CROSS-SECTIONS OF ORBITS AND THEIR APPLICATION TO DENSITIES OF MAXIMAL INVARIANTS

R. A. WIJSMAN
UNIVERSITY OF ILLINOIS

1. Introduction and summary

Let G be a group of one-to-one transformations of the sample space X onto itself. A maximal invariant is a function constant on orbits and distinguishing orbits. If G leaves a certain statistical problem invariant and an invariant procedure is to be selected, it is necessary first to solve the problem of how to obtain the distribution of a maximal invariant, given any distribution on X. One of the possible methods consists of writing this distribution as an integral over the group G. This method has been promoted notably by Stein [10], [11], Karlin [6], and James [5], but does not seem to have been used very much in the literature (among the exceptions, see [3], [7]) in spite of the fact that the method has several advantages. Unfortunately, although some specific problems have thus been treated, there does not seem to exist much in the form of a general theory. This paper is intended as a step in that direction. Some new theorems will be presented and several examples given.

The principal tool used in this paper that makes things work is the so-called *cross-section* of orbits, local or global (precise definitions of various terms will be given in section 2). A global cross-section is a subset Z of X such that every orbit intersects Z at exactly one point, in addition to a few other properties to be defined in section 2. A local cross-section at x is a global cross-section for an open, invariant neighborhood of the orbit passing through x. If a global cross-section Z exists, it is possible to convert an integral $\int_X p\, d\mu$ (μ is Lebesgue measure) into an iterated integral of the form $\int_Z \nu_Z(dz) \int_G p(gz)\nu_G(dg)$, where ν_Z and ν_G are certain measures on Z, G, respectively. For any global cross-section Z there is a natural maximal invariant, namely the function that associates to every orbit its intersection with Z. For any distribution P on X, with density p with respect to Lebesgue measure, the distribution of the maximal invariant is then a distribution on Z given by $\nu_Z(dz) \int p(gz)\nu_G(dg)$. The exact nature of the measures ν_Z and ν_G will be given in sections 4 and 5.

In many statistical problems the primary interest is in the probability ratio of a maximal invariant, given any two densities p_1 and p_2. It is then not necessary

Research supported, in part, by the National Science Foundation under Grant GP-3814.

to obtain a global cross-section; one can get by with a local cross-section at every x, and the probability ratio at x is then given by $\int p_2(gx)\nu_G(dg)/\int p_1(gx)\nu_G(dg)$.

There is another function served by a cross-section, global or local. If the principle of invariance is invoked and statistical procedures restricted to depend only on a maximal invariant, this amounts to demand that the procedures be measurable with respect to the sigma-field $\mathfrak{A}^I$ of invariant measurable sets. There is a priori no guarantee that this leaves the statistician with enough procedures to choose from. An obvious situation of this kind arises if G is transitive on X, for then X is one orbit and $\mathfrak{A}^I$ is trivial. The same thing could happen without G being transitive on X, if every orbit is dense in X. Such "misbehavior" of orbits is excluded if a cross-section exists, and we have then a guarantee that $\mathfrak{A}^I$ is "rich" enough. This will be shown in more detail in theorem 5.

Among practitioners of invariance it is customary to choose the range space of any maximal invariant to be Euclidean. On the other hand, if a global cross-section Z exists, it is in general not Euclidean (it is an analytic manifold under the conditions to be imposed presently). This seems contradictory, but, in fact, the Euclidean choice is often possible only after removing from X a set of Lebesgue measure 0. For example, let X be Euclidean n-space with the origin deleted, that is, X consists of the points $x = (x_1, \cdots, x_n) \neq 0$. Let G consist of the transformations $x \rightarrow cx$, where c runs through the positive reals. The orbits are then the rays emanating from the origin. It is customary to choose as a maximal invariant the function $(x_1/x_n, \cdots, x_{n-1}/x_n, \operatorname{sgn} x_n)$, but this is possible only when the collection of rays with $x_n = 0$ is removed from X. There are, of course, many other choices of removal of a null set from X to make the maximal invariant Euclidean. On the other hand, any $(n-1)$-sphere concentric with the origin is a global cross-section and provides a natural maximal invariant that does not suffer from the defect of the Euclidean maximal invariants mentioned above.

Invariance considerations in statistics have been useful in parametric and in nonparametric problems. In this paper, only the parametric case will be considered. In the bulk of applications to parametric problems X is Euclidean and G is a Lie group consisting of translations and/or linear transformations. The translations are trivial to deal with, so they will not be considered here. We shall therefore make the following restrictions throughout this paper: X is a nonempty open subset of Euclidean n-space E^n, and G is a Lie subgroup of the general linear group $GL(n, R)$ of $n \times n$ real nonsingular matrices. Thus, $x \in X$ is an n-vector (taken to be a column vector), g an $n \times n$ nonsingular matrix, and the transformation of X by g given by $x \rightarrow gx$. The subset X is called a *linear G-space.*

It will be understood throughout that all Lie groups of $n \times n$ matrices that arise (including G) are endowed with the usual topology inherited from E^{n^2}.

2. Definitions and notation

The action gx of $g \in G$ on $x \in X$ was already defined in section 1. Then $g \in G$ acts in a natural way on sets, families of sets, measures, and functions. That is, if $A \subset X$, define $gA = \{gx: x \in A\}$; if $\mathfrak{A}$ is a family of sets, define $g\mathfrak{A} = \{gA: A \in \mathfrak{A}\}$; if P is a measure on $\mathfrak{A}$, define gP on $g\mathfrak{A}$ by $gP(gA) = P(A)$; if f is a function on X, define gf by $gf(gx) = f(x)$. We say that $A \subset X$ is invariant if $gA = A$ for every $g \in G$, and that f on X is invariant if $gf = f$ for every $g \in G$. The *orbit* of x (more precisely the G-orbit of x) is $Gx = \{gx: g \in G\}$. Thus, a function f on X is invariant if and only if it is constant on each orbit. The space of orbits, considered as an abstract space, is written X/G. If for any $x \in X$, $Gx = X$, we say that G is transitive on X. For any $A \subset X$, $GA = \{gA: g \in G\}$; GA can also be considered as the union of the orbits that intersect A.

If G_0 is a subgroup of G, then for any $g \in G$, gG_0 is called the left coset of g (relative to G_0), sometimes written $[g]$. If g_1 and g_2 are in the same coset, we shall sometimes write $g_1 \sim g_2$. The space of left cosets relative to G_0 is written G/G_0. The natural map $\varphi: G \to G/G_0$ is defined by $\varphi(g) = [g]$. The group G acts in a natural way on G/G_0 by $g_1[g_2] = [g_1g_2]$. The *isotropy group* G_x of $x \in X$ is defined as $G_x = \{g \in G: gx = x\}$. It is easily verified that there is a one-to-one correspondent between Gx and G/G_x.

We shall denote by $\mathfrak{A}$ the sigma-field of Borel subsets of X. Since for every $g \in G$, $x \to gx$ is a homeomorphism, $g\mathfrak{A} = \mathfrak{A}$. Define $\mathfrak{A}^I = \{A: A \in \mathfrak{A}$ and A is invariant$\}$, then $\mathfrak{A}^I$ is a sub-sigma-field of $\mathfrak{A}$. For any P on $\mathfrak{A}$, let P^I be the restriction of P to $\mathfrak{A}^I$. A *maximal invariant* is a triple $(Z, \mathfrak{B}, t)$, Z a space, $\mathfrak{B}$ a sigma-field of subsets of Z, and t a function: $X \to$ onto Z such that if $z = t(x)$, $x = X$, then $t^{-1}(z) = Gx$, and such that $t^{-1}\mathfrak{B} = \mathfrak{A}^I$. Thus, t is invariant and takes different values on different orbits. Consequently, Z is in one-to-one correspondence with X/G. Furthermore, the sets of $\mathfrak{B}$ are in one-to-one correspondence with those of $\mathfrak{A}^I$. Given any distribution P on $\mathfrak{A}$, t induces on $\mathfrak{B}$ the distribution $Pt^{-1} = P^It^{-1}$, which we shall denote by P^Z.

If G acts on two spaces, X and Y, then $f: X \to Y$ is called *equivariant* if $f(gx) = gf(x)$ for all $x \in X$, $g \in G$ (that is, f commutes with g). A *map* is a continuous function. Before defining cross-sections, it is convenient to define first the somewhat more general object of a *slice*. (A good reference is Palais [9], who also gives references to earlier work. For references to cross-sections see [4] and [8].) Definitions 1, 4, and 5 are taken from Palais [9].

DEFINITION 1. *A slice at x is a set $Z \subset X$ such that* (i) *$x \in Z$;* (ii) *GZ is open in X;* (iii) *there exists an equivariant map $f: GZ \to G/G_x$ such that $f^{-1}(G_x) = Z$.*

A slice Z has the property that $G_xZ = Z$, and if $g \notin G_x$, then $gZ \cap Z = \varnothing$. However, an orbit may intersect a slice in more than one point. This is not allowed in a local cross-section.

DEFINITION 2. *A local cross-section at x is a slice Z at x such that if $z \in Z$ and $gz \in Z$ for some $g \in G$, then $gz = z$.*

DEFINITION 3. *A global cross-section is a local cross-section Z such that $GZ = X$.*

DEFINITION 4. *A neighborhood V of x is called* thin *if the closure of $\{g \in G: gV \cap V \neq \varnothing\}$ is compact.*

DEFINITION 5. *The space X is called a* Cartan G-space *if every $x \in X$ has a thin neighborhood.*

We shall say that a k-dimensional slice or cross-section is *flat* if it is contained in the translate of a k-dimensional linear space. We shall denote n-dimensional Lebesgue measure by μ_n, left Haar measure on G by μ_G. Let G_0 be a compact subgroup of G, $Y = G/G_0$, and φ the natural map $G \to Y$. Then φ induces on Y the left Haar measure $\mu_Y = \mu_G \varphi^{-1}$. Finally, $|g|$ stands for the absolute value of the determinant of g, and the $n \times n$ identity matrix will be denoted I_n or e.

3. Existence of local cross-sections

In section 4 it will be shown how to use local cross-sections in integration. Here we shall deal with their existence. One cannot expect to be able to put a local cross-section at every x; for example, any point that lies on an orbit of less than maximum dimension has no local cross-section. For the purpose of integration with respect to Lebesgue measure μ_n, it would be sufficient to show that the set of exceptional points is of μ_n measure 0. Unfortunately, this is not so in general [12]. The extra condition that makes things work is the assumption that X be a Cartan G-space (definition 5). The following theorem is proved in [12].

THEOREM 1. *If $X \subset E^n$ is a linear Cartan G-space, there is an open linear Cartan G-space $X^0 \subset X$ such that $\mu_n(X - X^0) = 0$ and there is a flat local cross-section at every $x \in X^0$.*

In order to apply theorem 1, it is necessary to verify the Cartan condition. In many applications this can be done directly without difficulty, but in others it could conceivably be troublesome. There are no known easy general sufficient conditions (a necessary condition is, of course, that G_x be compact for every $x \in X$ (see definitions 4 and 5)). Fortunately, there is an important class of applications of invariance to problems in multivariate normal analysis where the Cartan property can be proved once and for all. This follows from theorem 2.

THEOREM 2. *Let $X = X_1 \times X_2$, where $X_1 \subset E^{n_1}$, and X_2 is a space of $k \times k$ positive definite matrices, so that $X_2 \subset E^{n_2}$ with $n_2 = k(k+1)/2$. For any $x \in X$ put $x = (r, s)$, $r \in X_1$, $s \in X_2$, so that x is an n-vector, where $n = n_1 + n_2$. Let G^* be a closed subgroup of $GL(k, R)$, F^* a continuous homomorphism of G^*, and let a group G of linear transformations on X be defined by $r \to Br$, $s \to CsC'$, $C \in G^*$, $B = B(C) \in F^*$. Then X is a linear Cartan G-space.*

PROOF. Let $V = V_1 \times V_2$ be a neighborhood of (r, s). A simple argument, using the continuity of $B(C)$, shows that if V_2 is a thin neighborhood of s for the transformations $s \to CsC'$, $C \in G^*$, then V is a thin neighborhood of (r, s). Therefore, in the proof we may assume $X = X_2$. Furthermore, it follows from definition 4 that it is sufficient to give the proof for X being the space of all $k \times k$ positive definite matrices. In order to show that every $s \in X$ has a thin neighborhood, it is sufficient to show this for I_k, which we shall abbreviate I. We have to show that there is a neighborhood V of I in X such that if $M = \{C \in G^*: CVC' \cap V \neq \varnothing\}$, then the closure of M in G^* is compact:

equivalently (since G^* is closed in $GL(k, R)$), that the closure of M in $GL(k, R)$ is compact. Take any $0 < a < 1$ and define $\mathfrak{J}_a$ = set of all $k \times k$ lower triangular matrices whose elements are in absolute value less than a. Let $V = \{(I + T)(I + T)' : T \in \mathfrak{J}_a\}$; then V is a neighborhood of I. If $C \in M$, then there exist $T_1, T_2 \in \mathfrak{J}_a$ such that $(I + T_2)(I + T_2)' = C(I + T_1)(I + T_1)'C'$ so that there exists a $k \times k$ orthogonal matrix Ω such that $C(I + T_1) = (I + T_2)\Omega$; that is $C = (I + T_2)\Omega(I + T_1)^{-1}$. It can easily be verified that the elements of $(I + T)^{-1}$ are uniformly bounded for $T \in \mathfrak{J}_a$ so that M is bounded. Moreover, for $T \in \mathfrak{J}_a$, $(1 - a)^k < |I + T| < (1 + a)^k$ so that $|C| > (1 - a)^{2k} > 0$ if $C \in M$. It follows that the closure of M in $GL(k, R)$ coincides with the closure of M in E^{k^2}, which is compact as a closed, bounded subset of E^{k^2}.

EXAMPLE 1. In the derivation of Hotelling's T^2, r is the sample mean and s the sample covariance matrix in a sample from a multivariate normal distribution. The group $G^* = F^* = GL(k, R)$ so that $gx = (Cr, CsC')$. Since theorem 2 applies, and therefore theorem 1, there is a local cross-section at almost every x.

4. Application of local cross-sections to the probability ratio of a maximal invariant

It is proved in ([12], lemma 3) that if Z is a local cross-section at x, then $G_z = G_x$ for every $z \in Z$. Putting $Y = G/G_x$, every orbit intersecting Z is now a copy of Y. Thus, there is a one-to-one correspondence between GZ and $Y \times Z$.

THEOREM 3. *Let Z be a flat k-dimensional* ($0 < k < n$) *local cross-section at a point $x_0 \in X$ such that G_{x_0} is compact, and let p be a real-valued function on GZ, integrable with respect to μ_n. Then there exists an analytic, real-valued function ψ on Z such that*

$$\int_{GZ} p(x)\mu_n(dx) = \int_Z \psi(z)\mu_k(dz) \int_G p(gz)|g|\mu_G(dg). \tag{1}$$

PROOF. Denote $Y = G/G_{x_0}$, φ the natural map $G \to Y$, and $\mu_Y = \mu_G\varphi^{-1}$. Being compact, G_{x_0} is conjugate to an orthogonal group [1]. It follows that $|g_0| = 1$ for every $g_0 \in G_{x_0}$ so that $g_1 \sim g_2$ implies $|g_1| = |g_2|$. The common value of $|g|$ for all $g \in \varphi^{-1}y$ will be denoted $|y|$. Now the function $(y, z) \to gz$, where g is any member of $\varphi^{-1}y$, is an analytic homeomorphism of $Y \times Z$ onto GZ ([9], proposition 2.1.2, [12], section 2). This permits writing the integral of p over GZ as an integral over the product space $Y \times Z$, and the latter as an iterated integral, using Fubini's theorem. The volume element $\mu_n(dx)$ is expressible as $\mu_n(dx) = \psi(y, z)\mu_Y(dy)\mu_k(dz)$, $\psi > 0$ analytic. Making the transformation $x \to gx$ which transforms $dx \to |g|\,dx$ and leaves μ_Y and z invariant, we readily deduce $\psi(y, z) = |y|\psi(z)$, $\psi > 0$ analytic on Z. Thus we have $\int_{GZ} p(x)\mu_n(dx) = \int_Z \psi(z)\mu_k(dz) \int_Y p(gz)|y|\mu_Y(dy)$, in which g is any member of $\varphi^{-1}y$. The integral over Y equals $\int_G p(gz)|g|\mu_G(dg)$, and the theorem is proved.

If p is the density with respect to μ_n of a probability distribution P on X, then it follows from (1) that

$$P^Z(dz)/\mu_k(dz) = \psi(z) \int_G p(gz)|g|\mu_G(dg) \tag{2}$$

is the density with respect to μ_k of the maximal invariant defined locally by the local cross-section Z.

THEOREM 4. *Let $X \subset E^n$ be a linear Cartan G-space, and let $p_i \geq 0$, $\int_X p_i(x)\mu_n(dx) = 1$, $i = 1, 2$, be two given probability densities. Then for any maximal invariant $(Z, \mathcal{B}, t)$ its probability ratio is for almost all (μ_n) x given by*

$$\frac{dP_2^Z}{dP_1^Z}(t(x)) = \frac{dP_2^I}{dP_1^I}(x) = \frac{\int p_2(gx)|g|\mu_G(dg)}{\int p_1(gx)|g|\mu_G(dg)}. \tag{3}$$

PROOF. The first inequality in (3) is clear, and it implies that the probability ratio does not depend on the choice of maximal invariant. If at x a local cross-section Z exists we may take a maximal invariant defined locally by Z (with $\mathcal{B}$ the Borel subsets of Z; the measurability question will be settled in theorem 5). Writing (2) for P_2 and P_1 and taking the ratio gives (3). According to theorem 1, we may exclude from X a set of μ_n-measure 0 such that in the remaining X^0 there is a local cross-section at every point x, concluding the proof.

Note that in (3) the extreme left member is constant along each orbit, so that this ought to be true also for the ratio of integrals on the extreme right. That this is indeed so can be verified directly by replacing x with $g_1^{-1}x$, for any fixed $g_1 \in G$; then numerator and denominator are both multiplied by the same constant $|g_1|\ \Delta(g_1)$, where Δ is the modular function.

One of the great advantages of the expression (3) is that it is not necessary to find an explicit expression for a maximal invariant which in some cases may be quite a hard problem. Expression (3) is especially useful in cases where G is not specified completely so that it is out of the question to give an explicit expression for a maximal invariant. Yet, even in such cases, (3) may give sufficient information. For instance, when the p_i also depend on an integer m, we may be able to study the asymptotic behavior of (3) as $m \to \infty$ for arbitrary G (within the restrictions imposed on G). An application of this kind to the question of termination with probability one of a certain class of sequential probability ratio tests of composite hypotheses will be made in a future paper.

5. Global cross-sections

A global cross-section gives more but is also harder to come by than a local one, and the theorems guaranteeing the existence of a global cross-section (theorems 6 and 7) are much more restricted in their generality than theorem 1 on the existence of local cross-sections. First we shall deal with the measurability question.

THEOREM 5. *Let Z be a global cross-section; then Z is closed in X and is therefore a Borel set. Let $\mathcal{B}$ be the sigma-field of Borel subsets of Z. Define $t: X \to Z$ by $t(x) = Gx \cap Z$; then $t^{-1}\mathcal{B} = \mathfrak{A}^I$ so that $(Z, \mathcal{B}, t)$ is a natural maximal invariant.*

PROOF. Let x_0 be an arbitrary point of Z. Then Z is a slice at x_0. With x, in definition 1, replaced by x_0, let f be the equivariant map of definition 1. Suppose $x_m \in Z$, $x_m \to x$; then $G_{x_0} = f(x_m) \to f(x)$ so that $f(x) = G_{x_0}$, proving $x \in Z$. Putting $y = f(x)$, $z = t(x)$, it follows from a result of Palais ([9], proposition 2.1.2; see also [12], section 2) that the one-to-one correspondence $x \leftrightarrow (y, z)$ is a homeomorphism. Under this homeomorphism there is a one-to-one correspondence between the Borel sets of X and those of $Y \times Z$, and to the invariant Borel sets of X correspond sets of the form $Y \times B$, $B \in \mathfrak{B}$, in $Y \times Z$, proving $t^{-1}\mathfrak{B} = \mathfrak{A}^I$.

The basic method for finding a global cross-section is to find another group H that also acts on X and such that the combined action of G and H is transitive on X. Then, under certain additional conditions to be specified in theorems 6 and 7, any H-orbit is a global cross-section.

THEOREM 6. *Let G and H be two commuting Lie groups of linear transformations on X, and x_0 a point of X, such that the following conditions are fulfilled:* (i) *G_{x_0} and H_{x_0} are compact;* (ii) *if $g \in G$, $h \in H$, then $gx_0 = hx_0$ only if $gx_0 = hx_0 = x_0$;* (iii) *the dimensions of the orbits Gx_0 and Hx_0 are positive;* (iv) *GH is transitive on X. Put $Y = G/G_{x_0}$, $Z = H/H_{x_0}$, and identify Z with Hx_0. Then Z is a global cross-section, and if the real-valued function p is μ_n-integrable on X, we have*

$$\int_X p(x)\mu_n(dx) = c\int_Z |h|\mu_Z(dz)\int_G p(ghx_0)|g|\mu_G(dg), \tag{4}$$

in which h is any member of H such that $[h] = z$, and the constant c is given by the Radon-Nikodym derivative

$$c = \mu_n(dx)/\mu_Y(dy)\mu_Z(dz) \qquad \text{evaluated at} \quad x_0. \tag{5}$$

PROOF. Let φ_G be the natural map $G \to Y$, and similarly, $\varphi_H: H \to Z$. Suppose x has two representations: $x = ghx_0 = g_1h_1x_0$; then, using the commutativity of G and H, we have $g_1^{-1}gx_0 = h^{-1}h_1x_0$. Since $g_1^{-1}g \in G$, $h^{-1}h_1 \in H$, it follows from (ii) that $g_1^{-1}g \in G_{x_0}$, $h^{-1}h_1 \in H_{x_0}$, that is, $g \sim g_1$ and $h \sim h_1$. Consequently, there is a one-to-one correspondence between $Y \times Z$ and X given by $(y, z) \leftrightarrow ghx_0$, where g is any member of $\varphi_G^{-1}y$, h any member of $\varphi_H^{-1}z$.

We shall show now that it is an analytic homeomorphism. It is sufficient to do this in a neighborhood of x_0. Let $\{K_\alpha\}$, α in a finite set of integers, be a basis for the Lie algebra of G such that $K_1x_0, \cdots, K_kx_0$ are linearly independent and $K_\alpha x_0 = 0$ for the remaining α's. Similarly, let $\{L_\beta\}$ be a basis for the Lie algebra of H such that $L_1x_0, \cdots, L_\ell x_0$ are linearly independent and $L_\beta x_0 = 0$ for the remaining β's. By (iii), $k > 0$, $\ell > 0$. If W is any submanifold of X (such as Y, Z, or X itself), we shall denote by W_x the tangent space to W at x. With a slight abuse of notation, any tangent vector at x_0 to a submanifold of X is of the form $\sum v^i\, \partial/\partial x^i$, where the v^i and x^i are the components of vectors v, x, and the differentiations are to be performed at x_0. For convenience of notation, however, we shall identify such a tangent vector with v. With this convention, the tangent space Y_{x_0} at x_0 to $Y = Gx_0$ is spanned by the vectors $K_1x_0, \cdots, K_kx_0$, Z_{x_0} by $L_1x_0, \cdots, L_\ell x_0$, $(GHx_0)_{x_0}$ by $K_1x_0, \cdots, K_kx_0, L_1x_0, \cdots, L_\ell x_0$, while X_{x_0} is all of E^n.

Although by (iv) GHx_0 and X are the same set of points, we have not shown that they are the same analytic manifold, that is, carry the same analytic structure. Actually, we shall only need to know that GHx_0 as an analytic manifold has dimension n. To show this, suppose $\dim GHx_0 = m < n$ (for the following argument I am indebted to R. L. Bishop and N. T. Hamilton). Each element of GH has a neighborhood V small enough that Vx_0 is homeomorphic to an m-cell, so that Vx_0 is a Borel set of μ_n-measure 0 (since $m < n$). The group GH can be covered by a countable family of such neighborhoods V since the topology of GH is the relativized topology of E^{n^2} (see section 1) and has therefore a countable base. Then GHx_0 is covered by a countable family of sets of the form Vx_0. It would follow that $\mu_n(GHx_0) = 0$ which is impossible since $GHx_0 = X$ as a point set, and $\mu_n(X) > 0$.

It was shown above that $\dim GHx_0 = n$. Since $\dim GHx_0 = \dim (GHx_0)_{x_0}$, the vectors $K_1x_0, \cdots, L_\ell x_0$ must span n-space (so that $k + \ell \geq n$). We shall show now that the vectors are actually linearly independent (implying $k + \ell = n$). Suppose the contrary; then there exist $K = \sum_1^k a_iK_i$ and $L = \sum_1^\ell b_jL_j$ with $Kx_0 \neq 0$, $Lx_0 \neq 0$, and $(K - L)x_0 = 0$. For any real t we have then $e^{t(K-L)}x_0 = x_0$, or $e^{-tL}e^{tK}x_0 = x_0$ (making use of the commutativity of G and H), or $e^{tK}x_0 = e^{tL}x_0$. Now $e^{tK} \in G$ and $e^{tL} \in H$, and then it follows from (ii) that $e^{tK}x_0 = e^{tL}x_0 = x_0$ for every t. Using the latter of these two equalities, it follows that $Lx_0 = 0$, which is a contradiction.

We have shown now that $K_1x_0, \cdots, K_kx_0, L_1x_0, \cdots, L_\ell x_0$ is a basis for E^n. Remembering that $K_1x_0, \cdots, K_kx_0$ is a basis for Y_{x_0}, $L_1x_0, \cdots, L_\ell x_0$ for Z_{x_0}, and keeping in mind that $(Y \times Z)_{x_0} = Y_{x_0} \times Z_{x_0}$ and $X_{x_0} = E^n$, we have established that $(Y \times Z)_{x_0}$ and X_{x_0} are linearly isomorphic. It follows then from ([1], proposition 3, p. 80), that $Y \times Z$ and X are locally analytically homeomorphic at x_0, as was to be proved.

With the correspondence $x \leftrightarrow (y, z)$ define f by $f(x) = y$; then f is continuous by the above result, f is equivariant (for G), and $f^{-1}(G_{x_0}) = Z$. Therefore, f can be taken as the function f in definition 1 (with x in definition 1 replaced by x_0).

We conclude that Z is a slice at x_0. But Z also satisfies definition 2, because the orbit of x intersects Z in the unique point hx_0, where h is any member of $\varphi_H^{-1}z$. Therefore, Z is a local cross-section at x_0, and since $GHx_0 = X$, Z is a global cross-section. The proof of (4) and (5) rests on the fact that $\mu_n(dx) = c|g|\mu_Y(dy)|h|\mu_Z(dz)$ (g any member of $\varphi_G^{-1}y$, h any member of $\varphi_H^{-1}z$) and is essentially the same as the proof of theorem 3.

REMARKS. 1. If the conditions of theorem 6 hold for some x_0, they hold for every $x_0 \in X$ so that every H-orbit is a global cross-section for (X, G). Furthermore, the statement of the theorem is symmetric in G and H, so that every G-orbit is a global cross-section for (X, H).

2. If p in theorem 6 is the density with respect to μ_n of a distribution P on $\mathfrak{A}$, then (4) gives the density of P^Z with respect to $\mu_Z(dz)$ as $c|h| \int p(ghx_0)|g|\mu_G(dg)$.

3. In many applications $H_{x_0} = \{e\}$ in which case $Z = H$.

4. To determine c by (5) amounts essentially to differentiation. Alternatively, c can be determined by integrating the right-hand side of (4), with p any manageable function, and setting the result equal to the left-hand side (which equals 1 if p is a probability density).

Some of the examples that follow have also been treated by Karlin [6], using integration over invariant measures on groups and arriving at the same results along a slightly different path.

EXAMPLE 2 (ratio of two variables, noncentral t). Let $n = 2$, $x = (x_1, x_2)'$, $X = \{x: x_2 > 0\}$. Let G consist of the matrices $g = aI_2$, $a > 0$, with $\mu_G(dg) = da/a$ and $|g| = a^2$. Let H be the group of 2×2 triangular matrices with 1's on the diagonal and b above the diagonal, $-\infty < b < \infty$. Then $\mu_H(dh) = db$ and $|h| = 1$. Clearly, G and H commute, and it is easy to check that GH is transitive on X. Choose $x_0 = (0, 1)'$ so that $x = ghx_0 = (ab, a)'$; G_{x_0} and H_{x_0} are trivial, so $Z = H$, and $gx_0 \neq hx_0$, unless $g = h = e$. Therefore, all conditions of theorem 6 are met. We compute c from $x_1 = ab$, $x_2 = a$, so that at $a = 1$, $b = 0$ we have $dx_1 = db$, $dx_2 = da$ so $dx_1dx_2 = \mu_G(dg)\mu_H(dh)$; hence $c = 1$. Substitution into (4) gives $\int p(x)\mu_2(dx) = \int_{-\infty}^{\infty} db \int_0^{\infty} p(ab, a)a\, da$. We observe that a maximal invariant under G is $x_1/x_2 = b$. If x_1 and x_2 are considered random variables, p their joint density with respect to μ_2, then we read off the density of x_1/x_2 at b with respect to μ_1 as $\int_0^{\infty} p(ab, a)a\, da$. In particular, if x_1 is the sample mean, x_2 the sample standard deviation in a sample from a normal distribution, we get an integral for the noncentral t-density.

EXAMPLE 3 (noncentral Wishart). Consider all $k \times n$ matrices $\mathfrak{X}$ that are of rank k, $k \leq n$, and let x be the kn-vector obtained from $\mathfrak{X}$ by writing the elements of $\mathfrak{X}$ in some arbitrary but fixed order (note that the n in our general theory is replaced by kn). Let X be the totality of all such x. Let G correspond to all transformations $\mathfrak{X} \to \mathfrak{X}\Omega$, with Ω an $n \times n$ orthogonal transformation. Haar measure on G can be chosen normalized so that $\mu_G(G) = 1$. Furthermore, $|g| = 1$. Choose H to be the group corresponding to all transformations $\mathfrak{X} \to T\mathfrak{X}$, with T a $k \times k$ lower triangular matrix with positive diagonal elements, then $|h| = |T|^n$. For left Haar measure on H we may take $\mu_H(dh) = d(TT')/|TT'|^{(k+1)/2}$ (here $d(TT')$ is short for the product of differentials of the elements on, and on one side of, the diagonal of the symmetric matrix TT'). Choose x_0 to correspond to $\mathfrak{X}_0 = (I, 0)$, where $I = I_k$ and 0 denotes a $k \times (n - k)$ matrix of 0's. We have then $\mathfrak{X} = T\mathfrak{X}_0\Omega$. All conditions of theorem 6 can be verified to hold. In particular, H_{x_0} is trivial so that $Z = H$. Since there is a one-to-one correspondence between $h = T$ and TT', we may take as maximal invariant $TT' = \mathfrak{X}\mathfrak{X}'$. This is the Wishart matrix if the columns of $\mathfrak{X}$ form a sample from a multivariate normal distribution with 0 mean vector, and (4) provides an easy way to evaluate the Wishart density. If the columns of $\mathfrak{X}$ are independently multivariate normal with common covariance matrix but arbitrary means, then (4) yields an integral for the noncentral Wishart density.

In order to compute c using (5) note that Y corresponds to all $k \times n$ matrices which have orthonormal rows, that is, Y is the Stiefel manifold of k-frames in n-space [5]. Writing an element of this manifold as a $k \times n$ matrix A, with rows a_i', an invariant differential form on Y is given by James ([5], (4.39)), as

$$\prod_{j=1}^{n-k} \prod_{i=1}^{k} b_j' \, da_i \prod_{i<j\leq k} a_j' \, da_i, \tag{6}$$

where the b_j form with the a_i an orthonormal set. At $\tilde{x}_0$ this form reduces to $\prod da_{ij}$ where the product is over all (i, j) with $i < j \leq k$ and all (i, j) with $j > k$. It should also be noted that in order that $A + dA$ has orthonormal rows, we must have $dA = (d\Sigma, dC)$ with $d\Sigma$ skew-symmetric and dC arbitrary. Taking the normalization factor into account, taken from ([5], (5.10)), we find at x_0

$$\mu_Y(dy) = 2^{-k} \left[\prod_{\nu=n-k+1}^{n} \pi^{-\nu/2} \Gamma(\nu/2) \right] \prod da_{ij}. \tag{7}$$

In order to find $\mu_Z(dz) = \mu_H(dh)$, let dT be a lower triangular matrix with elements dt_{ij}, $i \geq j$. Then at x_0 (where $T = I$),

$$\mu_H(dh) = \prod_{i\geq j} (dT + dT')_{ij} = 2^k \prod_{i\geq j} dt_{ij}. \tag{8}$$

Finally, we obtain $\mu_n(dx)$ in terms of the dt_{ij} and da_{ij} as follows: $\tilde{x}_0 + d\tilde{x} = (I + dT)(I + d\Sigma, dC)$, so that $d\tilde{x} = (dT + d\Sigma, dC)$, omitting higher order differentials which will not contribute to the exterior differential form. We get

$$\mu_n(dx) = \prod dx_i = \prod d\tilde{x}_{ij} = \prod dt_{ii} \prod_{i>j} (dt_{ij} + d\sigma_{ij}) \prod_{i<j} d\sigma_{ij} \prod dc_{ij}. \tag{9}$$

Now

$$\prod_{i>j} (dt_{ij} + d\sigma_{ij}) \prod_{i<j} d\sigma_{ij} = \prod_{i>j} dt_{ij} \prod_{i<j} d\sigma_{ij} \tag{10}$$

since $d\sigma_{ij} = -d\sigma_{ji}$ and any exterior differential form with a repeated differential is 0. Therefore,

$$\mu_n(dx) = \prod_{i\geq j} dt_{ij} \prod_{i<j} d\sigma_{ij} \prod c_{ij} = \prod_{i\geq j} dt_{ij} \prod da_{ij}. \tag{11}$$

Substituting all this into (5) yields $c = \prod_{\nu=n-k+1}^{n} \pi^{\nu/2} [\Gamma(\nu/2)]^{-1}$.

The value of c in example 3 can of course be obtained much more simply by relating it to the known multiplicative constant of the central Wishart density. The point of the above computation is to illustrate how (5) can be used directly to compute c, even in a fairly complicated case. In such cases the use of exterior differential forms may be of some practical advantage over the type of computation that uses Jacobians.

A special case of example 3 arises when $k = n$. In that case (4) provides a decomposition of an integral over all $k \times k$ nonsingular matrices into an integration over the orthogonal group and an integration over the identity component of the lower triangular group. This decomposition was also derived by Stein in [10] by a different method.

There is no guarantee that in every problem one is successful in finding a group H such that the hypotheses of theorem 6 are satisfied. The following theorem gives a result similar to (4) (but not as easy to apply) under weaker conditions. Specifically, it is no longer required that G and H commute.

THEOREM 7. *Assume the same conditions as in theorem* 6, *except that the commutativity of G and H is replaced by the following conditions: G is closed in GH, for each $h \in H$, $hGh^{-1} = G$ (that is, G is a closed normal subgroup of GH), and $hG_{x_0}h^{-1} = G_{x_0}$. Then Z is a global cross-section, and for any μ_n-integrable p we have*

$$\int p(x)\mu_n(dx) = c \int \psi(h)|h|\mu_Z(dz) \int p(ghx_0)|g|\mu_G(dg) \tag{12}$$

in which c is given by (5) *and $\psi(h)$ is the Radon-Nikodym derivative*

$$\psi(h) = \mu_Y(dy)/\mu_Y(hdyh^{-1}) \qquad \text{evaluated at} \quad y = G_{x_0}. \tag{13}$$

PROOF. It can be checked algebraically as in the proof of theorem 6 that there is a one-to-one correspondence $x \leftrightarrow (y, z)$. It is still true that a G-orbit transforms into a G-orbit under the transformation $x \to hx$, but it is no longer true that y remains constant under this transformation. The hypothesis $hGh^{-1} = G$ implies that each $h \in H$ acts on G by $g \to hgh^{-1}$, and the hypothesis $hG_{x_0}h^{-1} = G_{x_0}$ implies that each h even acts on G/G_{x_0}, by $h[g]h^{-1} = [hgh^{-1}]$. It is immediately verified that to $x \to hx$ corresponds $(y, z) \to (hyh^{-1}, hz)$. The proof of theorem 7 is the same as the proof of theorem 6, except for the factor $\psi(h)$ and the proof of the linear independence of the vectors $K_1x_0, \cdots, L_lx_0$.

To establish (13), let dy and dz be "small" neighborhoods of $[e]$ in Y, Z, respectively. Then the volume element is $\mu_n(dx) = c\mu_Y(dy)\mu_Z(dz)$. Under the transformation $x \to hx$, $\mu_n(hdx) = |h|\mu_n(dx) = c|h|\mu_Y(dy)\mu_Z(dz)$. On the other hand, under this transformation $dy \to hdyh^{-1}$, $dz \to hdz$, so $\mu_n(hdx) = c|h|\psi(h)\mu_Y(hdyh^{-1})\mu_Z(dz)$, where we have made use of the left invariance of μ_Z. Equating the two expressions for $\mu_n(hdx)$ yields (13).

In the part of the proof of theorem 6 where the linear independence of $K_1x_0, \cdots, L_lx_0$ was established we used the fact that $e^{t(K-L)} = e^{-tL}e^{tK}$. This is no longer true in general if G and H do not commute. However, the proof goes through in exactly the same way after we have shown that $e^{t(K-L)} = e^{-tL}g$ for some $g \in G$ (g may depend on t). In order to establish this fact, denote by $\Lambda(G)$ the Lie algebra of G; $\Lambda(H)$ and $\Lambda(GH)$ are similarly defined. Then $\Lambda(GH)$ consists of all $K + L$, $K \in \Lambda(G)$, $L \in \Lambda(H)$. Let φ be the natural map: $GH \to GH/G$, that is, $\varphi(hg) = hG$. The differential $d\varphi$ maps $\Lambda(GH)$ onto $\Lambda(GH)/\Lambda(G)$ ([1], p. 115, proposition 1; [2], p. 132, theorem 6.6.4). More specifically, if $M \in \Lambda(GH)$, then $d\varphi(M)$ depends on M only through its residue class mod $\Lambda(G)$ ([1], pp. 114–115). That is, if $K \in \Lambda(G)$, then

$$d\varphi(M + K) = d\varphi(M). \tag{14}$$

Furthermore, if $M \in \Lambda(GH)$, we have (see [1], p. 118, proposition 1; [2], p. 129, (26))

$$\varphi(e^M) = e^{d\varphi(M)}. \tag{15}$$

Taking in (14) $M = L \in \Lambda(H)$ and in (15) first $M = K + L$ and then $M = L$, we obtain

$$\varphi(e^{K+L}) = \varphi(e^L). \tag{16}$$

Now $e^L \in H$, so $\varphi(e^L) = e^L G$. Substituting this into (16) we get $\varphi(e^{K+L}) = e^L G$, so that $e^{K+L} = e^L g$ for some g. If t is any real number, replacing K by tK, L by $-tL$, we have $e^{t(K-L)} = e^{-tL} g$ for some $g \in G$, as was to be shown. This concludes the proof of theorem 7.

EXAMPLE 4. Consider all $(p+q) \times (p+q)$ positive definite matrices $\mathfrak{x}$, partitioned as

$$\mathfrak{x} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{17}$$

in which S_{11} is $p \times p$ positive definite, S_{22} is $q \times q$ positive definite, $S_{12} = S_{21}'$ is $p \times q$, and let x be the corresponding n-vector, where $n = p(p+1)/2 + pq + q(q+1)/2$. Let G correspond to the transformations $\mathfrak{x} \to \tilde{C}\mathfrak{x}\tilde{C}'$, where

$$\tilde{C} = \begin{bmatrix} I_p & 0 \\ C & I_q \end{bmatrix} \tag{18}$$

and C runs through all $q \times p$ matrices. We can take $\mu_G(dg) = \prod dC_{ij}$ where the C_{ij} are the elements of C. If we take x_0 corresponding to $\mathfrak{x}_0 = I_{p+q}$, then G_{x_0} is trivial, so $Y = G$. Define H by the transformations $\mathfrak{x} \to \tilde{D}\mathfrak{x}\tilde{D}'$, where $\tilde{D} = \operatorname{diag}(A, B)$, and A runs through $GL(p, R)$, B through $GL(q, R)$. All conditions of theorem 7 can be verified to hold. We shall pursue this example only to the extent of computing $\psi(h)$. For notational economy, denote by $(C)^*$ a $(p+q) \times (p+q)$ matrix that has C as its last q rows and first p columns, and zeros otherwise. Then if $dy = dg$ corresponds to $(dC)^*$, and h to $\operatorname{diag}(A, B)$, $hdgh^{-1}$ corresponds to $\operatorname{diag}(A, B)(dC)^* \operatorname{diag}(A^{-1}, B^{-1}) = (BdCA^{-1})^*$. We have then $\mu_G(dg) = \prod dC_{ij}$ and $\mu_G(hdgh^{-1}) = \prod (BdCA^{-1})_{ij} = |B|^p|A|^{-q} \prod dC_{ij}$. Thus, $\psi(h) = |A|^q|B|^{-p}$.

REFERENCES

[1] C. CHEVALLEY, *Theory of Lie Groups* I, Princeton, Princeton University Press, 1946.

[2] P. M. COHN, *Lie Groups*, Cambridge, Cambridge University Press, 1957.

[3] N. GIRI, "On the likelihood ratio test of a normal multivariate testing problem," *Amer. Math. Statist.*, Vol. 35 (1964), pp. 181–189.

[4] A. GLEASON, "Spaces with a compact Lie group of transformations," *Proc. Amer. Math. Soc.*, Vol. 1 (1950), pp. 35–43.

[5] A. T. JAMES, "Normal multivariate analysis and the orthogonal group," *Ann. Math. Statist.*, Vol. 25 (1954), pp. 40–75, and many later papers, mostly integration over the orthogonal group.

[6] S. KARLIN, "Notes on multivariate analysis," Stanford University, unpublished.

[7] S. KARLIN and D. TRUAX, "Slippage problems," *Ann. Math. Statist.*, Vol. 31 (1960), pp. 296–324.

[8] D. MONTGOMERY and L. ZIPPIN, *Topological Transformation Groups*, New York, Interscience, 1955.

[9] R. S. PALAIS, "On the existence of slices for actions of noncompact Lie groups," *Ann. of Math.*, Vol. 73 (1961), pp. 295–323.

[10] C. M. STEIN, "Notes on multivariate analysis," Stanford University, unpublished.

[11] ———, "Some problems in multivariate analysis," Part I, Technical Report No. 6, Department of Statistics, Stanford University, 1956.

[12] R. A. WIJSMAN, "Existence of local cross-sections in linear Cartan G-spaces under the action of non-compact groups," *Proc. Amer. Math. Soc.*, Vol. 17 (1966), pp. 295–301.