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 $X = (X_1, \ldots, X_t)$ have a multinomial distribution based on $N$ trials with unknown vector of cell probabilities $p = (p_1, \ldots, p_t)$. This paper derives admissibility and complete class results for the problem of simultaneously estimating $p$ under entropy loss (EL) and squared error loss (SEL). Let $\mathcal{S}$, $\mathbb{S}$ and $f(x \mid p)$ denote the $(t-1)$-dimensional simplex, the support of $X$ and the probability mass function of $X$, respectively. First it is shown that $\delta$ is Bayes w.r.t. EL for prior $P$ if and only if $\delta$ 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 $\mathcal{S}$ of the form $T = \bigcup_{j=1}^{k} F_j$ where $k \geq 1$ and each $F_j$ is a facet of $\mathcal{S}$ which satisfies: $F$ a facet of $\mathcal{S}$ such that $F \nsubseteq F_j \Rightarrow F \nsubseteq T$. The minimal complete class of rules w.r.t. EL when $N \geq t-1$ is characterized as the class of Bayes rules with respect to priors $P$ which satisfy $P(\mathcal{S}^0) = 1$, $\xi(x) = \int f(x \mid p) P(dp) > 0$ for all $x$ in $\{x \in \mathbb{S} : \sup_{\mathcal{S}^0} f(x \mid p) > 0\}$ for some $\mathcal{S}^0$ in $Q$ containing all the vertices of $\mathcal{S}$. As an application, the maximum likelihood estimator is proved to be admissible w.r.t. EL when the estimation problem has parameter space $\Theta = \mathcal{S}$ but it is shown to be inadmissible for the problem with parameter space $\Theta = (\mathcal{S}$ minus its vertices). This is a severe form of "tyranny of boundary." Finally it is shown that when $N \geq t-1$ any estimator $\delta$ which satisfies $\delta(x) > 0 \ \forall x \in \mathbb{S}$ 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 $X = (X_1, \ldots, X_t)$ has a multinomial distribution based on $N$ trials with unknown vector of cell probabilities $p = (p_1, \ldots, p_t)$ in $\mathcal{S} \equiv \{p \mid p_i \geq 0$
\( \forall i, \sum p_i = 1 \). Denote the mass function of \( X \) by
\[
f(x \mid p) = N! \prod (p_i/x_i!), \quad x \in \mathcal{E}
\]
where \( \mathcal{E} = \{x = (x_1, \ldots, x_t) \mid \sum x_i = N, \ x_i \geq 0 \ \text{is integer} \ \forall i \} \) and our convention is that the range of any product or summation over the integers \( \{1, \ldots, t\} \) will be suppressed.

This paper derives admissibility and complete class results for the problem of simultaneously estimating \( p \) under entropy loss (EL); it uses these results to establish relationships between the admissible rules under EL and squared error loss (SEL). For \( p, s \in \mathcal{S} \), SEL is defined by
\[
L_S(p, s) = N \sum (p_i - s_i)^2 = N \| p - s \|^2 \tag{1.1}
\]
and EL by
\[
L_E(p, s) = N \sum p_i \ln p_i - \ln s_i \tag{1.2}
\]
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 \( \delta = \delta(X) \) will be denoted by \( R_E(p, \delta) \) and \( R_S(p, \delta) \), 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 \( \mathcal{S}^+ \equiv \{q \in \mathcal{S} : q_i > 0\} \) and \( \partial \mathcal{S} \equiv \{q \in \mathcal{S} : q \in \mathcal{S}^+ \} \) denote the relative interior and boundary of \( \mathcal{S} \), respectively. If \( p \in \mathcal{S}^+ \), then a guess \( a^1 \in \partial \mathcal{S} \) is equivalent to any \( a^2 \in \partial \mathcal{S} \) satisfying \( \|p - a^1\|^2 = \|p - a^2\|^2 \). In contrast, EL differentiates between positive and zero guesses of \( p_i > 0 \); \( L_E(p, a) = +\infty \) for all \( p \in \mathcal{S}_+ \) and \( a \in \partial \mathcal{S} \). Alternatively, Akaike \([1, 2]\) motivates EL by the premise that the reason for estimating \( p \) by \( a \) is to base decisions about \( f(x \mid p) \) on \( f(x \mid a) \) since \( L_E(p, a) \) is the Kullback–Leibler mean information for discrimination between \( f(x \mid p) \) and \( f(x \mid a) \) \([8]\). Asymptotically \( L_E(p, a) \) is roughly the negative of the logarithm of the probability of observing a sample distribution closely approximated by \( f(x \mid p) \) when a large number of observations are independently drawn from \( f(x \mid a) \) \([9]\). See \([1, 2]\) and the references therein for additional motivation and examples.

Since the action space \( \mathcal{A} = \mathcal{S} \) is a convex compact subset of Euclidean \( t \)-space, \( \mathcal{R}_t \), and both \( L_S(p, a) \) and \( L_E(p, a) \) are convex in \( a \) \( \forall p \in \mathcal{S} \), the nonrandomized decision rules form an essentially complete class \([6]\). Throughout this paper attention will be restricted to nonrandomized
MULTINOMIAL ESTIMATION

estimators \( \delta : \mathcal{X} \rightarrow \mathcal{S} \) except for the class \( \mathcal{D}_c \) defined in the proof of Theorem 2.1 where the convexity of the corresponding class of risk functions, \( I(\mathcal{D}_c) \), will require the inclusion of randomized rules.

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

\[
\int R(p, \delta^B) P(dp) \leq \int R(p, \delta) P(dp)
\]

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

\[
v(x) = \int p P(dp | x)
\]

where \( P(- | x) \) denotes the posterior distribution of \( P \) given \( x \) or an arbitrary probability measure according to whether \( \int f(x | p) P(dp) > 0 \) 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 \( p \) is proved to be admissible for the problem with parameter space \( \mathcal{S} \) and inadmissible for the problem with parameter space \( \mathcal{S} \) 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 \( \mathcal{S}^+ \) (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 \( \mathcal{S} \). The Bayes risk of the estimator \( \delta \) under \( L_E \) is

\[
r_E(p, \delta) = \sum_{\mathcal{S}} L_E(p, \delta(x)) f(x | p) P(dp).
\]

If \( P(- | x) \) is defined as in Section 1, then

\[
r_E(p, \delta) = \sum_{\mathcal{S}} E[L_E(p, \delta(x)) | x] \xi(x)
\]
where \( \xi(x) = \int f(x \mid p) P(dp) \) is the marginal distribution of \( X \) and

\[
E[L_E(p, \delta(x)) \mid x] = \int \sum_i p_i \ln(p_i/\delta_i(x)) P(dp \mid x)
\] (2.1)

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

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

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 \( \mathcal{A} = \mathcal{S} \) and the parameter space \( \Theta = \mathcal{S} \) are convex compact subsets of \( \mathbb{R}^l \). Since \( L_A(p, a) \) is bounded, convex, and continuous in \( a \) for each \( p \), 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 \( A \subset \mathcal{A} \) let \( \mathcal{F}(A) \equiv \{ x \in \mathcal{X} : \sup \rho_r(x \mid p) > 0 \} \) denote the set of outcomes which can be "seen" under \( A \). For arbitrary sets \( E \) and \( F \) let \( E \setminus F \equiv \{ w \in E : w \notin F \} \).

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

**Proof.** \( \delta^0 \) must be admissible for the problem \( \mathcal{S} = (\mathcal{S}, \Theta = \mathcal{S}, \mathcal{A} = \mathcal{S}, L_E) \), \( \mathcal{A} = \mathcal{S}, L_E \) where \( \mathcal{A}^0 = \mathcal{F}(\mathcal{S}^0) \); if not, there exists a \( \delta \) such that \( R_E(p, \delta) < R_E(p, \delta^0) \) \( \forall p \in \mathcal{S}^0 \) with \( \forall \) for some \( p \in \mathcal{S}^0 \Rightarrow \delta \) is better than \( \delta^0 \) for the problem \( \mathcal{S} \) since \( R_E(p, \delta) = +\infty \) \( \forall p \in \mathcal{S} \setminus \mathcal{S}^0 \).

Let \( \mathcal{D}^0 = \{ \delta : \mathcal{S}^0 \to \mathcal{S} \} \), \( 0 < c < 1 \) and \( \mathcal{D}^c = \{ \delta \ a \ \text{randomized rule on} \ \mathcal{S}^0 \} \) where \( \mathcal{D}^c \) a randomized rule on \( \mathcal{S}^0 \) is a mapping from \( \mathcal{S}^0 \) to the set of all probability measures on \( \mathcal{S}^0 \) is a continuous, real valued function on \( \mathcal{S}^0 \), \( R_E(\cdot, \delta) \) is a continuous function on \( \mathcal{S}^0 \), that the risk set \( \mathcal{R}(\mathcal{D}^0) = \{ R_E(\cdot, \delta) : \delta \in \mathcal{D}^0 \} \) is convex and that \( \tilde{R}(\mathcal{D}^0) = \{ h : \mathcal{S}^0 \to [0, \infty] : \delta \in \mathcal{D}^0 \} \) is closed in \( \prod_{\mathcal{S}^0} [0, \infty] \) under the topology defined by pointwise convergence of sequence of functions. By Theorem 3.5 of [5], \( \delta^0 \) is Bayes relative to \( \mathcal{D}^c \) for some prior \( P \) such that \( P(\mathcal{S}^0) = 1 \). We
claim $\delta^0$ is Bayes with respect to $P$ relative to ${\cal D}^0$; if not, then $\exists \delta \in {\cal D}^0 \backslash {\cal D}^0_\epsilon$ and $x^* \in {\cal D}^0_\epsilon \ni \xi(x^*) > 0$ and $\phi(\delta(x^*)) > \phi(\delta^0(x^*))$ where $\phi(\delta) = \int_{\Theta} \sum_i p_i \ln \delta_i P(dp \mid x^*)$. Let $\delta^1(x) = \lambda \delta(x) + (1 - \lambda) \delta^0(x)$ for $\lambda \in (0, 1)$. By concavity of $\phi(\delta)$, $\phi(\delta^1(x)) \geq \lambda \phi(\delta(x)) + (1 - \lambda) \phi(\delta^0(x)) > \phi(\delta^0(x))$ for $\lambda \in (0, 1)$. Choose $\lambda$ sufficiently small so that $\forall i \in \{1, \ldots, t\}$, $\delta^1_i(x) = \delta^0_i(x) + \lambda (\delta_i(x) - \delta^0_i(x)) > \delta^0_i(x) \quad \forall x \in {\cal D}^0$ since $c \in (0, 1) \Rightarrow \delta^1 \in {\cal D}^0 \Rightarrow \delta^0$ is not Bayes with respect to $P$ relative to ${\cal D}^0_\epsilon$. Since $P$ has support in $Y^\theta$, we have $\forall \delta \in {\cal D} = \{\delta : \Theta \to {\cal S}\}$ that the Bayes risk relative to $P$ is

$$r_E(P, \delta) = \int_{\Theta} \sum_{x \in \Theta} L_E(p, \delta(x)) P(dp \mid x) \xi(x)$$

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, $\delta^0$ denotes a nonrandomized estimator for the problem $\mathcal{S} = (\Theta, \Theta = \mathcal{S}, \mathcal{K} = \mathcal{S}, L_E)$ and $\mathcal{S}^0 = \{p \in \mathcal{S} : R_E(p, \delta^0) < \infty\}$. We begin by describing the structure of $\mathcal{S}^0$ for admissible $\delta^0$ in Lemmas 3.1 and 3.2 and then characterize the minimal complete class for $\mathcal{S}$ in Theorems 3.1 and 3.2.

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

For any facet $F$ of $\mathcal{S}$ let $I(F)$ be defined by $p \in F \Leftrightarrow p_i = 0 \forall i \in I(F)$; set $I(\mathcal{S}) = \emptyset$. Let $Q$ denote the collection of subsets $T$ of $\mathcal{S}$ which can be written in the form $T = \bigcup_{i=1}^k F_i$ for some positive integer $k$ where $F_1, \ldots, F_k$ are facets of $\mathcal{S}$ such that $\forall i = 1, \ldots, k$ if $F$ is a facet of $\mathcal{S}$ satisfying $F \supseteq F_i$, then $F \in T$.

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

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


Lemma 3.2. \{e_1, \ldots, e_t\} \subset \mathcal{S}^0 for any admissible \delta^0 where e_i is the unit vector with 1 in the ith component and 0 in the remaining components.

Proof. Suppose w.l.o.g. e_1 \notin \mathcal{S}^0, then \( R_F(e_1, \delta^0) = +\infty \iff \delta^0_1(\mathbb{N}e_1) = 0 \). Define a new estimator \( \bar{\delta} \) by
\[
\bar{\delta}(x) = e_1, \quad x = Ne_1
\]
\[
\equiv \delta^0(x), \quad x \neq Ne_1,
\]
then \( 0 = R_F(e_1, \bar{\delta}) < \infty = R_F(e_1, \delta^0), \quad R_F(p, \bar{\delta}) \leq \infty = R_F(p, \delta^0) \) for all \( p \) with \( 0 < p_1 < 1 \) and \( R(p, \bar{\delta}) = R(p, \delta^0) \) for all \( p \) with \( p_1 = 0 \). Therefore \( \bar{\delta} \) dominates \( \delta^0 \) which contradicts the admissibility of \( \delta^0 \) and completes the proof. One consequence of Lemma 3.2 is that for admissible \( \delta^0 \) every facet of \( \mathcal{S} \) has a nonempty intersection with \( \mathcal{S}^0 \).

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

Theorem 3.1. Every admissible \( \delta^0 \) is Bayes w.r.t. some prior \( \mathbb{P} \) with support in \( \mathcal{S}^0 \) which satisfies \( \{x \in \mathcal{S}; \xi(x) > 0\} \subset \mathbb{E}(\mathcal{S}^0) \). Furthermore if \( N \geq t - 1 \), then \( \{x \in \mathcal{S}; \xi(x) > 0\} = \mathbb{E}(\mathcal{S}^0) \).

Proof. Theorem 2.1 guarantees \( \delta^0 \) is Bayes w.r.t. some prior \( \mathbb{P} \) with support in \( \mathcal{S}^0 \). If \( x \in \mathbb{E} \setminus \mathbb{E}(\mathcal{S}^0) \), then \( f(x \mid p) = 0 \forall p \in \mathcal{S}^0 \Rightarrow \xi(x) = 0 \) and so \( \{x \in \mathcal{S}; \xi(x) > 0\} \subset \mathbb{E}(\mathcal{S}^0) \). Now suppose \( N \geq t - 1 \); if \( \{F_j\}_{j=1}^J \) is the collection of facets such that \( P(F_j^+) > 0 \) for \( j = 1, \ldots, J \) where \( F_j^+ \) denotes the relative interior of \( F_j \), then Theorem A.1 in the Appendix proves that \( \mathcal{S}^0 = \bigcup_{j=1}^J F_j \). If \( x \in \mathbb{E}(\mathcal{S}^0) \), then \( \exists p \in \mathcal{S}^0 \) such that \( f(x \mid p) > 0 \). Assume w.l.o.g. that \( x_i > 0 \) for \( i = 1, \ldots, r \) and \( = 0 \) for \( i = r + 1, \ldots, t \) and that \( p \in F_1 \). Then \( p_i > 0 \) for \( i = 1, \ldots, r \) and \( p^i \in F_j^+ \Rightarrow p^i > 0 \) for \( i = 1, \ldots, r \). Hence \( \xi(x) > 0 \) and the proof is finished.

Theorem 3.2. Suppose \( N \geq t - 1 \) and \( T \subset Q \) contains the vertices \( e_1, \ldots, e_t \). If \( \mathbb{P} \) is a probability measure on \( \mathcal{S} \) satisfying \( P(T) = 1 \) and \( \xi(x) = \int f(x \mid p) P(dp) > 0 \forall x \in \mathbb{E}(T) \), then any Bayes estimator \( \delta \) w.r.t. \( \mathbb{P} \) is admissible under EL.

Proof. Suppose \( \delta' \) satisfies \( R(p, \delta) \leq R(p, \delta) \forall p \in \mathcal{S} \), then \( \delta' \) is also Bayes w.r.t. \( \mathbb{P} \). By Remark 2.1 the Bayes estimator is unique for \( x \) such that \( \xi(x) > 0 \) and hence \( \delta(x) = \delta'(x) \forall x \in \mathbb{E}(T) \). So \( R_F(p, \delta) = R_F(p, \delta') \forall p \in T \). Let \( \{F_j\} \) be the set of facets of \( \mathcal{S} \) such that \( P(F_j^+) > 0 \forall j \); it is easy to see that \( \bigcup F_j = T \). Furthermore, Theorem A.1 guarantees \( \{p \in \mathcal{S}: \)
\[ R_F(p, \delta') < \infty \} = \bigcup F_j = \{ p \in \mathcal{F}: R_F(p, \delta) < \infty \} \] since \( N \geq t - 1 \) and hence \( R_F(p, \delta') = \infty = R_F(p, \delta) \) \( \forall p \in \mathcal{F} \setminus \mathcal{E} \) so that \( \delta \) is admissible.

**Remark 3.1.** Any estimator \( \delta \) satisfying \( \delta(Ne_i) = e_i \) for \( i = 1, ..., t \) is admissible under EL since \( d \) is Bayes with respect to the prior \( P \) putting mass \( 1/t \) at each point of \( T = \{e_1, ..., e_t\} \).

**Remark 3.2.** The proof of Theorem 3.2 shows that if \( \zeta(x) > 0 \) \( \forall x \in \mathcal{E} \), then any Bayes rule is admissible for all \( N \geq 1 \).

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

\[ R_S(p, \delta_{mle}) = 1 - \|p\|^2 \]

is small when \( p \) 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, \( \delta_{mle} \) is

(i) admissible when \( \Theta = \mathcal{S} \) for \( N \geq t - 1 \),

(ii) inadmissible when \( \Theta = \mathcal{S} \setminus \{e_1, ..., e_t\} \) for \( N \geq 1 \).

**Proof.** Part (i) follows from Remark 3.1. To prove part (ii) fix \( p \in \mathcal{S} \setminus \{e_1, ..., e_t\} \); assume w.l.o.g. that \( p_j > 0 \) for \( j = 1, 2 \) since \( p \) has at least two nonzero components. Choose \( x^* = x^*(p) \) so that \( P_p[X = x^*] > 0 \) and \( \delta_{mle}^1(x^*) > 0 \Rightarrow R(p, \delta_{mle}) = \infty \). \( \delta_{mle} \) is clearly dominated by the constant estimator \( \delta_c(x) = (1/t, ..., 1/t) \) \( \forall x \in \mathcal{E} \).

**Remark 3.3.** In contrast, \( \delta_{mle} \) is admissible under SEL when \( \Theta = \mathcal{S} \setminus \{e_1, ..., e_t\} \) and even when \( \Theta = \mathcal{S}^+ \). See [4, especially Proposition 1.7].

The admissibility of \( \delta_{mle} \) 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 \( x \in \mathcal{E} \) via Lemmas 4.1 and 4.2; it then considers estimators which allow zero guesses.
Lemma 4.1. Let $\delta^0$ be admissible under EL. Then there is an estimator $\delta'$ which is admissible under SEL such that $\delta^0(x) = \delta'(x)$ $\forall x \in \mathcal{B}(\mathcal{F}^0)$ when $N \geq t - 1$.

Proof. From Theorem 3.1, $\delta$ is unique Bayes on $\mathcal{B}(\mathcal{F}^0)$ relative to some prior $P$ on $\mathcal{F}^0$. Let $P_1$ be a prior with marginal distribution $\zeta_i(x) = \int f(x | p) P_1(dp | x) > 0$ $\forall x \in \mathcal{B}$. Define $\delta'(x) = \delta(x)$ for $x \in \mathcal{B}(\mathcal{F}^0)$ and $\delta'(x) = \int p P_1(dp | x)$ for $x \in \mathcal{B}(\mathcal{F}^0)$. Then $\delta'$ is admissible for SEL by [4].

Lemma 4.2. Suppose $N \geq t - 1$ and $\delta$ is admissible under SEL. If $\delta(x) > 0$ $\forall x \in \mathcal{B}$, then $\delta$ is unique Bayes on all of $\mathcal{B}$ for some prior $P$. (Equivalently $\delta$ is Bayes for a prior $P$ having $\xi(x) > 0$ $\forall x \in \mathcal{B}$.)

Proof. Any $\delta$ admissible under SEL is Bayes from Section 2; let $P$ be a prior such that $\delta$ is Bayes w.r.t. $P$, Suppose $\exists x^* \in \mathcal{B} \ni \xi(x^*) = 0$; w.l.o.g., assume $x_i^* > 0$ for $i = 1, ..., s$ and $x_i^* = 0$ for $i = s + 1, ..., t$ $(1 \leq s \leq t)$. Then $P$ has support in $T = \{p \in \mathcal{F} : p_i = 0 \text{ for some } i = 1, ..., s\} = \bigcup F_j$, 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 $\mathcal{F} = \{F \in T : F \text{ a facet, } P(F) > 0\}$. Partially order the sets $I(F)$, $F \in \mathcal{F}$ by inclusion and let $I_0$ be a minimal set in this ordering. Now $I_0 \neq \emptyset$ since $I_0 = \emptyset \Rightarrow \exists F \in \mathcal{F} \ni I(F) = I_0 = \emptyset$ which contradicts $F \in T$; assume w.l.o.g. that $I_0 = \{r + 1, ..., s\}$ with $1 \leq r + 1 \leq s$. Fix $\hat{x} \in \mathcal{B}$ so that $\hat{x}_i = 0$ for $i = r + 1, ..., s$ and $\hat{x}_i > 0$ otherwise; this is possible since $N \geq t - 1$. Clearly $\xi(\hat{x}) > 0$ since $\xi(\hat{x}) = 0$ (together with $\xi(x^*) - 0 \ni \text{Supp}(P) \subset \{p \in \mathcal{F} : p_i = 0 \text{ for some } 1 \leq i \leq r\} = I(F) \subset \{1, ..., r\}$ $\forall F \in \mathcal{F}$ which is impossible since $\exists F' \in \mathcal{F} \ni I(F') = I_0 = \{r + 1, ..., s\}$). To complete the proof it suffices to show $p_i = 0$ for $i = r + 1, ..., s$ whenever $f(\hat{x} | p) > 0$ and $p \in F \in \mathcal{F}$ and hence $\delta_i(\hat{x}) = 0$ for $i = r + 1, ..., s$ which contradicts the assumption $\delta(x) > 0$ $\forall x \in \mathcal{B}$. But $f(\hat{x} | p) > 0 \Rightarrow p_i > 0$ for $1 \leq i \leq r$ and $s + 1 \leq i \leq t \Rightarrow I(F) \subset \{r + 1, ..., s\} = I_0 \Rightarrow I(F) = I_0$ 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 $N \geq t - 1$, a positive estimator $\delta$ is admissible under SEL if and only if $\delta$ is admissible under EL. Suppose $\delta(x) > 0$ for all $x \in \mathcal{B}$; then $\mathcal{F}^0 = \{p \in \mathcal{F} | R_E(p, \delta) < \infty\} = \mathcal{F}$. Hence if $\delta$ is admissible under SEL, then $\delta$ is Bayes for some prior $P$ satisfying $\xi(x) > 0$ $\forall x \in \mathcal{B}(\mathcal{F}^0)$ and hence $\delta$ is admissible under EL by Theorem 3.2. Conversely, if $\delta$ is admissible under EL, then $\delta(x) = \delta'(x)$ $\forall x \in \mathcal{B}(\mathcal{F}^0) = \mathcal{B}$ where $\delta'$ is admissible under SEL so that $\delta$ is admissible under SEL.

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

**Example 4.1.** Consider the estimator for the trinomial problem defined by

$$\delta(x_1, x_2, x_3) = (0, 1/2, 1/2), \quad x_1 = 0$$

$$= (1/3, 1/3, 1/3), \quad x_1 > 0.$$ 

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

$$\bar{\delta}(x_1, x_2, x_3) = (0, 1/2, 1/2), \quad x_1 = 0$$

$$= (1/3, 1/3, 1/3), \quad 0 < x_1 < N$$

$$= (1, 0, 0), \quad x_1 = N.$$ 

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

Conversely, an estimator $\delta$ which guards against the states of nature $e_1, \ldots, e_t$ by guessing $\delta(Ne_i) = e_i$ for $i = 1, \ldots, t$ will be admissible under EL. However $\delta$ can be inadmissible under SEL by making counter intuitive guesses at other $x$ values.

**Example 4.2.** By Remark 3.1 the estimator defined by

$$\delta(x) \equiv e_i, \quad x = Ne_i \quad (1 \leq i \leq t)$$

$$= e_1 \quad \text{otherwise}$$

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

**Appendix**

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

**Theorem A.1.** Let $\delta^0$ be any Bayes rule w.r.t. $P$ under EL. Then
\( T \subset \mathcal{S}^0 \equiv \{ \mathbf{p} \in \mathcal{S} : R_E(\mathbf{p}, \delta^0) < \infty \} \) for all \( N \geq 1 \). Furthermore \( T = \mathcal{S}^0 \) when \( N \geq t - 1 \).

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

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

**Remark A.1.** When \( N < t - 1 \), strict containment can occur in \( T \subset \mathcal{S}^0 \).

For example let \( t = 3, N = 1 \) and \( P \) be a prior with support contained in the interior of \( F_1 = \{ \mathbf{p} \in \mathcal{S} : p_1 = 0 \} \). Then \( \min\{ \xi(e_1), \xi(e_2) \} > 0 = \xi(e_3) \) where \( e_i \) is the \( i \)th unit vector. Every Bayes rule has \( \delta^0(e_2) \) and \( \delta^0(e_3) \) uniquely determined; the rule with \( \delta^0(e_i) = (1/3, 1/3, 1/3) \), say, gives \( \mathcal{S}^0 = \{ e_1 \} \cup F_1 \neq T = F_1 \). However, the authors conjecture that if in addition to the hypotheses of Theorem A.1 either \( \xi(\mathbf{x}) > 0 \ \forall \mathbf{x} \in \mathcal{X}(\mathcal{S}^0) \) or \( \delta^0 \) is admissible w.r.t. entropy loss, then \( T = \mathcal{S}^0 \) and hence Theorems 3.1, 3.2 and 4.1 would hold for all \( N \geq 1 \).

**References**


