JOURNAL OF MULTIVARIATE ANALYSIS **24**, 46–52 (1988)

Proof of a Necessary and Sufficient Condition for Admissibility in Discrete Multivariate Problems*

L. D. BROWN AND R. H. FARRELL

Cornell University

Communicated by A. Cohen

The proof of Farrell (1968. *Ann. Math. Statist.* **26** 518–522) is adapted to the special problems presented by discrete problems. Continuity of the risk functions is verified, sequential subcompactness is verified, and a necessary and sufficient condition for admissibility proven. In the discrete problems considered one obtains pointwise convergence of the sequence of Bayes estimators to the admissible estimator. This last property is crucial to further development of the decision theory given in Brown and Farrell (1985. *Ann. Math. Statist.* **13** 706–726). © 1988 Academic Press, Inc.

1. INTRODUCTION

The purpose of this paper is to provide a self-contained and elementary proof of a necessary and sufficient condition for admissibility of estimators of a multivariate parameter vector λ in certain discrete problems relative to quadratic type losses. We will discuss p-dimensional (column) parameter vectors and the lattice N^p of $p \times 1$ vectors x having nonnegative integer coordinates. For $1 \leqslant i \leqslant p$ let e_i be the $p \times 1$ unit vector with ith coordinate equal to one. Then we may describe the discrete probabilities as

$$c(\lambda)\, h(x) \prod_{i=1}^{p} (e_i'\lambda)^{(e_i'x)}, \tag{1.1}$$

a description which includes the multivariate Poisson and negative binomial families, having differing parameter spaces. The necessary and sufficient condition is stated as Theorem 2.6, is similar to Stein [6] and Farrell [4]. In fact, to apply Farrell [4] it is necessary to prove the first four lemmas of Section 2, that is, write most of this paper.

Received August 12, 1983; revised August 20, 1984.
AMS 1980 subject classification: Primary 62C07; Secondary 62F10.
Key words and phrases: Estimation, multivariate discrete probabilities, decision theory.
* Research supported in part by the National Science Foundation Grant M.C.S. 8200031.

0047-259X/88 $3.00

This paper was originally Section 2 of Brown and Farrell [2] but was taken out of that paper in order to shorten that paper. Brown and Farrell, op. cit., needed only a necessary condition, pointwise limits of sequences of Bayes estimators, condition (2.5(2)) of this paper, not the full strength of Theorem 2.6.

The complete Theorem 2.6 is needed as a foundation for Johnstone [5], who must know the existence of a sequence of a priori measures putting unit mass at a specified parameter value. The original motivation for this paper was Johnstone's [5] Ph.D. thesis.

In applications such as Brown [1] it was necessary to restrict consideration to estimators having bounded risk functions. In fact Brown, op. cit., believed that he had proved admissible estimators had everywhere finite risk. But Johnstone, in a personal communication, has shown Brown's argument to be incorrect and the question remains open for the multivariate normal mean vector estimation problem. For the discrete problems that we consider, with the parameter vector λ restricted to have only positive coordinates, in Section 3 an example is given of an admissible estimator having infinite risk on a half-space.

2. Notations, Preliminary Decision Theory

In this paper x, y will be p-dimensional column vectors, x having nonnegative integer entries, y having nonnegative real number entries. We write R^p_+ for the parameter space. For the multivariate Poisson measures $R^p_+ = (0, \infty) \times \cdots \times (0, \infty)$ and for the multivariate negative binomial $R^p_+ = (0, 1) \times \cdots \times (0, 1)$. The decision theory depends on the lattice of x values being unbounded but needs only that R^p_+ be an open set containing intervals $\{\lambda \mid \lambda_0 \leqslant \lambda \leqslant \lambda_1\}$ whenever $\lambda_0, \lambda_1 \in R^p_+$. Limit measures obtained will in general be supported on the closure of R^p_+ but in this paper it is not necessary to name the closure set.

The discrete probability density functions are of the geometric form

$$c(\lambda)\, h(x)\, \lambda^{(x)}, \lambda^{(x)} = \prod_{i=1}^{p} (e_i'\lambda)^{(e_i'x)}. \tag{2.1}$$

Then $1 = \sum_{x \geqslant 0} c(\lambda)\, h(x)\, \lambda^{(x)}$. In estimation of the parameter vector λ, the loss and risk are computed by

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

where $\delta(\ )$ is a $p \times 1$ column vector and the loss function being used is L, the risk function R.

General theory is presented in this paper but applications are dependent on the specification of the parameters $\alpha_1, \ldots, \alpha_p$ in the function. In particular the character of the stepwise reduction of the sample space discussed in Brown and Farrell [2] is dependent on the choice of these parameters and they restrict the discussion 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 2 loss is used by Johnstone [5] and Tsui [7].

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

$$R(\lambda, \delta) = \sum_{i=1}^{p} \sum_{x \geqslant 0} c(\lambda)\, h(x)\, \lambda^{(x+\alpha_i e_i)} (e_i'(\delta(x) - \lambda))^2. \tag{2.3}$$

In the sequel we will say a parameter set Λ is monotone provided $\lambda \in \Lambda$ and $\lambda' \leqslant \lambda$ implies $\lambda' \in \Lambda$. The finite risk set of an estimator is defined by $\Lambda^*(\delta) = \{\lambda \mid R(\lambda, \delta) < \infty\}$. It is understood that $\Lambda^*(\delta) \subset R_+^p$.

LEMMA 2.1. *If δ is admissible then $\Lambda^*(\delta)$ is nonempty and monotone. If λ_0 is a boundary point of $\Lambda^*(\delta)$ then* $\lim_{\lambda \to \lambda_0\, \lambda \leqslant \lambda_0} R(\lambda, \delta) = R(\lambda_0, \delta)$. *If this limit is infinite then* $\lim_{\lambda \to \lambda_0} R(\lambda, \delta) = R(\lambda_0, \delta)$.

Proof. If $R(\lambda_0, \delta) < \infty$ then $\sum_{i=1}^{p} \sum_{x \geqslant 0} (e_i' \delta(x))^2 h(x) \lambda_0^{(x)} < \infty$. This series is absolutely convergent and $\lambda_0 \geqslant 0$ so that if $0 \leqslant \lambda \leqslant \lambda_0$ by the comparison test $\sum_{i=1}^{p} \sum_{x \geqslant 0} (e_i' \delta(x))^2 h(x) \lambda^{(x)} < \infty$, the risk is an infinite series that is dominated term by term by the terms $2((e_i'\delta(x))^2 + (e_i'\lambda_0)^2)\, c(\lambda_0)\, h(x)\, \lambda_0^{(x+\alpha_i e_i)}$ so that for λ having only nonzero entries, $\lambda \leqslant \lambda_0$, the M-test for uniform convergence applies and $\lim_{\lambda \to \lambda_0} R(\lambda, \delta) = R(\lambda_0, \delta)$. The risk function is lower semicontinuous on the closure of R_+^p so at a boundary point λ_0 of $\Lambda^*(\delta)$ such that $R(\lambda_0, \delta) = \infty$, it follows that $\lim_{\lambda \to \lambda_0} \inf R(\lambda, \delta) = \lim_{\lambda \to \lambda_0} R(\lambda, \delta) = \infty$. ∎

LEMMA 2.2. *If λ_0 is interior to $\Lambda^*(\delta)$ then $R(\ , \delta)$ is continuous at λ_0.*

Proof. $c(\lambda)^{-1} = \sum_{x \geqslant 0} h(x) \lambda^{(x)}$ is an analytic function of λ, hence continuous in $\lambda \in R_+^p$. $\lambda^{(\alpha_i e_i)}$ is likewise continuous on R_+^p. If λ_0 is interior to $\Lambda^*(\delta)$ then there exist $\lambda_1 < \lambda_0 < \lambda_2$, λ_1 and λ_2 in $\Lambda^*(\delta)$, such that $\{\lambda \mid \lambda_1 \leqslant \lambda \leqslant \lambda_2\}$ has interior, and λ_1 has no zero entries. By the argument for Lemma 2.1, the series $\sum_{x \geqslant 0} (e_i'(\delta(\lambda) - \lambda))^2 h(x) \lambda^{(x)}$ converges uniformly in λ in the rectangle. Hence continuity at λ_0 follows.

LEMMA 2.3. *Let $\mathscr{R}$ be the set of risk functions of nonrandomized estimators. The set $\mathscr{R}$ is sequentially subcompact (c.f., Farrell [4] for a definition).*

Proof. Let $\{\delta_n, n \geqslant 1\}$ be a sequence of estimators. If $\lim_{n\to\infty} \inf R(\lambda, \delta_n) = \infty$ for all $\lambda \in R^p_+$ then the conclusion of the lemma follows. Suppose for some λ that $\lim_{n\to\infty} \inf R(\lambda, \delta_n) < \infty$. We may choose a subsequence $\{m'_n\}$ such that for all i,

$$\sup_n \sum_{x \geqslant 0} h(x)\, \lambda^{(x)}(e_i^t(\delta_{m'_n}(x) - \lambda))^2 < \infty.$$

For those x such that $h(x) > 0$, since $\lambda^{(x)} > 0$, it follows that $\sup_n |e_i^t \delta_{m'_n}(x)| < \infty$. By a diagonalization argument there exists a subsequence $\{m_n\}$ such that if $h(x) > 0$ then $\lim_{n\to\infty} e_i^t \delta_{m_n}(x) = e_i^t \delta(x)$ exists. By Fatou's lemma, for all $\lambda \in R^p_+$,

$$\begin{aligned} &\lim_{n\to\infty} \inf R(\lambda, \delta_{m_n}) \\ &\qquad \geqslant \sum_{i=1}^{p} \sum_{x \geqslant 0} c(\lambda)\, h(x)\, \lambda^{(x+\alpha_i e_i)} \lim_{n\to\infty} \inf\, (e_i^t(\delta_{m_n}(x) - \lambda))^2 \\ &\qquad = R(\lambda, \delta). \end{aligned} \tag{2.4}$$

Since the estimator δ is nonrandomized the risk function $R(\ ,\delta) \in \mathscr{R}$ and the lemma is proven. ∎

In the sequel we let $\Lambda = \{\lambda_i, i \geqslant 1\}$ be a countable dense subset of R^p_+, and $\Lambda_n = \{\lambda_1, ..., \lambda_n\}$. $\{\lambda_1\} = \Lambda_1$ will be the "preferred" parameter value. Let $\mathscr{R}_n$ be the set of vectors $\mathscr{R}_n(\delta)$ with

$$\mathscr{R}_n(\delta)^t = (R(\lambda_1, \delta), ..., R(\lambda_n, \delta))$$

taken over all nonrandomized estimators δ. We let $\mathscr{R}_{n\geqslant}$ be the set $\{y|$ exists $x \in \mathscr{R}_n, y \geqslant x\}$. The ordering here means componentwise the components of y are greater than those of x.

LEMMA 2.4. *$\mathscr{R}_{n\geqslant}$ is a closed and convex set.*

Proof. Let $y_q \in \mathscr{R}_{n\geqslant}$, and $y_q \geqslant \mathscr{R}_n(\delta_q)$. Let $y = \lim_{q\to\infty} y_q$. On a subsequence δ_{m_q}, for all λ, (2.4) holds and $\delta(x) = \lim_{q\to\infty} \delta_{m_q}(x)$ is a nonrandomized estimator. Thus $R_n(\delta) \in \mathscr{R}_{n\geqslant}$ and $R_n(\delta) \leqslant y$. Hence $y \in \mathscr{R}_{n\geqslant}$, and $\mathscr{R}_{n\geqslant}$ is closed. To show convexity note that $\alpha R(\lambda, \delta) + (1-\alpha) R(\lambda, \tilde{\delta}) \geqslant R(\lambda, a\delta + (1-a)\tilde{\delta})$, so the mixture is in $\mathscr{R}_{n\geqslant}$. ∎

LEMMA 2.5. *If $\varepsilon > 0$ and δ is admissible then there exists an n_0 such that if $n > n_0$ then $R_n(\delta) - \varepsilon \tilde{e}_n \notin \mathscr{R}_{n\geqslant}$, where $\tilde{e}_n$ is the $n \times 1$ unit vector with one in the first coordinate position.*

Proof. By contradiction. If there is a sequence m_n on which $R_{m_n}(\delta) - \varepsilon \tilde{e}_{m_n} \in \mathscr{R}_{m_n \geqslant}$ then $R_{m_n}(\delta) \geqslant \varepsilon \tilde{e}_{m_n} + R_{m_n}(\delta_{m_n})$ for $\lambda \in \Lambda_{m_n}$. By subcompactness there exists a further subsequence and an estimator $\tilde{\delta}$ such that in the limit $R(\lambda_1, \delta) \geqslant \varepsilon + R(\lambda_1, \tilde{\delta})$ and $R(\lambda, \delta) \geqslant R(\lambda, \tilde{\delta})$ for all $\lambda \in \Lambda$. By Lemma 2.2, $R(\, , \delta)$ is continuous on the interior of the finite risk set. Then $R(\, , \tilde{\delta})$ is finite valued on this set and, by Lemma 2.2, continuous on $\Lambda^*(\delta)$. Since Λ is dense, it follows that by taking limits

$$R(\lambda, \delta) \geqslant R(\lambda, \tilde{\delta}) \qquad \text{for all } \lambda \in \Lambda^*(\delta)$$

and

$$R(\lambda_1, \delta) \geqslant R(\lambda_1, \tilde{\delta}) + \varepsilon.$$

This contradiction proves the lemma.

THEOREM 2.6. *Let δ be an estimator and $\lambda_1 \in \Lambda^*(\delta)$ be a preferred parameter and $\Lambda^*(\delta)$ and Λ_n be as above. A necessary and sufficient condition that δ be admissible is that there exist a sequence of finite measures $\{\nu_n, n \geqslant 1\}$ with ν_n discrete and supported on Λ_n such that $\nu_n(\{\lambda_1\}) = 1$ and such that if δ_n is Bayes relative to ν_n then*

$$\lim_{n \to \infty} \int (R(\lambda, \delta) - R(\lambda, \delta_n))\, \nu_n(d\lambda) = 0; \tag{2.5(1)}$$

$$\lim_{n \to \infty} \delta_n(x) = \delta(x) \qquad \text{for all } x \text{ such that } h(x) > 0. \tag{2.5(2)}$$

Proof. By Lemma 2.5, we may suppose the sequence $\varepsilon_n \downarrow 0$ such that $R_n(\delta) - \varepsilon_n \tilde{e}_n \notin \mathscr{R}_{n\geqslant}$. Then, since $\mathscr{R}_{n\geqslant}$ is closed and convex, there exists a vector $a_n \in R^p$ and a constant c such that

$$a_n^t(R_n(\delta) - \varepsilon_n \tilde{e}_n) < c \leqslant \inf_{x \in \mathscr{R}_{n\geqslant}} a_n^t x. \tag{2.6}$$

Consequently if $b \in R_+^p$, $m \geqslant 0$ is an integer, and $x \in \mathscr{R}_{n\geqslant}$ then $c \leqslant a_n^t(x + mb)$ and $m^{-1}(c - a_n^t x) \leqslant a_n^t b$. Thus in the limit as $m \to \infty$, $0 \leqslant a_n^t b$ for all $b \in R_+^n$. Thus the components of a_n are nonnegative and condition (2.6) implies $a_n^t \tilde{e}_n \neq 0$. Thus we may renormalize so the first coordinate of a_n is 1 and consider the entries of a_n as the masses of the measure ν_n. Then (also change the value of c) $\int R(\lambda, \delta)\, \nu_n(d\lambda) - \varepsilon_n \leqslant c \leqslant \int R(\lambda, \delta_n)\, \nu_n(d\lambda)$, from which (2.5(1)) follows. By a standard calculation

$$0 = \lim_{n \to \infty} \sum_{i=1}^{p} \sum_{x \geqslant 0} \int c(\lambda)\, h(x)\, \lambda^{(x + \alpha_i e_i)} \times (e_i^t(\delta(x) - \delta_n(x)))^2\, \nu_n(d\lambda)$$

$$\geqslant \lim_{n \to \infty} \sum_{i=1}^{p} \sum_{x \geqslant 0} c(\lambda_1)\, h(x)\, \lambda_1^{(x + \alpha_i e_i)} \times (e_i^t(\delta(x) - \delta_n(x)))^2.$$

Hence if $h(x) > 0$ it follows that $\delta(x) = \lim_{n \to \infty} \delta_n(x)$.

Conversely, to show admissibility, suppose δ' is as good as δ. Then (2.5(1)) implies

$$\lim_{n \to \infty} \int (R(\lambda, \delta') - R(\lambda, \delta_n))\, \nu_n(d\lambda) = 0, \tag{2.7}$$

so that as argued above,

$$0 = \lim_{n \to \infty} \sum_{i=1}^{p} \sum_{x \geqslant 0} c(\lambda_1)\, h(x)\, \lambda_1^{(x + \alpha_i e_i)} (e_i^t(\delta'(x) - \delta_n(x)))^2. \tag{2.8}$$

Thus, if $h(x) > 0$ it follows that

$$\delta'(x) = \lim_{n \to \infty} \delta_n(x) = \delta(x). \tag{2.9}$$

This proves δ to be admissible. ∎

3. An Example

In this section a two-dimensional ($p = 2$) example is given of an admissible estimator having infinite risk on a half-space. This is same as Example 7.4 of Brown and Farrell [2] but is given a different treatment here. To simplify notation in this section we will write $\lambda^t = (\lambda_1, \lambda_2)$, $\delta^t = (\delta_1, \delta_2)$, and (x, y) for points of the lattice. Let $\{a_n, n \geqslant 1\}$ be a positive real number sequence such that $\sum_y a_y^2 \lambda^y / y!$ converges if $0 \leqslant \lambda < 1$ and diverges if $\lambda \geqslant 1$. We show for squared error loss that the estimator $\delta'(x, y)^t = (a_y, 0)$ is admissible and has infinite risk on the half-space $\lambda_2 \geqslant 1$. In fact the risk function is $\lambda_2^2 + \sum_y (a_y - \lambda_1)^2 \exp(-\lambda_2)\, \lambda_2^y / y!$ which converges if and only if $\lambda_2 < 1$.

To show admissibility, suppose δ is as good as δ'. Let $y_0 \geqslant 0$ be the least integer y such that there exists an integer $x \geqslant 0$ with $\delta(x, y_0) \neq \delta'(x, y_0)$. We show $y_0 \geqslant 0$ cannot exist as follows. After cancellation of equal terms in the risk functions the comparison is

$$\sum_{y \geqslant y_0} \sum_{x} \|\delta - \lambda\|^2 c(\lambda_1, \lambda_2) h(x, y) \lambda_1^x \lambda_2^y \leqslant \sum_{y \geqslant y_0} \sum_{x} \|\delta' - \lambda\| c(\lambda_1, \lambda_2) h(x, y) \lambda_1^x \lambda_2^y.$$

Using continuity of the risk functions we may divide by $\lambda_2^{y_0}$ and set $\lambda_2 = 0$ to obtain

$$\begin{aligned} &\sum_{x} (\delta_1(x, y_0) - \lambda_1)^2 c(\lambda_1, 0) h(x, y_0) \lambda_1^x \\ &\quad \leqslant \sum_{x} ((\delta_1(x, y_0) - \lambda_1)^2 + (\delta_2(x, y_0))^2) c(\lambda_1, 0) h(x, y_0) \lambda_1^x \\ &\quad \leqslant \sum_{x} (a_{y_0} - \lambda_1)^2 c(\lambda_1, 0) h(x, y_0) \lambda_1^x. \end{aligned} \tag{3.1}$$

For the one-dimensional problem, since $a_{y_0} > 0$, this constant is a uniquely determined Bayes estimator, hence is admissible. Then (3.1) implies $\delta_1(x, y_0) = a_{y_0}$ for all $x \geqslant 0$, which implies $\delta_2(x, y_0) = 0$ for all $x \geqslant 0$. This contradicts the definition of y_0. Consequently $y_0 \geqslant 0$ cannot exist and $\delta = \delta'$ for all (x, y).

References

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

[2] Brown, L. D. and Farrell, R. H. (1985). Complete class theorems for estimation of multivariate Poisson means and related problems. *Ann. Math. Statist.* **13** 706–726.

[3] Brown, L. D. and Farrell, R. H. (1985). All admissible linear estimators of the multivariate Poisson mean vector. *Ann. Math. Statist.* **13** 282–294.

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

[5] Johnstone, I. M. (1981). *Admissible Estimation of Poisson Means, Birth–Death Processes, and Discrete Dirichlet Problems.* Ph.D. thesis, Cornell University.

[6] Stein, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Statist.* **26** 518–522.

[7] Tsui, K. W. (1982). *Robustness of Clevenson–Zidek Type Estimators.* Technical Report 691, Department of Statistics, University of Wisconsin.

[8] Wald, A. (1949). *Statistical Decision Functions.* Wiley, New York.