THE SEQUENTIAL COMPOUND DECISION PROBLEM WITH $m \times n$ FINITE LOSS MATRIX[1]

By J. Van Ryzin

Argonne National Laboratory

1. Introduction and summary. Consideration of a sequence of statistical decision problems having identical generic structure constitutes a sequential compound decision problem. The risk of a sequential compound decision problem is defined as the average risk of the component problems. In the case where the component decisions are between two fully specified distributions P_1 and P_2, $P_1 \neq P_2$, Samuel (Theorem 2 of [9]) gives a sequential decision function whose risk is bounded from above by the risk of a best "simple" procedure based on knowing the proportion of component problems in which P_2 is the governing distribution plus a sequence of positive numbers converging to zero uniformly in the space of parameter-valued sequences as the number of problems increases. Related results are abstracted by Hannan in [2] for the sequential compound decision problem where the parameter space in the component problem is finite. The decision procedures in both instances rely on the technique of "artificial randomization," which was introduced and effectively used by Hannan in [1] for sequential games in which player I's space is finite. In the game situation such randomization is necessary. However, in the compound decision problem such "artificial randomization" is not necessary as is shown in this paper.

Specifically, we consider the case where each component problem consists of making one of n decisions based on an observation from one of m distributions. Theorems 4.1, 4.2, and 4.3 give upper bounds for the difference in the risks (the regret function) of certain sequential compound decision procedures and a best "simple" procedure which is Bayes against the empirical distribution on the component problem parameter space. None of the sequential procedures presented depend on "artificial randomization." The upper bounds in these three theorems are all of order $N^{-\frac{1}{2}}$ and are uniform in the parameter-valued sequences. All procedures depend at stage k on substitution of estimates of the $k - 1$st (or kth) stage empirical distribution p_{k-1} (or p_k) on the component parameter space into a Bayes solution of the component problem with respect to p_{k-1} (or p_k). Theorem 4.1 (except in the case where the estimates are degenerate) and Theorem 4.3 when specialized to the compound testing case between P_1 and P_2 (Theorems 5.1 and 5.2) yield a threefold improvement of Samuel's results mentioned above by simultaneously eliminating the "artificial randomization," by improving the convergence rate of the upper bound of the regret function to $N^{-\frac{1}{2}}$, and by widening the class of estimates.

Received 5 February 1965; revised 4 March 1966.

[1] Work performed under the auspices of the United States Atomic Energy Commission.

Higher order uniform bounds on the regret function in the sequential compound testing problem are also given. The bounds in Theorems 5.3 and 5.4 (or Theorems 5.5 and 5.6) are respectively of $O((\log N)N^{-1})$ and $o(N^{-\frac{1}{2}})$ and are attained by imposing suitable continuity assumptions on the induced distribution of a certain function of the likelihood ratio of P_1 and P_2. Theorem 6.1 extends Theorems 4.1, 4.2, and 4.3 to the related "empirical Bayes" problem. Also lower bounds of equivalent or better order are given for all theorems.

The next section introduces notation and preliminaries to be used in this paper and in the following paper [15].

2. Preliminaries.

2.1. *The Problem.* Consider the following finite statistical decision problem. Let X be a random variable taking values in a measure space $(\mathfrak{X}, \mathfrak{F}, \mu)$ which is known to have one of m possible distributions in the finite class $\mathcal{P} = \{P_\theta ; \theta \varepsilon \Omega\}$, $\Omega = \{1, \cdots, m\}$. Based on observing X we are required to make a decision $d \varepsilon \mathfrak{D} = \{1, \cdots, n\}$ incurring loss $L(\theta, d)$ if we make decision $d \varepsilon \mathfrak{D}$ when X is distributed as P_θ, $\theta = 1, \cdots, m$, $d = 1, \cdots, n$. Let $\mathbf{L} = (L(\theta, d))$ be the $m \times n$ matrix of losses.

Suppose that such a statistical decision problem is repeated sequentially with Ω, $\mathfrak{D}$, $(\mathfrak{X}, \mathfrak{F}, \mu)$ and $\mathbf{L}$ remaining the same in each problem. Specifically, let $\mathbf{X} = \{X_k ; k = 1, 2, \cdots\}$ be a sequence of independent random variables X_k with values in $(\mathfrak{X}, \mathfrak{F}, \mu)$ where X_k is distributed as P_{θ_k}, θ_k ranging in Ω. Let $\boldsymbol{\theta} = \{\theta_k ; \theta_k \varepsilon \Omega, k = 1, 2, \cdots\}$ be a sequence of parameter values on Ω. Let $\boldsymbol{\Omega}$ be the space of all such sequences.

In the sequential situation $\mathbf{X}_k = (X_1, \cdots, X_k)$ is known to the statistician at stage k. Hence, a decision procedure for the statistician in such a repetitive sequential situation is a sequence $\mathbf{t} = \{\mathbf{t}_k\}$ of decision functions $\mathbf{t}_k = \mathbf{t}_k(\mathbf{X}_k) = (\mathbf{t}_{k1}(\mathbf{X}_k), \cdots, \mathbf{t}_{kn}(\mathbf{X}_k))$ where $\mathbf{t}_{kd}(\mathbf{X}_k) \geqq 0$ represents the probability of selecting decision d after $\mathbf{X}_k$ is observed. For slightly greater generality, so that the notation will apply also in the problem of [15] where the statistician is given both $\boldsymbol{\theta}_{k-1} = (\theta_1, \cdots, \theta_{k-1})$ and $\mathbf{X}_k$ at the kth stage we may write

$$(2.1) \qquad \mathbf{t}_k = \mathbf{t}_k(\boldsymbol{\theta}, \mathbf{X}_k) = (\mathbf{t}_{k1}(\boldsymbol{\theta}, \mathbf{X}_k), \cdots, \mathbf{t}_{kn}(\boldsymbol{\theta}, \mathbf{X}_k)),$$

$$\mathbf{t}_{kd}(\boldsymbol{\theta}, \mathbf{X}_k) \geqq 0, \qquad \textstyle\sum_{d=1}^{n} \mathbf{t}_{kd}(\boldsymbol{\theta}, \mathbf{X}_k) = 1,$$

where $\mathbf{t}_k$ is a probability distribution on $\mathfrak{D} = \{1, \cdots, n\}$ with the dependence on $\boldsymbol{\theta}$ in accord with the knowledge of the θ's available to the statistician at stage k. Thus in this paper we shall always have $\mathbf{t}_k = \mathbf{t}_k(\mathbf{X}_k)$ while in [15] we shall always have $\mathbf{t}_k = \mathbf{t}_k(\boldsymbol{\theta}_{k-1}, \mathbf{X}_k)$. Such a procedure is said to be *simple* if there exists a vector of n real-valued measurable functions on $(\mathfrak{X}, \mathfrak{F}, \mu)$, $t(x) = (t_1(x), \cdots, t_n(x))$, with $\sum_{d=1}^{n} t_d(x) = 1$, $t_d(x) \geqq 0$, such that $\mathbf{t}_{kd}(\boldsymbol{\theta}, \mathbf{X}_k) = t_d(X_k)$, $k = 1, 2, \cdots; d = 1, \cdots, n$. Hence, the kth decision depends only on X_k, the kth observed random variable. A simple rule will be denoted by the vector function $t = (t_1, \cdots, t_n)$.

Before discussing such procedures further we shall introduce some notation.

2.2. *Notation.* For $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, let $\mathbf{P}$ denote the product probability measure $\times_{i=1}^{\infty} P_{\theta_i}$. For fixed k, let $\boldsymbol{\theta}_k = (\theta_1, \cdots, \theta_k)$ and $\mathbf{P}_k = \times_{i=1}^{k} P_{\theta_i}$. Expectation with respect to $\mathbf{P}$ and $\mathbf{P}_k$ will be denoted by $\mathbf{E}$ and $\mathbf{E}_k$ respectively. Also, expectation with respect to P_θ will be denoted by E_θ for each $\theta \varepsilon \Omega$.

The following notation relating to the $m \times n$ matrix of losses $\mathbf{L} = (L(\theta, d))$ is convenient. Let L_θ and L^d denote respectively the θth row and dth column of this matrix, $\theta = 1, \cdots, m$, $d = 1, \cdots, n$. The ordered difference of the dth and d'th columns $L^d - L^{d'}$ will be denoted $L^{dd'}$ with components $L_\theta^{dd'} = L(\theta, d) - L(\theta, d')$. Let $L = \max_{\theta, d, d'} |L_\theta^{dd'}|$ and $L^* = \max_{\theta, d} |L(\theta, d)|$.

Let E^m be m-dimensional Euclidean space. If $u = (u_1, \cdots, u_m)$ and $v = (v_1, \cdots, v_m)$ are two vectors in E^m, define $uv = (u_1v_1, \cdots, u_mv_m)$ and $(u, v) = \sum_{i=1}^m u_iv_i$ as the vector of componentwise products and the usual inner product respectively. The usual norm in E^m is given by $\|u\| = (u, u)^{\frac{1}{2}}$. Let $\epsilon_i = (\delta_{i1}, \cdots, \delta_{im})$, where $\delta_{ij} = 0$ or 1 as $i \neq j$ or $i = j$, be the ith basis vector in E^m, $i = 1, \cdots, m$. The vector of all zeros in E^m is denoted by $\mathbf{0}$ and the vector of all ones by $\mathbf{1}$. The dimension in which these operations are carried out will always be clear from the context.

Next, for each positive integer and each $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, define $p_k(\boldsymbol{\theta}) = k^{-1}\sum_{i=1}^k \epsilon_{\theta_i}$, where ϵ_i is the ith basis vector in E^m. Note that $p_k(\boldsymbol{\theta}) = (p_{k1}(\boldsymbol{\theta}), \cdots, p_{km}(\boldsymbol{\theta}))$, where $p_{k\theta}(\boldsymbol{\theta}) = k^{-1}\sum_{i=1}^k \delta_{\theta_i\theta}$ is the relative frequency of θ_i's equal to θ in the first k members of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_k = (\theta_1, \cdots, \theta_k)$, $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. We shall call $p_k(\boldsymbol{\theta})$ the *k-stage empirical distribution* on Ω for $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. We shall often write $p_k(\boldsymbol{\theta})$ simply as $p_k = (p_{k1}, \cdots, p_{km})$ suppressing $\boldsymbol{\theta}$ therein.

We shall assume throughout that the measure μ boundedly dominates the class P. That is, for every $\theta \varepsilon \Omega$,

$$f_\theta(x) = (dP_\theta/d\mu)(x) \leqq K \qquad \text{a.e.}\mu \tag{2.2}$$

for some positive finite K. There is no loss in generality in this assumption since μ may be taken as $\sum_{\theta=1}^m P_\theta$ with $K = 1$. Let $f(x) = (f_1(x), \cdots, f_m(x))$ be the m-vector of densities defined by (2.2).

2.3. *Decision Procedures.* We now return to discussion of procedures $\mathbf{t}$ in (2.1). For fixed N and $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, the risk function of the procedure $\mathbf{t}$ over the N problems is denoted by $R_N(\boldsymbol{\theta}, \mathbf{t})$ and is defined to be the average of the risks in the component problems,

$$\mathbf{E}\{\textstyle\sum_{d=1}^n L(\theta_k, d)\mathbf{t}_{kd}(\boldsymbol{\theta}, \mathbf{X}_k)\} = \mathbf{E}(L_{\theta_k}, \mathbf{t}_k(\boldsymbol{\theta}, \mathbf{X}_k)), \quad k = 1, \cdots, N. \tag{2.3}$$

Hence,

$$R_N(\boldsymbol{\theta}, \mathbf{t}) = \mathbf{E}W_N(\boldsymbol{\theta}, \mathbf{t}), \tag{2.4}$$

where

$$W_N(\boldsymbol{\theta}, \mathbf{t}) = N^{-1}\textstyle\sum_{k=1}^N (L_{\theta_k}, \mathbf{t}_k(\boldsymbol{\theta}, \mathbf{X}_k)) \tag{2.5}$$

is the average loss over the first N problems.

If one uses a simple decision procedure $t = (t_1, \cdots, t_n)$ the risk function (2.4) has a particularly simple form. The loss incurred in using t is

$$W_N(\boldsymbol{\theta}, t) = \textstyle\sum_{\theta=1}^{m} p_{N\theta}(\boldsymbol{\theta})\{\mathrm{AV}_{\theta_k=\theta}(L_{\theta_k}, t(X_k))\}, \tag{2.6}$$

where $\mathrm{AV}_{\theta_k=\theta}$ indicates the numerical average of the $Np_{N\theta}(\boldsymbol{\theta})$ values for which $\theta_k = \theta$. Now since the $(L_{\theta_k}, t(X_k))$ for $\theta_k = \theta$ are independent identically distributed random variables with mean $\rho_\theta(t) = E_\theta(L_\theta, t(X))$, we may express their expected average as $\rho_\theta(t)$ to obtain from (2.6),

$$R_N(\boldsymbol{\theta}, t) = \textstyle\sum_{\theta=1}^{m} p_{N\theta}(\boldsymbol{\theta})\rho_\theta(t) = (p_N(\boldsymbol{\theta}), \rho(t)), \tag{2.7}$$

where $\rho(t) = (\rho_1(t), \cdots, \rho_m(t))$.

The problem of selecting a simple decision procedure to minimize the risk (2.7) is straightforward for fixed $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ and N, i.e., choose t as a Bayes solution in the single stage component problem with $p_N(\boldsymbol{\theta})$ considered as an *a priori* distribution on Ω. However, we shall need slightly more general results later for which we choose to minimize $(\xi, \rho(t))$ for any fixed $\xi \varepsilon E^m$. Recalling (2.2) we may express $\rho_\theta(t) = E_\theta(L_\theta, t) = \int \sum_{d=1}^{n} L(\theta, d) f_\theta(x) t_d(x)\, d\mu(x)$ and hence

$$r(\xi, t) = (\xi, \rho(t)) = \int \textstyle\sum_{d=1}^{n} (\xi, L^d f(x)) t_d(x)\, d\mu(x). \tag{2.8}$$

Therefore (2.8) is minimized for fixed ξ by any vector function $t(\xi, x) = (t_1(\xi, x), \cdots, t_n(\xi, x))$ which is chosen as a probability distribution concentrating on those d's for which $(\xi, L^d f(x))$ is a minimum. That is, $t(\xi, x)$ has components

$$\begin{aligned} t_d(\xi, x) &= 0 && \text{if} \quad (\xi, L^d f(x)) > \min_j(\xi, L^j f(x)) \\ &= 1 && \text{if} \quad (\xi, L^d f(x)) < \min_{j \neq d}(\xi, L^j f(x)) \\ &= \text{arbitrary} && \text{if} \quad (\xi, L^d f(x)) = \min_j(\xi, L^j f(x)), \end{aligned} \tag{2.9}$$

where $\sum_{d=1}^{n} t_d(\xi, x) = 1$ and $t_d(\xi, x) \geqq 0$ a.e.μ. When convenient we shall simply refer to the rule in (2.9) as $t\{\xi\}$. If ξ is an *a priori* distribution on Ω, then $t\{\xi\}$, defined by (2.9) is a Bayes solution in the component problem with respect to ξ. We may always replace (2.9) with the following particular minimizing non-randomized version of $t\{\xi\}$ given by $t^*\{\xi\}$, where

$$\begin{aligned} t_d^*(\xi, x) &= 1 \text{ if } d \text{ is the smallest integer for which} \\ &\qquad (\xi, L^d f(x)) = \min_j (\xi, L^j f(x)) \\ &= 0 \text{ otherwise.} \end{aligned} \tag{2.10}$$

For each $\xi \varepsilon E^m$ define the following function

$$\phi(\xi) = \inf_t r(\xi, t) = r(\xi, t\{\xi\}), \tag{2.11}$$

for any $t\{\xi\}$ given by (2.9). For ξ which are *a priori* distributions on Ω, the function $\phi(\xi)$ in (2.11) is referred to as the *Bayes envelope functional.*

We are now in a position to discuss the criterion by which we shall examine the efficacy of a sequential procedure $\mathbf{t}$ in (2.1). For fixed N, if $p_N(\boldsymbol{\theta})$ were known the best simple rule one can use is a componentwise Bayes rule $t\{p_N(\boldsymbol{\theta})\}$ defined by (2.9) with $\xi = p_N(\boldsymbol{\theta})$. In general, this procedure is unavailable since both N and $p_N(\boldsymbol{\theta})$ must be known in advance. However, consider the function

$$R_N(\boldsymbol{\theta}, \mathbf{t}) - \phi(p_N(\boldsymbol{\theta})). \tag{2.12}$$

This function is called the *regret function* of the sequential procedure $\mathbf{t}$ against the class of all simple procedures. We shall examine sequential procedures $\mathbf{t}$ for which (2.12) converges to zero uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ as N, the number of problems, increases. Hence, in a limiting risk sense such sequential procedures have the advantage of performing for the first N problems about as well as an optimal simple rule based on knowing $p_N(\boldsymbol{\theta})$ for all N sufficiently large, no matter what the sequence $\boldsymbol{\theta} \in \boldsymbol{\Omega}$.

Before stating the procedures $\mathbf{t}$ in (2.1) explicitly and examining the convergence of (2.12) we close this section with the following additional notation and preliminary lemma.

The characteristic function of a set A will be denoted simply by A enclosed in square brackets; that is,

$$\begin{aligned}[A][a] &= 1 \quad \text{if} \quad a \in A \\ &= 0 \quad \text{if} \quad a \notin A.\end{aligned} \tag{2.13}$$

The following useful lemma is a simple consequence of the Berry-Esseen normal approximation theorem (see Loève [5], p. 288).

LEMMA 2.1. *If $Y_1, \cdots, Y_n$ are n independent random variables with mean* 0, *variance $\sigma_i^2 = EY_i^2$ and third absolute moment $\gamma_i = E|Y_i|^3$, then for α and l real, $l \geqq 0$,*

$$\Pr\{\alpha \leqq \textstyle\sum_{i=1}^n Y_i \leqq \alpha + l\} \leqq l\sigma^{-1}(2\pi)^{-\frac{1}{2}} + 2\beta\sigma^{-3}\gamma,$$

where β is an absolute constant, $\sigma^2 = \sum_{i=1}^n \sigma_i^2$, and $\gamma = \sum_{i=1}^n \gamma_i$.

3. Sequential compound decision procedures. Consider again the problem stated in Section 2. If in such a problem the statistician knows at stage k only the first k observations $\mathbf{X}_k = (X_1, \cdots, X_k)$ and has *no* knowledge of the sequence $\boldsymbol{\theta}$, then such a problem is called a *sequential compound decision problem.* Thus the procedure $\mathbf{t}$ in (2.1) at stage k will be functions of $\mathbf{X}_k$ only, i.e., $\mathbf{t}_k(\boldsymbol{\theta}, \mathbf{X}_k) = \mathbf{t}_k(\mathbf{X}_k)$ in (2.1) (see (3.1)).

Such problems were considered previously for $m = n = 2$ by Samuel [9] and [10] and for general m and n results are abstracted in [2]. The relationship between these earlier results and our results will be made specific later. Also see Samuel [11] for results on the related sequential compound estimation problem.

In [2], [9] and [10], the procedures $\mathbf{t}$ in (2.1) were obtained by substituting estimates of $p_k(\boldsymbol{\theta})$ (or $p_{k-1}(\boldsymbol{\theta})$) at stage k and using the corresponding Bayes rule against such an estimate. Specifically, let $\bar{h}_k$ be an unbiased estimate of $p_k(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, that is, $\mathbf{E}\bar{h}_k = p_k(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, $k = 1, 2, \cdots$. Define $\bar{h}_0 = \bar{\epsilon}_0 = \mathbf{0}$, the zero vector in E^m. Consider the following two strongly (N need not be known in advance) sequential procedures:

$$\begin{aligned} \text{(3.1)} \qquad \mathbf{t}^* &= \{t^*(\bar{h}_{k-1}, X_k); k = 1, 2, \cdots\}, \\ \mathbf{t}^{**} &= \{t^*(\bar{h}_k, X_k); k = 1, 2, \cdots\}, \end{aligned}$$

where $t^*(\xi, x) = (t_1^*(\xi, x), \cdots, t_n^*(\xi, x))$ with $t_d^*(\xi, x)$ defined in (2.10) for $\xi \varepsilon E^m$, $x \varepsilon \mathfrak{X}$, $d = 1, \cdots, n$. The regret function (2.12) of the procedures $\mathbf{t}^*$ and $\mathbf{t}^{**}$ will be shown to have upper bounds $c'N^{-\frac{1}{2}}$ and $c''N^{-1}$, respectively, for suitable chosen classes of estimates $\bar{h}_k$. The c' and c'' in both cases will be independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.

The question of obtaining estimates $\bar{h}_k$ has already been solved in [14] (see also [8]) in the following sense. If there exists a vector function $h(x) = (h_1(x), \cdots, h_m(x))$ on $\mathfrak{X}$ such that $E_\theta\{h_j(X)\} = \delta_{\theta j}$, where $\delta_{\theta j} = 0$ or 1 according as $\theta \neq j$ or $\theta = j$, then the estimate

$$\text{(3.2)} \qquad \bar{h}_k = \bar{h}_k(\mathbf{X}_k) = k^{-1}\textstyle\sum_{i=1}^k h(X_i)$$

is an unbiased estimate of $p_k(\boldsymbol{\theta})$ for all k and $\boldsymbol{\theta}$, that is $\mathbf{E}\bar{h}_k(\mathbf{X}_k) = p_k(\boldsymbol{\theta})$. It is shown in ([14], Lemma 3 and Corollary 3) that such a function $h = h(x)$ exists if and only if the set of densities $\{f_1, \cdots, f_m\}$ in (2.2) are linearly independent in $L_1(\mu)$, the space of all μ-integrable functions. In fact, because of (2.2) there always exists an h such that $\max_j |h_j(x)| \leqq H < \infty$ a.e.μ if the set $\{f_1, \cdots, f_m\}$ are linearly independent in $L_1(\mu)$. (See Corollary 2 of [14]). For a method of obtaining such functions h as well as the proofs of the results stated above see Section 3 of [14]. Henceforth in this paper we make the assumption that

(A_1) The set of densities $\{f_1, \cdots, f_m\}$ is a set of linearly independent functions in $L_1(\mu)$.

Under (A_1) we may define by the results of [14] a non-empty class of vector functions

$$\mathfrak{H} = \{h = (h_1, \cdots, h_m) \mid h_j \varepsilon L_2(\mu), E_\theta h_j(X) = \delta_{\theta j}, \theta, j = 1, \cdots, m\}$$

with $L_2(\mu)$ the space of all μ-square integrable functions. Throughout the remainder of the paper we shall confine ourselves to estimates $\bar{h}_k$ of the form (3.2) for suitably chosen $h \varepsilon \mathfrak{H}$. For such estimates the following simple inequality is verified in [14] (Lemma 4): If $h \varepsilon \mathfrak{H}$, then for $k \geqq 1$,

$$\text{(3.3)} \qquad \mathbf{E}\|\bar{h}_k(\mathbf{X}_k) - p_k(\boldsymbol{\theta})\| \leqq \sigma k^{-\frac{1}{2}},$$

where $\sigma^2 = \max_\theta E_\theta\|h(X) - \epsilon_\theta\|^2$.

We comment that the functions $h \varepsilon \mathfrak{H}$ have (necessarily to satisfy $E_\theta h_j(X) = \delta_{\theta j}$) components h_j which can take negative values and thus the reason for defining (2.8)–(2.11) for all $\xi \varepsilon E^m$ as well as ξ's which are proper *a priori* distributions.

We now give a number of useful preliminary results, some of which have appeared previously in [1], [9], or [14] in the same, more restricted, or more general form. They are included here in the present notation for sake of completeness and later use.

The first equality is a direct consequence of (2.8) and the fact that $\sum_{d=1}^n t_d(x)$

$= 1$ a.e.μ., and is given by

$$(3.4)\qquad (\xi, \rho(t) - \rho(t')) = \int \sum_{d<d'} (\xi, L^{dd'}f(x))\{t_d(x)t'_{d'}(x) - t_{d'}(x)t_d'(x)\}d\mu(x).$$

The Bayes envelope functional satisfies the following properties:

LEMMA 3.1. *If* ξ, $\xi' \varepsilon E^m$, *then*

$$(3.5)\qquad 0 \leqq (\xi, \rho(t\{\xi'\})) - \phi(\xi) \leqq (\xi - \xi', \rho(t\{\xi'\}) - \rho(t\{\xi\})),$$

$$(3.6)\qquad (\xi' - \xi, \rho(t\{\xi'\})) \leqq \phi(\xi') - \phi(\xi) \leqq (\xi' - \xi, \rho(t\{\xi\}))$$

$$(3.7)\qquad |\phi(\xi') - \phi(\xi)| \leqq \|\xi - \xi'\|B, \text{ where } B = \|\bar{L}\| \text{ with } \bar{L} = (\bar{L}_1, \cdots, \bar{L}_m), \bar{L}_\theta = \max_d L(\theta, d).$$

PROOF. See Corollary 1 of [14] for (3.5). The right-hand side of (3.6) follows from the left-hand side of (3.5), which implies $\phi(\xi') \leqq (\xi', \rho(t\{\xi\}))$. The left-hand side of (3.6) follows by interchange of ξ' and ξ in the right-hand side of (3.6). Inequality (3.6), the Schwarz inequality and noting that range $\rho_\theta \leqq \bar{L}_\theta$ yield (3.7).

LEMMA 3.2. *Let* $\xi_1, \cdots, \xi_N \varepsilon E^m$, $\bar{\xi}_i = i^{-1} \sum_{j=1}^i \xi_j$, *and* $\bar{\xi}_0 = \mathbf{0}$, *the zero vector in* E^m. *Then*

$$(3.8)\qquad N^{-1} \sum_{i=1}^N (\xi_i, \rho(t\{\bar{\xi}_i\})) \leqq \phi(\bar{\xi}_N) \leqq N^{-1} \sum_{i=1}^N (\xi_i, \rho(t\{\bar{\xi}_{i-1}\})).$$

PROOF. Let $\zeta_i = \sum_{j=1}^i \xi_j$ and define $\zeta_0 = \bar{\xi}_0 = \mathbf{0}$. Next observe that definition (2.9) implies $t\{\xi\}$ is invariant under scale change, that is, $t\{a\xi\} = t\{\xi\}$ for $a > 0$. Hence $\rho(t\{\bar{\xi}_i\}) = \rho(t\{\zeta_i\})$ for $i = 0, 1, \cdots, N$, and we have

$$\begin{aligned}\sum_{i=1}^N (\xi_i, \rho(t\{\bar{\xi}_{i-1}\})) &= \sum_{i=1}^N (\zeta_i - \zeta_{i-1}, \rho(t\{\zeta_{i-1}\})) \\ &= \sum_{i=1}^N (\zeta_i, \rho(t\{\zeta_{i-1}\}) - \rho(t\{\zeta_i\})) + N\phi(\bar{\xi}_N).\end{aligned}$$

The right-hand side of (3.8) now follows from the left-hand side of (3.5). Similarly,

$$\begin{aligned}\sum_{i=1}^N (\xi_i, \rho(t\{\bar{\xi}_i\})) &= \sum_{i=1}^N (\zeta_i - \zeta_{i-1}, \rho(t\{\zeta_i\})) \\ &= \sum_{i=1}^{N-1} (\zeta_i, \rho(t\{\zeta_i\}) - \rho(t\{\zeta_{i+1}\})) + N\phi(\bar{\xi}_N) \\ &\leqq N\phi(\bar{\xi}_N).\end{aligned}$$

Before stating and proving our theorems we point out that the results here given and their proofs were inspired by the analogous results in the fixed sample size compound decision problem where for fixed N the N decisions to be made can be held in abeyance until all N random variables $\mathbf{X}_N = (X_1, \cdots, X_N)$ are known. See [3], [4], and [6] for the case $m = n = 2$ and [14] for the more general m and n. Familiarity with these papers would be extremely helpful in what follows. We point out the specific relationships between the present paper and [2], [3], [4], [6], [9], and [14] as the results are given.

4. Main theorems. We shall now state certain sufficient conditions used to obtain upper bounds in (2.12) of order $N^{-\frac{1}{2}}$ for the sequential procedure $\mathbf{t}^*$ in Theorems 4.1 and 4.2 and for $\mathbf{t}^{**}$ in Theorem 4.3. Lower bounds for the regret function of equivalent orders are given in Theorem 4.1.1 and Theorem 4.3.1.

Let $L_3(\mu)$ be the space of measurable functions g on $\mathfrak{X}$ such that $\int |g|^3 \, d\mu < \infty$. The following conditions will be pertinent to what follows.

(A_2) The function $h = (h_1, \cdots, h_m)$ is such that h_j is in $L_3(\mu)$ for $j = 1, \cdots, m$.

(A_3) The columns of the loss matrix $(L(\theta, d))$ are mutually non-dominated; that is, there is no pair of columns L^d and $L^{d'}$ such that $L(\theta, d) \geqq L(\theta, d')$ for $\theta = 1, \cdots, m$.

(A_4) The function $h = (h_1, \cdots, h_m)$ is such that $(\mathbf{1}, h) = 1$ a.e.μ (where $\mathbf{1}$ is the vector of 1's in E^m) and V_θ, the covariance matrix of h under P_θ, is of rank $m - 1$ for $\theta = 1, \cdots, m$.

(A_5) The function $h = (h_1, \cdots, h_m)$ is such that V_θ, the covariance matrix of h under P_θ, is of rank m for $\theta = 1, \cdots, m$.

We now prove the following three results.

THEOREM 4.1. *If $h \in \mathfrak{H}$ and (A_1), (A_2), (A_3), and (A_4) hold, then there exists a positive constant c_1 independent of $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ such that $R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(p_N(\boldsymbol{\theta})) \leqq c_1 N^{-\frac{1}{2}}$.*

PROOF. By independence of $h(X_k)$ and $\rho(t^*\{\bar{h}_{k-1}\})$ and by unbiasedness of $h(X_k)$, we have $\mathbf{E}(\epsilon_{\theta_k}, \rho(t^*\{\bar{h}_{k-1}\})) = \mathbf{E}(h(X_k), \rho(t^*\{\bar{h}_{k-1}\}))$ for $k = 1, \cdots, N$. Then, since $R_N(\boldsymbol{\theta}, \mathbf{t}^*) = N^{-1} \sum_{k=1}^N \mathbf{E}(\epsilon_{\theta_k}, \rho(t^*\{\bar{h}_{k-1}\}))$, we have

$$R_N(\boldsymbol{\theta}, \mathbf{t}^*) = N^{-1} \sum_{k=1}^N \mathbf{E}(h(X_k), \rho(t^*\{\bar{h}_{k-1}\}) - \rho(t^*\{\bar{h}_k\})) + N^{-1} \sum_{k=1}^N \mathbf{E}(h(X_k), \rho(t^*\{\bar{h}_k\})). \tag{4.1}$$

Apply equality (3.4) to the first term and Lemma 3.2 with $\xi_k = h(X_k)$ to the second term in (4.1) to obtain

$$R_N(\boldsymbol{\theta}, \mathbf{t}^*) \leqq A_N(\boldsymbol{\theta}) + \mathbf{E}\phi(\bar{h}_N), \tag{4.2}$$

where

$$A_N(\boldsymbol{\theta}) = N^{-1} \sum_{k=1}^N \sum_{d<d'} \mathbf{E} \int |(h(X_k), L^{dd'}f(x))| \{t_d^*(\bar{h}_{k-1}, x) t_{d'}^*(\bar{h}_k, x) + t_d^*(\bar{h}_k, x) t_{d'}^*(\bar{h}_{k-1}, x)\} \, d\mu(x).$$

Observe that inequality (3.6) and unbiasedness of $\bar{h}_N$ imply

$$\mathbf{E}\{\phi(\bar{h}_N) - \phi(p_N)\} \leqq 0. \tag{4.3}$$

Hence, to complete the proof of the theorem we need only show in (4.2) that

$$A_N(\boldsymbol{\theta}) \leqq c_1 N^{-\frac{1}{2}} \quad \text{uniformly in} \quad \boldsymbol{\theta} \in \boldsymbol{\Omega}. \tag{4.4}$$

To do this fix $k, d, d' (d < d')$, $k \geqq 2$, x_k and x and let $u = L^{dd'}f(x)$. Observe that with the bracket notation for indicator functions we have

$$\begin{aligned} t_d^*(\bar{h}_{k-1}, x) t_{d'}^*(\bar{h}_k, x) + t_d^*(\bar{h}_k, x) t_{d'}^*(\bar{h}_{k-1}, x) \\ \leqq [-(h(x_k), u) < \textstyle\sum_{\nu=1}^{k-1}(h(X_\nu), u) \leqq 0] \\ + [0 < \textstyle\sum_{\nu=1}^{k-1}(h(X_\nu), u) \leqq -(h(x_k), u)]. \end{aligned} \tag{4.5}$$

If $u = \alpha\mathbf{1}$, α a constant, we see that since under (A_4) $(h, \mathbf{1}) = 1$, (4.5) implies the left-hand side is zero. Therefore, assume $u \neq \alpha\mathbf{1}$. The right-hand side of (4.5) is zero or one and it can be one only if $(\sum_{\nu=1}^{k-1} h(X_\nu), u)$ falls into an interval of length $|(h(x_k), u)|$. Applying the Berry-Esseen normal approximation theorem in the form of Lemma 2.1 to the sum $\sum_{\nu=1}^{k-1} (h(X_\nu) - \epsilon_{\theta_\nu}, u)$ with $s_{k-1}^2 = \sum_{\nu=1}^{k-1} uV_{\theta_\nu}u'$ and β as the Berry-Esseen constant, we have

$$(4.6)\quad \mathbf{E}_{k-1}\{t_d^*(\bar{h}_{k-1}, x))t_{d'}^*(\bar{h}_k, x) + t_d^*(\bar{h}_k, x)t_{d'}^*(\bar{h}_{k-1}, x)\} \leqq s_{k-1}^{-1}(2\pi)^{-\frac{1}{2}}|(h(x_k), u)| + 2s_{k-1}^{-3}\beta \sum_{\nu=1}^{k-1} E_{\theta_\nu}|(h(X_\nu) - \epsilon_{\theta_\nu}, u)|^3.$$

Let $\bar{u} = m^{-1} \sum_{j=1}^{m} u_j$. Note that under ($A_4$) the null space of the matrix V_θ is the subspace spanned by the vector $\mathbf{1}$. Hence,

$$\begin{aligned} uV_\theta u' &= (u - \bar{u}\mathbf{1})V_\theta(u - \bar{u}\mathbf{1})' \\ &\geqq \|u - \bar{u}\mathbf{1}\|^2 \min\{yV_\theta y' \mid \|y\| = 1, (y, \mathbf{1}) = 0\} \\ &= \|u - \bar{u}\mathbf{1}\|^2\lambda_\theta^2, \end{aligned}$$

where λ_θ^2 is the minimum positive eigenvalue of V_θ. Therefore, we have

$$(4.7)\qquad s_{k-1}^2 = \sum_{\nu=1}^{k-1} uV_{\theta_\nu}u' \geqq \|u - \bar{u}\mathbf{1}\|^2\lambda^2(k-1),$$

where $\lambda^2 = \min_\theta \lambda_\theta^2 > 0$. This inequality together with the inequalities $|(h(x_k), u)| \leqq \|h(x_k)\| \|u - \bar{u}\mathbf{1}\| + |\bar{u}|$, $\sum_{\nu=1}^{k-1} E_{\theta_\nu}|(h(X_\nu) - \epsilon_{\theta_\nu}, u)|^3 \leqq \|u - \bar{u}\mathbf{1}\|^3(k-1) \max_\theta E_\theta\|h(X) - \epsilon_\theta\|^3$, and $\|u - \bar{u}\mathbf{1}\|^{-1} |\bar{u}| \leqq 1$ (which follows from assumption (A_3) and the fact that $\sum_{i=1}^{m} (u_i - \bar{u})^2 \geqq \bar{u}^2$ if not all the u_i are the same sign), yields by (4.6) for $k \geqq 2$,

$$(4.8)\quad \mathbf{E}_{k-1}\{t_d^*(\bar{h}_{k-1}, x)t_{d'}^*(\bar{h}_k, x) + t_d^*(\bar{h}_k, x)t_{d'}^*(\bar{h}_{k-1}, x)\} \leqq (k-1)^{-\frac{1}{2}}\{\alpha_1(\|h(x_k)\| + 1) + \alpha_2\},$$

where $\alpha_1 = (2\pi\lambda^2)^{-\frac{1}{2}}$ and $\alpha_2 = 2\beta\lambda^{-3} \max_\theta E_\theta\|h - \epsilon_\theta\|^3$, with $\alpha_2 < \infty$ since $h_j \,\varepsilon\, L_3(\mu)$, $j = 1, \cdots, m$, implies $\|h\| \,\varepsilon\, L_3(\mu)$. The bound in (4.8) together with $\int|(h(x_k), L^{dd'}f(x)) \mid d\mu(x) \leqq \|h(x_k)\|m^{\frac{1}{2}}L$, implies for $k \geqq 2$, $d < d'$,

$$(4.9)\quad \mathbf{E}\int|(h(X_k), L^{dd'}f(x))|\{t_d^*(\bar{h}_{k-1}, x)t_{d'}^*(\bar{h}_k, x) + t_d^*(\bar{h}_k, x)t_{d'}^*(\bar{h}_{k-1}, x)\}d\mu(x) \leqq (k-1)^{-\frac{1}{2}}\alpha_3,$$

where $\alpha_3 = m^{\frac{1}{2}}L \max_\theta \{\alpha_1 E_\theta\|h\|^2 + (\alpha_1 + \alpha_2)E_\theta\|h\|\}$. Taking the bound in (4.9) to be $m^{\frac{1}{2}}L \max_\theta E_\theta\|h\| \leqq \alpha_3$ (since $\alpha_2 > 1$) if $k = 1$, and summing (4.9) over all k, d, d', $d < d'$, yields

$$A_N(\boldsymbol{\theta}) < N^{-1}\binom{n}{2}\{1 + \sum_{k=2}^{N}(k-1)^{-\frac{1}{2}}\}\alpha_3.$$

Observe that for $N \geqq 2$, we have

$$\begin{aligned} (4.10)\qquad \sum_{k=2}^{N} (k-1)^{-\frac{1}{2}} &= 1 + \sum_{k=2}^{N-1} \int_{k-1}^{k} k^{-\frac{1}{2}}\, dx \\ &\leqq 1 + \int_1^{N-1} x^{-\frac{1}{2}}dx = 2(N-1)^{\frac{1}{2}} - 1. \end{aligned}$$

Hence, (4.4) is satisfied with $c_1 = 2\alpha_3\binom{n}{2}$, and the proof is completed.

THEOREM 4.1.1. *If $h \varepsilon \mathcal{H}$ and* (A_1) *holds, then there exists $c_1' > 0$ independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that* $R_N(\boldsymbol{\theta}, \mathbf{t}) - \phi(p_N(\boldsymbol{\theta})) \geqq -c_1' N^{-\frac{1}{2}}$.

PROOF. As in the proof of the above theorem $R_N(\boldsymbol{\theta}, \mathbf{t}^*) = N^{-1}\mathbf{E} \sum_{k=1}^{N} (h(X_k), \rho(t^*\{\bar{h}_{k-1}\}))$. Hence by Lemma 3.2 with $\xi_k = h(X_k)$ followed by use of inequalities (3.7) and (3.3) we have

$$(4.11) \qquad R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(p_N(\boldsymbol{\theta})) \geqq \mathbf{E}\{\phi(\bar{h}_N) - \phi(p_N)\} \geqq -\sigma B N^{-\frac{1}{2}}.$$

Theorem 4.1.1 follows with $\sigma B = c_1'$.

THEOREM 4.2. *If $h \varepsilon \mathcal{H}$ and* (A_1), (A_2) *and* (A_5) *hold, then there exists a positive constant c_2 independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that* $R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(p_N(\boldsymbol{\theta})) \leqq c_2 N^{-\frac{1}{2}}$.

PROOF. The proof follows exactly as in Theorem 4.1 except we need only assume $u \neq \mathbf{0}$ ($u = \mathbf{0}$ implies that the left-hand side of (4.5) is zero) in developing (4.6), since by (A_5) the null space of V_θ is the zero vector for $\theta = 1, \cdots, m$. Hence, (4.7) is replaced by $s_{k-1}^2 \geqq \|u\|^2 \lambda^2 (k-1)$, which when carried through the remainder of the proof yields the same result as in Theorem (4.1) with $\alpha_3^* = m^{\frac{1}{2}} L \max_\theta \{\alpha_1 E_\theta \|h\|^2 + \alpha_2 E_\theta \|h\|\}$ replacing α_3 in (4.9).

From the proofs of Theorems (4.1) and (4,2), it is obvious that $c_2 < c_1$.

Observe that Thorem 4.1.1 may be used in conjunction with either Theorem 4.1 or Theorem 4.2 to form a uniform (in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$) convergence theorem for the absolute value of the regret function of $\mathbf{t}^*$.

One can delete the need for assumptions (A_3) and (A_4) (or (A_5)) by using the procedure $\mathbf{t}^{**}$ instead of $\mathbf{t}^*$. This result is given by the following theorem.

THEOREM 4.3. *If $h \varepsilon \mathcal{H}$ and* (A_1) *and* (A_2) *hold, then there exists a positive constant c_3 independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that* $R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) - \phi(p_N(\boldsymbol{\theta})) \leqq c_3 N^{-\frac{1}{2}}$.

PROOF. By a change of variable we may express for each k, $\mathbf{E}(L_{\theta_k}, t^*(\bar{h}_k, X_k)) = \mathbf{E}E_{\theta_k}(L_{\theta_k}, t^*(\bar{h}^{(k)}, X))$, where $\bar{h}^{(k)} = k^{-1}\{\sum_{\nu=1}^{k-1} h(X_\nu) + h(X)\}$ and $\mathbf{E}E_{\theta_k}$ is an interated integral with E_{θ_k} on X. Subtracting and adding $\mathbf{E}E_{\theta_k}(L_{\theta_k}, t^*(\bar{h}_k, X)) = \mathbf{E}(\epsilon_{\theta_k}, \rho(t^*\{\bar{h}_k\}))$ from the above equality and averaging the result on k yields

$$(4.12) \qquad \begin{aligned} R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) &= N^{-1} \textstyle\sum_{k=1}^{N} \mathbf{E}(L_{\theta_k}, t^*(\bar{h}_k, X_k)) \\ &= B_N(\boldsymbol{\theta}) + B_N'(\boldsymbol{\theta}), \end{aligned}$$

where $B_N(\boldsymbol{\theta}) = N^{-1} \sum_{k=1}^{N} \mathbf{E}E_{\theta_k}(L_{\theta_k}, t^*(\bar{h}^{(k)}, X) - t^*(\bar{h}_k, X))$ and $B_N'(\boldsymbol{\theta}) = N^{-1} \sum_{k=1}^{N} \mathbf{E}(\epsilon_{\theta_k}, \rho(t^*\{\bar{h}_k\}))$. We show: (i) $B_N(\boldsymbol{\theta}) \leqq b^* N^{-\frac{1}{2}}$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ and (ii) $B_N'(\boldsymbol{\theta}) - \phi(p_N) \leqq b' N^{-\frac{1}{2}}$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.

(i) Observe that under P_θ, $h(X) - \epsilon_\theta = (h_1(X), \cdots, h_\theta(X) - 1, \cdots, h_m(X))$ is an m-dimensional random variable with mean zero and covariance matrix V_θ of rank r_θ. Hence, if under P_θ, $r_\theta > 0$, then there exists an $m \times r_\theta$ matrix W_θ with transpose W_θ' such that $W_\theta W_\theta' = V_\theta$ and $W_\theta Z_\theta'(X) = (h(X) - \epsilon_\theta)'$, where $Z_\theta(X)$ is an r_θ-dimensional random variable with mean zero and identity covariance matrix. Therefore, if X is distributed as P_θ and g is an r_θ-vector with $\|g\| = 1$, then

$$(4.13) \qquad E_\theta(Z_\theta(X), g)^2 = \|g\|^2 = 1$$

and

$$E_\theta \|Z_\theta(X)\|^2 = r_\theta . \tag{4.14}$$

Now express the difference of each term in $B_N(\boldsymbol{\theta})$ as

$$\begin{aligned} &\mathbf{E}E_{\theta_k}(L_{\theta_k}, t^*(\bar{h}^{(k)}, X) - t^*(\bar{h}_k, X)) \\ &\quad = \textstyle\sum_{d<d'} \mathbf{E}E_{\theta_k} L_{\theta_k}^{dd'}\{t_d^*(\bar{h}^{(k)}, X)t_{d'}^*(\bar{h}_k, X) - t_{d'}^*(\bar{h}^{(k)}, X)t_d^*(\bar{h}_k, X)\}. \end{aligned} \tag{4.15}$$

Fix θ, d, $d'(d < d')$ and $k \varepsilon I_\theta = \{k \mid \theta_k = \theta\}$. Using our bracket notation for characteristic functions and considering when $t_d^* t_{d'}^* = 1$, we obtain

$$\begin{aligned} &t_d^*(\bar{h}^{(k)}, X)t_{d'}^*(\bar{h}_k, X) + t_{d'}^*(\bar{h}^{(k)}, X)t_d^*(\bar{h}_k, X) \\ &\leqq [-(h(X_k), L^{dd'}f(X)) < \textstyle\sum_{\nu=1}^{k-1} (h(X_\nu), L^{dd'}f(X)) \\ &\leqq -(h(X), L^{dd'}f(X))] + [-(h(X), L^{dd'}f(X)) \\ &< \textstyle\sum_{\nu=1}^{k-1} (h(X_\nu), L^{dd'}f(X)) \leqq -(h(X_k), L^{dd'}f(X))]. \end{aligned} \tag{4.16}$$

If h is degenerate under P_θ, then the $\mathbf{E} \times E_{\theta_k}$ integral of the right-hand side of (4.16) is zero since $\theta_k = \theta$. If h is non-degenerate under $P_\theta(r_\theta > 0)$, but $L^{dd'}f(X)W_\theta = \mathbf{0}$ for fixed $X = x$, then the right-hand side of (4.16) is zero at such x because $(h(x'), L^{dd'}f(x)) = (Z_\theta(x'), L^{dd'}f(x)W_\theta) + (\epsilon_\theta, L^{dd'}f(x)) = (\epsilon_\theta, L^{dd'}f(x)) = (\epsilon_\theta, L^{dd'}f(x))$ for both $x' = x$ and X_k. Omitting these degenerate cases we shall bound the right-hand side of (4.16) using Lemma 2.1.

Specifically, assume $r_\theta > 0$, $p_{k-1,\theta}(\boldsymbol{\theta}) > 0$ and fix $X = x$ such that $L^{dd'}f(x)W_\theta \neq \mathbf{0}$. Define $g(x) = \|L^{dd'}f(x)W_\theta\|^{-1}L^{dd'}f(x)W_\theta$, noting that $\|g(x)\| = 1$. Next fix $X_k = x_k$, $X_\nu = x_\nu$, $\nu \not\varepsilon I_\theta$, $\nu < k$. We observe that for the right-hand side of (4.16) not to vanish, the sum

$$\|L^{dd'}f(x)W_\theta\|^{-1} \textstyle\sum_{\nu<k,\nu\varepsilon I_\theta}(h(X_\nu) - \epsilon_\theta, L^{dd'}f(x)) = \textstyle\sum_{\nu<k,\nu\varepsilon I_\theta}(Z_\theta(X_\nu), g(x))$$

of $(k - 1)p_{k-1,\theta}(\boldsymbol{\theta}) \geqq 1$ terms must fall into an interval of length $|(Z_\theta(x) - Z_\theta(x_k), g)| \leqq \|Z_\theta(x) - Z_\theta(x_k)\|$. But the terms $(Z_\theta(X_\nu), g(x))$ are independent and identically distributed with the mean zero and variance 1 (by (4.13)). Hence, the Berry-Esseen result of Lemma 2.1 bounds the probability of the event on the right-hand side of (4.16) by

$$\{(k - 1)p_{k-1,\theta}(\boldsymbol{\theta})\}^{-\frac{1}{2}}\{(2\pi)^{-\frac{1}{2}}\|Z_\theta(x_k) - Z_\theta(x)\| + 2\beta E_\theta|(Z_\theta(X), g(x))|^3, \tag{4.17}$$

where the expectation is on X in the second term. Integrating with respect to $E_{\theta_k} \times E_\theta$ in (4.17) and weakening by $E_{\theta_k}E_\theta\|Z_\theta(X_k) - Z_\theta(X)\| \leqq 2r_\theta^{\frac{1}{2}}$ (implied by (4.14)) and by $E_\theta|(Z_\theta(X), g(x))|^3 \leqq E_\theta\|Z_\theta(X)\|^3$, we have from (4.16) and (4.17) that

$$\begin{aligned} &\mathbf{E}E_{\theta_k}\{t_d^*(\bar{h}^{(k)}, X)t_{d'}^*(\bar{h}_k, X) + t_{d'}^*(\bar{h}^{(k)}, X)t_d^*(\bar{h}_k, X)\} \\ &\qquad\qquad \leqq \min\{1, \{(k - 1)p_{k-1,\theta}(\boldsymbol{\theta})\}^{-\frac{1}{2}}q_\theta\}, \end{aligned} \tag{4.18}$$

where we take $q_\theta = \max\{1, (2\pi^{-1}r_\theta)^{\frac{1}{2}} + 2\beta E_\theta\|Z_\theta(x)\|^3\}$ to include the cases $r_\theta = 0$

and $(k-1)p_{k-1,\theta}(\boldsymbol{\theta}) = 0$ with the convention that $1/0 = \infty$. Note that $q_\theta < \infty$ since, $h_j \varepsilon L_3(\mu)$ for $j = 1, \cdots, m$ implies $\|Z_\theta(x)\| \varepsilon L_3(\mu)$. Multiplying both sides of (4.16) by $L = \max_{\theta,d,d'} |L_\theta^{dd'}|$ and noting that the left-hand side is then an upper bound for the d, d' term of (4.15), sum over all θ, d, d', k, $k \varepsilon I_\theta$, $d < d'$ to obtain from (4.15) and (4.18),

$$(4.19)\qquad B_N(\boldsymbol{\theta}) \leqq L\tbinom{n}{2}N^{-1} \textstyle\sum_{\theta=1}^m \sum_{j=1}^{Np_{N\theta}(\boldsymbol{\theta})} \min\{1, q_0(j-1)^{-\frac{1}{2}}\},$$

where $q_0 = \max_\theta q_\theta$. Let $[N/m]^*$ be the integral part of N/m. Observe that the double sum in (4.19) is bounded by $m \sum_{j=1}^{[N/m]^*+1} \min\{1, q_0(j-1)^{-\frac{1}{2}}\} \leqq m(1 + q_0 \sum_{j=2}^{[N/m]^*+1} (j-1)^{-\frac{1}{2}}) < m(1 + 2q_0\{[N/m]^*\}^{\frac{1}{2}})$, where the last inequality follows from (4.10) with $[N/m]^* + 1$ replacing N. Finally, since $[N/m]^* \leqq N/m$ we have

$$B_N(\boldsymbol{\theta}) = \tbinom{n}{2}L\{mN^{-1} + 2q_0m^{\frac{1}{2}}N^{-\frac{1}{2}}\}$$

from whence (i) follows.

(ii) Express $B_N'(\boldsymbol{\theta}) = N^{-1} \sum_{k=1}^N \mathbf{E}(\epsilon_{\theta_k} - h(X_k), \rho(t^*\{\bar{h}_k\})) + N^{-1} \sum_{k=1}^N \mathbf{E}(h(X_k), \rho(t^*\{\bar{h}_k\}))$. Note that by independence of $h(X_k)$ and $\rho(t^*\{\bar{h}_{k-1}\})$ and unbiasedness of $h(X_k)$, the first term is equal to

$$(4.20)\quad B_N''(\boldsymbol{\theta}) = N^{-1} \textstyle\sum_{k=1}^N \mathbf{E}(\epsilon_{\theta_k} - h(X_k), \rho(t^*\{\bar{h}_k\}) - \rho(t^*\{\bar{h}_{k-1}\})),$$

while Lemma 3.2 implies that the second term is bounded from above by $\mathbf{E}\phi(\bar{h}_N)$. Hence, by inequality (4.3) we have

$$(4.21)\qquad B_N'(\boldsymbol{\theta}) - \phi(p_N(\boldsymbol{\theta})) \leqq B_N''(\boldsymbol{\theta}) + \mathbf{E}\{\phi(\bar{h}_N) - \phi(p_N(\boldsymbol{\theta}))\} \leqq B_N''(\boldsymbol{\theta}),$$

with $B_N''(\boldsymbol{\theta})$ given by (4.20).

Now to bound $B_N''(\boldsymbol{\theta})$ use (3.4) to write

$$(4.22)\quad B_N''(\boldsymbol{\theta}) \leqq N^{-1} \textstyle\sum_{k=1}^N \sum_{d<d'} \mathbf{E} \int |(h(X_k) - \epsilon_{\theta_k}, L^{dd'}f(x))| \\ \{t_d^*(\bar{h}_{k-1}, x)t_{d'}^*(\bar{h}_k, x) + t_d^*(\bar{h}_k, x)t_{d'}^*(\bar{h}_{k-1}, x)\}d\mu(x).$$

Fix θ, d, d', k, $d < d'$, $k \varepsilon I_\theta$, $X_k = x$, $X = x$ and let $u = L^{dd'}f(x)$ as in the proof of Theorem 4.1. Using the transformation $(h - \epsilon_\theta)' = W_\theta Z_\theta'$ as in (i) and noting that the coefficient $|(h(x_k) - \epsilon_{\theta_k}, u)| = |(Z_\theta(x_k), uW_\theta|$, we see that only the terms for which $uW_\theta \neq 0$ and $r_\theta > 0$ contribute to the bound. However, if $uW_\theta \neq 0$ and $r_\theta > 0$, we can apply a Berry-Esseen argument from Lemma 2.1 for fixed X_ν, $\nu \varepsilon I_\theta$, $\nu < k$ in the bound (4.5) to the sum $\|uW_\theta\|^{-1} \sum_{\nu\varepsilon I_\theta,\nu<k} (h(X_\nu) - \epsilon_\theta, u) = \sum_{\nu\varepsilon I_\theta,\nu<k} (Z_\theta(X_\nu), g)$ (with $g = \|uW_\theta\|^{-1}uW_\theta$) which falls into an interval of length $|(Z_\theta(x_k), g)|$, and we obtain by arguments similar to those used in (i) and Theorem 4.1,

$$(4.23)\qquad \begin{aligned} &\mathbf{E} \int |(h(X_k) - \epsilon_{\theta_k}, L^{dd'}f(x)| \\ &\{t_d^*(\bar{h}_{k-1}, x)t_{d'}^*(\bar{h}_k, x) + t_d^*(\bar{h}_k, x)t_{d'}^*(\bar{h}_{k-1}, x)\}d\mu(x) \\ &\qquad \leqq \sigma m^{\frac{1}{2}}\mathrm{L} \min\{1, \{(k-1)p_{k-1,\theta}\}^{-\frac{1}{2}}q_\theta^*\}, \end{aligned}$$

where

$$q_\theta^* = \max\{1, ((2\pi)^{-1}r_\theta)^{\frac{1}{2}} + 2\beta E_\theta\|Z_\theta(X)\|^3\} \text{ and } 1/0 = \infty.$$

The remainder of the proof follows as in (i) with q_θ^* of (4.23) playing a role similar to that of q_θ in (4.18).

THEOREM 4.3.1. *If* $h \in \mathcal{H}$ *and if* (A_1), (A_2), *and* (A_5) (*or* (A_3) *and* (A_4)) *hold, then there exists a* $c_3' > 0$ *independent of* $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ *such that* $R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) - \phi(p_N(\boldsymbol{\theta})) \geqq -c_3'N^{-\frac{1}{2}}$.

PROOF. The bound on $B_N(\boldsymbol{\theta})$ in (i) holds for $|B_N(\boldsymbol{\theta})|$ as well, while by independence of $h(X_k)$ and $\rho(t^*\{\bar{h}_{k-1}\})$ and unbiasedness of $h(X_k)$ we may write

$$B_N'(\boldsymbol{\theta}) = B_N^*(\boldsymbol{\theta}) + N^{-1} \textstyle\sum_{k=1}^N \mathbf{E}(h(X_k), \rho(t^*\{\bar{h}_{k-1}\})), \tag{4.24}$$

where $B_N^*(\boldsymbol{\theta}) = N^{-1} \sum_{k=1}^N \mathbf{E}(\epsilon_{\theta_k}, \rho(t^*\{\bar{h}_k\}) - \rho(t^*\{\bar{h}_{k-1}\}))$. By Lemma 3.2 the second term in (4.24) is bounded from below by $\mathbf{E}\phi(\bar{h}_N)$. Hence, Equations (3.7), (3.3), (4.12), (4.24), and part (i) of Theorem 4.3 yield

$$R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) - \phi(p_N) \geqq -(\sigma B + b^*)N^{-\frac{1}{2}} + B_N^*(\boldsymbol{\theta}). \tag{4.25}$$

But $-B_N^*(\boldsymbol{\theta})$ is the same as the first term on the right-hand side of (4.1) with $h(X_k)$ replaced by ϵ_{θ_k}. Hence, lower bounds for $B_N^*(\boldsymbol{\theta})$ can be obtained under assumptions (A_3) and (A_4) (or (A_5)) by arguments similar to those used for the upper bounds on the term $A_N(\boldsymbol{\theta})$ in Theorem 4.1 (or Theorem 4.2).

If one compares the statements and proofs of Theorems 4.1, 4.2, and 4.3 with Theorem 2 of Van Ryzin [14], he sees that these theorems are results which are completely analogous to the results obtained therein for the fixed sample size (where all N decisions are held in abeyance until all N random variables have been observed) compound decision problem with $m \times n$ loss matrix. Note that the bound is of the same order ($O(N^{-\frac{1}{2}})$) uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Omega}$ and hence with regard to regret convergence the rate is as good for the sequential compound decision problem as in the fixed sample size case.

5. Sequential compound testing for two specified distributions. We now specialize the problem to the situation where $m = n = 2$, $L(1, 1) = L(2, 2) = 0$, $L(2, 1) = a > 0$, and $L(1, 2) = b > 0$. This situation is called "compound testing between two completely specified distributions" because each component problem consists of testing a simple hypothesis $H_1 : \theta = 1$ against a simple alternative $H_2 : \theta = 2$ incurring no loss for correct decision and a loss $a > 0$ or $b > 0$ according as we commit a type II or type I error. For a treatment of this problem in the fixed sample size case, see [3], [4], [6], and [12]. The sequential case, which we consider here, has been treated in this testing situation by Samuel in [9] and [10]. We will point out how reductions of our results (Theorems 4.1 and 4.3 of previous section) to the sequential compound testing case furnish improvements of Samuel's results in [9].

A convenient simplification of notation is possible in this testing situation. Let P_1 and P_2 be the two distributions in question and assume $P_1 \neq P_2$ (assumption (A_1) when $m = 2$). In the notation of Section 2 it is convenient to make the following identifications: $\xi = (\xi_1, \xi_2) = (1 - \eta, \eta)$ for η real, $(t_1, t_2) = (1 - t, t)$ where $t(x) = \Pr\{\text{deciding } \theta = 2 \mid x\}$. Hence, a simple rule may be specified by the single function t only, rather than the vector function (t_1, t_2). A Bayes rule against prior distribution $(1 - \eta, \eta)$ is then given by $(1 - t(\eta, x), t(\eta, x))$, where $t(\eta, x) = t_2(\xi, x)$ is given by (2.9), that is

$$(5.1) \qquad t(\eta, x) = 1, 0 \text{ or arbitrary according as } Z(x) <, >, \text{or} = \eta,$$

where $Z(x) = \{af_2(x) + bf_1(x)\}^{-1}bf_1(x)$ and we assume without loss of generality that $af_2(x) + bf_1(x) > 0$ for almost all x. The non-randomized version of $t(\eta, x)$ which corresponds to (2.10) in this case becomes

$$(5.2) \qquad t^*(\eta, x) = 1 \quad \text{or} \quad 0 \quad \text{according as} \quad Z(x) < \eta \text{ or } \geqq \eta.$$

The Bayes envelope functional in (2.11) depends now only on η real and is given by

$$(5.3) \quad \psi(\eta) = \phi((1 - \eta, \eta)) = \min_t \{a\eta E_2(1 - t) + b(1 - \eta)E_1 t\}$$
$$= a\eta E_2(1 - t(\eta, X)) + b(1 - \eta)E_1 t(\eta, X).$$

Furthermore, in considering sequential rules for the testing case, it suffices to consider sequences of scalar functions $\mathbf{t}(\mathbf{X}) = \{\mathbf{t}_k(\mathbf{X}_k), k = 1, 2, \cdots\}$, where $\mathbf{t}_k(\mathbf{X}_k) = \Pr\{\text{deciding } \theta_k = 2 \mid \mathbf{X}_k\}$, and $1 - \mathbf{t}_k(\mathbf{X}_k) = \Pr\{\text{deciding } \theta_k = 1 \mid \mathbf{X}_k\}$. Note that if $h(x)$ is any function such that $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$ and if $\bar{\delta}_k = \bar{\delta}^k(\boldsymbol{\theta}) = k^{-1} \sum_{j=1}^k \delta_{\theta_j 2}$, then $(1 - \bar{h}_k, \bar{h}_k)$ with $\bar{h}_k = k^{-1} \sum_{j=1}^k h(X_k)$ is an unbiased estimate of $p_k = (p_{k1}, p_{k2}) = (1 - \bar{\delta}_k, \bar{\delta}_k)$ for all $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ and we may define the sequential procedure analogous to that give in (3.1). In particular, we consider the sequential procedure (now a sequence of scalar functions) $\mathbf{t}^* = \{t^*(\bar{h}_{k-1}, X_k), k = 1, 2, \cdots\}$, where $\bar{h}_0 = 0$ and $t^*(\eta, x)$ is defined by (5.2). This sequential procedure satisfies the following result, which is only a restatement of Theorem 4.1 for the special case of compound testing.

THEOREM 5.1. *If h is such that $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$, $h \varepsilon L_3(\mu)$ and h is nondegenerate under P_1 and P_2, then there exists a constant $c_1 > 0$ independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that $R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \leqq c_1 N^{-\frac{1}{2}}$.*

It is interesting to see precisely how Theorem 5.1 generalizes and strengthens Theorem 2 of Samuel [9] if h is non-degenerate under P_1 and P_2. The sequential rule given in Theorem 2 of [9] (in the present notation) is as follows: Let $h(X)$ be an unbiased estimate of $\theta - 1$ for $\theta = 1, 2$ (that is $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$). Define $p_k(\mathbf{X}_k) = 0, \bar{h}_k, 1$ according as $\bar{h}_k < 0$, $0 \leqq \bar{h}_k \leqq 1$, or $\bar{h}_k > 1$. Then define $g(k, W, \mathbf{X}_k) = \{1 + k^{-\frac{1}{4}}(W_1 + W_2)\}^{-1} \{p_k(\mathbf{X}_k) + k^{-\frac{1}{4}}W_2\}$ with $W = (W_1, W_2)$ a random variable uniformly distributed over the unit square. Then, define $\hat{\mathbf{t}} = \{\hat{\mathbf{t}}_k(\mathbf{X}_k), k = 1, 2, \cdots\}$ where $\hat{\mathbf{t}}_k(\mathbf{X}_k) = t^*(g_{k-1}, X_k)$ with $g_{k-1} = g(k, W, \mathbf{X}_{k-1})$ for $k = 2, 3, \cdots$ and $g_0 = \frac{1}{2}$ and $t^*(\eta, x)$ given by (5.2). Then Theorem 2 of [9] states that if h is bounded, then $\hat{R}_N(\boldsymbol{\theta}, \hat{\mathbf{t}}) - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \leqq c_N$ where $c_N \to 0$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ with the risk $\hat{R}_N$ taken with respect to $E \times E_W$, E_W indicating expectation on W. Thus, if h is non-degenerate under P_1 and P_2 Theorem 5.1 furnishes a threefold improvement of Theorem 2 of [9] by removing the need for the "artificial randomization" introduced through W, while simultaneously strengthening the convergence rate to $N^{-\frac{1}{2}}$ and widening the class of estimates from those with bounded h to those with finite third absolute moment under P_1 and P_2.

The improvement in rate is not new, however, since the result announced by Hannan in [2] states that if $k^{-\frac{1}{2}}$ is used instead of $k^{-\frac{1}{4}}$ in defining $g(k, W, \mathbf{X}_k)$ then the rate $N^{-\frac{1}{2}}$ holds. In fact, Hannan's result as stated in [2] is much more extensive and applies to any $m \times n$ sequential compound decision problem, where n may even be infinite under certain regularity conditions. However, his results rely upon analogous results in his game paper ($p_{k-1}(\boldsymbol{\theta})$ known at stage k) [1], where "artificial randomization" is essential. A main contribution of the present paper is the removal of this "artificial randomization" in the sequential compound testing problem (as well as in the more general $m \times n$ case of the previous section). The only randomization necessary is that provided by $\bar{h}_k$, $k = 1, 2, \cdots$, as shown by Theorems 4.1–4.3, Theorem 5.1 and results yet to follow.

We also state Theorem 4.3 of the previous section for the compound testing problem. Define $\mathbf{t}^{**}$ as the sequence of scalar functions $\{t^*(\bar{h}_k, X_k), k = 1, 2, \cdots\}$, where $t^*(\eta, x)$ is given by (5.2). Then Theorem 4.3 yields

THEOREM 5.2. *If h is such that $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$ and $h \,\varepsilon\, L_3(\mu)$, then there exists a constant $c_3 > 0$ independent of $\boldsymbol{\theta} \,\varepsilon\, \boldsymbol{\Omega}$ such that $R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \leqq c_3 N^{-\frac{1}{2}}$.*

The decision rule $\mathbf{t}^{**}$ was proposed by Samuel in Equation (12) or [9] as the most natural choice of a rule for the strongly sequential compound testing problem, but no results are proved therein. Hence, Theorem 5.2 furnishes the desired result in this direction. In view of Theorems 5.1 and 5.2 the assumption about differentiability of the Bayes envelope $\psi(\eta)(R(\eta)$ in [9]) on [0, 1] imposed in Theorem 1 of [9] is unnecessary to achieve regret convergence of the procedures $\mathbf{t}^*$ and $\mathbf{t}^{**}$.

Finally, we wish to point out that the assumptions of Theorem 5.2 may be *always* satisfied for *any* pair of distributions $P_1, P_2 (P_1 \neq P_2)$ by choosing $h(x)$ as

$$h^*(x) = (c_{11}c_{22} - c_{12}^2)^{-1}\{c_{11}f_2(x) - c_{12}f_1(x)\}, \tag{5.4}$$

where $c_{\theta j} = E_\theta\{f_j(X)\}$ for $\theta, j = 1, 2$. Note that h^* is bounded (and hence in $L_3(\mu)$) since $f_i(x) \leqq K$ by (2.2) and that $E_\theta\{h^*(X)\} = \theta - 1$ for $\theta = 1, 2$. See Theorem 1 of [14] for the generalization of h^* when $m > 2$.

An example will now be given in which the lack of randomization results in *strict* inequalities on both sides of $R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) \leqq \psi(\bar{\delta}_N(\boldsymbol{\theta})) \leqq R_N(\boldsymbol{\theta}, \mathbf{t}^*)$. The randomization in our problem is furnished by the assumption of non-degeneracy of h in Theorem 5.1. The analog of this assumption in Theorem 4.1 (or Theorem 4.2) is (A_4) (or (A_5)) and hence this example motivates imposing of this assumption. See Theorem 4.3.1 in this regard also.

EXAMPLE. Let $a = b = 1$ and let $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots\}$ be such that $\theta_i = 1$ or 2 as i is even or odd. Let P_1 and P_2 be discrete probability distributions on $\{-1, 0, 1\}$ with probabilities given by:

X	-1	0	1
$P_1\{X = x\}$	.1	.5	.4
$P_2\{X = x\}$	.4	.5	.1
$Z(x)$	.2	.5	.8

Let $\bar{h}_k = \bar{\delta}_k$, $\bar{\delta}_0 = 0$, in the definitions of $\mathbf{t}^*$ and $\mathbf{t}^{**}$. $R_N(\boldsymbol{\theta}, \mathbf{t}^*) = 1$ or $.6 + (.8)N^{-1}$ as $N = 1$ or $N \geqq 2$, $R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) = .1 - (.1)N^{-1}$, and $\psi(\bar{\delta}_N(\boldsymbol{\theta})) = 0$, $.35$ or $.35 - (.25)N^{-1}$ as $N = 1$, even, or odd and $\geqq 3$. Thus, the result of Theorem 5.1 (Theorem 4.1 for the testing case) is violated in this example, that is, $R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \psi(\bar{\delta}_N) \rightarrow .25$ as $N \rightarrow \infty$. Also Theorem 4.3.1 is not true for this example since $R_N(\boldsymbol{\theta}, \mathbf{t}^{**}) - \psi(\bar{\delta}_N) \rightarrow -.25$ as $N \rightarrow \infty$.

By this example we see that some sort of randomization here furnished by an assumption like (A_4) (or (A_5)) is necessary in Theorem 4.1 (or Theorem 4.2) and Theorem 4.3.1. Whether assumptions somewhat weaker than these (but stronger than degeneracy of h under P_θ, $\theta = 1, \cdots, m$) can yield these same results has not been established.

This example is the same phenomenon as that in sequential game theory, where in the game of "matching pennies" the strategy for player II of playing Bayes against player I's empirical distribution of prior moves will not guarantee that the average risk over N-repetitions of the game approaches the Bayes envelope functional of the N-stage empirical distribution of player I's move. See [1] and [9] in this respect.

We now present additional theorems of higher convergence rates in the case of sequential compound testing. Specify $\mu = aP_2 + bP_1$ (that is, $af_2(x) + bf_1(x) = 1$ a.e.μ) and let $Z(x) = bf_1(x)$. For $\theta = 1, 2$, let P_θ^* be the probability measure induced under P_θ on the unit interval by the measurable transformation $Z(x)$ from $\mathfrak{X}$ into $[0, 1]$. We shall modify (5.2) as follows:

$$\text{(5.5)} \quad \begin{aligned} t'(\eta, x) &= 1 \quad \text{if} \quad Z(x) < \eta,\ Z(x) \in (0, 1) \text{ or } Z(x) = 0 \\ &= 0 \quad \text{if} \quad Z(x) \geqq \eta,\ Z(x) \in (0, 1) \text{ or } Z(x) = 1, \end{aligned}$$

for η real, $x \in \mathfrak{X}$. The motivation behind (5.5) is that when $\eta = 0$, (5.2) is an inadmissible Bayes procedure in the component problem if there exists an $x \in \mathfrak{X}$ with $P_1\{X = x\} = 0$, $P_2\{X = x\} > 0$. However, (5.5) is always admissible Bayes in the component problem. In view of this remark, (5.5) should also be used in place of (5.2) when defining the rules in Theorems 5.1 and 5.2. It is easy to see that Theorems 5.1 and 5.2 continue to hold if this is done.

The following conditions are pertinent to what follows:

(I) The induced distribution of Z under P_θ is continuous on the open interval $(0, 1)$ for $\theta = 1, 2$.

Observe that (I) implies uniform continuity on $[0, 1]$ (and hence the whole real line) of the non-normed distribution function $H(z) = \int [0 < Z(x) < 1, Z(x) < z](af_2(x) + bf_1(x))d\mu(x)$.

(II) For $\theta = 1, 2$, P_θ^* is absolutely continuous with respect to Lebesgue measure λ on the real line and there exist a $K < \infty$ such that $p_\theta^*(z) = (dP_\theta^*/d\lambda)(z) \leqq K'$.

For examples of pairs of distributions (P_1, P_2) for which assumptions (I) or (II) hold and for a further discussion of these conditions see Sections 5 and 6 of Hannan and Van Ryzin [4].

The following theorems hold. Let $\mathbf{t}' = \{t'(\bar{h}_{k-1}, X_k), k = 1, 2, \cdots\}$ be the strongly sequential rule where $t'(\eta, x)$ is defined by (5.5) with $\bar{h}_0 = 0$. Theorems (5.3) and (5.4) (or Theorems (5.5) and (5.6)) are the analog of Theorems 2 and 3 of Hannan and Van Ryzin [4] in the fixed sample size compound testing problem.

THEOREM 5.3. *Let h be such that $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$, and* $|h(x)| \leqq H < \infty$ a.e.μ. *If* (II) *holds, then there exists a constant c' independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that $R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \leqq c'$* $(\log N)N^{-1}$.

PROOF. Let $W = \{x \mid 0 < Z(x) < 1\}$. By (5.5) and our choice of μ, we have from (4.2),

$$(5.6) \quad A_N(\boldsymbol{\theta}) = N^{-1} \sum_{k=1}^N \mathbf{E} \int_W |h(X_k) - Z(x)|\{[\bar{h}_k \leqq Z(x) < \bar{h}_{k-1}] + [\bar{h}_{k-1} \leqq Z(x) < \bar{h}_k]\}\, d\mu(x)$$
$$\leqq (H + 1)N^{-1}\{\textstyle\sum_{k=2}^N \mathbf{E} \int_W [|Z(x) - \bar{h}_{k-1}| \leqq (2H)k^{-1}]\, d\mu(x) + (a + b)\}'$$

where the inequality follows by boundedness of h implying $|h(X_k) - Z(x)| \leqq H + 1$ and $|\bar{h}_k - \bar{h}_{k-1}| \leqq (2H)k^{-1}$, $k \geqq 2$.

Note that in (5.6), assumption (II) implies that for $k \geqq 2$,

$$\int_W[|Z(x) - \bar{h}_{k-1}| \leqq (2H)k^{-1}]\, d\mu(x)$$
$$\leqq \int [\bar{h}_{k-1} - (2H)k^{-1} \leqq z \leqq \bar{h}_{k-1} + (2H)k^{-1}]\{ap_1^*(z) + bp_2^*(z)\}\, dz$$
$$\leqq \{4(a + b)HK'\}k^{-1}.$$

Hence, (5.6) and the fact that for $N \geqq 2$, $\sum_{k=2}^N k^{-1} < \int_1^N x^{-1}\, dx = \log N$ imply

$$A_N(\boldsymbol{\theta}) \leqq (H + 1)(a + b)\{4HK'(\log N)N^{-1} + N^{-1}\}$$

from whence (4.2), (4.3) and (5.3) complete the proof.

The restriction to boundedness of h in Theorem 5.3 is not very stringent since, as was pointed out earlier, it may always be satisfied by choosing h as h^* in (5.4). This comment pertains to theorems which are to follow also.

THEOREM 5.3.1. *Let h be such that $h \,\varepsilon\, L_2(\mu)$ and $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$. If* (II) *holds, then there exists a positive constant c_0' independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that $R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \geqq -c_0'N^{-1}$.*

PROOF. To obtain the necessary lower inequality note that by (4.11) of Theorem 4.1.1. and inequality (3.6), we have

$$(5.7) \quad R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N) \geqq \mathbf{E}\{\psi(\bar{h}_N) - \psi(\bar{\delta}_N)\}$$
$$\geqq \mathbf{E}(\bar{\delta}_N - \bar{h}_N)\{bE_1t'(\bar{h}_N, X) - aE_2(1 - t'(\bar{h}_N, X))\}.$$

Subtracting from the lower bound in (5.7) the quantity $\mathbf{E}(\bar{\delta}_N - \bar{h}_N)\cdot \{bE_1(t'(\bar{\delta}_N, X)) - aE_2(1 - t'(\bar{\delta}_N, X))\}$ which equals zero by unbiasedness of $\bar{h}_N$, we obtain from (5.7) and the definition of μ,

$$(5.8)\quad R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N) \geqq \mathbf{E}(\bar{\delta}_N - \bar{h}_N) \int \{t'(\bar{h}_N, x) - t'(\bar{\delta}_N, x)\}\, d\mu(x).$$

From the definition (5.5), (5.8) yields

$$(5.9)\quad \begin{aligned} R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N) &\leqq \mathbf{E}(\bar{\delta}_N - \bar{h}_N) \textstyle\int_W \{[\bar{\delta}_N \leqq Z(x) < \bar{h}_N] \\ &\qquad - [\bar{h}_N \leqq Z(x) < \bar{\delta}_N]\}\, d\mu(x) \\ &= -\mathbf{E}|\bar{h}_N - \bar{\delta}_N| \textstyle\int_W \{[\bar{\delta}_N \leqq Z(x) < \bar{h}_N] \\ &\qquad + [\bar{h}_N \leqq Z(x) < \bar{\delta}_N]\}\, d\mu(x). \end{aligned}$$

Assumption (II) implies that

$$\textstyle\int_W \{[\bar{\delta}_N \leqq Z(x) < \bar{h}_N] + [\bar{h}_N \leqq Z(x) < \bar{\delta}_N]\}\, d\mu(x) \leqq |\bar{h}_N - \bar{\delta}_N|(a + b)K'$$

and the result follows from (5.9) with

$$c_0' = (a + b)K' \max\{E_1h^2(X), E_2(h(X) - 1)^2\}$$

THEOREM 5.4. *Let h be non-degenerate under P_θ and such that $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$ and $|h(x)| \leqq H < \infty$ a.e.μ. If* (I) *holds, then for every $\epsilon > 0$ there exists an $N' = N'(\epsilon)$ independent of $\boldsymbol{\theta} \,\varepsilon\, \boldsymbol{\Omega}$ such that $R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \leqq \epsilon N^{-\frac{1}{2}}$ for $N \geqq N'$.*

PROOF. In the kth, $k \geqq 2$, term of the sum on the right-hand side of (5.6), partition the space under the μ integral into $W \cap W_k = \{x \mid |Z(x) - \bar{\delta}_{k-1}| \leqq (k-1)^{-1/8} + (2H)\, k^{-1}\}$ and $W \cap W_k'$ where W_k' is the complement of W_k. Fix $x \,\varepsilon\, W_k$, let $v = \min\{E_1h^2, E_2(h - 1)^2\}$ and note that Lemma 2.1 applied to the normalized sum $s_{k-1}^{-1} \sum_{\nu=1}^{k-1} (h(X_\nu) - \bar{\delta}_{k-1})$, with $s_{k-1}^2 = (k-1)\cdot \{(1 - \bar{\delta}_{k-1})E_1h^2 + \bar{\delta}_{k-1}E_2(h - 1)^2\}$, yields for $k \geqq 2$,

$$(5.10)\quad \begin{aligned} \mathbf{E}[|Z(x) - \bar{h}_{k-1}| \leqq (2H)k^{-1}] &\leqq s_{k-1}^{-1}(4H)(2\pi)^{-\frac{1}{2}} \\ &\qquad + 2\beta s_{k-1}^{-3} \textstyle\sum_{\nu=1}^{k-1} E_{\theta_\nu} |h(X_\nu) - \theta_\nu + 1|^3 \\ &< (k-1)^{-\frac{1}{2}}\, c^*, \end{aligned}$$

where $c^* = v^{-1}\{(4H)(2\pi)^{-\frac{1}{2}} + 2\beta(H + 1)\}$, β the Berry-Esseen constant. Hence, by (5.10) applied for $x \,\varepsilon\, W \cap W_k$ we have for $k \geqq 2$,

$$(5.11)\quad \mathbf{E} \textstyle\int_{W \cap W_k} [|Z(x) - \bar{h}_{k-1}| \leqq (2H)k^{-1}]\, d\mu(x) \leqq (k-1)^{-\frac{1}{2}} c^* \int_{W \cap W_k} d\mu(x).$$

For $x \,\varepsilon\, W \cap W_k'$, we have $[|Z(x) - \bar{h}_{k-1}| \leqq (2H)k^{-1}] \leqq [|\bar{h}_{k-1} - \bar{\delta}_{k-1}| \geqq (k-1)^{-1/8}]$ and hence for such x Tchebichev's inequality implies

$$(5.12)\quad \begin{aligned} \mathbf{E} \textstyle\int_{W \cap W_k'} [|Z(x) - \bar{h}_{k-1}| &\leqq (2H)k^{-1}]\, d\mu(x) \\ &\leqq (a + b)\mathbf{E}[|\bar{h}_{k-1} - \bar{\delta}_{k-1}| \geqq (k-1)^{-1/8}] \\ &\leqq (k-1)^{-3/4}(a + b)V^2, \end{aligned}$$

where $V^2 = \max\{E_1h^2, E_2(h - 1)^2\}$. Now (5.11) and (5.12) together with uniform continuity of $H(z)$, a consequence of (I), furnishing $\int_{W \cap W_k} d\mu(x)$ uniformly (in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$) small for k sufficiently large implies

$$\mathbf{E} \int_W [|Z(x) - \bar{h}_{k-1}| \leqq (2H)k^{-1}] \, d\mu(x) \leqq y_k(k-1)^{-\frac{1}{2}}, \qquad k \geqq 2,$$

where $y_k \to 0$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ as $k \to \infty$. Hence, we have by Toeplitz's lemma (see Loève [5], p. 238) with $a_k = (k-1)^{-\frac{1}{2}}$ and $b_N = \sum_{k=2}^N a_k \to \infty$ that $b_N^{-1} \sum_{k=2}^N a_k y_k \to 0$ independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. Since by (4.10) $b_N < 2N^{\frac{1}{2}}$ we see from (5.6) and the above that

$$\begin{aligned} (5.13) \quad N^{\frac{1}{2}} A_N(\boldsymbol{\theta}) &\leqq N^{-\frac{1}{2}}(H+1)\{\textstyle\sum_{k=2}^N y_k(k-1)^{-\frac{1}{2}} + (a+b)\} \\ &< 2\,(H+1) b_N^{-1} \textstyle\sum_{k=2}^N a_k y_k + N^{-\frac{1}{2}}(H+1)(a+b) \to 0 \end{aligned}$$

uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.

The theorem now follows by (5.13), (4.2), (4.3), and (5.3).

THEOREM 5.4.1. *Let h be such that $h \varepsilon L_2(\mu)$ and $E_\theta\{h(X)\} = \theta - 1$ for $\theta = 1, 2$. If* (I) *holds, then for every $\epsilon > 0$ there exists an $N_0' = N_0'(\epsilon)$ independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that $R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \geqq -\epsilon N^{-\frac{1}{2}}$ for $N \geqq N_0'$.*

PROOF. Partition the space under the $\mathbf{E}$ integral of (5.9) into $F_N = \{|\bar{h}_N - \bar{\delta}_N| \leqq N^{-\frac{1}{4}}\}$ and its complement F_N'. Note that Tchebichev's inequality implies $\mathbf{E}[F_N'] \leqq N^{\frac{1}{2}}\mathbf{E}(\bar{h}_N - \bar{\delta}_N)^2 = N^{-\frac{1}{2}}V^2$, while on F_N, $[\bar{\delta}_N \leqq Z(x) < \bar{h}_N] + [\bar{h}_N \leqq Z(x) < \bar{\delta}_N] \leqq [|Z(x) - \bar{\delta}_N| \leqq N^{-\frac{1}{4}}] = [G_N]$. Hence we have from (5.9) and Schwarz inequality

$$\begin{aligned} R_N(\boldsymbol{\theta}, \mathbf{t}') - \psi(\bar{\delta}_N) &\geqq -\mathbf{E}|\bar{h}_N - \bar{\delta}_N|\{(a+b)[F_N'] + \textstyle\int_W [G_N] \, d\mu(x)\} \\ &\geqq -\{\mathbf{E}(\bar{h}_N - \bar{\delta}_N)^2\}^{\frac{1}{2}}\{(a+b)(\mathbf{E}[F_N'])^{\frac{1}{2}} + \textstyle\int_W [G_N] \, d\mu(x)\} \\ &\geqq -\{N^{-\frac{3}{4}}(a+b)V^2 + N^{-\frac{1}{2}}V \textstyle\int_W [G_N] \, d\mu(x)\}. \end{aligned}$$

Since uniform continuity of $H(z)$, a consequence of (I), implies that with arguments of the type used in obtaining lower bounds on $\mathbf{E}\{\psi(\bar{h}_N) - \psi(\bar{\delta}_N)\}$ in Theorems 5.3.1 and 5.4.1 and a treatment of the $-B_N^*(\boldsymbol{\theta})$ term in (4.25) similar to that used for $A_N(\boldsymbol{\theta})$ in the proofs of Theorems 5.3 and 5.4.

THEOREM 5.5.1. *If the assumptions of Theorem* 5.5 *hold, then there exists a* $c_0'' > 0$ *independent of* $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ *such that* $R_N(\boldsymbol{\theta}, \mathbf{t}'') - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \geqq -c_0'' (\log N)N^{-1}$.

THEOREM 5.6.1. *Let the assumptions of Theorem* 5.6 *hold and assume h is non-degenerate under* P_1 *and* P_2. *Then, for every* $\epsilon > 0$ *there exists an* $N_0'' = N_0''(\epsilon)$ *independent of* $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ *such that* $R_N(\boldsymbol{\theta}, \mathbf{t}'') - \psi(\bar{\delta}_N(\boldsymbol{\theta})) \geqq -\epsilon N^{-\frac{1}{2}}$ *for* $N \geqq N_0'$.

There is an interesting comparison of Theorems 5.3 and 5.4 (or Theorems 5.5 and 5.6) with the analogous results of Hannan and Van Ryzin ([4], Theorems 2 and 3). In their paper it is shown that for the fixed sample size compound testing problem if one uses for the kth component problem the procedure $t'(\bar{h}_N, X_k)$, $k = 1, \cdots, N$, where $t'(\eta, x)$ is given by (5.5), then the regret function of this procedure is of $O(N^{-1})$ or of $o(N^{-\frac{1}{2}})$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ under condition (II) and (I) respectively. Hence, Theorem 5.4 (or Theorem 5.6) achieves exactly the same result in the sequential case, while Theorem 5.3 (or Theorem 5.5) does not achieve the same order $O(N^{-1})$ but the lesser order $O((\log N)N^{-1})$. It is not surprising that some rate of convergence is lost in the sequential case. But the fact that for Theorem 5.4 (or Theorem 5.6) and for Theorem 5.1 (or 5.2) the *same* rates of $o(N^{-\frac{1}{2}})$ and $O(N^{-\frac{1}{2}})$ respectively were attainable in the sequential case as in the fixed sample size case (Theorems 3 and 1 of [4] respectively) is worth noting.

6. Comments. We make the following comments:

(i) All the results in this paper can be extended to the class of randomized rules which in (2.9) assign the arbitrary value as r^{-1}, where $r = r(\xi, x)$ is the number of columns minimizing the quantities $(\xi, L^j f(x))$. This result follows by the symmetry argument used by the author in Section 7 of [14] and hence is not repeated here. In particular, the component rule in (5.5) would be replaced by

$$\begin{aligned} t^{\#}(\eta, x) &= 1 \quad \text{if} \quad Z(x) < \eta, \quad Z(x) \,\varepsilon\, (0, 1) \quad \text{or} \quad Z(x) = 0 \\ &= 0 \quad \text{if} \quad Z(x) > \eta, \quad Z(x) \,\varepsilon\, (0, 1) \quad \text{or} \quad Z(x) = 1 \\ &= \tfrac{1}{2} \quad \text{if} \quad Z(x) = \eta, \quad Z(x) \,\varepsilon\, (0, 1), \end{aligned}$$

for η real, $x \,\varepsilon\, \mathfrak{X}$. Note that $t^{\#}(\eta, x)$ is admissible Bayes for $0 \leqq \eta \leqq 1$.

(ii) The results given here are all non-Bayesian; that is, the parameter θ is not assumed to be random. However, suppose we do assume a Bayesian situation in which $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots\}$ is a sequence of independent identically distributed random variables with $\xi_\theta' = \Pr\{\theta_k = \theta\}$, $\theta = 1, \cdots, m$, $\sum_{\theta=1}^m \xi_\theta' = 1$ Let $\mathbf{E}_N^*$ denote expectation with respect to the N independent random variables $\theta_1, \cdots, \theta_N$ and define

$$\mathbf{R}_N(\xi', \mathbf{t}) = \mathbf{E}_N^* R_N(\boldsymbol{\theta}, \mathbf{t}), \tag{6.1}$$

where $R_N(\boldsymbol{\theta}, \mathbf{t})$ is given by (2.4) and $\xi' = (\xi_1', \cdots, \xi_m')$.

Let $\Xi = \{\xi' = (\xi_1', \cdots, \xi_m') \mid \sum_{\theta=1}^m \xi_\theta' = 1, \xi_i' \geqq 0\}$ be the class of distributions on Ω. The following theorem holds.

THEOREM 6.1. *Let* $\mathbf{t}$ *be of the form* $\mathbf{t}^*$ (*or* $\mathbf{t}^{**}$) *and let the assumptions of Theorem* 4.1 *or* 4.2 (*or Theorem* 4.3) *hold. Then there exists a constant* c *independent of* $\xi' \varepsilon \Xi$ *such that*

$$0 \leqq \mathbf{R}_N(\xi', \mathbf{t}) - \phi(\xi') \leqq cN^{-\frac{1}{2}}. \tag{6.2}$$

PROOF. Observe that by optimality of $t(\xi', x)$ (see (2.9)) with $\xi = \xi'$) in the component problem, we have $\mathbf{E}_k^*\mathbf{E}(L_{\theta_k}, \mathbf{t}_k) \geqq \phi(\xi')$, where $\mathbf{E}(L_{\theta_k}, \mathbf{t}_k)$ is as in (2.3) (see also (14) on p. 3 of [8]). Hence by interchanging the order of integration in (6.1) we have $\mathbf{R}_N(\xi', \mathbf{t}) = N^{-1} \sum_{k=1}^N \mathbf{E}_k^*\mathbf{E}(L_{\theta_k}, \mathbf{t}_k) \geqq \phi(\xi')$, and the left-hand side of (6.2) is proved.

To obtain the right-hand inequality in (6.2) note that

$$\mathbf{R}_N(\xi', \mathbf{t}) - \phi(\xi') = \mathbf{E}_N^*\{R_N(\boldsymbol{\theta}, \mathbf{t}) - \phi(p_N(\boldsymbol{\theta}))\} + \mathbf{E}_N^*\{\phi(p_N(\boldsymbol{\theta})) - \phi(\xi')\}. \tag{6.3}$$

The integrand in the first term of (6.3) is bounded from above by $c_0N^{-\frac{1}{2}}$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, where $c_0 = \max\{c_1, c_2, c_3\}$ where c_i as the constant of Theorem 4.i, $i = 1, 2, 3$. Hence, the first term in (6.3) is bounded by $c_0N^{-\frac{1}{2}}$ uniformly in $\xi' \varepsilon \Xi$. The second term in (6.3) is bounded from above by $B\mathbf{E}_N^* \|p_N(\boldsymbol{\theta}) - \xi'\|$ by inequality (3.7). But by the Schwarz inequality and independence of the θ's, we have $\{\mathbf{E}_N^* \|p_N(\boldsymbol{\theta}) - \xi'\|\}^2 \leqq \mathbf{E}_N^* \|p_N(\boldsymbol{\theta}) - \xi'\|^2 = \sum_{\theta=1}^m \mathbf{E}_N^*(p_{N\theta}(\boldsymbol{\theta}) - \xi_\theta')^2 = N^{-1} \sum_{\theta=1}^m \xi_\theta'(1 - \xi_\theta') < N^{-1}$. Hence the second term in (6.3) is bounded by $\mathrm{B}N^{-\frac{1}{2}}$ uniformly in $\xi' \varepsilon \Xi$. Taking $c = c_0 + B$ completes the proof.

Results of higher order analogous to Theorems 5.3–5.6 are also true in this Bayesian situation. However, we shall forego their statement and proof.

The Bayesian results described herein are closely related to the "empirical Bayes" approach of Robbins in [7] and [8].

(iii) All the convergence theorems in this paper are concerned with risk convergence and are of order at least $N^{-\frac{1}{2}}$. This poses the interesting problem of whether the corresponding losses (see (2.5) for definition) have a convergence rate on them in the sense that $N^\epsilon\{W_N(\boldsymbol{\theta}, \mathbf{t}) - \phi(p_N(\boldsymbol{\theta}))\} \to 0$ in probability (or with probability 1), $0 \leqq \epsilon < 1$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. A recent paper by Samuel [10] with $\epsilon = 0$ gives non-uniform results and calculations of this type for the rules she proposed in [9] for the sequential compound testing problem. Hannan and Robbins in [3] (Theorem 3) give one-sided uniform results of this type with almost sure convergence for the fixed sample size compound testing problem with $\epsilon = 0$.

The results of this paper hint that in the sequential compound decision problem results with an $\epsilon > 0$ might be possible, while the papers by Hannan and Van Ryzin [4] and Van Ryzin [14] point to the same possibilities in the fixed sample size case for the compound testing problem and the general $m \times n$ compound decision problem, respectively. Uniform (in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$) theorems in this direction will be given in a forthcoming paper [13].

Acknowledgment. The author wishes to thank the referee for pointing out changes which considerably improved an earlier version of the paper.

REFERENCES

[1] HANNAN, JAMES F. (1957). Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games*, **3** 97–139. Princeton University Press.

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

[3] HANNAN, JAMES F. and ROBBINS, HERBERT (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37–51.

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

[5] LOÈVE, MICHEL (1960). *Probability Theory* (2nd ed.). Van Nostrand, Princeton.

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

[7] ROBBINS, HERBERT (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 157–163. Univ. of California Press.

[8] ROBBINS, HERBERT (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[9] SAMUEL, ESTER (1963). Asymptotic solutions of the sequential compound decision problem. *Ann. Math. Statist.* **34** 1079–1094.

[10] SAMUEL, ESTER (1964). Convergence of the losses of certain decision rules for the sequential compound decision problem. *Ann. Math. Statist.* **35** 1606–1621.

[11] SAMUEL, ESTER (1965). Sequential compound estimators. *Ann. Math. Statist.* **36** 879–889.

[12] VAN RYZIN, J. R. (1964). Asymptotic solutions to compound decision problems. Ph.D. Thesis, Department of Statistics, Michigan State University.

[13] VAN RYZIN, J. (1965). Uniform convergence of losses in the fixed and sequential $m \times n$ compound decision problem (Abstract). *Ann. Math. Statist.* **36** 1323–1324.

[14] VAN RYZIN, J. R. (1966). The compound decision problem with $m \times n$ finite loss matrix. *Ann. Math. Statist.* **37** 412–424.

[15] VAN RYZIN, J. (1966). Repetitive play in finite statistical games with unknown distribution. *Ann. Math. Statist.* **37** 976–994.