EQUIVARIANT ESTIMATORS AND A SPECIAL GROUP STRUCTURE

by

Thomas H. Woteki

Dissertation submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
Statistics

APPROVED:

Dr. Lawrence S. Mayer, Chairman
Department of Statistics

Dr. Raymond H. Myers
Department of Statistics

Dr. Arlo D. Hendrickson
Department of Statistics

Dr. Charles J. Parry
Department of Mathematics

Dr. Ezra A. Brown
Department of Mathematics

April, 1974

Blacksburg, Virginia

ACKNOWLEDGEMENTS

The author would like to express his thanks to the following people for their help and encouragement throughout the preparation of this manuscript:

To Dr. Lawrence S. Mayer, the author's Research Adviser, without whose inspiration, enthusiasm and confidence this dissertation would not have been possible.

To the author's Graduate Committee for their time and interest.

To Dr. Stanley G. Wayment for his patience and for providing clerical help and other such services related to the preparation of this manuscript.

To Catherine E. Woteki, the author's wife, for her encouragement and understanding throughout the author's career as a graduate student.

Finally, to Mrs. Susan Gonzales, who quickly and accurately typed this manuscript, for her cooperation and always pleasant attitude in the face of numerous corrections and revisions.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION AND REVIEW OF LITERATURE

Section 1.1 Introduction

The principle of defining a group of transformations on a sample space together with a family of distributions such that the group carries the probability structure in a natural way has been considered by several investigators. In this dissertation we adopt this principle which leads to the consideration of equivariant estimators, discussed in Chapter V, and maximal invariant functions together with some associated concepts, discussed in Chapter VI.

In Chapter II we present some definitions and results from group theory which will be helpful in later chapters. In particular, the notion of semi-direct product is discussed. Mac Lane and Birkhoff [10] and Hall [6] are suggested as references.

The ideas and definitions upon which the principle of invariance is based are discussed in Chapter III. For detailed discussions of this concept see Ferguson [2], Lehmann [9] and Zacks [16].

In Chapter IV we introduce the Haar measure, a measure defined on a locally compact topological group having certain properties with respect to the group. We use the Haar integral only as a tool to obtain, in a natural way, densities for maximal invariants. For other applications of the Haar integral in the literature, the reader is referred to Fraser, [3], [4], [5]; Hora and Buehler, [8]; Lehmann, [9]; Wijsman, [14]; Zacks, [16]. Nachbin, in his presentation in [12], discusses the existence (and uniqueness, up to a constant of proportionality) of the Haar measure and develops a theory for the Haar integral.

In Chapter V, discussion centers on an invariant family of distributions, the goal being to characterize equivariant estimators under suitable hypotheses. First we present the work of Berk [1] who characterized equivariant estimators under certain hypotheses. We relax those hypotheses and obtain the desired characterization.

In obtaining the above characterization we also gain information concerning maximal invariants and global cross-sections, two ideas which were considered in detail by Wijsman [14]. In particular, Wijsman made certain assumptions concerning the group of transformations in order to obtain a global cross-section, maximal invariant and a density for the maximal invariant. In Chapter VI we consider an invariant family of distributions similar to that introduced in Chapter V and obtain results comparable to Wijsman's. Then, we use the density of the maximal invariant to develop a test of hypothesis concerning a coset of the true parameter, a problem considered briefly by Fraser [5]. Related to the problem of testing the coset of the parameter is that of estimating the coset or orbit. In Chapter VII we discuss this in the context of invariant estimation. Using results from Chapters V and VI we develop a loss function for invariant estimators.

Section 1.2 Review of Literature

As mentioned previously, both Lehmann [9] and Ferguson [2] will serve as general introductions to the principle of invariance. In particular, see Chapters I and VI of the former and Chapters I and IV of the latter. Zacks [16] consolidates some of the results presented in the literature and also provides a general introduction to invariance; of particular note are Chapters I and VII.

Fraser [3], [4] considered a group structure on a probability space and used these concepts to introduce the structural distribution and the pivotal quantity. Later, Fraser [5] considered in detail some of the ideas suggested in his earlier work.

As indicated in Section 1.1, Wijsman [14] discussed the relation between the group structure, maximal invariants and global cross-sections. He also obtained a density for the maximal invariant.

Hora and Buehler [8], under assumptions similar to Fraser's, introduced the fiducial distribution. Moreover, they characterized the form of invariant loss functions and showed that for those loss functions the best equivariant estimator is also best with respect to the fiducial distribution.

As mentioned in Section 1.1, Berk [1] obtained a characterization of equivariant estimators while Staudte [13] obtained a characterization of invariant loss functions similar to Hora-Buehler above.

Finally, Younger [15] used methods similar to those discussed in Chapter VII applied to linear models to obtain an estimator of the standardized regression coefficient. This estimator was shown to have certain desirable properties when compared to the usual estimator.

CHAPTER II

A REVIEW OF GROUP THEORY

The concept of a group of transformations is fundamental to the notion of invariant and equivariant estimation. Herein we review some elementary concepts of group theory and in particular those topics which will be used in our work on equivariant estimation. The reader is referred to Hall [6] and Mac Lane and Birkhoff [10]. We begin with the definitions of group and subgroup.

Section 2.1 Groups and Subgroups

Definition 2.1. A group is a non-empty set K together with a binary operation $K \times K \to K$ written $(k,k') \to kk'$ such that:

i) This operation is associative. That is $k_1(k_2k_3) = (k_1k_2)k_3$.

ii) There is an element $e \in K$ with $ek = k = ke$ for all $k \in K$. That is, e is an identity element of K.

iii) For this element e, there is to each element $k \in K$ an element $k^{-1} \in K$ with $kk^{-1} = e = k^{-1}k$. That is every element of K has an inverse in K.

Direct applications of the definitions establish that the identity element e is unique. Similarly, for each $k \in K$ the element k^{-1} is unique and is referred to as the inverse of k. The element kk' is referred to as the product of k and k' and hence K is referred to as a multiplicative group.

Much information about a group can be obtained from a knowledge of its subgroups.

Definition 2.2. A subset G of a multiplicative group K is closed under the operation of multiplication in K provided that $g, g' \varepsilon G$ imply that $gg' \varepsilon G$.

Definition 2.3. A subset G of a multiplicative group K is said to be a subgroup of K, provided that G is closed under the operation of multiplication and is itself a group with respect to the restriction of the multiplication in K to elements of G.

The definition amounts to saying that G is a subgroup of K if and only if

i) $g, g' \varepsilon G$ imply $gg' \varepsilon G$.

ii) e, the identity element of K, is an element of G.

iii) $g \varepsilon G$ implies $g^{-1} \varepsilon G$.

See for example Mac Lane and Birkhoff [10].

Section 2.2 Normal Subgroups

The normal subgroups of a group play an important part in the theory. In particular, in order that we may define a "natural" group structure on the quotient of K by H, K/H, it is sufficient that H be a normal subgroup of K. Normal subgroups are also important in the development of some of the later chapters.

Definition 2.4. A subgroup H of K is said to be normal in K, denoted $H < K$, provided that $khk^{-1} \varepsilon H$ whenever $k \varepsilon K$ and $h \varepsilon H$.

The following elementary ideas follow directly from Definition 2.4. See for example Mac Lane and Birkhoff [10].

Proposition 2.1. H is a normal subgroup of K if and only if $k^{-1} hk \varepsilon H$ whenever $k \varepsilon K$ and $h \varepsilon H$.

Proposition 2.2. H is a normal subgroup of K if and only if for every $h \varepsilon H$ and $k \varepsilon K$ there is an $h' \varepsilon H$ such that $kh = h'k$.

It is well known that an equivalence relation E defined on a set X partitions X into mutually disjoint equivalence classes. That is, if $x \varepsilon X$ then $\{y \varepsilon X | y E x\}$, where $y E x$ denotes y is equivalent to x under E, is the equivalence class containing x. The set of all equivalence classes partitions X. Now if K is a group and G a subgroup, we may define an equivalence relation on K as follows: for $k_1, k_2 \varepsilon K$, k_1 is equivalent to k_2 provided there exists a $g \varepsilon G$ such that $k_1 = gk_2$ or, in other words, provided $k_1k_2^{-1} \varepsilon G$.

Definition 2.5. If G is a subgroup of a group K then the left coset of $k \varepsilon K$ is the set $Gk = \{gk | g \varepsilon G\}$.

It is easily seen that the collection of all left cosets Gk, $k \varepsilon K$, is the set of equivalence classes for the equivalence relation of the preceeding paragraph so that k_1 is equivalent to k_2 if and only if $Gk_1 = Gk_2$.

In a similar way we could define right cosets $kG = \{kg | g \varepsilon G\}$ and an equivalence relation: k_1 is equivalent to k_2 provided $k_2^{-1} k_1 \varepsilon G$.

In general $kG \neq Gk$. In fact, a necessary and sufficient condition for $kG = Gk$ for any $k \varepsilon K$ is that $G < K$.

Proposition 2.3. If H is a subgroup of the group K then $H < K$ if and only if every right coset is a left coset. That is $Hk = kH$ for all $k \varepsilon K$.

Definition 2.6. For H a subgroup of K let $K/H = \{Hk|k \varepsilon K\}$ be the set of all left cosets. Furthermore, let $[k]$ denote the coset Hk.

Further consequences of the normality of H in K are that we may impose a natural group structure on K/H as follows:

Proposition 2.4. If K is a group and $H < K$ then K/H is a group under the operation $[k_1][k_2] = [k_1k_2]$ for any $[k_1], [k_2] \varepsilon K/H$.

Detailed discussions of the ideas and results of this section may be found in any of the references cited.

Section 2.3 The Direct Product

Earlier it was stated that the subgroups of a group K can provide information about the structure of K. This is particularly true when K is the direct product of two of its subgroups.

Definition 2.7. The multiplicative group K is said to be the direct product of the subgroups H and G, denoted $K = H \times G$, provided that

i) $G \cap H = \{e\}$.

ii) For each $h \varepsilon H$ and $g \varepsilon G$, $gh = hg$.

iii) For each $k \varepsilon K$ there is an $h \varepsilon H$ and $g \varepsilon G$ such that $k = hg$.

Condition i) of Definition 2.7 states that the only element common to H and G is the identity; ii) says that elements of H and G commute with respect to multiplication and iii) says that any element of K can be factored into the product of an element of H and an element of G. Moreover, conditions i) and iii) imply the factorization is unique as follows:

Proposition 2.5. If K is a group and G and H are subgroups satisfying iii) of Definition 2.7, then $G \cap H = e$ if and only if the factorization guaranteed in iii) is unique.

Proof. Suppose K, G and H satisfy the hypothesis of the theorem and let $G \cap H = \{e\}$. Let $k \varepsilon K$ and suppose $g, g' \varepsilon G$ and $h, h' \varepsilon H$ such that $h'g' = k = hg$. It follows then that $g'g^{-1} = (h')^{-1}h \varepsilon G \cap H$ so that $g'g^{-1} = (h')^{-1}h = e$ and $g = g'$ and $h = h'$. Thus the factorization is unique. Conversely, suppose the factorization is unique and let $k \varepsilon G \cap H$. Then $k \varepsilon H$ and $k = h\,e$ where $h \varepsilon H$ and $e \varepsilon G \subset K$. Moreover, $k \varepsilon G$ so $k = e\,g$ where $e \varepsilon H \subset K$ and $g \varepsilon G$. Since $h\,e = k = e\,g$ and the factorization of k is unique we have $h = e = g$ and so $k = e$. That is $G \cap H = \{e\}$. Q.E.D.

Thus, as a corollary to Proposition 2.5, we have that the factorization in a direct product is unique.

Proposition 2.6. If K is the direct product of H and G then for each $k \varepsilon K$ there is a unique $h \varepsilon H$ and a unique $g \varepsilon G$ such that $k = hg$.

Proof. This result follows directly from Proposition 2.5 and Definition 2.7. Q.E.D.

Condition ii) of the definition of direct product implies that both G and H are normal in K. Subsequently we will see that under certain other additional assumptions the normality of G and H is sufficient for $K = HxG$.

Proposition 2.7. If $K = HxG$ then $H < K$ and $G < K$.

Proof. Let $g \varepsilon G$ and $k \varepsilon K$. $k = st$ for some $s \varepsilon H$ and $t \varepsilon G$ and so $kgk^{-1} = stgt^{-1}s^{-1}$. By ii) of Definition 2.5 $s(tgt^{-1})s^{-1} = ss^{-1}(tgt^{-1}) = e(tgt^{-1}) = tgt^{-1}$ and clearly $tgt^{-1} \varepsilon G$. Therefore, $kgk^{-1} \varepsilon G$ and so $G < K$. Similarly $H < K$.

Section 2.4 The Semi-Direct Product

In view of Proposition 2.7, both of the subgroups involved in a direct product group must be normal and moreover this was seen to be a direct consequence of part ii) of Definition 2.7. We may now ask what structure results when the requirement of commutativity of elements of G and H is dropped. In fact, we will require only that one of the subgroups be normal in K. This leads to the definition of semi-direct product. Not only is this interesting from an algebraic point of view, but we note also that the transformation group of scale and location changes on a sample space can be represented as a semi-direct product. We begin with a formal definition.

(We assume that the reader is familiar with the notions of homomorphisms and isomorphisms of groups. Also, that an automorphism of a group K is simply an isomorphism from K into K and that the collection of all automorphisms of K is itself a group under composition.)

Definition 2.8. Let G and H be groups, let $\mathrm{Aut}(H)$ denote the group of all automorphisms of H and let $\theta : G \to \mathrm{Aut}(H)$ be a homomorphism. By the semi-direct product of H and G relative to θ, denoted by $Hx_\theta G$, we mean the group formed by the collection of ordered pairs

$\{(h,g) | h \varepsilon H, g \varepsilon G\}$ under the binary operation given by

$$(h_1, g_1)(h_2, g_2) = (h_1((\theta g_1)h_2), g_1g_2).$$

That $Hx_\theta G$ is a group under the above operation, with identity (e_H,e_G), where e_H and e_G are the identities of H and G respectively, and inverse of (h,g) given by $((\theta(g^{-1}))h^{-1}, g^{-1})$, is verifiable by direct calculation. See, for example, Hall [6] or Mac Lane and Birkhoff [10]. We make the following observations:

Proposition 2.8. If $K = Hx_\theta G$ then

i) $H' = \{(h,e_G) | h \varepsilon H\}$ is a normal subgroup of K.

ii) $G' = \{(e_H,g) | g \varepsilon G\}$ is a subgroup of K.

iii) $H' \cap G' = \{(e_H,e_G)\}$.

Proof. Verifiable by direct appeal to the appropriate definitions.

The following theorems serve to characterize semi-direct products:

Proposition 2.9. Any group K with subgroups $H < K$ and G satisfying both

i) $H \cap G = \{e\}$ and

ii) For each $k \varepsilon K$ there is $h \varepsilon H$ and $g \varepsilon G$ such that $k = hg$,

is a semi-direct product. Explicitly, if $\theta:G\to Aut(H)$ is defined for each g by $(\theta(g))h = ghg^{-1}$ then an isomorphism $\phi:Hx_\theta G \cong K$ is given by $\phi(h,g) = hg$.

Proof. See Mac Lane and Birkhoff [10].

Proposition 2.10. If a group K is isomorphic to a semi-direct product $H_1x_\theta G_1$ then there exists subgroups H and G of K such that:

i) $H < K$.

ii) $H \cap G = \{e\}$.

iii) For each $k \in K$ there exists an $h \in H$ and $g \in G$ such that $k = hg$.

Proof. Let $\phi : H_1 x_\theta G_1 \cong K$ be the given isomorphism. By Proposition 2.8, $H_1' = \{x,e_G) | x \in H_1\} < H_1 x_\theta G_1$ and $G_1' = \{(e_H,x) | x \in G_1\}$ is a subgroup of $H_1 x_\theta G_1$ and moreover $H_1 \cap G_1 = \{(e_H,e_G)\}$ the identity of $H_1 x_\theta G_1$.

Consider $H = \{k \in K|$ there is a $z \in H_1'$ such that $\phi(z) = k\}$ and $G = \{k \in K|$ there is a $z \in G_1'$ such that $\phi(z) = k\}$.

Now ϕ is an isomorphism so that:

a) G is a subgroup of K because G_1' is a subgroup of $H_1 x_\theta G_1$;

b) $H < K$ because $H_1 < H_1' x_\theta K_1'$ and

c) $H \cap G = \{e_k\}$ because $H_1' \cap G_1' = \{(e_H,e_G)\}$. Thus, i and ii of the conclusion are established.

Now, let $k \in K$. There exists an element $(x,y) \in H_1 x_\theta G_1$ such that $\phi((x,y)) = k$. Now $(x,y) = (x,e_G)(e_H,y)$ so that $(x,e_G) \in H_1'$ and $(e_H,y) \in G_1'$ and $h = \phi((x,e_G)) \in H$ and $g = \phi((e_H,y)) \in G$. Therefore, $k = \phi((x,y)) = \phi((x,e_G)(e_H,y)) = \phi((x,e_G))\,\phi((e_H,y)) = hg$ and we see that if $k \in K$ there exist $h \in H$ and $g \in G$ with $k = hg$ and so, part iii of the proposition is proved. Q.E.D.

In view of Propositions 2.9 and 2.10, if K is a group with subgroups G and $H < K$ satisfying the hypotheses of Proposition 2.9 then we say K is the semi-direct product of H and G and we write $K = H\times_s G$. Recalling our remarks at the beginning of this section we see that the two preceding propositions generalize the definition of direct product, Definition 2.7, having replaced ii) of that definition by i) of Proposition 2.10.

The following two propositions develop some properties of the semi-direct product which we will find useful in the sequel.

Proposition 2.11. If $K = H\times_s G$ is the semi-direct product of two subgroups H and G (so that $H < K$) then for each $k \in K$ there is an $h \in H$ and $g \in G$ such that $k = gh$.

Proof. Let $k \in K$. Since $K = H\times_s G$ there exist $h' \in H$ and $g \in G$ so that $k = h'g$. Now $h'g = g(g^{-1} h'g)$ and $g^{-1}h'g \in H$ since $H < K$. Letting $h = g^{-1}h'g$ we have $k = gh$ as desired. Q.E.D.

Proposition 2.12. If $K = H\times_s G$ then for each $k \in K$ there exist unique elements $g \in G$ and h and $h' \in H$ such that $hg = k = gh'$.

Proof. For each $k \in K$ the existence of the elements $g \in G$ and h and $h' \in H$ is a direct corollary to the definition of $K = H\times_s G$ and Proposition 2.11. The uniqueness of the factorizations follows from Proposition 2.5. Q.E.D.

Note that, in some sense, Proposition 2.12 stipulates two unique factorizations for an element k - a "left hand" and a "right hand" factorization.

The following propositions display the relationship between the semi-direct product and the direct product.

Proposition 2.13. If $K = Hx_sG$ (so that $H < K$) and if also $G < K$ then $K = H \times G$.

Proof. That conditions i) and iii) of Definition 2.7 obtain follows immediately from the definition of $K = Hx_sG$. It remains to show thatii) of Definition 2.7 is fulfilled.

Let $h \in H$ and $g \in G$. Then $hg = k \in K$ and $(hgh^{-1})h = k = g(g^{-1}hg)$. Since both of H and G are normal in K we have $g' = hgh^{-1} \in G$ and $h' = g^{-1}hg \in H$. By Proposition 2.12 the two factorizations $g'h = k = gh'$ are the same. That is $hgh^{-1} = g' = g$ or, equivalently, $hg = gh$. Thus, ii) of Definition 2.7 is satisfied so $K = HxG$. Q.E.D.

Finally, the following proposition notes that any direct product is trivially a semi-direct product.

Proposition 2.14. If $K = HxG$ for subgroups H and G then also $K = Hx_sG$.

Proof. If $K = HxG$ then $H < K$ by Proposition 2.7. Moreover, $H \cap G = \{e\}$ and the factorization of any $k \in K$ is provided for by i) and iii) of Definition 2.7. It follows at once from Proposition 2.9 that $K = Hx_\theta G$ where $\Theta: G \to \mathrm{Aut}(H)$ is such that $\Theta(g)$ is the identity function on H for all $g \in G$. That is, $K = Hx_sG$. Q.E.D.

CHAPTER III

THE PRINCIPLE OF INVARIANCE

Lehmann [9] points out that statistical problems often display certain symmetries which give rise to natural restrictions to impose on the statistical procedures to be employed. The mathematical expression of symmetry is invariance under a group of transformations while the natural restrictions we impose comprise the principle of invariance.

Section 3.1 Invariant Families of Distributions

Virtually all of our discussions will revolve around an invariant family of distributions.

Definition 3.1. A family of distributions is a triple (X,Ω,P) where

i) Ω is a non-empty set called the parameter space.

ii) $P = \{P_\omega | \omega \in \Omega\}$ is a collection of probability distributions defined on a σ-algebra A of subsets of X and indexed by the elements of Ω such that $\omega_1 \neq \omega_2$ implies $P_{\omega_1} \neq P_{\omega_2}$.

With reference to Definition 3.1, we take the σ-algebra A to be the Borel subsets of X, if X carries some topology. We say that ω is the parameter of the distribution P_ω and we use X both to denote the space of a random variable having distribution P_ω and the random variable itself. Moreover, $X \sim P_\omega$ means X has the given distribution.

The following is the basis of the principle of invariance:

Definition 3.2. A G-invariant family of distributions is a family of distributions (X,Ω,P) together with a group of transformations G on

X such that: for every $g \in G$ and $\omega \in \Omega$ there is a unique $\omega^* \in \Omega$ such that if $X \sim P_\omega$ then $gX \sim P_{\omega^*}$.

In Definition 3.2, gX is the random variable which equals gx when $X=x$. Definition 3.2 implies certain measurability conditions for $g \in G$. In fact, we assume that each $g \in G$ is a measurable transformation with respect to A.

Definition 3.2 allows us to define a group of transformations on Ω as follows: for each $g \in G$ let $g^*: \Omega \to \Omega$ be given by $g^*\omega = \omega^*$ if and only if $gX \sim P_{\omega^*}$ when $X \sim P_\omega$. That $G^* = \{g^* | g \in G\}$ forms a group of transformations under the composition $g_1^* g_2^* = (g_1 g_2)^*$ is the subject of the next proposition.

Proposition 3.1. If (X,Ω,P) is a G-invariant family of distributions then $G^* = \{g^* | g \in G\}$ is a group of transformations on Ω. Furthermore, $g \to g^*$ is a homomorphism $G \to G^*$ of groups.

Proof. See Lehmann [9] or Ferguson [2].

We note the following useful fact:

Proposition 3.2. If (X,Ω,P) is a G-invariant family of distributions, and if $A \in A$, $P_\omega \in P$, $\omega \in \Omega$, $g \in G$, all arbitrary, then $P_\omega(A) = P_{g^*\omega}(gA)$.

Proof. Let the hypothesis be satisfied and suppose $X \sim P_\omega$. Then $gX \sim P_{g^*\omega}$ and under the transformation g, A is transformed to gA. Thus $P_\omega(A) = P_\omega(X \in A) = P_{g^*\omega}(g\,X \in g\,A) = P_{g^*\omega}(gA)$ as desired. Q.E.D.

Section 3.2 The Invariance Principle: Equivariant Estimation

Given a G-invariant family of distributions (X,Ω,P), the principle of invariance suggests that we restrict our attention to equivariant estimators and invariant loss functions.

By an estimator we mean a function $\phi: X \to \Omega$. Invariance considerations ($gX \sim P_{g^*\omega}$ whenever $X \sim P_\omega$) suggest that if we estimate the parameter to be ω when x is observed, then we should use the estimate $g^*\omega$ when gx is observed. That is, we should require our estimators to have the property $\phi(gx) = g^*\phi(x)$ for all $g \in G$ and all $x \in X$.

Definition 3.3. Given a G-invariant family of distributions (X,Ω,P), a G-G* equivariant estimator is an estimator ϕ such that $\phi(gx) = g^*\phi(x)$ for all $g \in G$ and $x \in X$.

Again, consideration of the transformation structure leads to the definition of natural loss functions for equivariant estimators: Invariant loss functions. By a loss function we mean a non-negative real valued function L on $\Omega \times \Omega$. Invariant loss functions are those loss functions which have the property:

$$L(g^*\omega, \phi(gx)) = L(g^*\omega, g^*\phi(x)) = L(\omega, \phi(x)).$$

Equivariant estimators have the property that the risk associated with any ϕ under the invariant loss L is constant on the G^* orbits of Ω. Letting

$$R(\omega, \phi) = \int L(\omega, \phi(x))\, dP_\omega(x),$$

we have asserted the following:

Proposition 3.3. If (X,Ω,P) is a G-invariant family of distributions, and if ϕ is a G-G* equivariant estimator and if L is an invariant loss function then $R(g^*\omega, \phi) = R(\omega, \phi)$ for all $g \in G$ and $\omega \in \Omega$.

Proof. Let the hypotheses be satisfied. Now

$$R(g^*\omega, \phi) = \int L(g^*\omega, \phi(x))\, dP_{g^*\omega}(x),$$

so making the change of variable $y = g^{-1}x$ we have

$$\begin{aligned} R(g^*\omega, \phi) &= \int L(g^*\omega, \phi(x))\, dP_{g^*\omega}(x) \\ &= \int L(g^*\omega, \phi(gy))\, dP_{g^*\omega}(gy) \\ &= \int L(g^*\omega, \phi(gy))\, dP_{\omega}(y) \quad \text{by Proposition 3.2} \\ &= \int L(\omega, \phi(y))\, dP_{\omega}(y) \quad \text{by invariance of } L \\ &= R(\omega, \phi) \quad \text{as required.} \end{aligned}$$

Q.E.D.

The term equivariant estimator derives from the fact that the risk function is constant on orbits. That is, $R(g^*\omega, \phi) = R(\omega, \phi)$ as in Proposition 3.3.

Section 3.3 The Invariance Principle: Invariant Tests

Just as invariance suggests that we restrict attention to equivariant estimators it also prescribes that invariant tests be used to test certain hypotheses. See Lehmann [9].

Definition 3.5. Given a family of distributions (X,Ω,P) and a hypothesis $H: \omega \in \Omega_H$ vs $K: \omega \in \Omega_K$, a test for H vs K is a function $\phi: X \to [0, 1]$ such that, having observed $X=x$, the probability with which we reject H is $\phi(x)$.

Definition 3.6. Given a family of distributions (X,Ω,P) and a test ϕ for $H: \omega \varepsilon \Omega_H$ vs $K: \omega \varepsilon \Omega_K$ then we say

i) ϕ is a size α test provided that $\int\phi(x)\, dP_\omega(x) \leqslant \alpha$ for all $\omega \varepsilon \Omega_H$.

ii) The power of ϕ at $\omega \varepsilon \Omega_K$ is $\beta_\phi(\omega) = \int\phi(x)\, dP_\omega(x)$

The following definition describes a certain type of hypothesis which leads to another application of the invariance principle.

Definition 3.7. Given a G-invariant family of distributions (X,Ω,P), the hypothesis $H: \omega \varepsilon \Omega_H$ vs $K: \omega \varepsilon \Omega_K$ is said to be G-invariant provided that both $g^*\Omega_H = \Omega_H$ and $g^*\Omega_K = \Omega_K$ for all $g \varepsilon G$.

Recalling again that $gX \sim P_{g^*\omega}$ when $X \sim P_\omega$ and given a G-invariant hypothesis, adoption of the invariance principle suggests that we consider only those tests ϕ which are G-invariant. That is, those tests for which $\phi(gx) = \phi(x)$ for all $g \varepsilon G$, $x \varepsilon X$.

CHAPTER IV

HAAR MEASURE

We have mentioned that Chapter VI will include a discussion of densities of maximal invariants. These densities will be obtained with respect to Haar measure. Haar measure is a measure μ on the Borel sets of a locally compact topological group G such that $\mu(gB) = \mu(B)$. Nachbin [12] gives a complete account of Haar measure and integration.

Section 4.1 Haar Measure and Differentials

Definition 4.1. A Haar measure μ on a locally compact topological group G is a measure defined on the Borel sets of G having the property that $\mu(gB) = \mu(B)$ for all $g \varepsilon G$ and measurable B.

The measure μ is often referred to as left invariant or left Haar measure. Analogous to Definition 4.1 we may define right Haar measure. Nachbin [12] shows that a left Haar measure and a right Haar measure are related by a function called the modular function. We need only consider left Haar measure which we refer to as Haar measure. Nachbin also shows that if G is a locally compact topological group, then there exists a Haar measure on G and this measure is unique up to a positive constant of proportionality.

Nachbin goes on to develop an integral with respect to Haar measure, denoted $\int f(g)d\mu(g)$ where f is a (real-valued) integrable function on G. It follows from the definition of Haar measure that, under the transformation $k = tg$, $t \varepsilon G$ arbitrary but fixed, we have:

$$\int f(g)d\mu(g) = \int f(t^{-1}k)d\mu(t^{-1}k) = \int f(t^{-1}k)d\mu(k).$$

That is, $d\mu(tg) = d\mu(g)$ for all $t \in G$. We refer to $d\mu(g)$ as an invariant Haar differential.

Section 4.2 Haar Measure on Semi-Direct Products

The following result arises in the development of the Haar integral. We also use it to obtain an invariant measure on a semi-direct product.

Theorem 4.1. Let μ be a Haar measure on K and let a be a topological automorphism of K. There exists a unique positive real number $\delta(a)$ such that

$$\int f(a^{-1}(x))d\mu(x) = \delta(a) \int f(x)d\mu(x)$$

for all integrable f. Moreover, $a \to \delta(a)$ is a homomorphism from the group of topological automorphisms of K into the multiplicative group of positive real numbers.

Proof. See Nachbin [12].

Recall Definition 2.8 and suppose $Hx_\theta G$ is the semi-direct product of H and G relative to θ. If, in addition, H and G are topological groups and $Hx_\theta G$ is endowed with the product topology and $\theta(g)$ is a topological automorphism of H for each $g \in G$ then $Hx_\theta G$ is called a topological semi-direct product. If $Hx_\theta G$ is a topological semi-direct product and μ is a Haar measure on H then by Theorem 4.1 there is a certain homomorphism $\theta(g) \to \delta(\theta(g))$ defined for all $g \in G$. It is this homomorphism δ that we refer to in the following.

Theorem 4.2. Let $K = Hx_\theta G$ be the topological semi-direct product of H and G with respect to θ. If $d\mu(h)$ and $d\nu_1(g)$ are invariant

Haar differentials on H and G respectively then

$$d\lambda(k) = \frac{d\mu(h)d\nu_1(g)}{\delta(\theta(g))}, \quad k = (h,g),$$

is an invariant Haar differential on K.

Proof. See Nachbin [12].

Let us write $d\nu(g) = d\nu_1(g) / \delta(\theta(g))$ so that $d\lambda(k) = d\mu(h)d\nu(g)$. Since $d\nu_1(g)$ is an invariant differential on G it follows that

$$\begin{aligned} d\nu(g_1g) &= d\nu_1(g_1g) / \delta(\theta(g_1g)) \\ &= d\nu_1(g) / \delta(\theta(g_1)\theta(g)) \\ &= \frac{d\nu_1(g)}{\delta(\theta(g))\delta(\theta(g_1))} = d\nu(g) / \delta(\theta(g_1)). \end{aligned}$$

Recall from Definition 2.8 that $g_1k = (1,g_1)(h,g) = (\theta(g_1)h,g_1g)$. Since the differential $d\lambda(k)$ in Theorem 4.2 is K-invariant, it is also G-invariant. Thus

$$\begin{aligned} d\mu(h)d\nu(g) &= d\lambda(k) \\ &= d\lambda(g_1k) \\ &= d\mu(\theta(g_1)h)d\nu(g_1g) = \frac{d\mu(\theta(g_1)h)d\nu(g)}{\delta(\theta(g_1))} \end{aligned}$$

and we conclude that $d\mu(\theta(g_1)h) = \delta(\theta(g_1))d\mu(h)$ for all $g_1 \in G$.

We remark that if we are considering a G-invariant family of distributions (X,Ω,P), and if X is isomorphic to G, then it seems natural to try to find densities for the members of P with respect to a Haar measure on $X=G$. See Fraser [3], [4], [5]; Hora and Buehler [8]; Lehmann [9]; Wijsman [14] and Zacks [16].

CHAPTER V

EQUIVARIANT ESTIMATION AND SPECIAL GROUP STRUCTURES

R. H. Berk [1] gave necessary and sufficient conditions for the existence of G-G^* equivariant estimators given the G-invariant family (X,Ω,P). Moreover, he characterized the form of these estimators provided the group G is a direct factor of a group of transformations K on X such that K is transitive on X. In this chapter we show that it is sufficient for G to be a semi-direct factor of a group of transformations K on X such that K is transitive on X.

Section 5.1 Existence of Equivariant Estimators

Definition 5.1. Let (X,Ω,P) be a G-invariant family of distributions, for $y \in X$ let $\Omega_y = \{\omega \in \Omega \mid (G_y)^*\omega = \omega\}$ where $G_y = \{g \mid gy = y\}$.

Theorem 5.1. Let (X,Ω,P) be a G-invariant family of distributions. There exist G-G^* equivariant functions $\phi: X \to \Omega$ if and only if for any $y \in X$, $\Omega_y \neq \phi$.

Proof. See Berk [1].

Theorem 5.1 is the basic existence theorem. Theorem 5.2 follows as a corollary:

Theorem 5.2. Let (X,Ω,P) be a G-invariant family of distributions and $y \in X$. There exist G-G^* equivariant functions $\phi: Gy \to \Omega$ if and only if $\Omega_y \neq \phi$. Moreover, ϕ is determined by its value at any point of Gy, say y, as follows: $\phi(gy) = g^*\phi(y)$ where $\phi(y) \in \Omega_y$.

Proof. Let (X,Ω,P) be a G-invariant family of distributions and $y \in X$.

Suppose $\phi: Gy \to \Omega$ is a G-G^* equivariant function. Consider

$\phi(y)$ and let $g \in G_y$. Then $g^*\phi(y) = \phi(gy) = \phi(y)$. That is, $(G_y)^*\phi(y) = \phi(y)$ or $\phi(y) \in \Omega_y \neq \phi$. Moreover, ϕ is determined by its value at the point y for if $gy \in Gy$ then $\phi(gy) = g^*\phi(y)$.

Conversely, suppose $\Omega_y \neq \phi$ and let $\omega_o \in \Omega_y$. Define $\phi: Gy \to \Omega$ to be $\phi^*(gy) = g^*\omega_o$. Note that if $g_1G_y = g_2G_y$ then $(g_2^{-1}g_1)^* = (g_2^*)^{-1}g_1^* \in (G_y)^*$ so that $g_1^*\omega_o = g_2^*\omega_o$. That is, ϕ is well defined. Moreover, ϕ is G-G* equivariant by construction. Q.E.D.

Section 5.2 Equivariant Estimators and the Direct Product

Berk [1] discusses "a special group structure for which the characterization" given in Theorem 5.1 "assumes a particularly interesting form." To wit:

Let (X,Ω,P) be a G-invariant family of distributions and suppose there is a further group H of transformations of X such that:

i) Every element of G commutes with every element of H.

ii) If $g \in G$, $h \in H$, $x \in X$ such that $gx = hx$ then $gx = x$.

iii) The group $K = GH$ is transitive on X.

Now, if, in addition to i and iii above, we also have that $H \cap G = \{e\}$, where e is the identity transformation on X then K is the direct product of G and H. Berk says in fact that ii is a strong condition for $G \cap H = \{e\}$. Strictly speaking, this is not true unless G is effective on X: Suppose $t \in G \cap H$. Then if $x \in X$, $tx = tx$ implies $tx = x$ by ii. That is $tx = x$ for every $x \in X$. Now if G is effective, we may conclude $t = e$. This leads us to a reformulation of the above.

Definition 5.2. The G-invariant family of distributions (X,Ω,P) is said to have a direct product structure with respect to (K,H) provided that K is a group of transformations on X which is transitive on X and H is a subgroup of K such that

i) K is the direct product of G and H.

ii) If $g \in G$, $h \in H$ and $x \in X$ such that $gx = hx$ then $gx = x$.

We are now ready to prove Berk's main theorem.

Theorem 5.3. Let (X,Ω,P) be a G-invariant family of distributions with a direct product structure with respect to (K,H) and let $y \in X$. To every G-G* equivariant function $\phi:X\to\Omega$ there corresponds a function $\phi_y:Hy\to\Omega_y$ such that at $X = ghy$, $\phi(x) = g^*\phi_y(hy)$. Conversely, if $\Omega_y \neq \phi$ then any function $\phi_y:Hy\to\Omega_y$ determines a G-G* equivariant function $\phi:X\to\Omega$ such that at $X = ghy$, $\phi(x) = g^*\phi_y(hy)$.

Proof. Let $\phi:X\to\Omega$ be a G-G* equivariant function. Define ϕ_y on Hy to be the restriction of ϕ to Hy. That is $\phi_y(z) = \phi(z)$, $z \in Hy$. It remains to see that $\phi_y(z) \in \Omega_y$. Let $g \in G_y$. Since $z = hy$ for some $h \in H$, we see

$$g^*\phi_y(z) = g^*\phi_y(hy) = g^*\phi(hy) = \phi(ghy) =$$
$$\phi(hgy) = \phi(hy) = \phi_y(hy) = \phi_y(z).$$

Therefore, $\phi_y:Hy\to\Omega_y$ as required.

Conversely, suppose $\Omega_y \neq \phi$ and let $\phi_y:Hy\to\Omega_y$. Define $\phi:X\to\Omega$ as follows: For $x \in X$ let $k = gh \in K$ so that $x = ghy$ and let

$\phi(x) = g^*\phi_y(hy)$. If $g_1h_1y = k_1y = x = ghy$ then by commutativity of G and H and ii of Definition 5.2 we have $h_1y = hy$ and $g^{-1}g_1\varepsilon G_y$. Thus $\phi_y(hy) = \phi_y(h_1y)$ and $g^*\phi_y(hy) = g_1^*\,\phi_y(h_1y)$ since $\phi_y(hy)\varepsilon\Omega_y$.

Q.E.D.

We note here the importance of Part ii of Definition 5.2 and also that of the commutativity of G and H in the proof of Theorem 5.3.

Section 5.3 Measurability and the Direct Product Structure

At this point we will consider conditions under which the G-G^* equivariant functions specified in Theorem 5.3 are measurable. Thus, we will introduce into consideration σ-fields A and B for X and Ω respectively in the G-invariant family of distributions (X,Ω,P).

The following theorem will be useful in obtaining a characterization theorem for measurable equivariant functions.

Theorem 5.4. Let (X,Ω,P) be a G-invariant family of distributions having a direct product structure with respect to (K,H) and let $y \varepsilon X$. Then there is a bijection $\sigma:X\to Gy \times Hy$ given by $\sigma(x) = (gy, hy)$ where $k = gh\varepsilon K$ such that $x = ky$.

Proof. If $x \varepsilon X$ then there is some $k = gh\varepsilon K$ such that $ky = x$. Suppose $k_1 = g_1h_1\varepsilon K$ and $k_1y = x$. Then $g_1h_1y = x = ghy$ implying $g_1h_1y = ghy$. Since G and H commute $h^{-1}h_1y = g_1^{-1}gy$. By ii of Definition 5.2 $g_1^{-1}gy = y$ or $hy = h_1y$ and $gy = g_1y$. In other words, σ is well defined.

Now, if $(gy, hy) \varepsilon GyxHy$ then $\sigma(ghy) = (gy, hy)$, so σ is onto. Now suppose $x = ghy$ and $x_1 = g_1h_1y$ are such that $\sigma(x) =$

$\sigma(x_1)$ so that $gy = g_1y$ and $hy = h_1y$. Then

$$x = ghy = g(hy) = g(h_1y) = (gh_1)y = (h_1g)y = h_1(gy) = h_1(g_1y) = g_1h_1y = x_1.$$

That is, σ is one-one also. Q.E.D.

Note again the importance of Part ii of Definition 5.2 and also that of the commutativity of elements of G and H.

In view of the bijection $X \to Gy \times Hy$ given in Theorem 5.4, we may use A, the σ-field for X, to induce σ-fields A_G and A_H on Gy and Hy respectively and on $Gy \times Hy$.

The following is due to Berk:

Theorem 5.5. Let (X,Ω,P) be a G-invariant family of distributions which has a direct product structure with respect to (K,H), let $y \in X$ and let the function $Gy \times \Omega_y \to \Omega$ given by $(gy, \omega) \mapsto g^*\omega$ be $A_G \times \mathcal{B}$ measurable. Then a G-G* equivariant function $\phi : X \to \Omega$ (necessarily of the form given in Theorem 5.3) is measurable if and only if the corresponding ϕ_y is measurable.

Proof. Let the hypothesis of the theorem be satisfied. If $\phi : X \to \Omega$ is a measurable G-G* equivariant function, then ϕ_y, the restriction of ϕ to Hy, is measurable with respect to the σ-field A_H.

Conversely, suppose $\phi_y : Hy \to \Omega_y$ is measurable and let $\phi : X \to \Omega$ be the associated G-G* equivariant function. Letting $B \in \mathcal{B}$ and identifying X with $Gy \times Hy$ as in Theorem 5.4 we have

$$\begin{aligned}\{x\varepsilon X|\phi(x)\varepsilon B\} &= \{(gy, hy)\ \varepsilon\ Gy \times Hy \mid \phi(gy, hy)\ \varepsilon\ B\}\\ &= \{(gy, hy)\ \varepsilon\ Gy \times Hy \mid g^*\phi_y(hy)\ \varepsilon\ B\}\\ &= (e x \phi_y)^{-1}\{(gy, \omega)\ \varepsilon\ Gy \times \Omega_y \mid g^*\omega\ \varepsilon\ B\}\end{aligned}$$

where e is the identity function on Gy. Since the function $(gy, \omega) \mapsto g^*\omega$ is measurable, the set $C = \{(gy, \omega)\ \varepsilon\ Gy \times \Omega_y \mid g^*\omega\ \varepsilon\ B\}$ is measurable. Since ϕ_y is measurable by assumption, the function $e \times \phi_y\colon Gy \times Hy \to Gy \times \Omega_y$ is measurable. Thus,

$$\{x\varepsilon X|\phi(x)\ \varepsilon\ B\} = (e x \phi_y)^{-1}C$$

is measurable in A and hence ϕ is measurable. Q.E.D.

Thus, under the hypotheses of the theorem, all functions of the form given by Theorem 5.3 having measurable ϕ_y are measurable and conversely.

Section 5.4 Equivariant Estimators and a Semi-Direct Product

In Section 5.3 we presented some results which serve to characterize G-equivariant estimators under the condition $K = GxH$. In this section we will present results which give the same characterization but require only a semi-direct product structure for K. Since any direct product is trivially a semi-direct product, the theorems of Section 5.3 will then be seen to be a special case of the theorems to follow. A corollary to this statement is that in theory we are able to characterize estimators for a larger class of groups.

We begin this section with some preliminary results.

Definition 5.3. If G is a group of transformations on X then G is said to have a semi-direct product structure with respect to

(K,H) provided that K is a group of transformations on X which is transitive on X and H is a normal subgroup of K such that

i) $K = H x_s G$.

ii) $h \varepsilon H$, $g \varepsilon G$, $x \varepsilon X$ such that if $gx = hx$ then $gx = x$.

Proposition 5.6. Let G be a group of transformations on X having a semi-direct product structure with respect to (K,H). If $g_1, g_2 \varepsilon G$, $h_1, h_2 \varepsilon H$ and $x \varepsilon X$, such that either $g_1h_1x = g_2h_2x$ or $h_1g_1x = h_2g_2x$ then, $g_1x = g_2x$.

Proof. Let $g_1, g_2 \varepsilon G$, $h_1, h_2 \varepsilon H$ and $x \varepsilon X$ and suppose $g_1h_1x = g_2h_2x$. Then

$$h_2x = (g_2^{-1}g_1)h_1x = h_3(g_2^{-1}g_1)x$$

where

$$(g_2^{-1}g_1)h_1(g_2^{-1}g_1)^{-1} = h_3 \varepsilon H \text{ since } H \lhd K.$$

Thus, $g_2^{-1}g_1x = h_3^{-1}h_2x$ so by ii of Definition 5.3, $g_2^{-1}g_1x = x$ or $g_1x = g_2x$.

On the other hand, suppose $h_1g_1x = h_2g_2x$. Then

$$x = g_2^{-1}h_2^{-1}h_1g_1x = h_3g_2^{-1}g_1x$$

where

$$g_2^{-1}h_2^{-1}h_1g_2 = h_3 \varepsilon H \text{ since } H \lhd K.$$

Thus, $h_3^{-1}x = g_2^{-1}g_1x$ so by ii of Definition 5.3, $g_2^{-1}g_1x = x$ or $g_1x = g_2x$ as required. Q.E.D.

Proposition 5.7. Let G be a group of transformations on X having a semi-direct product structure with respect to (K,H) and let

$x_o \in X$. If $x \in X$ then $Gx \cap Hx_o \neq \phi$ and if z, $y \in Gx \cap Hx_o$ then $z = y$. That is, the H-orbit of x_o intersects each G-orbit in X in one and only one point.

Proof. Fix $x_o \in X$ and let $x \in X$. Since K is transitive on X there is a $k = hg \in K$ such that $hgx = x_o$. Thus, $gx \in Gx \cap Hx_o \neq \phi$. Now suppose z, $y \in Gx \cap Hx_o$. Then there exist elements $g_1, g_2 \in G$ and $h_1, h_2 \in H$ such that $g_1x = z = h_1x_o$ and $g_2x = y = h_2x_o$. Hence

$$h_1^{-1}g_1x = x_o = h_2^{-1}g_2x$$

so by Proposition 5.6

$$z = g_1x = g_2x = y. \qquad \text{Q.E.D.}$$

The phenomenon referred to in Proposition 5.7 is often described by saying that the H-orbit of x_o provides a cross section for the G-orbits. In particular, they have been considered by Wijsman [14].

Proposition 5.8. Let G be a group of transformations on X having a semi-direct product structure with respect to (K,H) and let $x_o \in X$. Then the assignment $x \to z$ where $z \in Gx \cap Hx_o$ is a function $\sigma_H: X \to Hx_o$. Furthermore:

i) σ_H is onto

ii) if $z \in Hx_o$ then $\sigma_H(z) = z$

iii) $\sigma_H(x) = \sigma_H(y)$ if and only if $Gx = Gy$

iv) if $z \in Hx_o$ and $g \in G$ such that $gz \in Hx_o$ then $gz = z$.

Proof. It follows immediately from Proposition 5.7 that the assignment $x \to z$, $z \in Gx \cap Hx_o$, is a function $\sigma_H: X \to Hx_o$.

Now suppose $z \in Hx_o$. Since $Gz \cap Hx_o = \{z\}$, $\sigma_H(z) = z$ by definition. Thus, we have established both i and ii.

In order to establish iii consider x, $y \varepsilon X$ such that $Gx = Gy$. Then $Gx \cap Hx_o = Gy \cap Hx_o$ so $\sigma_H(x) = \sigma_H(y)$. Conversely, suppose $\sigma_H(x) = \sigma_H(y)$. Then $Gx \cap Hx_o = Gy \cap Hx_o$ so there exist elements $g_1, g_2 \varepsilon G$ such that $g_1x = z = g_2y$ from which it follows that $Gx = Gy$.

Finally, suppose that $z \varepsilon Hx_o$ and $g \varepsilon G$ such that $gz \varepsilon Hx_o$.

By ii $\qquad \sigma_H(z) = z$ and $\sigma_H(gz) = gz$

and by iii $\qquad \sigma_H(gz) = \sigma_H(z)$.

Thus, $gz = z$ as required. Q.E.D.

Although the semi-direct product is not symmetric, Propositions 5.7 and 5.8 are still true when the roles of G and H are interchanged. Thus:

Proposition 5.9. Let G be a group of transformations on X having a semi-direct structure with respect to (K,H) and let $x_o \varepsilon X$. If $x \varepsilon X$, then $Hx \cap Gx_o \neq \phi$ and if z, $y \varepsilon Hx \cap Gx_o$, then $z = y$.

Proof. This is simply a restatement of Proposition 5.7 and establishes the fact that the G-orbit of x_o provides a cross section for the H orbits in X.

Proposition 5.10. Let G be a group of transformations on X having a semi-direct product structure with respect to (K,H) and let $x_o \varepsilon X$. Then the assignment $x \to z$ where $z \varepsilon Hx \cap Gx_o$ is a function $\sigma_G : X \to Gx_o$. Furthermore,

i) σ_G is onto

ii) if $z \varepsilon Gx_o$ then $\sigma_G(z) = z$

iii) $\sigma_G(x) = \sigma_G(y)$ if and only if $Hx = Hy$

iv) if $z \varepsilon Gx_o$ and $h \varepsilon H$ such that $hz \varepsilon Gx_o$ then $hz = z$.

v) $\sigma_G(gx) = g\, \sigma_G(x)$ for all $g \varepsilon G$ and $x \varepsilon X$.

Proof. It follows at once from Proposition 5.9 that σ_G is a function. Since for any $z \varepsilon Gx_o$ we have $z \varepsilon Hz \cap Gx_o$ it follows that $\sigma_G(z) = Hz \cap Gx_o = z$ which establishes i and ii.

In order to establish iii suppose $x, y \varepsilon X$ with $Hx = Hy$. Then $Hx \cap Gx_o = Hy \cap Gx_o$ so $\sigma_G(x) = \sigma_G(y)$. Conversely, if $\sigma_G(x) = \sigma_G(y)$, then $Hx \cap Gx_o = Hy \cap Gx_o$. Letting $z \varepsilon Hx \cap Gx_o$ there exist $h_1, h_2 \varepsilon H$ so that $h_1x = z = h_2y$ and thus $Hx = Hy$.

In order to establish iv, let $z \varepsilon Gx_o$ and $h \varepsilon H$ and suppose $hz \varepsilon Gx_o$. By ii, $\sigma_G(hz) = hz$. By iii, $\sigma_G(z) = \sigma_G(hz)$ so $z = \sigma_G(z) = \sigma_G(hz) = hz$ as required in iv.

Finally, let $g \varepsilon G$ and $x \varepsilon X$. Now $\sigma_G(x) \varepsilon Hx \cap Gx_o$ so let $hx = \sigma_G(x) = gx_o$. Then

$$g_1gx_o = g_1hx = (g_1hg_1^{-1})g_1x$$

so that

$$g_1gx_o = \sigma_G(g_1x) \varepsilon H(g_1x) \cap Gx_o.$$

But $g_1gx_o = g_1\sigma_G(x)$. Therefore, $g_1\sigma_G(x) = \sigma_G(g_1x)$ for all $g_1 \varepsilon G$ and $x \varepsilon X$ as required. Q.E.D.

We may now use the functions σ_G and σ_H to establish a bijection from X to $Hx_o \times Gx_o$. This bijection will be helpful in obtaining a characterization of G equivariant estimators.

Theorem 5.11. Let G be a group of transformations on X having a semi-direct product structure with respect to (K,H) and let $x_o \in X$. Then the assignment $x \to (\sigma_H(x), \sigma_G(x))$ establishes a bijection $\sigma : X \to Hx_o \times Gx_o$.

Proof. That the assignment determines a function follows from the fact that σ_G and σ_H are functions. Now suppose $x_1, x_2 \in X$ such that $\sigma(x_1) = \sigma(x_2)$. Then $\sigma_G(x_1) = \sigma_G(x_2)$ and $\sigma_H(x_1) = \sigma_H(x_2)$. By Part iii of Proposition 5.8, there is $g \in G$ such that $x_1 = gx_2$. Similarly, by Proposition 5.10, $x_1 = hx_2$ for some $h \in H$. Thus, $gx_2 = x_1 = hx_2$ which implies $x_1 = gx_2 = x_2$ by ii of Definition 5.3. This demonstrates that σ is one to one. It remains to show that σ is onto. Let $(z_1, z_2) \in Hx_o \times Gx_o$. Then there is $g \in G$ and $h \in H$ such that $gx_o = z_2$ and $hx_o = z_1$. Let $x = ghx_o$. Then

$$\sigma_H(x) = \sigma_H(ghx_o) = \sigma_H(hx_o)$$

by iii of Proposition 5.8. But $\sigma_H(hx_o) = hx_o = z_1$ by ii of that proposition. On the other hand, $x = ghx_o = (ghg^{-1})gx_o$ and $ghg^{-1} \in H$ since H is normal in K. Therefore,

$$\sigma_G(x) = \sigma_G((ghg^{-1})gx_o) = \sigma_G(gx_o) = gx_o = z_2$$

by Parts iii and ii of Proposition 5.10. Thus

$$\sigma(x) = (\sigma_H(x), \sigma_G(x)) = (hx_o, gx_o) = (z_1, z_2)$$

establishing that σ is onto. Q.E.D.

Proposition 5.12. Let G be a group of transformations on X having a semi-direct product structure with respect to (K,H), let $x_o \varepsilon X$ and let $\sigma:X \to Hx_o \times Gx_o$ be the bijection of Theorem 5.11. Then σ has an inverse function $\sigma^{-1}:Hx_o \times Gx_o \to X$ and the assignment $(g,(z_1, z_2)) \to \sigma(g\sigma^{-1}(z_1, z_2))$ for each $g \varepsilon G$ and $(z_1, z_2) \varepsilon Hx_o \times Gx_o$ defines an action G on $Hx_o \times Gx_o$. Furthermore, σ^{-1} is G-equivariant.

Proof. Since σ is a bijection, it has a unique inverse σ^{-1} given by $\sigma^{-1}(z_1, z_2) = x$ where $\sigma(x) = (z_1, z_2)$.

Now let $g \varepsilon G$ and $(z_1, z_2) \varepsilon Hx_o \times Gx_o$. Under the given assignment

$$\sigma^{-1}(g(z_1, z_2)) = \sigma^{-1}(\sigma(g(\sigma^{-1}(z_1, z_2)))) = g(\sigma^{-1}(z_1, z_2))$$

so that σ^{-1} is equivariant as required.

We are now ready to prove the analog of Theorem 5.3 under the condition that G has a semi-direct product structure.

Definition 5.4. The G-invariant family of distributions (X,Ω,P) is said to have a semi-direct product structure with respect to (K,H) provided that G has a semi-direct product structure with respect to (K,H). That is, G must satisfy Definition 5.3.

Theorem 5.12. Let (X,Ω,P) be a G-invariant family of distributions which has a semi-direct product structure with respect to (K,H) and let $y \varepsilon X$. To every G-G* equivariant function $\phi:X\to\Omega$ there corresponds a function $\phi_y:Hy \to \Omega_y$ such that at $x \varepsilon X$,

$\phi(x) = g^* [\phi_y(\sigma_H(x))]$ where $gy = \sigma_G(x)$. Conversely, if $\Omega_y \neq \phi$ then any function $\phi_y: Hy \to \Omega_y$ determines a G-G* equivariant function $\phi: X \to \Omega$ such that $\phi(x) = g^* \phi_y(\sigma_H(x))$ where $gy = \sigma_G(x)$.

Proof. Let (X, Ω, P) and G satisfy the hypotheses. Let $y \in X$ and let σ be the bijection $X \to Hy \times Gy$ given in Theorem 5.11 with $x_o = y$.

Let $\phi: X \to \Omega$ be a G-G* equivariant function. Induce a function $\hat{\phi}: Hy \times Gy \to \Omega$ by $\hat{\phi}(z) = \phi\sigma^{-1}(z)$. Note that $\hat{\phi}$ is well defined and is G-G* equivariant because σ^{-1} is G equivariant and ϕ is G-G* equivariant. For each $t \in H_y$ let $\hat{\phi}_t: Gy \to \Omega$ be given by $\hat{\phi}_t(s) = \hat{\phi}(t,s)$ for each $s \in Gy$. Note that $\hat{\phi}_{t_1} = \hat{\phi}_{t_2}$ if $t_1 = t_2 \in Hy$ and also that

$$\hat{\phi}_t(gs) = \hat{\phi}(t, gs) = \phi\sigma^{-1}((t, gs)) = \phi(g\sigma^{-1}(t, s)) =$$

$$g^*\phi\sigma^{-1}((t, s)) = g^*\hat{\phi}_t(s)$$

so that $\hat{\phi}$ is G-G* equivariant. Then by Theorem 5.2, $\hat{\phi}_t(y) \in \Omega_y$ for any $t \in Hy$. Now define $\phi_y: Hy \to \Omega_y$ by $\phi_y(t) = \hat{\phi}_t(y) = \hat{\phi}(t, y)$. Thus to ϕ there corresponds $\phi_y: Hy \to \Omega_y$. Moreover, if $x \in X$ and $\sigma(x) = (t, s)$ then $\phi(x) = \hat{\phi}(t, s) = \hat{\phi}_t(s)$. Since $\sigma_G(x) = s \in Gy$ there is $g \in G$ and $gy = s$ and since $\hat{\phi}_t$ is G-G* equivariant it follows that $\phi(x) = \hat{\phi}_t(gy) = g^*\hat{\phi}_t(y)$. Moreover, if also $g_1 \in G$ such that $g_1 y = s$ then $g_1^* \hat{\phi}_t(y) = g^* \hat{\phi}_t(y)$ since $\hat{\phi}_t(y) \in \Omega_y$. Finally, $\hat{\phi}_t(y) = \phi_y(t)$ where $t = \sigma_H(x)$ so that $\sigma(x) = g^*\phi_y(t)$ as required.

Conversely, suppose $y \in X$ such that $\Omega_y \neq \phi$ and let $\phi_y : Hy \to \Omega_y$ be given. Then for any $x \in X$ let $\phi(x) = g^*\phi_y(t)$ where $\sigma(x) = (t, s) = (\sigma_H(x), \sigma_G(x))$ and $g \in G$ such that $gy = s = \sigma_G(x)$. If also $g_1 \in G$ such that $g_1 y = s$ then $g^*\phi_y(t) = g_1^*\phi_y(t)$ since $\phi_y(t) \in \Omega_y$. Thus ϕ is well defined. It remains to show that this ϕ is G-G* equivariant. Thus let $g_1 \in G$ and $x \in X$. Now, $\phi(x) = g^*\phi_y(t)$ where $gy = \sigma_G(x)$ and $t = \sigma_H(x)$. By iii, Proposition 5.8, $\sigma_H(g_1 x) = \sigma_H(x)$ and by v, Proposition 5.10, $\sigma_G(g_1 x) = g_1 \sigma_G(x)$. Thus $g_1 gy = \sigma_G(g_1 x)$ while

$$\phi(g_1 x) = (g_1 g)^*\phi_y(t) = g_1^*(g^*\phi_y(t)) = g_1^* \phi(x).$$

Hence ϕ is G-G* equivariant. Q.E.D.

We remark here that ϕ agrees with ϕ_y on Hy.

Section 5.5 Measurability and the Semi-Direct Product

Theorem 5.12 provides a characterization of all G-G* equivariant functions $\phi : X \to \Omega$ under the semi-direct product structure. We now wish to consider conditions which assure the measurability of such functions. That is, we seek the analog of Berk's Theorem 5.5 for the semi-direct product structure.

Given the G-invariant family of distributions (X, Ω, P) with σ-algebras $\mathcal{A}$ and $\mathcal{B}$ on X and Ω respectively and given that the family of distributions has a semi-direct product structure with respect to (K,H) let $y \in X$ and let $\mathcal{A}_G$ denote the σ-algebra induced on Gy by $\mathcal{A}$. That is, $A' \in \mathcal{A}_G$ if and only if $A' = Gy \cap A$ where $A \in \mathcal{A}$. Similarly, let $\mathcal{A}_H$ be the induced σ-algebra on Hy. Recalling the bijection $\sigma: X \to Hy \times Gy$ given by Theorem 5.11 let $\mathcal{A}_X$ be the σ-algebra induced on the product space by σ. That is, $A' \varepsilon \mathcal{A}_X$ if and only if $\sigma^{-1}(A') \varepsilon \mathcal{A}$.

Preliminary to characterizing the measurable equivariant functions we consider the following lemma:

Proposition 5.13. Let (X,Ω,P) be a G-invariant family of distributions having a semi-direct product structure with respect to (K,H), let $y \varepsilon X$ and let $\sigma: X \to Hy \times Gy$ be the bijection of Theorem 5.11. Then to each G-G* equivariant function $\phi: X \to \Omega$ there corresponds a unique function $\hat{\phi}: Hy \times Gy \to \Omega$ given by $\hat{\phi}(s, t) = \phi(\sigma^{-1}(s, t))$. Furthermore, ϕ is $(\mathcal{A}, \mathcal{B})$ measurable if and only if $\hat{\phi}$ is $(\mathcal{A}_X, \mathcal{B})$ measurable.

Proof. We consider only the measurability requirements. Suppose ϕ is $(\mathcal{A}, \mathcal{B})$ measurable and let $B \varepsilon \mathcal{B}$. Now $\hat{\phi}^{-1}(B) = \sigma\phi^{-1}(B)$ and $\phi^{-1}(B) \varepsilon \mathcal{A}$ since ϕ is measurable. By definition of $\mathcal{A}_X$, $\hat{\phi}^{-1}(B) \varepsilon \mathcal{A}_X$. Thus $\hat{\phi}$ is $(\mathcal{A}_X, B)$ measurable.

Conversely, suppose $\hat{\phi}$ is measurable and $B \varepsilon \mathcal{B}$. Now $\phi^{-1}(B) = \sigma^{-1}(\hat{\phi}^{-1}(B))$ where $\hat{\phi}^{-1}(B) \varepsilon \mathcal{A}_X$. By definition of $\mathcal{A}_X$, $\phi^{-1}(B) = \sigma^{-1}(\hat{\phi}^{-1}(B)) \varepsilon \mathcal{A}$ so ϕ is $(\mathcal{A}, \mathcal{B})$ measurable. Q.E.D.

We now characterize all measurable G-G* equivariant functions ϕ in terms of their corresponding ϕ_y. This is the semi-direct product version of Berk's Theorem 5.5.

Theorem 5.14. Let (X,Ω,P) be a G-invariant family of distributions having a semi-direct product structure with respect to (K,H), let $y \in X$, let $\phi: X \to \Omega$ be a G-G^* equivariant function and let $\phi_y: Hy \to \Omega_y$ be the corresponding function given by Theorem 5.12. Under the conditions

i) $A_X = A_G \times A_H$ and

ii) the function $f: Gy \times \Omega_y \to \Omega$ given by $(gy, \omega) \to g^* \omega$ is $(A_G \times \mathcal{B}, \mathcal{B})$ measurable,

the function ϕ is $(A, \mathcal{B})$ measurable if and only if ϕ_y is $(A_H, \mathcal{B})$ measurable.

Proof. We first show the forward implication. Thus, let the hypotheses of the theorem be satisfied and let $\phi: X \to \Omega$ be G-G^* equivariant and $(A, \mathcal{B})$ measurable with corresponding $\phi_y: Hy \to \Omega_y$. We claim that $\phi_y^{-1}(B) = \phi^{-1}(B) \cap Hy$ for any $B \subset \Omega$. Indeed, let $x \in \phi_y^{-1}(B)$. Then $x \in Hy$ and $\phi_y(x) \in B$. Recalling the remark following Theorem 5.12 that ϕ agrees with ϕ_y on Hy we get that $\phi(x) = \phi_y(x) \in B$. That is, $x \in X$ and $\phi(x) \in B$ or $x \in \phi^{-1}(B)$. That is $\phi_y^{-1}(B) \subset \phi^{-1}(B) \cap Hy$. On the other hand, if $x \in \phi^{-1}(B) \cap Hy$, then $x \in Hy$ and $\phi_y(x) = \phi(x) \in B$. Thus $\phi_y^{-1}(B) = \phi^{-1}(B) \cap Hy$. Consequently, if ϕ is measurable and $B \in \mathcal{B}$ then, $\phi_y^{-1}(B) = \phi^{-1}(B) \cap Hy \in A_H$ by definition. Thus ϕ measurable implies ϕ_y measurable which completes the forward implication.

Conversely, suppose that $\phi y: Hy \to \Omega_y$ is $(A_H, \mathcal{B})$ measurable. Consider momentarily the identity function $e: Gy \to Gy$ and the function $(\phi_y x e): Hy \times Gy \to \Omega_y \times G_y$ given by $(\phi_y x e)(hy, gy) = (\phi_y(hy), gy)$ and note that $(\phi_y x e)$ is $A_H \times A_G$ measurable. Furthermore, recall Proposition 5.13 which states that the $(A, \mathcal{B})$ measurability of ϕ is equivalent to the $A_H \times A_G$ measurability of the induced function $\hat{\phi}$. We will establish the desired implication then if we show that ϕ_y is measurable implies $\hat{\phi}$ is measurable.

Thus, let $B \varepsilon \mathcal{B}$. Then,

$$\hat{\phi}^{-1}(B) = \{(hy, gy) \varepsilon Hy \times Gy | \hat{\phi}(hy, gy) \varepsilon B\} =$$

$$\{(hy, gy) \varepsilon Hy \times Gy | g^*\phi_y(hy) \varepsilon B\} =$$

$$(\phi_y \times e)^{-1}\{(\omega, gy) \varepsilon \Omega_y \times G_y | g^* \omega \varepsilon B\}.$$

Now the measurability of $\{(\omega, gy) | g^* \omega \varepsilon B\}$ follows from ii of the hypotheses. Furthermore, since $\phi_y \times e$ is $(A_H \times A_G, \mathcal{B} \times A_G)$ measurable we see that $\hat{\phi}^{-1}(B)$ is $A_H \times A_G$ measurable. Since $A_H \times A_G = A_X$, by i of the hypothesis we may conclude that $\hat{\phi}$ is A_X measurable and hence that ϕ is A measurable. Q.E.D.

CHAPTER VI

DENSITIES FOR MAXIMAL INVARIANTS

In this chapter we explore further the information provided by the existence of the so-called global cross-section mentioned in Chapter V. In particular, we show how a global cross-section gives rise to a function on the space X called a maximal invariant. The relationship between the distributions of the maximal invariant and the distribution on X are displayed. In addition, we give the density of the maximal invariant with respect to a certain measure. Finally, we consider a test of hypothesis involving cosets in the parameter space.

In later sections we will follow Fraser ([3], [4], [5]) and Younger ([15]) in that we will assume that (X,Ω,P) is a K-invariant family of distributions and furthermore that $K = X = \Omega$ and in this way we will identify sample points, parameter points and group elements. In addition, we will assume that K is isomorphic to the semi-direct product of two subgroups H and G. Initially however, we will not require $K = X = \Omega$.

Section 6.1 Global Cross Sections and Maximal Invariants

Following the proof of Proposition 5.7, we noted that the H-orbit of the point x_o provided a cross-section for the G-orbits in X. We now formalize this notion.

Definition 6.1. Let K be a group of transformations on X and let L be a subgroup of K. Then $Z \subset X$ is said to be a global cross-section for the L-orbits in X provided that:

i) for every $x \in X$ there is a unique $z_x \in Z$ such that $Lx \cap Z = \{z_x\}$.

ii) $LZ = \{\ell z | \ell \in L, z \in Z\} = X$.

Thus, as noted in the remarks following Propositions 5.7 and 5.9, global cross-sections exist when K is transitive on X, $K = H x_s G$ and $hgx = x$ implies $gx = x$. In the context of Definition 5.3, these conditions are that G has a semi-direct product structure with respect to (K,H). We now restate these properties formally.

Theorem 6.1. Let K be a group of transformations on X such that: K is transitive on X; there exist subgroups G and H such that $K = H x_s G$ and $hgx = x$ implies $gx = x$. If $x_o \varepsilon X$ then:

i) Hx_o is a global cross-section for the G-orbits in X.

ii) Gx_o is a global cross-section for the H-orbits in X.

Proof. As noted above, the hypothesis means that G has a semi-direct product structure with respect to (K,H) as described in Definition 5.3. Thus the conclusions of Propositions 5.7 and 5.9 hold which is to say the conclusions of this Theorem hold. Q.E.D.

Closely associated with the notion of global cross-section is that of maximal invariant.

Definition 6.2. Let K be a group of transformations on X and let L be a subgroup of K. Then an L-invariant is a function f on X such that $f(x) = f(y)$ whenever $Lx = Ly$.

Thus an L-invariant is simply a function which is constant on the L-orbits in X.

Definition 6.3. Let K be a group of transformations on X and let L be a subgroup of K. A maximal L-invariant is any function T

on X such that $T(x) = T(y)$ if and only if $Lx = Ly$.

This definition follows Lehmann [12]. Thus it is seen that a maximal L-invariant is constant on an L-orbit and takes distinct values on distinct orbits. Furthermore, a maximal invariant is unique in the sense that any invariant function depends only on the maximal invariant.

Proposition 6.2. Let K be a group of transformations on X and let L be a subgroup of K and let T be a maximal L-invariant. A necessary and sufficient condition for f to be an L-invariant is that $f(x)$ depend only on $T(x)$ for each $x \in X$. That is, there exists a function h such that $f(x) = h(T(x))$.

Proof. If f is a function of $T(x)$, say $f(x) = h(T(x))$ for every $x \in X$, then f is L-invariant since $x = \ell y$ gives $f(x) = hT(x) = hT(\ell y) = hT(y) = f(y)$ by the L-invariance of T. This establishes the sufficiency of the condition.

Conversely, suppose f is L-invariant. Let $A = \{t \mid t = T(x)$ for some $x \in X\}$ and $B = \{s \mid s = f(x)$ for some $x \in X\}$. Define $h: A \to B$ by $h(t) = f(x)$ where $x \in X$ such that $T(x) = t$. That h is well defined follows from the fact that T is a maximal L-invariant: Suppose x, y both satisfy $T(x) = t = T(y)$. Then

$$x = \ell y \quad \text{and} \quad h(t) = f(x) = f(\ell y) = f(y).$$

Finally, by definition, $f(x) = h(T(x))$ which establishes the necessity of the condition. Q.E.D.

Now, when a global cross-section exists, a maximal invariant can be defined in a natural way as follows:

Proposition 6.3. Let K be a group of transformations on X, let L be a subgroup of K and let Z be a global cross-section for the L-orbits in X. Then $T:X\to Z$ given by $T(x) = Lx\cap Z$ for each $x \in X$ is a maximal L-invariant.

Proof. Given the hypotheses it is clear that T is well defined by i of Definition 6.1. This same property also gives that T is L-invariant for if $\ell x = y$ then $Lx = Ly$ and so $T(x) = Lx\cap Z = Ly\cap Z = T(y)$. The maximality of T follows similarly: Suppose $Lx\cap Z = T(x) = T(y) = Ly\cap Z$. Then there is a $z \in Z$ with $\ell x = z = \ell' y$ for some $\ell, \ell' \in L$. Thus $Lx = Ly$ which establishes that T is a maximal L-invariant. Q.E.D.

Recalling Propositions 5.8 and 5.10 we see that the functions σ_H and σ_G are maximal invariants. That this is true also follows from Proposition 6.3.

Theorem 6.4. Let K a group of transformations which is transitive on X and such that $K = H x_s G$ and $hgx = x$ implies $gx = x$. If $x_o \in X$ then the functions $\sigma_H:X\to Hx_o$ and $\sigma_G:X\to GX_o$ where $\sigma_H(x) = Gx\cap Hx_o$ and $\sigma_G(x) = Hx\cap Gx_o$ are, respectively, maximal G-invariant and maximal H-invariant.

Proof. Suppose the hypotheses are satisfied. The conclusions follow from Propositions 5.8 (iii) and 5.10 (iii) or by applying Proposition 6.3 and Theorem 6.1. Following the second line of reasoning, Theorem 6.1 asserts that Hx_o is a global cross-section for the G-orbits in X and Gx_o is a global cross-section for the H-orbits in X. It then follows by Proposition 6.3 that $\sigma_H:X\to Hx_o$, where $\sigma_H(x) = Gx\cap Hx_o$, is a maximal G-invariant and that $\sigma_G:X\to GX_o$, where $\sigma_G(x) = H\cap Gx_o$, is a maximal G-invariant as required. Q.E.D.

Section 6.2 Maximal Invariant Parametric Functions and Distributions of Maximal Invariants

The definitions and results of the preceeding sections did not involve an invariant family of distributions. Their applications will, however, so that our discussions will once again revolve around some family of distributions (X,Ω,P) invariant under a group. Also, since we wish to find distributions for maximal invariants we will introduce measurability considerations concerning them.

Thus, we will consider a K-invariant family of distributions (X,Ω,P) where $P_\omega \varepsilon P$ is defined on some σ-algebra A of subsets of X. Following Wijsman [14] we say:

Definition 6.4. $A \subset X$ is L-invariant (L a subgroup of K) if $\ell A = \{\ell a \mid a \varepsilon A\} = A$ for all $\ell \varepsilon L$.

Denoting by A_I all sets $A \varepsilon A$ which are L-invariant we again follow Wijsman and say:

Definition 6.5. A maximal L-invariant statistic is a triple $(Z, \mathcal{B}, T)$ where $T: X \to Z$ is a maximal L-invariant function and $\mathcal{B}$ is a σ-algebra of subsets of Z such that $T^{-1}(B)\varepsilon A_I$ for all $B \varepsilon \mathcal{B}$.

Now for any $P_\omega \varepsilon P$, T induces on $\mathcal{B}$ the distribution P_ω^T given by $P_\omega^T(B) = P_\omega(T^{-1}(B))$. The most useful fact concerning such distributions is that they depend only on some maximal invariant on the parameter space.

Definition 6.6. Given the K-invariant family of distributions (X,Ω,P) and a subgroup L of K, a function π on Ω is a maximal L-invariant parametric function provided π is a maximal invariant with respect to the group L^* on Ω.

Theorem 6.5. Let (X,Ω,P) be a K-invariant family of distributions, π a maximal L-invariant parametric function and $(Z, \mathcal{B}, T)$ a maximal L-invariant statistic. Then the induced distribution of T depends only on $\pi(\omega)$.

Proof. If $\omega_1, \omega_2 \in \Omega$ such that $\pi(\omega_1) = \pi(\omega_2)$, then $\omega_1 = \ell^*\omega_2$ since π is a maximal L-invariant function. For $B \in \mathcal{B}$ we have $P^T_{\omega_1}(B) = P_{\omega_1}(T^{-1}(B)) = P_{\omega_1}(A)$ where $A = T^{-1}(B) \in \mathcal{A}_I$ since T satisfies Definition 6.5. Moreover, $P_{\omega_1}(A) = P_{\ell^*\omega_2}(A) = P_{\omega_2}(A)$ by the K-invariance of the family P and the fact that $A \in \mathcal{A}_I$. Finally,

$$P_{\omega_2}(A) = P_{\omega_2}(T^{-1}(B)) = P^T_{\omega_2}(B).$$

Therefore, $\pi(\omega_1) = \pi(\omega_2)$ implies $P^T_{\omega_1} = P^T_{\omega_2}$ as required. Q.E.D.

We wish now to apply Theorem 6.5 in the case where (X,Ω,P) is a K-invariant family, $K = H \times_s G$, $K^* = H^* \times_s G^*$ and K and K^* are transitive on X and Ω respectively such that $hgx = x$ implies $gx = x$ and $h^* g^* \omega = \omega$ implies $g^* \omega = \omega$. Letting $\mathcal{A}$ be the σ-algebra of subsets of X and $\mathcal{A}_H$ and $\mathcal{A}_G$ be as in Theorem 5.14, we have as maximal invariant statistics the functions $\sigma_H : X \to Hx_o$ and $\sigma_G : X \to Gx_o$ and as maximal invariant parametric functions we have

$\pi_H:\Omega \to H^*\omega_o$ and $\pi_G:\Omega \to G^*\omega_o$ given by $\pi_H(\omega) = G^*\omega \cap H^*\omega_o$ and $\pi_G(\omega) = H^*\omega \cap G^*\omega_o$. In these cases x_o and ω_o represent arbitrary but fixed points. As corollaries to the preceeding result, we have:

Theorem 6.6. If (X,Ω,P) is a K-invariant family of distributions as described above then the distribution of the maximal G-invariant $\sigma_H:X \to (Hx_o, A_H)$ depends only on the value of the maximal G-invariant parametric function $\pi_H:\Omega \to H^*\omega_o$.

Proof. Follows immediately as a corollary to Theorem 6.5.

Theorem 6.7. If (X,Ω,P) is a K-invariant family of distributions as described above, then the distribution of the maximal H-invariant $\sigma_G:X \to (Gx_o, A_G)$ depends only on the value of the maximal H-invariant parametric function $\pi_G:\Omega \to G^*\omega_o$.

Proof. Again, a corollary to Theorem 6.5.

Section 6.3 Densities for Maximal Invariants

In this section we will consider the problem of obtaining densities for maximal invariants, a problem originally investigated by Wijsman [14]. Following Fraser [3], [5] and Younger [15] we restrict ourselves to a K-invariant family (X,Ω,P) such that $X \cong \Omega \cong K = K^*$. In this way we identify points in X, K and Ω where x_o and ω_o correspond to $e \in K$ and $x = kx_o$ corresponds to k_x and $\omega = k_\omega \omega_o$ corresponds to k_ω.

Concerning the group K, we will assume that K is topologically isomorphic to the semi-direct product of two locally compact topological groups H and G relative to some Θ. That is, $K = H x_\Theta G$.

As is Chapter IV, we let μ be a left invariant Haar measure on H and ν_1 be a left invariant Haar measure on G. Then for $\delta: \mathrm{Aut}\, H \to R^+$, R^+ the positive reals as a group under multiplication,

$$d\lambda(k) = d\mu(h)\ d\nu_1(g)\ /\ \delta(\theta(g))$$

is left invariant Haar differential on $K = H x_\theta G$. We denote $d\nu_1(g)\ /\ \delta(\theta(g))$ by $d\nu(g)$.

Concerning the distributions $P_\omega \varepsilon P$ we will assume that each P_ω has a density with respect to λ. That is, for $P_\omega \varepsilon P$ we may write

$$P_\omega(A) = \iint_A f(h, g; \omega)\ d\mu(h) d\nu(g)$$

where it is understood that the measurable sets are the Borel sets and that A, the σ-algebra on $X = K$, is such that $A = A_H \times A_G$. Here again we follow Fraser [4], [5].

Without further qualification we will assume that these conditions are satisfied.

In this context, a maximal H-invariant is a function T on X = K such that

$$T(h_x, g_x) = T(x) = T(y) = T(h_y, g_y)$$

if and only if

$$(h_x, g_x) = (h, 1_G)\ (h_y, g_y)$$

for some $h \varepsilon H$. In particular, the function σ_G discussed earlier becomes $\sigma_G: K \to G$ given by $\sigma_G(h, g) = g$. Wijsman, in [14], considers the problem of obtaining a density for σ_G:

Theorem 6.8. If (X,Ω,P) is a K-invariant family of distributions, then a density for $P_\omega^{\sigma_G}$ with respect to the measure ν on G is given by $\int_H f(h, g; \omega)\, d\mu(h)$ where $f(h, g; \omega)$ is the density of P_ω with respect to λ.

Proof. By definition, for

$$B\varepsilon A_G,\ P_\omega^{\sigma_G}(B) = P_\omega\{(h, g) \mid h\varepsilon H,\ g\varepsilon B\}.$$

Now P_ω has density $f(h, g; \omega)$ with respect to λ so that

$$P_\omega^{\sigma_G}(B) = \int_B \int_H f(h, g; \omega)\, d\mu(h)\, d\nu(g)$$

from which it follows that $\int_H f(h, g; \omega)\, d\mu(h)$ is a density for $P_\omega^{\sigma_G}$ with respect to ν. Q.E.D.

In fact, we see that $P_\omega^{\sigma_G}$ has a density which is given by the marginal density obtained from $f(h, g; \omega)$ by integrating out the first coordinate.

In addition, since $\pi_G(\omega) = \pi_G(h_\omega, g_\omega) = g_\omega$ is a maximal H-invariant parametric function it follows from Theorem 6.7 that the density depends only on the point g_ω.

As a corollary, we obtain the following concerning probability ratios.

Theorem 6.9. Let (X,Ω,P) be a K-invariant family of distributions and let $f(h, g; \omega_1)$ and $f(h, g; \omega_2)$ be probability densities for $P_{\omega_1}, P_{\omega_2} \varepsilon P$. Then, for the maximal H-invariant $\sigma_G(x) = \sigma_G(h_x, g_x) = g_x$ its probability ratio is:

$$\frac{dP_{\omega_1}^{\sigma_G}}{dP_{\omega_2}^{\sigma_G}}(\sigma_G(x)) = \frac{\int\int (h,\ g_x;\ \omega_1)\ d\mu\ (h)}{\int\int (h,\ g_x;\ \omega_2)\ d\mu\ (h)}$$

Proof. Applying Theorem 6.8 with P_{ω_1} and P_{ω_2} and forming the ratio the result follows at once. Q.E.D.

Our objective now is to find a density for the maximal G-invariant. In the terms used in this section the maximal G-invariant corresponds to $\sigma_H(x) = \sigma_H(h_x, g_x) = \Theta(g_x^{-1})(h_x)$.

Theorem 6.10. Let (X,Ω,P) be a K-invariant family of distributions. Then the function $\sigma_H\colon K = X \to H$ given by $\sigma_H(h, g) = \Theta(g^{-1})(h)$ is a maximal G-invariant statistic.

Proof. Consider $h \in H$. Then

$$\sigma_H^{-1}(h) = \{(\Theta(g)(h), g) \mid g \in G\},$$

which is the G-orbit of $(h, 1)$ in this context. Thus, σ_H is G-invariant. On the other hand, suppose $h, h' \in H$. If

$$\{(\Theta(g)(h), g) \mid g \in G\} = \{(\Theta(g)(h'), g) \mid g \in G\}$$

we may write $(h, e) = (\Theta(g)(h), g)$ for some $g \in G$, implying in fact $g = e_G$ so that $h = \Theta(e_G)(h) = h'$. Thus σ_H is maximal G-invariant. Furthermore,

$$\sigma_H^{-1}(B) = \{(\Theta(g)(b), g) \mid g \in G,\ b \in B\} \in A_I.$$

Thus, σ_H is a maximal H-invariant statistic. Q.E.D.

It follows from Theorem 6.6 that the distribution $P_{\omega}^{\sigma_H}$ induced by P_{ω} and σ_H depends only on $\Theta(g_{\omega}^{-1})(h_{\omega}) = \pi_H(\omega)$ the maximal G-invariant parametric function. We now turn to finding the density of σ_H.

Theorem 6.11. If (X,Ω,P) is a K-invariant family of distributions, then a density for $P_{\omega}^{\sigma_H}$ with respect to the measure μ on H is given by

$$g(t;\omega) = \int f(\Theta(s)(t), s; \omega)\ \delta(\Theta(s))\ d\nu(s)$$

where $f(h, g; \omega)$ is a density for P_{ω} with respect to λ on K.

Proof. By definition,

$$P_{\omega}^{\sigma_H}(B) = P_{\omega}(\sigma_H^{-1}(B)) = \int_A\int f(h, g; \omega)\ d\mu(h)\ d\nu(g)$$

where

$$A = \{(\Theta(g)(b), g) \mid b\in B,\ g\in G\} = \sigma_H^{-1}(B).$$

Consider the transformation

$$(t, s) = \sigma(h, g) = (\sigma_H(h, g), g) = (\Theta(g^{-1})(h), g).$$

This transformation is one-one onto with inverse $(h, g) = (\Theta(s)(t), s)$. Under this transformation

$$d\mu \times d\nu(h, g) = d\mu(h)\ d\nu(g) =$$

$$d\mu(\Theta(s)(t))\ d\nu(s) = \delta(\Theta(s))\ d\mu(t)\ d\nu(s).$$

Moreover, under this transformation the set A is transformed to $B \times G =$

$\{(b, g) \mid b \varepsilon B, g \varepsilon G\}$. Thus,

$$P_\omega^{\sigma_H}(B) = \int\int_A f(h, g; \omega)\, d\mu(h)\, d\nu(g) =$$

$$\int_B(\int_G f\,(\Theta(s)(t), s; \omega)\, \delta(\Theta(s))\, d\nu(s)\, d\mu(t).$$

Therefore, $P_\omega^{\sigma_H}$ has a density with respect to μ given by

$$g(t; \omega) = \int_G f\,(\Theta(s)(t), s; \omega)\, \delta(\Theta(s))\, d\nu(s)$$

which is the desired conclusion. Q.E.D.

Corresponding to Theorem 6.9 we obtain the following concerning probability ratios:

Theorem 6.12. Let (X,Ω,P) be a K-invariant family of distributions and let $f(h, g; \omega_1)$ and $f(h, g; \omega_2)$ be probability densities for P_{ω_1}, $P_{\omega_2} \varepsilon P$ respectively. Then, for the maximal G-invariant

$$\sigma_H(x) = \sigma_H(h_x, g_x) = \Theta(g_x^{-1})(h_x)$$

its probability ratio is:

$$\frac{dP_{\omega_1}^{\sigma_H}}{dP_{\omega_2}^{\sigma_H}}(\sigma_H(x)) = \frac{\int f\,(\Theta(s)(\Theta(g_x^{-1})(h_x)), s; \omega_1)\, \delta(\Theta(s))\, d\nu(s)}{\int f\,(\Theta(s)(\Theta(g_x^{-1})(h_x)), s; \omega_2)\, \delta(\Theta(s))\, d\nu(s).}$$

Section 6.4 A Test of Hypothesis Concerning Cosets

Fraser [5] assumes the same basic structure considered in the previous section and discusses two tests of hypotheses concerning cosets.

In particular, Fraser indicates a method of testing the hypothesis

$H_o: \omega \varepsilon hG\omega_o$ or equivalently that the H-component of the parameter is a specified point. The second test Fraser discusses is $H_o: \omega \varepsilon Hg\omega_o$ or equivalently that the G-component of the parameter is a specified point.

As Fraser points out, these two tests may be regarded in other terms as tests about cosets. That is, $H_o: \omega \varepsilon Hg\omega_o$ is taken to mean that the parameter is in the left coset of H indexed by g. Since H is normal under our assumptions we have $Hg = gH$ so a test about a left coset of H is a test about a right coset and vice versa.

Finally, Fraser considers the coset $H_o: \omega \varepsilon hG\omega_o$. Here, however, it is not necessary that $hG = Gh$ and so this test is not also a test about left cosets of G. In this section we propose to consider such a test. Recall Definitions 3.5, 3.6 and 3.7.

Definition 6.7. A size α test, π, of the hypothesis $H_o: \omega \varepsilon S$ vs $H_1: \omega \varepsilon T$ is uniformly most powerful for testing H_o vs H_1 provided that if π_1 is any other size α test of H_o vs H_1 then:

$$\int \pi(x) P_\omega d(x) \geqslant \int \pi_1(x) P_\omega d(x) \quad \text{for all} \quad \omega \varepsilon T.$$

What follows is the fundamental lemma of Neyman-Pearson:

Theorem 6.13. If P_{ω_o} and P_{ω_1} are two probability distributions on a space having densities p_o and p_1 respectively with respect to some measure μ then there exists a most powerful test π of size α for testing $H_o: \omega = \omega_o$ vs $H_1: \omega = \omega_1$. Furthermore, there exists a constant c such that $\pi(x) = 1$ if $p_1(x) > cp_o(x)$ and $\pi(x) = o$ if $p_1(x) < cp_o(x)$.

Proof. See Lehmann [9].

If a hypothesis is invariant under a group, then the principle of invariance suggests that tests be invariant also. That is, that $\pi(kx) = \pi(x)$ for all x and k. This is discussed in detail by Lehmann [9] and in Chapter III.

Returning to the problem of testing cosets, consider the hypothesis H_o: $\omega\varepsilon G\omega_1$ vs H_1: $\omega\varepsilon G\omega_2$. It is seen that the hypothesis is invariant under G. Moreover, if the conditions on a family of K-invariant distributions (X,Ω,P) are assumed to be as in the preceeding section, then it follows from the results of that section that the distribution of the maximal G-invariant statistic σ_H is constant for all ω satisfying H_o and constant for all ω satisfying H_1. We have:

Theorem 6.14. Let (X,Ω,P) be a K-invariant family of distributions as specified in Section 6.4. Then:

i) there exists a size α test, π, for testing H_o: $\omega\varepsilon G\omega_1$ vs H_1: $\omega\varepsilon G\omega_2$.

ii) π depends only on the value of the maximal G-invariant statistic $\sigma_H(x) = \sigma_H(h_x, g_x) = \Theta(g_x^{-1})(h_x)$. That is, π is G-invariant.

iii) π is most powerful for testing H_o vs H_1 among tests which depend only on the maximal G-invariant.

Proof. (i and ii) Given the assumptions of the preceeding section, it follows from Theorem 6.6 that the distribution of the maximal G-invariant statistic σ_H is constant for all ω satisfying H_o since the maximal G-invariant parametric function is constant on $G\omega_1$ having the value $\Theta(g_{\omega_1}^{-1})(h_{\omega_1})$ where $(h_{\omega_1}, g_{\omega_1}) = \omega_1$. Similarly, it follows that the induced distribution for σ_H is constant for all ω satisfying H_1.

Thus, the hypothesis given is equivalent to the hypothesis H_o^T: $T = \sigma_H(x)$ has distribution $P_{\omega_1}^T$ vs H_1^T: $T = \sigma_H(x)$ has distribution $P_{\omega_2}^T$.

By Theorem 6.11, $P_{\omega_1}^T$ and $P_{\omega_2}^T$ have densities $g(t; \omega_1)$ and $g(t; \omega_2)$, respectively, with respect to a left invariant Haar measure μ on H where $t = \sigma_H(x)$. Thus, by the Neyman-Pearson Lemma, Theorem 6.13, there exists a most powerful test π of size α for testing H_o^T vs H_1^T of the form $\pi(t) = 1$ if $g(t; \omega_2) > cg(t; \omega_1)$ and $\pi(t) = 0$ if $g(t; \omega_2) < cg(t; \omega_1)$ for some constant c. Note that π depends only on the value $t = \sigma_H(x)$ and that π determines a test for H_o vs H_1 by accepting H_o if and only if one accepts H_o^T. This establishes i and ii.

To establish iii, suppose that π_1 is a test for H_o vs H_1 which depends only on $t = \sigma_H(x)$. Then π_1 determines a test for H_o^T vs H_1^T by: accept H_o^T if and only if one accepts H_o. But by the Neyman-Pearson Lemma π is most powerful for H_o^T vs H_1^T. Thus, π is most powerful for H_o vs H_1 among those tests which depend only on t, that is, among the G-invariant tests. Q.E.D.

As an illustration, consider the family of normal distributions $P'_{(\mu,\,\sigma)}$ and the corresponding distributions $P_{(\mu,\,\sigma)}$ on the space of the sufficient statistic $(\overline{X}, S)$. This family is invariant under the group K of location-scale transformations and K is topologically isomorphic to the semi-direct product $R \times_{\Theta} R^{+}$, where R is the set of real numbers considered as a topological group under addition, R^{+} is the positive reals under multiplication and $\Theta: R^{+} \to \mathrm{Aut}(R)$ is given by $\Theta(\beta)(\alpha) = \alpha\beta$ for all $\beta \varepsilon R^{+}$ and $\alpha \varepsilon R$.

In this case, $d\mu(\bar{x}) = d\bar{x}$, the usual differential on R, is left invariant on R and $d\nu_1(s) = \frac{ds}{s}$, where ds is again the lebesque differential on R, is a left invariant differential on R^{+} and $d\mu d\nu = \frac{d\bar{x}ds}{s^2}$ is a left invariant differential on $K = R \times_{\Theta} R^{+}$ so that $\delta(\Theta(s)) = s$.

In this situation, the maximal G-invariant statistic obtained in Theorem 6.10 is $\sigma_H(\bar{x}, s) = \bar{x}/s$ and by Theorem 6.6 the distribution of $\bar{x}/s$ induced by $P_{(\mu,\ \sigma)}$ for $(\bar{x}, s)$ depends only on μ/σ, the value of the maximal invariant parametric function.

Thus, the G-invariant hypothesis:

H_o: $(\bar{X}, S)$ follows P_ω, $\omega \varepsilon \{(\beta\mu, \beta\sigma) \mid \beta \varepsilon R^{+} = G\} = G(\mu, \sigma)$ vs

H_1: $(\bar{X}, S)$ follows P_ω, $\omega \varepsilon \{(\beta\mu_1, \beta\sigma_1) \mid \beta \varepsilon R^{+} = G\} = G(\mu_1, \sigma_1)$

may, according to Theorem 6.14, be tested by considering the test:

H_o^T: $T = \bar{X}/S$ follows $P^T_{(\mu,\ \sigma)}$ which depends only on μ/σ vs

H_1^T: $T = \bar{X}/S$ follows $P^T_{(\mu_1,\ \sigma_1)}$ which depends only on μ_1/σ_1

based on the densities (with respect to $d\bar{x}$)

$$g(t;\ (\mu,\ \sigma)) = \int f(st,\ s;\ (\mu,\ \sigma))\ s\ \frac{ds}{s^2}$$

and

$$g(t;\ (\mu_1,\ \sigma_1)) = \int f(st,\ s;\ (\mu_1,\ \sigma_1))\ s\ \frac{ds}{s^2}$$

where $f(\bar{x},\ s;\ (\mu,\ \sigma))$ is the density of $P_{(\mu,\ \sigma)}$ with respect to the K-invariant differential $d\bar{x}\ \frac{ds}{s^2}$

Note that H_o vs H_1 is equivalent to H_o: $(\mu/\sigma) = c_1$ vs H_1: $(\mu/\sigma) = c_2$ and of course $T = \bar{X}/S$ is a form of the T statistic.

CHAPTER VII

INVARIANT ESTIMATION AND SOME EXAMPLES

In this chapter we will discuss some examples and propose a method of evaluating invariant estimators.

Section 7.1 An Illustration Dealing with the Normal Family

In order to illustrate the application of Theorem 5.12 we will discuss in some detail the example mentioned briefly in Section 6.4 when we considered a test of hypothesis concerning cosets.

Throughout this chapter let R denote the real numbers and let R^+ denote the multiplicative group of positive reals and let $X_1, \ldots, X_n$ be i.i.d. normal random variables with mean μ and standard deviation σ.

The distribution of the X_i's induces a distribution, $P_{(\mu, \sigma)}$, on the space of the sufficient statistic $(\overline{X}, S)$, $S = \sqrt{\Sigma(X_i - \overline{X})^2/(n-1)}$. Let (X,Ω,P) then be the family of distributions where

$X = (\overline{X}, S)$ the space of the sufficient statistic

$\Omega = \{(\mu, \sigma) \mid \mu\varepsilon R, \sigma\varepsilon R^+\}$

$P = \{P_{(\mu, \sigma)} \mid (\mu, \sigma) \varepsilon\Omega\}$

Consider the group $G = R^+$. We may define an action of G on the space X by

$$gx = g(\bar{x}, s) = (g\bar{x}, gs) \text{ for all } g\varepsilon G \text{ and } (\bar{x}, s) = x\varepsilon X.$$

It may be seen that if the variable

$$X = (\overline{X}, S) \sim P_{(\mu, \sigma)}, \text{ then}$$

$$gX = g(\overline{X}, S) = (g\overline{X}, gS) \sim P_{(g\mu,\ g\sigma)}.$$

This follows from the fact that if X_i is normal with mean μ and standard deviation σ, that is $X_i \sim N(\mu, \sigma)$, then $gX_i \sim N(g\mu, g\sigma)$. Thus, we may define a group of transformations $G^* = G$ on Ω such that

$$g^*(\mu, \sigma) = g(\mu, \sigma) = (g\mu, g\sigma).$$

It follows from the above discussion that (X,Ω,P) is a G-invariant family of distributions with X,Ω,P and $G = G^*$ as above. Recall Definition 5.1 and Theorem 5.1 which deal with the existence of G-equivariant functions $\phi\colon X \to \Omega$. For the example under discussion, we have the following:

For any $y = (\bar{x}, s) \varepsilon X$

$$G_y = \{g \mid gy = y\} = \{g \mid (g\bar{x}, gs) = (\bar{x}, s)\} = \{1\}$$

$$G_y^* = G_y = \{1\}$$

and

$$\begin{aligned} \Omega_y &= \{\omega \varepsilon \Omega \mid (G_y)^* \omega = \omega\} \\ &= \{(\mu, \sigma) \varepsilon \Omega \mid 1\cdot(\mu, \sigma) = (\mu, \sigma)\} \\ &= \Omega \neq \phi \end{aligned}$$

thus it follows from Theorem 5.1 that G-equivariant functions do exist.

Our aim now is to apply Theorem 5.12 to describe these G-equivariant functions. In order to do this, we must first find groups K and H acting on X such that the family (X,Ω,P) has a semi-direct product structure with respect to (K,H) (Definition 5.4). Thus, we consider the following:

Let $K = \{(h, g) | h \epsilon R, g \epsilon R^+\}$. K is a group with identity element $(o, 1)$ and multiplication given by

$$(h, g)(s, t) = (h+gs, gt).$$

The inverse of (h, g) is $(h/g, g^{-1})$.

We may define an action of K on $X = (\overline{X}, S)$ as follows:

$(h, g)(\bar{x}, s) = (h+g\bar{x}, gs)$ so that K is a group of transformations on X.

We note that

$G' = \{(o, g) | g \epsilon R^+\}$ is a subgroup of K

that $H = \{(h, 1) | h \epsilon R\}$ is a normal subgroup of K

and that any element, (h, g), of K may be written uniquely as $(h, g) = (h, 1)(o, g)$. That is, $K = Hx_s G'$.

Finally, we note that $H \cong R$ and that $G' \cong G$ our original group. Thus, we can and will identify the element $(h, 1) \epsilon H$ with $h \epsilon R$ and the element $(o, g) \epsilon G'$ with $g \epsilon G$. With this in mind, we write $K = Hx_s G$.

We claim that K and H are groups such that the requirements of Definition 5.4, and thus the hypothesis of Theorem 5.12, are met. In order to see this, we must show the following:

i) $K = Hx_s G$

ii) if $h \epsilon H$, $g \epsilon G$, $x \epsilon X$ such that $gx = hx$ then $gx = x$

iii) K is transitive on X. That is, if $x, y \epsilon X$ then there is a $k \epsilon K$ such that $kx = y$.

We verified i above. We now show ii. Let $x = (\bar{x}, s) \epsilon X$ and $h \epsilon H$ and $g \epsilon G$ such that $gx = hx$. We have

$$gx = (o, g)(\bar{x}, s) = (g\bar{x}, gs)$$

and

$$hx = (h, 1)(\bar{x}, s) = (h+\bar{x}, s)$$

so that $gx = hx$ implies $gs = s$ or $g = 1$. Therefore,

$$gx = (o, 1)(\bar{x}, s) = (\bar{x}, s) = x \text{ as required for ii.}$$

In order to show iii we need only remark that K and the space X are "isomorphic" -- in fact, they are equal. For illustrative purposes we show it by direct calculation. Let $x = (\bar{x}, s)$, $y = (u, v)\varepsilon X$. We require $k = (h, g)\varepsilon K$ such that

$$(u, v) = y = kx = (h, g)(\bar{x}, s) = (h+g\bar{x}, gs).$$

That is $\quad v = gs \text{ or } g = s^{-1}v$

and $\quad h = u-g\bar{x} = u-s^{-1}v\bar{x}.$

That is $k = (u-s^{-1}v\bar{x}, s^{-1}v)$ is the required group element. Hence, K is transitive on X and so, according to our previous remarks, the hypothesis of Theorem 5.12 is satisfied.

When we apply Theorem 5.12 we note that we must pick a point $y\varepsilon X$ and then calculate the following:

i) The subset Hy of X

ii) the subset Ω_y of Ω

iii) the function $\sigma_G: X \to Gy$

iv) the function $\sigma_H: X \to Hy$

v) for each $x\varepsilon X$ the element $g\varepsilon G$ such that $gy = \sigma_G(x)$.

We find it convenient to do so now. Thus, let $y = (o, 1)\varepsilon X$. Then:

i) $Hy = \{hy|h\varepsilon H\} = \{(h, 1)(o, 1)|h\varepsilon H\} = \{(h, 1)|h\varepsilon H\}$

ii) $\Omega_y = \Omega$. We verified this earlier when applying Theorem 5.1

iii) $\sigma_G: X \to Gy$ is defined by

$$\begin{aligned}\sigma_G(x) &= \sigma_G(\bar{x}, s) = H(\bar{x}, s) \cap Gy \\ &= \{(h+\bar{x}, s) | h\varepsilon H\} \cap \{(o, g) | g\varepsilon G\} \\ &= \{(o, s)\}\end{aligned}$$

iv) $\sigma_H: X \to Hy$ is defined by

$$\begin{aligned}\sigma_H(x) &= \sigma_H(\bar{x}, s) = G(\bar{x}, s) \cap Hy \\ &= \{(g\bar{x}, gs) | g\varepsilon G\} \cap \{(h, 1) | h\varepsilon H\} \\ &= \{(\bar{x}/s, 1)\}\end{aligned}$$

v) Let $(\bar{x}, s)\varepsilon X$. Then from iii $\sigma_G(\bar{x}, s) = (o, s)$. Since $y = (o, 1)$, we see $(o, s) = \sigma_G(\bar{x}, s) = (o, s)(o, 1) = (o, s)y$ so that $g = (o, s)$ is the required element of G.

Finally, we may apply Theorem 5.12. Suppose we take our G-equivariant $\phi: X \to \Omega$ to be $\phi(\bar{x}, s) = (\bar{x}, s)$. According to the theorem there is a function $\phi_y: Hy \to \Omega_y$ such that at $x = (\bar{x}, s)\varepsilon X$

$$\phi(x) = g^*\phi_y(\sigma_H(x)) \quad \text{where} \quad gy = \sigma_G(x) \quad \text{and}$$
$$g = g^* \quad \text{since} \quad G = G^*.$$

From the preceeding discussions we have that

$$(\bar{x}, s) = \phi(\bar{x}, s) = \phi(x) = g^*\phi_y(\sigma_H(x)) = (o, s)\ \phi_y\ (\bar{x}/s, 1).$$

Multiplying both sides by the inverse of (o, s), namely (o, s^{-1}), we get

$$(\bar{x}/s, 1) = (o, s^{-1})(\bar{x}, s) =$$
$$(o, s^{-1})[(o, s)\ \phi_y(\bar{x}/s, 1)] = \phi_y\ (\bar{x}/s, 1).$$

Thus, we see that $\phi_y: Hy \to \Omega_y = \Omega$ is given by

$$\phi_y(hy) = \phi_y((h, 1)(o, 1)) = \phi_y(h, 1) = (h, 1).$$

Consider another G-equivariant estimator $\phi(\bar{x}, s) = (\bar{x}, cs)$ where

$$c = ((n-1)/2)^{1/2} \Gamma((n-1)/2) / \Gamma(n/2)$$

and n is the sample size.

Applying Theorem 5.12 we get

$$(\bar{x}, cs) = \phi(\bar{x}, s) = (o, s)\,\phi_y(\bar{x}/s, 1)$$

and multiplying both sides by $(o, s)^{-1} = (o, s^{-1})$ we have

$$(\bar{x}/s, c) = (o, s^{-1})(\bar{x}, cs) = (o, s^{-1})\phi(x) =$$
$$(o, s^{-1})(o, s)\,\phi_y(\bar{x}/s, 1) = \phi_y\,(\bar{x}/s, 1)$$

from which it follows that $\phi_y\colon Hy \to \Omega_y = \Omega$ is given by $\phi_y\,(h, 1) = (h, c)$, c as above.

Section 7.2 A Loss Function for an Invariant Estimation Problem

We will illustrate the general idea to be considered in this section by continuing the discussion of the example from the preceeding section.

First, we note that the family (X,Ω,P) from above is not only G invariant, but K invariant as well. That is, if

$$X = (\overline{X}, S) \sim P_{(\mu,\, \sigma)} \text{ and if } k = (h, g)\varepsilon K \text{ then}$$
$$kX = (h, g)(\overline{X}, S) = (h+g\overline{X}, gs) \sim P_{(h+g\mu,\, g\sigma)}.$$

Thus, K induces a suitable group K^* of transformations on Ω, namely $K^* = K$, which makes (X,Ω,P) into a K-invariant family of distributions. This is evident if we note that K is simply the group of location-scale changes.

Observe also that $K = X = \Omega$ so that we are in the situation described at the beginning of Section 6.3. Thus, the theorems and remarks of that section and of Section 6.4 apply.

In Section 7.1 we discussed the functions $\sigma_H: X \to Hy$, a maximal G-invariant, and $\sigma_G: X \to Gy$, a maximal H invariant. Since we have $K = X = \Omega$, we may define analogous maximal invariant parametric functions. Let $\omega_o = (o, 1)\varepsilon\Omega$ and

$$\pi_H: \Omega \to H\omega_o \quad \text{given by} \quad \pi_H(\mu, \sigma) = \pi(\omega) = G\omega \cap H\omega_o,$$

and

$$\pi_G: \Omega \to G\omega_o \quad \text{given by} \quad \pi_G(\mu, \sigma) = \pi(\omega) = H\omega \cap G\omega_o.$$

We see that for $\omega\varepsilon\Omega$

$$\begin{aligned} G\omega \cap H\omega_o &= \{(o, g)(\mu, \sigma)|g\varepsilon G\} \cap \{(h, 1)(o, 1)|h\varepsilon H\} \\ &= \{(g\mu, g\sigma|g\varepsilon G\} \cap \{(h, 1)|h\varepsilon H\} \\ &= \{(\mu/\sigma, 1)\} \end{aligned}$$

so that $\pi_H(\mu, \sigma) = (\mu/\sigma, 1)$ and is a maximal G-invariant parametric function. Similarly $\pi_G(\mu, \sigma) = (o, \sigma)$ and π_G is a maximal H-invariant parametric function.

With these preliminaries out of the way we now return to our example in order to motivate the general discussion which follows.

In our example suppose we are interested in the parameter (μ, σ) only up to its H-orbit. Now the H-orbit of (μ, σ) in Ω is

$$H(\mu, \sigma) = \{(h, 1)(\mu, \sigma)|h\varepsilon H\} = \{(h+\mu, \sigma)|h\varepsilon H\}.$$

Note that for all $(\xi, \delta)\varepsilon H(\mu, \sigma)$ we have $\delta = \sigma$ so that estimating the H orbit of (μ, σ) is equivalent to estimating σ. As Hora-Buehler [8] have discussed this is an equivariant estimation problem. Note that $\sigma = \pi_G(\mu, \sigma)$, the value of the maximal H-invariant parametric function.

Similarly, suppose we are interested in the parameter (μ, σ) only up to its G-orbit. The G-orbit of (μ, σ) in Ω is

$$G(\mu, \sigma) = \{(o, g)(\mu, \sigma) | g \varepsilon G\} = \{(g\mu, g\sigma) | g \varepsilon G\}.$$

Note that $(\xi, \delta) \varepsilon G(\mu, \sigma)$ if and only if $\xi/\delta = \mu/\sigma$. Thus, analogous to the above, estimation of the G-orbit of (μ, σ) is equivalent to estimating $\mu/\sigma = \pi_H(\mu, \sigma)$ the value of the maximal G-invariant parametric function.

As Hora-Buehler [8] specifically point out, this is not an equivariant estimation problem and they exclude it from consideration. However, it is an invariant estimation problem in the following sense.

We will identify μ/σ with $(\mu/\sigma, 1) \varepsilon \Omega = \{(\xi, \delta) | \xi \varepsilon R, \delta \varepsilon R^+\}$. Thus, estimation of the G-orbit of (μ, σ) is equivalent to selecting a function

$$\psi: X = (\overline{X}, S) \to \{(\mu/\sigma, 1) | (\mu, \sigma) \varepsilon \Omega\} = \Omega_o$$

as we have seen,

$$\text{if } X \sim P_{(\mu, \sigma)} \text{ then } gX \sim P_{(\xi, \delta)} \text{ where}$$

$$(\xi, \delta) = (g\mu, g\sigma) \text{ so that } \xi/\delta = \mu/\sigma.$$

Adoption of the principle of invariance therefore suggests that we consider only those $\psi: X \to \Omega_o$ for which $\psi(gx) = \psi(x)$. Following Definition

6.2 we see ψ is G-invariant. It follows from Proposition 6.2 that ψ is a function of a maximal G-invariant. From our discussions earlier in this chapter (or directly from Theorem 6.4), we know that $\sigma_H(x) = \sigma_H(\bar{x}, s) = (\bar{x}/s, 1)$ is a maximal G-invariant.

Thus, any G-invariant function

$$\psi: X \to \Omega_o$$

is a function of

$$\sigma_H: X \to Hy \qquad y = (o, 1)\varepsilon X$$

so that estimating the G-orbit of (μ, σ) is equivalent to estimating μ/σ is equivalent to finding a suitable function

$$\phi_y: Hy \to \Omega_o.$$

Recall our notation of Section 7.1 where we saw that $\Omega_y = \Omega$. Since $\Omega_o \subset \Omega_y$ and since the hypotheses of Theorem 5.12 are satisfied, we may infer that corresponding to each $\phi_y: Hy \to \Omega_o$ there is a G-equivariant estimator $\phi: X \to \Omega$. For example, as we saw in the example in Section 7.1, if

$$\phi_y(hy) = \phi_y(h, 1) = (h, 1)$$

then corresponding to ϕ_y is the G-equivariant estimator $\phi(\bar{x}, s) = (\bar{x}, s)$.

Since the ϕ corresponding to a ϕ_y is G-equivariant there do exist natural loss functions for ϕ namely the G-invariant loss functions. That is, loss functions L such that

$$L(\omega, \phi(x)) = L(g\omega, g\phi(x)) = L(g\omega, \phi(gx)).$$

We propose now to use such an L to define a "natural" loss function

L_I for ϕ_y.

Thus, let L be a given G-invariant loss function and consider an estimator of μ/σ. That is, a function

$$\phi_y: Hy = \{(\bar{x}/s, 1) | (\bar{x}, s) \varepsilon X\} \to \Omega_o \subset \Omega.$$

We now observe $t = \bar{x}/s$. Since Hy provides a global cross-section for the G-orbits in X, the value t specifies a G-orbit. For any $(\delta, 1) \varepsilon \Omega_o$ we define the loss $L_I(\delta, \phi_y(t))$ to be the average loss incurred using the corresponding G-equivariant estimator ϕ and the loss function L (evaluated at $(\delta, 1)$), the average being taken over the G-orbit in X specified by t.

We will now turn to the general case to make these ideas explicit.

Let (X, Ω, P) be a K invariant family of distributions, let $K = HX_\Theta G$, let $K^* \cong K$ where K^* is the group on Ω and suppose that each $P_\omega \varepsilon P$ has a density $f(x; \omega) = f((h_x, g_x); (h_\omega^*, g_\omega^*))$ with respect to a K-invariant Haar measure. In short, assume that we are in the situation described in Section 6.3.

Let us consider the problem of estimating the G-orbit of $\omega \varepsilon \Omega$. Since a maximal G-invariant parametric function will distinguish between G-orbits and since

$$\pi_H: \Omega \to H^* \omega_o \text{ given by } \pi_H(\omega) = \pi_H(h_\omega^*, g_\omega^*) = \Theta(g_\omega^{*-1})(h_\omega^*)$$

(where $\omega_o \varepsilon \Omega$ is arbitrary but fixed and $(h_\omega^*, g_\omega^*) \omega_o = \omega$) is a maximal G-invariant, the problem of estimating the G-orbit of Ω is equivalent to finding a suitable function $\psi: X \to H^* \omega_o$. By the invariance principle these functions should be G-invariant, that is $\psi(gx) = \psi(x)$.

Since ψ is invariant, it must be a function of the maximal invariant $\sigma_H(x) = \sigma_H(h_x, g_x) = \Theta(g_x^{-1})(h_x)\varepsilon Hy$ where $y\varepsilon X$ fixed but arbitrary and $(h_x, g_x)y = x$.

Thus, we need only consider functions ϕ_y: $Hy \to H^*\omega_o$.

If L is a G-invariant loss function, we may use L to define a loss L_I for ϕ_y as follows: Let ϕ be the G-equivariant estimator corresponding to ϕ_y given by Theorem 5.12. For $\omega\varepsilon H^*\omega_o$ and $t\varepsilon Hy$ let

$$L_I(\omega, \phi_y(t)) = \frac{\int L(\omega, \phi(\Theta(s)t, s))\, f(\Theta(s)t, s; \omega)\, \delta(\Theta(s))d\nu(s)}{\int f(\Theta(s)t, s; \omega)\, \delta(\Theta(s))d\nu(s)}$$

where $f(h, g; \delta)$ is the density of P_δ with respect to the left invariant Haar differential $d\lambda = d\mu d\nu$ as in Section 6.3.

Note the similarity of the above with Theorem 6.11. In fact, if we make the change of variable $(t, s) = (\Theta(g^{-1})(h), g)$ we get a density for (t, s) given by

$$g(t, s; \omega) = f(\Theta(s)t, s; \omega)\, \delta(\Theta(s))$$

and the marginal for t is the density of the maximal G-invariant as in Theorem 6.11.

Thus, we may write $L_I(\omega, \phi_y(t))$ as

$$\begin{aligned} L_I(\omega, \phi_y(t)) &= \int L(\omega, \phi(\Theta(s)t, s))g(s|t; \omega)d\nu(s) \\ &= E_\omega^{s|t}(L(\omega, \phi(\Theta(s)t, s)) \qquad \omega\varepsilon H^*\omega_o,\ t\varepsilon Hy. \end{aligned}$$

This is precisely what was meant by saying we would assign a loss $L_I(\omega, \phi_y(t))$ defined to be an average taken over the G-orbit in X specified by t.

Finally, we consider the risk, R_I, associated with the loss L_I. We have

$$\begin{aligned} R_I(\omega, \phi_y) &= E_\omega^t\, L_I(\omega, \phi_y(t)) \\ &= E_\omega^t\, [\, E_\omega^{s|t}\, L(\omega, \phi(\Theta(s)t, s))] \\ &= E_\omega\, L(\omega, \phi(\Theta(s)t, s)) \\ &= \iint L(\omega, \phi(\Theta(s)t, s))\; g(t, s; \omega) d\mu(t) d\nu(s) \\ &= \iint L(\omega, \phi(\Theta(s)t, s))\; f(\Theta(s)t, s; \omega)\; \delta(\Theta(s)) d\mu(t) d\nu(s). \end{aligned}$$

Making the transformation $(h, g) = (\Theta(s)t, s)$ we get

$$\begin{aligned} R_I(\omega, \phi_y) &= \iint L(\omega, \phi(h, g))\; f(h, g; \omega)\; d\mu(h) d\nu(g) \\ &= R\,(\omega, \phi). \end{aligned}$$

Now $R(\omega, \phi)$ is just the risk associated with the G-invariant loss function L and the G-equivariant estimator ϕ. Hence, L_I has been defined in such a way that the minimum risk G-invariant estimator ϕ_y corresponds to the minimum risk G-equivariant estimator ϕ for the loss function L. Thus, we may assert:

Theorem 7.1. Let (X, Ω, P) be a K-invariant family of distributions satisfying the conditions outlined in Section 6.3. Let L be a G-invariant loss function and let L_I be as above and let $y \varepsilon X$ and $\omega_o \varepsilon \Omega$. For the G-invariant estimation problem described by the functions ϕ_y: $Hy \to H^{*}\omega_o$, the minimum risk invariant estimator with respect to L_I corresponds to the minimum risk G-equivariant estimator with respect to

L within the class of G-equivariant estimators $\phi: X \to \Omega$ corresponding to the various ϕ_y.

REFERENCES

[1] Berk, R. H. (1967). A special group structure and equivariant estimation. Ann. Math. Statist., 38, 1436-1445.

[2] Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

[3] Fraser, D. A. S. (1961). On fiducial inference. Ann. Math. Statist., 32, 661-676.

[4] Fraser, D. A. S. (1961). The fiducial method and invariance. Biometrika, 48, 261-280.

[5] Fraser, D. A. S. (1968). The Structure of Inference. John Wiley and Sons, New York.

[6] Hall, M., Jr. (1959). The Theory of Groups. Macmillan, New York.

[7] Halmos, P. R. (1950). Measure Theory. D. Van Nostrand, Princeton, New Jersey.

[8] Hora, R. B. and Buehler, R. J. (1967). Fiducial theory and invariant estimation. Ann. Math. Statist., 37, 643-656.

[9] Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley and Sons, New York.

[10] MacLane, S. and Birkhoff, G. (1967). Algebra. Macmillan, New York.

[11] Montgomery D. and Zippin, L. (1955). Topological Transformation Groups. Interscience, New York.

[12] Nachbin, L. (1965). The Haar Integral. D. Van Nostrand, Princeton, New Jersey.

[13] Staudte, R. G., Jr. (1971). A characterization of invariant loss functions. Ann. Math. Statist., 42, 1322-1327.

[14] Wijsman, R. A. (1965). Cross-sections of orbits and their application to densities of maximal invariants. Proc. Fifth Berkeley Symp. Math. Statist. Prob., 1, 389-400.

[15] Younger, M. S. (1972). Invariant Estimation with Application to Linear Models. Doctor of Philosophy Dissertation, Virginia Polytechnic Institute and State University.

[16] Zacks, S. (1971). The Theory of Statistical Inference. John Wiley and Sons, New York.

EQUIVARIANT ESTIMATORS AND A SPECIAL GROUP STRUCTURE

by

Thomas H. Woteki

(ABSTRACT)

Given a G-invariant family of distributions and under suitable hypotheses concerning G, we characterize the form of G-equivariant estimators. In fact, corresponding to each G-equivariant estimator is an appropriate G-invariant function and conversely.

In the course of characterizing the G-equivariant estimators, we obtain two maximal invariant functions. Some properties of these functions are obtained and in particular we calculate their densities with respect to an appropriate Haar measure.

Finally, we consider an invariant estimator problem, the problem of estimating the orbit of a parameter. It is seen that this invariant problem may be referred back to an equivariant one. A loss function for the invariant problem is defined in such a way that the minimum risk invariant estimator corresponds to the minimum risk equivariant estimator within a subclass of all equivariant estimators.