CHAPTER 2

Group Actions and Relatively Invariant Integrals

In this chapter, group actions are reviewed and are illustrated with examples of relevance for statistical applications. Relatively invariant integrals (measures) are defined and examples are given. An important result, due to Weil, gives necessary and sufficient conditions for the existence and uniqueness of relatively invariant integrals when the group action is transitive. A discussion of invariant and equivariant functions closes out the chapter.

2.1. Group actions. In many examples, the elements of a group G are functions which are one-to-one and onto from a set $\mathbf{X}$ to itself. Further, the group operation in G is just function composition when the elements of G are regarded as functions on $\mathbf{X}$. A typical example of this is the group Gl_n of $n \times n$ nonsingular matrices and the space R^n. Each $g \in \mathrm{Gl}_n$ is a one-to-one onto map from R^n to R^n and

$$g_1(g_2(x)) = (g_1 \circ g_2)(x),$$
$$ex = x.$$

That is, g_1 evaluated at the point $g_2(x) \in R^n$ is equal to $g_1 \circ g_2 \in \mathrm{Gl}_n$ evaluated at $x \in R^n$ and the identity in G is the identity function. Thus, function composition is "the same as" the group operation. In many circumstances the group operation is defined so that the above relationship holds. The idea of a group action on a set simply abstracts the essentials of this situation.

Let $\mathbf{X}$ be a set and let G be a group with identity e.

DEFINITION 2.1. A function F defined on $G \times \mathbf{X}$ to $\mathbf{X}$ satisfying

(i) $F(e, x) = x$, $x \in \mathbf{X}$,
(ii) $F(g_1 g_2, x) = F(g_1, F(g_2, x))$, $g_1, g_2 \in G$, $x \in \mathbf{X}$,

specifies G *acting on the left of* $\mathbf{X}$.

Although Definition 2.1 captures mathematically what one means by a left group action, in concrete examples, the explicit use of F can be mathematical overkill. In the statistical literature, the most common verbiage to indicate left group action is "Suppose G acts on the left of $\mathbf{X}$ with action $x \to gx$." What this means is that the value of F at (g, x) is denoted by gx (whose definition is suppose to be clear from context), so in this notation, conditions (i) and (ii) of Definition 2.1 are

$$ex = x, \qquad x \in \mathbf{X},$$
$$(g_1 g_2)x = g_1(g_2 x).$$

That is, $g \in G$ is thought of as defining a function of $\mathbf{X}$ to $\mathbf{X}$ and the value of g at x is written gx or sometimes $g(x)$. The equation $(g_1g_2)x = g_1(g_2x)$ then means that "function composition" and the group operation are "the same." This can actually be made precise using Definition 2.1. For each $g \in G$, define a function T_g on $\mathbf{X}$ to $\mathbf{X}$ by

$$T_g(x) \equiv F(g, x).$$

Then T_e is the identity function and (ii) simply means

$$T_{g_1}\big(T_{g_2}(x)\big) = T_{g_1g_2}(x).$$

That each T_g is one-to-one and onto is easily verified as is the equation

$$T_{g^{-1}} = T_g^{-1}.$$

Thus $\{T_g | g \in G\}$ is a group under function composition and function composition in this group corresponds to group composition in G.

The reason for the adjective "left" in Definition 2.1, is that there is also a definition of a right group action in which condition (ii) becomes

(ii′) $F(g_1g_2, x) = F(g_2, F(g_1, x))$.

In these notes, all group actions are defined so that they are left group actions. Some care must be taken in certain examples to insure that a group action is a left group action.

Here is our first example.

Example 2.1. Consider $G = \mathrm{Gl}_n$ and $\mathbf{X} = R^n$. Define F on $\mathrm{Gl}_n \times R^n$ by

$$F(g, x) = gx,$$

where gx means the matrix g times the vector x. That F defines a left group action is immediate. In less formal notation, one would say "Gl_n acts on the left of R^n via the group action

$$x \to gx,"$$

where gx means what it did before. Since Gl_n acts on the left of R^n, so does every subgroup of Gl_n—just restrict the action to the subgroup. In particular, O_n and $G_T^+ \subset \mathrm{Gl}_n$ both act on the left of R^n via the action

$$x \to gx. \qquad \square$$

Henceforth we will switch to the less formal description of left group actions and just say something like "consider G acting on the left of $\mathbf{X}$ with action

$$x \to gx."$$

A group G acting on the left of $\mathbf{X}$ induces a natural equivalence relationship among the elements of $\mathbf{X}$, namely x_1 is equivalent to x_2 iff $x_1 = gx_2$ for some $g \in G$. This equivalence relationship divides $\mathbf{X}$ into disjoint subsets called *orbits*, that is

$$O_x = \{gx | g \in G\}$$

is called the *orbit of* x and consists of exactly those elements in $\mathbf{X}$ which are equivalent to x. Hence two points are equivalent iff they are in the same orbit.

EXAMPLE 2.2. Take $\mathbf{X}$ to be the real vector space S_n of all $n \times n$ real symmetric matrices and take $G = O_n$. The left action of O_n on $\mathbf{X}$ is defined by

$$x \to gxg',$$

where g' is the transpose of $g \in O_n$ and gxg' means the product of the three matrices g, x and g'. To describe the orbit of $x \in S_n$, recall the spectral theorem for elements of S_n which asserts that for each x, there is a $g \in O_n$ such that

$$x = g\lambda g',$$

where the diagonal matrix $\lambda \in S_n$ has diagonal elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ which are the ordered eigenvalues of x. Thus, two points x_1 and x_2 are equivalent iff x_1 and x_2 have the same vector of ordered eigenvalues. This follows because the group action

$$x \to gxg'$$

does not change the eigenvalues of x and if x_1 is equivalent to x_2, then x_1 and x_2 have the same eigenvalues. Thus, we can say that the eigenvalues provide an index for the orbits in S_n under the action of O_n—an index in the sense that there is a one-to-one correspondence between orbits and vectors (ordered) of eigenvalues. Later in this chapter, we introduce *maximal invariants* which are just functions which provide one-to-one orbit indices. Thus for this example, the vector of eigenvalues provides a maximal invariant. □

EXAMPLE 2.3. For this example, take $\mathbf{X}$ to be $F_{p,n}$ which is the set of $n \times p$ real matrices x which satisfy $x'x = I_p$, the $p \times p$ identity matrix. Thus $x \in F_{p,n}$ iff the p columns of x are the first p columns of some $n \times n$ orthogonal matrix. Let $G = O_n$ which acts on the left of $F_{p,n}$ via

$$x \to gx,$$

where gx means matrix multiplication. Note that $F_{1,n}$ is the sphere of radius 1 on R^n and $F_{n,n} = O_n$. Given $x \in F_{p,n}$, let $g \in O_n$ have as its first p rows the transposes of the first p columns of x. Then the orthogonality of g implies that

$$gx = x_0 = \begin{pmatrix} I_p \\ 0 \end{pmatrix} : n \times p.$$

Thus there is only one orbit in $F_{p,n}$ since every element x is equivalent to x_0. More generally, when G acts on $\mathbf{X}$ and there is only one orbit, we say G acts *transitively* on $\mathbf{X}$. □

EXAMPLE 2.4. Let $\mathscr{L}_{p,n}$ be the vector space of all $n \times p$ real matrices, $p \leq n$, and consider the product group $O_n \times \mathrm{Gl}_p$ whose elements are written (γ, g). The group action is defined by

$$x \to \gamma x g'$$

for $x \in \mathscr{L}_{p,n}$, $\gamma \in O_n$ and $g \in \mathrm{Gl}_p$, where g' means the transpose of g. The reason for the transpose on g is so that the action is a *left* action. Without the transpose, Definition 2.1(ii) does not hold. This action is not transitive, but it "almost" is transitive in the following sense. Let $\mathbf{X} \subset \mathscr{L}_{p,n}$ be all the rank p elements in $\mathscr{L}_{p,n}$ so the complement of $\mathbf{X}$ in $\mathscr{L}_{p,n}$ has Lebesgue measure 0. Then $O_n \times \mathrm{Gl}_p$ acts on $\mathbf{X}$ with the action defined above. To see that this action on $\mathbf{X}$ is transitive, consider $x \in \mathbf{X}$ and write x as

$$x = uv',$$

where $u \in F_{p,n}$ (of Example 2.3) and $v \in G_T^+ \subset \mathrm{Gl}_p$. [This is the so-called *Q-R* decomposition which is usually proved via the Gram–Schmidt orthogonalization procedure; for an example, see Proposition 5.2 in Eaton (1983), page 160.] Hence, for $(\gamma, g) \in O_n \times \mathrm{Gl}_p$,

$$(\gamma, g)x = \gamma x g' = \gamma u v' g' = \gamma u (gv)'.$$

By picking $g = v^{-1}$ and γ so that

$$\gamma u = \begin{pmatrix} I_p \\ 0 \end{pmatrix} \equiv x_0 \in \mathbf{X},$$

it follows that

$$(\gamma, g)x = x_0.$$

Thus, every x is equivalent to x_0, so the group action is transitive. □

The issue of removing a "small" subset from a space in order to make a group action "nicer" in some sense occurs in many examples. In the above example, "small" means a set of Lebesgue measure 0 and "nice" means transitive. However, in other examples, these words can have different meanings.

2.2. Relatively invariant integrals. Consider a space $\mathbf{X}$ (space in the sense introduced in Chapter 1, so $\mathbf{X}$ is a locally compact Hausdorff space whose topology has a countable base) and let G be a topological group.

DEFINITION 2.2. The group G acts *topologically on the left of* $\mathbf{X}$ if G acts on the left of $\mathbf{X}$ and if the action of G, say F: $G \times \mathbf{X} \to \mathbf{X}$, is continuous.

Because all the actions we consider are left actions, the phrase "on the left" is deleted in what follows and we simply say G acts topologically on $\mathbf{X}$. As usual,

$K(\mathbf{X})$ denotes the vector space of continuous functions with compact support defined on $\mathbf{X}$, so integrals are defined on $K(\mathbf{X})$. When G acts topologically on $\mathbf{X}$, then L_g defined on $K(\mathbf{X})$ by

$$(L_g f)(x) = f(g^{-1}x)$$

maps $K(\mathbf{X})$ onto $K(\mathbf{X})$. Note that

$$L_g L_h = L_{gh}$$

so that G acts on the left of $K(\mathbf{X})$ with the action

$$f \to L_g f.$$

DEFINITION 2.3. Let χ be a multiplier for G. An integral J on $K(\mathbf{X})$ is *relatively* (left) *invariant with multiplier* χ if

$$J(L_g f) = \chi(g) J(f)$$

for $f \in K(\mathbf{X})$ and $g \in G$. Equivalently, if J is represented by the Radon measure m, then m is *relatively* (left) *invariant with multiplier* χ if

$$\int f(g^{-1}x) m(dx) = \chi(g) \int f(x) m(dx)$$

for $f \in K(\mathbf{X})$ and $g \in G$.

Here are a couple of examples.

EXAMPLE 2.5. With $\mathbf{X} = R^n$ and $G = \mathrm{Gl}_n$ as in Example 2.1, G acts topologically on $\mathbf{X}$ and Lebesgue measure dx is relatively invariant with multiplier

$$\chi(g) = |\det(g)|. \qquad \square$$

EXAMPLE 2.6. With $\mathbf{X} = R^n$ and $G = O_n$, consider a probability measure P on $\mathbf{X}$ and let $X \in \mathbf{X}$ be a random vector with distribution P—this we write as $\mathscr{L}(X) = P$. Recall that X has a *spherical* distribution if $\mathscr{L}(X) = \mathscr{L}(gX)$ for $g \in O_n$. In terms of P, X is spherical iff P is invariant (with multiplier $\chi \equiv 1$) under O_n, that is, iff

$$P(B) = P(g^{-1}B)$$

for all Borel sets B and $g \in O_n$. If we define the probability measure gP by

$$(gP)(B) = P(g^{-1}B)$$

for all Borel sets B, then X is spherical iff $P = gP$ for $g \in O_n$. Since $\mathscr{L}(X) = P$ implies $\mathscr{L}(gX) = gP$, we have that X is spherical iff P is O_n invariant (i.e., $P = gP$). Notice that the group O_n acts on all the probability measures on R^n via the definition of gP. $\square$

EXAMPLE 2.7. For this example, let S_n be the vector space of $n \times n$ real symmetric matrices and let $G = \mathrm{Gl}_n$ act on S_n via

$$x \to gxg'.$$

With dx denoting Lebesgue measure on S_n, define an integral J by

$$J(f) = \int_{S_n} f(x)\,dx.$$

To see if J is relatively invariant (for some multiplier), consider

$$J(L_g f) = \int_{S_n} f\big(g^{-1}x(g^{-1})'\big)\,dx$$

and introduce the change of variables $y = g^{-1}x(g^{-1})'$ so $x = gyg'$. This change of variables defines a nonsingular linear transformation on S_n whose determinant is

$$(\det(g))^{n+1}$$

[see Eaton (1983), page 169 for a proof]. Thus

$$dx = |\det(g)|^{n+1}\,dy,$$

so $J(L_g f) = |\det(g)|^{n+1}J(f)$ and J is relatively invariant with the given multiplier. This example is considered again later. □

EXAMPLE 2.8. As in Example 2.4, let $\mathbf{X} \subset \mathscr{L}_{p,n}$ be all the $n \times p$ real matrices of rank p (so $p \leq n$) and take $G = O_n \times \mathrm{Gl}_p$ with the group action defined in Example 2.4. Let dx denote Lebesgue measure restricted to $\mathbf{X}$ and define the integral J by

$$J(f) = \int f(x)\frac{dx}{(\det x'x)^{n/2}} = \int f(x)m(dx)$$

for $f \in K(\mathbf{X})$. We now show J is invariant (relatively invariant with multiplier $\chi \equiv 1$). For $(\gamma, g) \in O_n \times \mathrm{Gl}_p$,

$$J(L_{(\gamma,g)}f) = \int f\big(\gamma' x(g^{-1})'\big)m(dx)$$

and the change of variable $y = \gamma' x(g^{-1})'$ yields $x = \gamma y g'$. This linear transformation $(y \to \gamma y g')$ on $\mathscr{L}_{p,n}$ has a Jacobian given by

$$|\det(g)|^n$$

[see Eaton (1983), page 168]. Substitution now yields

$$\begin{aligned} J(L_{(\gamma,g)}f) &= \int f(y)|\det(g)|^n \frac{dy}{|\det(g'y'yg')|^{n/2}} \\ &= \frac{|\det(g)|^n}{|\det gg'|^{n/2}} J(f) = J(f). \end{aligned}$$

Thus J is invariant. □

Now we turn to the question of existence and uniqueness of relatively invariant integrals with a given multiplier. Assume G acts topologically on $\mathbf{X}$ and assume G acts transitively on $\mathbf{X}$ (this is the only case where we can hope for uniqueness). Fix x_0 in $\mathbf{X}$ and assume that the function

$$\pi: g \to gx_0$$

on G to $\mathbf{X}$ is an open mapping (forward images of open sets are open). Note that π maps G onto X because G is assumed to be transitive. Further, let

$$H = \{g | gx_0 = x_0\}$$

so H is a closed subgroup of G. This subgroup is often called the *isotropy subgroup of* x_0. Let Δ_H denote to modulus of H and let Δ denote the modulus of G.

THEOREM 2.1 (Weil). *In order that there exists a relatively invariant integral J with multiplier χ on G, it is necessary and sufficient that χ satisfy the equation*

$$\Delta_H(h) = \chi(h)\Delta(h) \quad \textit{for } h \in H. \tag{2.1}$$

When J exists, it is unique up to a positive constant.

PROOF. See Nachbin (1965), Chapter 3, especially pages 125–141. □

A few remarks are in order concerning Theorem 2.1. The validity of the theorem does not depend on the choice of x_0, that is, if the result is true for one x_0, then it is true for all x_0's. The verification of (2.1) requires the calculation of both the modulus for G and the modulus for H. In the special case when $\mathbf{X}$ is compact and G is compact, Theorem 2.1 guarantees the existence of a unique G-invariant probability measure on $\mathbf{X}$. For future reference, we state this as:

THEOREM 2.2. *Assume the conditions of Theorem* 2.1 *and assume both* $\mathbf{X}$ *and G are compact. Then there exists a unique G-invariant probability measure on* $\mathbf{X}$.

PROOF. Because $\mathbf{X}$ is compact, all integrals are finite as the function $f \equiv 1$ is integrable. Since G is compact, H is compact so $\Delta = \Delta_H \equiv 1$ and χ must be identically 1 also. Thus, (2.1) holds so there exists a finite G-invariant integral (measure) on $\mathbf{X}$. Normalizing this to be a probability measure gives the uniqueness. □

Here are a couple of examples.

EXAMPLE 2.9. With Gl_n acting on R^n, observe that the basic assumption of transitivity of the group action is not satisfied. However Gl_n does act transitively on $R^n - \{0\} \equiv \mathbf{X}$. With

$$x_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in R^n,$$

the isotropy subgroup of x_0 is

$$H = \{h | h \in \mathrm{Gl}_n,\ hx_0 = x_0\}.$$

It is easily verified that $h \in H$ iff

$$h = \begin{pmatrix} 1 & b \\ 0 & a \end{pmatrix} \in \mathrm{Gl}_n,$$

where $a \in \mathrm{Gl}_{n-1}$ and b is $1 \times (n-1)$. The modulus of Gl_n is $\Delta \equiv 1$. To compute the modulus of H, a left invariant measure on H is first computed. Let $da\,db$ be Lebesgue measure restricted to H. For $k \in H$, consider

$$J(f) = \int f\left(k^{-1}\begin{pmatrix} 1 & b \\ 0 & a \end{pmatrix}\right) da\,db.$$

Set

$$\begin{pmatrix} 1 & c \\ 0 & d \end{pmatrix} = k^{-1}\begin{pmatrix} 1 & b \\ 0 & a \end{pmatrix},$$

where

$$k = \begin{pmatrix} 1 & \beta \\ 0 & \alpha \end{pmatrix}.$$

Then,

$$\begin{pmatrix} 1 & b \\ 0 & a \end{pmatrix} = k\begin{pmatrix} 1 & c \\ 0 & d \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ 0 & \alpha \end{pmatrix}\begin{pmatrix} 1 & c \\ 0 & d \end{pmatrix} = \begin{pmatrix} 1 & c + \beta d \\ 0 & \alpha d \end{pmatrix}.$$

The Jacobian of this transformation is

$$\phi(k) = |\det(\alpha)|^{n-1}.$$

Thus J is relatively invariant with multiplier ϕ, so from Theorem 1.6,

$$\nu_l(da, db) = \frac{da\,db}{|\det(b)|^{n-1}}$$

is a left-invariant measure on H. Thus, to compute Δ_H, consider

$$J_1(fR_k) = \int f(xk^{-1})\nu_l(dx)$$

with ν_l given above. Here, dx has been written for $da\,db$. The usual change of variable–Jacobian argument yields

$$J_1(fR_k) = |\det(\alpha)| J_1(f).$$

Therefore

$$\Delta_H(k) = |\det(\alpha)|,$$

where

$$k = \begin{pmatrix} 1 & \beta \\ 0 & \alpha \end{pmatrix}.$$

Now consider a multiplier

$$\chi_\delta(g) = |\det(g)|^\delta$$

on G where δ is a fixed real number. Then (2.1) holds iff

$$\Delta_H(h) = \chi_\delta(h) \quad \text{for } h \in H.$$

With

$$h = \begin{pmatrix} 1 & \beta \\ 0 & \alpha \end{pmatrix},$$

$$\chi_\delta(h) = |\det(\alpha)|^\delta,$$

so that (2.1) becomes

$$|\det(\alpha)| = |\det(\alpha)|^\delta$$

for all $\alpha \in \mathrm{Gl}_{n-1}$. Thus δ must be 1 and

$$\chi(g) = |\det(g)|$$

is the only multiplier on Gl_n which satisfies (2.1). Of course, Lebesgue measure on $R^n - \{0\}$ is relatively invariant with multiplier χ and is unique up to a positive constant. □

EXAMPLE 2.10. As in Example 2.3, consider the group O_n acting on $F_{p,n}$. This group action is transitive and both O_n and $F_{p,n}$ are compact sets. Thus there exists a unique O_n-invariant probability distribution on $F_{p,n}$ which we call the *uniform distribution* on $F_{p,n}$. When $p = 1$, this is just the uniform distribution on the sphere of radius 1 in R^n. When $p = n$, then we just get Haar measure on $O_n = F_{n,n}$.

This example is related to the problem of how to define what one means by "a randomly chosen subspace of dimension p in R^n." To see the connection, let $S_{p,n}$ denote all the rank p orthogonal projection matrices defined on R^n. Then $S_{p,n}$ is the image of $F_{p,n}$ under the mapping

$$x \to xx', \qquad x \in F_{p,n}.$$

This map is onto but, of course, not one-to-one. The group O_n acts on $S_{p,n}$ by

$$u \to gug', \qquad u \in S_{p,n},$$

as is suggested by mapping from $F_{p,n}$ to $S_{p,n}$ and the action of O_n on $F_{p,n}$. This action is transitive because for each $u \in S_{p,n}$, there is a $g \in O_n$ such that

$$gug' = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}.$$

That $S_{p,n}$ is a compact subset of $\mathscr{L}_{n,n}$ is easily established. Thus, there exists a unique O_n-invariant probability on $S_{p,n}$ which we call the *uniform distribution* on $S_{p,n}$ Since there is a one-to-one map between subspaces of dimension p and elements of $S_{p,n}$, this defines a uniform distribution on p dimensional subspaces. □

2.3. Maximal invariants. In this section we consider the problem of calculating explicit representations of orbit indices when a group G acts on a set $\mathbf{X}$. Recall that for $x \in \mathbf{X}$,

$$O_x = \{gx | g \in G\}$$

is the *orbit of* x and these orbits are the equivalence class of points which are equivalent under the action of G. That is, x_1 is equivalent to x_2 if $x_1 = gx_2$ for some $g \in G$. The problem is to describe what the orbits are for some examples. This was done in Example 2.2, but here general methods are described along with the examples. The main interest in orbit indices arises from problems in the construction of best invariant tests which is discussed in later chapters.

Let the group G act on $\mathbf{X}$ and suppose a function f maps $\mathbf{X}$ into $\mathbf{Y}$.

DEFINITION 2.4. The function f is *invariant* if $f(x) = f(gx)$. The function f is *maximal invariant* if f is invariant and if $f(x_1) = f(x_2)$ implies $x_1 = gx_2$ for some $g \in G$.

Thus, f is invariant iff f is constant on each orbit in $\mathbf{X}$. Also f is maximal invariant iff f is constant on each orbit and takes different values on different orbits. That is, maximal invariant functions provide an orbit index, namely, knowing the value of maximal invariant f at x is equivalent to knowing O_x. Notice that the image space $\mathbf{Y}$ plays no role in Definition 2.4. In examples, $\mathbf{Y}$ is ordinarily chosen with convenience in mind.

THEOREM 2.3. *Suppose f: $\mathbf{X} \to \mathbf{Y}_1$ is maximal invariant under the action of G on $\mathbf{X}$. Then a function h: $\mathbf{X} \to \mathbf{Y}_2$ is invariant iff there exists a function k, k: $\mathbf{Y}_1 \to \mathbf{Y}_2$, such that $h(x) = k(f(x))$.*

PROOF. If k exists, obviously h is invariant since f is invariant. Conversely, suppose h is invariant and define k on $\mathbf{Y}_1$ as follows:

(i) If $y \in \mathbf{Y}_1$ is given by $y = f(x)$, set $k(y) = h(x)$.
(ii) If $y \in \mathbf{Y}_1$ is not in the range of f, define k arbitrarily.

That k is well defined follows from the invariance of h and the maximal invariance of f. Obviously $h(x) = k(f(x))$. □

This result shows that once a maximal invariant is known, then all the invariant functions are known, namely, they are just the functions of a maximal invariant.

The first method we use to construct a maximal invariant might be termed "the reduction method." The idea is to "reduce" or "transform" a point x to a canonical form via elements of G is such a way as to pick out a particular point from each orbit.

EXAMPLE 2.11. Consider the group O_n acting on R^n, in the usual way. Given $x \in R^n$, there exists a $g \in O_n$ such that

$$gx = \|x\|x_0,$$

where $\|x\|$ is the length of x and x_0 is the first standard unit vector

$$x_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

For example, take g to have first row equal to $x'/\|x\|$. That the function

$$f(x) = \|x\|x_0$$

is maximal invariant is proved as follows. Clearly f is invariant since x and gx have the same length. Now $f(x) = f(y)$ iff $\|x\| = \|y\|$. Thus when $\|x\| = \|y\|$, we must find a $g \in G$ so that $gx = y$. Pick g_1 and g_2 such that

$$g_1 x = \|x\|x_0 = \|y\|x_0 = g_2 y.$$

Then $y = g_2^{-1} g_1 x$, so x and y are in the same orbit. Clearly, any one-to-one function of $\|x\|$ is also a maximal invariant, and a function is invariant iff it can be written as a function of $\|x\|$. □

EXAMPLE 2.12. As in Example 2.2, let $G = O_n$ act on S_n with action

$$x \to gxg', \qquad g \in O_n.$$

Given $x \in S_n$, the spectral theorem shows there is a $g \in O_n$ such that

$$gxg' = \lambda(x),$$

where $\lambda(x)$ is an $n \times n$ diagonal matrix with diagonal elements $\lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_n(x)$, which are the ordered eigenvalues of x. The claim is that

$$f(x) = \lambda(x)$$

is maximal invariant. Since eigenvalues are invariant under the group action $x \to gxg'$, f is invariant. If $f(x_1) = f(x_2)$, we must show that $x_2 = gx_1g'$ for some $g \in G$. Pick $g_1, g_2 \in G$ such that

$$g_i x_i g_i' = \lambda(x_i), \qquad i = 1,2.$$

With $g = g_2' g_1$,

$$gx_1g' = g_2'g_1x_1g_1'g_2 = g_2'\lambda(x_1)g_2 = g_2'\lambda(x_2)g_2 = x_2.$$

Thus the vector of ordered eigenvalues of x is a maximal invariant. □

EXAMPLE 2.13. For this example the product group $O_n \times O_p$ acts on $\mathscr{L}_{p,n}$ via

$$x \to gxh', \qquad g \in O_n, \ h \in O_p,$$

where we assume $p \le n$. The reduction argument here follows from the singular value decomposition theorem [for the version used here, see Anderson (1984), page 590]. According to this theorem, given $x \in \mathscr{L}_{p,n}$, there exists $g \in O_n$ and $h \in O_p$ such that

$$x = g\lambda(x)h',$$

where $\lambda(x)$: $n \times p$ has the form

$$\lambda(x) = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ 0 & & \ddots & \\ & & \cdots & \lambda_p \\ & & 0 & \end{pmatrix}$$

and $\lambda_1 \ge \cdots \ge \lambda_p \ge 0$ are the square roots of the eigenvalues of $x'x$. (The numbers $\lambda_1, \dots, \lambda_p$ are often called the *singular values* of x.) The claim is that

$$f(x) = \lambda(x)$$

is maximal invariant. Obviously f is invariant since singular values do not change under the given group action. If $f(x_1) = f(x_2)$ [so $\lambda(x_1) = \lambda(x_2)$], pick $g_1, g_2 \in O_n$ and h_1, $h_2 \in O_p$ such that

$$g_i x_i h_i' = \lambda(x_i), \qquad i = 1, 2.$$

Since $\lambda(x_1) = \lambda(x_2)$, this implies that

$$x_2 = g_2' g_1 x_1 h_1' h_2,$$

so x_1 and x_2 are in the same orbit. Thus f is maximal invariant. □

The next method of constructing maximal invariants involves finding a function τ mapping $\mathbf{X}$ into G, which has the property

$$\tau(gx) = g\tau(x), \tag{2.2}$$

where $g\tau(x)$ means the composition of the two group elements g and $\tau(x)$. Temporarily assume we can find such a τ and consider

$$f(x) = (\tau(x))^{-1}x,$$

where $(\tau(x))^{-1}$ is the inverse in G of $\tau(x)$. Using (2.2),

$$\begin{aligned} f(gx) &= (\tau(gx))^{-1}gx = (g\tau(x))^{-1}gx \\ &= (\tau(x)^{-1})g^{-1}gx = (\tau(x))^{-1}x = f(x), \end{aligned}$$

so f is invariant. To show f is a maximal invariant, suppose $f(x_1) = f(x_2)$. Then

$$(\tau(x_1))^{-1}x_1 = (\tau(x_2))^{-1}x_2,$$

which yields

$$x_2 = \left[\tau(x_2)(\tau(x_1))^{-1}\right]x_1 = gx_1.$$

This shows x_1 and x_2 are in the same orbit which entails the maximal invariance of f. Here are some examples where τ can be constructed.

EXAMPLE 2.14. With $G = R^1$ and $\mathbf{X} = R^n$, consider the action

$$x \to x + ge_n, \qquad x \in R^n,$$

where e_n is the vector of 1's in R^n and $g \in R^1$. Define τ by

$$\tau(x) = \bar{x},$$

where as usual $\bar{x} = n^{-1}\Sigma_1^n x_i$. An easy calculation shows (2.2) holds and thus

$$f(x) = (\tau(x))^{-1}x = x - \bar{x}e_n$$

is a maximal invariant. □

EXAMPLE 2.15. For this example elements of G consist of pairs (a, b) with $a > 0$ and $b \in R^1$ and the group operation is

$$(a_1, b_1)(a_2, b_2) = (a_1a_2, a_1b_2 + b_1).$$

With e_n denoting the vector of 1's in R^n, the set $\mathbf{X}$ is

$$\mathbf{X} = R^n - \operatorname{span}\{e_n\},$$

so $\mathbf{X}$ is R^n with a line removed. The action of G on $\mathbf{X}$ is

$$(a, b)x = ax + be_n.$$

For $x \in \mathbf{X}$, $\bar{x} \in R^1$ is as in the last example and

$$s(x) = \left[\sum_1^n (x_i - \bar{x})^2\right]^{1/2}.$$

Define τ on $\mathbf{X}$ to G by

$$\tau(x) = (s(x), \bar{x}).$$

Note that $s(x) > 0$ because the line where s vanishes has been removed from R^n. (This is another example of removing a "small" set.) An easy calculation shows that

$$\tau((a, b)x) = (a, b)(s(x), \bar{x}).$$

Since (2.2) holds,

$$f(x) = (\tau(x))^{-1}x = \frac{x - \bar{x}e_n}{s(x)}$$

is maximal invariant. □

In some examples, it is possible to calculate a maximal invariant by decomposing the group into subgroups and doing the calculation for each subgroup, that is, doing the calculation in steps. The applicability of this stepwise procedure depends on the notion of moving a group action from one space to another. Here is the idea. Suppose G acts on $\mathbf{X}$ and a fixed function f maps $\mathbf{X}$ onto $\mathbf{Y}$. In order to try to define a group action on $\mathbf{Y}$, one possibility is to write $y = f(x)$ and then define gy by

$$gy = f(gx). \tag{2.3}$$

Conditions under which this gives an unambiguous definition are provided by:

THEOREM 2.4. *Suppose that $f(x_1) = f(x_2)$ implies that $f(gx_1) = f(gx_2)$ for all $g \in G$. Then* (2.3) *defines G acting on $\mathbf{Y}$.*

PROOF. To see that gy is well defined, if $y = f(x_1) = f(x_2)$, then by assumption $f(gx_1) = f(gx_2)$ for all $g \in G$. Thus $gy = f(gx_1)$ for all $g \in G$ is an unambiguous specification of gy. Obviously, $ey = y$ for all $y \in \mathbf{Y}$. Also for $y = f(x)$ and $g_1, g_2 \in G$,

$$\begin{aligned} g_1(g_2 y) &= g_1 f(g_2 x) = f(g_1 g_2 x) \\ &= (g_1 g_2) f(x) = (g_1 g_2) y. \end{aligned}$$

Thus G acts on $\mathbf{Y}$. □

To describe the stepwise calculation of a maximal invariant, suppose G acts on $\mathbf{X}$ and suppose H and K are subgroups of G which *generate* G, that is, each $g \in G$ can be written in the form $g = h_1 k_1 h_2 k_2 \cdots h_r k_r$ for some integer r where $h_i \in H$ and $k_i \in K$, $i = 1, \dots, r$.

THEOREM 2.5. *Suppose that under the action of H on $\mathbf{X}$, the function f_1 mapping $\mathbf{X}$ onto $\mathbf{Y}$ is a maximal invariant and satisfies*

$$f_1(x_1) = f_1(x_2) \quad \textit{implies} \quad f_1(kx_1) = f_1(kx_2)$$

for all $k \in K$. Also suppose that f_2 mapping $\mathbf{Y}$ into $\mathbf{Z}$ is a maximal invariant under the induced action of K acting on $\mathbf{Y}$ (as described in Theorem 2.4). *Then $f(x) = f_2(f_1(x))$ is a maximal invariant under the action of G on $\mathbf{X}$.*

PROOF. Recall that the action of K on $\mathbf{Y}$ is defined by: write $y = f_1(x)$ and set $ky = f_1(kx)$. To show f is invariant, consider $h \in H$. Then

$$f(hx) = f_2(f_1(hx)) = f_2(f_1(x))$$

since f_1 is H invariant. For $k \in K$,

$$f(kx) = f_2(f_1(kx)) = f_2(kf_1(x)) = f_2(f_1(x))$$

by definition of $kf_1(x)$ and the invariance of f_2. It is now an easy induction argument to show that

$$f(h_1 k_1 h_2 k_2 \cdots h_r k_r x) = f(x)$$

for all $r = 1, 2, \dots$ and $h_i \in H$, $k_i \in K$. Thus $f(gx) = f(x)$ since H and K generate G. To show f is maximal, suppose $f(x_1) = f(x_2)$. Hence

$$f(x_1) = f_2(f_1(x_1)) = f_2(f_1(x_2))$$

which, by the maximality of f_2 implies there is a $k \in K$ so that

$$kf_1(x_1) = f_1(x_2) = f_1(kx_1).$$

The final equality follows from the definition of the action of K on $\mathbf{Y}$. The maximality of f_1 implies there is an $h \in H$ so that

$$x_2 = h(kx_1) = (hk)x_1.$$

Thus, $x_2 = gx_1$ for some g so x_1 and x_2 are in the same orbit. Thus f is maximal. □

The advantages of reducing in steps are most apparent when dealing with rather complicated problems. The next example, in which the correlation coefficient is a maximal invariant, well illustrates this situation.

EXAMPLE 2.16. With $n \geq 3$, let e_n denote the vector of 1's in R^n and set

$$Q = I_n - n^{-1}e_n e_n'.$$

Clearly, Q is the orthogonal projection onto the subspace perpendicular to e_n. Let $\mathbf{X} \subset \mathscr{L}_{2,n}$ be the set of $n \times 2$ real matrices x such that Qx has rank 2. Note that

$$s(x) = x'Qx$$

is the (unnormalized) sample covariance matrix when one observes n two dimensional vectors and arranges them into x: $n \times 2$ whose rows are the transposes of the data vectors. Points in $\mathbf{X}$ are those sample points such that $s(x)$ has rank 2, that is, $s(x)$ is positive definite.

The group for this example is a product group $G = G_1 \times G_2$ with

$$G_1 = \{\gamma | \gamma \in O_n, \gamma e_n = e_n\}.$$

The group G_2 is a subgroup of Al_2. An element of G_2 is a pair (a, b) with $b \in R^2$ and

$$a = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \in \mathrm{Gl}_2, \qquad a_i > 0,\ i = 1,2.$$

The action of an element $(\gamma, (a, b))$ in $G_1 \times G_2$ on $x \in \mathscr{L}_{2,n}$ is

$$(\gamma, (a, b))x = \gamma x a' + e_n b',$$

where the prime denotes transpose. Even though $a' = a$, we write a' to remind the reader that transposes are necessary in such situations to insure that actions are indeed left actions. That we have a left action is easy to check, as is the fact that points in $\mathbf{X}$ are mapped into points in $\mathbf{X}$.

To construct a maximal invariant, consider the two subgroups H and K defined as follows. With $e \in \mathrm{Gl}_2$ denoting the 2×2 identity matrix and $I \in O_n$ denoting the $n \times n$ identity matrix, let

$$H = \{(\gamma, (e, b)) | \gamma \in G_1, (e, b) \in G_2\}$$

and

$$K = \{(I, (a, 0)) | (a, 0) \in G_2\}.$$

Since every element $(\gamma, (a, b))$ in G can be expressed as

$$(\gamma, (a, b)) = (\gamma, (e, b))(I, (a, 0)),$$

H and K generate G.

Our first claim is that f_1 defined on $\mathbf{X}$ onto S_2^+ (the set of 2×2 positive definite matrices) by

$$f_1(x) = s(x)$$

is maximal invariant under the action of H on $\mathbf{X}$. To see that f_1 is H-invariant, consider $(\gamma, (e, b)) \in H$ so

$$(\gamma, (e, b))x = \gamma x + e_n b'.$$

Thus

$$Q(\gamma, (e, b))x = Q\gamma x = \gamma Q x$$

because $Qe_n = 0$ and Q commutes with each $\gamma \in G_1$. Because $Q = Q' = Q^2$,

$$\begin{aligned} f_1((\gamma, (e, b))x) &= [(\gamma, (e, b))x]'Q[(\gamma, (e, b))x] \\ &= (\gamma Q x)'(\gamma Q x) = x'Qx = f_1(x) \end{aligned}$$

and hence f_1 is invariant. For the maximality of f_1, suppose

$$f_1(x) = x'Qx = f_1(y) = y'Qy$$

with $x, y \in \mathbf{X}$. Since $Q^2 = Q = Q'$,

$$(Qx)'Qx = (Qy)'Qy,$$

which implies there is a $\gamma \in G_1$ such that

$$\gamma Q y = Qx.$$

The existence of this γ follows from a minor modification of Proposition 1.20 in Eaton (1983). From the definition of Q we have

$$x = x - e_n\bar{x}' + e_n\bar{x}' = Qx + e_n\bar{x}',$$

where $\bar{x} = x'e_n/n \in R^2$. Therefore,

$$\begin{aligned} x &= Qx + e_n\bar{x}' = \gamma Q y + e_n\bar{x}' = \gamma(y - e_n\bar{y}') + e_n\bar{x}' \\ &= (\gamma, (0, \bar{x} - \bar{y}))y. \end{aligned}$$

Hence x and y are in the same H orbit, so f_1 is maximal.

To apply Theorem 2.5, it must be verified that

$$f_1(x) = f_1(y) \quad \text{implies} \quad f_1(kx) = f_1(ky)$$

for all $k \in K$. With

$$k = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \in K,$$

it is an easy calculation to verify that

$$s(kx) = ks(x)k'.$$

Thus $f_1(x) = f_1(y)$ implies that

$$f_1(kx) = s(kx) = ks(x)k' = ks(y)k' = s(ky) = f(ky),$$

so Theorem 2.5 applies.

Finally, a maximal invariant for the action of K on S_2^+ given by

$$s \to ksk', \qquad k \in K,$$

needs to be found. Writing

$$s = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \in S,$$

an easy argument shows that

$$f_2(s) = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}$$

is a maximal invariant on S_2^+. Hence by Theorem 2.5,

$$f(x) = \frac{s_{12}(x)}{\sqrt{s_{11}(x)s_{22}(x)}}$$

is a maximal invariant under the action of G on $\mathbf{X}$. Of course, $f(x)$ is just the sample correlation coefficient. □

2.4. Induced group actions: Equivariance. The situation described in Theorem 2.4 is one circumstance in which it is possible to induce a group action on one space, given a group action on anther space. In this section other circumstances are discussed where group actions can be induced on a space, based on a group action on an associated space. Applications of these ideas occur throughout these notes.

First, suppose that a group G acts on $\mathbf{X}$ and let f be a function on $\mathbf{X}$ to $\mathbf{Y}$. Here is a useful definition of a new function, denoted by gf for $g \in G$, which also maps $\mathbf{X}$ to $\mathbf{Y}$:

$$(gf)(x) = f(g^{-1}x). \tag{2.4}$$

It is clear that $ef = f$ where e is the identity in G and

$$g_1(g_2 f) = (g_1 g_2)f, \tag{2.5}$$

that is

$$\begin{aligned}(g_1(g_2 f))(x) &= (g_2 f)(g_1^{-1}x) = f(g_2^{-1}g_1^{-1}x) = f((g_1g_2)^{-1}x) \\ &= ((g_1g_2)f)(x).\end{aligned}$$

Of course, the reason for the inverse in the definition (2.4) is so that (2.5) is valid. Hence if $\mathbf{Z}$ is a set of functions from $\mathbf{X}$ to $\mathbf{Y}$ such that $f \in \mathbf{Z}$ implies $gf \in \mathbf{Z}$ for all $g \in G$, then G acts on $\mathbf{Z}$ via (2.4). This group action on the set of functions $\mathbf{Z}$ should not be confused with the group action induced on $\mathbf{Y}$ which is described in Theorem 2.4.

EXAMPLE 2.17. Take $\mathbf{X} = \{1, 2, \ldots, n\}$ and let G be the group of all one-to-one functions from $\mathbf{X}$ to $\mathbf{X}$ (the permutation group of $\mathbf{X}$). Let $\mathbf{Z}$ be the set of all real valued functions defined on $\mathbf{X}$. A point in $\mathbf{Z}$ can be thought of as a vector in

R^n, namely, $f \in \mathbf{Z}$ corresponds to a vector in R^n given by

$$\begin{pmatrix} f(1) \\ f(2) \\ \vdots \\ f(n) \end{pmatrix} \in R^n.$$

Conversely, every point in R^n is a point in $\mathbf{Z}$. According to (2.4), if $f \in R^n$ has ith coordinate $f(i)$, then gf has ith coordinate $(gf)(i) = f(g^{-1}i)$. This definition yields the left group action of the permutation group G on R^n. There is some confusion in the literature concerning the action of the permutation group on R^n. □

EXAMPLE 2.18. Consider a measurable space $(\mathbf{X}, \mathscr{B})$ and suppose that G acts on $\mathbf{X}$ in such a way that each g is a bimeasurable mapping. In this circumstance, we say that G acts *measurably* on $(\mathbf{X}, \mathscr{B})$. Thus, G acts on the σ-algebra $\mathscr{B}$ in the obvious way:

$$gB = \{x | x = gy,\ y \in B\}.$$

Now, let P be a probability measure and think of P as a map from $\mathscr{B}$ into $[0,1]$. Then, according to (2.4) (with $\mathbf{X} = \mathscr{B}$ and $f = P$),

$$(gP)(B) = P(g^{-1}B).$$

In other words, gP is a probability measure on $\mathscr{B}$ whose value at B is $P(g^{-1}B)$.

In terms of random variables, the above means that if $\mathscr{L}(X) = P$, then $\mathscr{L}(gX) = gP$ for X taking values in $\mathbf{X}$. To see this, first observe that $\mathscr{L}(X) = P$ means

$$\mathrm{Prob}(X \in B) = P(B).$$

Therefore,

$$\mathrm{Prob}(gX \in B) = \mathrm{Prob}(X \in g^{-1}B) = P(g^{-1}B) = (gP)(B),$$

so $\mathscr{L}(gX) = gP$. This definition of gP appeared in Example 2.10. □

EXAMPLE 2.19. This example deals with induced group actions for randomized decision functions (also known as Markov kernels, transition probability kernels, among other things). Consider two measurable spaces $(\mathbf{X}, \mathscr{B}_1)$ and $(\mathbf{Y}, \mathscr{B}_2)$. A *Markov kernel* δ is a function defined on $\mathscr{B}_2 \times \mathbf{X}$ into $[0,1]$ such that:

(i) For each $x \in \mathbf{X}$, $\delta(\cdot|x)$ is a probability measure on $\mathscr{B}_2$.
(ii) For each $B \in \mathscr{B}_2$, $\delta(B|\cdot)$ is a $\mathscr{B}_1$ measurable function.

Suppose that the group G acts measurably on both $(\mathbf{X}, \mathscr{B}_1)$ and $(\mathbf{Y}, \mathscr{B}_2)$ so G acts on $\mathscr{B}_2 \times \mathbf{X}$ via

$$g(B, x) = (gB, gx).$$

If δ is Markov kernel, then according to (2.4) (with $\mathbf{X} = \mathbf{B}_2 \times \mathbf{X}$ and $f = \delta$) G

acts on δ via

$$(g\delta)(B|x) = \delta(g^{-1}B|g^{-1}x).$$

This example occurs later in the discussion of invariance and decision theory. □

Yet another method for inducing a group action concerns what might be called the "kernel method." Suppose G acts on $\mathbf{X}$ and $K(x, y)$ is a function of two variables defined on $\mathbf{X} \times \mathbf{Y}$. The range space of the kernel K is not relevant. The idea here is to try to use K to induce a group action on $\mathbf{Y}$ in such a way that K becomes an invariant function. To say it another way, when can a group action on $\mathbf{Y}$ be specified so that $K(x, y) = K(gx, gy)$ for all x, y, g?

THEOREM 2.6. *Consider G acting on $\mathbf{X}$ and suppose K is defined on $\mathbf{X} \times \mathbf{Y}$. Assume that for each $g \in G$ and each $y \in \mathbf{Y}$, there exists a unique $y' \in \mathbf{Y}$ such that*

$$K(gx, y') = K(x, y) \tag{2.6}$$

for all $x \in \mathbf{X}$. Then G acts on $\mathbf{Y}$ via the defined group action $gy = y'$. With this group action,

$$K(gx, gy) = K(x, y),$$

so K is an invariant function on $\mathbf{X} \times \mathbf{Y}$.

PROOF. That $ey = y$ for $y \in Y$ is clear by the uniqueness of y'. To verify we have a left action, consider $(g_1g_2)y$ and use (2.6) to compute as follows:

$$K((g_1g_2)x, (g_1g_2)y) = K(x, y) \quad \text{for } x \in \mathbf{X}$$

implies that

$$K(z, (g_1g_2)y) = K\big(g_2^{-1}(g_1^{-1}z), y\big) = K\big(g_1^{-1}z, g_2y\big) = K(z, g_1(g_2y))$$

for all $z \in \mathbf{X}$. Again uniqueness implies

$$(g_1g_2)y = g_1(g_2y).$$ □

EXAMPLE 2.20. Consider a parametric family of probability measures $\mathscr{P} = \{P(\cdot|\theta)|\theta \in \Theta\}$, defined on a measurable space $(\mathbf{X}, \mathscr{B})$. Suppose that G acts measurably on $(\mathbf{X}, \mathscr{B})$. The family $\mathscr{P}$ is *G-invariant* if $P \in \mathscr{P}$ implies $gP \in \mathscr{P}$ for all $P \in \mathscr{P}$ and $g \in G$. Here the notation gP of Example 2.18 is used.

Assuming the family $\mathscr{P}$ is G-invariant, also assume that

$$P(B|\theta_1) = P(B|\theta_2) \quad \text{for all } B \in \mathscr{B}$$

implies that $\theta_1 = \theta_2$, that is, the points in Θ are in one-to-one correspondence with the elements of the family $\mathscr{P}$. To apply Theorem 2.6, define K on $\mathbf{B} \times \Theta$ by

$$K(B, \theta) = P(B|\theta).$$

To show that the assumptions on K assumed in Theorem 2.6 hold, consider $\theta \in \Theta$ and $g \in G$. Since $\mathscr{P}$ is invariant, $P(g^{-1}B|\theta) = (gP(\cdot|\theta))(B)$ is in $\mathscr{P}$, so there exists a $\theta' \in \Theta$ such that

$$(gP(\cdot|\theta))(B) = P(B|\theta')$$

for all $B \in \mathscr{B}$. In terms of K, this means that there is a θ' such that

$$K(B, \theta') = K(g^{-1}B, \theta)$$

for all B and by the assumption on the family $\mathscr{P}$, θ' is unique. Theorem 2.6 implies that the natural induced group action on Θ yields

$$P(gB|g\theta) = P(B|\theta)$$

for $g \in G$, $B \in \mathscr{B}$ and $\theta \in \Theta$. Thus, if $\mathscr{L}(X) = P(\cdot|\theta)$, then $\mathscr{L}(gX) = P(\cdot|g\theta)$ because

$$\mathscr{L}(gX) = gP(\cdot|\theta).$$

This example is treated more completely in the next lecture. □

Finally, the notation of an equivariant function is introduced. A special case of this notion arose in the construction of a maximal invariant via the function τ in Equation (2.2). For the general case, suppose a group G acts on both $\mathbf{X}$ and $\mathbf{Y}$.

DEFINITION 2.5. A function f on $\mathbf{X}$ to $\mathbf{Y}$ is *equivariant* if

$$f(gx) = gf(x) \quad \text{for } g \in G,\ x \in \mathbf{X}. \tag{2.7}$$

The terminology in the statistical literature is not consistent. In some works, condition (2.5) is called invariance, but recently the tendency has been to the word equivariance. Note that when the group action of G on $\mathbf{Y}$ is trivial (that is, $gy = y$ for all g and all y), then equivariance reduces to invariance.

Given G acting on $\mathbf{X}$ and $\mathbf{Y}$, it seems rather difficult to give a description of all the equivariant functions. However, given G acting on $\mathbf{X}$ and given a function f, the results of Theorem 2.4 give the necessary and sufficient condition for the existence of a group action on $\mathbf{Y}$ such that (2.7) holds. In fact the condition of Theorem 2.4,

$$f(x_1) = f(x_2) \quad \text{implies that} \quad f(gx_1) = f(gx_2) \quad \text{for all } g \in G, \tag{2.8}$$

is precisely the necessary and sufficient condition that f be equivariant according to (2.7). Theorem 2.4 establishes the implication in one direction. That (2.7) implies (2.8) is obvious.

Equivariant functions arise naturally in estimation problems which are invariant under a group (these are discussed in detail later). Here are some examples which are related to estimation problems.

EXAMPLE 2.21. Take $\mathbf{X} = \mathscr{L}_{p,n}$ as the vector space of $n \times p$ real matrices and $\mathbf{Y} = S_p$ as the vector space of $p \times p$ real symmetric matrices. The group Gl_p acts on $\mathbf{X}$ by

$$x \to g(x) = xg', \qquad g \in \mathrm{Gl}_p$$

and Gl_p acts on S_p by

$$y \to g(y) = gyg', \qquad g \in \mathrm{Gl}_p.$$

Fix an $n \times n$ symmetrix matrix B and define f by

$$f(x) = x'Bx.$$

Then

$$f(g(x)) = f(xg') = gx'Bxg' = g(f(x))$$

so f is equivariant. □

EXAMPLE 2.22. For this example, G is the group G_T^+ of $p \times p$ lower triangular matrices with positive diagonal elements, $\mathbf{X}$ is the set of all $n \times p$ real matrices of rank p and $\mathbf{Y}$ is G_T^+. Recall that each $x \in \mathbf{X}$ can be written uniquely as

$$x = \gamma g',$$

where γ is an $n \times p$ matrix which satisfies $\gamma'\gamma = I_p$ and $g \in G_T^+$ [see Proposition 5.2 in Eaton (1983)]. Define f on $\mathbf{X}$ to G_T^+ by $f(x)$ is the unique element in G_T^+ such that

$$x = \gamma(f(x))'$$

as above. With G_T^+ acting on $\mathbf{X}$ by

$$x \to g(x) = xg'$$

and with G_T^+ acting on itself via left multiplication

$$h \to gh, \qquad h \in G_T^+,$$

it is easily verified that

$$f(g(x)) = gf(x).$$

Thus f is equivariant. □

Equivariant functions can, under certain conditions, be used in conjunction with Haar measure arguments to define invariant integrals.

EXAMPLE 2.23. Let S_p^+ be the set of positive definite matrices and let G_T^+ be the group of $p \times p$ lower triangular matrices with positive diagonal elements. The function ϕ on G_T^+ to S_p^+ defined by

$$\phi(h) = hh'$$

is one-to-one, onto, bicontinuous (a homeomorphism) and is equivariant,

$$\phi(gh) = ghh'g' = g(hh'),$$

where G_T^+ acts on S_p^+ in the usual way:

$$x \to g(x) = gxg'.$$

The group G_T^+ acts transitively on S_p^+ and Theorem 2.1 together with Example 2.7 shows that the integral

$$J_1(f) = \int_{S_p^+} f(x) \frac{dx}{(\det(x))^{(p+1)/2}}$$

is invariant under this group action.

However, consider the integral

$$J_2(f) = \int_{G_T^+} f(\phi(h)) \nu_l(dh) = \int f(hh') \nu_l(dh)$$

for $f \in K(S_p^+)$, where ν_l is a left-invariant measure on G_T^+. The equivariance of ϕ shows that

$$J_2(L_g f) = \int f(g^{-1}(\phi(h)))\nu_l(dh) = \int f(\phi(g^{-1}h))\nu_l(dh) = J_2(f)$$

for $g \in G_T^+$ and $f \in K(S_p^+)$. The uniqueness assertion of Theorem 2.1 shows there is a fixed constant $c > 0$ such that

$$\int f(x)\frac{dx}{(\det(x))^{(p+1)/2}} = c\int f(hh')\nu_l(dh)$$

for all $f \in K(S_p^+)$ and hence for all f which are integrable. The value of the constant c depends on the explicit choice for ν_l. With the choice

$$\nu_l(dh) = \frac{dh}{\prod_{i=1}^p h_{ii}^i}$$

as in Example 1.10, the constant c is 2^p. This is proved by choosing

$$f(x) = |\det(x)|^r \exp[-\tfrac{1}{2}\operatorname{tr} x]$$

and evaluating the two integrals for some convenient choice of the number r. □