1978

# Admissible Solutions of Finite State Sequence Compound Decision Problems

Stephen B. Vardeman, *Purdue University*

*The Annals of Statistics*
1978, Vol. 6, No. 3, 673-679

# ADMISSIBLE SOLUTIONS OF FINITE STATE SEQUENCE COMPOUND DECISION PROBLEMS

BY STEPHEN B. VARDEMAN

*Purdue University*

A general method of constructing procedures which are both admissible and asymptotically optimal in finite state sequence compound decision problems is suggested and applied to the situation of a two state classification component. When used in an empirical Bayes setting, procedures so constructed are seen to be both admissible and asymptotically optimal.

**1. Introduction.** We consider a situation in which independent structurally identical decision problems are to be faced serially. Numerous authors have produced procedures for various types of component problems satisfying the classical compound optimality criterion but there has been little study of the finite $N$ properties of such asymptotically optimal rules. Indeed it is possible that in some cases procedures exist with better $N$ problem average risk functions for each $N$.

In this paper we give a natural notion of admissibility for sequence compound rules and in the case of a finite state component problem, suggest a method of producing procedures which satisfy both the admissibility criterion and the classical asymptotic optimality criterion. The method is applied to the situation where the component problem is a two state classification problem. The proof of the asymptotic optimality of the resulting admissible sequence compound rule is carried out under a smoothness condition on the two possible distributions of the component problem likelihood ratio statistic and depends upon an estimation result of Gilliland, Hannan and Huang (1976), developed in their study of Bayes procedures in nonsequential versions of the compound problem. Finally we note that the suggested method of constructing good sequence compound rules can also produce admissible, asymptotically optimal empirical Bayes procedures.

**2. Notation and generalities.** We consider a component decision problem with states $\theta \in \Theta$ indexing distributions $P_\theta$ on a sample space $(\mathscr{X}, \mathscr{F})$, possible actions $a \in \mathscr{A}$, loss function $L(\cdot, \cdot)$ and decision rules $d(\cdot)$, measurable functions on $\mathscr{X}$ into $\mathscr{A}$. The risk of a rule $d(\cdot)$ when state $\theta$ holds will be denoted by $R(\theta, d) = \int L(\theta, d(x))\, dP_\theta(x)$ and for a signed measure $G$ on $\Theta$, $R(G, d)$ will abbreviate $\int R(\theta, d)\, dG(\theta)$. $d_G(\cdot)$ will stand for a Bayes rule versus $G$ in the component problem and $R(G)$ will denote the minimum Bayes risk against $G$, $R(G, d_G)$.

Received July 1976; revised July 1977.

*AMS* 1970 *subject classification.* Primary 62C25.

*Key words and phrases.* Sequence compound decision problem, sequence compound admissibility, empirical Bayes decision problem.

The problem addressed here is: What are good procedures when one is to face a sequence of independent decision problems, all with the above structure? Decision rules in such a situation are sequences $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots)$ of measurable functions, $\delta_i(\cdot)$ mapping the first $i$ observations $\mathbf{X}_i = (X_1, \cdots, X_i)$ into an action $a_i$ to be taken in the $i$th problem. For a sequence of states $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)$ and a sequence compound decision rule $\boldsymbol{\delta}$, we will denote the average risk of $\boldsymbol{\delta}$ through the first $N$ problems when $\boldsymbol{\theta}$ holds as

$$R_N(\boldsymbol{\theta}, \boldsymbol{\delta}) = \frac{1}{N} \sum_{i=1}^{N} EL(\theta_i, \delta_i(\mathbf{X}_i))$$

$$= \frac{1}{N} \sum_{i=1}^{N} \int L(\theta_i, \delta_i(\mathbf{x}_i))\, dP_{\boldsymbol{\theta}_i}(\mathbf{x}_i)$$

where $\boldsymbol{\theta}_i$ denotes $(\theta_i, \cdots, \theta_i)$ and $P_{\boldsymbol{\theta}_i} = P_{\theta_1} \times \cdots \times P_{\theta_i}$, the distribution of $\mathbf{X}_i$. $R_N(\boldsymbol{\theta}, \boldsymbol{\delta})$ clearly depends on $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}_N$. For $\mathbf{G}_N$ a signed measure on $\Theta^N$ let $\mathbf{G}_N{}^i$ denote the marginal of $\mathbf{G}_N$ on the first $i$ coordinates of $\Theta^N$. In notation similar to that in the component problem, take as the $N$ problem Bayes risk of the rule $\boldsymbol{\delta}$ against $\mathbf{G}_N$

$$R_N(\mathbf{G}_N, \boldsymbol{\delta}) = \int \frac{1}{N} \sum_{i=1}^{N} \int L(\theta_i, \delta_i(\mathbf{x}_i))\, dP_{\boldsymbol{\theta}_i}(\mathbf{x}_i)\, d\mathbf{G}_N(\boldsymbol{\theta}_N)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \int\int L(\theta_i, \delta_i(\mathbf{x}_i))\, dP_{\boldsymbol{\theta}_i}(\mathbf{x}_i)\, d\mathbf{G}_N{}^i(\boldsymbol{\theta}_i)\,.$$

The classical optimality criterion for a sequence compound procedure is that its $N$ problem risk be asymptotically no larger than the minimum that could be obtained if before facing any decisions one was furnished with $E_N$, the empiric distribution of states $\theta_1$ through $\theta_N$, and determined to choose a fixed $d(\cdot)$ and in the $i$th problem take action $d(X_i)$. That is,

DEFINITION 2.1. A sequence compound procedure $\boldsymbol{\delta}$ is called s.c. optimal provided

$$\limsup_N (R_N(\boldsymbol{\theta}, \boldsymbol{\delta}) - R(E_N)) \leqq 0\,.$$

As indicated before, taken alone such a definition of optimality is open to criticism on the basis that when considered as a function of $\boldsymbol{\theta}_N$, the $N$ problem risk function of an s.c. optimal rule $\boldsymbol{\delta}$, $R_N(\boldsymbol{\theta}, \boldsymbol{\delta})$ may well be inadmissible for each $N$. Hence

DEFINITION 2.2. A sequence compound procedure $\boldsymbol{\delta}$ will be called s.c. admissible provided that for each $N$ there is no $\boldsymbol{\delta}^*$ with $R_N(\boldsymbol{\theta}, \boldsymbol{\delta}^*) \leqq R_N(\boldsymbol{\theta}, \boldsymbol{\delta})$ for all $\boldsymbol{\theta}$ with strict inequality for some $\boldsymbol{\theta}$.

S.c. admissibility does not imply s.c. optimality. The main result of this paper is the demonstration of sequence compound rules for a two state classification problem component which satisfy both Definitions 2.1 and 2.2.

**3. Considerations for finite $\Theta$.** For $\Theta = \{1, 2, \cdots, m\}$ a method of showing the s.c. admissibility of a procedure $\boldsymbol{\delta}^\circ$ would be to produce for each $N$ a

distribution $\mathbf{G}_N$ on $\Theta^N$ such that $\mathbf{G}_N(\boldsymbol{\theta}_N) > 0$ for each $\boldsymbol{\theta}_N \in \Theta^N$ and such that $\boldsymbol{\delta}^\circ$ minimizes $R_N(\mathbf{G}_N, \cdot)$. For $\boldsymbol{\delta}^\circ$ to minimize $R_N(\mathbf{G}_N, \cdot)$ it is necessary and sufficient that $\delta_i{}^\circ$ minimize

$$(1) \qquad \sum_{\boldsymbol{\theta}_i \in \Theta^i} \int L(\theta_i, \delta_i(\mathbf{x}_i))\, dP_{\boldsymbol{\theta}_i}(\mathbf{x}_i) \mathbf{G}_N{}^i(\boldsymbol{\theta}_i)$$

for each $i = 1, 2, \cdots, N$ over choices of measurable maps $\delta_i$ from $\mathscr{X}^i$ to $\mathscr{A}$. (In the terminology of Gilliland and Hannan (1969) such a $\delta_i{}^\circ$ would be Bayes versus $\mathbf{G}_N{}^i$ in a $\Gamma^i$ decision problem.) Notice that for $\mu$ a sigma finite measure dominating $P_1, \cdots, P_m$ and $f_\theta = dP_\theta/d\mu$, subject to measurability considerations, the choice of $\delta_i$ as

$$(2) \qquad \delta_i(\mathbf{x}_i) = \text{an } a \text{ which minimizes } \sum_{\boldsymbol{\theta}_i \in \Theta^i} L(\theta_i, a) \mathbf{G}_N{}^i(\boldsymbol{\theta}_i) \prod_{j=1}^i f_{\theta_j}(x_j)$$

will minimize (1). It is informative to rewrite (2) as

$$(3) \qquad \delta_i(\mathbf{x}_i) = \text{an } a \text{ which minimizes}$$
$$\sum_{k=1}^m f_k(x_i) L(k, a) \Big( \sum_{\boldsymbol{\theta}_i \in \Theta^i \ni \theta_i = k} \mathbf{G}_N{}^i(\boldsymbol{\theta}_i) \prod_{j=1}^{i-1} f_{\theta_j}(x_j) \Big)$$

(interpreting the empty product as 1 in the case $i = 1$), because abbreviating $\sum_{\boldsymbol{\theta}_i \in \Theta^i \ni \theta_i = k} \mathbf{G}_N{}^i(\boldsymbol{\theta}_i) \prod_{j=1}^{i-1} f_{\theta_j}(x_j)$ to $w_{k,i}(\mathbf{G}_N{}^i)$, it is then apparent that for fixed $\mathbf{x}_{i-1}$, $\delta_i(\mathbf{x}_i)$ is a component problem Bayes rule against a measure giving mass $w_{k,i}(\mathbf{G}_N{}^i)$ to each state $k = 1, \cdots, m$.

Now a standard method of producing s.c. optimal rules in finite state settings is, at problem $i$, to estimate $E_{i-1}$ in some consistent fashion, say by $\hat{E}_{i-1}$ and to take action $d_{\hat{E}_{i-1}}(X_i)$ (see for example Hannan (1956), (1957), Van Ryzin (1966) or Vardeman (1975)). This suggests that to produce a procedure satisfying Definitions 2.1 and 2.2 one might search for a sequence of distributions $(\mathbf{G}_1, \mathbf{G}_2, \cdots)$ such that

(a) $\mathbf{G}_i$ is a distribution on $\Theta^i$ such that $\mathbf{G}_i(\boldsymbol{\theta}_i) > 0$ for all $\boldsymbol{\theta}_i \in \Theta^i$,

(b) $\mathbf{G}_{i-1}$ is the marginal of $\mathbf{G}_i$ on the first $i - 1$ coordinates of $\Theta^i$, and

(c) when normalized the weights $w_{1,i}(\mathbf{G}_i), \cdots, w_{m,i}(\mathbf{G}_i)$ give a consistent estimate of $E_{i-1}$.

One might then take $\delta_i$ to be of form (3) with $\mathbf{G}_i$ replacing $\mathbf{G}_N{}^i$ and attempt to prove s.c. optimality for the resulting rule. We proceed to carry out such a program for a two state classification component problem.

**4. Admissible, asymptotically optimal two state classification rules.** For this section we specialize to the case where $\Theta = \{0, 1\}$ and $P_0$ and $P_1$ are distinct probability measures on $(\mathscr{X}, \mathscr{F})$. Let $\mu = P_0 + P_1$, $0 \leqq f_i = dP_i/d\mu \leqq 1$, $i = 1, 2$ and note that the extended real valued $\rho = f_1/f_0$ is well defined (i.e., not 0/0) a.e. $\mu$. Take $\mathscr{A} = \Theta$ and assume the loss structure

$$\begin{aligned} L(\theta, a) &= 0 \quad \text{for} \quad \theta = a \\ &= L_0 \quad \text{for} \quad \theta = 0 \quad \text{and} \quad a = 1 \\ &= L_1 \quad \text{for} \quad \theta = 1 \quad \text{and} \quad a = 0 \end{aligned}$$

where $L_0$ and $L_1$ are positive real numbers. $L$ will abbreviate the ratio $L_0/L_1$.

An intuitively appealing sequence of distributions on $\Theta, \Theta^2, \cdots$, for which there is available the kind of consistency result for the weights $w_{k,i}(\mathbf{G}_i)$ alluded to in Section 3, was suggested by Robbins (1951) in his original treatment of the nonsequential compound decision problem. Denote by $\mathbf{H}_i$ the probability on $\Theta^i$ which is symmetric and places equal mass on the subsets of $\Theta^i$ defined by $\sum_{j=1}^{i}\theta_j = \alpha$, $\alpha = 0, 1, \cdots, i$ (that is for which $\mathbf{H}_i(\{(\theta_1, \cdots, \theta_i)\}) = ((i+1)\binom{i}{\alpha})^{-1}$). It is a simple exercise to show that $\mathbf{H}_{i-1}$ is the marginal distribution of $\mathbf{H}_i$ on the first $i-1$ coordinates of $\Theta^i$ and $\mathbf{H}_i(\{\boldsymbol{\theta}_i\}) > 0$ for each $\boldsymbol{\theta}_i \in \Theta^i$. Further, with $p_i$ denoting $\alpha/i$, $\Delta$ abbreviating $\int f_1\, d(P_1 - P_0)$, and $w_{k,i}$ standing for $\sum_{\boldsymbol{\theta}_i \in \Theta^i \ni \theta_i = k} \mathbf{H}_i(\boldsymbol{\theta}_i) \prod_{j=1}^{i-1} f_{\theta_j}(X_j)$ for $k = 0$ and 1, Gilliland, Hannan, and Huang (1976) prove the following.

PROPOSITION 4.1. *For $\mathbf{X}_{i-1}$ distributed as $P_{\boldsymbol{\theta}_{i-1}}$*

$$E\left|\frac{w_{0,i}}{w_{0,i}+w_{1,i}} - (1-p_{i-1})\right| = E\left|\frac{w_{1,i}}{w_{0,i}+w_{1,i}} - p_{i-1}\right| \leqq \left(\frac{4\Delta^{-1}(2\pi(i-1))^{\frac{1}{2}}+1}{i+1}\right)$$

*for any $\boldsymbol{\theta}_{i-1} \in \Theta^{i-1}$.*

Thus with $\hat{p}_{i-1} = w_{1,i}(w_{0,i}+w_{1,i})^{-1}$, the sequence compound procedure $\phi = (\phi_1, \phi_2, \cdots)$ defined by

$$\phi_i(\mathbf{X}_i) = I\left[\rho(X_i) \geqq L\,\frac{1-\hat{p}_{i-1}}{\hat{p}_{i-1}}\right]$$

is s.c. admissible and the estimation result lends hope of proving uniform s.c. optimality at a good rate. Adding a condition on the possible distributions of $\rho(X)$ we will prove,

THEOREM 4.2. *For $k = 0, 1$ let $\nu_k$ denote the distribution of $L(L+\rho(X))^{-1}$ for $X$ with distribution $P_k$. If there exists a $\gamma \in (0, 1]$ and real number $C$ such that for any two real numbers $0 \leqq a \leqq b \leqq 1$, $\nu_k([a, b]) \leqq C(b-a)^{\gamma}$, then there exists a real number $\mathscr{K}$ depending only on $\Delta$, $\max(L_0, L_1)$, and $C$ such that*

$$R_N(\boldsymbol{\theta}, \boldsymbol{\phi}) - R(E_N) \leqq \mathscr{K} N^{-\gamma/2}.$$

PROOF. It is standard in proofs of s.c. optimality (see for example Vardeman (1975)) to note that $(1/N)\sum_{i=1}^{N} R(\theta_i, d_{E_i}) \leqq R(E_N)$ so that

$$R_N(\boldsymbol{\theta}, \boldsymbol{\phi}) - R(E_N) \leqq \frac{1}{N}\sum_{i=1}^{N} E(L(\theta_i, \phi_i(\mathbf{X}_i)) - L(\theta_i, d_{E_i}(X_i))). \tag{4}$$

The function $d_{E_i}(x) = I[\rho(x) \geqq L(1-p_i)p_i^{-1}]$ is component Bayes versus $E_i$ so that the right side of (4) may be bounded by

$$\frac{1}{N}\{L_0 \sum_{i \ni \theta_i = 0} EI[L(1-\hat{p}_{i-1})\hat{p}_{i-1}^{-1} \leqq \rho(X_i) < L(1-p_i)p_i^{-1}] + L_1 \sum_{i \ni \theta_i = 1} EI[L(1-p_i)p_i^{-1} \leqq \rho(X_i) < (1-\hat{p}_{i-1})\hat{p}_{i-1}^{-1}]\}. \tag{5}$$

But now consider a typical summand above.

$$EI[L(1-\hat{p}_{i-1})\hat{p}_{i-1}^{-1} \leqq \rho(X_i) < L(1-p_i)p_i^{-1}] = EE[I[p_i < L(L+\rho(X_i))^{-1} \leqq \hat{p}_{i-1}] \mid \mathbf{X}_{i-1}] \leqq EC|\hat{p}_{i-1} - p_i|^{\gamma}$$

by the assumption on the distribution of $\rho(X_i)$. Further for $i > 1$

$$E|\hat{p}_{i-1} - p_i|^\gamma \leqq E(|\hat{p}_{i-1} - p_{i-1}| + |p_i - p_{i-1}|)^\gamma$$
$$\leqq (E|\hat{p}_{i-1} - p_{i-1}| + |p_i - p_{i-1}|)^\gamma .$$

So applying Proposition 4.1 together with the fact that $|p_i - p_{i-1}| \leqq (i-1)^{-1}$ for $i > 1$, we have $R_N(\boldsymbol{\theta}, \boldsymbol{\phi}) - R(E_N)$ is no larger than

$$C \max(L_0, L_1) \frac{1}{N} (1 + \textstyle\sum_{i=2}^{N} ((4\Delta^{-1}(2\pi(i-1))^{\frac{1}{2}} + 1)(i+1)^{-1} + (i-1)^{-1})^\gamma)$$

and the result follows. □

Several comments are in order. The first is that the $\gamma = 1$ version of the condition on the distribution of $L(L + \rho(X))^{-1}$ is similar to one used by Hannan and Van Ryzin (1965) in an investigation of a nonsequential compound classification problem and can be verified by showing $\nu_0$ and $\nu_1$ have bounded densities with respect to Lebesgue measure. Such is the case for example for $P_0$ the normal $(0, 1)$ distribution and $P_1$ the normal $(\beta, 1)$ distribution. The second is that proof obviously carries over practically verbatim to any other sequence of distributions $G_1, \mathbf{G}_2, \cdots$ for which the weights $w_{k,i}(\mathbf{G}_i)$ are consistent for $E_{i-1}$ at an $i^{-\frac{1}{2}}$ rate. Gilliland, Hannan and Huang (1976) have considered distributions $\mathbf{G}_i$ defined by $\mathbf{G}_i(\boldsymbol{\theta}_i) = \int t^\alpha (1-t)^{i-\alpha}\, d\Lambda(t)$ with $\alpha = \sum_{j=1}^{i} \theta_j$ for a wide class of probabilities $\Lambda$ on $(0, 1)$ and proved analogous of Proposition 4.1 for such $\mathbf{G}_i$. Note that since for $\boldsymbol{\theta}_{i-1} \in \Theta^{i-1}$ with

$$\beta = \textstyle\sum_{j=1}^{i-1} \theta_j \quad \text{we have}$$
$$\mathbf{G}_i^{i-1}(\boldsymbol{\theta}_{i-1}) = \int (t^\beta (1-t)^{i-\beta} + t^{\beta+1}(1-t)^{i-\beta-1})\, d\Lambda(t) = \mathbf{G}_{i-1}(\boldsymbol{\theta}_{i-1})$$

such sequences of distributions can be used to produce whole classes of s.c. admissible, s.c. optimal classification rules.

**5. Admissible finite state empirical Bayes procedures.** The attractive sequence compound properties of the kind of procedures discussed in the previous section can carry over to empirical Bayes problems. That is, consider a situation where $(\theta_1, X_1), (\theta_2, X_2), \cdots, (\theta_N, X_N)$ are i.i.d. with $\theta_i$ distributed according to some unknown prior $G$ on $\Theta$ and the conditional distribution of $X_i$ given $\theta_i$ is $P_{\theta_i}$. Based on the observable $\mathbf{X}_N = (X_1, \cdots, X_N)$ an action is to be taken and loss $L(\theta_N, a)$ suffered. Many authors have considered procedures of the form $d_{\hat{G}}(X_N)$ where $\hat{G}$ is an estimate of $G$ based on $\mathbf{X}_{N-1}$ and proved

$$\limsup EL(\theta_N, d_{\hat{G}}(X_N)) - R(G) \leqq 0 .$$

Such an asymptotic optimality property does not guarantee empirical Bayes admissibility. That is, there may well be $\delta(\mathbf{X}_N)$ for which

$$E[L(\theta_N, d_{\hat{G}}(X_N)) - L(\theta_N, \delta(\mathbf{X}_N))] \geqq 0$$

for each $G$ with strict inequality for at least one $G$. (This notion of empirical

Bayes inadmissibility was used by Meeden (1972) and recently explored extensively in finite $\Theta$ problems by Boyer (1976).) But for finite $\Theta$ component problems, carrying out the program described in Section 3 can produce rules that are not only s.c. admissible and s.c. optimal but also both admissible and asymptotically optimal in the empirical Bayes problem.

Boyer (1976) has noted that in finite state problems, decision rules which are "second level Bayes" versus priors with large support are admissible empirical Bayes rules. That is, with $\mathscr{G}$ standing for the $m$ dimensional probability simplex and $\mathbf{g} \in \mathscr{G}$ the $m$ vector $(g_1, g_2, \cdots, g_m)$, if $\Lambda$ is a probability with support all of $\mathscr{G}$ and $\delta^\circ(\cdot)$ minimizes

$$\int \sum_{\boldsymbol{\theta}_N \in \Theta^N} \int L(\theta_N, \delta(\mathbf{X}_N))\, dP_{\boldsymbol{\theta}_N}(\mathbf{X}_N) \prod_{i=1}^N g_{\theta_i}\, d\Lambda(\mathbf{g}) \tag{6}$$

and has finite risk (6), then $\delta^\circ$ is an admissible empirical Bayes rule. But interchanging the sum and integration with respect to $\Lambda$ it is immediate that if $\mathbf{G}_N$ is of the form $\mathbf{G}_N(\boldsymbol{\theta}_N) = \int \prod_{i=1}^N g_{\theta_i}\, d\Lambda(\mathbf{g})$ and $\delta_N(\cdot)$ is of the form (3) with $i = N$, then $\delta_N(\mathbf{X}_N)$ is an admissible empirical Bayes rule. If in addition a result like Proposition 4.1 holds, subject to measurability considerations the following lemma can be used to establish a rate of convergence to asymptotic optimality for $\delta_N(\mathbf{X}_N)$.

LEMMA 5.1. *Let $X$ have conditional distribution $P_\theta$ given $\theta$, $\theta$ have distribution $G$ on $\Theta = \{1, 2, \cdots, m\}$ and the pair $(\theta, X)$ be independent of a random vector $V = (V_1, \cdots, V_m)$. Suppose $0 \leqq L(\theta, a)$, for any $W \in R^m$ $d_W$ exists and $\int L(\theta, d_W(x))\, dP_\theta(x) \leqq B$. Then*

$$EL(\theta, d_V(X)) - R(G) \leqq B \sum_{k=1}^m E|G(\{k\}) - V_k| .$$

PROOF. Iterating expectations

$$\begin{aligned} EL(\theta, d_V(X)) - R(G) &= EE[L(\theta, d_V(X)) - R(G) \mid V] , \\ &= E(R(G, d_V) - R(G, d_G)) . \end{aligned} \tag{7}$$

But by the minimizing property of a component Bayes rule the right side of (7) is bounded by

$$\begin{aligned} \text{r.h.s. } (7) &\leqq E(R(G, d_V) - R(G, d_G) - (R(V, d_V) - R(V, d_G))) . \\ &= \sum_{k=1}^m E(G(\{k\}) - V_k)(R(k, d_V) - R(k, d_G)) . \\ &\leqq BE \sum_{k=1}^m |G(\{k\}) - V_k| . \end{aligned}$$

□

For example, returning to the two state classification component of Section 4, since the Robbins prior can be expressed in the form $\mathbf{H}_i(\boldsymbol{\theta}_i) = \int t^\alpha (1 - t)^{1-\alpha}\, dt$ with $\alpha = \sum_{j=1}^i \theta_j$, $\phi_N(\mathbf{X}_N)$ is admissible in the empirical Bayes classification problem, and since for fixed $\mathbf{X}_{N-1}$, $\phi_N(\mathbf{X}_N)$ is component Bayes versus the prior giving weights $1 - \hat{p}_{N-1}$ and $\hat{p}_{N-1}$ to states 0 and 1 respectively, by the lemma

$$EL(\theta_N, \phi_N(\mathbf{X}_N)) - R(G) \leqq 2 \max(L_0, L_1) E|G(\{1\}) - \hat{p}_{N-1}| .$$

But triangulating about $p_{N-1} = (1/(N-1)) \sum_{i=1}^{N-1} \theta_i$, applying Proposition 4.1

and the moment inequality

$$E|G(\{1\}) - \hat{p}_{N-1}| \leqq \frac{4\Delta^{-1}(2\pi(N-1))^{\frac{1}{2}} + 1}{N+1} + \left(\frac{G(\{1\})(1-G(\{1\}))}{N-1}\right)^{\frac{1}{2}}.$$

It is typical of results in this area that the empirical Bayes optimality of $\phi$ follows under less stringent assumptions than the s.c. optimality, that is, no regularity of the distribution of $\rho(X)$ is needed.

**Acknowledgment.** The author wishes to thank Professor James Hannan for his help in simplifying a version of Lemma 5.1 developed while the author was a student at Michigan State University.

## REFERENCES

BOYER, JOHN (1976). Some admissibility considerations in the finite state component compound and empirical Bayes decision problems. Ph. D. thesis, Michigan State University.

GILLILAND, DENNIS C. and HANNAN, JAMES F. (1969). On the extended compound decision problem. *Ann. Math. Statist.* **40** 1536–1541.

GILLILAND, DENNIS C., HANNAN, JAMES F. and HUANG, J. S. (1976). Asymptotic solutions to the two state component compound decision problems, Bayes versus diffuse priors on proportions. *Ann. Statist.* **4** 1101–1112.

HANNAN, JAMES F. (1956). The dynamical statistical decision problem when the component problem involves a finite number, $m$, of distributions. (Abstract.) *Ann. Math. Statist.* **27** 212.

HANNAN, JAMES F. (1957). Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games* **3** 97–139 (M. Dresher and A. W. Tucker, eds.). Princeton University Press.

HANNAN, JAMES F. and VAN RYZIN, J. R. (1965). Rates of convergence in the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **36** 1743–1752.

MEEDEN, GLEN (1972). Some admissible empirical Bayes procedures. *Ann. Math. Statist.* **43** 96–101.

ROBBINS, HERBERT (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probability* 131–148. Univ. California Press.

VAN RYZIN, J. R. (1966). The sequential compound decision problem with $m \times n$ finite loss matrix. *Ann. Math. Statist.* **37** 954–975.

VARDEMAN, STEPHEN B. (1975). $O(N^{\frac{1}{2}})$ convergence in the finite state restricted risk component sequence compound decision problem. RM-349, Statistics and Probability, Michigan State Univ.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
MATHEMATICAL SCIENCES BUILDING
WEST LAFAYETTE, INDIANA 47907