5) or (4.6). The further restriction that these families satisfy condition L does not further reduce the size of the complete class to an important extent.

THEOREM 4.1. *Suppose δ is admissible. Then there exist a well-ordered decreasing sequence, $\{X_\beta\}$, of convex, monotone, bounded subsets of Q^p, indexed by ordinals, and corresponding nontrivial, compatible families of measures satisfying condition L, $\{\omega_{\beta y} : y \in \mathscr{M}(X_\beta)\}$ on $\bar{\Lambda}$ with the following properties: $\cap_\beta X_\beta = \emptyset$, with X_β^0 as defined following* (4.1),

$$X_B = \cap_{\beta < B} X_\beta^0 (= X_{B-1}^0 \text{ if } B \text{ is not a limit ordinal}), \tag{4.4}$$

$\pi_\beta(x) < \infty$ for all $x \in X_\beta$, π_β defined by (4.2) *relative to the family $\{\omega_{\beta y} : y \in \mathscr{M}(x_\beta)\}$, and in*

Case 1 ($\alpha_1 = \cdots = \alpha_p = 0$): *$X_1 = X$ and*

$$e_i^t\delta(x) = \pi_\beta(x + e_i)/\pi_\beta(x) \quad x \in X_\beta^+ = X_\beta - X_{\beta+1}, \quad \textit{or in} \tag{4.5}$$

Case 2 $(\alpha_1 = \cdots = \alpha_p = -1)$: $X_1 = \{z\colon \exists i = 1, \cdots, p$ *and* $z + e_i \in X\}$ *and*

$$e_i^t\delta(x) = \pi_\beta(x)/\pi_\beta(x - e_i), \quad x - e_i \in X_\beta^+. \tag{4.6}$$

See Example 7.2 for a case where the well-ordered sequence $X_1, X_2, \cdots$, requires transfinite ordinals.

The following lemmas are used in the proof of Theorem 4.1.

LEMMA 4.2. *Let $C \subset Q^p$ be monotone and bounded below, $C \neq \emptyset$. Then $\mathscr{M}(C)$ is finite.*

PROOF. It suffices to consider the case where $C \subset Q_+^p$ and we do so below. The proof is by induction on the dimension p.

Case $p = 2$. We suppose $p = 2$ and C is nonempty. The hypotheses then imply that C is countably infinite. There exist points $(a_1, b_1)^t$ and $(a_2, b_2)^t$ in C such that $a_1 = \min\{e_1^t x\colon x \in C\}$ and $b_2 = \min\{e_2^t x\colon x \in C\}$. Consider (a_3, b_3). If $(a_3, b_3) \notin \{x : x^t \geq (a_1, b_1)\}$ then $b_3 < b_1$. If $(a_3, b_3) \notin \{x : x^t \geq (a_2, b_2)\}$ then $a_3 < a_2$. Thus

$$z \notin \{x : x^t \geq (a_1, b_1)\} \cup \{x : x^t \geq (a_2, b_2)\}$$

implies $z \leq (a_2, b_1)$. The set of such z is finite and contains $\mathscr{M}(C)$. Hence $\mathscr{M}(C)$ is finite.

Inductive step that truth for $p - 1$ implies truth for p. Pick $z \in C$. Then $x \notin \{x : x \geq z\}$ means $e_i^t x < e_i^t z$ for some $i = 1, \cdots, p$. Consider the hyperplanes $H_{i,c} = \{x : e_i^t x = c\}$ with c an integer satisfying $0 \leq c \leq e_i^t z$. There are only a finite number of such hyperplanes. The set $C \cap H_{i,c}$ is a convex monotone subset of $H_{i,c}$. Hence $\mathscr{M}(C \cap H_{i,c})$ is finite by the induction hypothesis. If $y \in \mathscr{M}(C)$ then either $y = z$ or $y \in H_{i,c}$ for some i, $0 \leq c \leq e_i^t z$. In the latter case $y \in \mathscr{M}(C \cap H_{i,c})$. Hence

$$\mathscr{M}(C) \subset \{z\} \cup (\cup_{i,c} \mathscr{M}(C \cap H_{i,c})).$$

This proves that $\mathscr{M}(C)$ is finite for the given p. □

LEMMA 4.3. *Let C be any set of lattice points, C_s a nonempty subset of C, and $\{\omega_y, y \in C_s\}$ a family of nonnegative measures on $\bar{\Lambda}$, and let π be determined as in* (4.2). *Then C^0 is also convex and monotone.*

PROOF. Let $y \in C_s$. For $x \geq y$ write $z = x - y$. Then

$$\pi(x) = \pi(y + z) = \int \lambda^{(z)}\, \omega_y(d\lambda)$$

Clearly $\{z\colon \pi(y + z) = 0\}$ is convex and monotone. Consequently $\{x\colon x \geq y, \int \lambda^{(x-y)}\, \omega_y(d\lambda) = 0\}$ also has these properties, as does C^0 since

$$C^0 = \cap_{y \in C_s} \left\{x\colon x \geq y, \int \lambda^{(x-y)}\, \omega_y(d\lambda) = 0\right\}. \quad \square$$

LEMMA 4.4. *Let C be convex, monotone, and bounded below. Let $\{\omega_{iy}: y \in \mathscr{M}(C)\}$, $i = 1, \cdots$ be a sequence of compatible families of measures on $\overline{\Lambda}$, with corresponding functions π_i, such that*

$$\sum_{y \in \mathscr{M}(C)} \pi_i(y) = 1. \tag{4.7}$$

and

$$\limsup_{i \to \infty} \pi_i(x) < \infty \quad x \in C. \tag{4.8}$$

Then there exists a subsequence $\{i'\} \subset \{i\}$ and a nontrivial compatible family $\{\omega_y: y \in \mathscr{M}(C)\}$, with corresponding function π, such that

$$\omega_{i'y} \to \omega_y \text{ (weakly)}, \quad y \in \mathscr{M}(C), \quad \text{and} \quad \pi_{i'}(x) \to \pi(x) < \infty, \quad x \in C. \tag{4.9}$$

PROOF. Fix $y \in \mathscr{M}(C)$. Let $m_{iy}(z) = \int \lambda^{(z)}\omega_{iy}(d\lambda) = \Pi_i(y + z)$. By (4.8) $\limsup_{i\to\infty} m_{iy}(z) < \infty$. Hence there is a subsequence $\{i'\}$ such that $\omega_{i'y}$ converges, say $\omega_{i'y} \to \omega_y$, and such that $m_{i'y}(z) \to m_y(z)$ where $m_y(z)$ corresponds to ω_y. The subsequence $\{i'\}$ can be chosen so that this convergence holds for each $y \in \mathscr{M}(C)$.

If $u, v \in \mathscr{M}(C)$ and $x \geq u$, $x \geq v$, $m_{i,u}(x - u) = m_{i,v}(x - v)$ since $\{\omega_{iy}\}$ is compatible. It follows that $m_u(x - u) = m_v(x - v)$ so that the limit family $\{\omega_y\}$ is compatible. It obviously then satisfies (4.9). It is nontrivial since $\sum_{y \in \mathscr{M}(C)} \pi(y) = 1$ by (4.9) and (4.7). □

PROOF OF THEOREM 4.1. The construction in Case 1 proceeds inductively on X_β, beginning with $X_1 = X$. Suppose the construction has been carried out, and (4.5) holds, for all $\beta < B$. Define X_B by (4.4). Assume $X_B \neq \emptyset$. Note that X_B is convex and monotone by Lemma 4.3 and X_B is bounded below.

By Theorem 2.4 δ is the limit of procedures δ_n, each of which is Bayes with respect to a prior ν_n with compact support in Λ.

Let $M = \mathscr{M}(X_B)$ and

$$\gamma_n = \sum_{y \in M} \pi_{\nu_n}(y) \tag{4.10}$$

with π_{ν_n} defined from ν_n via formula (3.1). (Note that $\pi_{\nu_n}(x) > 0$ for all $x \in X$, so that $\gamma_n > 0$.) Define the compatible family $\{\omega_{ny}: y \in M\}$ by

$$\omega_{ny}(d\lambda) = c(\lambda)\lambda^{(y)}\nu_n(d\lambda)/\gamma_n. \tag{4.11}$$

Let π_n denote the function defined from $\{\omega_{ny}\}$ via formula (4.2). Then for $x \geq y$

$$\pi_n(x) = \pi_{\nu_n}(x)/\gamma_n.$$

Also,

$$\sum_{y \in M} \pi_n(y) = 1. \tag{4.12}$$

Now, for $x \in X_B$

$$e_i^t\delta_n(x) = \pi_{\nu_n}(x + e_i)/\pi_{\nu_n}(x) = \pi_n(x + e_i)/\pi_n(x) \to e_i^t\delta(x).$$

By Lemma 4.4 there is a subsequence $\{n'\}$ and a compatible family $\{\omega_y: y \in M\}$

and corresponding π such that $\pi_{n'}(x) \to \pi(x)$ for $x \in X_B$. Consequently

$$e_i^t\delta(x) = \pi(x + e_i)/\pi(x) = \lim_{n'\to\infty}\pi_{n'}(x + e_i)/\pi_{n'}(x) \quad \text{for} \quad x \in X_B^+. \tag{4.13}$$

This is the desired result, (4.5).

The proof in Case 2 is similar. Because of the representation (3.2) the construction here begins with $X_1 = \{x : x + e_i \in X, \text{ for some } i = 1, \cdots, p\}$. The argument then proceeds as above with (4.13) replaced by

$$\begin{aligned} e_i^t\delta(x) &= \pi(x)/\pi(x - e_i) \\ &= \lim_{n'\to\infty}\pi_{n'}(x)/\pi_{n'}(x - e_i) \quad \text{for} \quad x - e_i \in X_B^+. \end{aligned} \tag{4.14}$$

This yields (4.6), and completes the proof. □

Note that the families $\{\omega_{\beta y}: y \in \mathscr{M}(X_\beta)\}$ constructed above satisfy

$$\textstyle\sum_{y\in\mathscr{M}(X_\beta)} \pi_\beta(y) = 1 \tag{4.15}$$

by (4.12) and Lemma 4.4.

5. The one-dimensional problem. For $p = 1$ it is easy to give a concise and more precise statement of Theorem 4.1. Let Δ_k denote the set of estimators, δ, such that $\delta(0) = \cdots = \delta(k) = 0$.

THEOREM 5.1. *When $p = 1$ δ is admissible only if for some $k = -1, \cdots$ $\delta \in \Delta_k - \Delta_{k+1}$ and there is a finite measure ω on $\bar{\Lambda}$ such that*

$$\delta(x) = \int \lambda^{(x-k)}\,\omega(d\lambda) \Big/ \int \lambda^{(x-k-1)}\,\omega(d\lambda) > 0, \quad x \geq k + 1. \tag{5.1}$$

PROOF. If $\delta \in \Delta_k - \Delta_{k+1}$ and is admissible, then by Theorem 4.1 there must be a measure ω such that on $X_{k+2} = \{x \in X: x \geq k + 1\}$

$$\begin{aligned} \delta(x) &= \pi(x + 1)/\pi(x) \\ &= \int \lambda^{(x-k)}\,\omega(d\lambda) \Big/ \int \lambda^{(x-k-1)}\,\omega(d\lambda) \quad \text{for} \quad x \in X_{k+2}^+. \end{aligned} \tag{5.2}$$

Since $\delta \notin \Delta_{k+1}$, $\delta(k + 1) > 0$. Hence $\pi(k + 1) > 0$. It follows that $\omega(\bar{\Lambda} - \{0\}) > 0$ so that $\pi(x) > 0$ for all $x \geq k + 1$ (i.e., $X_{k+2}^+ = X_{k+2}$). (5.2) is thus valid whenever $\delta(x) \neq 0$. □

Whenever $k = -1$ or ω in (5.1) satisfies $\omega(\{0\}) = 0$, then (5.1) can be written in a more conventional form. Let

$$\nu(d\lambda) = c^{-1}(\lambda)\lambda^{-(k+1)}\omega(d\lambda) \tag{5.3}$$

with $c(\lambda)$ given in the definition (2.1) of p_λ. Then (5.1) says that $\delta(0) = \cdots = \delta(k) = 0$ and for $x \geq k + 1$

$$\delta(x) = \int \lambda p_\lambda(x)\,\nu(d\lambda) \Big/ \int p_\lambda(x)\,\nu(d\lambda). \tag{5.4}$$

Thus, δ is determined in the usual manner as a generalized Bayes rule on the set where it is not 0.

It is useful to note that if $\delta \in \Delta_k$ and δ' dominates δ then $\delta' \in \Delta_k$.

LEMMA 5.2. *Let $\delta(0) = \cdots = \delta(k) = 0$ and δ have finite risk for some $\lambda > 0$. Let δ' be as good as δ. Then $\delta'(0) = \cdots = \delta'(k) = 0$.*

PROOF. By induction. Since $c(\lambda) > 0$ comparison of the risk functions gives

$$\sum_{x=0}^{\infty} h(x)\lambda^x(\delta'(x) - \lambda)^2 \le \sum_{x=0}^{\infty} h(x)\lambda^x(\delta(x) - \lambda)^2.$$

Since both risk functions are finite for some $\lambda > 0$, both are analytic functions of λ on a neighborhood of 0. By continuity at $\lambda = 0$, the right side, hence the left side, vanishes. Since $h(0) > 0$, it follows that $\delta'(0) = 0$. Thus $\lambda^2 h(0)$ may be cancelled from both sides and then both sides divided by λ. This establishes the inductive step which if repeated $(k + 1)$ times gives the result. □

The following example shows that not every admissible estimator can be determined as in (5.4).

EXAMPLE 5.3. Let

$$\delta(0) = 0, \quad \delta(1) = ½, \quad \delta(x) = 1 \quad \text{for} \quad x \ge 2. \tag{5.5}$$

This estimator cannot be described as in (5.4) since every estimator determined by (5.4) is either identically constant on $\{x \mid x \ge k + 1\}$ or is strictly increasing in this set.

On the other hand, δ in (5.5) is the member of the complete class of Theorem 4.1 described by $X_1 = X$, $\omega_1(\{0\}) = 1$; $X_2 = \{x \mid x \ge 1\}$, $\omega_2(\{0\}) = \omega_2(\{1\}) = ½$.

Furthermore, δ is admissible. Therefore, if δ' dominates δ, then by Lemma 5.2 $\delta'(0) = 0$. Hence, δ' must also dominate δ in the conditional problem given $x \in X_2$. The probability function in this problem is

$$d(\lambda)h(x)\lambda^x, \quad x \ge 1, \quad \text{with} \quad d^{-1}(\lambda) = \sum_{x=1}^{\infty} h(x)\lambda^x = O(\lambda) \tag{5.6}$$

But δ is unique Bayes in this conditional problem relative to the prior distribution

$$\nu_2(d\lambda) \propto (d(\lambda)\lambda)^{-1}\omega_2(d\lambda). \tag{5.7}$$

(In this formula when $\lambda = 0$ interpret the factor $(d(\lambda)\lambda)^{-1}$ in the natural way as

$$\lim_{\lambda\downarrow 0}(d(\lambda)\lambda)^{-1} = \lim_{\lambda\downarrow 0}(\lambda^{-1}(\textstyle\sum_{x=1}^{\infty} h(x)\lambda^x)) = h(1).)$$

Thus δ is admissible in this conditional problem and $\delta'(x) = \delta(x)$, $x \in X_2$. It follows that δ is admissible in the original problem, as asserted. □

The converse assertion in Theorem 6.2 of the next section generalizes and codifies the procedure used above to prove admissibility.

6. Estimators when $p \ge 2$: Theory. Theorem 5.1 and Lemma 5.2 provide a structural analysis of admissible estimators when $p = 1$. A complete structural analysis when $p \ge 2$ is also possible but we have found no concise description.

(For example, when $p = 2$ we find it necessary to examine sixteen separate cases at each stage of the reduction.) Instead, in this section we present several partial structural results based on the zeros of the estimator.

We first state the main results and then provide the proofs.

THEOREM 6.1. *Let C be convex, monotone and bounded below. In the following δ and δ' are two estimators with δ' as good as δ and $R(\lambda, \delta) < \infty$ for some $\lambda \in \Lambda$. Let*

$$S_1 = \{\delta: \delta(x) = 0 \ \forall x \notin C\}$$

and

$$S_2 = \{\delta: e_i^t\delta(x + e_i) = 0 \ \forall x \in C, \ x + e_i \in X, \ i = 1, \cdots, p\}.$$

In Case j, $j = 1, 2$, suppose $\delta \in S_j$. Then $\delta' \in S_j$. Also, if δ is admissible relative to the estimators in S_j then δ is admissible (and, of necessity, $\delta' = \delta$). □

A trivial application of the theorem shows that the estimator $\delta(\cdot) \equiv 0$ is admissible.

THEOREM 6.2. *Let δ be a given estimator. Define*

$$C_1 = \{x \in X: e_i^t\delta(x) > 0, \ i = 1, \cdots, p\} \tag{6.1}$$

and

$$C_2 = \{x: x + e_i \in X, \ e_i^t\delta(x + e_i) > 0, \ i = 1, \cdots, p\}. \tag{6.2}$$

In Case 1 (Case 2, resp.) suppose C_1 (C_2) is convex and monotone. Then, if δ is admissible there exists a compatible family,

$$\{\omega_y: y \in \mathscr{M}(C_j)\}, \quad j = 1 \ (j = 2),$$

satisfying condition L and $\pi(x) < \infty$, $x \in C_1$ ($x \in C_2$) such that $C_1^+ = C_1$ ($C_2^+ = C_2$) and

$$e_i^t\delta(x) = \pi(x + e_i)/\pi(x), \quad x \in C_1 \tag{6.3}$$

$$(e_i^t\delta(x + e_i) = \pi(x + e_i)/\pi(x), \ x \in C_2). \tag{6.4}$$

As a partial converse, suppose $C_1^+ = C_1$ ($C_2^+ = C_2$) and (6.3) ((6.4)) holds with $\delta(x) = 0$ for $x \notin C_1$ ($e_i^t\delta(x + e_i) = 0$ for $x \notin C_2$). Suppose $\omega_y(\overline{\Lambda} - \hat{\Lambda}) = 0$, $y \in \mathscr{M}(C_j)$, $j = 1$ ($j = 2$). Define

$$g_{1y}(\lambda) = \textstyle\sum_{x \geq y} h(x)(\| \delta(x) \|^2 + \| \lambda \|^2)\lambda^{(x-y)}, \quad y \in \mathscr{M}(C_1),$$

and

$$g_{2y}(\lambda) = \textstyle\sum_{x \geq y} \sum_{j=1}^{p} h(x + e_j)((e_j^t\delta(x + e_j))^2 + (e_j^t\lambda)^2)\lambda^{(x-y)}, \quad y \in \mathscr{M}(C_2).$$

Assume

$$g_{jy}(\lambda) < \infty, \quad \lambda \in \hat{\Lambda}, \quad y \in \mathscr{M}(C_j), \quad \text{for} \quad j = 1 \ (j = 2) \tag{6.5}$$

and

$$\int g_{jy}(\lambda)\ \omega_y(d\lambda) < \infty, \quad y \in \mathscr{M}(C_j) \quad \text{for} \quad j = 1\ (j = 2). \tag{6.6}$$

Then δ is admissible.

REMARKS. Note that (6.3) agrees with (4.5) and (6.4) agrees with (4.6).

In general C_1 and C_2 need not be convex and monotone even when δ is admissible. See Example 7.3.

Theorem 6.2 does not constitute a necessary and sufficient condition for admissibility for two important reasons. The most conspicuous reason is that the converse in Case 1 applies only to estimators for which $\delta(x) = 0$ for $x \notin C_1$, with a similar condition for Case 2. There exist many admissible estimators which do not satisfy this condition. See Examples 7.1 and 7.3. The second reason lies in the finiteness condition (6.6). (The condition (6.5) is a minor and perhaps removeable technical condition.) Condition (6.6) is akin to requiring that the estimator be Bayes with finite Bayes risk. There are very many estimators which are admissible and are useful in applications which do not satisfy such a condition, for example, the estimator $\delta(x) = x$. See Example 7.1.

To see more clearly why (6.6) conveys this finite Bayes risk character, consider the Case 1, multidimensional Poisson problem. Suppose $C_1 = X$ and $\delta(x) = O(1 + \| x \|)$. Then $g_1(\lambda) = O((1 + \| \lambda \|^2)c^{-1}(\lambda))$ and (6.6) requires that $\int \| \lambda \|^2 c^{-1}(\lambda)\ \omega(d\lambda) < \infty$. This is precisely the condition for δ to have finite Bayes risk under the finite prior measure $\nu(d\lambda) = c^{-1}(\lambda)\omega(d\lambda)$.

Johnstone (1981) has given a shorter, direct proof of the special case of Theorem 6.2 which follows.

COROLLARY 6.3. *Suppose $\hat{\Lambda} = \bar{\Lambda}$. In Case 2 suppose δ is admissible and $e_i^t\delta(x) = 0$ if and only if $e_i^t x = 0$. Then δ is generalized Bayes in the sense that there exists a σ-finite locally finite measure ν on $\bar{\Lambda}$ such that*

$$e_i^t\delta(x) = \pi_\nu(x)/\pi_\nu(x - e_i) \quad \text{if} \quad e_i^t x \geq 1. \tag{6.7}$$

PROOF. Apply the first part of Theorem 6.2. Here $C_2 = X$. Hence $\mathscr{M}(C_2) = \{0\}$ and there exists a measure ω_0 for which the corresponding π function—call it π_0—satisfies (6.4). Let $\nu(d\lambda) = c^{-1}(\lambda)\omega_0(d\lambda)$. Then $\pi_\nu = \pi_0$ and (6.4) is the same as (6.7). □

The analogous result in Case 1 is

COROLLARY 6.4. *Suppose $\hat{\Lambda} = \bar{\Lambda}$. In Case 1 suppose δ is admissible and $e_i^t\delta(x) > 0$ for all x and all $i = 1, \cdots, p$. Then δ is generalized Bayes for a σ-finite locally finite measure ν on $\bar{\Lambda}$. Thus*

$$e_i^t\delta(x) = \pi_\nu(x + e_i)/\pi_\nu(x) \quad x \in X. \tag{6.8}$$

PROOF. Similar to Corollary 6.3. □

Proof of Theorem 6.1. Consider Case 1. Let $\delta \in S_1$. Suppose $\delta' \notin S_1$. Then there is a value $x' \notin C$ such that $\delta'(x') \neq 0$. Since $x' \notin$ (convex hull of C) it may be strictly separated from C; so there is a vector $a \neq 0$ and a constant c such that $a^t x' < c$ and $a^t x \geq c$ for all $x \in C$. Since C is monotone $a^t(x + ne_i) \geq c$ for all $n \geq 0$, $x \in C$. This implies that $a^t e_i \geq 0$ for all $i = 1, \cdots, p$.

Let λ_τ be the vector with $e_i^t \lambda_\tau = \tau^{e_i^t a}$; so that $\lambda_\tau^{(x)} = \tau^{a^t x}$. Let

$$d = \inf\{a^t x \colon x \in X,\ \delta'(x) \neq 0\} \geq 0. \tag{6.9}$$

Then $d \leq a^t \delta'(x') < c$ and there exists an $x'' \notin C$ such that $\delta(x'') \neq 0$ and $a^t x'' = d$ since there can be only a finite number of values of $a^t x < c$, $x \in X$. Then

$$\begin{aligned} c^{-1}(\lambda_\tau)(R(\lambda_\tau, \delta') &- R(\lambda_\tau, \delta)) \\ &= \sum (\| \delta'(x) - \lambda_\tau \|^2 - \| \delta(x) - \lambda_\tau \|^2) h(x) \tau^{a^t x} \\ &\geq \textstyle\sum_{x: a^t x = d} (\| \delta'(x) - \lambda_\tau \|^2 - \| \lambda_\tau \|^2) h(x) \tau^{a^t x} \\ &\quad - \textstyle\sum_{x: a^t x > d} \| \delta(x) - \lambda_\tau \|^2 h(x) \tau^{a^t x}. \end{aligned} \tag{6.10}$$

$R(\lambda_{\tau_0}, \delta) < \infty$ for some $\tau_0 > 0$ by Lemma 2.1 and the fact that $R(\lambda, \delta) < \infty$ for some $\lambda \in \Lambda$. Hence

$$\sum (\| \delta(x) \|^2 + \| \lambda_{\tau_0} \|^2) h(x) \tau_0^{a^t x} < \infty.$$

It then follows from the dominated convergence theorem that

$$\lim_{\tau \downarrow 0} \textstyle\sum_{x: a^t x > d} (\| \delta(x) \|^2 + \| \lambda_\tau \|^2) h(x) \tau^{a^t x - d} = 0. \tag{6.11}$$

Combining (6.10) and (6.11) yields

$$\lim_{t \downarrow 0} (c^{-1}(\lambda_\tau) \tau^{-d} (R(\lambda_\tau, \delta') - R(\lambda_\tau, \delta))) \geq \| \delta'(x'') \|^2 h(x'') > 0. \tag{6.12}$$

This contradicts the assumption that δ' is as good as δ. Consequently $\delta'(x) = 0$ for all $x \notin C$ as asserted in the first conclusion of the theorem. The second conclusion of the theorem is a logical consequence of the first conclusion.

The proof for Case 2 is similar. Note that now

$$R(\lambda_\tau, \delta) = \textstyle\sum_{x + e_i \in X} (e_i^t \delta(x + e_i) - e_i^t \lambda_\tau)^2 h(x + e_i) \tau^{a^t x}. \tag{6.13}$$

Let $x' \notin C$ be a value such that $e_i^t \delta'(x' + e_i) \neq 0$ for some $i = 1, \cdots, p$. Define the vector a as before and

$$d = \inf\{a^t x \colon x + e_i \in X,\ e_i^t \delta'(x + e_i) \neq 0 \text{ for some } i = 1, \cdots, p\}. \tag{6.14}$$

It is no longer always true that $d \geq 0$. Nevertheless, using (6.14) and proceeding as in (6.10) and (6.11) yields

$$\begin{aligned} \lim_{\tau \downarrow 0} (c^{-1}(\lambda_\tau) \tau^{-d} (R(\lambda_\tau, \delta') &- R(\lambda_\tau, \delta))) \\ &\geq (e_j^t \delta'(x'' + e_j))^2 h(x'' + e_j) \\ &\quad - \lim_{\tau \downarrow 0} \textstyle\sum_{a^t x > d, x + e_i \in X} (\delta_i(x + e_i) - \tau^{e_i^t a})^2 h(x + e_i) \tau^{a^t x - d} \\ &= (e_j^t \delta'(x'' + e_j))^2 h(x'' + e_j) > 0 \end{aligned} \tag{6.15}$$

as the analog of (6.12), where by construction x'' and $x'' + e_j$ satisfy $a^t x'' = d$ and $e_j^t \delta'(x'' + e_j) \neq 0$. This then yields the conclusions of the theorem in Case 2. □

PROOF OF THEOREM 6.2. Consider Case 1. Suppose δ is admissible and examine the characterization of Theorem 4.1. Let $\beta_0 = \inf\{\beta: X_\beta^+ \cap C_1 \neq \emptyset\}$. Let $x_1 \in X_{\beta_0} \cap C_1$. Then $x \geq x_1$ implies $x \in X_{\beta_0} \cap C_1$ since both X_{β_0} and C_1 are monotone. Let $x_1 \geq y_1$ for some $y_1 \in \mathscr{M}(X_{\beta_0})$. Then

$$0 < \pi(x_1 + e_i) = \int \lambda^{(x_1 - y_1 + e_i)} \, \omega_{\beta_0 y_1}(d\lambda), \quad i = 1, \cdots, p, \tag{6.16}$$

by (4.5) since $\delta_i(x_1) > 0$, $i = 1, \cdots, p$. This shows $x_2 = x_1 + e_i \in X_{\beta_0}^+ \cap C_1$, and repeating the above reasoning with x_2 and replacing x_1 shows $x_1 + e_i + e_j \in X_{\beta_0}^+ \cap C_1$, etc. It follows by induction that $(x_1 + \mathbf{1}) \in X_{\beta_0}^+ \cap C_1$. Hence

$$0 < \pi(x_1 + \mathbf{1}) = \int \lambda^{(x_1 + \mathbf{1} - y_1)} \, \omega_{\beta_0 y_1}(d\lambda). \tag{6.17}$$

Note that $e_i^t(x_1 + 1 - y_1) \geq 1$ for $i = 1, \cdots, p$. Hence (6.17) implies

$$\omega_{\beta_0 y_1}(\{\lambda: e_i^t \lambda > 0, \, i = 1, \cdots, p\}) > 0. \tag{6.18}$$

The compatibility condition then yields that (6.17) is satisfied by every $\omega_{\beta_0 y}$, $y \in \mathscr{M}(X_{\beta_0})$. Consequently $X_{\beta_0}^+ = X_{\beta_0}$ and $C_1 \supset X_{\beta_0}$. The minimality of β_0 then yields $C_1 = X_{\beta_0}$. Choose the compatible family $\{\omega_y: y \in \mathscr{M}(C_1)\}$ to be $\{\omega_{\beta_0 y}: y \in \mathscr{M}(X_{\beta_0})\}$, and then (6.3) is satisfied.

To prove the converse assertion in Case 1, begin by letting $\mathscr{M}(C_1) = \{y_i: i = 1, \cdots, I\}$ and defining

$$D_i = \{x: x \geq y_i, \, x \ngeq y_j, \, j = 1, \cdots, i - 1\} \subset C_1. \tag{6.19}$$

Let $\{\nu_n'\}$ and $\{\omega_{ny}': y \in \mathscr{M}(C_1)\}$ be the sequences promised by condition L, satisfying (4.3) and $\omega_{ny}' \to \omega_y'$, $y \in \mathscr{M}(C_1)$. If necessary, one may define sequences $\{\nu_n\}$ and $\{\omega_{ny}\}$ by truncating ν_n' to the set $\{\lambda: \|\lambda\| \leq m_n, \, g_{1y}(\lambda) \leq m_n, \, y \in \mathscr{M}(C_1)\}$ and letting m_n tend to infinity sufficiently slowly so that $\{\nu_n\}$ and $\{\omega_{ny}\}$ satisfy (4.3) and $\omega_{ny}' \to \omega_y'$, $y \in \mathscr{M}(C_1)$,

$$\int \lambda^{(t)} \, \omega_{ny}(d\lambda) \to \int \lambda^{(t)} \, \omega_y(d\lambda), \quad t \geq 0, \tag{6.20}$$

and

$$\int g_{1y}(\lambda) \, \omega_{ny}(d\lambda) \to \int g_{1y}(\lambda) \, \omega_y(d\lambda). \tag{6.21}$$

(In establishing (6.21) we use (6.5) and the fact that $\omega_y(\bar{\Lambda} - \hat{\Lambda}) = 0$, and also that g_{1y} is continuous on $\hat{\Lambda}$ by virtue of an argument like that in Lemma 2.1.)

Then

$$\lim_{n\to\infty} \int \sum_{x\in C_1} \| \delta(x) - \lambda \|^2 h(x)\lambda^{(x)} c(\lambda)\ \nu_n(d\lambda)$$

$$(6.22) \qquad = \lim_{n\to\infty} \sum_{i=1}^{I} \sum_{x\in D_i} \int \| \delta(x) - \lambda \|^2 h(x)\lambda^{(x-y_i)}\ \omega_{ny_i}(d\lambda)$$

$$= \sum_{i=1}^{I} \sum_{x\in D_i} \int \| \delta(x) - \lambda \|^2 h(x)\lambda^{(x-y_i)}\ \omega_{y_i}(d\lambda)$$

by (6.6).

Now, suppose δ' is better than δ. Then $\delta'(x) = 0$ for $x \notin C_1$, by Theorem 6.1. Thus

$$0 \le \liminf \int (R(\lambda, \delta) - R(\lambda, \delta'))\ \nu_n(d\lambda)$$

$$(6.23) \qquad = \liminf \sum_{i=1}^{I} \sum_{x\in D_i} \int (\| \delta(x) - \lambda \|^2 - \| \delta'(x) - \lambda \|^2) h(x)\lambda^{(x-y_i)}\ \omega_{ny_i}(d\lambda)$$

$$\le \sum_{i=1}^{I} \sum_{x\in D_i} h(x) \int (\| \delta(x) - \lambda \|)^2 - \| \delta'(x) - \lambda \|^2)\lambda^{(x-y_i)}\ \omega_{y_i}(d\lambda)$$

by Fatou's lemma. The condition that $C_1^{\dagger} = C_1$ guarantees that $\omega_{y_i}(\Lambda) > 0$, $i = 1, \cdots, I$. This, together with (6.3), shows that each integral on the right of (6.23) is nonpositive, and at least one is strictly negative unless $\delta'(x) = \delta(x)$. It thus follows that $\delta'(x) = \delta(x)$, $x \in C_1$, which shows that δ is admissible.

The proof of Theorem 6.2 for Case 2 is similar to the above. It requires only modifications like those which appear in the Case 2 proof of Theorem 6.1 in order to accommodate the changed form of the loss and of the corresponding estimator (6.4). We omit the details. □

REMARK 6.5. The method of proof in the converse to Theorem 6.2 can be used to yield somewhat more. Let δ be an estimator constructed according to the paradigm in Theorem 4.1. Suppose at each stage of the process the sets X_β ($= C_i$) and corresponding measures $\{\omega_{\beta y}\}$ satisfy (6.5) and (6.6). Then δ is admissible. In this case it is not necessary either that $\delta(x) = 0$ for $x \notin X_\beta$, etc., or that $X_\beta^{\dagger} = X_\beta$ as would be required by a straightforward reading of Theorem 6.2.

7. Estimators when $p \ge 2$: Examples. The following examples show how Theorems 4.1, 6.1 and 6.2 may be applied and demonstrate some of the peculiarities and technicalities associated with these theorems. Except where noted (as in the second paragraph of Example 7.1) these examples concern the p-dimensional multivariate Poisson problem in Case 1. In the following examples, measures ω_y are specified by giving their value on their support. Their value is zero elsewhere. We also write λ_i instead of $e_i^t\lambda$, etc.

EXAMPLE 7.1. The estimator $\delta(x) = x$ is in the complete class of Theorem 4.1 (in Case 1). Let $p = 2$. Set $X_1 = X$, $X_2 = X - \{\mathbf{0}\}$, $X_3 = \{x \in X: x \ge (1, 1)\}$.

Then $\mathscr{M}(X_1) = \mathbf{0}$, $\mathscr{M}(X_2) = \{(1, 0), (0, 1)\}$, $\mathscr{M}(X_3) = (1, 1)$. Set $\omega_{1,0}(\{\mathbf{0}\}) = 1$. For $A \subset \mathbb{R}$, a measurable set, let

$$\omega_{2,(1,0)}((A, 0)) = \omega_{2,(0,1)}((0, A)) = \int_A e^{-t}\, dt/2,$$

and for $A \subset \mathbb{R}^2$, a measurable set, let

$$\omega_{3,(1,1)}(A) = \int_A \int e^{-(t_1+t_2)}\, dt_1\, dt_2. \tag{7.1}$$

Easy calculations in (4.5) show that the corresponding estimator is $\delta(x) = x$. The argument for $p \geq 3$ is similar but requires $p + 1$ steps.

$\delta(x) = x$ is also in the complete class in Case 2. Again, let $p = 2$. Here, $X_1 = \{x \in Q: x > (-1, -1)\}$. Set $X_2 = \{x: x \geq \mathbf{0}\}$, and

$$\begin{aligned} &\omega_{1,(-1,0)}((0, 1)) = \omega_{1,(0,-1)}((1, 0)) = \tfrac{1}{2} \\ &\omega_{2,(0,0)}(A) = \int_A \int e^{-(t_1+t_2)}\, dt_1\, dt_2. \end{aligned} \tag{7.2}$$

(Other choices of ω_{1y} are possible as long as $\omega_{1,(-1,0)}$ is concentrated on the line $(0, [0, \infty))$ and gives mass to $(0, (0, \infty))$; and symmetrically for $\omega_{1,(0,-1)}$.) (4.6) then easily yields that $\delta(x) = x$. The argument for $p \geq 3$ is entirely similar.

Note that $\delta(x) = x$ is admissible in Case 1 only for $p \leq 2$ and in Case 2 only for $p = 1$. A comprehensive treatment of admissibility questions for $\delta(x) = x$ and all other linear estimators can be found in Brown and Farrell (1985). □

EXAMPLE 7.2. Consider $p = 2$ and

$$\delta(x) = \begin{cases} \mathbf{0} & x_1 x_2 = 0 \\ v & \text{otherwise} \quad (v \in (0, \infty)^2). \end{cases} \tag{7.3}$$

Set $X_1 = X$, $X_2 = X - \{\mathbf{0}\}$, $X_3 = X_2 - \{(1, 0), (0, 1)\}$, $\cdots$ and $X_\infty = \cap_{i=1}^{\infty} X_i = \{x: x \geq (1, 1)\}$. Then $\mathscr{M}(X_i) = \{(i - 1, 0), (0, i - 1), (1, 1)\}$ for $\infty > i \geq 2$. Set

$$\begin{aligned} \omega_{1,0}(\{\mathbf{0}\}) &= 1 \\ \omega_{i,y}(\{\mathbf{0}\}) &= \tfrac{1}{2}, \quad 2 \leq i < \infty, \quad y = (i - 1, 0), (0, i - 1), \\ \omega_{i,(1,1)} &= 0, \quad 2 \leq i < \infty, \\ \omega_{\infty,(1,1)}(\{v\}) &= 1. \end{aligned}$$

Then $\delta(x)$ satisfies (4.6) so it is in the complete class of Theorem 4.1. Furthermore, it is admissible by Theorem 6.2 since $\omega_{\infty,(1,1)}$ trivially satisfies (6.6) by virtue of having bounded support.

Suppose, more generally, that $\delta(x) = \mathbf{0}$ if $x_1 x_2 = 0$ and δ is admissible for the conditional problem given $X_\infty = \{x: x \geq (1, 1)\}$. Then Theorem 6.1 yields that δ is admissible. For example, the estimator

$$\delta(x) = \begin{cases} \mathbf{0} & x_1 x_2 = 0 \\ x & \text{otherwise} \end{cases} \tag{7.4}$$

is admissible.

Note that the estimator (7.3) requires a transfinite sequence of reductions, $X_1, X_2, \cdots, X_\infty$, in order to satisfy the characterization of Theorem 4.1. □

EXAMPLE 7.3. A peculiar admissible estimator, mentioned also in Brown and Farrell (1983), is (for $p = 2$)

$$\delta(x) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x. \tag{7.5}$$

This fits the characterization of Theorem 4.1 upon setting $X_i = \{x: x \geq (i-1, 0)$ or $x \geq (0, 1)\}$, $i = 1, 2, \cdots$, and $X_{\infty+i} = \{x: x \geq (0, i+1)\}$, $i = 0, 1, \cdots$, and

$$\begin{aligned} \omega_{i,(i-1,0)}(\mathbf{0}) &= 1 \quad i = 1, 2, \cdots \\ \omega_{i,(0,1)} &= 0 \quad i = 2, \cdots \\ \omega_{\infty+i,(0,i+1)}((i+1, 0)) &= 1 \quad i = 0, 1, \cdots . \end{aligned} \tag{7.6}$$

Admissibility of δ follows via Remark 6.5 from this stepwise representation.

A minor modification of this estimator, while still admissible, shows yet another peculiarity. Let

$$\delta(x) = \begin{cases} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x & x^t \neq (0, 2) \\ 0 & x^t = (0, 2). \end{cases} \tag{7.7}$$

(An easy addition to the preceding argument shows this estimator is admissible.) Then

$$C = \{x: \delta(x) \neq \mathbf{0}\} = X - (\{(i, 0), i = 0, 1, \cdots, \} \cup \{(0, 2)\})$$

is not monotone or convex.

EXAMPLE 7.4. Here is an admissible estimator whose risk is somewhere infinite. Let $f(y) > 0$ be real valued such that $\sum_{k=0}^{\infty} f(k)\tau^k$ is convergent if $\tau \leq 1$ and is divergent if $\tau > 1$. Let the estimator be

$$\delta((x_1, x_2)) = (f(x_2), 0). \tag{7.8}$$

This estimator can be seen to be admissible by applying the argument of Example 7.3 with $\omega_{\infty+i}$ in (7.6) changed to be

$$\omega_{\infty+i,(0,i+1)}((f(i+1), 0)) = 1 \quad i = 0, 1, \cdots . \tag{7.9}$$

Then

$$R((\lambda_1, \lambda_2), \delta) = \sum_{k=0}^{\infty} (f(k) - \lambda_2)^2 e^{-\lambda_2} \lambda_2^k / k! + \lambda_1^2$$

$$\begin{cases} < \infty & \text{if} \quad \lambda_2 \leq 1 \\ = \infty & \text{if} \quad \lambda_2 > 1. \end{cases}$$

Thus, $\Lambda^* = \{\lambda: \lambda_2 \leq 1\} \neq \hat{\Lambda}$. □

EXAMPLE 7.5. When $p = 3$ the above can be used to lead to yet more pathological behavior. Define

$$\delta((x_1, x_2, x_3)) = \begin{cases} 0 & \text{if} \quad x_1x_2 + x_2x_3 + x_1x_3 = 0 \\ (0, f(x_3), 0) & \text{if} \quad x_2x_3 > 0, \quad x_1 = 0 \\ (0, 0, f(x_1)) & \text{if} \quad x_1x_3 > 0, \quad x_2 = 0 \\ (f(x_2), 0, 0) & \text{if} \quad x_1x_2 > 0, \quad x_3 = 0 \\ v & \text{if} \quad x_1x_2x_3 > 0. \end{cases} \tag{7.10}$$

This results in an estimator for which

$$\Lambda^* = \{\lambda: \lambda_1 = 0,\ \lambda_3 \le 1 \text{ or } \lambda_2 = 0,\ \lambda_1 \le 1 \\ \text{or } \lambda_3 = 0,\ \lambda_2 \le 1 \text{ or } \lambda_i \le 1\ i = 1, 2, 3\}.$$

We think $v_i \le 1$, $i = 1, 2, 3$, is the correct condition here. Otherwise the estimator is not admissible relative to $\bar{\Lambda}$. Here $((\Lambda)^*)^\circ)^- = \{\lambda: \lambda_i \le 1, i = 1, 2, 3\} \not\supset \Lambda)^*$. If $v_i \le 1$, $i = 1, 2, 3$, then this estimator is admissible relative to $\hat{\Lambda}$, but is not admissible (relative to Λ). (If it were admissible (relative to Λ) it would have to be a limit of Bayes procedures for priors concentrated on $(\Lambda^* \cap \Lambda)$. All such Bayes procedures have $\delta_i \le 1$, $i = 1, 2, 3$. But, f is an unbounded function. So, δ in (7.10) cannot be a limit of such Bayes procedures. Admissibility relative to $\hat{\Lambda}$ follows from an extension of Remark 6.5.) □

EXAMPLE 7.6. A compatible family need not satisfy condition L. Let $p = 2$ and

$$C = \{x: x \ge (0, 2) \text{ or } x \ge (1, 1) \text{ or } x \ge (2, 0)\}.$$

Let

$$\begin{gathered} \omega_{(0,2)}(\mathbf{0}) = \omega_{(2,0)}(\mathbf{0}) = 0 \\ \omega_{(1,1)}(\mathbf{0}) = 1. \end{gathered} \tag{7.11}$$

Then $\{\omega_y\}$ is compatible. However, if ν_n is any finite measure on Λ and $\omega_{ny}(d\lambda) = c(\lambda)\lambda^{(y)}\nu_n(d\lambda)$ then

$$\ln \omega_{n,(2,0)}(\Lambda) + \ln \omega_{n,(0,2)}(\Lambda) \ge 2 \ln \omega_{n,(1,1)}(\Lambda) \tag{7.12}$$

by Hölder's inequality (or, here, by the Cauchy-Schwarz inequality). If ω satisfies condition L then $\omega_{ny}(\Lambda) \to \omega_y(\mathbf{0})$ so (7.12) yields $\ln \omega_{(0,2)}(\mathbf{0}) + \ln \omega_{(2,0)}(\mathbf{0}) \ge 2 \ln \omega_{(1,1)}(\mathbf{0})$ as necessary for condition L. $\{\omega_y\}$ defined by (7.11) does not satisfy this condition.

Consider the estimator

$$\delta(x) = \begin{cases} \mathbf{0} & x_1 + x_2 \le 1 \quad \text{or} \quad x = (1, 1) \\ (1, 1) & \text{otherwise.} \end{cases} \tag{7.13}$$

This estimator can be described by a sequence of compatible measures (of length four) in the manner of Theorem 4.1. However it is not in the complete class of that theorem, and hence is not admissible, since the sequence of compatible measures would have to contain (7.11) which does not satisfy condition L.

EXAMPLE 7.7. When $\hat{\Lambda} \neq \bar{\Lambda}$ it may be necessary to have $\omega_{\beta y}(\bar{\Lambda} - \hat{\Lambda}) > 0$ in Theorem 4.1. The simplest example involves the one-dimensional geometric distribution, whose probability function is

$$p_\lambda(x) = (1 - \lambda)\lambda^x \quad x = 0, 1, \cdots . \tag{7.14}$$

Then $\Lambda = (0, 1)$, $\hat{\Lambda} = [0, 1)$, and $\bar{\Lambda} = [0, 1]$. The estimator

$$\delta(x) \equiv 1 \tag{7.15}$$

is admissible and is produced in the manner of Theorem 4.1 by setting $\omega_{10}(\{1\}) = 1$. (δ cannot be proved admissible by Theorem 6.2 since that theorem requires that $\omega_{\beta y}((\bar{\Lambda} - \hat{\Lambda}) = 0$.) Similarly the estimator

$$\delta^*(x) = (2^x + (½))/(2^x + 1) \quad x = 0, 1, \cdots \tag{7.16}$$

corresponds to $\omega_{10}(\{½\}) = \omega_{10}(\{1\}) = 1$. Again Theorem 6.2 does not apply. Nevertheless this estimator is admissible. To see this note that

$$\textstyle\sum_{x \geq y} (\delta^*(x) - \lambda)^2\lambda^x \leq 2 \sum_{x \geq y} (\delta^*(x) - 1)^2\lambda^x + 2 \sum_{x \geq y} (1 - \lambda)^2\lambda^x \tag{7.17}$$

converges to zero uniformly on (0, 1) as $y \to \infty$. It follows that δ^* uniquely minimizes

$$½ \lim \sup_{\lambda \to 1}(1 - \lambda)^{-1}R(\lambda, \delta) + (½)(1 - ½)^{-1}R(½, \delta) \tag{7.18}$$

and hence must be admissible. (When $\delta = \delta^*$ in (7.18), the limit as $\lambda \to 1$ exists, not merely the lim sup.)

8. Estimation of the negative binomial mean. In this section we give a brief, nonformal, reduction of the problem to a ratio of integrals similar to the ratio obtained for Case 2 loss when estimating the power series parameter, λ.

Consider the case $p = 1$ and the negative binomial distribution

$$p_\lambda(x) = \binom{x + k - 1}{k - 1}(1 - \lambda)^k\lambda^x, \quad x = 0, 1, \cdots . \tag{8.1}$$

The preceding sections give results for estimating the parameter λ. However for various applications it is more natural to estimate the expectation parameter,

$$\theta(\lambda) = E_\lambda(X) = k\lambda/(1 - \lambda). \tag{8.2}$$

For example, Tsui (1982) has considered this problem, using the loss function $(\delta - \theta)^2/\theta$.

Given an a priori measure ν on (0, 1) the Bayes estimator δ then minimizes

$$\int (\delta(x) - \theta(\lambda))^2\theta^{-1}(\lambda)\lambda^x c(\lambda)\nu(d\lambda) \tag{8.3}$$

with $c(\lambda) = (1 - \lambda)^k$ so that

$$\delta(x) = k \int \lambda^x c(\lambda)\nu(d\lambda) \Big/ \int (1 - \lambda)\lambda^{x-1}c(\lambda)\nu(d\lambda). \tag{8.4}$$

Defining $\xi(x) = \delta(x)/(\delta(x) + k)$ and manipulating (8.4) yields

$$\xi(x) = \delta(x)/(\delta(x) + k) = \int \lambda^x c(\lambda)\nu(d\lambda) \Big/ \int \lambda^{x-1} c(\lambda)\nu(d\lambda). \tag{8.5}$$
$$= \pi_\nu(x)/\pi_\nu(x-1).$$

Thus $\xi(x)$ has the same expression here as did $\delta(x)$ in (3.2) for Case 2 loss in estimating λ. The same is, of course, true in the multivariate case where

$$p_\lambda(x) = c(\lambda) \prod_{i=1}^{p} \binom{x_i + k_i - 1}{k_i - 1} \lambda_i^{x_i}, \quad x \in Q_+^p, \tag{8.6}$$
$$c(\lambda) = \prod_{i=1}^{p} (1 - \lambda_i)^{k_i},$$

and the loss function is

$$\sum (\delta_i - \theta_i)^2/\theta_i \quad \text{with} \quad \theta_i(\lambda_i) = k_i\lambda_i/(1 - \lambda_i),$$

and the Bayes procedure δ for ν satisfies

$$\xi_i(x) = \delta_i(x)/(\delta_i(x) + k_i) = \pi_\nu(x)/\pi_\nu(x - e_i). \tag{8.7}$$

Thus, Theorem 4.1 for Case 2 describes a complete class by virtue of describing the functions $\xi(\cdot)$ which correspond to estimators δ in this class. (One minor change may be noted. The value $\xi(x) = 1$ is impossible for an admissible estimator, hence measures, such as $\omega(\{\mathbf{1}\}) = 1$, which lead to such results need not be considered.) The main results of Section 5 and 6 also carry over to this problem; the only major change necessary is that the definition of g_{2y} in Theorem 6.2 must be modified to match the new loss function. □

REFERENCES

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

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

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

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

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

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.

FARRELL, R. H. (1968). Towards a theory of generalized Bayes tests. *Ann. Math. Statist.* **39** 1–22.

GHOSH, M., HWANG, J. T. and TSUI, K. W. (1983). Construction of improved estimators in multiparameter estimation for exponential families. *Ann. 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-death processes, and discrete Dirichlet problems. Ph.D. Thesis, Cornell University.

JOHNSTONE, I. M. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. Submitted, *Ann. Statist.* **12** 1173–1198.

JOHNSTONE, I. M. (1985). Admissible estimation, Dirichlet principles and recurrence of birth-death chains on Z_+^p. To appear in *Z. Wahrsch. verw. Gebiete.*

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

TSUI, K. W. (1982). Robustness of Clevenson-Zidek type estimators. Technical Report No. 691, Dept. of Statistics, University of Wisconsin.

WALD, A. (1949). *Statistical Decision Functions.* Wiley, New York.

HEBREW UNIVERSITY
MT. SCOPUS CAMPUS
91905 JERUSALEM
ISRAEL

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853