The Annals of Statistics
1979, Vol. 7, No. 5, 1127–1135

ADMISSIBLE DECISION RULES FOR THE COMPOUND DECISION PROBLEM: THE TWO-ACTION TWO-STATE CASE

By James Inglis

University of Rochester

For the two-action two-state compound decision problem, a class of decision rules is found that are both asymptotically optimal and admissible at each stage n. These rules are Bayes rules with respect to appropriate exchangeable prior distributions.

1. Introduction. Consider simultaneously n decision problems with identical structure: χ the sample space, Ω the parameter space indexing a family of probability distributions $\{P_\theta, \theta \in \Omega\}$ over χ, A the action space, and $L(a, \theta)$ the loss function defined on $A \times \Omega$.

If $\boldsymbol{\theta}_n = (\theta_1, \cdots, \theta_n)$ is a set of parameter values, where θ_i is the parameter value in the ith problem, and $\mathbf{a}_n = (a_1, \cdots, a_n)$ is a corresponding set of actions, the overall loss is defined to be the average loss

$$L(\mathbf{a}_n, \boldsymbol{\theta}_n) = (1/n)\Sigma_{i=1}^n L(a_i, \theta_i).$$

Let $\boldsymbol{\psi}_n(\mathbf{x}_n)$ be a (nonrandomized) compound decision rule specifying action $\psi_i(\mathbf{x}_n)$ for the ith component problem when $\mathbf{x}_n = (x_1, x_2, \cdots, x_n)$ is observed, where $x_i \in \chi$. Then

$$R(\boldsymbol{\psi}_n, \theta_n) = (1/n)\Sigma_{i=1}^n R(\psi_i, \theta_i) = (1/n)\Sigma_{i=1}^n E_{\boldsymbol{\theta}_n} L(\psi_i(\mathbf{x}_n), \theta_i).$$

The rule $\boldsymbol{\psi}_n(\cdot)$ is called *simple* if $\psi_i(\mathbf{x}_n) = \psi_i(x_i)$ and *simple symmetric* if, in addition, $\psi_i(x_i) = \psi(x_i)$. If $\boldsymbol{\psi}_n(\cdot)$ is simple symmetric, then

$$R(\boldsymbol{\psi}_n, \boldsymbol{\theta}_n) = (1/n)\Sigma_{i=1}^n R(\psi, \theta_i), \tag{1.1}$$

which is also the Bayes risk of the decision rule ψ in the single component case with respect to the prior that puts probability $1/n$ at each of the values $\theta_1, \cdots, \theta_n$. This empirical probability density function will be called h_n. For a prior density function g on Ω, $R(g)$ denotes the risk of the Bayes rule with respect to g. If h_n were known and $\boldsymbol{\psi}_n$ were required to be a simple symmetric rule, (1.1) would have minimum value $R(h_n)$.

Now consider an infinite sequence of such decision problems with an infinite sequence of parameter values $\boldsymbol{\theta}_\infty$.

Definition. A sequence of decision rules, $\{\boldsymbol{\psi}_n\}_{n=1}^\infty$, is called *asymptotically optimal* (hereafter abbreviated AO) if $R(\boldsymbol{\psi}_n, \boldsymbol{\theta}_n) - R(h_n) \to 0$ uniformly in $\boldsymbol{\theta}_\infty$. That

Received January 1974; revised December 1976.
AMS 1970 *subject classifications.* Primary 62C25; secondary 62C15.
Key words and phrases. Compound decision problem, decision theory, admissibility, exchangeable random variables.

is, given $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for all $n > N(\varepsilon)$, $|R(\psi_n, \boldsymbol{\theta}_n) - R(h_n)| < \varepsilon$, and the choice of N does not depend on $\boldsymbol{\theta}_\infty$.

Repeatedly assertions will be made that limits are approached "uniformly in $\boldsymbol{\theta}_\infty$." As above, this will mean that the choice of N, or δ, etc., does not depend on $\boldsymbol{\theta}_\infty$.

(Note: there is some variation in the notation and terminology for the compound decision problem. For more complete discussions, see Copas [1969] or Oaten [1972].)

Most compound decision rules that have been proposed were developed mainly on consideration of the AO criterion. For example, Robbins [1951] mentioned that the rule he proposed as AO is not admissible. He mentioned a possible competitor that is admissible, but stated that its risk function "seems difficult to compute." Hannan and Robbins [1955] mentioned that the rules they found in the solution to the two-action two-state problem are clearly not admissible. Samuel [1961, 1963] mentioned one admissible solution for the two-action two-state problem but gave an incomplete proof of the AO criterion. In a later paper [1967], Samuel gave a proof of asymptotic optimality for an admissible rule for a specific two-action two-state case, the uniform $(0, \theta_1)$ vs. uniform $(0, \theta_2)$. Copas [1970] also mentioned the question of admissibility briefly.

In this paper admissible Bayes rules with respect to suitably defined exchangeable priors will be defined. The proof of the AO criterion will involve a modification of techniques developed by Berk [1966].

2. Preliminaries. Here $\Omega = \{0, 1\}$. P_0 and P_1 are known probability distribution functions ($P_0 \neq P_1$) with associated probability measures p_0 and p_1. Consider their associated probability density functions $f_0(x)$ and $f_1(x)$ with respect to $\lambda = p_0 + p_1$. Without loss of generality, assume $f_i(x) > 0$ for all x in χ, $i = 1, 2$. Also assume P_0, P_1 nonatomic and

$$(2.1)\qquad \int_\chi f_i(x)|\log f_j(x)|\, d\lambda(x) < \infty \qquad \text{for } i, j = 1, 2.$$

Assume $0 - 1$ loss; i.e., $A = \{0, 1\}$ and $L(0, 0) = L(1, 1) = 0$ and $L(1, 0) = L(0, 1) = 1$. (Modifications to what follows are straightforward for the case $L(0, 1) = a$ and $L(1, 0) = b$).

In the single component case, if ψ is a decision rule, then

$$(2.2)\qquad R(\psi, \theta) = (1 - 2\theta)E(\psi(x)|\theta) + \theta.$$

If there is a prior distribution on Ω with $P(\theta = 1) = \pi$, then a Bayes rule, ψ_π, with respect to that prior is

$$(2.3)\qquad \begin{aligned} \psi_\pi(x) &= 1 \qquad \text{if } E(\theta|x) \geq \tfrac{1}{2} \\ &= 0 \qquad \text{otherwise,} \end{aligned}$$

where $E(\theta|x) = \pi f_1(x)\{\pi f_1(x) + (1 - \pi)f_0(x)\}^{-1}$. This rule has Bayes risk $R(\psi_\pi, \pi) = R(\pi)$.

In the compound decision (multicomponent) case, $\boldsymbol{\theta}_\infty$ is an infinite sequence of 0's and 1's. Let r be the number of 1's in $\boldsymbol{\theta}_n$, the first n terms of $\boldsymbol{\theta}_\infty$. (The dependence of r on n is suppressed.) For $\boldsymbol{\theta}_n$, h_n puts probability mass r/n on 1. If r is known, a best simple symmetric rule is

$$\text{(2.4)} \quad \begin{aligned} \psi(x_i) &= 1 \quad \text{if} \quad (r/n)f_1(x_i)\{(r/n)f_1(x_i) + (1 - r/n)f_0(x_i)\}^{-1} \geq \tfrac{1}{2} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

This is just (2.3) with π replaced by r/n. The best simple symmetric decision rule, ψ_n, then, has risk $R(r/n)$, and this is the criterion with which compound decision rules are compared. In addressing this problem, Hannan and Robbins [1955] suggested estimating r/n from $\mathbf{x}_n$, and then using a Bayes rule with respect to that estimate in each component problem. They proved the following theorem (Theorem 4, page 46, Hannan and Robbins [1955]), presented here in slightly different notation.

THEOREM 1. *If* $0 \leq p_n(\mathbf{x}_n) \leq 1$, *and* $|p_n(\mathbf{x}_n) - (r/n)| \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ *uniformly in* $\boldsymbol{\theta}_\infty$, *then* $\psi_{p_n(\mathbf{x}_n)}$ (*the rule described above*) *is* AO.

Hannan and Robbins [1955] proposed estimators of r/n that satisfy the theorem; they are averages of bounded unbiased estimators of the θ_i. (Note: the boundedness is not necessary. Appendix A, Inglis [1973] applied to this case, shows that only the unbiasedness is required.)

A Bayes estimator for $\boldsymbol{\theta}_n$ will be developed. The prior density for each $\boldsymbol{\theta}_n$ will be exchangeable. Here there is a slight notational difficulty to be overcome. Later the properties of the Bayes estimator of $\boldsymbol{\theta}_n$ will be examined with respect to the true, but unknown, $\boldsymbol{\theta}_n$. To differentiate between the real $\boldsymbol{\theta}_n$ and the ones in the Bayes model, call the parameter values in the model $\mathbf{y}_n$. Then consider the following prior: $P(y_i = 1) = g = 1 - P(y_i = 0)$, where g is a random variable taking values in [0, 1] according to a probability measure μ, which has an absolutely continuous probability density function greater than zero for all g in (0, 1).

Denote expectation for this Bayes prior model by E'. Straightforward calculation using (2.2) yields a Bayes rule with ith component

$$\text{(2.5)} \quad \begin{aligned} \psi_i(\mathbf{x}_n) &= 1 \quad \text{if} \quad E'(y_i|\mathbf{x}_n) \geq \tfrac{1}{2} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

The expression $E'(y_i|\mathbf{x}_n)$ can be rewritten

$$\text{(2.6)} \quad E'(y_i|\mathbf{x}_n) = \frac{f_1(x_i)\int_0^1 g\Pi_{j\neq i}(gf_1(x_j) + (1-g)f_0(x_j))\,d\mu(g)}{\begin{aligned} &f_1(x_i)\textstyle\int_0^1 g\Pi_{j\neq i}(gf_1(x_j) + (1-g)f_0(x_j))\,d\mu(g) \\ &+f_0(x_i)\textstyle\int_0^1(1-g)\Pi_{j\neq i}(gf_1(x_j) + (1-g)f_0(x_j))\,d\mu(g)\end{aligned}}.$$

Let a superscript (i) on a vector signify that vector with the ith element deleted.

The right hand side of (2.6) can be rewritten as

$$\frac{f_1(x_i)E'(g|x_n^{(i)})}{f_1(x_i)E'(g|x_n^{(i)}) + f_0(x_i)(1 - E'(g|x_n^{(i)}))},$$

where

$$E'(g|\mathbf{x}_n^{(i)}) = \frac{\int_0^1 g\Pi_{j\neq i}(gf_1(x_j) + (1-g)f_0(x_j))\,d\mu(g)}{\int_0^1 \Pi_{j\neq i}(gf_1(x_j) + (1-g)f_0(x_j))\,d\mu(g)}. \tag{2.7}$$

The Bayes rule ψ_n, then, is as follows: for each component problem use the other $n-1$ observations to estimate g (by $E'(g|\mathbf{x}_n^{(i)})$) and then use the single component Bayes rule (2.3) with respect to that estimated g. From the conditions on μ stated at the beginning of this section, for each n this exchangeable prior for $\mathbf{y}_\infty$ puts positive probability on every possible $\mathbf{y}_n$, and so ψ_n is admissible.

3. Establishing asymptotic optimality. The estimates of g, $E'(g|\mathbf{x}_n^{(i)})$, are always in the unit interval. By a generalization of Theorem 1, if $\max_{i\leqslant n}|E'(g|\mathbf{x}_n^{(i)}) - (r/n)| \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$, then the above Bayes rule ψ_n will be AO. (Note that the above statement involves the two probability structures. The conditional expectation $E'(\cdot|\cdot)$ is for the Bayes model with prior density. The convergence, however, is with respect to the true compound decision structure.)

The method of proof is to first establish that $|E'(g|\mathbf{x}_n) - (r/n)| \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$. Then established results about the behavior of certain maximum partial sums with respect to the strong law of large numbers will apply to verify that $\max_{i<n}|E'(g|\mathbf{x}_n^{(i)}) - (r/n)| \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$.

Refer to (2.7), delete the superscript (i), and consider the full expression for $E'(g|\mathbf{x}_n)$ and the associated posterior probability measure $\mu_n(g)$ for g. For C, a measurable subset of [0, 1],

$$\int_C d\mu_n(g) = \frac{\int_C \Pi_{i=1}^n(gf_1(x_i) + (1-g)f_0(x_i))\,d\mu(g)}{\int_0^1 \Pi_{i=1}^n(gf_1(x_i) + (1-g)f_0(x_i))\,d\mu(g)}. \tag{3.2}$$

Consider the behavior of this posterior probability distribution. What follows is a modification in two directions of an approach used by Berk [1966]: (i) to independent, nonidentically distributed random variables; (ii) to uniformity over all $\boldsymbol{\theta}_\infty$ sequences.

Let $w_n(g) = (1/n)\Sigma_{i=1}^n \log(gf_1(x_i) + (1-g)f_0(x_i))$, for g in [0, 1]. Then $w_n(\cdot)$ is a random variable defined on χ_n taking values in the separable Banach space, $C[0, 1]$ with the sup norm. The distribution of $w_n(\cdot)$ depends, of course, on $\boldsymbol{\theta}_n$, but that dependence is not explicit in the notation. Let $v_s(g) = sE_1(\log(gf_1(x) + (1-g)f_0(x))) + (1-s)E_0(\log(gf_1(x) + (1-g)f_0(x)))$ for all s, g in [0, 1], where E_i means expectation with respect to the corresponding f_i. Assumption (2.1) guarantees that $v_s(g)$ always exists.

LEMMA 1.

(i) *For fixed* $s \in [0, 1]$, $v_s(\cdot)$ *is in* $C[0, 1]$.

(ii) $E_{\theta_n}(w_n(\cdot)) = v_{r/n}(\cdot)$, *with expectation here being the Bochner integral in the separable Banach space* $C[0, 1]$. (*Equivalent to the statement* $E_{\theta_n}(w_n(g)) = v_{r/n}(g)$ *for all* $g \in [0, 1]$.)

(iii) *For each* $s \in [0, 1]$, $v_s(g) \leqslant v_s(s)$, *with equality only when* $g = s$.

(iv) *The function* $v_s(s)$ *is in* $C[0, 1]$ *and is bounded away from* $-\infty$.

PROOF. The lemma is a straightforward consequence of the definitions. See page 13, Inglis [1973] for details.

If $u(g)$ is a function in $C[0, 1]$ and B a measurable subset of $[0, 1]$, then define $B\|u\|_n$ as the L^n norm of u over B; i.e.,

$$B\|u\|_n = \left[\int_B |u(g)|^n \, d\mu(g)\right]^{1/n}.$$

The sup norm over B is $B\|u\|_\infty = \sup_{g \in B}|u(g)|$. If $B = [0, 1]$, simply write $\|u\|_n$ and $\|u\|_\infty$, respectively. Given $\delta > 0$ and $g \in [0, 1]$, let $U_\delta(g) = \{g' : g' \in [0, 1]$ and $v_g(g') > v_g(g) - \delta\}$.

LEMMA 2.

(i) *The element* $g \in U_\delta(g)$.

(ii) *The set* $\{g\} = \cap_{\delta>0} U_\delta(g)$.

(iii) *For fixed* $\delta > 0$, $\mu(U_\delta(g))$ *is a continuous function of* g.

(iv) *For* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that* $g' \in U_\delta(g)$ *implies* $|g' - g| < \varepsilon$ *uniformly in* g.

PROOF. Part (i) holds because $v_g(g')$ is a concave function of g'. (Establishing the concavity is straightforward.) Thus $v_g(g')$ is increasing on the interval $[0, g]$ and decreasing on the interval $[g, 1]$. (ii) is a consequence of Lemma 1 (iii). From (i), $U_\delta(g)$ is an interval with upper and lower endpoints $e_\delta^u(g)$ and $e_\delta^l(g)$. The upper endpoint $e_\delta^u(g)$ satisfies two conditions: (i) $e_\delta^u(g) \geqslant g$ and (ii) $v_g(e_\delta^u(g)) = \max(v_g(g) - \delta, \inf_{g' \geqslant g}\{v_g(g')\})$. The lower endpoint $e_\delta^l(g)$ satisfies two similar conditions. (Note: The conditions are a bit complicated because, for g close to 1 (or 0), $U_\delta(g)$ will include 1 (or 0) and thus be half-open.) Because of the continuity of $v_s(g)$ in g for each s and the continuity of $v_s(s)$ (Lemma 1 (i) and (iv)), it is easy to see that $e_\delta^u(g)$ and $e_\delta^l(g)$ are continuous functions of g. Because of the absolute continuity of μ, $\mu(U_\delta(g)) = \mu\{[0, e_\delta^u(g))\} - \mu\{[0, e_\delta^l(g))\}$, and each of the terms on the right is continuous in g. Therefore (iii) holds. The function $e_\delta^u(g) - g$ is also continuous in g. For each g, it is monotone decreasing in δ to zero by (iii) above. Because a monotone decreasing sequence of continuous functions, converging pointwise to a continuous function on a compact set, converges uniformly; for $\varepsilon > 0$, there exists $\delta(u)$ such that, for $g' \geqslant g$, $v_g(g) - v_g(g') < \delta(u)$ implies $g' - g < e_{\delta(u)}^u(g) - g < \varepsilon$ uniformly in g. An analogous chain of reasoning shows there exists a $\delta(l)$ such that for $g' - g$, $v_g(g) - v_g(g') < \delta(l)$ implies $g - g' < g - e_{\delta(l)}^l(g) < \varepsilon$ uniformly in g. Now let $\delta = \min(\delta(u), \delta(l))$ and (iv) is proved.

Hereafter $v_{r/n}(\cdot)$ is written as $v_n(\cdot)$. In this notation Lemma 1 (ii) becomes $E_{\theta_n}(w_n(\cdot)) = v_n(\cdot)$ and Lemma 1 (iii) implies $\sup_{0 \leqslant g \leqslant 1} v_n(g) = v_n(r/n)$.

THEOREM 2. *Fix $\delta > 0$. Consider the sequence of intervals $U_\delta(r/n)$, henceforth simply called $U_\delta(n)$. (Note that $r/n \in U_\delta(n)$ for all n.) Then $\mu_n(U_\delta(n)) \to 1$* a.e. $[\theta_\infty]$ *uniformly in θ_∞.*

PROOF. Referring to (3.2), consider the ratio

$$\frac{\mu_n(U_\delta^c(n))}{\mu_n(U_\delta(n))} = \frac{\int_{U_\delta^c(n)} \Pi_{i=1}^n (g f_1(x_i) + (1-g) f_0(x_i))\, d\mu(g)}{\int_{U_\delta(n)} \Pi_{i=1}^n (g f_1(x_i) + (1-g) f_0(x_i))\, d\mu(g)}, \tag{3.3}$$

where U^c means the complement of U. If this ratio converges to zero in an appropriate manner, the theorem will hold. The integrand for the numerator and denominator can be rewritten as

$$\left[\exp\{(1/n)\Sigma_{i=1}^n \log(g f_1(x_i) + (1-g) f_0(x_i))\}\right]^n = \left[\exp\{w_n\}\right]^n.$$

The nth roots of the integrals are L^n norms over $U_\delta(n)$ and $U_\delta^c(n)$. So (3.3) can be expressed as

$$\frac{\mu_n(U_\delta^c(n))}{\mu_n(U_\delta(n))} = \left[\frac{U_\delta^c(n)\|\exp w_n\|_n}{U_\delta(n)\|\exp w_n\|_n}\right]^n. \tag{3.4}$$

Consider the denominator in the brackets first. Note that

$$\begin{aligned} U_\delta(n)\|\exp w_n\|_n - U_\delta(n)\|\exp v_n\|_\infty &= (U_\delta(n)\|\exp w_n\|_n - U_\delta(n)\|\exp v_n\|_n) \\ &\quad + (U_\delta(n)\|\exp v_n\|_n - U_\delta(n)\|\exp v_n\|_\infty). \end{aligned} \tag{3.5}$$

Several lemmas are needed.

LEMMA 3. *Suppose $\cup_{\phi \in \Phi}\{\{a_n(\phi)\}_{n=1}^\infty\}$ and $\cup_{\phi\in\Phi}\{\{b_n(\phi)\}_{n=1}^\infty\}$ are two sets of sequences of elements belonging to an L^∞ space which satisfy* (i) $\|a_n(\phi) - b_n(\phi)\|_\infty \to 0$ *uniformly in ϕ as $n \to \infty$, and* (ii) $\sup_{\phi, n}\|a_n(\phi)\|_\infty < \infty$ *or* $\sup_{\phi, n}\|b_n(\phi)\|_\infty < \infty$. *(Note that* (i) *and either one of the boundedness conditions implies the other.) Then* $\|\exp\{a_n(\phi)\} - \exp\{b_n(\phi)\}\|_\infty \to 0$ *uniformly in ϕ.*

PROOF. Exp$(\cdot)$ is a uniformly continuous function on any bounded subset of the real line. Hence if C is a set of uniformly bounded elements of the L^∞ space, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $c_1, c_2 \in C$ and $\|c_1 - c_2\|_\infty < \delta$ imply $\|\exp c_1 - \exp c_2\|_\infty < \varepsilon$. Assumption (ii) shows that $\cup_{n,\phi}\{a_n(\phi), b_n(\phi)\}$ is a set of uniformly bounded elements. Assumption (i) then shows that for the above δ there exists N such that for all $n \geqslant N$ and for all ϕ in Φ, $\|a_n(\phi) - b_n(\phi)\|_\infty < \delta$. Then for all $n \geqslant N$, $\|\exp a_n(\phi) - \exp b_n(\phi)\|_\infty < \varepsilon$, and the lemma is proved.

LEMMA 4. $U_\delta(n)\|\exp w_n\|_n - U_\delta(n)\|\exp v_n\|_n \to 0$ a.e. $[\theta_\infty]$ *uniformly in θ_∞.*

PROOF. Note that

$$(3.6)\quad |U_\delta(n)\|\exp w_n\|_n - U_\delta(n)\|\exp v_n\|_n| \leqslant U_\delta(n)\|\exp w_n - \exp v_n\|_n$$
$$\leqslant U_\delta(n)\|\exp w_n - \exp v_n\|_\infty \leqslant \|\exp w_n - \exp v_n\|_\infty.$$

First consider $\|w_n - v_n\|_\infty$. By Lemma 1 (ii) $v_n = E(w_n)$ so set

$$Z_i(g) = \log(gf_1(x_i) + (1-g)f_0(x_i)) - E_{\theta_i}(\log(gf_1(x_i) + (1-g)f_0(x_i))).$$

$Z_i(\cdot)$ is a random variable in $C[0, 1]$. Then $w_n - v_n = (1/n)\Sigma_{i=1}^n Z_i$. $E(Z_i) = \mathbf{0}$ (the zero function), and (2.1) implies that $E(\|Z_i\|) < \infty$ for all i. Then Theorem 2, Appendix A, Inglis [1973], a uniform strong law of large numbers on separable Banach spaces, applies; and $\|w_n - v_n\|_\infty \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$. Now applying Lemma 3 to $\|w_n - v_n\|_\infty$ completes the proof.

LEMMA 5. $U_\delta(n)\|\exp v_n\|_n - U_\delta(n)\|\exp v_n\|_\infty \to 0$ *uniformly in* $\boldsymbol{\theta}_\infty$. (Note that this lemma does not involve probabilities.)

PROOF. A straightforward generalization of a standard proof on the convergence of L^n norms to L^∞ norms; e.g. page 91, Taylor [1958]. See page 19, Inglis [1973] for details.

Combining (3.5) and Lemmas 4 and 5 establishes that

$$(3.7)\quad U_\delta(n)\|\exp w_n\|_n - U_\delta(n)\|\exp v_n\|_\infty \to 0 \text{ a.e. } [\boldsymbol{\theta}_\infty] \text{ uniformly in } \boldsymbol{\theta}_\infty.$$

Next consider the numerator within the brackets of (3.4), which can be written as $U_\delta^c(n)\|\exp w_n\|_n$. The same argument as in Lemma 4 yields

$$(3.8)\quad U_\delta^c(n)\|\exp w_n\|_n - U_\delta^c(n)\|\exp v_n\|_n \to 0 \text{ a.e. } [\boldsymbol{\theta}_\infty] \text{ uniformly in } \boldsymbol{\theta}_\infty.$$

On $U_\delta^c(n)$, by definition, $v_n(g) \leqslant v_n(r/n) - \delta$. In addition, $U_\delta^c(n)\|\exp v_n\|_n \leqslant U_\delta^c(n)\|\exp v_n\|_\infty$ for all n, so

$$(3.9)\quad U_\delta^c(n)\|\exp v_n\|_n \leqslant \exp\{v_n(r/n) - \delta\}.$$

Combining (3.8) and (3.9) establishes that

$$(3.10)\quad \limsup_n U_\delta^c(n)\|\exp w_n\|_n \leqslant \exp\{v_n(r/n) - \delta\} \text{ a.e. } [\boldsymbol{\theta}_\infty] \text{ uniformly in } \boldsymbol{\theta}_\infty.$$

Now refer to (3.4). Combine (3.7) and (3.10) and note that Lemma 1 (iv) ensures that $\exp\{v_n(r/n)\}$ is bounded away from zero. Then it follows that

$$\limsup_n \frac{U_\delta^c(n)\|\exp w_n\|_n}{U_\delta(n)\|\exp w_n\|_n} \leqslant \frac{\exp\{v_n(r/n) - \delta\}}{\exp\{v_n(r/n)\}}$$
$$= \exp\{-\delta\} \text{ a.e. } [\boldsymbol{\theta}_\infty] \text{ uniformly in } \boldsymbol{\theta}_\infty.$$

Therefore, from (3.4),

$$(3.11)\quad \limsup_n \frac{\mu_n(U_\delta^c(n))}{\mu_n(U_\delta(n))} \leqslant \exp\{-n\delta\} \to 0 \text{ a.e. } [\boldsymbol{\theta}_\infty] \text{ uniformly in } \boldsymbol{\theta}_\infty.$$

For any series $\{a_n\}_{n=1}^\infty$, $0 < a_n < 1$ for all n, $\lim_{n\to\infty}(1 - a_n)/a_n = 0$ implies $\lim_{n\to\infty} a_n = 1$. Then (3.11) implies $\mu_n(U_\delta(n)) \to 1$ a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$ and Theorem 2 is proved.

THEOREM 3. *With the conditions as previously prescribed for* μ *and the assumption that* $f_1(x)$ *and* $f_2(x)$ *satisfy* (2.1), $|E'(g|\mathbf{x}_n) - (r/n)| \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ *uniformly in* $\mathbf{C}_\infty$.

PROOF. Note that

$$E'(g|\mathbf{x}_n) = \int_{U_\delta(n)} g\,d\mu_n(g) + \int_{U_\delta^c(n)} g\,d\mu_n(g).$$

Then

$$|E'(g|\mathbf{x}_n) - (r/n)| \leq \left|\int_{U_\delta(n)} g\,d\mu_n(g) - (r/n)\right| + \left|\int_{U_\delta^c(n)} g\,d\mu_n(g)\right|.$$

For any $\varepsilon > 0$, from Lemma 2 (iv) there exists δ so that $|g - r/n| < \varepsilon$ for all $g \in U_\delta(n)$. Hence

$$\begin{aligned}|E'(g|\mathbf{x}_n) - r/n| &\leq \int_{U_\delta(n)} |g - r/n|\, d\mu_n(g) + \int_{U_\delta^c(n)} |g|\, d\mu_n(g)\\ &\leq \varepsilon\mu_n(U_\delta(n)) + \mu_n(U_\delta^c(n)) \to \varepsilon\end{aligned}$$

a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$ by Theorem 2. The theorem is proved.

THEOREM 4. *With the conditions as described in Theorem* 3, $\max_{i \leq n}|E'(g|\mathbf{x}_n^{(i)}) - (r/n)| \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ *uniformly in* $\boldsymbol{\theta}_\infty$.

SKETCH OF THE PROOF. It is a modification of the material leading up to Theorem 3. The desired result will follow if $\max_{i \leq n}\mu_n^{(i)}(U_\delta^c(n))/\mu_n^{(i)}(U_\delta(n)) \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$ where the superscript (i) indicates, as before, that the ith observation is deleted. (cf. (3.3)). The previously stated arguments holds with straightforward changes. The crucial step is verifying that $\max_{i \leq n}\|w_n^{(i)} - v_n\|_\infty \to 0$ a.e. $[\boldsymbol{\theta}_\infty]$ uniformly in $\boldsymbol{\theta}_\infty$ for modifying Lemma 4. This result is a consequence of the Banach space version of the following easily verified fact (equivalent to the SLLN) for i.i.d. real random variables with mean μ: $\max_{i \leq n}(S_n^{(i)}/(n-1)) \to \mu$ a.e.

THEOREM 5. ψ_n *with* $\psi_i(\mathbf{x}_n)$ *defined by* (2.5) *for each i is admissible at each stage n and is* AO.

PROOF. That ψ_n is admissible at each stage n is implied by its definition. Combining Theorem 4 with the generalized Theorem 1 establishes the proof.

4. Some remarks. (1) In seeking admissible AO rules, Robbins [1951] and Samuel [1963, 1967] also developed Bayes rules with respect to a certain prior distribution. (See the introduction above.) The prior they used was the following: put prior probability $1/\left[(n+1)\binom{n}{k}\right]$ on each $\boldsymbol{\theta}_n$, where k is the number of 1's in $\boldsymbol{\theta}_n$. In the notation of this paper, this is the case when the prior μ for g is uniform over [0, 1]. Theorem 5 proves that this admissible (at each stage n) rule is indeed AO for f_1 and f_2 satisfying (2.1). Thus the rules Robbins and Samuel proposed are indeed AO and admissible (at each stage n). And in addition there is an entire class of admissible (at each stage n) AO decision rules; one for each possible prior μ on (0, 1).

(2) Huang [1972] considered the original problem investigated by Robbins [1951] where $P_0 = $ Normal $(-1, 1)$ and $P_1 = $ Normal $(1, 1)$. Using Robbins' notation,

Huang demonstrated that, for $n = 2$, the "average of unbiased estimators" rule that Robbins suggested as AO was not dominated by the admissible rule Robbins mentioned (see Remark 1 above) as a possible better one.

Huang considered the following symmetric priors on Ω^2. For $\frac{1}{4} \leqslant \lambda \leqslant \frac{1}{2}$,

$$P_\lambda\{\boldsymbol{\theta}_2\} = \lambda \text{ for } \boldsymbol{\theta}_2 = (1, 1), (-1, 1)$$
$$= \left(\tfrac{1}{2}\right) - \lambda \text{ for } \boldsymbol{\theta}_2 = (1, -1), (-1, 1).$$

When $\lambda = \frac{1}{3}$, the result is Robbins' prior. Huang added, however, that $\{P_\lambda | \lambda' \leqslant \lambda \leqslant \lambda''\}$ for some λ', λ'' with $\frac{1}{3} < \lambda' < \lambda'' < \frac{1}{2}$ is a set of priors on $\boldsymbol{\theta}_2$ for which the corresponding (Bayes) admissible rules will dominate the "average of unbiased estimators" rule. It is easy to see that for any $\lambda \in (\frac{1}{3}, \frac{1}{2})$, there exists a μ that yields a prior on $\mathbf{y}_2$ equal to p_λ. Thus for this special case and sample size, the class of admissible AO rules does contain rules that dominate the usual one.

(3) Gilliland, et al., [1974] have also considered this two-action two-state compound decision problem and in several cases have some bounds on the rates of convergence of $R(\psi_n, \psi_n) - R(h_n)$.

Acknowledgment. The author would like to thank Professor M. V. Johns, Jr. under whom the dissertation of which this is a part was written. He would also like to thank the Associate Editor for many useful comments.

REFERENCES

[1] Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

[2] Copas, J. B. (1969). Compound decisions and empirical Bayes. *J. Roy. Statist. Soc. Ser. B.* **31** 397–425.

[3] Gilliland, D. C., Hannan, J. and Huang, J. S. (1974). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. Tech. Report No. RM-320, Dept. Statist., Michigan State Univ.

[4] Hannan, J. F. and Robbins, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **36** 1743–1752.

[5] Huang, J. S. (1972). A note on Robbins' compound decision procedure. *Ann. Math. Statist.* **43** 348–350.

[6] Inglis, J. (1973). Admissible decision rules for the compound decision problem. Tech. Report No. 37, Dept. Statist., Stanford Univ.

[7] Oaten, A. (1972). Approximation to Bayes risk in compound decision problems. *Ann. Math. Statist.* **43** 1164–1184.

[8] Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probability*. Univ. California Press, 131–148.

[9] Samuel, E. (1961). On the compound decision problem in the nonsequential and sequential case. Ph.D. thesis, Columbia Univ.

[10] Samuel, E. (1963). Asymptotic solutions of the sequential compound decision problem. *Ann. Math. Statist.* **34** 1079–1084.

[11] Samuel, E. (1967). The compound statistical decision problem. *Sankhyā Ser. A* **29** 123–140.

[12] Taylor, A. E. (1958). *Introduction to Functional Analysis*. Wiley, New York.

WB 1J311
Bell Laboratories
Holmdel, New Jersey 07733