JOURNAL OF MULTIVARIATE ANALYSIS 12, 469–479 (1982)

Admissibility and Complete Class Results for the Multinomial Estimation Problem with Entropy and Squared Error Loss

AYODELE IGHODARO,* THOMAS SANTNER,* AND LAWRENCE BROWN†

Cornell University

Communicated by A. Cohen

Let $\mathbf{X} \equiv (X_1, ..., X_t)$ have a multinomial distribution based on N trials with unknown vector of cell probabilities $\mathbf{p} \equiv (p_1, ..., p_t)$. This paper derives admissibility and complete class results for the problem of simultaneously estimating $\mathbf{p}$ under entropy loss (EL) and squared error loss (SEL). Let $\mathscr{S}$, $\mathscr{X}$ and $f(\mathbf{x} \mid \mathbf{p})$ denote the $(t-1)$-dimensional simplex, the support of $\mathbf{X}$ and the probability mass function of $\mathbf{X}$, respectively. First it is shown that δ is Bayes w.r.t. EL for prior P if and only if δ is Bayes w.r.t. SEL for P. The admissible rules under EL are proved to be Bayes, a result known for the case of SEL. Let Q denote the class of subsets of $\mathscr{S}$ of the form $T = \bigcup_{j=1}^{k} F_j$ where $k \geqslant 1$ and each F_j is a facet of $\mathscr{S}$ which satisfies: F a facet of $\mathscr{S}$ such that $F \supsetneqq F_j \Rightarrow F \not\subset T$. The minimal complete class of rules w.r.t. EL when $N \geqslant t-1$ is characterized as the class of Bayes rules with respect to priors P which satisfy $P(\mathscr{S}^0) = 1$, $\xi(\mathbf{x}) \equiv \int f(\mathbf{x} \mid \mathbf{p}) P(d\mathbf{p}) > 0$ for all $\mathbf{x}$ in $\{\mathbf{x} \in \mathscr{X}: \sup_{\mathscr{S}^0} f(\mathbf{x} \mid \mathbf{p}) > 0\}$ for some $\mathscr{S}^0$ in Q containing all the vertices of $\mathscr{S}$. As an application, the maximum likelihood estimator is proved to be *admissible* w.r.t. EL when the estimation problem has parameter space $\Theta = \mathscr{S}$ but it is shown to be *inadmissible* for the problem with parameter space $\Theta = (\mathscr{S}$ minus its vertices). This is a severe form of "tyranny of boundary." Finally it is shown that when $N \geqslant t-1$ any estimator δ which satisfies $\delta(\mathbf{x}) > 0$ $\forall \mathbf{x} \in \mathscr{X}$ is admissible under EL if and only if it is admissible under SEL. Examples are given of nonpositive estimators which are admissible under SEL but not under EL and vice versa.

1. INTRODUCTION

Suppose $\mathbf{X} = (X_1, ..., X_t)$ has a multinomial distribution based on N trials with unknown vector of cell probabilities $\mathbf{p} = (p_1, ..., p_t)$ in $\mathscr{S} \equiv \{\mathbf{p} \mid p_i \geqslant 0$

Received March 13, 1981; revised August 10, 1981.

AMS 1980 subject classifications: Primary 62C15; Secondary 62C07, 62G05, 62F10.

Key words and phrases: Simultaneous estimation, multinomial distribution, minimal complete class, entropy loss, square error loss, maximum likelihood estimator, admissible.

* Research supported in part by NSF grant ENG 7906914.

† Research supported in part by NSF grant MCS 7824175.

0047-259X/82/040469–11\$02.00/0

$\forall i, \sum p_i = 1\}$. Denote the mass function of $\mathbf{X}$ by

$$f(\mathbf{x} \mid \mathbf{p}) = N! \prod (p_i^{x_i}/x_i!), \qquad \mathbf{x} \in \mathscr{X}$$

where $\mathscr{X} = \{\mathbf{x} = (x_1,..., x_t) \mid \sum x_i = N, \ \ x_i \geqslant 0$ is integer $\forall i\}$ and our convention is that the range of any product or summation over the integers $\{1,..., t\}$ will be suppressed.

This paper derives admissibility and complete class results for the problem of simultaneously estimating $\mathbf{p}$ under entropy loss (EL); it uses these results to establish relationships between the admissible rules under EL and squared error loss (SEL). For $\mathbf{p}, \mathbf{s} \in \mathscr{S}$, SEL is defined by

$$L_S(\mathbf{p}, \mathbf{a}) = N \sum [p_i - a_i]^2 = N \|\mathbf{p} - \mathbf{a}\|^2 \tag{1.1}$$

and EL by

$$L_E(\mathbf{p}, \mathbf{a}) = N \sum p_i[\ln p_i - \ln a_i] \tag{1.2}$$

where $b \ln 0$ is defined to be 0 and $+\infty$ for $b = 0$ and $b < 0$, respectively, and $\| \cdot \|$ denotes the usual Euclidean norm. The risk functions corresponding to L_E and L_S for an estimator $\boldsymbol{\delta} = \boldsymbol{\delta}(\mathbf{X})$ will be denoted by $R_E(\mathbf{p}, \boldsymbol{\delta})$ and $R_S(\mathbf{p}, \boldsymbol{\delta})$, respectively.

SEL has been used widely, initially because of mathematical convenience and later because of historical momentum, although in some problems symmetry considerations might justify it. However, SEL is inappropriate in problems where it is important to differentiate between zero and positive guesses of $p_i > 0$. Let $\mathscr{S}^+ \equiv \{\mathbf{q} \in \mathscr{S}\colon q_i > 0\}$ and $\partial\mathscr{S} \equiv \{\mathbf{q} \in \mathscr{S}\colon \mathbf{q} \notin \mathscr{S}^+\}$ denote the relative interior and boundary of $\mathscr{S}$, respectively. If $\mathbf{p} \in \mathscr{S}^+$, then a guess $\mathbf{a}^1 \in \partial\mathscr{S}$ is equivalent to any $\mathbf{a}^2 \in \partial\mathscr{S}$ satisfying $\|\mathbf{p} - \mathbf{a}^1\|^2 = \|\mathbf{p} - \mathbf{a}^2\|^2$. In contrast, EL differentiates between positive and zero guesses of $p_i > 0$; $L_E(\mathbf{p}, \mathbf{a}) = +\infty$ for all $\mathbf{p} \in \mathscr{S}^+$ and $\mathbf{a} \in \partial\mathscr{S}$. Alternatively, Akaike [1, 2] motivates EL by the premise that the reason for estimating $\mathbf{p}$ by $\mathbf{a}$ is to base decisions about $f(\mathbf{x} \mid \mathbf{p})$ on $f(\mathbf{x} \mid \mathbf{a})$ since $L_E(\mathbf{p}, \mathbf{a})$ is the Kullback–Leibler mean information for discrimination between $f(\mathbf{x} \mid \mathbf{p})$ and $f(\mathbf{x} \mid \mathbf{a})$ [8]. Asymptotically $L_E(\mathbf{p}, \mathbf{a})$ is roughly the negative of the logarithm of the probability of observing a sample distribution closely approximated by $f(\mathbf{x} \mid \mathbf{p})$ when a large number of observations are independently drawn from $f(\mathbf{x} \mid \mathbf{a})$ [9]. See [1, 2] and the references therein for additional motivation and examples.

Since the action space $\mathscr{A} = \mathscr{S}$ is a convex compact subset of Euclidean t-space, R^t, and both $L_S(\mathbf{p}, \mathbf{a})$ and $L_E(\mathbf{p}, \mathbf{a})$ are convex in $\mathbf{a}$ $\forall \mathbf{p} \in \mathscr{S}$, the nonrandomized decision rules form an essentially complete class [6]. Throughout this paper attention will be restricted to nonrandomized estimators $\delta: \mathscr{X} \to \mathscr{S}$ except for the class $\mathscr{D}_c^0$ defined in the proof of Theorem 2.1 where the convexity of the corresponding class of risk functions, $\Gamma(\mathscr{D}_c^0)$, will require the inclusion of randomized rules.

Let $L(\cdot, \cdot)$ denote an arbitrary loss function on $\mathscr{S} \times \mathscr{A}$ and $R(\mathbf{p}, \delta)$ be the risk of an estimator δ at $\mathbf{p} \in \mathscr{S}$ under $L(\cdot, \cdot)$. An estimator δ^B is Bayes under $L(\cdot, \cdot)$ with respect to (w.r.t.) a prior P on $\mathscr{S}$ means

$$\int R(\mathbf{p}, \delta^B)\, P(d\mathbf{p}) \leqslant \int R(\mathbf{p}, \delta)\, P(d\mathbf{p}) \tag{1.3}$$

for every δ. For notational simplicity the domain of integration is suppressed throughout when it is $\mathscr{S}$. In particular, it is well known that a Bayes rule under L_S w.r.t. P is given by

$$\mathbf{v}(\mathbf{x}) = \int \mathbf{p} P(d\mathbf{p} \mid \mathbf{x}) \tag{1.4}$$

where $P(\cdot \mid \mathbf{x})$ denotes the posterior distribution of P given $\mathbf{x}$ or an arbitrary probability measure according to whether $\int f(\mathbf{x} \mid \mathbf{p})\, P(d\mathbf{p}) > 0$ or $= 0$, respectively.

We conclude this section by outlining the remainder of the paper. Section 2 proves that (1.4) is also Bayes under L_E w.r.t. P; it shows that every admissible rule under L_E is Bayes. Section 3 characterizes the minimal complete class under L_E. As an application, the maximum likelihood estimator (mle) of $\mathbf{p}$ is proved to be admissible for the problem with parameter space $\mathscr{S}$ and inadmissible for the problem with parameter space $\mathscr{S}$ minus its vertices. This result is an extreme case of "tyranny of the vertices." Reference [7] characterized the behavior of the mle under SEL by the same language; however its behavior is quite different under SEL since the mle is admissible for SEL even over the parameter space $\mathscr{S}^+$ (see Section 3). The final section details the relationship between the admissible rules under SEL and EL.

2. Admissible and Bayes Rules under EL

Let P be a prior on $\mathscr{S}$. The Bayes risk of the estimator δ under L_E is

$$r_E(P, \delta) \equiv \int \sum_{\mathscr{X}} L_E(\mathbf{p}, \delta(\mathbf{x}))\, f(\mathbf{x} \mid \mathbf{p})\, P(d\mathbf{p}).$$

If $P(\cdot \mid \mathbf{x})$ is defined as in Section 1, then

$$r_E(P, \delta) = \sum_{\mathscr{X}} E[L_E(\mathbf{p}, \delta(\mathbf{x})) \mid \mathbf{x}]\, \xi(\mathbf{x})$$

where $\xi(\mathbf{x}) = \int f(\mathbf{x} \mid \mathbf{p}) P(d\mathbf{p})$ is the marginal distribution of $\mathbf{X}$ and

$$E[L_E(\mathbf{p}, \boldsymbol{\delta}(\mathbf{x})) \mid \mathbf{x}] = \int \sum p_i \ln(p_i/\delta_i(\mathbf{x}))\, P(d\mathbf{p} \mid \mathbf{x}). \tag{2.1}$$

From (2.1) a Bayes rule is any $\boldsymbol{\delta}$ which maximizes the multinomial log-likelihood kernel $\sum \nu_i \ln \delta_i$ where $\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{x})$ is given by (1.4). It is well known that $\boldsymbol{\delta} = \boldsymbol{\nu}(\mathbf{x})$ is the required maximum and hence is Bayes under L_E w.r.t. P.

Remark 2.1. The argument above also shows that $\boldsymbol{\nu}(\mathbf{x})$ is unique Bayes under L_S or $L_E \Leftrightarrow \xi(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathscr{X}$.

Since good decision rules are usually (extended) Bayes the equivalence of the Bayes rules established above suggests that the classes of admissible rules under the two losses are related. We study this relationship by first establishing that admissible rules under SEL or EL must be Bayes with respect to SEL or EL, respectively.

In both cases the action space $\mathscr{A} = \mathscr{S}$ and the parameter space $\Theta = \mathscr{S}$ are convex compact subsets of R^t. Since $L_S(\mathbf{p}, \mathbf{a})$ is bounded, convex, and continuous in $\mathbf{a}$ for each $\mathbf{p}$, every admissible rule under SEL is Bayes with respect to SEL [10, Theorem 3.20]. Even though EL is not bounded, Theorem 2.1 shows the above result holds for L_E by a generalization in [5] of Wald's theorem. For any $A \subset \mathscr{S}$ let $\mathscr{X}(A) \equiv \{x \in \mathscr{X}: \sup_A r(\mathbf{x} \mid \mathbf{p}) > 0\}$ denote the set of outcomes which can be "seen" under A. For arbitrary sets E and F let $E \backslash F \equiv \{w \in E: w \notin F\}$.

THEOREM 2.1. *Suppose $\boldsymbol{\delta}^0 = \boldsymbol{\delta}^0(\mathbf{x})$ is admissible for the problem $\mathscr{P} \equiv (\mathscr{X}, \Theta = \mathscr{S}, \mathscr{A} = \mathscr{S}, L_E)$, then $\boldsymbol{\delta}^0$ is Bayes with respect to $\mathscr{D} \equiv \{\boldsymbol{\delta}: \mathscr{X} \to \mathscr{S}\}$ for some prior P such that $P(\mathscr{S}^0) = 1$ where $\mathscr{S}^0 \equiv \{\mathbf{p}: R_E(\mathbf{p}, \boldsymbol{\delta}^0) < \infty\}$.*

Proof. $\boldsymbol{\delta}^0$ must be admissible for the problem $\mathscr{P}^0 \equiv (\mathscr{X}^0, \Theta = \mathscr{S}^0, \mathscr{A} = \mathscr{S}, L_E)$ where $\mathscr{X}^0 \equiv \mathscr{X}(\mathscr{S}^0)$; if not, there exists a $\bar{\boldsymbol{\delta}}$ such that $R_E(\mathbf{p}, \bar{\boldsymbol{\delta}}) \leqslant R_E(\mathbf{p}, \boldsymbol{\delta}^0)\ \forall \mathbf{p} \in \mathscr{S}^0$ with $<$ for some $\mathbf{p} \in \mathscr{S}^0 \Rightarrow \bar{\boldsymbol{\delta}}$ is better than $\boldsymbol{\delta}^0$ for the problem $\mathscr{P}$ since $R_E(\mathbf{p}, \boldsymbol{\delta}^0) = +\infty\ \forall \mathbf{p} \in \mathscr{S} \backslash \mathscr{S}^0$.

Let $\mathscr{D}^0 \equiv \{\boldsymbol{\delta}: \mathscr{X}^0 \to \mathscr{S}\}$, $0 < c < 1$ and $\mathscr{D}_c^0 \equiv \{\boldsymbol{\delta}$ a randomized rule on $\mathscr{X}^0$: $\delta_i(\mathbf{x}) \equiv \int a_i \delta(d\mathbf{a}, \mathbf{x}) \geqslant c\boldsymbol{\delta}^0(\mathbf{x})\ \forall \mathbf{x} \in \mathscr{X}\}$ where a randomized rule on $\mathscr{X}^0$ is a mapping from $\mathscr{X}^0$ to the set of all probability measures on $(\mathscr{A} = \mathscr{S}, \mathscr{B}(\mathscr{A}))$ and $\mathscr{B}(\mathscr{A})$ denote the Borel σ-field in $\mathscr{A}$. For every $A \in \mathscr{B}(\mathscr{A})$, $\delta(A, \cdot)$ is assumed measurable in $(\mathscr{X}^0, \mathscr{B}(\mathscr{X}^0))$ where $\mathscr{B}(\mathscr{X}^0)$ is the power set of $\mathscr{X}^0$. It is straightforward to verify that $\forall \boldsymbol{\delta} \in \mathscr{D}_c^0$, $R_E(\cdot, \boldsymbol{\delta})$ is a continuous, real valued function on $\mathscr{S}^0$, that the risk set $\Gamma(\mathscr{D}_c^0) \equiv \{R_E(\cdot, \boldsymbol{\delta}): \boldsymbol{\delta} \in \mathscr{D}_c^0\}$ is convex and that $\tilde{\Gamma}(\mathscr{D}_c^0) \equiv \{h: \mathscr{S}^0 \to [0, \infty]: \exists \boldsymbol{\delta} \in \mathscr{D}_c^0$ with $R_E(\mathbf{p}, \boldsymbol{\delta}) \leqslant h(\mathbf{p}) \forall \mathbf{p} \in \mathscr{S}^0\}$ is closed in $\prod_{\mathscr{S}^0} [0, \infty]$ under the topology defined by pointwise convergence of sequence of functions. By Theorem 3.5 of [5], $\boldsymbol{\delta}^0$ is Bayes relative to $\mathscr{D}_c^0$ for some prior P such that $P(\mathscr{S}^0) = 1$. We claim $\boldsymbol{\delta}^0$ is Bayes with respect to P relative to $\mathscr{D}^0$; if not, then $\exists \bar{\boldsymbol{\delta}} \in \mathscr{D}^0 \backslash \mathscr{D}_c^0$ and $\mathbf{x}^* \in \mathscr{S}^0 \ni \xi(\mathbf{x}^*) > 0$ and $\phi(\bar{\boldsymbol{\delta}}(\mathbf{x}^*)) > \phi(\boldsymbol{\delta}^0(\mathbf{x}^*))$ where $\phi(\boldsymbol{\delta}) = \int_{\mathscr{S}^0} \sum_i p_i \ln \delta_i P(d\mathbf{p} \mid \mathbf{x}^*)$. Let $\boldsymbol{\delta}^\lambda(\mathbf{x}) \equiv \lambda \bar{\boldsymbol{\delta}}(\mathbf{x}) + (1-\lambda)\, \boldsymbol{\delta}^0(\mathbf{x})$ for $\lambda \in (0, 1)$. By concavity of $\phi(\boldsymbol{\delta})$, $\phi(\boldsymbol{\delta}^\lambda(\mathbf{x}^*)) \geqslant \lambda \phi(\bar{\boldsymbol{\delta}}(\mathbf{x}^*)) + (1-\lambda)\, \phi(\boldsymbol{\delta}^0(\mathbf{x}^*)) > \phi(\boldsymbol{\delta}^0(\mathbf{x}))$ $\forall \lambda \in (0, 1)$. Choose λ sufficiently small so that $\forall i \in \{1,..., t\}$, $\delta_i^\lambda(\mathbf{x}) = \delta_i^0(\mathbf{x}) + \lambda(\bar{\delta}_i(\mathbf{x}) - \delta_i^0(\mathbf{x})) > c\delta_i^0(\mathbf{x})$ $\forall \mathbf{x} \in \mathscr{X}^0$ since $c \in (0, 1) \Rightarrow \boldsymbol{\delta}^\lambda \in \mathscr{D}_c^0 \Rightarrow \boldsymbol{\delta}^0$ is not Bayes with respect to P relative to $\mathscr{D}_c^0$. Since P has support in $\mathscr{S}^0$, we have $\forall \boldsymbol{\delta} \in \mathscr{D} \equiv \{\boldsymbol{\delta}: \mathscr{X} \to \mathscr{S}\}$ that the Bayes risk relative to P is

$$\begin{aligned} r_E(P, \boldsymbol{\delta}) &= \sum_{\mathscr{X}} \int_{\mathscr{S}} L_E(p, \boldsymbol{\delta}(\mathbf{x}))\, P(d\mathbf{p} \mid \mathbf{x})\, \xi(\mathbf{x}) \\ &= \sum_{\mathscr{X}^0} \int_{\mathscr{S}^0} L_E(\mathbf{p}, \boldsymbol{\delta}(\mathbf{x}))\, P(d\mathbf{p} \mid \mathbf{x})\, \xi(\mathbf{x}) \\ &\geqslant r_E(P, \boldsymbol{\delta}^0) \end{aligned}$$

and the proof is completed. Section 4 will use Theorem 2.1 to establish the relationship between the admissible rules under EL and SEL.

3. The Minimal Complete Class under EL

Throughout Sections 3 and 4, $\boldsymbol{\delta}^0$ denotes a nonrandomized estimator for the problem $\mathscr{P} = (\mathscr{X},\ \Theta = \mathscr{S},\ \mathscr{A} = \mathscr{S}, L_E)$ and $\mathscr{S}^0 \equiv \{\mathbf{p} \in \mathscr{S}: R_E(p, \boldsymbol{\delta}^0) < \infty\}$. We begin by describing the structure of $\mathscr{S}^0$ for admissible $\boldsymbol{\delta}^0$ in Lemmas 3.1 and 3.2 and then characterize the minimal complete class for $\mathscr{P}$ in Theorems 3.1 and 3.2.

Definition 3.1. $F \subset S$ is a *facet* of $\mathscr{S}$ means either $F = \mathscr{S}$ or $\exists l \in \{1,..., t-1\}$ and integers $1 \leqslant i_1 < i_2 < \cdots < i_l \leqslant t$ such that $\mathbf{p} \in F \Leftrightarrow p_{i_j} = 0$ for $j = 1,..., l$.

For any facet F of $\mathscr{S}$ let $I(F)$ be defined by $\mathbf{p} \in F \Leftrightarrow p_i = 0$ $\forall i \in I(F)$; set $I(\mathscr{S}) = \varnothing$. Let Q denote the collection of subsets T of $\mathscr{S}$ which can be written in the form $T = \bigcup_{i=1}^k F_i$ for some positive integer k where $F_1,..., F_k$ are facets of $\mathscr{S}$ such that $\forall i = 1,..., k$ if F is a facet of $\mathscr{S}$ satisfying $F \supsetneqq F_i$, then $F \not\subset T$.

Lemma 3.1. *$\mathscr{S}^0$ is in Q for any admissible $\boldsymbol{\delta}^0$.*

Proof. Clearly $\mathscr{S}^0 \neq \varnothing$ since $\boldsymbol{\delta}^0$ is admissible. If $\mathscr{S}^0 = \mathscr{S}$, then the result is true. If $\mathscr{S}^0 \neq \mathscr{S}$, then suppose $\mathscr{S}^0 \cap F^+ \neq \varnothing$ for some facet F where $F_+ = \{\mathbf{p} \in F: p_j > 0$ for $j \notin I(F)\}$ is the relative interior of F. A straightforward argument proves $F \subset \mathscr{S}^0$.

LEMMA 3.2. *$\{\mathbf{e}_1,..., \mathbf{e}_t\} \subset \mathscr{S}^0$ for any admissible $\boldsymbol{\delta}^0$ where $\mathbf{e}_i$ is the unit vector with* 1 *in the* ith *component and* 0 *in the remaining components.*

Proof. Suppose w.l.o.g. $\mathbf{e}_1 \notin \mathscr{S}^0$, then $R_E(\mathbf{e}_1, \boldsymbol{\delta}^0) = +\infty \Leftrightarrow \delta_1^0(N\mathbf{e}_1) = 0$. Define a new estimator $\bar{\boldsymbol{\delta}}$ by

$$\begin{aligned} \bar{\boldsymbol{\delta}}(\mathbf{x}) &\equiv \mathbf{e}_1, & \mathbf{x} &= N\mathbf{e}_1 \\ &\equiv \boldsymbol{\delta}^0(\mathbf{x}), & \mathbf{x} &\neq N\mathbf{e}_1, \end{aligned}$$

then $0 = R_E(\mathbf{e}_1, \bar{\boldsymbol{\delta}}) < \infty = R_E(\mathbf{e}_1, \boldsymbol{\delta}^0)$, $R_E(\mathbf{p}, \bar{\boldsymbol{\delta}}) \leqslant \infty = R_E(\mathbf{p}, \boldsymbol{\delta}^0)$ for all $\mathbf{p}$ with $0 < p_1 < 1$, and $R(\mathbf{p}, \bar{\boldsymbol{\delta}}) = R(\mathbf{p}, \boldsymbol{\delta}^0)$ for all $\mathbf{p}$ with $p_1 = 0$. Therefore $\bar{\boldsymbol{\delta}}$ dominates $\boldsymbol{\delta}^0$ which contradicts the admissibility of $\boldsymbol{\delta}^0$ and completes the proof. One consequence of Lemma 3.2 is that for admissible $\boldsymbol{\delta}^0$ *every* facet of $\mathscr{S}$ has a nonempty intersection with $\mathscr{S}^0$.

Theorems 3.1 and 3.2 below characterize an estimator $\boldsymbol{\delta}^0$ as admissible when $N \geqslant t-1$ if there exists a set T in Q containing the vertices and a prior P with support in T so that (1) $\boldsymbol{\delta}^0$ is Bayes w.r.t. P and (2) $\mathscr{X}(T)$ is the support of $\xi(\cdot)$. The symbol Supp P will denote the support of P in the following.

THEOREM 3.1. *Every admissible $\boldsymbol{\delta}^0$ is Bayes w.r.t. some prior P with support in $\mathscr{S}^0$ which satisfies $\{\mathbf{x} \in \mathscr{X}: \xi(\mathbf{x}) > 0\} \subset \mathscr{X}(\mathscr{S}^0)$. Furthermore if $N \geqslant t-1$, then $\{\mathbf{x} \in \mathscr{X}: \xi(\mathbf{x}) > 0\} = \mathscr{X}(\mathscr{S}^0)$.*

Proof. Theorem 2.1 guarantees $\boldsymbol{\delta}^0$ is Bayes w.r.t. some prior P with support in $\mathscr{S}^0$. If $\mathbf{x} \in \mathscr{X} \backslash \mathscr{X}(\mathscr{S}^0)$, then $f(\mathbf{x} \mid \mathbf{p}) = 0$ $\forall\, \mathbf{p} \in \mathscr{S}^0 \Rightarrow \xi(\mathbf{x}) = 0$ and so $\{\mathbf{x} \in \mathscr{X}: \xi(\mathbf{x}) > 0\} \subset \mathscr{X}(\mathscr{S}^0)$. Now suppose $N \geqslant t-1$; if $\{F_j\}_{j=1}^J$ is the collection of facets such that $P(F_j^+) > 0$ for $j = 1,..., J$ where F_j^+ denotes the relative interior of F_j, then Theorem A.1 in the Appendix proves that $\mathscr{S}^0 = \bigcup_{j=1}^J F_j$. If $\mathbf{x} \in \mathscr{X}(\mathscr{S}^0)$, then $\exists\, \mathbf{p} \in \mathscr{S}^0$ such that $f(\mathbf{x} \mid \mathbf{p}) > 0$. Assume w.l.o.g. that $x_i > 0$ for $i = 1,..., r$ and $= 0$ for $i = r+1,..., t$ and that $\mathbf{p} \in F_1$. Then $p_i > 0$ for $i = 1,..., r$ and $\mathbf{p}' \in F_1^+ \Rightarrow p_i' > 0$ for $i = 1,..., r$. Hence $\xi(\mathbf{x}) > 0$ and the proof is finished.

THEOREM 3.2. *Suppose $N \geqslant t-1$ and $T \in Q$ contains the vertices $\mathbf{e}_1,..., \mathbf{e}_t$. If P is a probability measure on $\mathscr{S}$ satisfying $P(T) = 1$ and $\xi(\mathbf{x}) \equiv \int f(\mathbf{x} \mid \mathbf{p})\, P(d\mathbf{p}) > 0$ $\forall\, \mathbf{x} \in \mathscr{X}(T)$, then any Bayes estimator $\boldsymbol{\delta}$ w.r.t. P is admissible under EL.*

Proof. Suppose $\boldsymbol{\delta}'$ satisfies $R(\mathbf{p}, \boldsymbol{\delta}') \leqslant R(\mathbf{p}, \boldsymbol{\delta})$ $\forall\, \mathbf{p} \in \mathscr{S}$, then $\boldsymbol{\delta}'$ is also Bayes w.r.t. P. By Remark 2.1 the Bayes estimator is unique for xsuch that $\xi(\mathbf{x}) > 0$ and hence $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}'(\mathbf{x})$ $\forall\, \mathbf{x} \in \mathscr{X}(T)$. So $R_E(\mathbf{p}, \boldsymbol{\delta}) = R_E(\mathbf{p}, \boldsymbol{\delta}')$ $\forall\, \mathbf{p} \in T$. Let $\{F_j\}$ be the set of facets of $\mathscr{S}$ such that $P(F_j^+) > 0$ $\forall\, j$; it is easy to see that $\bigcup F_j \subset T$. Furthermore, Theorem A.1 guarantees $\{\mathbf{p} \in \mathscr{S}:$

$R_E(\mathbf{p}, \boldsymbol{\delta}') < \infty\} = \bigcup F_j = \{\mathbf{p} \in \mathscr{S}: R_E(\mathbf{p}, \boldsymbol{\delta}) < \infty\}$ since $N \geqslant t-1$ and hence $R_E(\mathbf{p}, \boldsymbol{\delta}') = \infty = R_E(\mathbf{p}, \boldsymbol{\delta})\ \forall\, p \in \mathscr{S} \backslash \mathscr{F}$ so that $\boldsymbol{\delta}$ is admissible.

Remark 3.1. Any estimator $\boldsymbol{\delta}$ satisfying $\boldsymbol{\delta}(N\mathbf{e}_i) = \mathbf{e}_i$ for $i = 1, ..., t$ is admissible under EL since $\mathbf{d}$ is Bayes with respect to the prior P putting mass $1/t$ at each point of $T = \{\mathbf{e}_1, ..., \mathbf{e}_t\}$.

Remark 3.2. The proof of Theorem 3.2 shows that if $\xi(\mathbf{x}) > 0\ \forall\, \mathbf{x} \in \mathscr{X}$, then any Bayes rule is admissible for all $N \geqslant 1$.

The maximum likelihood estimator $\boldsymbol{\delta}^{\mathrm{mle}}(\mathbf{X}) = \mathbf{X}/N$ is admissible under SEL [3, 4, 7]. As is well known, this admissibility is related to the fact that the risk

$$R_S(\mathbf{p}, \boldsymbol{\delta}^{\mathrm{mle}}) = 1 - \|\mathbf{p}\|^2$$

is small when $\mathbf{p}$ is near vertices. Johnson [7] refers to this behavior as the "tyranny of the boundary." For the case of EL the boundary exerts a much stronger tyranny as evidenced by the following result.

THEOREM 3.3. *Under EL,* $\boldsymbol{\delta}^{\mathrm{mle}}$ *is*

(i) *admissible when* $\Theta = \mathscr{S}$ *for* $N \geqslant t-1$,

(ii) *inadmissible when* $\Theta = \mathscr{S} \backslash \{\mathbf{e}_1, ..., \mathbf{e}_t\}$ *for* $N \geqslant 1$.

Proof. Part (i) follows from Remark 3.1. To prove part (ii) fix $\mathbf{p} \in \mathscr{S} \backslash \{\mathbf{e}_1, ..., \mathbf{e}_t\}$; assume w.l.o.g. that $p_i > 0$ for $i = 1, 2$ since $\mathbf{p}$ has at least two nonzero components. Choose $\mathbf{x}^* = \mathbf{x}^*(\mathbf{p})$ so that $P_{\mathbf{p}}[\mathbf{X} = \mathbf{x}^*] > 0$ and $\delta_1^{\mathrm{mle}}(\mathbf{x}^*) = 0 \Rightarrow R(\mathbf{p}, \boldsymbol{\delta}^{\mathrm{mle}}) = \infty$. $\boldsymbol{\delta}^{\mathrm{mle}}$ is clearly dominated by the constant estimator $\boldsymbol{\delta}^c(\mathbf{x}) = (1/t, ..., 1/t)\ \forall\, \mathbf{x} \in \mathscr{X}$.

Remark 3.3. In contrast, $\boldsymbol{\delta}^{\mathrm{mle}}$ is admissible under SEL when $\Theta = \mathscr{S} \backslash \{\mathbf{e}_1, ..., \mathbf{e}_t\}$ and even when $\Theta = \mathscr{S}^+$. See [4, especially Proposition 1.7].

The admissibility of $\boldsymbol{\delta}^{\mathrm{mle}}$ under entropy loss can therefore be dismissed as being artificial.

The last section shows the admissible rules under EL and SEL coincide for estimators which are always positive but each class contains (nonpositives) estimators not contained by the other.

4. RELATIONSHIP BETWEEN ADMISSIBLE CLASSES UNDER SEL AND EL

The development below first studies estimators which are positive for all $\mathbf{x} \in \mathscr{X}$ via Lemmas 4.1 and 4.2; it then considers estimators which allow zero guesses.

LEMMA 4.1. *Let* δ^0 *be admissible under* EL. *Then there is an estimator* δ' *which is admissible under* SEL *such that* $\delta^0(\mathbf{x}) = \delta'(\mathbf{x})$ $\forall \mathbf{x} \in \mathscr{X}(\mathscr{S}^0)$ *when* $N \geqslant t - 1$.

Proof. From Theorem 3.1, δ is unique Bayes on $\mathscr{X}(\mathscr{S}^0)$ relative to some prior P on $\mathscr{S}^0$. Let P_1 be a prior with marginal distribution $\xi_1(\mathbf{x}) \equiv \int f(\mathbf{x} \mid \mathbf{p})$ $P_1(d\mathbf{p} \mid \mathbf{x}) > 0$ $\forall \mathbf{x} \in \mathscr{X}$. Define $\delta'(\mathbf{x}) = \delta(\mathbf{x})$ for $\mathbf{x} \in \mathscr{X}(\mathscr{S}^0)$ and $= \int \mathbf{p} P_1(d\mathbf{p} \mid \mathbf{x})$ for $\mathbf{x} \notin \mathscr{X}(\mathscr{S}^0)$. Then δ' is admissible for SEL by [4].

LEMMA 4.2. *Suppose* $N \geqslant t - 1$ *and* δ *is admissible under* SEL. *If* $\delta(\mathbf{x}) > 0$ $\forall \mathbf{x} \in \mathscr{X}$, *then* δ *is unique Bayes on all of* $\mathscr{X}$ *for some prior* P. *(Equivalently* δ *is Bayes for a prior* P *having* $\xi(\mathbf{x}) > 0$ $\forall \mathbf{x} \in \mathscr{X}$.)

Proof. Any δ admissible under SEL is Bayes from Section 2; let P be a prior such that δ is Bayes w.r.t. P. Suppose $\exists \mathbf{x}^* \in \mathscr{X} \ni \xi(\mathbf{x}^*) = 0$; w.l.o.g. assume $x_i^* > 0$ for $i = 1,..., s$ and $x_i^* = 0$ for $i = s + 1,..., t$ $(1 \leqslant s \leqslant t)$. Then P has support in $T \equiv \{\mathbf{p} \in \mathscr{S}: p_i = 0 \text{ for some } i = 1,..., s\} = \bigcup F_j$ where $F_j, j = 1,..., 2^s - 1$, have index sets $I(F_j)$, $j = 1,..., 2^s - 1$, which are the distinct nonempty subsets of $\{1,..., s\}$. Let $\mathscr{F} \equiv \{F \subset T\colon F \text{ a facet, } P(F) > 0\}$. Partially order the sets $I(F)$, $F \in \mathscr{F}$ by inclusion and let I_0 be a minimal set in this ordering. Now $I_0 \neq \varnothing$ since $I_0 = \varnothing \Rightarrow \exists F \in \mathscr{F} \ni I(F) = I_0 = \varnothing$ which contradicts $F \in T$; assume w.l.o.g. that $I_0 = \{r + 1,..., s\}$ with $1 \leqslant r + 1 \leqslant s$. Fix $\hat{\mathbf{x}} \in \mathscr{X}$ so that $\hat{x}_i = 0$ for $i = r + 1,..., s$ and $\hat{x}_i > 0$ otherwise; this is possible since $N \geqslant t - 1$. Clearly $\xi(\hat{\mathbf{x}}) > 0$ since $\xi(\hat{\mathbf{x}}) = 0$ (together with $\xi(\mathbf{x}^*) = 0) \Rightarrow \text{Supp}(P) \subset \{\mathbf{p} \in \mathscr{S}: p_i = 0 \text{ for some } 1 \leqslant i \leqslant r\} \Rightarrow I(F) \subset \{1,..., r\}$ $\forall F \in \mathscr{F}$ which is impossible since $\exists F' \in \mathscr{F} \ni I(F') = I_0 = \{r + 1,..., s\}$. To complete the proof it suffices to show $p_i = 0$ for $i = r + 1,..., s$ whenever $f(\hat{\mathbf{x}} \mid \mathbf{p}) > 0$ and $\mathbf{p} \in F \in \mathscr{F}$ and hence $\delta_i(\hat{\mathbf{x}}) = 0$ for $i = r + 1,..., s$ which contradicts the assumption $\delta(\mathbf{x}) > 0$ $\forall \mathbf{x} \in \mathscr{X}$. But $f(\hat{\mathbf{x}} \mid \mathbf{p}) > 0 \Rightarrow p_i > 0$ for $1 \leqslant i \leqslant r$ and $s + 1 \leqslant i \leqslant t \Rightarrow I(F) \subset \{r + 1,..., s\} = I_0 \Rightarrow I(F) = I_0$ since I_0 is minimal $\Rightarrow p_i = 0$ for $r + 1,..., s$ and hence the desired result.

Remark. One consequence of Lemmas 4.1 and 4.2 is that when $N \geqslant t - 1$, a positive estimator δ is admissible under SEL if and only if δ is admissible under EL. Suppose $\delta(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathscr{X}$; then $\mathscr{S}^0 \equiv \{\mathbf{p} \in \mathscr{S} \mid R_E(\mathbf{p}, \delta) < \infty\} = \mathscr{S}$. Hence if δ is admissible under SEL, then δ is Bayes for some prior P satisfying $\xi(\mathbf{x}) > 0$ $\forall \mathbf{x} \in \mathscr{X} = \mathscr{X}(\mathscr{S}^0)$ and hence δ is admissible under EL by Theorem 3.2. Conversely, if δ is admissible under EL, then $\delta(x) = \delta'(\mathbf{x})$ $\forall \mathbf{x} \in \mathscr{X}(\mathscr{S}^0) = \mathscr{X}$ where δ' is admissible under SEL so that δ is admissible under SEL.

There exist (nonpositive) estimators δ which are admissible under SEL but not EL and vice versa. However, in the latter case, a δ admissible under EL must coincide on $\mathscr{X}(\mathscr{S}^0)$ with an estimator which is admissible under SEL by Lemma 4.1. Intuitively, a (nonpositive) estimator δ admissible under SEL for which $\mathscr{S}^0 = \bigcup_{i=1}^k F_j$ has at least one facet, F_1 say, disjoint from the remaining ones is a candidate to be inadmissible under EL. In this case $\mathscr{X}(\mathscr{S}^0) = \bigcup_{j=1}^k \mathscr{X}(F_j)$ where $\mathscr{X}(F_1) \cap (\bigcap_{j=2}^k \mathscr{X}(F_j)) = \varnothing$. If $\bar{\delta}$ can be constructed to modify $\boldsymbol{\delta}$ at one or more $\mathbf{x} \in \mathscr{X}(F_1)$ to decrease $R_E(\mathbf{p}, \boldsymbol{\delta})$ for $\mathbf{p} \in F_1$, then since $R_E(\mathbf{p}, \boldsymbol{\delta}) = +\infty$ for $\mathbf{p} \notin \mathscr{S}^0$, the resulting estimator can improve $\boldsymbol{\delta}$. Example 4.1 illustrates this phenomenon.

EXAMPLE 4.1. Consider the estimator for the trinomial problem defined by

$$\begin{aligned} \boldsymbol{\delta}(x_1, x_2, x_3) &= (0, 1/2, 1/2), && x_1 = 0 \\ &= (1/3, 1/3, 1/3), && x_1 > 0. \end{aligned}$$

$\boldsymbol{\delta}$ is admissible under SEL [4]. Consider $\bar{\boldsymbol{\delta}}$ defined by

$$\begin{aligned} \bar{\boldsymbol{\delta}}(x_1, x_2, x_3) &= (0, 1/2, 1/2), && x_1 = 0 \\ &= (1/3, 1/3, 1/3), && 0 < x_1 < N \\ &= (1, 0, 0), && x_1 = N. \end{aligned}$$

Then, $p_1 = 0 \Rightarrow R_E(\mathbf{p}, \boldsymbol{\delta}) = R_E(\mathbf{p}, \bar{\boldsymbol{\delta}}) < \infty$, $p_1 = 1 \Rightarrow 0 = R_E(\mathbf{p}, \bar{\boldsymbol{\delta}}) < R_E(\mathbf{p}, \boldsymbol{\delta}) = N \ln(3)$, and $0 < p_1 < 1 \Rightarrow R_E(\mathbf{p}, \bar{\boldsymbol{\delta}}) = R_E(\mathbf{p}, \boldsymbol{\delta}) = \infty$; hence $\boldsymbol{\delta}$ is inadmissible under EL.

Conversely, an estimator $\boldsymbol{\delta}$ which guards against the states of nature $\mathbf{e}_1, \ldots, \mathbf{e}_t$ by guessing $\boldsymbol{\delta}(N\mathbf{e}_i) = \mathbf{e}i$ for $i = 1, \ldots, t$ will be admissible under EL. However $\boldsymbol{\delta}$ can be inadmissible under SEL by making counter intuitive guesses at other $\mathbf{x}$ values.

EXAMPLE 4.2. By Remark 3.1 the estimator defined by

$$\begin{aligned} \boldsymbol{\delta}(\mathbf{x}) &\equiv \mathbf{e}_i, && \mathbf{x} = N\mathbf{e}_i \quad (1 \leqslant i \leqslant t) \\ &\equiv \mathbf{e}_1 && \text{otherwise} \end{aligned}$$

is admissible under EL; it is inadmissible under SEL by Theorem 3.2 of [4].

APPENDIX

Let P be a given prior on $\mathscr{S}$ and $\{F_j\}_{j-1}^J$ the collection of facets such that $P(F_j^+) > 0$ where F_j^+ is the relative interior of F_j; denote $T \equiv \bigcup_{j=1}^J F_j$.

THEOREM A.1. *Let* $\boldsymbol{\delta}^0$ *be any Bayes rule w.r.t.* P *under* EL. *Then*

$T \subset \mathscr{S}^0 \equiv \{\mathbf{p} \in \mathscr{S}\colon R_E(\mathbf{p}, \boldsymbol{\delta}^0) < \infty\}$ for all $N \geqslant 1$. Furthermore $T = \mathscr{S}^0$ when $N \geqslant t-1$.

Proof. Suppose $\mathbf{p}^* \in T$, then assume w.l.o.g. that $\mathbf{p}^* \in F_1$. Fix $\mathbf{x}^* \in \mathscr{X}$ such the $f(\mathbf{x}^* \mid \mathbf{p}^*) > 0$; then $x_i^* = 0$ whenever $p_i^* = 0$. Consequently, $f(\mathbf{x}^* \mid \mathbf{p}) > 0$ $\forall \mathbf{p} \in F_1$ since $I(F_1) \subset \{i\colon p_i^* = 0\}$. Then $\xi(\mathbf{x}^*) \geqslant \int_{F_1^+} f(\mathbf{x}^* \mid \mathbf{p})\, P(d\mathbf{p}) > 0$ and $\forall i \notin I(F_1)$, $\delta_i^0(\mathbf{x}^*) \geqslant \int_{F_1^+} p_i f(\mathbf{x}^* \mid \mathbf{p})\, P(d\mathbf{p} \mid \mathbf{x}^*) > 0$ since $P(F_1^+) > 0$ and $p_i f(\mathbf{x}^* \mid \mathbf{p}) > 0$ $\forall \mathbf{p} \in F_1^+$ and $i \notin I(F_1)$. Since this is true for any $\mathbf{x}^*$ for which $f(\mathbf{x}^* \mid \mathbf{p}^*) > 0$, it follows that $\mathbf{p}^* \in \mathscr{S}^0$.

Conversely, if $N \geqslant t-1$, suppose $\mathbf{p}^* \in \mathscr{S}^0$ but $\mathbf{p}^* \notin T$. Assume w.l.o.g. $p_i^* > 0$ for $i = 1,..., q$ and $= 0$ for $i = q+1,..., t$ $(1 \leqslant q \leqslant t)$. Then $I(F_j) \cap \{1,..., q\} \neq \varnothing$ $\forall j = 1,..., J$ since $p^* \notin T$. Partially order the sets $I(F_j) \cap \{1,..., q\}$, $j = 1,..., J$, by inclusion and let I_0 be a minimal set; then $I_0 \neq \varnothing$. Assume w.l.o.g. that $I_0 = \{1,..., r\} = I(F_j) \cap \{1,..., q\}$ for $j = 1,..., \lambda$ but $I_0 \neq I(F_j) \cap \{1,..., q\}$ for $j = \lambda+1,..., J$ $(1 \leqslant r \leqslant q,\ 1 \leqslant \lambda \leqslant J)$. Choose $\mathbf{x}^* \in \mathscr{X}$ so that $x_i^* > 0$ for $i = r+1,..., q$ and $= 0$ otherwise; this is possible since $N \geqslant t-1$ and $1 \leqslant r \leqslant q \leqslant t$. Then $\mathbf{x}^* \in \mathscr{X}(F_j)$ for $j = 1,..., \lambda$ and $\mathbf{x}^* \notin \mathscr{X}(F_j)$ for $j = \lambda+1,..., J$ since I_0 is minimal. It follows that $\delta_i(\mathbf{x}^*) = 0$ for $i = 1,..., r$ since $p_i = 0$, $i = 1,..., r$ whenever $f(\mathbf{x}^* \mid \mathbf{p}) > 0$ and $p \in T \supset \operatorname{Supp} P$. Thus, $R(\mathbf{p}^*, \boldsymbol{\delta}) = +\infty$ since $p_i^* > 0$ for $i = 1,..., r$ which contradicts $\mathbf{p}^* \in \mathscr{S}^0$ and proves the theorem.

Remark A.1. When $N < t-1$, strict containment can occur in $T \subset \mathscr{S}^0$. For example let $t = 3$, $N = 1$ and P be a prior with support contained in the interior of $F_1 = \{\mathbf{p} \in \mathscr{S}\colon p_1 = 0\}$. Then $\min\{\xi(\mathbf{e}_1), \xi(\mathbf{e}_2)\} > 0 = \xi(\mathbf{e}_1)$ where $\mathbf{e}_i$ is the ith unit vector. Every Bayes rule has $\boldsymbol{\delta}^0(\mathbf{e}_2)$ and $\boldsymbol{\delta}^0(\mathbf{e}_3)$ uniquely determined; the rule with $\boldsymbol{\delta}^0(\mathbf{e}_1) = (1/3,\ 1/3,\ 1/3)$, say, gives $\mathscr{S}^0 = \{\mathbf{e}_1\} \cup F_1 \neq T = F_1$. However, the authors conjecture that if in addition to the hypotheses of Theorem A.1 either $\xi(\mathbf{x}) > 0$ $\forall \mathbf{x} \in \mathscr{X}(\mathscr{S}^0)$ or $\boldsymbol{\delta}^0$ is admissible w.r.t. entropy loss, then $T = \mathscr{S}^0$ and hence Theorems 3.1, 3.2 and 4.1 would hold for all $N \geqslant 1$.

References

[1] Akaike, H. (1977). On entropy maximization principle. In *Applications of Statistics* (P.R. Krishnaiah, Ed.), pp. 27–41. North-Holland, Amsterdam.

[2] Akaike, H. (1978). A new look at the Bayes procedure. *Biometrika* **65** 53–59.

[3] Alam, K. (1979). Estimation of multinomial probabilities. *Ann. Statist.* 7 282–283.

[4] Brown, L. D. (1979). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.*, in press.

[5] Brown, L. D. (1980). *Notes on Statistical Decision Theory.* Lecture Notes, Cornell University.

[6] Ferguson, T. (1967). *Mathematical Statistics.* Academic Press, New York.

[7] Johnson, B. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.* **42** 1579–1587.

[8] Kullback, S. and Liebler, R. (1951). On information and sufficiency. *Ann. Math. Statist.* **22** 79–86.
[9] Sanov, I. (1961). On the probability of large deviations of random variables. *Selected Transl. Math. Statist. Prob.* **1** 213–244.
[10] Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York.