A REPRESENTATION OF BAYES INVARIANT PROCEDURES IN TERMS OF HAAR MEASURE

JAMES V. ZIDEK

(Received Feb. 17, 1968; revised Apr. 26, 1969)

Summary

It is shown that under fairly general conditions the Bayes procedure among the class of procedures invariant under a group of transformations which leave the statistical problem invariant, is really a Bayes or formal Bayes procedure with respect to a prior measure which is constructed from the right Haar measure on the group and the specified prior. This result is useful in problems where the principle of invariance is applied. Such cases, involving the two action problem or a selection and ranking problem are given as examples.

1. Introduction

The purpose of this paper is to show that under fairly general conditions the Bayes invariant procedure with respect to a prior, λ, within the class of invariant procedures is really a formal Bayes (for a definition, see section 2, below) or Bayes procedure with respect to a certain prior measure. This measure is constructed from the right Haar measure on the group of transformations leaving the problem invariant (and giving rise to the class of invariant procedures in question) and the specified prior, λ.

This theorem eliminates the necessity of delineating the class of invariant procedures, among those available, before determining the Bayes procedure among them. Since the principle of invariance is often used as a means of reducing the class of decision procedures available (to a simpler one where the selection of a procedure is easier), the computation of the Bayes invariant procedure arises in a natural way. We give, as an example, the application of this theorem to the two action problem, where the group acting on the parameter space may create more than two orbits. A selection problem is also treated from the same point of view.

In as much as this theorem provides a representation of the Bayes invariant procedure, it yields, in the general problem, a precise mathematical form in place of an otherwise somewhat indefinite quantity. The theorem was used in this way by the author ([8], pp. 83–89) in proving the inadmissibility, under squared error loss, of Pitman's estimator of the scale parameter of the exponential density when its location parameter is unknown.

In particular contexts like that just described, it is often easy to deduce the consequences of the main result of this work. However, it seems desirable to record the result once and for all, in at least the generality given here, to avoid repetition of the argument involved, in these special cases.

The method of proof is a straightforward generalization of that used in proving the well-known result that if the group operating on the parameter space does so in a simply transitive manner, the "best" invariant procedure is the formal Bayes procedure with respect to the prior measure induced on the parameter space by the right Haar measure of this group. Some of the basic features of the proof are involved in a result proved by Stein [5].

In section two we develop the necessary notation and give some of the basic results from topological group theory that are used in the sequel. A suitable reference for this material is the book of Nachbin [3]. The notion of a problem's being invariant under a group is defined in this same section.

The third section is devoted to the main theorem. In section four, three applications of this theorem are given.

2. Notation and preliminary results

For convenience, the following convention is adopted. If D denotes an abstract space, let $s(D)$ be a, usually unspecified, σ-algebra of subsets of D. If D is a topological space, $\beta(D)$ will denote the σ-algebra of Borel subsets of D. The usefulness of this convention will become apparent below, where several point sets appear in the same context, each possessing a different σ-algebra.

Let $(\mathscr{X}, s(\mathscr{X}))$ denote a measurable space. Suppose a random variable, X, is observed and takes its values in $\mathscr{X}$. Assume X is distributed according to an unknown but unique member of a family,

$$\mathscr{P} = \{P_\theta ; \theta \in \Theta\} \tag{2.1}$$

of probability distributions on $s(\mathscr{X})$, which is indexed by a set Θ, called the parameter space. $(\Theta, s(\Theta))$ is a measurable space. After observing X, an element or "action" is chosen at random from a set, $\mathscr{A}$, of actions according to a certain probability distribution. More precisely, if

$(\mathcal{A}, s(\mathcal{A}))$ is a measurable space, this action is chosen by means of a decision procedure chosen from a class $\mathcal{D}$ of decision procedures available in the problem. Any element, $\delta(\circ, \circ)$ of $\mathcal{D}$ is a mapping of $s(\mathcal{A}) \times \mathcal{X}$ into the interval $[0, 1]$ which satisfies the conditions, for each $x \in \mathcal{X}$, $\delta(\circ, x)$ is a probability distribution on $s(\mathcal{A})$ and for each $A \in s(\mathcal{A})$, $\delta(A, \circ)$ is a measurable mapping of $\mathcal{X}$ into $[0, 1]$. If $X = x$ is observed, the action selected is a random variable distributed according to $\delta(\circ, x)$.

If the action chosen is $a \in \mathcal{A}$ when $X = x$ is observed, and X is distributed according to P_θ, $\theta \in \Theta$, a nonnegative loss, $L(a, x, \theta)$ is incurred. It is assumed that L is jointly measurable in its three arguments.

For any measurable function $f : \mathcal{X} \to (-\infty, \infty)$, let

$$E^\theta\{f(X)\} = \int_{\mathcal{X}} f(x) dP_\theta(x) , \tag{2.2}$$

whenever the latter quantity is defined. The risk, $r(\theta, \delta)$, of a procedure $\delta \in \mathcal{D}$, when X is distributed by P_θ, $\theta \in \Theta$, is defined by

$$r(\theta, \delta) = E^\theta\left\{\int_{\mathcal{A}} L(a, x, \theta)\delta(da; X)\right\} . \tag{2.3}$$

Suppose ψ is a σ-finite measure on $(\mathcal{X}, s(\mathcal{X}))$ which dominates the family, $\mathcal{P}$. Let, for each $\theta \in \Theta$, $p(\circ \mid \theta)$ denote the Radon-Nikodym derivative of P_θ with respect to ψ. Assume $p(\circ \mid \circ)$ is jointly measurable in its arguments.

Let Π be a probability measure on $s(\Theta)$. A Bayes procedure with respect to Π is defined to be any member of $\mathcal{D}$, assuming that at least one exists, which minimizes and makes finite

$$\int_\Theta r(\theta, \delta) d\Pi(\theta) \tag{2.4}$$

as a function of δ. If a Bayes procedure exists, it is a member of $\mathcal{D}$ which, at $X = x$, minimizes as a function of δ, the quantity,

$$\int_{\mathcal{A}} \int_\Theta L(a, x, \theta) p(x \mid \theta) d\Pi(\theta) \delta(da; x) \tag{2.5}$$

a.e. $[\psi]$. As is clear from expression (2.5), any Bayes procedure with respect to Π depends on Π only through the posterior probability distribution of θ,

$$P_\Pi(\circ \mid \circ) : s(\Theta) \times \mathcal{X} \to [0, 1]$$

defined by

$$P_\Pi(B \mid X = x) = \int_B p(x \mid \theta) d\Pi(\theta) \Big/ \int_\Theta p(x \mid \theta) d\Pi(\theta) , \tag{2.6}$$

a.e. $[\psi]$, for every $B \in s(\Theta)$.

From some points of view, it is reasonable to allow Π to be a σ-finite measure. Provided $P_\Pi(\Theta \mid X=x)<\infty$ a.e. $[\psi]$, Π is called a *prior measure* (improper if $\Pi(\Theta)=\infty$). We can define a *formal posterior distribution* of θ using equation (2.6). A *formal Bayes procdure* is defined as any member of $\mathscr{D}$ which, at $X=x$, minimizes, and makes finite

$$\int_{\mathscr{A}} \int_{\Theta} L(a, x, \theta) P_\Pi(d\theta \mid X=x) \delta(da; x) \tag{2.7}$$

a.e. $[\psi]$, as a function of δ. This is, of course, subject to the condition that such a member of $\mathscr{D}$ exists. The condition

$$\int_{\Theta} r(\theta, \delta_\Pi) d\Pi(\theta) < \infty \tag{2.8}$$

will usually not hold when Π is an improper prior measure and δ_Π denotes the formal Bayes procedure with respect to Π.

Before defining the notion of a problem's being invariant under a group of transformations, we give some definitions and results from the theory of topological groups.

Let G denote a locally compact topological group. μ and ν, respectively, will denote the left and right Haar measures on $\beta(G)$. Δ denotes the modular function of G, and is a continuous homomorphism of G into the multiplicative group of positive real numbers. For all $B \in \beta(G)$ and $g \in G$,

$$\begin{aligned} &\mu(gB) = \mu(B)\,, \\ &\nu(Bg) = \nu(B)\,, \\ &\nu(B) = \mu(B^{-1})\,, \end{aligned} \tag{2.9}$$

and

$$\Delta(g)\mu(B) = \mu(Bg)\,, \tag{2.10}$$

where $B^{-1} = \{y : y^{-1} \in B\}$ and if $B_i \subset G$, $i=1, 2$,

$$B_1 B_2 = \{y : y = g_1 g_2,\ \ g_i \in B_i,\ \ i=1, 2\}\,. \tag{2.11}$$

A locally compact group is called unimodular if $\Delta \equiv 1$. All locally compact Abelian and compact groups are unimodular.

If H is a subgroup of G, G/H denotes the space of left cosets of H in G. We define the natural mapping, f, of G onto G/H by

$$f(g) = gH\,, \tag{2.12}$$

where gH is that element of G/H which labels the coset

$$gH=\{y:y=gh,\ h\in H\}. \tag{2.13}$$

Suppose G/H is endowed with the quotient topology. Then f is both open and continuous.

The following lemma will be useful in section 3.

LEMMA 2.1. *If H is a compact subgroup of G, Δ is constant on left cosets of H.*

PROOF. Suppose $g_2\in g_1H$, g_1 and g_2 being in G. Then

$$g_1^{-1}g_2\in H.$$

According to a theorem of Weil (see, for example, Nachbin, p. 138) if $\Delta^H(\circ)$ denotes the modular function on H,

$$\Delta^H(g)=\Delta(g),\qquad g\in H.$$

Since H is compact, $\Delta(g)=1$, $g\in H$, and, in particular,

$$\Delta(g_1^{-1}g_2)=[\Delta(g_1)]^{-1}\Delta(g_2)=1.$$

Thus $\Delta(g_1)=\Delta(g_2)$.

Define a measure μ' on $(G/H, \beta(G/H))$ by

$$\mu'(B^*)=\mu(f^{-1}B^*),\qquad B^*\in\beta(G/H). \tag{2.14}$$

Then μ' is a left invariant measure with respect to the group G, provided that the operation of G on G/H is defined in the natural way, that is, if $g^*\in G/H$, and $g^*=f(g'H)$, then gg^* is defined by

$$gg^*=f(gg'H). \tag{2.15}$$

Furthermore μ' is finite on compact sets. For suppose $K^*\in\beta(G/H)$ is a compact subset of G/H. Then there exists a compact set $K\subset G$ for which $f(K)=K^*$ (see, for example, Loomis [2], p. 111). Since H is compact, KH is compact and

$$KH=f^{-1}(K^*)\in\beta(G). \tag{2.16}$$

But

$$\mu'(K^*)=\mu(f^{-1}f(KH))=\mu(KH)<\infty. \tag{2.17}$$

Suppose G is a group of one-to-one transformations of a measurable space $(M, s(M))$ onto itself. Then G is called, in our own terminology, a *measurable topological transformation group* acting on the left of M, provided the following conditions hold:

(i) for every element $g\in G$, the mapping $x\to g(x)$ is a measurable transformation of M onto itself.

(ii) for every pair of elements, g_1 and g_2, of G and every $x \in M$.

$$g_1(g_2(x)) = (g_1 g_2)(x) . \tag{2.18}$$

(iii) the bivariate mapping, $(g, x) \to g(x)$ is simultaneously measurable in g and x.

We will drop the term "topological" in "measurable topological transformation group" if we do not assume G is a topological group, but rather, that $(G, s(G))$ is a measurable space.

Observe, in particular, that if e denotes the identity of G, (ii) implies $ex = x$ for each $x \in M$.

If M^* is an invariant subspace of M (i.e., $GM^* = M^*$), then we can define the orbit space, M^*/G of G acting on M^*. Any point of M^*/G labels a subset of M^* which has the property that if x_1 and x_2 are two of its members, there exists an element, $g \in G$, such that $x_1 = gx_2$. If M^*/G is endowed with the quotient topology, the natural mapping of M^* onto M^*/G is continuous and open. In making this remark we are, of course, implicitly assuming M^* is a topological space.

If M^*/G reduces to a point, G is said to act *transitively* on M^*. If for every pair of points, x_1 and x_2 of M^*, there exists exactly one element, $g \in G$, for which $x_1 = gx_2$, G is said to act *simply transitively* or *exactly transitively* on M^*.

This concludes the summary of the basic elements of topological group and transformation group theory necessary for the results of the next section. We now discuss the notion of a statistical problem's being invariant under a group and the invariance of a statistical procedure.

Suppose $\mathcal{G}$ is a locally compact, measurable topological transformation group which acts on the left of the sample space, $\mathcal{X}$. Assume there exist measurable transformation groups $\bar{\mathcal{G}}$ and $\hat{\mathcal{G}}$, each homomorphic to $\mathcal{G}$ such that every element of $\bar{\mathcal{G}}$ and $\hat{\mathcal{G}}$ is a bimeasurable, one-to-one transformation of Θ and $\mathcal{A}$, respectively, onto itself. Under these homomorphic mappings, which are assumed to be measurable, let $\bar{h} \in \bar{\mathcal{G}}$ and $\hat{h} \in \hat{\mathcal{G}}$ correspond to the element $h \in \mathcal{G}$, and assume

$$P_{\bar{h}\theta}(hB) = P_\theta(B) , \qquad \theta \in \Theta,\ B \in s(\mathcal{X}) \tag{2.19}$$

and

$$L(\hat{h}a, hx, \bar{h}\theta) = L(a, x, \theta) , \qquad a \in \mathcal{A},\ x \in \mathcal{X},\ \theta \in \Theta . \tag{2.20}$$

The problem is then said to be invariant under the group $\mathcal{G}$.

An element $\delta \in \mathcal{D}$ is said to be an invariant procedure under the group $\mathcal{G}$ if

$$\delta(\bar{g}^{-1}A ; g^{-1}x) = \delta(A ; x) , \qquad x \in \mathcal{X},\ A \in s(\mathcal{A}) . \tag{2.21}$$

$\mathcal{D}_I \subset \mathcal{D}$ will denote the subclass of all those procedures which are invariant under $\mathcal{G}$. The maximal invariant under the group $\mathcal{G}$ is any random variable t for which $t(gx)=t(x)$, $x \in \mathcal{X}$ and $g \in \mathcal{G}$, and whose image space is isomorphic to $\mathcal{X}/\mathcal{G}$. We shall use the symbol $\simeq$ to mean "is isomorphic to."

3. A representation of Bayes invariant procedures

We now add the assumptions necessary to the present work:

ASSUMPTION I.

(3.1) $$\mathcal{X}=\mathcal{G}/\mathcal{H}\times\mathcal{X}^*$$

where $\mathcal{X}^*\simeq\mathcal{X}/\mathcal{G}$ and $\mathcal{H}$ is a compact subgroup of G which leaves a particular point, say x_0, invariant, that is,

(3.2) $$\mathcal{H}=\{h : h \in \mathcal{G},\ hx_0=x_0\}.$$

If $h \in \mathcal{H}$, $\hat{h}a=a$, $a \in \mathcal{A}$.

Topologize $\mathcal{G}/\mathcal{H}$ with the quotient topology and let π denote the continuous, open, natural mapping of $\mathcal{G}$ onto $\mathcal{G}/\mathcal{H}$. Assume

(3.3) $$s(\mathcal{X})=\beta(\mathcal{G}/\mathcal{H})\times s(\mathcal{X}^*),$$

where $s(\mathcal{X}^*)$ is unspecified.

If $\mathcal{G}$ operates exactly transitively on each orbit represented in $\mathcal{X}/\mathcal{G}$, Assumption I holds with $\mathcal{H}=\{e\}$, provided $\mathcal{X}$ is properly labelled. Keifer ([1], p. 584, p. 585) gives examples where, in one case, Assumption I holds with $\{e\}$ properly contained in $\mathcal{H}$, and another where it fails to hold.

Assumption I means that a random variable (G^*, X^*) is observed, where X^* is a maximal invariant and G^* takes its values in $\mathcal{G}/\mathcal{H}$. We define the operation of $\mathcal{G}$ on the range of this random variable as follows. If $g \in \mathcal{G}$ and $x=(g_\cdot^*, x^*) \in \mathcal{X}$,

(3.4) $$gx=(gg_\cdot^*, x^*),$$

where $gg_\cdot^*$ is defined as in equation (2.15).

In Assumption I $\mathcal{H}$ is assumed compact since we shall be interested in integrals of the form

(3.5) $$\int_{\mathcal{G}} f(\pi(g))d\mu(g),$$

where f is some nonnegative real valued measurable function, and if $\mathcal{H}$ were not compact, this integral would usually be infinite. μ, of course, denotes the left Haar measure of $\mathcal{G}$.

Define a measure μ^* on $\beta(\mathcal{G}/\mathcal{H})$ by

$$(3.6) \qquad \mu^*(B^*)=\mu(\pi^{-1}B^*) , \qquad B^* \in \beta(\mathcal{G}/\mathcal{H}) .$$

Let ρ be a given, σ-finite measure on $s(\mathcal{X}^*)$.

ASSUMPTION II. The family $\mathcal{P}$, given in equation (2.1), is dominated by $\mu^* \times \rho$.

The following lemma describes a property associated with the density functions, $p(g^*, x^* \mid \theta)$, of members of $\mathcal{P}$, which will be used frequently in this section.

LEMMA 3.1.

$$(3.7) \qquad p(g^*, x^* \mid \bar{g}_\circ\theta)=p(g_\circ^{-1}g^*, x^* \mid \theta) \qquad a.e. \ [\mu^* \times \rho] ,$$

when $(g^*, x^*) \in \mathcal{X}$, $\theta \in \Theta$, *and* $\bar{g} \in \bar{\mathcal{G}}$.

PROOF. Let $G^* \times B^*$ denote any measurable rectangle in $s(\mathcal{X})$. Then

$$\begin{aligned} &\int_{G^* \times B^*} p(g^*, x^* \mid \bar{g}_\circ\theta) d\mu^*(g^*) d\rho(x^*) \\ &= P_{\bar{g}_\circ\theta}(G^* \times B^*) \\ &= P_\theta(g_\circ^{-1}G^* \times B^*) \\ &= \int_{g_\circ^{-1}G^* \times B^*} p(g^*, x^* \mid \theta) d\mu^*(g^*) d\rho(x^*) \\ &= \int_{G^* \times B^*} p(g_\circ^{-1}g^*, x^* \mid \theta) d\mu^*(g^*) d\rho(x^*) . \end{aligned}$$

The conclusion is an immediate consequence of the uniqueness of the Radon-Nikodym derivative.

It is assumed that $p(\circ, \circ \mid \circ)$ can be chosen so that it is jointly measurable in its three arguments.

Denote $\Theta/\bar{\mathcal{G}}$ by Θ^*. $s(\Theta^*)$ is unspecified but assumed given. Let ϕ be any one-to-one measurable mapping (assuming one exists) of Θ^* into Θ for which

$$(3.8) \qquad \phi(\theta^*) \in \bar{\pi}^{-1}(\theta^*) , \qquad \theta^* \in \Theta^* ,$$

where $\bar{\pi}$ denotes the natural mapping of Θ onto Θ^*. Let T_ϕ denote the measurable mapping of $(\mathcal{G} \times \Theta^*, \beta(\mathcal{G}) \times s(\Theta^*))$ onto $(\Theta, s(\Theta))$ which is given by

$$(3.9) \qquad T_\phi(g, \theta^*)=\bar{g}\phi(\theta) .$$

ASSUMPTION III. There exists a one-to-one measurable transformation $\phi : \Theta^* \to \Theta$ satisfying condition (3.8) for which $\{g : \bar{g}\phi(\theta^*) = \phi(\theta^*)\}$ is a compact subset of $\mathcal{G}$, $\theta^* \in \Theta^*$.

Let λ be a given probability measure on $s(\Theta^*)$. Define a measure Λ on $s(\Theta)$ by

$$\Lambda(B) = (\nu \times \lambda) T_\phi^{-1} B , \qquad B \in s(\theta) , \tag{3.10}$$

where ν denotes the right Haar measure on $\mathcal{G}$.

LEMMA 3.2. *Λ is independent of the choice of ϕ.*

PROOF. Let ϕ_1 and ϕ_2 be two one-to-one measurable mappings of Θ^* into Θ, each of which satisfies (3.8). Define a one-to-one mapping, γ, of Θ^* into $\mathcal{G}$ which satisfies

$$\phi_2(\theta^*) = \overline{\gamma(\theta^*)}\phi_1(\theta^*) , \qquad \theta^* \in \Theta^* .$$

Then, if Λ_i $(i=1, 2)$ corresponds to T_{ϕ_i} in the manner described by equation (3.10), and $B \in s(\Theta)$,

$$\begin{aligned}\Lambda_2(B) &= \int_{\Theta^*} \nu\{g : \bar{g}\phi_2(\theta^*) \in B\} d\lambda(\theta^*) \\ &= \int_{\Theta^*} \nu\{g : \overline{g\gamma(\theta^*)}\phi_1(\theta^*) \in B\} d\lambda(\theta^*) \\ &= \Lambda_1(B)\end{aligned}$$

since

$$\begin{aligned}&\nu\{g : \overline{g\gamma(\theta^*)}\phi_1(\theta^*) \in B\} \\ &= \nu(\{g : \bar{g}\phi_1(\theta^*) \in B\}[\gamma(\theta^*)]^{-1}) \\ &= \nu\{g : \bar{g}\phi_1(\theta^*) \in B\} .\end{aligned}$$

The conclusion of this lemma follows from this observation.

We shall now show that it is meaningful to define a formal posterior distribution with respect to Λ. Let $(g_\cdot^*, x^*) \in \mathcal{X}$ and $g_\circ$ be any fixed element of $\pi^{-1}(g_\cdot^*)$. Then

$$\begin{aligned}&\int_\Theta p(g_\cdot^*, x^* \mid \theta) d\Lambda(\theta) \\ &= \int_{\Theta^*}\int_{\mathcal{G}} p(g_\cdot^*, x^* \mid \bar{g}\phi(\theta^*)) d\nu(g) d\lambda(\theta^*) \\ &= \int_{\Theta^*}\int_{\mathcal{G}} p(\pi(g^{-1}g_0), x^* \mid \phi(\theta^*)) d\nu(g) d\lambda(\theta^*) \\ &= (1/\Delta(g_0)) \int_{\Theta^*}\int_{\mathcal{G}} p(\pi(g), x^* \mid \phi(\theta^*)) d\mu(g) d\lambda(\theta^*) .\end{aligned}$$

Thus

$$\begin{aligned}(3.11)\qquad &\int_{\Theta} p(g_\circ^*, x^* \mid \theta) d\Lambda(\theta) \\ &= (1/\Delta(g_0)) \int_{\Theta^*} \int_{\mathcal{G}^*} p(g^*, x^* \mid \phi(\theta^*)) d\mu^*(g^*) d\lambda(\theta^*)\end{aligned}$$

where $\mathcal{G}^* = \mathcal{G}/\mathcal{H}$. Since $\mathcal{H}$ is compact, we conclude, from Lemma 3.1 that this last quantity does not depend on the choice of $g_\circ$ in $\pi^{-1}(g_\circ^*)$. Furthermore, if we integrate the left hand side of equation (3.11) with respect to ρ, we conclude from this equation, that the result is $1/\Delta(g_\circ)$ which is finite. Thus

$$(3.12)\qquad \int_{\Theta} p(g_\circ^*, x^* \mid \theta) d\Lambda(\theta) < \infty, \qquad \text{a.e. } [\mu^* \times \rho] .$$

For convenience, define a real valued function K^* on $\mathcal{X}^*$ by

$$(3.13)\qquad K(x^*) = \int_{\Theta^*} \int_{\mathcal{G}^*} p(g^*, x^* \mid \phi(\theta^*)) d\mu^*(g^*) d\lambda(\theta^*) ,$$

$x^* \in \mathcal{X}^*$. Then the formal posterior density function of θ given $(g_\circ^*, x^*)$ with respect to Λ, is given by

$$p(\theta \mid g_\circ^*, x^*) = p(g_\circ^*, x^* \mid \theta) \Big/ \int_{\Theta} p(g_\circ^*, x^* \mid \theta) d\Lambda(\theta) ,$$

or, using equation (3.11)

$$(3.14)\qquad p(\theta \mid g_\circ^*, x^*) = \Delta(g_0) p(g_\circ^*, x^* \mid \theta)/K(x^*) ,$$

where $g_\circ$ is any element of $\pi^{-1}(g_\circ^*)$.

Let $\mathcal{P}'$ denote the family of all probability distributions on $(\mathcal{A}, s(\mathcal{A}))$. To determine the formal Bayes procedure with respect to Λ consider, for $(g_\circ^*, x^*) \in \mathcal{X}$,

$$\begin{aligned}(3.15)\qquad &\inf_{P \in \mathcal{P}'} \int_{\Theta^*} \int_{\mathcal{G}} \int_{\mathcal{A}} L(a, (g_\circ^*, x^*), \bar{g}\phi(\theta^*)) \\ &\times dP(a) p(\bar{g}\phi(\theta^*) \mid g_\circ^*, x^*) d\nu(g) d\lambda(\theta^*) .\end{aligned}$$

We shall show that under Assumption IV, below, the formal Bayes procedure we obtain, is the Bayes invariant procedure with respect to λ.

Observe that

$$\begin{aligned}(3.16)\qquad &\int_{\Theta^*} \int_{\mathcal{G}} \int_{\mathcal{A}} L(\hat{h}^{-1}a, (h^{-1}g_\circ^*, x^*), \overline{(h^{-1}g)}\phi(\theta^*)) dP(a) \\ &\times \frac{p(h^{-1}g_\circ^*, x^* \mid \overline{(h^{-1}g)}\phi(\theta^*))}{\Delta(h^{-1})K(x^*)} d\nu(g) d\lambda(\theta^*) \\ &= \int_{\Theta^*} \int_{\mathcal{G}} \int_{\mathcal{A}} L(a, (h^{-1}g_\circ^*, x^*), \bar{g}\phi(\theta^*)) dP(\hat{h}a) \\ &\times p(\bar{g}\phi(\theta^*) \mid h^{-1}g_\circ^*, x^*) d\nu(g) d\lambda(\theta^*) ,\end{aligned}$$

$h \in \mathcal{G}$. Suppose, for almost all $[\mu^* \times \rho]$ pairs, $(g^*_\circ, x^*)$, the infimum in expression (3.15) is attained (and is finite) at a point of $\mathcal{P}'$ which we denote by $\delta(\circ; g^*_\circ, x^*)$. Then as a consequence of equation (3.16), this infimum is attained and is finite at the point, $\delta(\hat{h}^{-1}(\circ); h^{-1}g^*_\circ, x^*)$ for each $h \in \mathcal{G}$. If e denotes the identity of $\mathcal{G}$, $e^* = \pi(e)$, and we choose $g_0 \in \pi^{-1}(g^*_\circ)$, then the same is true, in particular, for

$$\delta(\hat{g}_\circ^{-1}(\circ); e^*, x^*) . \tag{3.17}$$

Suppose $g_{\circ 1}$ and $g_{\circ 2}$ are both elements of $\pi^{-1}(g^*_\circ)$. Then

$$\delta(\hat{g}_{\circ 2}^{-1}(\circ); e^*, x^*) = \delta(\hat{g}_{\circ 1}^{-1}\widehat{(g_{\circ 1}g_{\circ 2}^{-1})}(\circ); e^*, x^*)$$

and consequently

$$\delta(\hat{g}_{\circ 2}^{-1}(\circ); e^*, x^*) = \delta(\hat{g}_{\circ 1}^{-1}(\circ); e^*, x^*) , \tag{3.18}$$

since $g_{\circ 1}g_{\circ 2}^{-1} \in \mathcal{H}$ and, by Assumption I, $g_{\circ 1}g_{\circ 2}^{-1}a = a$ for every $a \in \mathcal{A}$. Define δ_F by

$$\delta_F(\circ; g^*_\circ, x^*) = \delta(\hat{g}_\circ^{-1}(\circ); e^*, x^*) \tag{3.19}$$

where $\pi(g_\circ) = g^*_\circ$. By equation (3.18), δ_F is well defined. Furthermore, for each $h \in \mathcal{G}$,

$$\begin{aligned} \delta_F(\hat{h}^{-1}(\circ); h^{-1}g^*_\circ, x^*) &= \delta(\hat{g}_\circ^{-1}\hat{h}\hat{h}^{-1}(\circ); e^*, x^*) \\ &= \delta_F(\circ; g^*_\circ, x^*) , \end{aligned} \tag{3.20}$$

so that if δ_F were a procedure, as we assume in Assumption IV, it would be invariant.

Assumption IV. For almost all $[\mu^* \times \rho]$ pairs, $(g^*, x^*) \in \mathcal{X}$, there exists at least one member of $\mathcal{P}'$, denoted by $\delta(\circ; g^*_\circ, x^*)$ which achieves and makes finite, the infimum in expression (3.15). δ_F, defined in equation (3.19) is an element of $\mathcal{D}$.

The result contained in the following lemma is well-known and its proof is omitted.

Lemma 3.3. *If* $\delta \in \mathcal{D}_T$, $r(\theta, \delta) = r(\bar{g}\theta, \delta)$, $g \in \mathcal{G}$, $\theta \in \Theta$.

An important consequence of this lemma is that we can, in speaking of Bayes invariant procedures, suppose the prior distribution is on $s(\Theta^*)$.

The main result of this paper now follows as a simple consequence of the foregoing lemmas.

Theorem 3.1. *Under Assumption* I–IV, *the Bayes or formal Bayes procedure with respect to the prior measure* Λ *(see equation* (3.10)), *defined* *by equation* (3.19), *is a Bayes invariant procedure with respect to the given prior* λ, *assuming one exists.*

PROOF. Let $R(\lambda, \delta)$ denote the Bayes risk of an invariant procedure, δ, with respect to the prior, λ, on $(\Theta^*, s(\Theta^*))$, that is,

$$R(\lambda, \delta) = \int_{\Theta^*}\int_{\mathcal{X}^*}\int_{\mathcal{G}}\int_{\mathcal{A}} L(a, (\pi(g), x^*), \phi(\theta^*)) \times \delta(da; \pi(g), x^*)p(\pi(g), x^* \mid \phi(\theta^*))d\mu(g)d\rho(x^*)d\lambda(\theta^*) .$$

Then for any element $g_\circ \in \mathcal{G}$,

$$\begin{aligned} R(\lambda, \delta) &= \int_{\mathcal{X}^*}\int_{\Theta^*}\int_{\mathcal{G}}\int_{\mathcal{A}} L(a, (\pi(g^{-1}g_\circ), x^*), \phi(\theta^*)) \\ &\quad \times \delta(da; \pi(g^{-1}g_\circ), x^*)\frac{p(\pi(g^{-1}g_\circ), x^* \mid \phi(\theta^*))\Delta(g_\circ)}{K(x^*)} \\ &\quad \times d\nu(g)d\lambda(\theta^*)K(x^*)d\rho(x^*) \\ &= \int_{\mathcal{X}^*}\int_{\Theta^*}\int_{\mathcal{G}}\int_{\mathcal{A}} L(a, (\pi(g_\circ), x^*), \bar{g}\phi(\theta^*))\delta(da; \pi(g_\circ), x^*) \\ &\quad \times p(\bar{g}\phi(\theta^*) \mid \pi(g_\circ), x^*)d\nu(g)d\lambda(\theta^*)K(x^*)d\rho(x^*) \\ &\geqq \int_{\mathcal{X}^*}\int_{\Theta^*}\int_{\mathcal{G}}\int_{\mathcal{A}} L(a, (\pi(g_\circ), x^*), \bar{g}\phi(\theta^*))\delta_F(da; \pi(g_\circ), x^*) \\ &\quad \times p(\bar{g}\phi(\theta^*) \mid \pi(g_\circ), x^*)d\nu(g)d\lambda(\theta^*)K(x^*)d\rho(x^*) \\ &= R(\lambda, \delta_F) . \end{aligned}$$

Thus, δ_F is a Bayes invariant procedure with respect to λ.

It seems likely that Theorem 3.1, where applicable, will be easy to apply, at least when ν is known explicitly. The method can be somewhat imprecisely summarized as follows.

Suppose $\mathcal{G}$, $\bar{\mathcal{G}}$, $\hat{\mathcal{G}}$, Θ, $\mathcal{A}$, $\mathcal{X} = G/\mathcal{H} \times \mathcal{X}/\mathcal{G}$, L and $\mathcal{P}$ are given. Assume $P_\theta \in \mathcal{P}$ is dominated by a σ-finite measure $\mu^* \times \rho$, where μ^* is the left invariant, positive, Borel measure on $\mathcal{G}/\mathcal{H}$. Let $p(g^*, x^* \mid \theta)$ be the density of P_θ with respect to $\mu^* \times \rho$, and ϕ denote any $1:1$ mapping of $\Theta^* = \Theta/\bar{\mathcal{G}}$ into Θ which maps points of Θ^* into the orbits they label. Then the Bayes invariant procedure with respect to a prior, λ, on Θ^*, evaluated at $(g^*_\circ, x^*) \in \mathcal{X}$, is that probability measure which minimizes as a function of P,

$$\int_{\Theta^*}\int_{\mathcal{G}}\int_{\mathcal{A}} L(a, (g^*_\circ, x^*), \bar{g}\phi(\theta^*))dP(a) \times p(\bar{g}\phi(\theta^*) \mid g^*_\circ, x^*)d\nu(g)d\lambda(\theta^*) ,$$

where $p(\theta \mid x)$ is the posterior density of θ given $X = x$.

4. Applications

We shall now discuss some examples in which this theorem is applied. The first concerns hypothesis testing and constitutes a special case of some interest in its own right.

Example 1. (Hypothesis Testing): let

$$\mathcal{A}=\{a_1, a_2\} \tag{4.1}$$
$$\Theta=\Theta_1 \cup \Theta_2 ,$$

where Θ_1 and Θ_2 are each nonempty and disjoint, and

$$L(a_i, x, \theta)=\begin{cases} 1, & \theta \notin \Theta_i \\ 0, & \theta \in \Theta_i \end{cases} \tag{4.2}$$

$i=1, 2$, $x \in \mathcal{X}$. Assume $\bar{\mathcal{G}}\Theta_i=\Theta_i$, $i=1, 2$. Let Θ_i^* denote $\Theta_i/\bar{\mathcal{G}}$ so that

$$\Theta^*=\Theta_1^* \cup \Theta_2^* . \tag{4.3}$$

In this case expression (3.15) can be written

$$\inf_{0\leq p\leq 1}\Big\{p \int_{\Theta_1*}\int_{\mathcal{G}} p(g_\cdot^*, x^* \mid \bar{g}\phi(\theta^*))\Delta(g_\cdot)/K(x^*)d\nu(g) \tag{4.4}$$
$$\times d\lambda(\theta^*)+(1-p)\int_{\Theta_2*}\int_{\mathcal{G}} p(g_\cdot^*, x^* \mid \bar{g}\phi(\theta^*))\Delta(g_\cdot)/K(x^*)$$
$$\times d\nu(g)d\lambda(\theta^*)\Big\} ,$$

where λ is the prior probability distribution on $(\Theta^*, s(\Theta^*))$. This infimum is attained by $p=1$ or $p=0$ according as the fuction $T: \mathcal{X} \rightarrow [0, \infty]$, defined by

$$T(g_\cdot^*, x^*)=\int_{\mathcal{G}}\int_{\Theta_2*} p(g_\cdot^*, x^* \mid \bar{g}\phi(\theta^*))d\lambda(\theta^*)d\nu(g)\Big/ \tag{4.5}$$
$$\int_{\mathcal{G}}\int_{\Theta_1*} p(g_\cdot^*, x^* \mid \bar{g}\phi(\theta^*))d\lambda(\theta^*)d\nu(g) ,$$

is greater than 1 or less than 1. When $T(g_\cdot^*, x^*)=1$, the infimum is attained at any point $p \in [0, 1]$. For definiteness we choose δ_F as

$$\delta_F(a_1; g_\cdot^*, x^*)=\begin{cases} 1, & T(g_\cdot^*, x^*)\leqq 1 \\ 0, & T(g_\cdot^*, x^*)>1 . \end{cases} \tag{4.6}$$

Since $T(\circ, \circ)$ is invariant any specification of δ_F will, of course, be invariant. δ_F is a Bayes invariant procedure with respect to λ by Theorem 3.1.

In the case where Θ_1^* and Θ_2^* both reduce to a point, say θ_i^* in the case of Θ_i^*, $i=1, 2$, T can be written,

$$T(g_\circ^*, x^*)=(1-\pi_1)\int_{\mathcal{G}} p(g_\circ^*, x^* \mid \bar{g}\phi(\theta_2^*))d\nu(g) \Big/ \pi_1 \int_{\mathcal{G}} p(g_\circ^*, x^* \mid \bar{g}\phi(\theta_1^*))d\nu(g) , \tag{4.7}$$

where $\pi_1=\lambda(\theta_1^*)$. One would expect $\pi_1 T(\circ, \circ)/(1-\pi_1)$ to represent the ratio of the densities of the maximal invariant under the respective hypothesis Θ_2 and Θ_1. This turns out to be the case under fairly general circumstances, and represents one, among several theorems which have come to be frequently called Stein's theorem. Versions of this theorem are given by Schwartz [4] and Wijsman [7].

In the present context, this theorem is easily proved. For if $f(\circ \mid \phi(\theta_i^*))$ denotes the density of the maximal invariant under Θ_i, $i=1, 2$, and $g_\circ \in \pi^{-1}(g_\circ^*)$,

$$\begin{aligned} f(x^* \mid \phi(\theta_i^*)) &= \int_{\mathcal{G}} p(\pi(g), x^* \mid \phi(\theta_i^*))d\mu(g) \\ &= \Delta(g_\circ)\int_{\mathcal{G}} p(\pi(g^{-1}g_\circ), x^* \mid \phi(\theta^*))d\nu(g) . \end{aligned}$$

Thus

$$f(x^* \mid \phi(\theta_i^*))=\Delta(g_\circ) \int_{\mathcal{G}} p(g_\circ^*, x^* \mid \bar{g}\phi(\theta_i^*))d\nu(g) , \tag{4.8}$$

and the asserted result follows.

Example 2. (Bayes Scale Invariant Estimators). Here

$$\begin{aligned} &\Theta=\mathcal{A}=(-\infty, \infty)\times(0, \infty) \\ &L(a, x, \theta)=h[(a_1-\mu)/\sigma, (a_2-\sigma)/\sigma] , \end{aligned} \tag{4.9}$$

where h is a nonnegative measurable function, $a=(a_1, a_2)$, $\theta=(\mu, \sigma)$, and

$$p(x \mid \theta)=f((x-\mathcal{E}\mu)/\sigma) , \tag{4.10}$$

where $x \in R^n=\mathcal{X}$, .and $\mathcal{E}=(1, 1, \cdots, 1) \in R^n$. The problem remains invariant under the multiplicative group, $\mathcal{G}$, of positive real numbers under which

$$x \rightarrow cx , \qquad (\mu, \sigma)\rightarrow(c\mu, c\sigma) , \qquad (a_1, a_2)\rightarrow(ca_1, ca_2) . \tag{4.11}$$

Observe

$$\mathcal{X}=\mathcal{G}\times R^{n-1}\times\{-1, 0, 1\} , \tag{4.12}$$

where $R^{n-1}\times\{-1,0,1\}$ constitutes the range of the maximal invariant which is $(x_2/s,\cdots,x_n/s,\operatorname{sgn}x_1)$ when $X=(x_1,\cdots,x_n)$ is observed, where $s=(\sum x_i^2)^{1/2}$ and $\operatorname{sgn}x=-1,0,1$ according as x is negative, zero, or positive. $\Theta/\bar{\mathcal{G}}=(-\infty,\infty)$ and a Bayes invariant procedure is obtained by minimizing, for each x,

$$\int_{-\infty}^{\infty}\int_0^{\infty}h\left(\frac{a_1}{c}-\eta,\frac{a_2}{c}-1\right)f\left(\frac{x}{c}-\mathcal{E}\eta\right)\frac{dc}{c}d\lambda(\eta)\,, \tag{4.13}$$

as a function of a_1 and a_2. In obtaining the expression, the mapping ϕ, of $\Theta/\bar{\mathcal{G}}$ into Θ has been taken to be

$$\phi(\eta)=(\eta,1)\,. \tag{4.14}$$

In the important special case of quadratic loss, where $h(x,y)=x^2+y^2$, it follows that the Bayes invariant estimators of μ an σ are

$$\int_{-\infty}^{\infty}\int_0^{\infty}\frac{\eta}{\sigma}f\left(\frac{x}{\sigma}-\mathcal{E}\eta\right)\frac{d\sigma}{\sigma}d\lambda(\eta)/D \tag{4.15}$$

and

$$\int_{-\infty}^{\infty}\int_0^{\infty}\frac{1}{\sigma}f\left(\frac{x}{\sigma}-\mathcal{E}\eta\right)\frac{d\sigma}{\sigma}d\lambda(\eta)/D \tag{4.16}$$

where

$$D=\int_{-\infty}^{\infty}\int_0^{\infty}\frac{1}{\sigma^2}f\left(\frac{x}{\sigma}-\mathcal{E}\eta\right)\frac{d\sigma}{\sigma}d\lambda(\eta)\,. \tag{4.17}$$

Example 3. (Selection and Ranking Problems). In selection and ranking problems, the reduction to invariant procedures seems quite natural. The results of Theorem 3.1 apply to this class of problems and, as an example, we treat the selection problem described by Studden [6] and obtain a slight generalization of a result given by him.

In this example $\mathcal{X}$ and Θ are arbitrary subsets of R^k which remain invariant under the permutation group $\mathcal{G}$. In a somewhat imprecise notation, assume, for $g\in\mathcal{G}$,

$$g\{i_1,\cdots,i_m\}=\{g(i_1),\cdots,g(i_m)\}\,, \tag{4.18}$$

and

$$gy=(y_{h1},\cdots,y_{hk})\,,\qquad y\in R^k\,, \tag{4.19}$$

where $i_r\in(1,2,\cdots,k)$, $r=1,2,\cdots,m$, and $h=g^{-1}$.

The action space $\mathcal{A}$ is the set of all 2^k subsets of $\{1,2,\cdots,k\}$, the objective of the statistician being to select a subset containing that integer for which the corresponding co-ordinate of θ is largest among its k co-ordinates.

For this problem, $\mathcal{G}=\bar{\mathcal{G}}=\hat{\mathcal{G}}$,

$$\Theta^*=\{\theta\in\Theta:\theta_1\leqq\cdots\leqq\theta_k\}\,,\qquad \mathcal{X}^*=\{x\in\mathcal{X}:x_1\leqq\cdots\leqq x_k\}\,,$$

and

$$\mathcal{X}\simeq\mathcal{G}\times\mathcal{X}^*\,. \tag{4.20}$$

The loss function, L, is given by

$$L(g\theta^*, a)=\sum_{i\in a}L_i(g\theta^*)+L[1-\chi_a(g(k))]\,, \tag{4.21}$$

where $\theta^*=(\theta_1^*,\cdots,\theta_k^*)\in\Theta^*$, $L_i(\theta)$ is the loss whenever the integer i is included in the statistician's selection, $\chi_a(\circ)$ is the characteristic function of $a\subset\mathcal{A}$, and L is a given positive constant which represents the loss for an incorrect selection.

Suppose $p^*(x_1,\cdots,x_k\,|\,\theta)$ denotes the density, with respect to a σ-finite measure ψ on $s(\mathcal{X})$, of a random variable X taking its values in $\mathcal{X}$. So that the problem may remain invariant under $\mathcal{G}$, we require ψ to be invariant under $\mathcal{G}$,

$$p(gx\,|\,g\theta)=p(x\,|\,\theta) \tag{4.22}$$

and

$$L_i(\theta)=L_{gi}(g\theta)\,,\qquad g\in\mathcal{G}\,. \tag{4.23}$$

Let Φ denote the isomorphic mapping of $\mathcal{X}$ onto $\mathcal{G}\times\mathcal{X}^*$. Then the random variable, $\Phi(X)$, has density $p(\circ,\circ\,|\,\theta)$ defined by

$$p(g, x^*\,|\,\theta)=p^*(x_{g1}^*,\cdots,x_{gk}^*\,|\,\theta)\,, \tag{4.24}$$

$g\in\mathcal{G}$, $x^*\in\mathcal{X}^*$, almost everywhere with respect to the measure $\mu\times\psi$ on $s(\mathcal{G})\times s(\mathcal{X}^*)$, where μ denotes the measure which assigns unit mass to each point of $\mathcal{G}$ and $s(\mathcal{G})$ is the set of all subsets of $\mathcal{G}$.

Let Ξ denote the class of all probability distributions on the set of all subsets of $\mathcal{A}$, and Ξ_1, the set $\{(\xi_1,\cdots,\xi_k):0\leqq\xi_i\leqq 1,\ i=1,2,\cdots,k\}$. According to Theorem 3.1, the Bayes invariant procedure with respect to a prior measure η on Θ^*, evaluated at $y_\circ=(g_\circ, x^*)$, is that member of Ξ which minimizes, as a function of ξ,

$$\sum_{a\in\mathcal{A}}\sum_{g}\int_{\Theta^*}\xi(a)\Big\{\sum_{i\in a}L_i(g\theta^*)-L\chi_a(g(k))\Big\}\times p(y_\circ\,|\,g\theta^*)d\eta(\theta^*)\,. \tag{4.25}$$

But the infimum of this last expression over Ξ is

$$\inf_{\Xi} \sum_{i=1}^{k} \sum_{g} \sum_{\substack{a \in \mathcal{A}: \\ i \in a}} \int_{\Theta^*} \xi(a)\{L_i(g\theta^*) - L\chi_{\{i\}}(g(k))\} \times p(y_\circ \mid g\theta^*) d\eta(\theta^*) = \inf_{\phi \in \Xi_1} \sum_{i=1}^{k} \phi(i) \int_{\Theta^*} \left\{ \sum_{j=1}^{k} L_j(\theta^*) p_{ij}(y_\circ \mid \theta^*) - L p_{ik}(y_\circ \mid \theta^*) \right\} \times d\eta(\theta^*) , \tag{4.26}$$

where

$$p_{ij}(y_\circ \mid \theta^*) = \sum_{g : g^{-1} i = j} p(y_\circ \mid g\theta^*) , \qquad i, j = 1, \cdots, k . \tag{4.27}$$

Let

$$T_i(y_\circ) = \sum_j \int_{\Theta^*} L_j(\theta^*) p_{ij}(y_\circ \mid \theta^*) d\eta(\theta^*) . \tag{4.28}$$

Then, it follows from equation (4.26), that if $\phi(y_\circ) = (\phi_1(y_\circ), \cdots, \phi_k(y_\circ))$ is the Bayes invariant procedure with respect to η,

$$\phi_i(y_\circ) = \begin{cases} 1 , & T_i(y_\circ) < L \int_{\Theta^*} p_{ik}(y_\circ \mid \theta^*) d\eta(\theta^*) \\ 0 , & \text{otherwise.} \end{cases} \tag{4.29}$$

In determining this procedure, it suffices to compute $\phi_k(y_\circ)$, since the remaining ϕ_i are determined by the invariance of ϕ. Furthermore, since $\mathcal{X} \simeq \mathcal{G} \times \mathcal{X}^*$, the procedure can readily be stated in terms of x. Making the obvious change in the domains of T_k and $p_{kk}(\circ, \circ \mid \theta^*)$,

$$\phi_k(x) = \begin{cases} 1 , & T_k(x) < L \int_{\Theta^*} p_{kk}(x \mid \theta^*) d\eta(\theta^*) \\ 0 , & \text{otherwise.} \end{cases} \tag{4.30}$$

A special case to which this result applies is the "random slippage problem." Here $\theta_j^* = \theta_\circ$, $j = 1, 2, \cdots, k-1$, where $\theta_\circ$ is a specified constant while $\theta_k^* = \theta_\circ + \Delta$, where Δ is a positive random variable with distribution η. It is assumed $L_i \equiv 1$, $i = 1, \cdots, k$, and

$$p(x \mid \theta) = \sum_{i=1}^{k} C(\theta_i) e^{\theta x_i} . \tag{4.31}$$

It is easily shown that the Bayes invariant procedure, ϕ, is specified by

$$\phi_k(x) = \begin{cases} 1 , & T^*(x) < L \int_0^\infty C(\theta_\circ + \delta) e^{\delta x_k} d\eta(\delta) \\ 0 , & \text{otherwise,} \end{cases} \tag{4.32}$$

where

$$T^*(x)=\sum_{j=1}^{k}\int_0^\infty C(\theta_\circ+\delta)e^{\delta x_j}d\eta(\delta)\,. \tag{4.33}$$

As Studden points out, ϕ can be interpreted as that procedure which minimizes, among all invariant procedures, the expected size of the selected subset, subject to the condition that the (unconditional) probability of correct selection is at least as great as that for ϕ.

5. Acknowledgements

The author is indebted to Professor C. Stein who suggested the problem considered here and pointed out several improvements on the original solution.

Most of the work reported in this paper was supported by the Office of Naval Research Contract, NONR 225 (72) and appeared in the author's Ph.D. thesis written at Stanford University.

UNIVERSITY OF BRITISH COLUMBIA

REFERENCES

[1] J. Keifer, "Invariance, minimax sequential estimation, and continuous time processes," *Ann. Math. Statist.*, 28 (1957), 573-601.

[2] L. H. Loomis, *An Introduction to Abstract Harmonic Analysis*, D. Van Nostrand Co., Inc., New York, 1953.

[3] Leopold Nachbin, *The Haar Integral*, D. Van Nostrand Co., Inc., New York, 1965.

[4] R. Schwartz, "Properties of invariant multivariate tests," Ph.D. Thesis, Cornell University, 1966.

[5] Charles Stein, "Approximation of improper prior measures by prior probability measures," *Bernoulli, Baves, Laplace, Anniversary Volume*, Springer Verlag Inc., New York, 1965.

[6] W. J. Studden, "On selecting a subset of k populations containing the best," *Ann. Math. Statist.*, 38 (1967), 1072-1078.

[7] R. A. Wijsman, "Cross-sections of orbits and their application to densities of maximal invariants," *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*

[8] J. V. Zidek, "On the admissibility of formal Bayes estimators," Ph.D. Thesis, Stanford University, 1967 (appeared as Tech. Rep. No. 58, Nonr-225(72), Dept. of Statistics, Stanford University).