The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 102-112

EQUIVARIANT PROCEDURES IN THE COMPOUND DECISION PROBLEM WITH FINITE STATE COMPONENT PROBLEM[1]

By James Hannan and J. S. Huang[2]

Michigan State University and University of Waterloo

0. Summary. Let $(\mathscr{X}, \mathscr{B}, P)$ be a probability measure space for each $P \in \mathscr{P} = \{F_0, \cdots, F_m\}$, $\mathscr{A}$ be an action space and L be a loss function defined on $\mathscr{X} \times \mathscr{P} \times \mathscr{A}$ such that for each i,

$$c_i = \int_- \vee_a L(x, F_i, a)\, dF_i(x) < \infty .$$

In the compound problem, consisting of N components each with the above structure, we consider procedures equivariant under the permutation group. With

$$\rho_{ij} = \vee_{B \in \mathscr{B}} |F_i(B) - F_j(B)| \qquad \text{and} \qquad K(\rho) = .5012\cdots \rho(1-\rho)^{-\frac{3}{2}},$$

we show that the difference between the simple and the equivariant envelopes is bounded by

$$\text{(T1)} \qquad \{2K(\rho) \textstyle\sum_i c_i^2\}^{\frac{1}{2}} N^{-\frac{1}{2}} \qquad \text{where} \quad \rho = \vee_{i,j} \rho_{ij},$$

and by

$$\text{(T2)} \qquad 2^m\{2K(\rho') \textstyle\sum_i c_i^2\}^{\frac{1}{2}} N^{-\frac{1}{2}} \quad \text{where} \quad \rho' = \vee\{\rho_{ij} \mid \rho_{ij} < 1\}.$$

The bound (T1) is finite iff the F_i are pairwise non-orthogonal and (T2) is designed to replace it otherwise.

1. Notations and history. Let $(\mathscr{X}, \mathscr{B}, P)$ be a probability measure space for each $P \in \mathscr{P} = \{F_0, F_1, \cdots, F_m\}$, $\mathscr{A}$ be an action space, L be a loss function which is defined on $\mathscr{X} \times \mathscr{P} \times \mathscr{A}$ to the nonnegative reals with value variously expressed

$$L(x, F_i, a) = L(x, F_i)(a) = {}_xL_i(a) . \tag{1}$$

We assume that for each i, $\vee_a L_i(a)$ has finite lower integral with respect to F_i,

$$c_i = \int_- \vee_a L_i(a)\, dF_i < \infty . \tag{2}$$

Since the space $\mathscr{A}$ serves only as a parameter space for the class $\mathscr{L} = \{L(a) \mid a \in \mathscr{A}\}$ of loss functions on $\mathscr{X} \times \mathscr{P}$, it is without loss of generality to assume that $\mathscr{A}$ contains no duplicates in this sense. To avoid the notational buildup attendant on the introduction of randomization at this and higher

Received February 6, 1970.

[1] This research was partially supported by NSF Grant GP-14591.

[2] Now at University of Guelph.

levels, we assume that $\mathscr{A}$ is its own extension to the class of all probability measures on the σ-field of subsets of $\mathscr{A}$ generated by $\{L(x, P) \mid x \in \mathscr{X}, P \in \mathscr{P}\}$: given any such probability measure ξ,

$$\exists\, a_\xi \in \mathscr{A} \ni L(a_\xi) = \int L(a)\, d\xi(a)\,, \qquad \forall x, P. \tag{3}$$

a_ξ is unique by the assumption of no duplicates. We observe that for each x and P, L is linear in a. For $a_t = ta_1 + (1-t)a_0$ is in $\mathscr{A}$ if a_0 and a_1 are, and

$$L(x, P, a_t) = \int L(x, P, \cdot)\, da_t = tL(x, P, a_1) + (1-t)L(x, P, a_0)\,, \tag{4}$$

where the first equality follows from (3) and the second one follows from linearity of integrals.

Hereafter we shall write integrals in operator notation, e.g. the integral in (3) would be expressed $\xi[L(a)]$ or, and preferably, $\xi[L]$.

Let $\mathscr{D}$ be the family of all functions d on $\mathscr{X}$ to $\mathscr{A}$ such that $L_i \circ d$, the function which maps x to ${}_xL_i(d(x))$, is $\mathscr{B}$-measurable for each i. For $d \in \mathscr{D}$ we define the risk of d at F_i as the integral of $L_i \circ d$ with respect to F_i,

$$R(F_i, d) = F_i[L_i \circ d] \leqq c_i\,. \tag{5}$$

If G is a distribution on $\{0, \cdots, m\}$ we define the Bayes risk against G by

$$\psi(G) = \wedge_{\mathscr{D}}\, G[F_i[L_i \circ d]]\,. \tag{6}$$

We refer to the decision problem described above as the component problem. When N decision problems each with above generic structure are considered simultaneously, the resulting N-fold global problem is called a compound decision problem with finite state components.

Specifically, let $\mathbf{x} \in \mathscr{X}^N$, $(\mathfrak{X}, \mathfrak{B}) = (\mathscr{X}, \mathscr{B})^N$, $\mathbf{P} \in \mathfrak{P} = \mathscr{P}^N$, $\mathbf{a} \in \mathfrak{A} = \mathscr{A}^N$ and let $\mathfrak{D}$ be the family of all functions $\mathbf{d} = (d_1, \cdots, d_N)$ from $\mathfrak{X}$ to $\mathfrak{A}$ such that

$${}_{x_\alpha}L_i \circ d_\alpha(\mathbf{x}) \tag{7}$$

is $\mathfrak{B}$-measurable for all α and i. Letting $N_i = \#\{\alpha \mid P_\alpha = F_i\}$ and

$$W(\mathbf{x}, \mathbf{P}, \mathbf{a}) = N^{-1} \textstyle\sum_{\alpha=1}^{N} L(x_\alpha, P_\alpha, a_\alpha)\,, \tag{8}$$

we define the risk of $\mathbf{d} \in \mathfrak{D}$ at $\mathbf{P} \in \mathfrak{P}$ by

$$R(\mathbf{P}, \mathbf{d}) = \mathbf{P}[W(\mathbf{x}, \mathbf{P}, \mathbf{d}(\mathbf{x}))] \leqq N^{-1} \textstyle\sum_i N_i c_i\,. \tag{9}$$

$\mathbf{d} \in \mathfrak{D}$ is called a simple (sometimes, simple symmetric) procedure if $d_\alpha(\mathbf{x}) = d(x_\alpha)$ for all α, for some $d \in \mathscr{D}$. Let $\mathbf{S}$ be the class of all simple procedures and let $\mathbf{d} \in \mathbf{S}$ be denoted by d^N. It will follow directly from the definition of $\mathfrak{E}$ in Section 2 that $\mathbf{S} \subset \mathfrak{E}$, the subclass of $\mathfrak{D}$ equivariant under the permutation group. As functions of $\mathbf{P}$, $\wedge_{\mathbf{S}} R(\mathbf{P}, \mathbf{d})$ and $\wedge_{\mathfrak{E}} R(\mathbf{P}, \mathbf{d})$ will be called the simple envelope and the equivariant envelope, respectively. It is well known (cf. (27) ff.) that the former coincides with the component Bayes risk $\phi(\mathbf{N})$ with $\mathbf{N}$ denoting the empiric distribution of $P_1, \cdots, P_N$.

The compound decision problem was introduced by Robbins (1951). He argued that a bootstrap procedure which first estimates the empiric distribution of $P_1, \cdots, P_N$ and then plays Bayes against the estimate within each component may have its compound risk uniformly close to the simple envelope.

Hannan and Robbins (1955) considered 2×2 $\mathscr{P} \times \mathscr{A}$ and (Theorem 3) bounded the average loss of a bootstrap procedure by the sum of an error of estimation and a loss-weighted Glivenko-Cantelli measure of deviations of the empiric distribution of $x_1, \cdots, x_N$, thus obtaining strong convergence to zero uniformly in $\mathbf{P}$ of the difference of the average loss from the simple envelope for all correspondingly good estimators. Risk convergence (Theorem 4) followed as a corollary. Oaten (1969) permits loss dependence on x, replaces Bayes by any of a wide class of ε-Bayes and otherwise generalizes these results to $m \times n$ $\mathscr{P} \times \mathscr{A}$ (Theorem 1 and its Corollary) and to certain compact $\mathscr{P} \times$ compact $\mathscr{A}$ with L continuous for each x (Theorems 4 and 5). Under continuity and other restrictions on densities, analogues of generalizations to $m \times n$ $\mathscr{P} \times \mathscr{A}$ were given earlier by Suzuki (1966a).

Hannan and Robbins (1955) also introduced the class of equivariant procedures and showed (Theorem 5) that the difference between the simple and equivariant envelopes converges to zero uniformly in $\mathbf{P}$ as $N \uparrow \infty$. The proof depended heavily on a measure theoretic lemma specializing Theorem II.1 of Hannan (1953). Our Theorems 1 and 2 ((T1) and (T2) of our Summary) are a strengthened generalization of their result, with Theorem 1 correspondingly related to Theorem 3 of Hannan and Huang (1972) and Theorem 2 following as a somewhat involved corollary to Theorem 1.

Hannan and Van Ryzin (1965), for 2×2 $\mathscr{P} \times \mathscr{A}$, and Van Ryzin (1966), for $m \times n$ $\mathscr{P} \times \mathscr{A}$, have established a rate of $O(N^{-\frac{1}{2}})$ (and under additional restrictions on $\mathscr{P}$ and L, $O(N^{-1})$) for uniform risk convergence of bootstrap procedures based on estimators which are averages over $x_1, \cdots, x_N$ of a suitable kernel.

The importance of our results stems from the basic character of equivariant procedures in the compound problem. Until Oaten's (1969) ε-Bayes relaxation, all of the bootstrap procedures considered were essentially equivalent (cf. Lemma 3 of Oaten (1969)) to equivariant procedures. The equivariant envelope is then a clearly more appropriate yardstick of performance than the simple one. The results themselves have already been used by Oaten (1969), together with his afore-mentioned Theorem 1, to prove risk convergence for a wide class of equivariant uniformly-ε-Bayes procedures (Theorem 2).

As noted in Section 3 of Hannan and Huang (1972), a generalization of the underlying measure theoretic lemma, Theorem 2 of Horn (1968), turns out to be distinctly improved by an immediate extension of the afore-mentioned

Theorem II.1. Her corresponding result on the difference between the envelopes (Theorem 1) inherits the deficiencies of her Theorem 2 and only shows convergence to zero for each $\mathbf{P}$ with a stronger restriction on $\mathscr{P}$ than mutual absolute continuity, with $\mathscr{A}$ finite and with L constant with respect to x.

Considerable other work relates to equivariant procedures in the compound problem. Stein (1956) and James and Stein (1961) (cf. Stein (1966), where the heuristic of the procedure is revealed, and Cogburn (1965)) obtained strong results with a Gaussian squared-deviation-loss estimation component problem. Section 4 of Samuel (1967) (cf. Robbins (1962)) investigates a procedure Bayes against uniform prior on proportions for the simplest 2×2 example.

In an important general development, Cogburn (1967) imbeds the compound and empirical Bayes problems in a general theory of stringency. The analogues of his Lemma 3.2 and Theorem 3.3, which follow from his methods in our case, conclude increasing convergence of the equivariant envelope on the special sequences $\{k\mathbf{N}_0 \mid k = 1, 2, \cdots\}$ and, under the very strong assumption that min-max regret relative to the equivariant envelope converges to 0 as $N \uparrow \infty$, identify the limit as the simple envelope. This assumption is not always satisfied in our case and, indeed, is only known to obtain as a consequence of the corresponding result for the simple envelope (Theorem 3 of Oaten (1969) when the F_i are linearly independent), together with weakened forms of our present theorems.

2. Equivariant decision procedures and symmetrizations in a compound decision problem. Let $\mathscr{G}$ be the permutation group on N objects. The generic element $g \in \mathscr{G}$ will also be used to denote the transformation induced by g:

$$gy = (y_{g1}, \cdots, y_{gN}) . \tag{10}$$

Letting $g(\mathbf{B}) = \{g\mathbf{x} \mid \mathbf{x} \in \mathbf{B}\}$ for $\mathbf{B} \in \mathfrak{B}$, it follows from the transformation theorem (Theorem 39. C, Halmos (1950)) that for each $\mathbf{P} \in \mathfrak{P}$ and $g \in \mathscr{G}$, $g\mathbf{P}$ is in $\mathfrak{P}$ and satisfies

$$\mathbf{P}(\mathbf{B}) = (g\mathbf{P})(g\mathbf{B}) , \qquad \mathbf{B} \in \mathfrak{B}. \tag{11}$$

Furthermore, it follows from the definition of W in (8) that, for each $\mathbf{a} \in \mathfrak{A}$, $g\mathbf{a}$ is in $\mathfrak{A}$ and

$$W(\mathbf{x}, \mathbf{P}, \mathbf{a}) = W(g\mathbf{x}, g\mathbf{P}, g\mathbf{a}) . \tag{12}$$

Thus the compound decision problem is invariant under $\mathscr{G}$.

$\mathbf{d} \in \mathfrak{D}$ is equivariant (under $\mathscr{G}$) if for all $g \in \mathscr{G}$,

$$\mathbf{d}(g\mathbf{x}) = g\mathbf{d}(\mathbf{x}) . \tag{13}$$

Hannan and Robbins (1955) and Ferguson (1967) use the term invariant procedures instead of equivariant procedures. The latter was suggested by

Wijsman (1968) to describe functions which transform properly rather than "invariantly." In our further references we will presuppose this change has been made.

It follows directly from definition (13) that $\mathbf{d} \in \mathfrak{E}$ if and only if there exists a function γ on $\mathscr{X} \times \mathscr{X}^{N-1}$ to $\mathscr{A}$, symmetric on $\mathscr{X}^{N-1}$, and such that for all α

$$d_\alpha(\mathbf{x}) = \gamma(x_\alpha, \check{\mathbf{x}}_\alpha) , \tag{14}$$

where

$$\check{\mathbf{x}}_\alpha = (x_1, \cdots, x_{\alpha-1}, x_{\alpha+1}, \cdots, x_N) . \tag{15}$$

Equivalently, $\mathbf{d} \in \mathfrak{E}$ if and only if there exists a function δ on $\mathscr{X} \times \mathscr{X}^N$ to $\mathscr{A}$, symmetric on $\mathscr{X}^N$, such that for all α,

$$d_\alpha(\mathbf{x}) = \delta(x_\alpha, \mathbf{x}) . \tag{16}$$

It follows from the definition of $\mathbf{S}$ and $\mathfrak{E}$ that

$$\mathbf{S} \subseteq \mathfrak{E} . \tag{17}$$

For each $\mathbf{d} \in \mathfrak{D}$ and $g \in \mathscr{G}$ define $\mathbf{d}^g$, the g-conjugate of $\mathbf{d}$, by

$$\mathbf{d}^g(\mathbf{x}) = g^{-1}[\mathbf{d}(g\mathbf{x})] . \tag{18}$$

Thus $\mathbf{d} \in \mathfrak{E}$ if and only if $\mathbf{d} = \mathbf{d}^g$ for all g. Define the symmetrization $\mathbf{d}^*$ of $\mathbf{d} \in \mathfrak{D}$ as the average of its conjugates,

$$\mathbf{d}^* = (N!)^{-1} \textstyle\sum_{\mathscr{G}} \mathbf{d}^g . \tag{19}$$

It follows immediately that

$$\mathfrak{E} = \mathfrak{D}^* \equiv \{\mathbf{d}^* \mid \mathbf{d} \in \mathfrak{D}\} . \tag{20}$$

Corresponding results hold for subgroups of the permutation group and relate to certain extended (cf. Swain (1965), Johns (1967), Gilliland and Hannan (1969) where only the sequence version is considered) compound decision problems.

3. Representation of equivariant risk. As a basis for comparing the simple and equivariant envelopes, we obtain in this section a convenient representation of the risk function of equivariant procedures and relate to the Bayes risk against a certain uniform prior.

For each $\mathbf{d} \in \mathfrak{D}$ and $g \in \mathscr{G}$, it follows directly from definition (9), (12) and the transformation theorem that the risk of the g-conjugate is

$$\begin{aligned} R(\mathbf{P}, \mathbf{d}^g) &= \mathbf{P}[W(\mathbf{x}, \mathbf{P}, \mathbf{d}^g(\mathbf{x}))] \\ &= \mathbf{P}[W(g\mathbf{x}, g\mathbf{P}, \mathbf{d}(g\mathbf{x}))] \\ &= R(g\mathbf{P}, \mathbf{d}) . \end{aligned} \tag{21}$$

Averaging the above over $\mathscr{G}$ gives the risk of the symmetrization $\mathbf{d}^*$,

(22) $$R(\mathbf{P}, \mathbf{d}^*) = (N!)^{-1} \sum_g R(g\mathbf{P}, \mathbf{d}) .$$

Since $\mathbf{d} = \mathbf{d}^g$ for equivariant $\mathbf{d}$, (21) also implies

(23) $$R(\mathbf{P}, \mathbf{d}) = R(g\mathbf{P}, \mathbf{d}) , \qquad \mathbf{d} \in \mathfrak{E}.$$

Letting $N_i(\mathbf{P}) = \#\{\alpha \mid P_\alpha = F_i\}$, and letting $\mathbf{N}(\mathbf{P}) = (N_0(\mathbf{P}), \cdots, N_m(\mathbf{P}))$ be a convenient index of the **P**-orbit, we shall hereafter write $R(\mathbf{N}, \mathbf{d})$ for LHS (23) with $\mathbf{N} = \mathbf{N}(\mathbf{P})$, and denote the equivariant envelope by

(24) $$\tilde{\phi}(\mathbf{N}) = \wedge_{\mathfrak{E}} R(\mathbf{N}, \mathbf{d}) .$$

Hannan and Robbins (1955) consider the class $\mathfrak{R}$ of all $\mathbf{d} \in \mathfrak{D}$ satisfying the constant risk property (23). They show that $\tilde{\phi}(\mathbf{N})$ coincides with $\wedge_{\mathfrak{R}} R(\mathbf{N}, \mathbf{d})$, the "risk-invariant" envelope. We wish to remark that the class $\mathfrak{R}$ need not be considered separately because, for each $\mathbf{d} \in \mathfrak{R}$, its symmetrization $\mathbf{d}^*$ has the same risk according to (22).

For $\boldsymbol{\delta} \in \mathfrak{D}$ it follows from the definition of risk (9) that

$$\begin{aligned} NR(\mathbf{P}, \boldsymbol{\delta}) &= \textstyle\sum_\alpha \mathbf{P}[L(x_\alpha, P_\alpha) \circ \delta_\alpha(\mathbf{x})] \\ &= \textstyle\sum_i \sum_{\alpha \mid P_\alpha = F_i} (F_i \times \check{\mathbf{P}}_\alpha)[L_i \circ \delta_\alpha] , \end{aligned}$$

where F_i acts on x_α and $\check{\mathbf{P}}_\alpha \equiv (P_1, \cdots, P_{\alpha-1}, P_{\alpha+1}, \cdots, P_N)$ acts on $\check{\mathbf{x}}_\alpha$.

In particular, if $\boldsymbol{\delta} \in \mathfrak{E}$, then by (14) δ_α (and therefore $L_i \circ \delta_\alpha$) is symmetric in $\check{\mathbf{x}}_\alpha$, and, with $N_{ji} \equiv N_j - 1$ or N_j depending on $j = i$ or $j \neq i$ and $\sum$ denoting sum over i such that $N_i > 0$, we have the following representation of equivariant risk,

(25) $$NR(\mathbf{N}, \boldsymbol{\delta}) = \textstyle\sum N_i F_i \times_j F_j^{N_{ji}}[L_i \circ \delta_1] ,$$

where F_i acts on x_1 and $\times_j F_j^{N_{ji}}$ acts on $\check{\mathbf{x}}_1$. The order of the F_j in $\times_j F_j^{N_{ji}}$ is immaterial since the integrand is symmetric. Let μ be any measure dominating $\mathscr{P}$ and let $f_i = dF_i/d\mu$. Abbreviating $\sum N_i f_i \times_i F_j^{N_{ji}}[L_i \circ \delta_1]$ by $T(\delta_1)$ for $\boldsymbol{\delta} \in \mathfrak{E}$, (25) is expressible as

(26) $$NR(\mathbf{N}, \boldsymbol{\delta}) = \mu[T(\delta_1)] .$$

If $\boldsymbol{\delta} \in \mathbf{S}$, say $\boldsymbol{\delta} = d^N$, then δ_1 is a function of x_1 alone. Thus

(27) $$R(\mathbf{N}, d^N) = \sum_i \frac{N_i}{N} F_i[L_i \circ d] .$$

The infimum over $\mathscr{D}$ of LHS (27) is, by definition, the simple envelope. The infimum over $\mathscr{D}$ of RHS (27) is, by (6), $\phi(\mathbf{N}/N)$, the Bayes risk against the prior $\mathbf{N}/N$. (This is also the infimum over $\mathfrak{D}$ of the $(\mathbf{N}/N)^N$-weighted risk but we shall make no use of this interpretation). Hereafter we abbreviate $\phi(\mathbf{N}/N)$ by $\phi(\mathbf{N})$.

We now show that $\tilde{\phi}(\mathbf{N})$ is the Bayes risk in the compound problem against the uniform prior on the orbit indexed by **N**. Let $U_\mathbf{N}$ denote such a prior.

Applying the transformation theorem to the mapping $g \to g\mathbf{P}$, we obtain from (22) that

$$R(\mathbf{N}, \mathbf{d}^*) = \sum_{\mathbf{Q}} R(\mathbf{Q}, \mathbf{d}) U_{\mathbf{N}}(\mathbf{Q}) . \tag{28}$$

The infimum over $\mathfrak{D}$ of the LHS above is $\tilde{\psi}(\mathbf{N})$ by (20). The infimum of the RHS is, by definition, the Bayes risk against $U_{\mathbf{N}}$, say, $R(U_{\mathbf{N}})$. Thus

$$\tilde{\psi}(\mathbf{N}) = R(U_{\mathbf{N}}) , \tag{29}$$

and therefore $\mathbf{d} \in \mathfrak{D}$ is ε-Bayes against $U_{\mathbf{N}}$ if and only if

$$R(\mathbf{N}, \mathbf{d}^*) \leqq \tilde{\psi}(\mathbf{N}) + \varepsilon . \tag{30}$$

4. The difference between the two envelopes when $\mathscr{P}$ is pairwise non-orthogonal. In this and the following section we bound the difference between the simple and equivariant envelopes. Since $\bigvee_{B \in \mathscr{B}} |F_i(B) - F_j(B)| = 1$ when $F_i \perp F_j$, Theorem 1 is of interest only when $\mathscr{P}$ is pairwise non-orthogonal.

THEOREM 1. *Let* $\rho_{ij} = \bigvee_{B \in \mathscr{B}} |F_i(B) - F_j(B)|$, c_i *satisfy* (2), $K(\rho) = .5012\ldots \rho(1-\rho)^{-\frac{3}{2}}$ *and let* $\rho = \bigvee_{i,j} \rho_{ij}$. *Then*

$$\psi(\mathbf{N}) - \tilde{\psi}(\mathbf{N}) \leqq \{2K(\rho) \textstyle\sum_i c_i^2\}^{\frac{1}{2}} N^{-\frac{1}{2}} . \tag{31}$$

PROOF. For each $\mathbf{N}$ and each equivariant $\boldsymbol{\delta}$, we will construct a simple procedure d^N whose risk at $\mathbf{N}$ is close to the risk of $\boldsymbol{\delta}$ at $\mathbf{N}$. To bound the difference in risks we use Theorem 3 of Hannan and Huang (1972), renotated here for our application by the use of relation (14) of that paper:

For any positive integer N and any nonnegative integral partitions $\mathbf{N}$ *and* $\mathbf{N}'$ *of* N,

$$\begin{aligned} \bigvee\{\times_i F_i^{N_i}[\varphi] - \times_i F_i^{N'_i}[\varphi] \mid 0 \leqq \varphi = \varphi^* \leqq 1\}^2 \\ \leqq nK(\rho) \textstyle\sum_i \wedge_i^{-1}(N_i' - N_i)^2 , \end{aligned} \tag{32}$$

with $n = \#\{k \mid N_k \neq N_k'\} - 1$, $\wedge_i = (N_i \wedge N_i') + 1$ *for all* i, *and* $\rho = \bigvee\{\rho_{ij} \mid N_i \neq N_i', N_j \neq N_j'\}$.

For $\boldsymbol{\delta} \in \mathfrak{E}$ consider $R(\mathbf{N}, \boldsymbol{\delta})$ in the form (25). For given i and x_1, let

$$\varphi = \text{the } x_1\text{-section of } \frac{L_i \circ \delta_1}{\bigvee_a L_i(a)} . \tag{33}$$

From (2), φ takes values in [0, 1] and, from (14), φ is symmetric in $\check{\mathbf{x}}_1$. Applying (32) to the integrand with respect to F_i in RHS (25) for each i yields

$$\begin{aligned} \times_j F_j^{N_{ji}}[L_i \circ \delta_1] \geqq \times_j F_j^{N_{jJ}}[L_i \circ \delta_1] \\ - \{\textstyle\bigvee_a L_i(a)\}\{K(\rho_{iJ})(N_i^{-1} + N_J^{-1})\}^{\frac{1}{2}} , \end{aligned} \tag{34}$$

for any $J \in \{0, \cdots, m\}$. With J such that $N_J = \bigvee_j N_j$, we weaken the bound (34) by simultaneously replacing N_J^{-1} by N_i^{-1} and $K(\rho_{iJ})$ by $K(\rho)$. Taking upper integrals with respect to F_i and weighting by N_i, we thus obtain

$$NR(\mathbf{N}, \boldsymbol{\delta}) \geqq \sum N_i F_i \times_j F_j^{N_j J}[L_i \circ \delta_1] - (2K(\rho))^{\frac{1}{2}} \sum_i N_i^{\frac{1}{2}} c_i \,, \tag{35}$$

with c_i given by (2).

We now construct the simple procedure d^N. Since for each x_1 and i, ${}_{x_1}L_i \circ \delta(x_1, \cdot)$ is a symmetric function of $\check{\mathbf{x}}_1$, it follows from the transformation theorem that

$$\times_j F_j^{N_j J}[{}_{x_1}L_i \circ \delta_1(x_1, \cdot)] = \xi[{}_{x_1}L_i(\cdot)] \,, \tag{36}$$

with $\xi = (\times_j F_j^{N_j J})[\delta_1(x_1, \cdot)]^{-1}$. We note that ξ depends on x_1 but not on i. By assumption (3), there exists $a_{x_1} \in \mathscr{A}$ such that RHS (36) $= {}_{x_1}L_i(a_{x_1})$, $\forall i$. Letting d be the function mapping x_1 to such a_{x_1}, we see that, for each x_1,

$${}_{x_1}L_i \circ d(x_1) = {}_{x_1}L_i(a_{x_1}) = \text{LHS } (36) \,, \tag{37}$$

which is $\mathscr{B}$-measurable. Therefore $d \in \mathscr{D}$, and d^N is simple. Also by (37) we recognize the first term of RHS (35) as N times RHS (27) which is bounded below by $N\psi(\mathbf{N})$. Thus

$$NR(\mathbf{N}, \boldsymbol{\delta}) \geqq N\psi(\mathbf{N}) - (2K(\rho))^{\frac{1}{2}} \sum_i N_i^{\frac{1}{2}} c_i \,. \tag{38}$$

Applying the Schwarz inequality to the sum on RHS (38) yields

$$R(\mathbf{N}, \boldsymbol{\delta}) \geqq \psi(\mathbf{N}) - (2K(\rho))^{\frac{1}{2}} N^{-\frac{1}{2}} (\textstyle\sum_i c_i^2)^{\frac{1}{2}} \,. \tag{39}$$

Since (39) holds for all $\boldsymbol{\delta} \in \mathfrak{E}$, this completes the proof of Theorem 1.

5. The difference between the two envelopes when $\mathscr{P}$ may have some pairwise orthogonality. In this section we derive, essentially as a corollary to Theorem 1, a useful bound for the difference when $\mathscr{P}$ may have some pairwise orthogonality.

THEOREM 2. *Let ρ_{ij}, c_i and K be as in theorem* 1 *and let $\rho = \vee\{\rho_{ij} \mid \rho_{ij} < 1\}$. Then*

$$\psi(\mathbf{N}) - \tilde{\psi}(\mathbf{N}) \leqq 2^m \{2K(\rho) \textstyle\sum_i c_i^2\}^{\frac{1}{2}} N^{-\frac{1}{2}} \,. \tag{40}$$

PROOF. The plan of our proof is first to decompose the whole problem into pieces of sub-problems, each satisfying the pairwise non-orthogonality condition. For arbitrary $\boldsymbol{\delta} \in \mathfrak{E}$ and a special choice of d to be $(2^{m+1} - 1)$ ε-Bayes with respect to $\mathbf{N}$, we represent the difference in risks of d^N and $\boldsymbol{\delta}$ as the sum of differences of simple and equivariant risks in the sub-problems with the simple procedure being ε-Bayes against the restriction of $\mathbf{N}$ to the sub-problems.

For each $I \subseteq \{0, \cdots, m\}$ let $\check{\mathscr{P}} = \{F_i \mid i \in I\}$ and $\check{N} = \sum_{i \in I} N_i$. The sub-problem determined by $\check{\mathscr{P}}^{\check{N}}$ will be called the I problem. Let $\check{\mathfrak{D}}$, $\check{\mathfrak{E}}$, $\check{\mathbf{N}}$, $\check{\psi}$, $\check{\tilde{\psi}}$, $\check{R}$, $\check{T}$ and $\check{*}$ denote the I problem counterpart of these symbols without the delete sign $\smile$. For simplicity we omit the delete sign on $\check{*}$ and $\check{R}$ hereafter.

Let μ be a measure dominating $\mathscr{P}$, $f_i = dF_i/d\mu$,

$$\mathscr{X}_I = \{x \in \mathscr{X} \mid f_i(x) > 0 \quad \text{iff} \quad i \in I\}, \tag{41}$$

and let μ_I be the restriction of μ to $\mathscr{X}_I$. Since $\mathscr{X} = \sum_I \mathscr{X}_I$ it follows that $\mu = \sum_I \mu_I$.

Let $\boldsymbol{\delta} \in \mathfrak{S}$. The risk of $\boldsymbol{\delta}$ is of form (25), which is expressible as integrals with respect to μ_I,

$$NR(\mathbf{N}, \boldsymbol{\delta}) = \textstyle\sum_I \mu_I\{\sum_{i \in I} N_i f_i \times_{j \in I} F_j^{N ji} \times_{j \notin I} F_j^{N j}[L_i \circ \delta_1]\}, \tag{42}$$

where we take $\times_{j \notin I} F_j^{Nj}$ to act on $(x_{\check{N}+1}, \cdots, x_N)$ for each I. For each I and each $(x_1, \cdots, x_{\check{N}})$ in RHS (42), there exists, for the same reason behind (36), a distribution ξ over $\mathscr{A}$ such that

$$\times_{j \notin I} F_j^{Nj}[L_i \circ \delta_1] = \xi[L_i], \qquad i \in I. \tag{43}$$

By assumption (3) there exists $a_\xi \in \mathscr{A}$ with $L_i(a_\xi) = \xi[L_i]$ for all $i \in I$. Letting $\check{\delta}_1$ be the function mapping $(x_1, \cdots, x_{\check{N}})$ to such a_ξ we see that $L_i \circ \check{\delta}_1$ gives RHS (43) and is thus $\mathscr{B}^{\check{N}}$-measurable for each $i \in I$. Furthermore, $\boldsymbol{\delta} \in \mathfrak{S}$ implies the symmetry of $\check{\delta}_1$ in $(x_2, \cdots, x_{\check{N}})$, and therefore $\check{\delta}_1$ is the first component of some $\check{\boldsymbol{\delta}} \in \check{\mathfrak{S}}$ constructable by the use of (14). Thus (42) yields the representation,

$$NR(\mathbf{N}, \boldsymbol{\delta}) = \textstyle\sum_I \mu_I[\check{T}(\check{\delta}_1)]. \tag{44}$$

For each I, let d_I be ε-Bayes in $\mathscr{D}$ against $\check{\mathbf{N}}/\check{N}$ and let $d = \sum_I \mathscr{X}_I d_I$ where $\mathscr{X}_I$ serves as the indicator function of itself. We note that $d \in \mathscr{D}$ and $d = d_I$ a.e. μ_I. Thus by (27), by $\mu = \sum \mu_I$ and by the fact that $L_i \circ d = L_i \circ d_I$ a.e. μ_I and is constant with respect to $(x_2, \cdots, x_{\check{N}})$, it follows that

$$NR(\mathbf{N}, d^N) = \textstyle\sum_I \mu_I[\check{T}(d_I)]. \tag{45}$$

The difference between (45) and (44) is

$$N\{R(\mathbf{N}, d^N) - R(\mathbf{N}, \boldsymbol{\delta})\} = \textstyle\sum_I \mu_I[\check{T}(d_I) - \check{T}(\check{\delta}_1)]. \tag{46}$$

For each I, define $\check{\mathbf{h}} = (\check{h}_1, \cdots, \check{h}_{\check{N}}) \in \check{\mathfrak{D}}$ by

$$\check{h}_\alpha(x_1, \cdots, x_{\check{N}}) = \begin{cases} \check{\delta}_\alpha(x_1, \cdots, x_{\check{N}}) & \text{if} \quad x_\alpha \in \mathscr{X}_I \\ d_I(x_\alpha) & \text{otherwise.} \end{cases} \tag{47}$$

By (14) we see that $\check{\mathbf{h}} \in \check{\mathfrak{S}}$. By direct calculation using (26) and the definition of $\check{\mathbf{h}}$,

$$\begin{aligned} \check{N}R(\check{\mathbf{N}}, \check{\mathbf{h}}) &= \mu_I[\check{T}(\check{\delta}_1)] + (\mu - \mu_I)[\check{T}(d_I)] \\ &= \check{N}R(\check{\mathbf{N}}, d_I^{\check{N}}) - \mu_I[\check{T}(d_I) - \check{T}(\check{\delta}_1)]. \end{aligned} \tag{48}$$

Since d_I is ε-Bayes with respect to $\check{\mathbf{N}}/\check{N}$ and $\check{\mathbf{h}} \in \check{\mathfrak{S}}$, (48) yields

(49) $$\mu_I[\check{T}(d_I) - \check{T}(\check{\delta}_1)] \leqq \check{N}(\check{\psi}(\check{\mathbf{N}}) - \check{\check{\psi}}(\check{\mathbf{N}})) + \check{N}\varepsilon\,.$$

It follows from (38) that, with $\check{\rho} = \vee\{\rho_{ij} \mid i, j \in I\}$,

(50) $$\text{RHS (49)} \leqq (2K(\check{\rho}))^{\frac{1}{2}} \sum_{i\in I} N_i^{\frac{1}{2}} c_i + \check{N}\varepsilon\,.$$

Summing (50) over all I with $\mu_I \neq 0$, we obtain an upper bound for RHS (46). Since $\mu_I \neq 0$ implies $\check{\rho} < 1$ we shall weaken (50) by replacing $\check{\rho}$ by ρ, and then dropping the restriction on the summand. Thus

(51) $$R(\mathbf{N}, d^N) - R(\mathbf{N}, \boldsymbol{\delta}) \leqq (2K(\rho))^{\frac{1}{2}} N^{-1} \sum_I \sum_{i\in I} N_i^{\frac{1}{2}} c_i + N^{-1}\varepsilon \sum_I \check{N}\,.$$

Since (51) holds for all $\boldsymbol{\delta} \in \mathfrak{E}$ and all $\varepsilon > 0$, and therefore for $\varepsilon = 0$, the proof is complete upon using the Schwarz inequality in (51):

(52) $$\sum_I \sum_{i\in I} N_i^{\frac{1}{2}} c_i = 2^m \sum_i N_i^{\frac{1}{2}} c_i \leqq 2^m N^{\frac{1}{2}} (\sum_i c_i^2)^{\frac{1}{2}}\,.$$

REFERENCES

COGBURN, ROBERT (1965). On the estimation of a multivariate location parameter with squared error loss in *Bernoulli* (1723), *Bayes* (1763), *Laplace* (1813), *Anniversary Volume*. Springer-Verlag, Berlin, 24–29.

COGBURN, ROBERT (1967). Stringent solutions to statistical decision problems. *Ann. Math. Statist.* **38** 447–464.

FERGUSON, THOMAS S. (1967). *Mathematical Statistics*. Academic Press, New York.

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

HALMOS, PAUL R. (1950). *Measure Theory*. Van Nostrand, Princeton.

HANNAN, JAMES F. (1953). Asymptotic solutions of compound decision problems. UNC Institute of Statistics Mimeo Series No. 68.

HANNAN, JAMES and HUANG, J.S. (1972). A stability of symmetrizations of product measures with few distinct factors. *Ann. Math. Statist.* **43** 308–319.

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.

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.

HORN, SUSAN DADAKIS (1968). The optimality criterion for compound decision problems. Tech. Report No. 10, Department of Statistics, Stanford University.

JAMES, W. and STEIN, CHARLES (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press, 361–379.

JOHNS, M.V., JR. (1967). Two-action compound decision problems. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press, **1** 463–478.

OATEN, ALLAN (1969). Approximation to Bayes risk in compound decision problems. RM-233, Department of Statistics and Probability, Michigan State Univ.

ROBBINS, HERBERT (1951). Asymptotically sub-minimax solutions of compound decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.*, Univ. of California Press, 131–148.

ROBBINS, HERBERT (1962). Some numerical results on a compound decision problem. *Recent Developments in Information and Decision Processes*. Macmillan, New York, 56–62.

SAMUEL, ESTER (1967). The compound statistical decision problem. *Sankhyā* **29** 123–140.

STEIN, CHARLES (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* Univ. of California Press, 197–206.

STEIN, CHARLES (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. *Research Papers in Statistics, Festschrift for J. Neyman.* Wiley, New York, 351–366.

SUZUKI, GIITIRO (1966a). Asymptotic solutions of the compound decision problem for many completely specified distributions. RM No. 1, The Institute of Statist. Math.

SUZUKI, GIITIRO (1966b). Discrete compound decision problem. *Ann. Inst. Statist. Math.* **18** 127–139.

SWAIN, DONALD D. (1965). Bounds and rates of convergence for the extended compound estimation problem in the sequence case. Tech. Report No. 81, Department of Statistics, Stanford Univ.

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

WIJSMAN, ROBERT A. (1968). Review of Thomas S. Ferguson "Mathematical Statistics." *Ann. Math. Statist.* **39** 2163–2167.