EQUIVARIANCE IN COMPOUND DECISION PROBLEMS AND A STABILITY OF SYMMETRIFICATIONS OF PRODUCT MEASURES

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY
JIN - SHENG HUANG

1970

This is to certify that the

thesis entitled

EQUIVARIANCE IN COMPOUND DECISION PROBLEMS
AND A STABILITY OF SYMMETRIFICATIONS OF
PRODUCT MEASURES

presented by

Jin-Sheng Huang

has been accepted towards fulfillment of the requirements for

Ph.D. **degree in** Statistics & Probability

James Hannan
Major professor

Date February 24, 1970

O-169

EQUIVARIANCE IN COMPOUND DECISION PROBLEMS AND A STABILITY OF SYMMETRIFICATIONS OF PRODUCT MEASURES

By

Jin-Sheng Huang

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1970

TO MY PARENTS

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Professor James F. Hannan who suggested the area of research and whose participation at each stage of the development actually amounts to co-authorship. I am also grateful to Professor Dennis C. Gilliland for reading the manuscript and for suggesting improvements. I am indebted to my former colleagues for their encouragements during my M.S.U. years. Finally I wish to thank Mrs. Noralee Barnes for her excellent typing and her cheerful attitude in preparing the manuscript.

TABLE OF CONTENTS

PART a

EQUIVARIANT PROCEDURES IN THE COMPOUND DECISION PROBLEM WITH FINITE STATE COMPONENT PROBLEM

0. SUMMARY

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

$c_i = \int \bigvee_a L(x,F_i,a)\,dF_i(x) < \infty.$

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

$$\rho_{ij} = \bigvee_{B \in \mathcal{B}} |F_i(B) - F_j(B)| \quad \text{and} \quad K(\rho) = \frac{12}{11}\left(\frac{3}{5}\right)^{3/2}\rho(1-\rho)^{-3/2},$$

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

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

and by

(T2) $\quad 2^m\{2K(\rho) \sum_i c_i^2\}^{\frac{1}{2}} N^{-\frac{1}{2}} \quad$ where $\quad \rho = \bigvee\{\rho_{ij} | \rho_{ij} < 1\}$.

The bound (T1) is infinite unless the F_i are pairwise non-orthogonal and (T2) is designed to replace it in this case.

1. NOTATIONS AND HISTORY

Let $(\mathfrak{X},\mathcal{B},P)$ be a probability measure space for each $P \in \mathcal{P} = \{F_0,F_1,\ldots,F_m\}$, $\mathcal{A}$ be an action space, L be a loss function which is defined on $\mathfrak{X} \times \mathcal{P} \times \mathcal{A}$ to the non-negative reals with value variously expressed

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

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

$$(2) \qquad c_i = \underline{\int} \bigvee_a L_i(a)dF_i < \infty.$$

Since the space $\mathcal{A}$ serves only as a parameter space for the class $\mathcal{L} = \{L(a) | a \in \mathcal{A}\}$ of loss functions on $\mathfrak{X} \times \mathcal{P}$, it is without loss of generality to assume that $\mathcal{A}$ contains no duplicates in this sense. To avoid the notational buildup attendant on the introduction of randomization at this and higher levels, we assume that $\mathcal{A}$ is its own extension to the class of all probability measures on the σ-field of subsets of $\mathcal{A}$ generated by $\{L(x,P) | x \in \mathfrak{X},\ P \in \mathcal{P}\}$: given any such probability measure ξ,

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

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

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

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

(5) $$R(F_i,d) = F_i[L_i \circ d] \le c_i.$$

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

(6) $$\psi(G) = \bigwedge_{\mathcal{D}} G[F_i[L_i \circ d]].$$

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

(7) $${}_{x_\alpha}L_i \circ d_\alpha(\underset{\sim}{x})$$

is $\underset{\sim}{\mathcal{B}}$-measurable for all α and i. Letting

$$(8) \qquad W(\underset{\sim}{x},\underset{\sim}{P},\underset{\sim}{a}) = N^{-1} \sum_{\alpha=1}^{N} L(x_\alpha, P_\alpha, a_\alpha),$$

we define the risk of $\underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}$ at $\underset{\sim}{P} \in \underset{\sim}{\mathcal{P}}$ by

$$(9) \qquad R(\underset{\sim}{P},\underset{\sim}{d}) = \underset{\sim}{P}[W(\underset{\sim}{x},\underset{\sim}{P},\underset{\sim}{d}(\underset{\sim}{x}))] \leq N^{-1} \sum_i N_i c_i.$$

$\underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}$ is called a simple (sometimes, simple symmetric) procedure if $d_\alpha(\underset{\sim}{x}) = d(x_\alpha)$ for all α, a.e. $\underset{\sim}{\mathcal{P}}$ for some $d \in \mathcal{D}$. Let $\underset{\sim}{S}$ be the class of all simple procedures and let $\underset{\sim}{d} \in \underset{\sim}{S}$ be denoted by d^N. It will follow directly from the definition of $\underset{\sim}{\mathcal{E}}$ in Section 2 that $\underset{\sim}{S} \subset \underset{\sim}{\mathcal{E}}$, the subclass of $\underset{\sim}{\mathcal{D}}$ equivariant under the permutation group. As functions of $\underset{\sim}{P}$, $\bigwedge_{\underset{\sim}{S}} R(\underset{\sim}{P},\underset{\sim}{d})$ and $\bigwedge_{\underset{\sim}{\mathcal{E}}} R(\underset{\sim}{P},\underset{\sim}{d})$ 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 $\psi(\underset{\sim}{N})$ with $\underset{\sim}{N}$ denoting the empiric distribution of $P_1,\ldots,P_N$.

The compound decision problem was introduced by Robbins (1951). He argued that a bootstrap procedure which first estimates the empiric distribution of $P_1,\ldots,P_N$ 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 $\mathcal{P} \times \mathcal{A}$ 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 $x_1,\ldots,x_N$, thus obtaining strong convergence to zero uniformly in $\underset{\sim}{P}$ 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 $m \times n$ $\mathcal{O} \times \mathcal{Q}$ (Theorem 1 and its Corollary) and to certain compact $\mathcal{O} \times$ compact $\mathcal{Q}$ with L continuous for each x (Theorems 4 and 5). Under continuity and other restrictions on densities, analogues of generalizations to $m \times n$ $\mathcal{O} \times \mathcal{Q}$ 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 $\underset{\sim}{P}$ 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 (1969b) and Theorem 2 following as a somewhat involved corollary to Theorem 1.

Hannan and Van Ryzin (1965), for 2×2 $\mathcal{O} \times \mathcal{Q}$, and Van Ryzin (1966), for $m \times n$ $\mathcal{O} \times \mathcal{Q}$, have established a rate of $O(N^{-\frac{1}{2}})$ (and under additional restrictions on $\mathcal{O}$ and L, $O(N^{-1})$) for uniform risk convergence of bootstrap procedures based on estimators which are averages over $x_1,\ldots,x_N$ 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 (1969b), 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 $\underset{\sim}{P}$ with a stronger restriction on $\mathcal{P}$ than pairwise non-orthogonality, with $\mathcal{A}$ 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. In an important general development, Cogburn (1967) imbeds the compound and Empirical Bayes problems in a general theory of stringency. Section 4 of Samuel (1967) (cf. Robbins (1962) and Suzuki (1966b)) investigates a procedure Bayes against uniform prior on proportions for the simplest 2×2 example.

2. EQUIVARIANT DECISION PROCEDURES AND SYMMETRIFICATIONS IN A COMPOUND DECISION PROBLEM

Let $\mathcal{G}$ be the permutation group on N objects. The generic element $g \in \mathcal{G}$ will also be used to denote the transformation induced by g:

$$g\,\underset{\sim}{y} = (y_{g1}, \ldots, y_{gN}). \tag{10}$$

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

$$\underset{\sim}{P}(\underset{\sim}{B}) = (g\underset{\sim}{P})(g\underset{\sim}{B}), \quad \underset{\sim}{B} \in \underset{\sim}{\mathcal{B}}. \tag{11}$$

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

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

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

$\underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}$ is equivariant (under $\mathcal{G}$) if for all $g \in \mathcal{G}$,

$$\underset{\sim}{d}(g\underset{\sim}{x}) = g\underset{\sim}{d}(\underset{\sim}{x}) \quad \text{a.e. } \mathcal{P}^N. \tag{13}$$

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 $\underset{\sim}{d} \in \underset{\sim}{\mathcal{S}}$ if and only if there exists a function γ on $\mathfrak{X} \times \mathfrak{X}^{N-1}$ to $\mathcal{A}$, symmetric on $\mathfrak{X}^{N-1}$, and such that for all α

$$(14) \qquad d_\alpha(\underset{\sim}{x}) = \gamma(x_\alpha, \check{\underset{\sim}{x}}_\alpha) \quad \text{a.e. } \mathcal{P}^N,$$

where

$$(15) \qquad \check{\underset{\sim}{x}}_\alpha = (x_1, \ldots, x_{\alpha-1}, x_{\alpha+1}, \ldots, x_N).$$

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

$$(16) \qquad d_\alpha(\underset{\sim}{x}) = \delta(x_\alpha, \underset{\sim}{x}) \quad \text{a.e. } \mathcal{P}^N.$$

It follows from the definition of $\underset{\sim}{S}$ and $\underset{\sim}{\mathcal{S}}$ that

$$(17) \qquad \underset{\sim}{S} \subseteq \underset{\sim}{\mathcal{S}}.$$

For each $\underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}$ and $g \in \mathcal{G}$ define $\underset{\sim}{d}^g$, the g-conjugate of $\underset{\sim}{d}$, by

$$(18) \qquad \underset{\sim}{d}^g(\underset{\sim}{x}) = g^{-1}[\underset{\sim}{d}(g\underset{\sim}{x})].$$

Thus $\underset{\sim}{d} \in \underset{\sim}{\mathcal{S}}$ if and only if $\underset{\sim}{d} = \underset{\sim}{d}^g$ for all g. Define the symmetrification $\underset{\sim}{d}^*$ of $\underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}$ as the average of its conjugates,

$$(19) \qquad \underset{\sim}{d}^* = (N!)^{-1} \sum_{\mathcal{G}} \underset{\sim}{d}^g.$$

It follows immediately that

$$(20) \qquad \underset{\sim}{\mathcal{S}} = \underset{\sim}{\mathcal{D}}^* \equiv \{\underset{\sim}{d}^* \mid \underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}\}.$$

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 $\underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}$ and $g \in \mathcal{G}$, it follows directly from definition (9), (12) and the transformation theorem that the risk of the g-conjugate is

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

Averaging the above over $\mathcal{G}$ gives the risk of the symmetrification $\underset{\sim}{d}^{*}$,

$$
R(\underset{\sim}{P},\underset{\sim}{d}^{*}) = (N!)^{-1} \sum_{g} R(g\underset{\sim}{P},\underset{\sim}{d}). \tag{22}
$$

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

$$
R(\underset{\sim}{P},\underset{\sim}{d}) = R(g\underset{\sim}{P},\underset{\sim}{d}), \quad \underset{\sim}{d} \in \underset{\sim}{\mathcal{E}}. \tag{23}
$$

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

(24) $$\tilde{\psi}(\underset{\sim}{N}) = \bigwedge_{\underset{\sim}{\mathcal{S}}} R(\underset{\sim}{N},\underset{\sim}{d}).$$

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

For $\underset{\sim}{\delta} \in \underset{\sim}{\mathcal{D}}$ it follows from the definition of risk (9) that

$$\begin{aligned} NR(\underset{\sim}{P},\underset{\sim}{\delta}) &= \sum_{\alpha} \underset{\sim}{P}[L(x_\alpha,P_\alpha) \circ \delta_\alpha(\underset{\sim}{x})] \\ &= \sum_{i} \sum_{\alpha | P_\alpha = F_i} (F_i \times \check{\underset{\sim}{P}}_\alpha)[L_i \circ \delta_\alpha], \end{aligned}$$

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

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

(25) $$NR(\underset{\sim}{N},\underset{\sim}{\delta}) = \Sigma\, N_i F_i \times_j F_j^{N_{ji}}[L_i \circ \delta_1],$$

where F_i acts on x_1 and $\times_j F_j^{N_{ji}}$ acts on $\check{\underset{\sim}{x}}_1$. 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 $\mathcal{P}$ and let $f_i = dF_i/d\mu$. Abbreviating $\Sigma\, N_i f_i \times_j F_j^{N_{ji}}[L_i \circ \delta_1]$ by $T(\delta_1)$ for $\underset{\sim}{\delta} \in \underset{\sim}{\mathcal{S}}$, (25) is expressible as

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

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

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

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

We now show that $\tilde{\psi}(\underset{\sim}{N})$ is the Bayes risk in the compound problem against the uniform prior on the orbit indexed by $\underset{\sim}{N}$. Let $U_{\underset{\sim}{N}}$ denote such a prior. Applying the transformation theorem to the mapping $g \to g\underset{\sim}{P}$, we obtain from (22) that

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

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

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

and therefore $\underset{\sim}{d} \in \underset{\sim}{\mathcal{D}}$ is ε-Bayes if and only if

$$R(\underset{\sim}{N}, \underset{\sim}{d}^*) \le \tilde{\psi}(\underset{\sim}{N}) + \varepsilon. \tag{30}$$

4. THE DIFFERENCE BETWEEN THE TWO ENVELOPES WHEN $\mathcal{P}$ IS PAIRWISE NON-ORTHOGONAL

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

Theorem 1. Let $\rho_{ij} = \bigvee_{B\in\mathcal{B}} |F_i(B) - F_j(B)|$, c_i satisfy (2), $K(\rho) = \frac{12}{11}(\frac{3}{5})^{3/2}\rho(1-\rho)^{-3/2}$ and let $\rho = \bigvee_{i,j} \rho_{ij}$. Then

(31) $$\psi(\underset{\sim}{N}) - \tilde{\psi}(\underset{\sim}{N}) \le \{2K(\rho) \sum_i c_i^2\}^{\frac{1}{2}} N^{-\frac{1}{2}}.$$

Proof. For each $\underset{\sim}{N}$ and each equivariant $\underset{\sim}{\delta}$, we will construct a simple procedure d^N whose risk at $\underset{\sim}{N}$ is close to the risk of $\underset{\sim}{\delta}$ at $\underset{\sim}{N}$. To bound the difference in risks we use Theorem 3 of Hannan and Huang (1969b), renotated here for our application by the use of relation (14) of that paper:

For any positive integer N and any non-negative integral partitions $\underset{\sim}{N}$ and $\underset{\sim}{N}'$ of N,

(32) $$\begin{aligned} &\bigvee\{\times_i F_i^{N_i}[\varphi] - \times_i F_i^{N_i'}[\varphi] \,|\, 0 \le \varphi = \varphi^* \le 1\}^2 \\ &\le n\,K(\rho) \sum_i \Lambda_i^{-1}(N_i' - N_i)^2, \end{aligned}$$

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

For $\underset{\sim}{\delta} \in \underset{\sim}{\mathcal{S}}$ consider $R(\underset{\sim}{N},\underset{\sim}{\delta})$ in the form (25). For given i and x_1, let

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

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

(34) $$\underset{i}{\times} F_j^{N_{ji}}[L_i \circ \delta_1] \geq \underset{j}{\times} F_j^{N_{jJ}}[L_i \circ \delta_1] - \{\bigvee_a L_i(a)\}\{K(\rho_{iJ})(\frac{1}{N_i} + \frac{1}{N_J})\}^{\frac{1}{2}},$$

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

(35) $$NR(\underset{\sim}{N},\underset{\sim}{\delta}) \geq \sum N_i F_i \underset{j}{\times} F_j^{N_{iJ}}[L_i \circ \delta_1] - (2K(\rho))^{\frac{1}{2}} \sum_i N_i^{\frac{1}{2}} c_i,$$

with c_i given by (2).

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

(36) $$\underset{j}{\times} F_j^{N_{jJ}}[{}_{x_1}L_i \circ \delta_1(x_1,\cdot)] = \xi[{}_{x_1}L_i(\cdot)],$$

with $\xi = (\underset{j}{\times} F_j^{N_{jJ}})[\delta_1(x_1,\cdot)]^{-1}$. We note that ξ depends on x_1 but not on i. By assumption (3), there exists $a_{x_1} \in \mathcal{A}$ 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,

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

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

$$(38) \qquad N R(\underset{\sim}{N},\underset{\sim}{\delta}) \geq N\,\psi(\underset{\sim}{N}) - (2K(\rho))^{\frac{1}{2}} \sum_i N_i^{\frac{1}{2}} c_i .$$

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

$$(39) \qquad R(\underset{\sim}{N},\underset{\sim}{\delta}) \geq \psi(\underset{\sim}{N}) - (2K(\rho))^{\frac{1}{2}} N^{-\frac{1}{2}} (\sum_i c_i^2)^{\frac{1}{2}} .$$

Since (39) holds for all $\underset{\sim}{\delta} \in \underset{\sim}{\mathcal{S}}$, this completes the proof of Theorem 1.

5. THE DIFFERENCE BETWEEN THE TWO ENVELOPES WHEN $\mathcal{P}$ MAY HAVE SOME PAIRWISE ORTHOGONALITY

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

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

$$\psi(\underset{\sim}{N}) - \tilde{\psi}(\underset{\sim}{N}) \leq 2^m \{2K(\rho) \sum_i c_i^2\}^{\frac{1}{2}} N^{-\frac{1}{2}}. \tag{40}$$

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 $\underset{\sim}{\delta} \in \underset{\sim}{\mathcal{S}}$ and a special choice of d to be $(2^{m+1} - 1)\varepsilon$ - Bayes with respect to $\underset{\sim}{N}$, we represent the difference in risks 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 $\underset{\sim}{N}$ to the sub-problems.

For each $I \subseteq \{0,\ldots,m\}$ let $\check{\mathcal{P}} = \{F_i | i \in I\}$ and $\check{N} = \sum_{i \in I} N_i$. The subproblem determined by $\check{\mathcal{P}}^{\check{N}}$ will be called the I problem. Let $\underset{\sim}{\check{\mathcal{D}}}$, $\underset{\sim}{\check{\mathcal{S}}}$, $\underset{\sim}{\check{N}}$, $U_{\underset{\sim}{\check{N}}}$, $\check{\psi}$, $\check{\tilde{\psi}}$, $\check{R}$, $\check{T}$ and $\check{*}$ denote the I problem counterpart of these symbols without the delete sign $\vee$. For simplicity we omit the delete sign on $\check{*}$ and $\check{R}$ hereafter.

Let

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

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

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

$$\text{(42)} \qquad NR(\underset{\sim}{N},\underset{\sim}{\delta}) = \sum_I \mu_I\{\sum_{i\in I} N_i f_i \underset{j\in I}{\times} F_j^{N_{ji}} \underset{j\notin I}{\times} F_j^{N_j}[L_i \circ \delta_1]\},$$

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

$$\text{(43)} \qquad \underset{j\notin I}{\times} F_j^{N_j}[L_i \circ \delta_1] = \xi[L_i], \quad i \in I.$$

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

$$\text{(44)} \qquad NR(\underset{\sim}{N},\underset{\sim}{\delta}) = \sum_I \mu_I[\check{T}(\check{\delta}_1)].$$

For each I, let d_I be ε-Bayes in $\mathcal{D}$ against $\check{\underset{\sim}{N}}/\check{N}$ and let $d = \sum_I \mathfrak{X}_I d_I$ where $\mathfrak{X}_I$ serves as the indicator function of itself. We note that $d \in \mathcal{D}$ 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 $(x_2,\ldots,x_{\check{N}})$, it follows that

$$\text{(45)} \qquad NR(\underset{\sim}{N},d^N) = \sum_I \mu_I[\check{T}(d_I)].$$

The difference between (45) and (44) is

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

For each I, define $\underset{\sim}{\check{h}} = (\check{h}_1,\ldots,\check{h}_{\check{N}}) \in \underset{\sim}{\check{\mathcal{D}}}$ by

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

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

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

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

$$\mu_I[\check{T}(d_I) - \check{T}(\check{\delta}_1)] \le \check{N}(\check{\psi}(\underset{\sim}{\check{N}}) - \tilde{\check{\psi}}(\underset{\sim}{\check{N}})) + \check{N}\epsilon. \tag{49}$$

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

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

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

$$R(\underset{\sim}{N},d^N) - R(\underset{\sim}{N},\underset{\sim}{\delta}) \le (2K(\rho))^{\frac{1}{2}} N^{-1} \sum_I \sum_{i\in I} N_i^{\frac{1}{2}} c_i + N^{-1}\epsilon \sum_I \check{N}. \tag{51}$$

Since (51) holds for all $\underset{\sim}{\delta} \in \underset{\sim}{\mathcal{S}}$ and all $\epsilon > 0$, and therfore for

$\varepsilon = 0$, the proof is complete upon using the Schwarz inequality in (51):

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

PART b

A STABILITY OF SYMMETRIFICATIONS OF PRODUCT MEASURES WITH FEW DISTINCT FACTORS

0. SUMMARY

Let $\mathcal{P} = \{F_0,\ldots,F_m\}$ be a class of probability measures on $(\mathcal{X},\mathcal{B})$. For any signed measure τ on $\mathcal{B}^N$, let τ^* be the average of τg over all $N!$ permutations g and let $\|\tau\| = \vee\{|\tau(C)|: C \in \mathcal{B}^N\}$. Let $d_{ij} = \|F_i - F_j\|$ and $K(x) = \frac{12}{11}(\frac{3}{5})^{3/2}x(1-x)^{-3/2}$. For any non-negative integral partitions $\underset{\sim}{N} = (N_0,\ldots,N_m)$ and $\underset{\sim}{N}' = (N_0',\ldots,N_m')$ of N, let $\delta_i = N_i' - N_i$ and $\wedge_i = (N_i' \wedge N_i) + 1$. With $\tau = \times F_i^{N_i} - \times F_i^{N_i'}$ and $n = \#\{i|\delta_i \neq 0\} - 1$, we bound $\|\tau^*\|^2$ by

(T3) $$n\,K(d)\Sigma\,\delta_i^2\,\wedge_i^{-1} \quad \text{with} \quad d = \vee\{d_{ij}|\delta_i \neq 0,\ \delta_j \neq 0\}$$

and, if $\mathcal{P}$ is internally connected by chains with non-orthogonal successive elements, by

(T4) $$\frac{1}{2}\,m\,K(\check{d})\,(\Sigma|\delta_i|)^2(\Sigma\,\wedge_i^{-1}) \quad \text{with} \quad \check{d} = \vee\{d_{ij}|F_i \not\perp F_j\}.$$

The bound (T3) is finite iff the F_i are pairwise non-orthogonal and (T4) is designed to replace it otherwise.

1. INTRODUCTION

Section 2 investigates some general properties of signed measures and their symmetrifications w.r.t. general groups. Section 3 specializes to the permutation group and notes a contraction effect of probability factors in product signed measures. The properties developed in Sections 2 and 3 will be used throughout the paper and, in particular, in Lemma 2, in the completion of the proof of Theorem 1, and in the proofs of Theorems 2, 3 and 4.

Section 4 proves Theorem 1, which is the special case of (T3) for $m = 1$ and $\delta_1 = 1$. This is the main result of the paper. Its proof contains a detailed outline of itself including Lemmas 2, 3 and 4. An example, consisting of the simplest special case, shows that the bound of Theorem 1 is sometimes asymptotically sharp to within a factor of 3.18... .

Section 5 proves Theorems 2, 3, and 4, all as corollaries to Theorem 1. Theorem 2 is the special case of (T3) for $m = 1$, Theorem 3 is (T3) and Theorem 4 is (T4).

Our main results are a strengthened generalization of Theorem II.1 of Hannan (1953). The latter is easily characterized in terms of the $m = 1$ case of (T3),

$$\text{(T2)} \qquad \|\tau^*\|^2 \le K(d_{01})\ \delta_1^2\ (\frac{1}{\Lambda_0} + \frac{1}{\Lambda_1}) ,$$

amounting to the assertion that, for $m = 1$ and fixed $F_0 \not\ll F_1$, (T II.1) $$\|\tau^*\|^2 \to 0 \quad \text{as} \quad \mathrm{RHS(T2)} \to 0.$$

As partially indicated in Section 3 and Section 5, the derivation in Section 5 of (T3) and (T4) as corollaries of Theorem 2 would equally well yield weakened (as (T II.1) weakens (T2)) forms of (T3) and (T4) as corollaries of (T II.1).

A corollary of the $\delta_1 = 1$ case of (T II.1), the Lemma of Hannan and Robbins (1955), was there used to show (Theorem 5) that the difference between the simple and equivariant envelopes converges to zero. A generalization of the Lemma, Theorem 2 of Horn (1968), is shown in Section 3 to be improved by the $\delta_1 = 1$ case of a rather immediate corollary to (T II.1). The special case of (T3) with two non-zero δ_i is used in Hannan and Huang (1969a) (Theorem 1) to bound a more general case of the difference. A similar application, Theorem 1 of Horn (1968), inherits the deficiencies of her Theorem 2.

2. PROPERTIES OF SYMMETRIFICATION AND OF AN $\mathcal{L}_1$-NORM

Although in this paper we will only be concerned with the difference between two symmetrified probability measures, some of the properties used in our proofs hold true and are easier to prove in a more general context. This section investigates some of these properties of symmetrified signed measures and their $\mathcal{L}_1$-norms.

Let $(\mathcal{Y},\mathcal{C})$ be a measurable space and $\mathcal{G}$ be a finite group of measurable transformations g on $(\mathcal{Y},\mathcal{C})$. For a signed measure τ on $(\mathcal{Y},\mathcal{C})$, we define τg as the induced signed measure and τ^* as the symmetrification of τ by

$$(\tau g)C = \tau(g(C)) \quad C \in \mathcal{C}, \quad \tau^* = AV(\tau g)$$

where AV denotes the average over $g \in \mathcal{G}$. Thus symmetrification (*, hereafter) is a linear operator.

For any real valued function f on $\mathcal{Y}$, define $f\circ g$ and f^* by

$$(f\circ g)y = f(gy) \ , \quad f^* = AV(f\circ g).$$

τ and f are said to be symmetric if $\tau = \tau^*$ and $f = f^*$, respectively. Since $\tau^* g \equiv \tau^*$, τ is symmetric iff $\tau \equiv \tau g$.

Throughout this paper we shall denote supremum and infimum by $\vee$ and $\wedge$, and express integrals as left operators by

$$\tau(f) = \int f(y)d\tau(y).$$

For any signed measure τ, define an $\mathcal{L}_1$-norm of τ by

(1) $$\|\tau\| = \vee\{|\tau(C)| : C \in \mathcal{C}\}.$$

It follows from the Jordan decomposition that

(2) $$\|\tau\| = \|\tau^+\| \vee \|\tau^-\|$$

and hence, if $\tau(1) = 0$,

(3) $$\|\tau\| = \|\tau^+\| = \|\tau^-\|.$$

In particular, if P and Q are probability measures, then

(4) $$0 \le \|P - Q\| \le 1,$$

with equality at 0 iff $P = Q$, and equality at 1 iff $P \perp Q$.

For use in the proof of Lemma 1, let μ be a measure $\ni d\tau/d\mu$ exists. Since $d\tau^+/d\mu = (d\tau/d\mu)^+$ and $d\tau^-/d\mu = (d\tau/d\mu)^-$,

$$\|\tau\| = \mu(d\tau/d\mu)^+ \vee \mu(d\tau/d\mu)^-.$$

Hence, if $\tau(1) = 0$,

(5) $$2\|\tau\| = \mu(|\frac{d\tau}{d\mu}|).$$

It follows from the transformation theorem (Theorem 39.C, Halmos (1950)) that

$$(\frac{d\tau}{d\mu})\circ g = \frac{d\tau g}{d\mu g}, \quad g \in \mathcal{G}.$$

In particular, if $\mu = \mu^*$, AV of the above equality yields

(6) $$(\frac{d\tau}{d\mu})^* = \frac{d\tau^*}{d\mu}.$$

Let μ be a measure and let h and f be such that the products h^*f and hf^* are μ^*-integrable. By the transformation theorem and symmetry, $\mu^*(h^*f\bullet g) = \mu^*(h^*f)$ for all $g \in \mathcal{G}$. Averaging over $\mathcal{G}$ and interchanging h and f yields

$$(7) \qquad \mu^*(h^*f) = \mu^*(h^*f^*) = \mu^*(hf^*).$$

For the special case of (7) obtained by letting $h = d\mu/d\mu^*$, we have $h^* = 1$ by (6) and therefore, if f (and hence also f^*) is μ^*-integrable, then

$$(8) \qquad \mu^*(f) = \mu^*(f^*) = \mu(f^*).$$

For use in the completion of proof of Theorem 1, we note that, by subadditivity of norm and by $\|\tau g\| \equiv \|\tau\|$ for a signed measure τ,

$$(9) \qquad \|\tau^*\| \le \text{AV}\ \|\tau g\| = \|\tau\|.$$

It follows from norm subadditivity and the Schwarz inequality (hereafter referred to as NS-SI) that if τ_i are signed measures then

$$(10) \qquad \|\Sigma\ \tau_i\|^2 \le (\Sigma\ 1)\Sigma\ \|\tau_i\|^2.$$

If $\tau = \tau_1 \times \tau_2$ is a product signed measure, $\|\tau\|$ is simply related to the corresponding norms of its factors. We abbreviate by omission subscripts on the norms. Since $\tau^+ = (\tau_1^+ \times \tau_2^+) + (\tau_1^- \times \tau_2^-)$, $\tau^- = (\tau_1^+ \times \tau_2^-) + (\tau_1^- \times \tau_2^+)$, it follows easily that

$$(11) \qquad \|\tau_1\|\cdot\|\tau_2\| \le \|\tau_1 \times \tau_2\| \le 2\|\tau_1\|\cdot\|\tau_2\|,$$

with the first equality iff either τ_1 or τ_2 is a measure or the negative of a measure, the second equality iff $\tau_1(1) = \tau_2(1) = 0$. In particular, if τ_2 is a probability measure, the first equality of (11) yields

$$\|\tau_1 \times \tau_2\| = \|\tau_1\|. \tag{12}$$

To assist in comparing our results with Theorem II.1 of Hannan (1953), consider a signed measure τ with $\tau(1) = 0$. By (3),

$$\|\tau\| = \|\tau^+\| = \vee\{\tau(f) \mid 0 \le f \le 1\}. \tag{13}$$

Since $\tau^*(1) = 0$ by linearity of $*$ operator, upon applying (13) to τ^*, and applying (8) to $\tau^*(f)$, it follows that

$$\|\tau^*\| = \vee\{\tau(f^*) \mid 0 \le f \le 1\} = \vee\{\tau(f) \mid 0 \le f^* = f \le 1\}, \tag{14}$$

with the second equality following from $\{f^* \mid 0 \le f \le 1\} = \{f \mid 0 \le f = f^* \le 1\}$.

3. PERMUTATIONS; CONTRACTION EFFECT OF PROBABILITY FACTORS

Henceforth we specialize $\mathcal{G}$ to be the group of transformations on $(\mathcal{X},\mathcal{B})^N$ induced by the group of permutations on N objects, where $(\mathcal{X},\mathcal{B})$ is a measurable space. We also let $\mathcal{G}$ denote the permutation group itself. Thus a generic element $g \in \mathcal{G}$ will be used both as a permutation and the transformation $g\underset{\sim}{x} = (x_{g1},\ldots,x_{gN})$.

The following lemma will be used in the successive extensions, in Section 5, from Theorem 2 to Theorem 3 to Theorem 4. Starting from Theorem II.1 of Hannan (1953) rather than Theorem 2, Lemma 1 together with NS-SI (10) would yield an extension paralleling our Theorem 3 extension from Theorem 2. It will be shown that Lemma 1 alone would yield an extension improving Theorem 2 of Horn (1968), where there is a stronger restriction on the F_i than pairwise non-orthogonality, and where the non-zero δ_i are 1 and -1 respectively.

Lemma 1. If $\tau = \check{\tau} \times P$ for a signed measure $\check{\tau}$ with $\check{\tau}(1) = 0$ and a probability measure P, then, abbreviating affixes on $*$ and on $\|\ \|$ by omission,

$$\|\tau^*\| \leq \|(\check{\tau})^*\|.$$

Proof. Since $\tau(1) = 0$, $\|\tau^*\| = \vee\{\check{\tau}(P(f))\,|\,0 \leq f = f^* \leq 1\}$ by (14). Since $\check{f} = P(f)$ is symmetric in the remaining variables and since $0 \leq \check{f} \leq 1$, one more application of (14) completes the proof.

For our extension of Theorem II.1, let $\tau = \times F_i^{N_i} - \times F_i^{N_i'}$ where the F_i are pairwise non-orthogonal and fixed, and the $\delta_i = N_i' - N_i$ are zero except for two i's. By judicious choice of g, we have $\tau g = \check{\tau} \times P$ with $P = \times\{F_i^{N_i} | \delta_i = 0\}$. By $(\tau g)^* \equiv \tau^*$ and Lemma 1, $\|\tau^*\| \le \|(\check{\tau})^*\|$ which, by (14) and Theorem II.1, converges to zero as this case of the bound in (T2) does.

4. TWO DISTINCT FACTORS WITH UNIT DIFFERENCES IN MULTIPLICITIES

Theorem 1. Let F_1 *and* F_0 *be p-measures and let* $\underset{\sim}{N} = (N_1, N_0)$ *be an integral partition of* N *with* $N_1 \geq 0 < N_0$. *With* $d = \|F_1 - F_0\|$, *with* $C(d) = d(1-d)^{-3/2}$ *and with* $K(d) = (.4)^2(6^{-1} + 5^{-1})^{-1}C(d)/C(.4)$

$$(15) \qquad \|(F_1^{N_1} \times F_0^{N_0})^* - (F_1^{N_1+1} \times F_0^{N_0-1})^*\|^2 \leq K(d)\left(\frac{1}{N_1+1} + \frac{1}{N_0}\right).$$

Proof. Since both sides of (15) vanish if $d = 0$, henceforth assume $F_1 \neq F_0$.

The proof proceeds according to the following outline: As in Hannan (1953), a parametric family of densities of the F_i is introduced and some of their moment properties are related to d. Starting as in Hannan (1953) but then weakening by the Schwarz inequality, Lemma 2 obtains a family of upper bounds for a slight generalization of LHS (15). Lemma 3 develops lower bounds for modal binomial probabilities for application to the denominators of a bound of Lemma 2. Lemma 4 uses characteristic function inversions to bound a generalization of the numerators of the bound of Lemma 2. The bound of (15) is then obtained as the minimum of d^2, from Lemma 1, and a bound resulting from the application of the other lemmas.

For $0 < p = 1-q < 1$ and $i = 1,0$, let

$$(16) \qquad F_p = pF_1 + qF_0\ , \quad f_i = dF_i/dF_p,$$

(17) $\theta = F_1(f_0)$, $P_i = p\, F_1(f_i) = 1 - Q_i (= 1 - q\, F_0(f_i))$,

and note

(18) $$F_i(f_i) = F_p(f_i^2) > (F_p(f_i))^2 = 1$$

by the Schwarz inequality, with equality eliminated by $F_p \neq F_i$. Thus

(19) $$\theta = p^{-1}(1 - q\, F_0(f_0)) < 1.$$

Let μ be any σ-finite measure dominating the F_i, let $h_i = dF_i/d\mu$ and let $h_p = ph_1 + qh_0$. Since $f_1 f_0 h_p = h_1 h_0 h_p^{-1} \geq h_1 \wedge h_0$, it follows that

(20) $$\theta \geq \mu(h_1 \wedge h_0) = 1 - d.$$

Finally, note that from (17) and (18)

(21) $P_1 - P_0 = 1 - \theta$, $P_i Q_i = pq\, F_1(f_i) F_0(f_i) > pq\, \theta$.

For integer N and $\underset{\sim}{p} \in [0,1]^N$, let $b(k;\underset{\sim}{p})$ denote the generalized binomial probability of k successes in N independent trials with success probabilities $\underset{\sim}{p}$.

At the cost of notational complications, the following lemma could be stated and proved for m instead of 2 probability measures.

<u>Lemma 2.</u> <u>For non-negative integral partitions of</u> N, $\underset{\sim}{N} = (N_0, N_1)$ <u>and</u> $\underset{\sim}{N}' = (N_0', N_1')$, <u>define</u> $\underset{\sim}{F} = F_1^{N_1} \times F_0^{N_0}$, $\underset{\sim}{F}' = F_1^{N_1'} \times F_0^{N_0'}$ <u>and</u> $\tau = \underset{\sim}{F} - \underset{\sim}{F}'$. <u>For</u> $0 < p < 1$

$$(22)\qquad \underset{\sim}{F}^*\left(\frac{d\underset{\sim}{F}'^*}{dF_p^N}\right) = \frac{b(N_1;P_1^{N_1'}\,P_0^{N_0'})}{b(N_1;p^N)} = \frac{b(N_1';P_1^{N_1}\,P_0^{N_0})}{b(N_1';p^N)}\,,$$

$$(23)\qquad 4\|\tau^*\|^2 < \left.\left.\frac{b(k;P_1^r\,P_0^{N-r})}{b(k;p^N)}\right]_{k=N_1}^{k=N_1'}\right]_{r=N_1}^{r=N_1'}\,.$$

Proof. Since $\underset{\sim}{F}^*(d\underset{\sim}{F}'^*/dF_p^N) = \underset{\sim}{F}'^*(d\underset{\sim}{F}^*/dF_p^N)$, the second equality in (22) will follow from the first.

By (6) the integrand in LHS (22) is $(d\underset{\sim}{F}'/dF_p^N)^*$, whence, by (8), LHS (22) is AV of

$$(24)\qquad \underset{\sim}{F}\left(\frac{d\underset{\sim}{F}'}{dF_p^N}\circ g\right)$$

Since the integrand in (24) is the product $\Pi_1^{N_1'} f_1(x_{g\alpha})\Pi_{N_1'+1}^N f_0(x_{g\alpha})$ of $\underset{\sim}{F}$-independent variables, the integral is expressible, in terms of $K = \#[\alpha > N_1' | g\alpha < N_1]$, as the product

$$(25)\qquad \begin{aligned}&[F_0(f_0)]^{N_0'-K}[F_1(f_0)]^K[F_0(f_1)]^{N_0-N_0'+K}\,[F_1(f_1)]^{N_1-K}\\ &= p^{-N_1}q^{-N_0}P_0^K Q_0^{N_0'-K}\,P_1^{N_1-K}\,Q_1^{N_1'-N_1+K} = H(K).\end{aligned}$$

Then (22) follows since

$$(26)\qquad \mathrm{AV}\ H(K(g)) = \sum_k \binom{N_0'}{k}\binom{N_1'}{N_1-k}\binom{N}{N_1}^{-1} H(k) = \frac{b(N_1;P_1^{N_1'}P_0^{N_0'})}{b(N_1;p^N)}$$

with the first equality following from the transformation theorem.

From (5) and the Schwarz inequality,

$$(27)\qquad 4\|\tau^*\|^2 = \left[F_p^N\left(\left|\frac{d\tau^*}{dF_p^N}\right|\right)\right]^2 < F_p^N\left(\left|\frac{d\tau^*}{dF_p^N}\right|^2\right) = \tau^*\left(\frac{d\tau^*}{dF_p^N}\right).$$

Four-fold application of (22) to this bound results in (23), completing the proof.

For integers $0 \le k \le N$, let

(28) $a_{Nk} = \{(N-1)pq\}^{\frac{1}{2}}\, b(k;p^N)$ with $p = (k+1)/(N+1)$.

Lemma 3. $a_{N\ N-1-k} = a_{Nk} \ge a_{N0} \uparrow e^{-1}$ as $N \uparrow \infty$.

Proof. Note that $f(x) = (1+x^{-1})^{x+\frac{1}{2}} \downarrow e$ as $0 < x \uparrow \infty$ since, with $t = (2x+1)^{-1}$, $\log f(x) = 1 + t^2/3 + t^4/5 + \ldots$. The lemma then follows from the representations,

$$(29)\qquad a_{Nk} = a_{Nk-1}\frac{f(k)}{f(N-k)} = a_{N0}\prod_1^k \frac{f(j)}{f(N-j)}\ ,\ a_{N0} = \{\frac{N-1}{N+1}\}^{\frac{1}{2}}\frac{1}{f(N)}\ .$$

The following lemma uses characteristic function inversions to bound a mixed second divided difference of generalized binomial probabilities. In our application to certain numerators of (23), the generalized binomials involve only two distinct probabilities but the added generality simplifies the proof and may serve to motivate other applications.

Lemma 4. For integers $0 \le k \le M$, $\underset{\sim}{P} = (P_1,\ldots,P_M) \in [0,1]^M$ and $(P,P+h) \in [0,1]^2$ less the diagonal, let Δ and $\boldsymbol{\Delta}$ respectively denote the divided difference operators from k to $k+1$ and from P to $P+h$. With $B(x) = (4/\pi)\int_0^1 y^{\frac{1}{2}}(1-y)^{-\frac{1}{2}}e^{-2xy}dy$ and $\sigma^2 = \Sigma_1^M P_\alpha(1-P_\alpha)$,

$$(30)\qquad |\Delta\ \boldsymbol{\Delta}\ b(k;(P,\underset{\sim}{P}))| \le B(\sigma^2),$$

$$(31)\qquad x^{3/2}B(x) \to (2\pi)^{-\frac{1}{2}} \text{ as } x \uparrow \infty,$$

$$(32)\qquad s = \sup\{x^{3/2}B(x)\,|\,0 \le x < \infty\} = .545447\ldots\ .$$

Proof. Note for later use that, with $\phi(u) = Pe^{iu} + Q$ the ch.f. of a Bernoulli variable with parameter $P = 1 - Q$,

(33) $$|\phi(u)|^2 = 1 - 4PQ \sin^2 \frac{u}{2} \le \exp - 4PQ \sin^2 \frac{u}{2} .$$

Since $\Delta e^{-iuk} \blacktriangle \phi(u) = e^{-iuk} 4 \sin^2 \frac{u}{2}$, the iterated divided difference in (30) is given by the Fourier inversion

(34) $$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-iuk} 4 \sin^2 \frac{u}{2} \prod_1^M \phi_\alpha(u) du.$$

By application of (33) to the ϕ_α, the modulus of (34) is bounded by

(35) $$\frac{4}{\pi} \int_0^{\pi} \sin^2 \frac{u}{2} e^{-2\sigma^2 \sin^2 \frac{u}{2}} du = B(\sigma^2).$$

(Since $v^{-1} \sin v \downarrow$ on $0 < v \le \pi/2$, RHS (33) $\le \exp - 4PQ(u/\pi)^2$ on $|u| \le \pi$. The corresponding weakening of (35) implies $s < \pi^{5/2} 2^{-7/2} = 1.553 \ldots$:

$$B(x) \le \frac{4}{\pi} \int_0^{\pi} \sin^2 \frac{u}{2} e^{-2x(u/\pi)^2} du < \frac{1}{\pi} \int_0^{\infty} u^2 e^{-2x(u/\pi)^2} du = \frac{\pi^{5/2}}{2^{7/2}} x^{-3/2}.)$$

With I_0, I_1 denoting the modified Bessel functions,

(36)
$$\begin{aligned} B(x) &= [- \frac{2}{\pi} \int_0^1 y^{-\frac{1}{2}}(1-y)^{-\frac{1}{2}} e^{-2xy} dy]' \\ &= [-2e^{-x} I_0(x)]' = 2e^{-x}[I_0(x) - I_1(x)], \end{aligned}$$

where the second equality follows by differentiation from the corresponding Laplace transform (cf. 29.3.124 of NBS-AMA55 (1964)), and the last from $I_0' = I_1$.

In view of (36), (31) is an immediate consequence of the usual asymptotic expansions of I_0 and I_1 (cf. 9.7.1 ibid).

To verify (32), first note that $[x^{3/2}B(x)]' = x^{\frac{1}{2}}e^{-x}F(x)$ with

$$F = (3-4x)I_0 + (4x-1)I_1.$$

Since $F(1) > 0 > F(2)$, the behavior of F' will imply that $\exists\, x_0 \in (1,2)$ with $F \gtrless 0$ on $[0,x_0]$ and $F < 0$ on (x_0,∞) as follows. Since $I_1 \le xI_0$, $xF' = x(4x-5)I_0 + (1-x)(4x+1)I_1 \le -(15/16)xI_0$ on $[0,1]$. Since $x(4x-1)F' = (4x+1)(1-x)F - 3I_0$, $1 \le x$ and $F(x) \ge 0 \Rightarrow F'(x) < 0$ and hence $1 \le x_0$ and $F(x_0) = 0 \Rightarrow F < 0$ on (x_0,∞).

Since, in addition, $.00012 > F(1.452) > 0 > F(1.453)$ (p. 228 BAAS v VI (1937)),

$$0 \le s - (1.452)^{3/2}B(1.452) \le (1.452)^{\frac{1}{2}}e^{-1.452}(.00012)(.001) < 4(10^{-8})$$

which results in (32) and thus completes the proof.

<u>Completion of proof of Theorem 1</u>. From Lemma 2 with $N_1' = N_1+1$ we have this case of the bounds (23). Let $p = (N_1+1)/(N+1)$ henceforth. This choice insures both denominators in (this case of) (23) are equal and, by Lemma 3, are bounded below by $\{(N-1)pq\}^{-\frac{1}{2}}a_{N0}$. Application of Lemma 4 to the mixed second difference remaining in the numerator results in the upper bound $s|h|\sigma^{-3}$, with

$$h = P_1 - P_0 = 1-\theta \le d \quad \text{and} \quad \sigma^2 = N_1P_1Q_1 + (N_0-1)P_0Q_0 > (N-1)pq(1-d)$$

by (20) and (21), so that, with $c_N = 4^{-1}s(N+1)^{3/2}(N-1)^{-3/2}f(N) \gtrless 4^{-1}se = .3706 \ldots$ as $N \uparrow \infty$,

$$(37) \qquad \|\tau^*\|^2 \le \frac{s}{4a_{N0}}\, C(d)\{(N-1)pq\}^{-1} = c_N\, C(d)\left(\frac{1}{N_1+1} + \frac{1}{N_0}\right).$$

From Lemma 1, $\|\tau^*\|^2 \le d^2$ independently of $\underset{\sim}{N}$. Since pq is maximal at $p = \frac{1}{2}$, (37) is minimal w.r.t. $\underset{\sim}{N}$ given N at $N_1 = [N/2]$ and, therefore, exceeds d^2 for all d iff

(38) $$\max_d d^2/C(d) = (.4)^2/C(.4) \le c_N([\tfrac{N+2}{2}]^{-1} + [\tfrac{N+1}{2}]^{-1}).$$

Since RHS (38) $\downarrow$ w.r.t. N with first violation of (38) at N = 10, c_N of (37) is replaceable by $c_9 = .5185\ldots$. This in turn can be improved by first increasing s to insure equality at N = 10, thus replacing c_N by $(.4)^2(6^{-1} + 5^{-1})^{-1}/C(.4) = .5070\ldots$ and completing the proof.

That the bound of the theorem is sometimes relatively sharp, will follow from examination of the simplest special case.

<u>Example</u>. Let F_0 and F_1 be Bernoulli probability measures $b(\ ;\xi_0)$ and $b(\ ;\xi_1)$ with $0 = \xi_0 < \xi_1 < 1$. Then $(F_0^{N_0} \times F_1^{N_1})^*$ and $(F_0^{N_0-1} \times F_1^{N_1+1})^*$ are symmetric probability measures on $\{0,1\}^N$, putting mass, $\binom{N}{k}^{-1} b(k;\xi_1^{N_1})$ and $\binom{N}{k}^{-1} b(k;\xi_1^{N_1+1})$ respectively, on every $\underset{\sim}{x}$ in $\{0,1\}^N$ with exactly k ones. Writing b_k for $b(k;\xi_1^{N_1})$ and using $b(k;\xi_1^{N_1+1}) = \xi_1 b_{k-1} + (1-\xi_1)b_k$,

(39) $$2\|\tau^*\| = \sum_k |b_k - \xi_1 b_{k-1} - (1-\xi_1)b_k| = \xi_1 \sum_k |b_k - b_{k-1}| = 2\xi_1 b_m$$

with m the greatest integer not exceeding $(N_1+1)\xi_1$. As $(N_1+1)\xi_1(1-\xi_1) \to \infty$, $(N_1+1)\xi_1(1-\xi_1)b_m^2 \sim (2\pi)^{-1}$ and therefore

(40) $$\|\tau^*\|^2 \sim \frac{1}{2\pi}\,\frac{\xi_1}{1-\xi_1}\,\frac{1}{N_1+1}\,.$$

When $\xi_1 \to 0$ and $N_0/N_1 \to \infty$ (which is the most favorable case for the following comparison), the bound of the theorem exceeds RHS (40) only by the factor $.8\pi\ (30/11)(.6)^{3/2} = 3.18\ldots$.

5. m+1 DISTINCT FACTORS

We now obtain various extensions of Theorem 1 as corollaries to that theorem. Theorems 1, 2 and 3 represent successive extensions each subsuming, yet corollary to, its predecessor. Theorem 3 is vacuous unless the F_i are pairwise non-orthogonal, and Theorem 4 is designed to replace it in this case. Thus, as implied in the summary, our final results are merely Theorems 3 and 4. Let

$$\tau = \underset{i}{\times} F_i^{N_i} - \underset{i}{\times} F_i^{N_i'} , \quad d_{ij} = \|F_i - F_j\| \tag{41}$$

$$\delta_i = N_i' - N_i , \quad \wedge_i = (N_i' \wedge N_i) + 1, \quad i,j = 0,\ldots,m.$$

<u>Theorem</u> <u>2</u>. <u>For</u> $m = 1$,

$$\|\tau^*\|^2 \leq K(d_{01})\delta_1^2\left(\frac{1}{\wedge_0} + \frac{1}{\wedge_1}\right). \tag{42}$$

<u>Proof</u>. Assume without loss of generality $\delta_1 \geq 1$ (we may rename $\underset{\sim}{N}$ and $\underset{\sim}{N}'$ otherwise), and for $j = 1,\ldots,\delta_1$, let

$$\tau_j = F_0^{N_0-j+1} \times F_1^{N_1+j-1} - F_0^{N_0-j} \times F_1^{N_1+j} . \tag{43}$$

Since $\tau = \Sigma\, \tau_j$, it follows from linearity of $*$ and NS-SI (10) that

$$\|\tau^*\|^2 \leq \delta_1\, \Sigma\, \|\tau_j^*\|^2. \tag{44}$$

Applying Theorem 1 to each summand in RHS (44) completes the proof.

Theorem 3. Let $n = \#\{i \mid \delta_i \neq 0\} - 1$ be positive (otherwise $\tau^* = 0$), and let $d = \vee\{d_{ij} \mid \delta_i \neq 0,\ \delta_j \neq 0\}$. Then

$$\|\tau^*\|^2 \leq n\ K(d)\ \Sigma\ \delta_i^2 \Lambda_i^{-1}.$$

Proof. Given $\underset{\sim}{N}$ and $\underset{\sim}{N}'$ we construct a sequence of partitions $\underset{\sim}{N} = \underset{\sim}{N}_0, \underset{\sim}{N}_1, \ldots, \underset{\sim}{N}_r = \underset{\sim}{N}'$, for some $r \leq n$, as follows. To construct $\underset{\sim}{N}_1 = (N_{10}, \ldots, N_{1m})$, let s be such that $|\delta_s| = \wedge\{|\delta_j| \mid \delta_j \neq 0\}$, and let t be such that δ_t has opposite sign with δ_s. Let $N_{1s} = N_s + \delta_s$, $N_{1t} = N_t - \delta_s$ and $N_{1j} = N_j$ for the other j's. Thus $\underset{\sim}{N}_1$ stays between $\underset{\sim}{N}$ and $\underset{\sim}{N}'$ coordinatewise and differs from $\underset{\sim}{N}$ in two coordinates. Repeating this construction, the process terminates in $r \leq n$ steps since each successive $\underset{\sim}{N}_i$ identifies at least one more coordinate with $\underset{\sim}{N}'$. Define

$$\tau_i = \underset{j=0}{\overset{m}{\times}} F_j^{N_{i-1j}} - \underset{j=0}{\overset{m}{\times}} F_j^{N_{ij}},\ i = 1, \ldots, r.$$

Since $\underset{\sim}{N}_i$ differs from $\underset{\sim}{N}_{i-1}$ in two coordinates, say s and t, the fact that $(\tau_i g)^* = (\tau_i)^*$ enables us to use Lemma 1 to obtain

$$\text{(45)} \qquad \|\tau_i^*\| \leq \|(F_s^{N_{i-1s}} \times F_t^{N_{i-1t}})^* - (F_s^{N_{is}} \times F_t^{N_{it}})^*\|.$$

Applying Theorem 1 to RHS (45), we note that, since each $\underset{\sim}{N}_i$ stays between $\underset{\sim}{N}_{i-1}$ and $\underset{\sim}{N}_{i+1}$ coordinatewise, the denominators in this application of RHS (42) are bounded below by Λ_s and Λ_t respectively. Since K is increasing, we further weaken this application of the bound (42) by replacing d_{st} by d. Thus

$$\text{(46)} \qquad [\text{RHS } (45)]^2 \leq K(d)(N_{i-1s} - N_{is})^2 \left(\frac{1}{\Lambda_s} + \frac{1}{\Lambda_t}\right).$$

Since $N_{i-1j} = N_{ij}$ except for $j \neq s,t$, RHS (46) is

$$K(d) \sum_{j=0}^{m} \Lambda_j^{-1}(N_{i-1j} - N_{ij})^2.$$

Since $\tau = \sum_{i=1}^{r} \tau_i$, by NS-SI and the above representation of RHS (46),

$$(47) \qquad \|\tau^*\|^2 \le r \sum \|\tau_i^*\|^2 \le r\, K(d) \sum_j \Lambda_j^{-1} \sum_i (N_{i-1j} - N_{ij})^2.$$

Since $\sum_i (N_{i-1j} - N_{ij}) = \delta_j$ and since $N_{i-1j} - N_{ij}$ are of the same sign for fixed j, the summation w.r.t. i in the last term of (47) is bounded by δ_j^2. Since $r \le n$ the proof is complete.

<u>Theorem 4</u>. <u>Let $F_0, F_1, \ldots, F_m$ be internally connected by chains with successive elements non-orthogonal and let</u> $\check{d} = \vee\{d_{ij} | F_i \not\perp F_j\}$. <u>Then</u>

$$\|\tau^*\|^2 \le \frac{1}{2}\, m\, K(\check{d}) (\sum_i |\delta_i|)^2 \sum_i \Lambda_i^{-1}.$$

<u>Proof</u>. For any connected graph of finitely many vertices there exists a vertex whose removal leaves the remaining graph connected. We shall rename $F_0, \ldots, F_m$ in such a way that successive removal of $F_0, F_1, \ldots$ leaves the remaining connected. For each i, let $t(i)$ be $\ni t(i) > i$ and $F_{t(i)} \not\perp F_i$.

Given $\underset{\sim}{N}$ and $\underset{\sim}{N}'$ we consider the partition which differs from either $\underset{\sim}{N}$ (if $\delta_0 \le 0$) or $\underset{\sim}{N}'$ (if $\delta_0 > 0$) only on the 0-th and the $t(0)$-th coordinates, where the 0-th coordinate is $N_0 \wedge N_0'$ and the $t(0)$-th coordinate, compared to that of $\underset{\sim}{N}$ or $\underset{\sim}{N}'$, is increased by $|\delta_0|$. By weakening Theorem 3 on both K and the second denominator (1 + the $t(0)$-th coordinate of the new partition $\ge \Lambda_{t(0)} + |\delta_0| \ge \Lambda_{t(0)}$), we see that the square norm of $*$ of the difference between the product measures associated with the two partitions is bounded by

$$K(\check{d})\ \delta_0^2\left(\frac{1}{\Lambda_0} + \frac{1}{\Lambda_{t(0)}}\right). \tag{48}$$

Iterate the process. Letting $\delta_j^{(i)}$ be the difference in the j-th coordinate at the i-th iteration of this process, we see that $\delta_j^{(i)} = 0$ for $j < i$ and

$$\delta_j^{(i)} = \delta_j + \Sigma\ \{\delta_r^{(r)} \mid 0 \le r < i,\ t(r) = j\} \tag{49}$$

for $j \ge i$. We also note that the $\delta_r^{(r)}$ above are disjoint sums of δ's from $\{\delta_0,\dots,\delta_{i-1}\}$.

Since $\Lambda_j^{(i)}$, the minimum of the j-th coordinates for the two partitions at i-th iteration plus 1, is increasing w.r.t. i, the bound corresponding to (48), further weakened by $\Lambda_i^{(i)} \ge \Lambda_i$, is

$$K(\check{d})[\delta_i^{(i)}]^2\left(\frac{1}{\Lambda_i} + \frac{1}{\Lambda_{t(i)}}\right). \tag{50}$$

Since each iteration results in reducing one coordinate difference to zero, the process terminates in m steps. By NS-SI as in the proof of Theorem 3, we see that, by (50),

$$\|\tau^*\|^2 \le m\ K(\check{d})\ \sum_i \delta_i^{(i)^2}\left(\frac{1}{\Lambda_i} + \frac{1}{\Lambda_{t(i)}}\right). \tag{51}$$

The coefficient of $\frac{1}{\Lambda_i}$ in the summation above is, with $\delta_m^{(m)} = 0$,

$$\delta_i^{(i)^2} + \sum_r\ \{\delta_r^{(r)^2} \mid t(r) = i\}. \tag{52}$$

By (49) and the comment following it, complemented by $\Sigma\ \delta_i = 0$, we see that $\delta_i^{(i)}$ and the $\delta_r^{(r)}$ in (52) are disjoint sums of $\{\delta_0,\dots,\delta_{i-1},\delta_{i+1},\dots,\delta_m\}$. Thus the maximum of (52) over all $\underset{\sim}{\delta}$ is $\le (\frac{\Sigma|\delta|}{2} - \delta_i^+)^2 + (\frac{\Sigma|\delta|}{2} - \delta_i^-)^2 \le \frac{1}{2} (\Sigma|\delta|)^2$, and the proof is complete.

The hypothesis of Theorem 4 fails to hold if and only if $\mathcal{P}$ is disconnected. Then either (i) there exists a component, i.e. a connected set of factors, whose $\underset{\sim}{N}$-multiplicity differs from its $\underset{\sim}{N}'$-multiplicity, in which case it follows easily that $\|\tau^*\| = 1$, or (ii) every component has identical $\underset{\sim}{N}$- and $\underset{\sim}{N}'$-multiplicity, in which case $\|\tau^*\|$ is simply related to the $\|(\tau_c)^*\|$ corresponding to the separate components:

$$\bigvee_c \|(\tau_c)^*\| \le \|\tau^*\| \le \sum_c \|(\tau_c)^*\|, \tag{53}$$

where the second inequality follows from Lemma 1 and the triangular inequality.

BIBLIOGRAPHY

British Association for the Advancement of Science (1937). Mathematical Tables, Vol. VI, Bessel Functions, Part 1, Functions of Order Zero and Unity. University Press, Cambridge.

Cogburn, Robert (1965). On the estimation of a multivariate location parameter with squared error loss. Bernoulli (1723), Bayes (1763), Laplace (1813), Anniversary Vol., 24-29. Springer-Verlag.

________ (1967). Stringent solutions to statistical decision problems. Ann. Math. Statist. **38**, 447-464.

Ferguson, Thomas S. (1967). Mathematical Statistics. Academic Press, New York and London.

Fox, Richard (1968). Contributions to compound decision theory and empirical Bayes squared error loss estimation. RM-214, Department of Statistics and Probability, Michigan State University.

Gilliland, Dennis C. and Hannan, James F. (1969). On an extended compound decision problem. Ann. Math. Statist. **40**.

Halmos, Paul R. (1950). Measure Theory. D. Van Nostrand, New York.

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

Hannan, James F. and Huang, J.S. (1969a). Equivariant procedures in the compound decision problem with finite state component problem. RM-237, Department of Statistics and Probability, Michigan State University.

Hannan, James F. and Huang, J.S. (1969b). A stability of symmetrications of product measures with few distinct factors. RM-238, Department of Statistics and Probability, Michigan State University.

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 for compound decision problems. Tech. Report No. 10, Department of Statistics, Stanford University.

James, W. and Stein, Charles (1961). Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 361-379. University of California Press.

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

National Bureau of Standards Applied Mathematics Series 55 (1964). Handbook of Mathematical Functions, U.S. Government Printing Office, Washington, D.C.

Oaten, Allan (1969). Approximation to Bayes risk in compound decision problems. RM-233, Department of Statistics and Probability, Michigan State University.

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

Robbins, Herbert (1962). Some numerical results on a compound decision problem. Recent Developments in Information and Decision Processes. The Macmillan Co., New York (1962), 56-62.

Samuel, E. (1965). On simple rules for the compound decision problem. J. R. Statist. Soc. (B), 27, 238-244.

Samuel, E. (1967). The compound statistical decision problem. Sankhya 29, 123-140.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 197-206. University of California Press.

Stein, Charles (1966). An approach to the recovery of inter-block information in balanced incomplete block design, Research Papers in Statistics, Festschrift for J. Neyman, 351-366. John Wiley & Sons, London, New York and Sydney.

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 University.

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

Suzuki, G. (1966b). Discrete compound decision problem. Ann. Inst. Stat. Math. **18**, 127-139.

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.