THE FINITE STATE COMPOUND DECISION PROBLEM,

EQUIVARIANCE AND RESTRICTED RISK COMPONENTS*

Dennis C. Gilliland and James Hannan

Michigan State University

The envelope results of Hannan and Huang (1972a) are generalized to arbitrary bounded risk components with simplified proofs. For equivariant "delete bootstrap" procedures, the excess compound risk over the simple envelope is bounded in terms of the L_1 error of estimation thus establishing a large class of asymptotic solutions to the compound decision problem with restricted risk component. This class includes the compound procedures which are Bayes versus certain symmetric priors (cf. Gilliland, Hannan and Huang, 1976). Asymptotic solutions and Bayes procedures in the empirical Bayes problem follow from those of the compound decision problem.

1. A finite state, restricted compound decision problem.

Consider a decision problem with states $P \in \mathcal{P} = \{F_0, F_1, \ldots, F_m\}$, where the F_i are distinct probability measures on $(\mathcal{X}, \mathcal{B})$. Let the risk set S of the decision problem be a bounded subset of $[0,\infty)^{m+1}$ and let $s = (s_0, s_1, \ldots, s_m)$ denote a generic point of S.

Consider a compound decision problem involving N independent repetitions of the above component structure. For $\underline{x} \in \mathcal{X}^N$ let $\check{\underline{x}}_\alpha$ denote $\underline{x}$ with the α-th component deleted and for $\underline{P} = \times_{\alpha=1}^N P_\alpha$ with $P_\alpha \in \mathcal{P}$, let $\check{\underline{P}}_\alpha$ denote $\underline{P}$ with the α-th factor deleted, α=1,2,...,N. Consider the class $\underline{S}$ of the compound rules $\underline{s} = (\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_N)$ where for each α, $\underline{s}_\alpha = (\underline{s}_{\alpha 0}, \underline{s}_{\alpha 1}, \ldots, \underline{s}_{\alpha m})$ denotes a

* Research supported by NSF Grant No. GP-33677X1.

AMS 1970 subject classification. Primary 62C25.

Key words & phrases. Compound decision problem, equivariance, equivariant envelope, simple envelope, restricted risk, empirical Bayes.

$\mathcal{B}^{N-1}$-measurable mapping into S. Here $\underline{s}_{\alpha i}(\check{\underline{x}}_\alpha)$ denotes the conditional on $\check{\underline{x}}_\alpha$ risk incurred by $\underline{s}$ in component α when $P_\alpha = F_i$. The (unconditional) component α risk of s at P with $P_\alpha = F_i$ is

$$(1) \qquad R_\alpha(\underline{P},\underline{s}) = \int \underline{s}_{\alpha i} d\underline{P}_\alpha .$$

The compound risk of s at P is defined to be

$$(2) \qquad \underline{R}(\underline{P},\underline{s}) = \sum_{\alpha=1}^{N} R_\alpha(\underline{P},\underline{s}).$$

For simplicity $\underline{R}$ is taken as the sum rather than the usual average.

When S is the largest possible risk set for a given component action space and loss function, the above problem is the usual compound problem with $\underline{S}$ being the largest class of compound decision rules. The above problem provides a setting in which there is control over the component risk behavior of compound rules by choice of S. This will be discussed more fully in Section 5.

The compound problem is invariant under the group of N! permutations of coordinates. A compound rule $\underline{s} \in \underline{S}$ is equivariant if and only if $\underline{s}_\alpha$ is constant with respect to α and a symmetric function of its argument $\check{\underline{x}}_\alpha$. Let $\underline{N} = (N_0, N_1, \ldots, N_m)$ where $N_i = \#\{\alpha | P_\alpha = F_i\}$ and let $N_{ji} = N_j$ if $j \neq i$, $N_{ji} = N_j - 1$ if $j = i$. The risk of equivariant $\underline{s}$ at $\underline{P}$ is a function of $\underline{P}$ through $\underline{N}$ and by (1) and (2) is given by

$$(3) \qquad \underline{R}(\underline{N},\underline{s}) = \sum_{i=0}^{m} N_i \int \underline{s}_{Ni} d\Big(\times_{j=0}^{m} F_j^{N_{ji}} \Big).$$

With $\underline{E}$ denoting the class of all equivariant rules in $\underline{S}$, the equivariant envelope is defined by the infimum

$$(4) \qquad \tilde{\psi}(N) = \bigwedge_{\underline{s} \in \underline{E}} \underline{R}(\underline{N},\underline{s}).$$

A rule $\underline{s} \in \underline{E}$ is said to be simple if $\underline{s}_N$ is a constant function; when $\underline{s}_N \equiv s$ we write $\underline{s} = s^N$. Specializing (3),

$$\underline{R}(\underline{N}, s^N) = \sum_{i=0}^{m} N_i s_i \tag{5}$$

which is also the component Bayes risk of s versus the empirical distribution of states $\underline{N}$. The simple envelope is defined by

$$\Psi(\underline{N}) = \bigwedge_{s \in S} \underline{R}(\underline{N}, s^N). \tag{6}$$

In Theorems 1 and 2 (Section 2) we show that $\Psi(\underline{N}) - \tilde{\Psi}(\underline{N}) = O(N^{1/2})$ uniformly in $\underline{N}$ if either P has no pairwise orthogonality or S satisfies a certain closure condition. The proofs are based on the Hannan and Huang (1972b) results on the stability of symmetrizations of product measures and are simpler than those given in their (1972a) paper for the envelope results and their compound problem. The results of the (1972a) paper are subsumed by the results of Section 2.

In Theorems 3 and 4 (Section 3) we develop bounds for $\underline{R}(\underline{N}, \underline{s}) - \Psi(\underline{N})$ for equivariant "delete bootstrap" rules. The bounds are used in Section 3 and the sequel Gilliland, Hannan and Huang (1976) to establish classes of such rules as asymptotic solutions to the compound decision problem.

In Remark 1 (Section 4) we observe that asymptotic solutions to our bounded risk compound decision problem are asymptotic average risk solutions to the empirical Bayes problem and that when equivariant they are asymptotic solutions to the empirical Bayes problem. In Remark 2 (Section 4) we show that the Bayes rules in the empirical Bayes problem are the Bayes rules versus an induced symmetric prior in the compound problem.

Finally, in Section 5 we show that our restricted compound decision problem subsumes the usual compound problem and give an example of a choice of S to control maximum component decision risk.

In what is to follow all sums Σ will be on i from 0 to m and all products $\times$ will be on j from 0 to m unless otherwise indicated.

2. Envelope results.

For each n and signed measure τ on $\mathcal{B}^n$ let τ^* denote the symmetrization of τ with respect to permutation of coordinates and $||\tau||$ denote the supremum $\vee\{|\tau(B)|:B \in \mathcal{B}^n\}$. Note that if g is a non-negative symmetric function whose τ-integral exists, then by the symmetry of g, $\int g d\tau = \int g d\tau^*$, so that, by the Jordan decomposition,

$$(7) \qquad |\int g d\tau| \le ||\tau^*|| \sup g.$$

THEOREM 1. Let $\rho_{ij} = ||F_i - F_j||$, $\rho = \vee\rho_{ij}$ and $K(y) = .5012\ldots y(1-y)^{-3/2}$. With $c_i = \vee\{s_i | s \in S\}$,

$$(8) \qquad \Psi(\underline{N}) - \tilde{\Psi}(\underline{N}) \le (2K(\rho))^{1/2}\Sigma\, c_i N_i^{1/2} \quad \text{for all } \underline{N}.$$

Proof. Let J be such that $N_J = \vee N_i$. By Theorem 3 of Hannan and Huang (1972b) with $\tau = \times F_j^{N_{ji}} - \times F_j^{N_{jJ}}$, $||\tau^*||^2 \le 2K(\rho_{iJ})N_i^{-1}$. Hence, for $\underline{s} \in \underline{E}$ it follows from (7) and the symmetry of $\underline{s}_{Ni}$ that

$$|\int \underline{s}_{Ni} d\tau| \le (2K(\rho_{iJ}))^{1/2} c_i N_i^{-1/2}$$

which together with (3) and the monotonicity of K on $[0,1]$ implies

$$(9) \qquad |\underline{R}(\underline{N},\underline{s}) - \int \Sigma\, N_i \underline{s}_{Ni}\, d(\times F_j^{N_{jJ}})| \le (2K(\rho))^{1/2}\Sigma\, c_i N_i^{1/2}.$$

Since $\Sigma N_i \underline{s}_{Ni} \ge \Psi(\underline{N})$, (9) implies

$$(10) \qquad \Psi(\underline{N}) - \underline{R}(\underline{N},\underline{s}) \le (2K(\rho))^{1/2}\Sigma\, c_i N_i^{1/2}.$$

Because $\underline{s}$ is an arbitrary element of $\underline{E}$, (8) is a consequence of (10) and the definition (4) of $\tilde{\Psi}$. Q.E.D.

The upper bound in (8) is $O(N^{1/2})$ uniformly in $\underline{N}$ provided $\rho < 1$. We now develop a useful bound in the case $\rho = 1$ for certain S by following the approach of Hannan and Huang (1972a, Section 5) to partition $\mathcal{X}$ relative to decision subproblems where Theorem 1 is useful.

For the rest of this section, suppose that T is a set of $\mathcal{B}$-measurable mappings $t = (t_0,\ldots,t_m)$ into $[0,\infty)^{m+1}$ and that S is the range of the function $t \to (\int t_0 dF_0,\ldots,\int t_m dF_m)$ defined on T. For the compound problem, let $\underline{T}$ be the set of $(\underline{t}_1,\underline{t}_2,\ldots,\underline{t}_N)$ where, for each α, $\underline{t}_\alpha = (\underline{t}_{\alpha 0},\ldots \underline{t}_{\alpha m})$ is a map into $[0,\infty)^{m+1}$ whose $\check{\underline{x}}_\alpha$-sections $\underline{t}_\alpha(\check{\underline{x}}_\alpha)$ belong to T and are such that the $\int t_{\alpha i}(\check{\underline{x}}_\alpha)dF_i$ are β^{N-1}-measurable. (As a consequence, given $\underline{s} \in \underline{S}$ there exists a $\underline{t} \in \underline{T}$ such that $\underline{s}_{\alpha i}(\check{\underline{x}}_\alpha) = \int \underline{t}_{\alpha i}(\check{\underline{x}}_\alpha)dF_i$ for all $i,\alpha,\underline{x}$.)
Let $f_i = dF_i/d\mu$, $i = 0,\ldots,m$ for some measure μ dominating $\mathcal{P}$ and consider the partition $\mathcal{X} = \Sigma_I \mathcal{X}_I$ where for each subset I of $\{0,1,\ldots,m\}$,

$$\chi_I = \bigcap_{i\in I} [f_i > 0] \bigcap_{i\notin I} [f_i = 0]. \tag{11}$$

We say that T is <u>support partition closed</u> if $t_I \in T$ for all I implies $\Sigma_I \chi_I t_I \in T$, where χ_I serves as its own indicator function.

THEOREM 2. With ρ_{ij} as defined in Theorem 1, let $\rho' = \vee\{\rho_{ij} | \rho_{ij} < 1\}$. With K and the c_i of Theorem 1 and T support partition closed,

$$\Psi(\underline{N}) - \tilde{\Psi}(\underline{N}) \leq 2^m (2K(\rho'))^{1/2} \Sigma c_i N_i^{1/2} \quad \text{for all } \underline{N}. \tag{12}$$

<u>Proof</u>. Let $\underline{s} \in \underline{E}$ and let $\underline{t} \in \underline{T}$ be such that $\underline{s}_{Ni}(\check{\underline{x}}_N) = \int \underline{t}_{Ni}(\check{\underline{x}}_N)dF_i$ for all i and $\check{\underline{x}}_N$. Partitioning $\chi = \Sigma_I \chi_I$ we can write

$$\underline{s}_{Ni} = \Sigma' \int \chi_I \underline{t}_{Ni} dF_i \tag{13}$$

where Σ' denotes the sum over I such that $\mu(\chi_I) > 0$. For each I let $t_I \in T$ and consider $t = \Sigma_I \chi_I t_I$ which belongs to T since T is support partition closed. Let $s_{Ii} = \int t_{Ii} dF_i$ and $s_i = \int t_i dF_i$, i=0,...,m and note that

$$(14) \qquad s_i = \Sigma' \int \chi_I t_{Ii} dF_i .$$

From (13) and (14) it follows that

$$(15) \qquad s_i - \underline{s}_{Ni} = \Sigma' \int \chi_I (t_{Ii} - \underline{t}_{Ni}) dF_i .$$

We can write $\int \chi_I (t_{Ii} - \underline{t}_{Ni}) dF_i$ as the difference of a simple and equivariant risk by noting that $\chi_I(t_{Ii} - \underline{t}_{Ni}) = t_{Ii} - \underline{t}'_{Ni}$ where $\underline{t}'_N = (1 - \chi_I)t_I + \chi_I \underline{t}_N$ has equivariant image since t_I and $\underline{t}_N$ have. With this substitution into (15) followed by integration with respect to $\times F_j^{N_{ji}}$ and N_i-weighted summation, we obtain

$$(16) \qquad \sum_i N_i s_i - \underline{R}(\underline{N},\underline{s}) = \Sigma N_i \int \Sigma' \int (t_{Ii} - \underline{t}'_{Ni}) dF_i d(\times F_j^{N_{ji}}).$$

But $\Sigma N_i s_i \geq \Psi(\underline{N})$, and by definition (11) of χ_I, $F_i(\chi_I) > 0$ only if $i \in I$, so that an interchange of the order of summation in the right hand side of (16) yields

$$(17) \qquad \Psi(\underline{N}) - \underline{R}(\underline{N},\underline{s}) \leq \Sigma' \sum_{i \in I} N_i \int (s_{Ii} - s'_{Ni}) d(\times F_j^{N_{ji}}).$$

For each I such that $\mu(\chi_I) > 0$, let $\check{\tilde{\Psi}}$ and $\check{\Psi}$ be the equivarient and simple envelopes respectively for the compound problem with P replaced by $\check{P} = \{F_i | i \in I\}$, which has no orthogonal elements. Also let $\check{N}_i = \#\{\alpha | P_\alpha = F_i\}$, $i \in I$. Because the non-I sections of the equivariant $\underline{s}'$ are equivariant in the sub-problem, it follows that

$$(18) \qquad \sum_{i \in I} N_i \int \underline{s}'_{Ni} d(\times F_j^{N_{ji}}) \geq \check{\tilde{\Psi}}(\check{\underline{N}})$$

and, since the s_I are arbitrary elements of S, (17) and (18) imply

$$\Psi(\underline{N}) - \underline{R}(\underline{N},\underline{s}) \leq \Sigma'\{\check{\Psi}(\check{\underline{N}}) - \tilde{\check{\Psi}}(\check{\underline{N}})\}. \tag{19}$$

The result (12) comes from applying Theorem 1 to each summand of the right hand side of (19), noting that for such I, $\check{\rho} \equiv \{||F_i - F_j||: i,j \in I\} \leq \rho'$, and using the monotonicity of K on [0,1] and the equality

$\sum_I \sum_{i\in I} c_i N_i^{1/2} = 2^m \Sigma c_i N_i^{1/2}$. Q.E.D.

Theorems 1 and 2 of Hannan and Huang (1972a) are corollaries to our Theorems 1 and 2 since their assumption (2) implies the boundedness of their risk set s, and since T induced by the class of all decision rules (cf. Section 5) is support partition closed. The Example of Section 5 shows that the conclusion of Theorem 2 does not follow from the boundedness of S alone.

3. Delete bootstrap rules.

For each $w = (w_0, w_1, \ldots, w_m) \in [0,\infty)^{m+1}$ the component Bayes risk of $s \in S$ versus w is given by

$$\Sigma\, w_i s_i . \tag{20}$$

In this section we assume that S is closed in R^{m+1} as well as bounded. Thus, for each $w \in [0,\infty)^{m+1}$ there exists a (Borel measurable in w) minimizer, say s^w, of (20) as s varies across S; and, when S is induced by a set T, we let t^w be such that $s_i^w = \int t_i^w \, dF_i$, $i = 0,1,\ldots,m$. (Hodges and Lehmann (1952) have used the term restricted Bayes to refer to Bayes procedures within classes S restricted by maximum risk.)

The assumption that S is closed is without essential loss of generality since, given any $\varepsilon > 0$ and $\underline{s}' \in \underline{S}'$ where S' is the R^{m+1} closure of S, there exist positive ε_α with $\sum_\alpha \varepsilon_\alpha < \varepsilon$ and $\underline{s} \in \underline{S}$ such that $|\underline{s}'_{\alpha i} - \underline{s}_{\alpha i}| \leq \varepsilon_\alpha$ for all i, thus ensuring that $|\underline{R}(\underline{P},\underline{s}') - \underline{R}(\underline{P},\underline{s})| < \varepsilon$ for all $\underline{P}$. Therefore, theorems concerning the risks of $\underline{S}'$ procedures have ε analogs for $\underline{S}$ procedures.

The relation

$$\Psi(\underline{N}) = \Sigma\, N_i s_i^{\underline{N}} \tag{21}$$

suggests that a compound rule $\underline{s}$ where for each α, $\underline{s}_\alpha = s^{\hat{\underline{N}}}$ with $\hat{\underline{N}} = \hat{\underline{N}}(\check{\underline{x}}_\alpha)$ an estimator of $\underline{N}$, might have compound risk asymptotically close to $\Psi(\underline{N})$ provided $\hat{\underline{N}}$ is consistent. Robbins (1951) suggested that playing Bayes versus an estimate $\hat{\underline{N}} = \hat{\underline{N}}(\underline{x})$ in each component might result in such compound risk behavior and used the term bootstrap to describe the effect. The use of x_α in both $\hat{\underline{N}}$ and the argument of the Bayes rule complicates the study of the risk. However, by deleting x_α from the data on which the estimate is based, we obtain a "delete bootstrap" rule in $\underline{S}$ whose risk behavior is easily studied. (The idea of deleting the component observation in the estimate of the empirical distribution of states (or functionals thereof) has been exploited in sequence compound problems, e.g., see Van Ryzin (1966b), Samuel (1965), Johns (1967), Gilliland (1968). At the game theoretic level it relates to play against the past strategies of Hannan (1957).) Formally we say that $\underline{s} \in \underline{S}$ is a delete bootstrap rule based on the estimators $\hat{\underline{N}}_1, \ldots \hat{\underline{N}}_N$, each a $\mathcal{B}^{N-1}$-measurable mapping into $[0,\infty)^{m+1}$, if $\underline{s}_\alpha = s^{\hat{\underline{N}}_\alpha}$ for all α and $\underline{x}$. When $\hat{\underline{N}}_\alpha = \underline{w}$, $\alpha=1,2,\ldots,N$, where $\underline{w} = (\underline{w}_0,\ldots,\underline{w}_m)$ is a symmetric function, the delete bootstrap rule, written $s^{\underline{w}}$, is equivariant. The next two theorems place useful bounds on the excess compound risk of $s^{\underline{w}}$, over the simple envelope.

THEOREM 3. With ρ, K and c_i as defined in Theorem 1 and J such that $N_J = \vee N_i$,

$$\underline{R}(\underline{N},s^{\underline{w}}) - \Psi(\underline{N}) \le \Sigma\, c_i \int |N_i - k\underline{w}_i|\, d(\times F_j^{N_jJ}) + (2K(\rho))^{1/2}\, \Sigma\, c_i N_i^{1/2} \tag{22}$$

for all $\underline{N}$ and every $\mathcal{B}^{N-1}$-measurable mapping k into $[0,\infty)$.

Proof. By (9)

$$\underline{R}(\underline{N},s^{\underline{w}}) \le \int \Sigma\, N_i s_i^{\underline{w}} d(\times F_j^{N_jJ}) + (2K(\rho))^{1/2}\, \Sigma\, c_i N_i^{1/2}. \tag{23}$$

From (21) and the defining properties of $\underline{s}^{\underline{w}}$, k and c_i it follows that

$$(24)\qquad \Sigma\, N_i s_i^{\underline{w}} - \Psi(\underline{N}) \leq \Sigma\, (N_i - k\underline{w}_i)(s_i^{\underline{w}} - s_i^{\underline{N}}) \leq \Sigma\, c_i |N_i - k\underline{w}_i|,$$

which together with (23) implies (22). Q.E.D.

THEOREM 4. With the hypothesis of Theorem 2 and Σ' denoting sum over I with $\mu(\chi_I) > 0$ and, for each such I, $\check{J}$ such that $N_{\check{J}} = \vee\{N_i \mid i \in I\}$,

$$(25)\ \underline{R}(\underline{N},\underline{s}^{\underline{w}}) - \Psi(\underline{N}) \leq \Sigma' \sum_{i\in I} c_i \int |N_i - k\underline{w}_i|\, d(\times\, F_j^{N_{j\check{J}}}) + 2^m(2K(\rho'))^{1/2}\, \Sigma\, c_i N_i^{1/2}$$

for all $\underline{N}$ and every $\mathcal{B}^{N-1}$ -measurable mapping k into $[0,\infty)$.

Proof. Since $\underline{s}^{\underline{w}} \in \underline{E}$, it follows from (3), (13) and the fact that $\mu(\chi_I) > 0$ if and only if $F_i(\chi_I) > 0$ for all $i \in I$, that

$$(26)\qquad \underline{R}(\underline{N},\underline{s}^{\underline{w}}) = \Sigma' \int [\sum_{i\in I} N_i \int\int \chi_I t_i^{\underline{w}}\, dF_i\, d(\mathop{\times}_{j\in I} F_j^{N_{ji}})]\, d(\mathop{\times}_{j\notin I} F_j^{N_j}).$$

Let D denote the difference between (26) and what is obtained by changing the N_{ji} to $N_{j\check{J}}$ for each I. Since the non-I sections of $\int \chi_I t_i^{\underline{w}}\, dF_i$ are symmetric and each I of the summation Σ' results in $\check{P}$ with $\check{\rho} \leq \rho'$, the technique leading to (9) shows that

$$(27)\qquad D \leq (2K(\rho'))^{1/2} \Sigma' \sum_{i\in I} c_i N_i^{1/2} \leq 2^m(2K(\rho'))^{1/2} \Sigma\, c_i N_i^{1/2}.$$

From the representation,

$$(28)\qquad \Psi(\underline{N}) = \Sigma' \sum_{i\in I} N_i \int \chi_I t_i^{\underline{N}}\, dF_i,$$

the definition of D and (26), $\underline{R}(\underline{N}, \underline{s}^{\underline{w}}) - \Psi(\underline{N}) - D$ is

$$\Sigma'\int \sum_{i\in I} N_i \int \chi_I (t_i^{\underline{w}} - t_i^{\underline{N}}) dF_i \; d(\times F_j^{N_{jJ}^{\vee}}). \tag{29}$$

Since T is support partition closed, it follows that for each $w \in [0,\infty)^{m+1}$,

$$\Sigma w_i s_i^{\underline{w}} = \Sigma w_i \int t_i^{\underline{w}} dF_i = \Sigma' \sum_{i\in I} w_i \int \chi_I t_i^{\underline{w}} dF_i = \Sigma' \wedge \{ \sum_{i\in I} w_i \int \chi_I t_i \; dF_i \,|\, t \in T\} \tag{30}$$

so that for each I,

$$k \sum_{i\in I} \underline{w}_i \int \chi_I (t_i^{\underline{N}} - t_i^{\underline{w}}) \; dF_i \geq 0. \tag{31}$$

Adding the Σ' summation of the integral of (31) with respect to $\times F_j^{N_{jJ}^{\vee}}$ to (29) we obtain

$$\Sigma'\int \sum_{i\in I} (N_i - k\underline{w}_i) \int \chi_I (t_i^{\underline{w}} - t_i^{\underline{N}}) dF_i \; d(\times F_j^{N_{jJ}^{\vee}}). \tag{32}$$

Since $|\int \chi_I (t_i^{\underline{w}} - t_i^{\underline{N}}) dF_i| \leq c_i$, we see that (25) is a consequence of the bound (32) and the inequality (27). Q.E.D.

The Example of Section 5 shows that the conclusion of Theorem 4 does not follow from the compactness of S alone.

Theorems 1 and 2 show that if S is bounded, and, in addition, T is support partition closed if $\mathcal{P}$ has pairwise orthogonality, then $\Psi(\underline{N}) - \tilde{\Psi}(\underline{N}) = O(N^{1/2})$ uniformly in $\underline{N}$. Theorems 3 and 4 show that if S is compact, and, in addition, T is support partition closed if $\mathcal{P}$ has pairwise orthogonality, then $\underline{R}(\underline{N},\underline{s}^{\underline{w}}) - \Psi(\underline{N})$ is bounded in (22) and (25) by estimation errors of $k\underline{w}$ for $\underline{N}$ and terms which are $O(N^{1/2})$ uniformly in $\underline{N}$. Since $|N_{iJ} - N_i|$, $|N_{iJ}^{\vee} - N_i| \leq 1$ for all i and I, the estimation errors are seen to be $O(N^{1/2})$ uniform in $\underline{N}$ provided $k\underline{w}$ is L_1 consistent for the empirical distribution of states governing the distribution of <u>its</u> argument $\underline{\check{x}}_N$ at a rate $O(N^{1/2})$ uniform in $\underline{N}$. Two important classes of symmetric estimators $k\underline{w}$ which achieve this rate of consistency are given below. For purposes of investigating consistency it is a notational convenience in the first example to let k and $\underline{w}$ be defined on χ^N rather than χ^{N-1}.

Kernel-type estimators. For linearly independent F_i and $\mu = \Sigma F_i$, dual bases in $L_2(\mu)$ furnish bounded $\mathcal{B}$-measurable mappings $h = (h_0,\ldots,h_m)$ such that for all i,j, $\int h_j dF_i = 1$ if $i = j$ and 0 otherwise (cf. Robbins, 1964, Section 7 and Van Ryzin, 1966a, Section 3). For such h, $\underline{h}(\underline{x}) = \sum_{\alpha=1}^{N} h(x_\alpha)$ satisfies $\int \underline{h}_i d(\times F_j^{N_j}) = N_i$ for all i and $\underline{N}$. If $\underline{w}$ is the retraction of $\underline{h}$ to $[0,\infty)^{m+1}$ and $k \equiv 1$, then for each i,

$$\int |N_i - k\underline{w}_i| d(\times F_j^{N_j}) \leq (\mathrm{Var}\ \underline{h}_i)^{1/2} = O(N^{1/2}) \text{ uniformly in } \underline{N}.$$

Estimators induced by priors. A second class of delete bootstrap rules consists of rules that are Bayes with respect to symmetric priors. To develop this important class of rules we need some additional notation. For each $\underline{K} = (K_0,K_1,\ldots,K_m)$ with nonnegative integers K_i, $\Sigma K_i = N$, let $[\underline{K}] \equiv \{\underline{P}|\underline{N} = \underline{K}\}$ and let $\binom{N}{\underline{K}}$ denote the multinomial coefficient $N!/(K_0!\ldots K_m!)$. A prior distribution $\underline{\beta}$ on the $(m+1)^N$ states $\underline{P}$ of the compound problem is symmetric if it distributes its mass uniformly within each orbit $[\underline{K}]$. Thus, a symmetric prior $\underline{\beta}$ can be identified with a distribution $\beta_{\underline{K}}$ on the $\binom{N+m}{m}$ different $\underline{K}$. By definition the Bayes component risk of $\underline{s} \in \underline{S}$ versus a symmetric $\underline{\beta}$ is

$$R_\alpha(\underline{\beta},\underline{s}) \equiv \sum_{\underline{K}} \beta_{\underline{K}} \binom{N}{\underline{K}}^{-1} \sum_{\underline{P}\in[\underline{K}]} R_\alpha(\underline{P},\underline{s}). \tag{33}$$

By (1) the inner summation is equal to

$$\sum_{\underline{P}\in[\underline{K}]} \sum_i \int \underline{s}_{\alpha i} d\check{\underline{P}}_\alpha \tag{34}$$

where $[\underline{K}]_i = \{\underline{P} \in [\underline{K}]|P_\alpha = F_i\}$. Let $\check{\underline{K}}_i = (K_{0i},\ldots,K_{mi})$ and note that, with $\times F_j^{K_{ji}}$ interpreted as 0 if $K_i = 0$, the inner sum of (34) is

$$\binom{N-1}{\check{\underline{K}}_i} \int \underline{s}_{\alpha i} d(\times F_j^{K_{ji}})^*, \tag{35}$$

where, as introduced in Section 2, * denotes symmetrization. Letting $(\times\, f_j^{K_{ji}})^*$ denote the symmetrization of $d(\times\, F_j^{K_{ji}})/d\mu^{N-1}$, (33) - (35) combine to give

$$\text{(36)}\qquad NR_\alpha(\underline{\beta},\underline{s}) = \sum_{\underline{K}} \beta_{\underline{K}} \sum K_i \int \underline{s}_{\alpha i}(\times\, f_j^{K_{ji}})^* d\mu^{N-1} = \int (\Sigma \underline{w}_i \underline{s}_{\alpha i}) d\mu^{N-1}$$

where

$$\text{(37)}\qquad \underline{w}_i = \sum_{\underline{K}} \beta_{\underline{K}} K_i (\times\, f_j^{K_{ji}})^*, \quad i=0,\ldots,m.$$

Thus (36) is minimized by taking $\underline{s}_\alpha = s^{\underline{w}}$ and therefore the equivariant delete bootstrap rule $\underline{s}^{\underline{w}}$ is Bayes versus $\underline{\beta}$.

For normalized estimators (37) induced by sufficiently diffuse symmetric priors in the two state problem, Gilliland, Hannan and Huang (1976) establish rates of L_1 consistency.

4. Empirical Bayes problem.

Results in compound decision theory have corollaries in the empirical Bayes theory of Robbins (1949) (also in the experience theory of Fabian and Spacek (1956)). If $P_1, P_2, \ldots$ are i.i.d. according to a probability distribution $\omega = (\omega_0, \ldots, \omega_m)$ on $\mathcal{P} = \{F_0, \ldots, F_m\}$, then component and compound risks are ω^N averages of (1) and (2), i.e. Bayes risks versus the symmetric prior with $\beta_{\underline{K}} = \binom{N}{\underline{K}} \omega_0^{K_0} \cdots \omega_m^{K_m}$. With this prior and $f_\omega \equiv \Sigma\, \omega_i f_i$, the induced estimator (37) is $\underline{w}_i = N\omega_i f_\omega^{N-1}$ and (36) is seen to be minimized by $\underline{s}_\alpha = s^\omega$. Thus, for each $\underline{s} \in \underline{S}$ and α

$$\text{(38)}\qquad R_\alpha(\omega^N, \underline{s}) \geq \Psi(\omega) \text{ for all } \omega.$$

Remark 1. If $\underline{s}$ is a compound rule which satisfies

$$\text{(39)}\qquad \underline{R}(\underline{P},\underline{s}) - \Psi(\underline{N}) = o(N) \text{ for all } \underline{P},$$

then

$$N^{-1}\underline{R}(\underline{\omega}^N,\underline{s}) \to \Psi(\omega) \text{ for all } \omega, \tag{40}$$

and if $\underline{s}$ is equivariant,

$$R_N(\omega^N,\underline{s}) \to \Psi(\omega) \text{ for all } \omega. \tag{41}$$

Proof. By (38) and (39) it follows that

$$\Psi(\omega) \le N^{-1}\underline{R}(\omega^N,\underline{s}) = \int [N^{-1}\Psi(\underline{N}) + o(1)]\, d\omega^N. \tag{42}$$

By Jensen's inequality and the bounded convergence theorem, RHS (42) is no greater than $\Psi(\omega) + o(1)$. If $\underline{s}$ is equivariant, then $R_\alpha(\omega^N,\underline{s})$ is constant with respect to α in which case (40) and (41) are equivalent. Q.E.D.

Hannan (Appendix, 1957) establishes (40) with rate for certain play against the past sequence strategies while Gilliland ((1.7), 1968) shows that (39) implies (40) rather generally in the sequence compound problem.

Remark 2. With $\beta_{\underline{K}}(\Lambda) = \int \binom{N}{\underline{K}}\omega_0^{K_0}\cdots\omega_m^{K_m}d\Lambda$, it follows from interchanging integration with respect to Λ and $\Sigma_{\underline{K}}$ in (33) that $\int R_\alpha(\omega^N,\underline{s})d\Lambda = R_\alpha(\underline{\beta}(\Lambda),\underline{s})$. Thus, Bayes rules versus Λ in the empirical Bayes problem are Bayes rules versus the symmetric prior $\underline{\beta}(\Lambda)$ in the compound problem.

5. Summary and final remarks.

The compound problem introduced in Section 1 and investigated in Sections 2, 3 and 4 subsumes the (finite state) compound decision problem of Hannan and Huang (1972a). There A is an action space L is a loss function on $\mathcal{X} \times \mathcal{P} \times \mathcal{A}$ to $[0,\infty)$, $\mathcal{D}$ is the set of all (decision) functions d on $\mathcal{X}$ to $\mathcal{A}$ such that the map $x \to {}_xL_i(d(x))$ is $\mathcal{B}$-measurable for all i, and $\underline{\mathcal{D}}$ is the set of all functions $\underline{d} = (d_1, d_2, \ldots, d_N)$ on $\mathcal{X}^N$ to $\mathcal{A}^N$ such that the map $\underline{x} \to {}_{x_\alpha}L_i(d_\alpha(\underline{x}))$ is $\mathcal{B}^N$-measurable for all i and α. With $t_i = L_i \circ d$, a class of decision rules $\mathcal{C} \subset \mathcal{D}$ determines a class $T = T(\mathcal{C})$ of $\mathcal{B}$-measurable mappings $t = (t_0, \ldots, t_m)$. Let $S = S(T)$ be the set of risk points determined by $s = (\int t_0 dF_0, \ldots, \int t_m dF_m)$ for $t \in T$. The the class $\underline{S}$ defined in Section 1 consists of the conditional component risks corresponding to the class of compound rules

$$\underline{\mathcal{C}} = \{\underline{d} \in \underline{\mathcal{D}} | \text{every } \underline{\check{x}}_\alpha\text{-section of } d_\alpha \text{ belongs to } \mathcal{C},\ \alpha=1,2,\ldots,N\}.$$

If $\mathcal{C} = \mathcal{D}$ then $\underline{\mathcal{C}} = \underline{\mathcal{D}}$ and the compound rules are not restricted.

The original motivation for the compound decision problem of this paper was the generality it provides for the envelope results (Section 2) and the fact that it is the natural setting in which to study delete bootstrap procedures (Section 3). Moreover, it allows for choice of S to control component risk behavior and the construction of asymptotically best equivariant procedures in $\underline{S}$.

As an example of a restricted class of some interest suppose c is some specified constant and consider $S = \{s \in S(T(\mathcal{D})) | s_i \leq c,\ i = 0, \ldots, m\}$. Compound rules $\underline{s}$ in $\underline{S}$ have, for each α and P_α, conditional on $\underline{\check{x}}_\alpha$ component risk no greater than c. In addition, since S is convex, $R_\alpha(\underline{P}, \underline{s}) \leq c$ for all $\underline{P}$.

Efron and Morris (1972) have proposed an interesting way to control component risks $R_\alpha(\underline{\theta}, \underline{d})$ for the compound problem with component squared error loss estimation of $\theta \in (-\infty, \infty)$ in $N(\theta, \sigma^2)$. For the known σ^2 case, wlog $\sigma^2 = 1$, they propose a compromise between the simple rule $(d^0)^N$ where $d^0(x) = x$, the component minimax estimator, and the James-Stein estimator $\underline{d}^1$, which is equivariant and possesses good compound risk behavior ($\underline{R}(\underline{\theta}, \underline{d}^1) < N$ for all $\underline{\theta}$ if $N \geq 3$). For $N \geq 3$ they study a class of equivariant "limited translation" estimators $\underline{d}^\rho$ where

$$\underline{d}_\alpha^\rho(\underline{x}) = (1-\rho) d^0(x_\alpha) + \rho\, \underline{d}_\alpha^1(\underline{x}),\ \alpha = 1,2,\ldots,N$$

and ρ is a function of $x_\alpha^2/\Sigma_1^N x_\alpha^2$ to $[0,1]$. For various choices ρ and $N > 3$ they compute maximum component risk $\vee\{R_\alpha(\underline{\theta},\underline{d}^\rho)|\underline{\theta}\}$ and compare $\underline{R}(\underline{\theta},\underline{d}^\rho)$ with $\underline{R}(\underline{\theta},\underline{d}^1)$.

We conclude with a simple example which illustrates that the conclusions of Theorems 2 and 4 do not follow from the compactness of S alone.

Example. Let the component decision problem be 0-1 loss classification with F_i degenerate at i, $i=0,1$ with $\chi = \{0,1\}$. Here there are but four non-randomized decision rules giving rise to risk points on the corners of a unit square. Consider the sub-class $C = \{d_0,d_1\}$ where d_i decides F_i regardless of the observed value x. Then $T = T(C) = \{t,t'\}$ where $t(x) \equiv (0,1) = s$ and $t'(x) \equiv (1,0) = s'$ and $S = S(T) = \{s,s'\}$. The set T is not support partition closed since, for example, $\chi_{\{0\}}t + \chi_{\{1\}}t'$ does not belong to T. Since the class of equivariant rules $\underline{E}$ includes the rules $\underline{s}^{(k)}$, $k=0,1,\ldots,N$ where

$$\underline{s}_N^{(k)}(\check{\underline{x}}_N) = s \text{ if } \sum_{\alpha=1}^{N-1} x_\alpha = k \text{ and } s' \text{ otherwise}$$

and since $\underline{R}(\underline{N},\underline{s}^{(N_1)}) = 0$, we see that the equivariant envelope $\tilde{\Psi}(N) \equiv 0$. However, the simple envelope is $\Psi(N) = N_0 \wedge N_1$ so that the conclusion of Theorem 2 fails to hold. Now let $\underline{w}_1(\check{\underline{x}}_N) = \Sigma_1^{N-1} x_\alpha$, $\underline{w}_0 = N - 1 - \underline{w}_1$. The delete bootstrap rule $s^{\underline{w}}$ with

$$s^{\underline{w}(\check{\underline{x}}_N)} = s \text{ if } \underline{w}_1(\check{\underline{x}}_N) < \tfrac{1}{2}(N-1) \text{ and } s' \text{ otherwise}$$

has $\underline{R}(\underline{N},s^{\underline{w}}) = N$ at $\underline{N}$ such that $N_0 = N_1$. At such $\underline{N}$, $\Psi(\underline{N}) = \frac{1}{2}N$ and the left hand side of (25) is equal to $\frac{1}{2}N$. Here $c_0 = c_1 = 1$, $\rho' = 0$ and $\underline{w}_i = N_{iJ}$ a.s. $\times F_j^{N_{jJ}}$ for all J and N so that with $k \equiv 1$, the right hand side of (25) is equal to 2 for all N.

Addendum. The thesis of Inglis [(1973). Admissible decision rules for the compound decision problem. TR 37, Dept. of Statist., Stanford] came to the authors' attention after the results of this paper were obtained. As detailed in Gilliland and Hannan (1974), the parts treating reduction to a consistency of the posterior mean seem to us to lack proof. (Our Theorems 3 and

4 and the last example of Section 3 do establish the reduction for a more general finite state component; their proofs do not extend to infinite state components.) For comment on the parts treating consistency of the posterior mean, see the Addendum to Gilliland, Hannan and Huang (1976).

REFERENCES

Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators - Part II: The empirical Bayes case. J. Amer. Statist. Assoc. **67** 130-139.

Fabian, V. and Spacek, A. (1956). Experience in statistical decision problems. Czechoslovak Math. Journal **6** 190-194, 434.

Gilliland, D.C. (1968). Sequential compound estimation. Ann. Math. Statist. **39** 1890-1904.

Gilliland, D.C. and Hannan, J. (1974). The finite state compound decision problem, equivariance and restricted risk components. RM-317, Dept. of Statist. and Prob., M.S.U.

Gilliland, D.C., Hannan, J. and Huang, J.S. (1976). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. Ann. Statist. **4** 1101-1112.

Hannan, J.F. (1957). Approximation to Bayes risk in repeated play. Contributions to the Theory of Games **3** 97-139.

Hannan, J. and Huang, J.S. (1972a). Equivariant procedures in the compound decision problem with finite state component problem. Ann. Math. Statist. **43** 102-112.

Hannan, J. and Huang, J.S. (1972b). A stability of symmetrizations of product measures with few distinct factors. Ann. Math. Statist. **43** 308-319.

Hodges, J.L. and Lehmann, E.L. (1952). The use of previous experience in reaching statistical decisions. Ann. Math. Statist. **23** 396-407.

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

Robbins, H. (1949). Unpublished lecture notes. Dept. of Statist., Univ. of North Carolina, Chapel Hill.

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

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. **35** 1-20.

Samuel, E. (1965). Sequential compound estimators. Ann. Math. Statist. **36** 879-889.

Van Ryzin, J.R. (1966a). The compound decision problem with $m \times n$ finite loss matrix. Ann. Math. Statist. **37** 412-424.

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