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

By J. R. Van Ryzin

Argonne National Laboratory

0. Summary. Simultaneous consideration of N statistical decision problems having identical generic structure constitutes a compound decision problem. The risk of a compound decision problem is defined as the average risk of the component problems. When the component decisions are between two fully specified distributions P_1 and P_2, $P_1 \neq P_2$, Hannan and Robbins [5] give a decision function whose risk is uniformly close (for N large) to the risk of a best "simple" procedure based on knowing the proportion of component problems in which P_2 is the governing distribution. This result was motivated by heuristic arguments and an example (component decisions between $\mathfrak{N}(-1, 1)$ and $\mathfrak{N}(1, 1)$) given by Robbins [8]. In both papers, the decision functions for the component problems depended on data from all N problems.

This paper generalizes and strengthens a result of Hannan and Robbins (Theorem 4, [5]) to the case where each component problem consists of making one of n decisions based on an observation from one of m distributions. Specifically, we find upper bounds for the difference in the risks (the regret function) of a certain compound procedure and a best "simple" procedure which is Bayes against the empirical distribution on the component parameter space. Theorem 2 gives sufficient conditions for a uniform (in parameter sequences) bound on the regret function of order $N^{-\frac{1}{2}}$, while Theorem 3 states sufficient conditions for a uniform bound of order N^{-1}. For $m = n = 2$, Theorem 2 furnishes a strengthening of Theorem 4 of [5]. More extensive results for the case $m = n = 2$ are given in a paper by Hannan and Van Ryzin [6].

Please note that the case considered here makes the N-decisions after the data from all N problems are available. The sequential case (kth decision after observations $1, 2, \cdots, k, k = 1, \cdots, N$) is treated by Hannan in [3] and by Samuel in [10].

1. Statement of the problem and notation.

Consider the following finite statistical decision problem. Let X be a random variable (or arbitrary dimensionality) known to have one of m possible distributions P_θ, θ in the finite parameter space $\Omega = \{1, \cdots, m\}$. Based on observing X we are required to make a decision $d \in \mathfrak{D} = \{1, \cdots, n\}$ incurring loss $L(\theta, d)$ if decision d is made when X is distributed as P_θ, $\theta = 1, \cdots, m$; $d = 1, \cdots, n$.

If we simultaneously consider N decision problems each with this generic

Received 23 October 1964; revised 5 November 1965.

[1] This paper is part of a Michigan State University doctoral thesis [12], supported in part by the National Science Foundation, Grant G-18976 and in part by the U. S. Atomic Energy Commission.

structure, then the N-fold global problem is called a finite compound decision problem. More precisely, let X_k, $k = 1, \cdots, N$, be N independent observations, X_k distributed according to P_{θ_k} with θ_k in Ω. Based on all N observations, a decision d_k in $\mathfrak{D}$ is to be made for each of the N component problems. For the kth subproblem, the decision $d_k = d$ represents selecting the dth column of the $m \times n$ loss matrix $(L(\theta, d))$. Note that in the case considered here all N decisions are held in abeyance until all random variables X_k, $k = 1, \cdots, N$, have been observed.

Before proceeding, we introduce the following notation. Let there exist a σ-finite measure μ dominating $\{P_1, \cdots, P_m\}$ such that the densities

$$(1) \qquad f_\theta(x) = (dP_\theta/d\mu)(x) \leqq K \qquad \text{a.e. } \mu$$

for some $K < \infty$. There is no loss of generality in this assumption since we may always choose $\mu = \sum_{\theta=1}^{m} P_\theta$ and $K = 1$. Also, let $f(x) = (f_1(x), \cdots, f_m(x))$ denote the vector of densities in (1).

In referring to the $m \times n$ matrix of losses $L(\theta, d)$, the rows will be denoted by L_θ, the columns by L^d, and the ordered difference of two columns by $L^{dd'} = L^d - L^{d'}$ with components $L(\theta, d) - L(\theta, d') = L_\theta^{dd'}$, $\theta = 1, \cdots, m$; $d, d' = 1, \cdots, n$.

The following notation will be used throughout the paper:

Let E^m be m-dimensional Euclidean space. Let $y = (y_1, \cdots, y_m)$ and $z = (z_1, \cdots, z_m)$ be vectors in E^m. Define the vector $yz = (y_1z_1, \cdots, y_mz_m)$. The inner product and norm of E^m will be denoted respectively by $(y, z) = \sum_{i=1}^{m} y_iz_i$ and $\|y\| = (y, y)^{\frac{1}{2}}$. The dimension of the space in which these operations are carried out will always be clear from the context.

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

$$\begin{aligned} [A](a) &= 1 \qquad \text{if } a \varepsilon A \\ &= 0 \qquad \text{if } a \not\varepsilon A. \end{aligned}$$

2. Decision procedures and some preliminaries. For the compound decision problem, a decision procedure may depend on the full observation $\mathbf{X} = (X_1, \cdots, X_N)$. Any $N \times n$ matrix of measurable functions $T(\mathbf{x}) = (t_d^k(\mathbf{x}))$ will be called a randomized decision function (procedure) for the compound decision problem if for $k = 1, \cdots, N$; $d = 1, \cdots, n$, $t_d^k(\mathbf{x}) = \Pr\{d_k = d \mid \mathbf{X} = \mathbf{x}\}$ and $\sum_{d=1}^{n} t_d^k(\mathbf{x}) = 1$. The kth row of $T(\mathbf{x})$ will be denoted by $t^{(k)}(\mathbf{x}) = (t_1^k(\mathbf{x}), \cdots, t_n^k(\mathbf{x}))$.

The decision function $T(\mathbf{x})$ is said to be *simple* if there exist functions $t_d(\cdot)$, $d = 1, \cdots, n$ such that $t^{(k)}(\mathbf{x}) = (t_1(x_k), \cdots, t_n(x_k))$ for $k = 1, \cdots, N$. A simple decision function will be denoted by $t = (t_1, \cdots, t_n)$.

Let $\mathbf{\Omega}$ be the set of all N-tuples $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_N)$, $\theta_k \varepsilon \Omega$. For each $\boldsymbol{\theta} \varepsilon \mathbf{\Omega}$, let $\mathbf{E}$ denote expectation with respect to $\times_{k=1}^{N} P_{\theta_k}$. Then denote by $R(\boldsymbol{\theta}, T)$ the risk function for the compound decision procedure $T(\mathbf{x})$. This risk is defined to be the average of the component risks $R_k(\boldsymbol{\theta}, T) = \mathbf{E}\{\sum_{d=1}^{n} L(\theta_k, d)t_d^k(\mathbf{X})\} = \mathbf{E}(L_{\theta_k}, t^{(k)}(\mathbf{X}))$, for each subproblem, $k = 1, \cdots, N$. Hence

$$(2) \qquad R(\boldsymbol{\theta}, T) = N^{-1}\sum_{k=1}^{N} R_k(\boldsymbol{\theta}, T) = \mathbf{E}N^{-1}\sum_{k=1}^{N} (L_{\theta_k}, t^{(k)}(\mathbf{X})).$$

The risk (2) may be considerably simplified in the case of a simple decision function. For $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ and $\theta = 1, \cdots, m$, define the relative frequencies $p_\theta(\boldsymbol{\theta}) = N^{-1}\{\# \text{ of } \theta_k = \theta, k \leqq N\}$, of problems in the first N problems in which the distribution P_θ governs. The vector $p(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \cdots, p_m(\boldsymbol{\theta}))$ will be called the empirical distribution on Ω.

Let $t = (t_1, \cdots, t_n)$ be a simple decision function and let E_θ denote expectation with respect to P_θ. Then, by (2), the risk incurred in using procedure t is

$$(3) \qquad \begin{aligned} R(\boldsymbol{\theta}, t) &= N^{-1}\sum_{k=1}^{N} \mathbf{E}(L_{\theta_k}, t(X_k)) \\ &= \sum_{\theta=1}^{m} p_\theta(\boldsymbol{\theta})\rho_\theta(t) = (p(\boldsymbol{\theta}), \rho(t)), \end{aligned}$$

where

$$(4) \qquad \begin{aligned} \rho_\theta(t) &= E_\theta(L_\theta, t(X)) = E_\theta\{\sum_{d=1}^{n} L(\theta, d)t_d(X)\}, \\ \rho(t) &= (\rho_1(t), \cdots, \rho_m(t)). \end{aligned}$$

Let $\xi = (\xi_1, \cdots, \xi_m)$ be any vector in m-dimensional Euclidean space. Let t be a simple decision function, that is, $t_d(x) \geqq 0$, $d = 1, \cdots, n$ is a set of measurable functions such that $\sum_{d=1}^{n} t_d(x) = 1$.

$$(5) \qquad \psi(\xi, t) = (\xi, \rho(t)).$$

Note that for $\xi = p(\boldsymbol{\theta})$ the function ψ becomes the risk function (3) for the simple decision procedure t.

The problem of choosing $t(x)$ to minimize $\psi(\xi, t)$ for fixed ξ is straightforward. From (1) and (5), we have

$$(6) \qquad \psi(\xi, t) = \int \{\sum_{d=1}^{n} (\xi, L^d f(x))t_d(x)\}\, d\mu(x).$$

Therefore, (6) is minimized for fixed ξ by any vector function t_ξ (defined a.e. μ) 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_ξ is of the form

$$(7) \qquad \begin{aligned} t_{\xi,d}(x) &= 0 \quad \text{if } (\xi, L^d f(x)) > \min_j (\xi, L^j f(x)) \\ &= 1 \quad \text{if } (\xi, L^d f(x)) < \min_{j \neq d} (\xi, L^j f(x)) \\ &= \text{arbitrary} \quad \text{if } (\xi, L^d f(x)) = \min_{j \neq d}(\xi, L^j f(x)), \end{aligned}$$

such that $t_{\xi,d}(x) \geqq 0$ for $d = 1, \cdots, n$ and $\sum_{d=1}^{n} t_{\xi,d}(x) = 1$ a.e. μ.

Note that if ξ is a bona fide *a priori* distribution, $(0 \leqq \xi_i, \sum_{i=1}^{m} \xi_i = 1)$, then such a t_ξ would be a decision procedure Bayes against ξ.

We observe that any randomized procedure of the form (7) minimizing $\psi(\xi, t)$ may be replaced by a non-randomized version which also minimizes $\psi(\xi, t)$ for fixed ξ. In particular, one such non-randomized version is given by the coordinate functions

$$(8)\qquad \begin{aligned} t'_{\xi,d}(x) &= 1 \quad \text{if } d \text{ is the smallest integer for which } (\xi, L^d f(x)) = \\ &\qquad\qquad \min_{j=1,\cdots,n} (\xi, L^j f(x)) \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

To see that (8) is of the form (7) we merely note that $t_\xi'(x) = (t'_{\xi,1}(x), \cdots, t'_{\xi,n}(x))$ is a probability distribution concentrating on the first column minimizing the quantities $(\xi, L^j f(x))$. It should be noted that t_ξ' may possibly be inadmissible, but this is not relevant to the results obtained herein. In what follows we restrict ourselves to the non-randomized version t_ξ' of the Bayes procedure t_ξ.

In [2], p. 102, Hannan has given a useful inequality for Bayes rules. A statement and proof of a similar result is given here.

LEMMA 1. *Let Y be a space closed under subtraction. Let $M(y, z)$ be a real-valued function on $Y \times Z$ such that $M(\cdot, z)$ is linear on Y for each $z \varepsilon Z$ and $\inf_z M(y, z)$ is attained for each $y \varepsilon Y$. Define $g(y) = \inf_z M(y, z)$ and let $z(y)$ be any Z-valued function such that $g(y) = M(y, z(y))$ on Y. Then, $y, y' \varepsilon Y$,*

$$0 \leqq M(y, z(y')) - g(y) \leqq M(y - y', z(y')) - M(y - y', z(y)).$$

PROOF. The lower inequality results from the definition of $g(y)$ and the upper inequality follows by adding the non-negative term $M(y', z(y)) - g(y')$.

Now define for $\xi \varepsilon E^m$ the function

$$(9)\qquad \phi(\xi) = \inf_t \psi(\xi, t) = (\xi, \rho(t_\xi)).$$

Observing that $(\xi, \rho(t))$ is linear in ξ and ρ, Lemma 1 and (9) yield

COROLLARY 1. *If $\xi, \xi^* \varepsilon E^m$, then*

$$(10)\qquad 0 \leqq \psi(\xi, t_{\xi^*}) - \phi(\xi) \leqq (\xi - \xi^*, \rho(t_{\xi^*}) - \rho(t_\xi)).$$

This corollary inspires the non-simple rule to be adopted later (see (12)). If $p^* \varepsilon E^m$ is a good approximation to $p(\boldsymbol{\theta})$ in the sense that $\|p^* - p(\boldsymbol{\theta})\|$ is small, then Corollary 1 says that a simple procedure $t_{p^*}(x)$ has risk within $\|p^* - p(\boldsymbol{\theta})\| \, \|\rho(t_{p^*}) - \rho(t_{p(\boldsymbol{\theta})})\|$ of the minimum attainable risk in the class of all simple procedures, given by $\phi(p(\boldsymbol{\theta}))$. Therefore, not knowing $p(\boldsymbol{\theta})$ in general, we seek estimates $\hat{p} = \hat{p}(X_1, \cdots, X_N)$ of $p(\boldsymbol{\theta})$ which with the aid of Lemma 5 take advantage of the risk approximation of Corollary 1.

In Section 5, we shall use the following lemma which is a simple consequence of the Berry-Esseen normal approximation theorem (see Loève [7], p. 288).

LEMMA 2. *If $Y_1, Y_2, \cdots, Y_n$ are independent identically distributed random variables with mean 0, variance 1 and third absolute moment γ, then for α and l real, $l \geqq 0$,*

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

where β is an absolute constant.

3. Estimation of empirical distributions on Ω. The results of this section are based on some unpublished lecture notes of Hannan [4]. See also Robbins [9], Section 7, and Teicher [11].

Let $L_1(\mu)$ and $L_2(\mu)$ be the function spaces of μ-integrable and μ-square integrable functions respectively. The usual norm and inner product for f, $g \varepsilon L_2(\mu)$ will be denoted respectively by $\|f\|_\mu$ and $(f, g)_\mu$. Note that the $f_\theta(x)$ as densities are in $L_1(\mu)$ and hence in $L_2(\mu)$ because they are bounded.

We make the following definitions. Let $S^{(m)} = \{\eta \mid \eta \varepsilon E^m, \eta_\theta \geqq 0, \sum_{\theta=1}^m \eta_\theta = 1\}$ be the simplex in E^m. For $\eta \varepsilon S^{(m)}$ define the probability *mixture* $P_\eta = \sum_{\theta=1}^m \eta_\theta P_\theta$ with μ-density $f_\eta(x) = (\eta, f(x))$. The class of all mixtures $\mathcal{P} = \{P_\eta \mid \eta \varepsilon S^{(m)}\}$ is said to be *identifiable* if for any $\eta, \eta' \varepsilon S^{(m)}$, $P_\eta = P_{\eta'}$ (or $f_\eta = f_{\eta'}$ a.e. μ) implies $\eta = \eta'$.

A vector function $h = (h_1(x), \cdots, h_m(x))$ into E^m with coordinate functions $h_j \varepsilon L_1(\mu)$ is an *unbiased estimate* for the class $\mathcal{P}$ if $E_\eta\{h(X)\} = \eta$ for all $\eta \varepsilon S^{(m)}$, where E_η denotes expectation with respect to the mixture P_η. If such an h exists the class $\mathcal{P}$ is said to be *estimable*. Let $\mathcal{E}$ be the class of all unbiased estimates for the class $\mathcal{P}$.

We state the following lemma without proof as its proof is simple and is essentially contained in Hannan [4] or Teicher [11], Theorem 1.

Lemma 3. *The class $\mathcal{P}$ is identifiable if and only if the set of densities $\{f_1, \cdots, f_m\}$ are linearly independent in $L_1(\mu)$.*

Let S be any linear subspace of $L_2(\mu)$ and $S^\perp$ be the orthogonal complement of S in $L_2(\mu)$. For any $g \varepsilon L_2(\mu)$, denote by g_S, $g_S^\perp$ the projection of g on S and $S^\perp$ respectively. Note that if $g \varepsilon L_2(\mu)$, $g = g_S + g_S^\perp$.

Let $\mathcal{H}$ denote the subclass of $\mathcal{E}$ for which $h_j \varepsilon L_2(\mu)$ for $j = 1, \cdots, m$. We now give a theorem which proves the existence of unbiased estimates for $\mathcal{P}$ and which yields the structure of the class $\mathcal{H}$. For $j = 1, \cdots, m$, let S_j be the subspace of $L_2(\mu)$ spanned by $\{f_\theta \mid \theta \neq j\}$. Let S be the subspace of $L_2(\mu)$ spanned by $\{f_1, \cdots, f_m\}$.

Theorem 1. *Let the set of densities $\{f_1, \cdots, f_m\}$ be linearly independent in $L_1(\mu)$. Then, the class $\mathcal{H}$ is non-empty. Furthermore, $h \varepsilon \mathcal{H}$ if and only if $h(x) = f^*(x) + g(x)$ a.e. μ, where $f_j^*(x) = (f_{jS_j^\perp}(x))(\|f_{jS_j^\perp}\|_\mu^2)^{-1}$ and $g_j(x) \varepsilon S^\perp$ for $j = 1, \cdots, m$.*

Proof. First note that linear independence of the densities $\{f_1, \cdots, f_m\}$ implies $\|f_{jS_j^\perp}\|_\mu > 0$, and hence $f_j^*(x)$ is well-defined. Furthermore, since $f_j^*(x) \varepsilon S_j^\perp$, $E_\theta f_j^*(X) = (f_j^*, f_\theta)_\mu = \delta_{\theta j}$, the Kronecker delta. Thus, $f^* \varepsilon \mathcal{E}$, $\mathcal{H}$ is non-empty, and $g_j \varepsilon S^\perp$, $j = 1, \cdots, m$ implies $f^* + g \varepsilon \mathcal{H}$. Conversely, if $h \varepsilon \mathcal{H}$, $E_\theta(h_j - f_j^*) = 0$ for all $\theta, j = 1, \cdots, m$ and $g_j = h_j - f_j^*$ is in $S^\perp, j = 1, \cdots, m$.

Observe that the functions f_j^* of Theorem 1 form the dual basis to $\{f_1, \cdots, f_m\}$ in the conjugate space of the subspace S.

Corollary 2. *There exist $h \varepsilon \mathcal{E}$ such that $|h_j(x)| \leqq M$ a.e. μ for $j = 1, \cdots, m$ and M finite.*

Proof. Choose $h_j(x) = f_j^*(x)$ for $j = 1, \cdots, m$. Then, since the f_j^*'s lie in S, they are essentially bounded as linear combinations of the essentially bounded densities $\{f_1, \cdots, f_m\}$.

Corollary 3. *$\mathcal{P}$ is identifiable if and only if $\mathcal{P}$ is estimable.*

Proof. If $\mathcal{P}$ is identifiable, Lemma 3 and Theorem 1 establish estimability and necessity follows. Sufficiency follows by noting that $\mathcal{P}$ estimable and $\eta_1 \neq \eta_2$ implies $E_{\eta_1}(h) = \eta_1 \neq \eta_2 = E_{\eta_2}(h)$ and hence $P_{\eta_1} \neq P_{\eta_2}$.

See also Choi [1] for a slightly different version of this corollary.

In view of Lemma 3, Theorem 1 and Corollary 3, and our dependence on the estimates h in defining our procedures (Section 4), we assume in the remainder of the paper that the set of densities $\{f_1, \cdots, f_m\}$ are linearly independent in $L_1(\mu)$.

The importance of the class $\mathcal{E}$ in estimating $p(\boldsymbol{\theta})$ can now be seen. With $\mathbf{X} = (X_1, \cdots, X_N)$ as before and $h \varepsilon \mathcal{E}$, define the random variable,

$$\bar{h}(\mathbf{X}) = N^{-1} \textstyle\sum_{k=1}^{N} h(X_k). \tag{11}$$

This equation yields an unbiased estimate of the empirical distribution $p(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, since $\mathbf{E}\bar{h}(\mathbf{X}) = N^{-1} \sum_{k=1}^{N} \epsilon_{\theta_k} = p(\boldsymbol{\theta})$. If $h \varepsilon \mathcal{E}$ and h is bounded as in Corollary 2, then $\bar{h}(\mathbf{X})$ inherits this boundedness through (11).

Consider now the subclass $\mathcal{H}$ of $\mathcal{E}$. If $h = (h_1, \cdots, h_m) \varepsilon \mathcal{H}$, then boundedness of the densities f_θ implies $E_\theta h_j^2(X) < \infty$. Denote the variance of h_j under P_θ for $\theta, j = 1, \cdots, m$ by $\sigma_\theta^2(h_j)$.

Lemma 4. *If $h \varepsilon \mathcal{H}$, then* $\mathbf{E}\, \|\bar{h} - p(\boldsymbol{\theta})\|^2 \leqq C^2N^{-1}$, *where* $C^2 = \max_\theta \sum_{j=1}^{m} \sigma_\theta^2(h_j)$.

Proof. By direct computation, we have

$$\begin{aligned}\mathbf{E}\, \|\bar{h} - p(\boldsymbol{\theta})\|^2 &= \textstyle\sum_{j=1}^{m} \mathbf{E}(\bar{h}_j - p_j(\boldsymbol{\theta}))^2 \\ &= N^{-1} \textstyle\sum_{j=1}^{m} \sum_{\theta=1}^{m} p_\theta(\boldsymbol{\theta})\sigma_\theta^2(h_j) \leqq C^2N^{-1}.\end{aligned}$$

4. Non-simple decision functions. With $h \varepsilon \mathcal{H}$ and the estimate $\bar{h}(\mathbf{X})$ of $p(\boldsymbol{\theta})$ given by (11), we now define a non-simple decision function which results from substituting $\bar{h}(\mathbf{X})$ for $p(\boldsymbol{\theta})$ in $t_{p(\boldsymbol{\theta})}$ as given by (7) (see [5], p. 44 and [6], Equation (12)). In so doing, we shall confine ourselves to that particular non-randomized version of $t_{\bar{h}}$ given by (8) and denoted by $t_{\bar{h}}'$. The resulting non-simple, non-randomized decision procedure T' consists of the N vector functions $t^{(k)}(\mathbf{x}) = t_{\bar{h}}'(x_k) = (t_{\bar{h},1}'(x_k), \cdots, t_{\bar{h},n}'(x_k))$ for $k = 1, \cdots, N$, where

$$\begin{aligned} t_{\bar{h},d}'(x_k) &= 1 \quad \text{if } d \text{ is the smallest integer for which } (\bar{h}, L^d f(x_k)) = \\ &\qquad \min\nolimits_{j=1,\cdots,n} (\bar{h}, L^j f(x_k)) \\ &= 0 \quad \text{otherwise,} \qquad\qquad d = 1, \cdots, n. \end{aligned} \tag{12}$$

The question immediately arises regarding optimality properties of the procedure T'. As a partial answer to this question, consider the function

$$R(\boldsymbol{\theta}, T) - \phi(p(\boldsymbol{\theta})) \tag{13}$$

for the decision function $T(\mathbf{x})$ and $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. This function will be called the *regret function* against simple decision functions for the decision procedure $T(\mathbf{x})$. We shall consider decision procedures $T(\mathbf{x})$ which makes the regret function small uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ for all N. In Theorems 2 and 3 it will be shown that the procedure T' has, under suitable conditions, good asymptotic properties in the sense that its regret function given by (13) has an upper bound approaching zero uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ as $N \to \infty$.

We now give a useful decomposition lemma for the risk $R(\boldsymbol{\theta}, T)$ in (13) for $T(\mathbf{x})$ such that

$$t^{(k)}(\mathbf{x}) = t_{\bar{h}}(x_k) = (t_{\bar{h},1}(x_k), \cdots, t_{\bar{h},n}(x_k)), \tag{14}$$

where $t_{\bar{h},d}(x_k)$ is defined by (7) with $\xi = \bar{h}$ and $x = x_k$.

Lemma 5. *Let X be a random variable independent of $\mathbf{X}$. If $T(\mathbf{x})$ is a compound decision function of the form* (14) *and $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, then,*

$$R(\boldsymbol{\theta}, T) = \mathbf{E}(p(\boldsymbol{\theta}), \rho(t_{\bar{h}})) + N^{-1} \textstyle\sum_{k=1}^{N} \sum_{d \neq d'} \mathbf{E}E_{\theta_k} L_{\theta_k}^{dd'} t_{\bar{h}^{(k)},d}(X) t_{\bar{h},d'}(X), \tag{15}$$

where $\rho_\theta(t_{\bar{h}}) = E_\theta(L_\theta, t_{\bar{h}}(X))$ and $\bar{h}^{(k)} = \bar{h} + N^{-1}(h(X) - h(X_k))$ and the E_{θ_k} integral in each of the N terms in the sum on the right-hand side of (15) *is on X.*

Proof. Fix $k = 1, \cdots, N$ and express $\mathbf{E}(L_{\theta_k}, t^{(k)}(\mathbf{X}))$ as an iterated integral, make a change of variable, and perform an added integration as follows,

$$\begin{aligned} \mathbf{E}(L_{\theta_k}, t^{(k)}(\mathbf{X})) &= \textstyle\int (L_{\theta_k}, t_{\bar{h}}(x_k))\, dP_{\theta_k}(x_k) \prod_{i \neq k} dP_{\theta_i}(x_i) \\ &= \textstyle\int (L_{\theta_k}, t_{\bar{h}^{(k)}}(x))\, dP_{\theta_k}(x) \prod_{i \neq k} dP_{\theta_i}(x_i) \\ &= \textstyle\int (L_{\theta_k}, t_{\bar{h}^{(k)}}(x))\, dP_{\theta_k}(x) \prod_{i} dP_{\theta_i}(x_i) \\ &= \mathbf{E}E_{\theta_k}(L_{\theta_k}, t_{\bar{h}^{(k)}}(X)), \end{aligned} \tag{16}$$

where $\mathbf{E}E_{\theta_k}$ represents an iterated integral. Writing $t_{\bar{h}^{(k)}}(x) = t_{\bar{h}^{(k)}}(x) - t_{\bar{h}}(x) + t_{\bar{h}}(x)$ in the right-hand side of (16) and averaging over all k, we have by (2)

$$\begin{aligned} R(\boldsymbol{\theta}, T) = N^{-1} \textstyle\sum_{k=1}^{N} \mathbf{E}E_{\theta_k}(L_{\theta_k}, t_{\bar{h}}(X)) \\ + N^{-1} \textstyle\sum_{k=1}^{N} \mathbf{E}E_{\theta_k}(L_{\theta_k}, t_{\bar{h}^{(k)}}(X) - t_{\bar{h}}(X)). \end{aligned} \tag{17}$$

The first term on the right-hand side of (17) may be simplified to $\mathbf{E}(p(\boldsymbol{\theta}), \rho(t_{\bar{h}}))$ by noting that for $\theta_k = \theta$, $E_{\theta_k}(L_{\theta_k}, t_{\bar{h}}(X))$ are pointwise equal to $\rho_\theta(t_{\bar{h}(\mathbf{x})})$.

Since the components of $t_{\bar{h}}(x)$ and of $t_{\bar{h}^{(k)}}(x)$ sum to unity, we have the simple equation

$$\begin{aligned} &(L_{\theta_k}, t_{\bar{h}^{(k)}}(x) - t_{\bar{h}}(x)) \\ &\qquad = \textstyle\sum_{d,d'} \{L(\theta_k, d) t_{\bar{h}^{(k)},d}(x) t_{\bar{h},d'}(x) - L(\theta_k, d') t_{\bar{h}^{(k)},d}(x) t_{\bar{h},d'}(x)\}, \end{aligned}$$

which shows that the second term in (17) is equal to the second term in (15).

Lemma 6. *Let $T'(\mathbf{x})$ be the procedure defined by* (12). *Then,*

$$R(\boldsymbol{\theta}, T') - \phi(p(\boldsymbol{\theta})) \leqq A_N + B_N, \tag{18}$$

where

$$A_N = \mathbf{E}(p(\boldsymbol{\theta}) - \bar{h}, \rho(t_{\bar{h}}') - \rho(t'_{p(\boldsymbol{\theta})}))$$

and

$$B_N = N^{-1} \textstyle\sum_{k=1}^{N} \sum_{d \neq d'} L_{\theta_k}^{dd'} \mathbf{E}E_{\theta_k} t'_{\bar{h}^{(k)},d}(X) t'_{\bar{h},d'}(X).$$

PROOF. Identify T' with T in Lemma 5. The result follows immediately from (5) and Corollary 1 by taking $\xi = p(\boldsymbol{\theta})$ and $\xi' = \bar{h}$.

5. A uniform bound of $O(N^{-\frac{1}{2}})$. The following theorem generalizes Theorem 1 of [6] (which strengthens Theorem 4 of [5]) to the case where the component problem has an $m \times n$ loss matrix.

THEOREM 2. *If $h \varepsilon \mathcal{E}$ and $E_\theta |h_j(X)|^3 < \infty$ for $\theta, j = 1, \cdots, m$, then $R(\boldsymbol{\theta}, T') - \phi(p(\boldsymbol{\theta})) \leqq cN^{-\frac{1}{2}}$ where c is independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ for all N.*

PROOF. In inequality (18) we show: (i) $A_N \leqq c_1 N^{-\frac{1}{2}}$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ for all N and (ii) $B_N \leqq c_2 N^{-\frac{1}{2}}$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ for all N.

(i) By the Schwarz m-space inequality, we have,

$$N^{\frac{1}{2}} A_N \leqq N^{\frac{1}{2}} \mathbf{E} \, \|\bar{h} - p\| \, \|\rho(t_{\bar{h}}') - \rho(t_p')\|, \qquad p = p(\boldsymbol{\theta}). \tag{19}$$

Let $\underline{L}_\theta = \min_d L(\theta, d)$ and $\bar{L}_\theta = \max_d L(\theta, d)$ and note that for every t $\underline{L}_\theta \leqq \rho_\theta(t) \leqq \bar{L}_\theta$, Then

$$\begin{aligned} \|\rho(t_{\bar{h}}') - \rho(t_p')\|^2 &= \textstyle\sum_{\theta=1}^m \{\rho_\theta(t_{\bar{h}}') - \rho_\theta(t_p')\}^2 \\ &\leqq \textstyle\sum_{\theta=1}^m (\bar{L}_\theta - \underline{L}_\theta)^2 \\ &= \|\bar{L} - \underline{L}\|^2, \end{aligned} \tag{20}$$

where $\bar{L} = (\bar{L}_1, \cdots, \bar{L}_m)$ and $\underline{L} = (\underline{L}_1, \cdots, \underline{L}_m)$.

Also, note that by the Schwarz integral inequality and Lemma 4,

$$N^{\frac{1}{2}} \mathbf{E} \, \|\bar{h} - p(\boldsymbol{\theta})\| \leqq \{N\mathbf{E} \, \|\bar{h} - p(\boldsymbol{\theta})\|^2\}^{\frac{1}{2}} \leqq C. \tag{21}$$

Inequalities (20) and (21), when substituted into (19), imply $N^{\frac{1}{2}} A_N \leqq C \, \|\bar{L} - \underline{L}\|$. Hence, (i) is proved.

(ii) 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 Λ_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 $W_\theta W_\theta' = \Lambda_\theta$ 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

$$E_\theta(Z_\theta(X), g)^2 = \|g\|^2 = 1 \tag{22}$$

and

$$E_\theta \, \|Z_\theta(X)\|^2 = r_\theta. \tag{23}$$

Now 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 $t_{\bar{h}(k),d}' = t_{\bar{h},d}' = 1$, we observe that

$$t'_{\bar{h}^{(k)},d}(X)t'_{\bar{h},d'}(X) + t'_{\bar{h}^{(k)},d'}(X)t'_{\bar{h},d}(X)$$

$$\leqq [0 < (\bar{h}, L^{dd'}f(X)) \leqq (\bar{h} - \bar{h}^{(k)}, L^{dd'}f(X))] \tag{24}$$

$$+ [(\bar{h} - \bar{h}^{(k)}, L^{dd'}f(X)) < (\bar{h}, L^{dd'}f(X)) \leqq 0].$$

If $h(X)$ is degenerate under P_θ and X is distributed as P_θ, then the $\mathbf{E} \times E_\theta$ integral of the right-hand side of (24) is zero. Also, if $h(X)$ is non-degenerate under $P_\theta(r_\theta > 0)$, but $L^{dd'}f(x)W_\theta = 0$ for fixed $X = x$, then the right-hand side of (24) is zero at x since $N(\bar{h} - \bar{h}^{(k)}, L^{dd'}f(X)) = (Z_\theta(X_k) - Z_\theta(X), L^{dd'}f(X)W_\theta)$. Omitting these degenerate cases, we shall bound the right-hand side of (24) by a Berry-Esseen normal approximation argument.

Specifically, assume $r_\theta > 0$, $Np_\theta(\boldsymbol{\theta}) > 1$ and fix $X = x$ such that $L^{dd'}f(x)W_\theta \neq 0$. Define $g(x) = \|L^{dd'}f(x)W_\theta\|^{-1} L^{dd'}f(x)W_\theta$ and note that $\|g(x)\| = 1$. Next, fix $X_k = x_k$ and $X_\nu = x_\nu$, $\nu \not\varepsilon I_\theta$. We observe that for the right-hand side of (24) not to vanish, the sum

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

of $Np_\theta(\boldsymbol{\theta}) - 1 > 0$ terms must fall into an interval of length

$$\|L^{dd'}f(x)W_\theta\|^{-1} |(\bar{h} - \bar{h}^{(k)}, L^{dd'}f(x))| = |(Z_\theta(x_k) - Z_\theta(x), g(x))|.$$

But the terms $(Z_\theta(X_\nu), g(x))$ are independent and identically distributed with mean zero and variance 1 (by (22)) and hence the Berry-Esseen result of Lemma 2 bounds the probability of the above event by

$$\{Np_\theta(\boldsymbol{\theta}) - 1\}^{-\frac{1}{2}}\{(2\pi)^{-\frac{1}{2}} |(Z_\theta(x_k) - Z_\theta(x), g(x))| + 2\beta E_\theta|(Z_\theta(X), g(x))|^3\}' \tag{25}$$

where the expectation is on X in the second term. Since (23) implies

$$E_{\theta_k}E_\theta |(Z_\theta(X_k) - Z_\theta(X), g(X))| \leqq E_{\theta_k}E_\theta \|Z_\theta(X_k) - Z_\theta(X)\|$$

$$\leqq 2\{E_\theta \|Z_\theta(X)\|^2\}^{\frac{1}{2}} = 2r_\theta^{\frac{1}{2}},$$

we see that if $r_\theta > 0$ and $Np_\theta(\boldsymbol{\theta}) > 1$, (24) and (25) yield

$$\mathbf{E}E_{\theta_k}\{t'_{\bar{h}^{(k)},d}(X)t'_{\bar{h},d'}(X) + t'_{\bar{h}^{(k)},d'}(X)t'_{\bar{h},d}(X)\} \leqq \{Np_\theta(\boldsymbol{\theta}) - 1\}^{-\frac{1}{2}}C_\theta^*,$$

where $C_\theta^* = (2r_\theta\pi^{-1})^{\frac{1}{2}} + 2\beta E_\theta \|Z_\theta(X)\|^3$ with $E_\theta \|Z_\theta(X)\|^3$ being finite since $Z_\theta(X)$ is a finite linear combination of random variables whose third absolute moments are finite under P_θ.

Since the right-hand side of (24) is zero if h is degenerate under P_θ and is always bounded by unity, we may trivially include the cases $r_\theta = 0$ and $Np_\theta(\boldsymbol{\theta}) \leqq 1$ by rewriting the above as

$$\mathbf{E}E_{\theta_k}\{t'_{\bar{h}^{(k)},d}(X)t'_{\bar{h},d'}(X) + t'_{\bar{h}^{(k)},d'}(X)t'_{\bar{h},d}(X)\} \tag{26}$$

$$\leqq \min\{1, |Np_\theta(\boldsymbol{\theta}) - 1|^{-\frac{1}{2}} C_\theta'\},$$

where $C_\theta' = \max\{1, C_\theta^*\}$.

Next observe that with $p_\theta = p_\theta(\boldsymbol{\theta})$ and $C' = (C_1', \cdots, C_m')$, we have

$$\begin{aligned} &\textstyle\sum_\theta Np_\theta \min\{1, |Np_\theta - 1|^{-\frac{1}{2}} C_\theta'\} \\ &\qquad \leqq \textstyle\sum_{\{\theta|Np_\theta \geqq 1\}} (Np_\theta - 1)\min\{1, |Np_\theta - 1|^{-\frac{1}{2}} C_\theta'\} + m \\ &\qquad \leqq \textstyle\sum_{\{\theta|Np_\theta \geqq 1\}} (Np_\theta - 1)^{\frac{1}{2}} C_\theta' + m \\ &\qquad \leqq \{\textstyle\sum_{\{\theta|Np_\theta \geqq 1\}} (Np_\theta - 1)\}^{\frac{1}{2}} \|C'\| + m \\ &\qquad \leqq N^{\frac{1}{2}} \|C'\| + m. \end{aligned} \tag{27}$$

Finally, noting that

$$B_N \leqq N^{-1} \textstyle\sum_{k=1}^N \sum_{d<d'} |L_{\theta_k}^{dd'}| \, \mathrm{E}E_{\theta_k}\{t'_{\bar{h}(k),d}(X)t'_{\bar{h},d'}(X) + t'_{\bar{h}(k),d'}(X)t'_{\bar{h},d}(X)\},$$

we see that (26) and (27) yield

$$N^{\frac{1}{2}}B_N \leqq \binom{n}{2} L(mN^{-\frac{1}{2}} + \|C'\|), \tag{28}$$

where $L = \max_{\theta,d,d'} |L_\theta^{dd'}|$.

Equation (28) implies (ii), which together with (i) and inequality (18) completes the proof.

6. A theorem of higher order.

In Hannan and Van Ryzin [6], Theorems 3 and 2, it is shown for the case $m = n = 2$, $L(1, 1) = L(2, 2) = 0$, $L(1, 2) = b > 0$, and $L(2, 1) = a > 0$, that uniform upper bounds of $o(N^{-\frac{1}{2}})$ and $O(N^{-1})$ respectively are obtained for the regret function under suitable continuity assumptions on the induced distribution of a certain function of the likelihood ratio of the pair of densities. Hence, it is a natural question to ask whether one can do better than $O(N^{-\frac{1}{2}})$ in the more general setting. The answer is no for general $m \times n$ loss matrices. In [12], the author gives two examples of loss matrices for which $O(N^{-\frac{1}{2}})$ is the *best* obtainable rate. These matrices are:

$$\begin{pmatrix} 0 & 2 \\ 2 & 0 \\ 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 1 \\ 1 & 7 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

The details of these examples can be found in Section 3, Chapter III of [12]. However, consider the following condition which is violated by the above two examples.

(C) For two distinct columns d, d' of the $m \times n$ loss matrix $(L(\theta, d))$, let $I_{dd'} = \{\theta \,|L(\theta, d) = L(\theta, d')\}$. For any d, d' and $\theta \, \varepsilon \, I_{dd'}$ there exists a $j = j(d, d', \theta)$ such that $L(\theta, d) > L(\theta, j)$ and $L(\theta', d) \geqq L(\theta', j)$ for all $\theta' \, \varepsilon \, I_{dd'}$.

The meaning of this condition is that in any subgame obtained by deleting rows from the $m \times n$ loss matrix, there remains a column strategy which is strictly preferable in the subgame to any pair of columns which are identical in the subgame.

Three important cases in which (C) is satisfied are concerned with the discrimination problem in which $m = n$, $L(\theta, d) = 0$ or >0 according as $\theta = d$ or $\theta \neq d$. These three cases are:

(i) Let $m = 2$ or 3. The case $m = 2$ is treated in detail in [5] and [6].

(ii) Define $L(\theta, d) = a(1 - \delta_{\theta d})$, where $\delta_{\theta d}$ is the Kronecker δ. Condition (C) is satisfied by choosing $j(d, d', \theta) = \theta$.

(iii) Let $w(t)$ be a strictly increasing function on $[0, \infty)$ with $w(0) = 0$. Define $L(\theta, d) = w(|\theta - d|)$. Since $L(\theta, d) = L(\theta, d')$ for $d \neq d'$ implies $d < \theta < d'$ or $d' < \theta < d$, condition (C) is satisfied by choosing $j(d, d', \theta) = \theta$.

Now under condition (C), we can give a continuity assumption (C′) on the family of distributions $\{P_1, \cdots, P_m\}$ under which Theorem 3 of $O(N^{-1})$ given below is valid. The condition is:

(C′) Let P_θ^* be the probability measure on E^m induced under P_θ by the measurable transformation $x \rightarrow f(x)$. The measure P_θ^* is absolutely continuous with respect to m-dimensional Lebesgue measure λ_m and $dP_\theta^*/d\lambda_m \leqq K$ for some $K < \infty$.

Assumption (C′) is closely related to condition (II) of Theorem 2 in [6]. For an exact discussion of this and two related conditions (one implied by (C′) and one equivalent to (II) of Theorem 2 in [6] when $m = 2$) see assumptions (II′) and (II″) of Section 3, Chapter III of [12] as well as Appendix 1 of [12].

We now state without proof the following theorem. For details of the proof see Theorem 6, Chapter III of [12].

THEOREM 3. *Let* (C) *and* (C′) *hold. If* $h \varepsilon \mathcal{E}$ *such that* $|h_j(x)| \leqq M$ a.e. μ *for* $j = 1, \cdots, m$ *and* $M < \infty$, *then* $R(\mathbf{\theta}, T') - \phi(p(\mathbf{\theta})) \leqq c^*N^{-1}$ *where* c^* *is independent of* $\mathbf{\theta} \varepsilon \mathbf{\Omega}$ *for all* N.

Observe that the assumption of bounded h's may always be satisfied in view of Corollary 2.

Assumption (C′) is rather unattractive and very stringent. Nonetheless, Theorem 3 (or Theorem 6 of [12]) is of interest in that it shows that a uniform bound of $O(N^{-1})$ is available in the more general $m \times n$ case as in the case $m = n = 2$ of Hannan and Van Ryzin [6], Theorem 2. That Theorem 3 is not vacuous is shown by the following example satisfying (C′).

EXAMPLE. Let $X = (X_1, \cdots, X_m)$ be the generic random variable for the component problem. Define for $\theta = 1, \cdots, m$, the probability measure P_θ having densities with respect to λ_m given by $f_\theta(x) = 2x_\theta$ if $x \varepsilon [0, 1]^m$, the unit m-cube. If we let $P_\theta^*(a_1, \cdots, a_m)$ be the cumulative distribution function corresponding to the induced probability measure P_θ^*, then $P_\theta^*(a_1, \cdots, a_m) = 2^{-(m+1)}(\prod_{i=1}^m a_i)a_\theta$ for $a \varepsilon [0, 2]^m$. Hence, P_θ^* is absolutely continuous with respect to λ^m and has λ_m-density $2^{-m}a_\theta$ on $[0, 2]^m$, which is bounded by 2^{-m+1} on $[0, 2]^m$. Therefore, (C′) is satisfied for this example.

7. Extension of results for a randomized procedure. We extend Theorems 2 and 3 to the non-simple, randomized procedure defined by substituting the estimate $\bar{h}$ for $p(\mathbf{\theta})$ in the simple randomized procedure which assigns equal probabilities of selection among all columns minimizing $(p(\mathbf{\theta}), L^d f)$ in (7). Such a randomized, non-simple rule is given by $N \times n$ matrix of functions $T^*(\mathbf{x}) = ({}^*t_d^{\,k}(\mathbf{x}))$, where for $d = 1, \cdots, n, k = 1, \cdots, N$,

$$
\begin{aligned}
{}^*t_d^{\,k}(\mathbf{x}) &= r^{-1} && \text{if } (\bar{h}, L^d f(x_k)) = \min_j (\bar{h}, L^j f(x_k)) \text{ and the minimum} \\
& && \text{is achieved by exactly } r \text{ indices} \\
&= 0 && \text{otherwise.}
\end{aligned} \tag{29}
$$

Theorems 2 and 3 hold for (29). To see this, let $\mathfrak{N}'$ be the class of all permutations $\pi = (\pi(1), \cdots, \pi(n))$ of the integers $\{1, 2, \cdots, n\}$. Let π' denote the identity permutation having $\pi'(j) = j, j = 1, \cdots, n$. Now define the following class of non-randomized rules T^π, $\pi \varepsilon \mathfrak{N}'$, given by the $N \times n$ functions

$$
\begin{aligned}
t^{\pi}_{\bar{h},d}(x_k) &= 1 && \text{if } \pi(d) \text{ is a minimum subject to the equality} \\
& && (\bar{h}, L^d f(x_k)) = \min_j (\bar{h}, L^j f(x_k)) \\
&= 0 && \text{otherwise.}
\end{aligned} \tag{30}
$$

Note that $T^{\pi'}(\mathbf{x}) = (t'_{\bar{h},d}(x_k))$ is that particular non-randomized, non-simple rule given by (12) for which Theorems 2 and 3 hold. It is clear by analogy with π' the rule T^π for each $\pi \varepsilon \mathfrak{N}'$ also satisfies Theorem 2 and 3. Hence, the rule

$$
(n!)^{-1} \textstyle\sum_{\pi \varepsilon \mathfrak{N}'} t_{\bar{h},d}(x_k), \qquad k = 1, \cdots, N, d = 1, \cdots, n, \tag{31}
$$

also satisfies Theorems 2 and 3. But with the aid of a simple combinatoric argument it is easy to show that (31) $= {}^*t_d^{\,k}(\mathbf{x})$ and thus Theorems 2 and 3 hold for the non-randomized, non-simple rule (29).

The author believes that Theorems 2 and 3 hold for any measurable, well-defined randomization in (12), but was unable to prove this.

8. Acknowledgments. I wish to thank Professor J. F. Hannan for suggesting this area of research and for his inspiring guidance throughout this investigation. I am particularly grateful to him for giving me the unpublished lecture notes [4] upon which Section 3 is based.

Thanks are also due to the referees for pointing out certain changes which greatly improved an earlier version of the manuscript.

REFERENCES

[1] Choi, Keewhan (1965). Strongly consistent estimates for finite mixtures of distribution functions. Unpublished.

[2] Hannan, James F. (1957). Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games,* **3** 97–139. Ann. of Math. Studies No. 39, Princeton Univ. Press.

[3] 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.

[4] Hannan, James F. (1957). Unpublished lecture notes. Michigan State Univ.

[5] 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.

[6] 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.

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

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

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

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

[11] TEICHER, HENRY (1963). Identifiability of finite mixtures. *Ann. Math. Statist.* **34** 1265–1269.

[12] VAN RYZIN, J. R. (1964). Asymptotic solutions to compound decision problems. Ph.D. thesis, Department of Statistics, Michigan State University. Also appears as an Argonne National Laboratory report, ANL-6820.