CHAPTER 4

Models Invariant under Compact Groups

Our goal here is to understand the structure of probability measures which are invariant under a compact group. In the first section, a basic representation theorem is proved and is interpreted in terms of random variables. Section 2 contains some basic examples while applications to some robustness problems are given in Section 3.

4.1. A representation theorem. Throughout this section, G is a compact group and ν is the unique left (and right) invariant probability measure on G. In some situations, it is convenient to express certain equations in terms of a "random group element" U which has distribution ν. This is written $\mathscr{L}(U) = \nu$ and we say U has a uniform distribution on G. What this means is that U is a random element of G and the expectation of any bounded measurable function of U is computed as

$$\mathscr{E}f(U) = \int_G f(u)\nu(du).$$

Naturally, the distribution of U is characterized by its invariance. That is, the equation

$$\mathscr{L}(U) = \mathscr{L}(gU), \qquad g \in G,$$

characterizes the distribution of U because of the uniqueness of the invariant measure ν.

EXAMPLE 4.1. Let G be the group of $n \times n$ real orthogonal matrices O_n. The existence of ν on O_n is given by general theory, but here we outline a "construction" of U, and hence ν, which uses the normal distribution. Let $X_1, \dots, X_n$ be iid $N(0, I_n)$ random vectors in R^n. That $X_1, \dots, X_n$ are linearly independent with probability 1 is not hard to prove [see Proposition 7.1 in Eaton (1983) for example]. Let $Y_1, \ldots, Y_n$ be an orthonormal basis for R^n obtained by performing the Gram–Schmidt orthogonalization procedure to $X_1, \ldots, X_n$ in that order. Consider $U \in O_n$ with columns $Y_1, \ldots, Y_n$ so that $U = U(X_1, \ldots, X_n)$ is a function of $X_1, \ldots, X_n$. Hence U is a random element of O_n.

Using the definition of the Gram–Schmidt procedure, it is easy to show

$$U(gX_1, \ldots, gX_n) = gU(X_1, \ldots, X_n) \tag{4.1}$$

for each $g \in O_n$. In other words, the orthonormal basis one obtains from $gX_1, \ldots, gX_n$ is the same as one obtains by first constructing an orthonormal basis from $X_1, \ldots, X_n$ and then transforming the basis by $g \in O_n$.

The claim is that U is uniform on O_n. To see this, first observe that

$$\mathscr{L}(X_1, \ldots, X_n) = \mathscr{L}(gX_1, \ldots, gX_n)$$

because $X_1, \ldots, X_n$ are iid $N(0, I_n)$. Using (4.1), we then have

$$\begin{aligned}\mathscr{L}(U) &= \mathscr{L}(U(X_1, \ldots, X_n))\\ &= \mathscr{L}(U(gX_1, \ldots, gX_n)) = \mathscr{L}(gU(X_1, \ldots, X_n))\\ &= \mathscr{L}(gU).\end{aligned}$$

But, as remarked earlier, the relation $\mathscr{L}(U) = \mathscr{L}(gU)$ for $g \in O_n$ characterizes the distribution of U as being uniform. With $\nu = \mathscr{L}(U)$, ν is the unique invariant probability on O_n. □

Now consider a space $\mathbf{X}$ and suppose the compact group G acts topologically on $\mathbf{X}$. Let P be a probability measure defined on the Borel σ-algebra of $\mathbf{X}$ and define a new probability measure P_1 by

$$P_1 = \int_G gP\nu(dg). \tag{4.2}$$

Equation (4.2) means

$$P_1(B) = \int_G (gP)(B)\nu(dg) = \int_G P(g^{-1}B)\nu(dg), \tag{4.3}$$

or in terms of $f \in K(\mathbf{X})$,

$$\int f(x)P_1(dx) = \int_G \int_{\mathbf{X}} f(gx)P(dx)\nu(dg). \tag{4.4}$$

Using (4.3), it is obvious that $hP_1 = P_1$ because ν is the invariant probability measure on G. Thus, averaging gP with respect to ν produces a G invariant probability. Also observe that if P in (4.2) is invariant, then $P = P_1$.

There is an alternative way to write (4.2) which is also interesting. To this end, let U be uniform on G and for $x \in \mathbf{X}$, consider the random element $Ux \in \mathbf{X}$. For a Borel set $B \subset \mathbf{X}$, the probability that Ux is in B is

$$\text{Prob}(Ux \in B) = \nu\{g\colon gx \in B\}.$$

The induced distribution of Ux on $\mathbf{X}$ is denoted by μ_x. In other words,

$$\mu_x(B) = \nu\{g | gx \in B\},$$

so that

$$\mathscr{E}f(Ux) = \int_{\mathscr{X}} f(z)\mu_x(dz) = \int_G f(gx)\nu(dg). \tag{4.5}$$

THEOREM 4.1. *For each $x \in \mathbf{X}$, the probability μ_x is G-invariant.*

PROOF. Since $\mathscr{L}(Ux) = \mu_x$,

$$g\mu_x = \mathscr{L}(g(Ux)) = \mathscr{L}((gU)x) = \mathscr{L}(Ux) = \mu_x,$$

where the next to the last equality follows from the uniformity of U on G. □

The random variable $Ux \in \mathbf{X}$ obviously takes its value in the orbit of x since $U \in G$. Thus $\mu_x = \mathscr{L}(Ux)$ is a G-invariant probability measure on $O_x = \{gx | g \in G\}$. Since G acts transitively on O_x and G is compact, Theorem 2.2 shows that μ_x is in fact the unique invariant probability measure on O_x (modulo checking the regularity conditions needed to apply Theorem 2.2). This turns out to be a useful way to think about μ_x.

Since μ_x is G-invariant for each $x \in \mathbf{X}$, the average of μ_x (over $\mathbf{X}$) is also G-invariant. In symbols,

$$P_1 = \int \mu_x P(dx)$$

is invariant. This equation means

$$P_1(B) \equiv \int \mu_x(B) P(dx)$$

for each Borel set B. The next result shows that every invariant probability has such a representation as an average of the μ_x's.

THEOREM 4.2. *Suppose P is a G-invariant probability measure on $\mathbf{X}$. Then*

$$P = \int_{\mathbf{X}} \mu_x P(dx). \tag{4.6}$$

PROOF. When P is invariant, Equation (4.2) is

$$P = \int gP\nu(dg),$$

which expressed as in (4.4), with $f = I_B$, is

$$P(B) = \int_G \int_{\mathbf{X}} I_B(gx) P(dx)\nu(dg).$$

Interchanging integrals and using the definition of μ_x yields

$$P(B) = \int_{\mathbf{X}} \int_G I_B(gx)\nu(dg) P(dx)$$
$$= \int_{\mathbf{X}} \mu_x(B) P(dx)$$

which is just (4.6). □

Theorem 4.2 is unsatisfactory because of the following. Notice that

$$\mu_{gx} = \mathscr{L}(Ugx) = \mathscr{L}(Ux) = \mu_x$$

because $\nu = \mathscr{L}(U)$ is both left- and right-invariant. Thus, $x \to \mu_x$ is an invariant function and hence can be written as a function of a maximal invariant, say $t(x)$. Thus it should be possible to express the average (4.6) as an average over the maximal invariant (i.e., over the orbits in $\mathbf{X}$) rather than over the whole space $\mathbf{X}$. In order to make this precise, we need the notion of a measurable cross section. The Borel σ-algebra of $\mathbf{X}$ is denoted by $\mathscr{B}$.

DEFINITION 4.1. A subset $\mathbf{Y} \subset \mathbf{X}$ is a *measurable cross section* if:

(i) $\mathbf{Y}$ is measurable.
(ii) For each x, $\mathbf{Y} \cap O_x$ consists of exactly one point, say $y(x)$.
(iii) The function t defined on $\mathbf{X}$ to $\mathbf{Y}$ by $t(x) = y(x)$ is $\mathscr{B}$ measurable when $\mathbf{Y}$ has the σ-algebra $\{B \cap Y | B \in \mathscr{B}\} = \mathscr{B}_1$.

Assume $\mathbf{Y}$ is a measurable cross-section.

THEOREM 4.3. *For each probability Q defined on $(\mathbf{Y}, \mathscr{B}_1)$, the measure*

$$P = \int_{\mathbf{Y}} \mu_y Q(dy) \tag{4.7}$$

is a G-invariant probability on $(\mathbf{X}, \mathscr{B})$. Conversely, if P is a G-invariant probability on $(\mathbf{X}, \mathscr{B})$, then there exists a probability Q on $(\mathbf{Y}, \mathscr{B}_1)$ such that (4.7) *holds.*

PROOF. Equation (4.7) means that

$$P(B) = \int_{\mathbf{Y}} \mu_y(B) Q(dy)$$

or in terms of a bounded measurable function f,

$$\int f(x) P(dx) = \int_{\mathbf{Y}} \int_G f(gy) \nu(dg) Q(dy). \tag{4.8}$$

Because $(g, x) \to gx$ is jointly continuous and hence jointly measurable, joint measurability of $(g, y) \to f(gy)$ is easily verified. Thus (4.8) makes sense. That (4.7) defines a G-invariant probability is easily checked because μ_y is G-invariant. For the converse, consider a probability P which is invariant. Then, for any bounded measurable function f defined on $\mathbf{X}$, we have

$$\int f(x) P(dx) = \int f(gx) P(dx)$$

for $g \in G$. Integration then yields

$$\int_{\mathbf{X}} f(x) P(dx) = \int_G \int_{\mathbf{X}} f(gx) P(dx) \nu(dg) = \int_{\mathbf{X}} \int_G f(gx) \nu(dg) P(dx). \tag{4.9}$$

For x fixed, there exists a $g_0 \in G$ such that $x = g_0 t(x)$ because x and $t(x)$ are in the same orbit. The invariance of ν gives

$$\int_G f(gx)\nu(dg) = \int_G f(gg_0 t(x))\nu(dg) = \int f(gt(x))\nu(dg).$$

The assumed measurability of t implies that

$$H(t(x)) = \int_G f(gt(x))\nu(dg)$$

is a measurable function. Define the measure Q on $(Y, \mathscr{B}_1)$ by

$$Q(B) = P(t^{-1}(B))$$

so that

$$\int_{\mathbf{X}} H(t(x))P(dx) = \int_{\mathbf{Y}} H(y)Q(dy).$$

Thus from (4.9), we have

$$\int_{\mathbf{X}} f(x)P(dx) = \int_{\mathbf{X}}\int_G f(gt(x))\nu(dg)P(dx),$$

which is just (4.8). □

The interpretation of Theorem 4.3 in terms of random variables is the following. In our earlier notation, let U be uniform on G. Also let Y be a random variable taking values in $(\mathbf{Y}, \mathscr{B}_1)$ which is independent of U. That is, U and Y are defined on some probability space for which they are independent. Now, form

$$X = UY \in \mathbf{X},$$

where by UY, we mean the group element U acting on Y. Thus X is a random variable in $\mathbf{X}$. Because U and Y are independent,

$$\mathscr{L}(gX) = \mathscr{L}(g(UY)) = \mathscr{L}(UY) = \mathscr{L}(X)$$

since

$$\mathscr{L}(gU) = \mathscr{L}(U), \qquad g \in G.$$

With $Q = \mathscr{L}(Y)$ and $P = \mathscr{L}(X)$, it follows immediately that for any bounded measurable function f,

$$\mathscr{E}f(X) = \int_{\mathbf{X}} f(x)P(dx) = \mathscr{E}f(UY) = \int_{\mathbf{Y}}\int_G f(gy)\nu(dg)Q(dy),$$

which is just (4.8) again. Conversely, suppose X takes values in $\mathbf{X}$ and $\mathscr{L}(X) = P = \mathscr{L}(gX)$. Then Equation (4.8) immediately implies the existence of two random variables $U \in G$ and $Y \in \mathbf{Y}$ which are independent, $\mathscr{L}(U) = \nu$ and $\mathscr{L}(Y) = Q$, such that $\mathscr{L}(X) = \mathscr{L}(UY)$. Summarizing this gives:

THEOREM 4.4. *Let U be uniform on G and let $\mathbf{Y}$ be a measurable cross section. For a random variable $X \in \mathbf{X}$, the following are equivalent*:

(i) $\mathscr{L}(X) = \mathscr{L}(gX)$, $g \in G$.

(ii) *There exists a random variable $Y \in \mathbf{Y}$ which is independent of U such that $\mathscr{L}(X) = \mathscr{L}(UY)$.*

Various versions of Theorems 4.3 and 4.4 and related results appear in the mathematical and statistical literature. Here are a few relevant references: Wijsman (1957), Farrell (1962), Hall, Wijsman and Ghosh (1965), Dawid (1977), Eaton and Kariya (1984) and the references therein.

4.2. Some standard examples. In the examples below, the random variable notation of Theorem 4.4 is used rather than the more cumbersome notation of Theorem 4.3.

EXAMPLE 4.2. This example is related to exchangeable 0-1 valued random variables. The space $\mathbf{X}$ consists of the set of all n vectors x whose coordinates are only 0 or 1. Thus $\mathbf{X}$ has 2^n elements and we usually write $\mathbf{X} = \{0,1\}^n$. The group $\mathscr{P}_n$ of $n \times n$ permutation matrices acts on the vectors x by matrix multiplication. Of course, if the random vector $X \in \mathbf{X}$ satisfies $\mathscr{L}(X) = \mathscr{L}(gX)$, $g \in \mathscr{P}_n$, then X is *exchangeable*. To apply Theorem 4.4, let $\mathbf{Y} = \{y_0, y_1, \dots, y_n\}$, where $y_i \in \mathbf{X}$ has its first i coordinates equal to 1 and the remaining coordinates equal to 0. That $\mathbf{Y}$ is a cross section is clear.

To say U is uniform on $\mathscr{P}_n$ means that U is picked at random from the set of $n!$ permutation matrices. Theorem 4.4 implies that X is exchangeable iff

$$\mathscr{L}(X) = \mathscr{L}(UY),$$

where Y has an arbitrary distribution on $\mathbf{Y}$ and is independent of U. In other words, X is exchangeable iff X is generated by first picking a $y_i \in \mathbf{Y}$ according to some distribution and then randomly permuting the elements of the picked y_i. □

EXAMPLE 4.3. Here we return to the spherical distributions on $R^n = \mathbf{X}$. With $G = O_n$, a random vector $X \in R^n$ which satisfies $\mathscr{L}(X) = \mathscr{L}(gX)$, $g \in O_n$ has a spherical distribution. Let y_0 be a fixed vector of length 1 in R^n and set

$$\mathbf{Y} = \{\alpha y_0 | \alpha \in R^1,\ \alpha \geq 0\}.$$

Since the orbits in R^n are spheres of a given radius, it is clear that $\mathbf{Y}$ intersects each orbit in exactly one point. In other words, an orbit is $\{gx | g \in O_n\}$ and $\|x\| y_0$ is the intersection of $\mathbf{Y}$ and $\{gx | g \in O_n\}$. That $\mathbf{Y}$ is measurable and that $t(x) = \|x\| y_0$ is a measurable function is clear, so $\mathbf{Y}$ is a measurable cross section.

As an Example 4.1, $U \in O_n$ is a random orthogonal matrix. For a random variable $Y \in \mathbf{Y}$, write $Y = R y_0$ where R is a nonnegative random variable and independent of U. Then X is spherical iff

$$\mathscr{L}(X) = \mathscr{L}(U(Ry_0)) = \mathscr{L}(RUy_0)$$

for some $R \geq 0$ independent of U. Note that Uy_0 has a uniform distribution on $\{x | x \in R^n, \|x\| = 1\}$ so that X is spherical iff X is generated by first picking a radius R and then independently picking a point Uy_0 at random on the surface of the unit sphere. □

EXAMPLE 4.4. This example generalizes the previous one. The space $\mathbf{X}$ of this example consists of the set of all $n \times p$ real matrices of rank p so $p \leq n$. The group $G = O_n$ acts topologically on the left of $\mathbf{X}$ by matrix multiplication, $x \to gx$ for $x \in \mathbf{X}$ and $g \in O_n$. There are two rather natural choices for a cross section in this example. The first is the set

$$\mathbf{Y}_1 = \left\{ \begin{pmatrix} s \\ 0 \end{pmatrix} \in \mathbf{X} \middle| s \text{ is } p \times p \text{ and positive definite} \right\}.$$

This choice stems from the factorization

$$x = g\begin{pmatrix} s \\ 0 \end{pmatrix}, \qquad g \in O_n$$

for each $x \in \mathbf{X}$ where

$$s = (x'x)^{1/2}$$

is the unique positive definite matrix satisfying $s^2 = x'x$. [One version of this well known result is Proposition 5.5 in Eaton (1983).] The uniqueness of s in this factorization shows that $\mathbf{Y}_1$ intersects each orbit in exactly one point. The measurability is easily checked since $\mathbf{Y}_1$ is a relatively open subset of $\mathbf{X}$ and the map

$$t(x) = \begin{pmatrix} (x'x)^{1/2} \\ 0 \end{pmatrix} \in \mathbf{Y}_1$$

is continuous.

Now, consider $X \in \mathbf{X}$ which satisfies $\mathscr{L}(X) = \mathscr{L}(gX)$. Random matrices with this property are sometimes said to have left-orthogonally invariant distributions. Examples of such distributions on $\mathbf{X}$ are provided by taking X to have density (with respect to Lebesgue measure) on $\mathbf{X}$ of the form $q(x'x)$ where q is a nonnegative function defined on $p \times p$ positive definite matrices. For example

$$q(x'x) = \frac{|\Sigma|^{-n/2}}{(\sqrt{2\pi})^{np}} \exp\left[-\frac{1}{2} \operatorname{tr} x'x\Sigma^{-1} \right],$$

where Σ is $p \times p$ and positive definite, correspond to X with iid rows which are $N_p(0, \Sigma)$.

When $\mathscr{L}(X) = \mathscr{L}(gX)$, $g \in O_n$, Theorem 4.4 implies that

$$\mathscr{L}(X) = \mathscr{L}\left(U\begin{pmatrix} S \\ 0 \end{pmatrix} \right),$$

where U is uniform on O_n and is independent of S. The distribution of S is arbitrary over $p \times p$ positive definites.

The representation

$$x = g\begin{pmatrix} s \\ 0 \end{pmatrix}$$

can also be written

$$x = g\begin{pmatrix} I_p \\ 0 \end{pmatrix} s = \psi s,$$

where

$$\psi = g\begin{pmatrix} I_p \\ 0 \end{pmatrix}$$

is an element of $F_{p,n}$ as described in Example 2.3. Thus

$$\mathscr{L}(X) = \mathscr{L}\left(U\begin{pmatrix} S \\ 0 \end{pmatrix}\right) = \mathscr{L}\left(U\begin{pmatrix} I_p \\ 0 \end{pmatrix} S\right) = \mathscr{L}(\Delta S)$$

with Δ and S independent. That Δ has a uniform distribution on $F_{p,n}$ follows from the transitivity of the action of O_n on $F_{p,n}$.

Much the same analysis as above can be given based on the representation

$$x = g\begin{pmatrix} u \\ 0 \end{pmatrix},$$

where u is an element of the group of $p \times p$ upper triangular matrices with positive diagonal elements, say G_U^+. In this case, a cross section is taken to be

$$\mathbf{Y}_2 = \left\{\begin{pmatrix} u \\ 0 \end{pmatrix} \in \mathbf{X} \middle| u \in G_U^+\right\}.$$

That u is unique in this representation is well known [for example, see Proposition 5.2 in Eaton (1983)]. The remainder of the analysis and application of Theorem 4.4 is left to the reader. □

EXAMPLE 4.5. Let S_p be the real vector space of the $p \times p$ symmetric matrices. The group O_p acts on S_p via

$$s \to gsg'$$

for $s \in S_p$ and $g \in O_p$. To say that a random element $X \in S_p$ has an invariant distribution is to say that

$$\mathscr{L}(X) = \mathscr{L}(gXg'), \qquad g \in O_p.$$

Examples of such distributions on S_p include the Wishart distribution with identity scale matrix as well as certain versions of the multivariate beta and multivariate F distributions. For example, see Olkin and Rubin (1964). Define $\mathbf{Y} \subset S_p$ by

$$\mathbf{Y} = \{D \in S_p | D \text{ is diagonal}, d_{11} \geq \cdots \geq d_{pp}\},$$

where the diagonal elements of D are $d_{11}, \ldots, d_{pp}$. Clearly $\mathbf{Y}$ is a closed subset of S_p. The spectral theorem shows that every $s \in S_p$ can be written

$$s = gDg'$$

for some $g \in O_p$ and some $D \in \mathbf{Y}$. Of course, the diagonal elements of D are the eigenvalues of s and the function

$$t(s) = D$$

is continuous and hence measurable. That $\mathbf{Y}$ is a measurable cross section is now apparent.

Theorem 4.4 asserts that if

$$\mathscr{L}(X) = \mathscr{L}(gXg') \quad \text{for } g \in O_p, \text{ then} \quad \mathscr{L}(X) = \mathscr{L}(UYU'),$$

where U is uniform on O_p and is independent of $Y \in \mathbf{Y}$. The distribution of Y is arbitrary. In other words X is generated by choosing the ordered eigenvalues according to some arbitrary distribution and then randomly "moving" Y via the map

$$Y \to UYU'$$

with U uniform on O_p. □

EXAMPLE 4.6. This example deals with yet another matrix decomposition result involving singular values. The space $\mathbf{X}$ is the vector space $\mathscr{L}_{p,n}$ of $n \times p$ real matrices with $p \leq n$. The group in question is the product group $O_n \times O_p$ which acts on $\mathscr{L}_{p,n}$ via

$$x \to gxh'$$

for $g \in O_n$ and $h \in O_p$. The singular value decomposition for x is

$$x = g\begin{pmatrix} D \\ 0 \end{pmatrix}h$$

with $g \in O_n$ and $h \in O_p$. Here D is a $p \times p$ diagonal matrix with diagonal elements $d_{11} \geq \cdots \geq d_{pp} \geq 0$. These diagonal elements are the square roots of the eigenvalues of $x'x$. Thus, a candidate for a cross section is

$$\mathbf{Y} = \left\{ \begin{pmatrix} D \\ 0 \end{pmatrix} \in \mathscr{L}_{p,n} \middle| D \text{ is diagonal, } d_{11} \geq \cdots \geq d_{pp} \geq 0 \right\}.$$

Arguments similar to those given previously show that indeed $\mathbf{Y}$ is a measurable cross section.

Now, consider $X \in \mathscr{L}_{p,n}$ such that $\mathscr{L}(X) = \mathscr{L}(gXh')$ for $g \in O_n$ and $h \in O_p$. Examples of such distributions include the multivariate normal distribution on $\mathscr{L}_{p,n}$ (with mean 0 and identity covariance) and certain versions of the multivariate t distribution [for example, see Dickey (1967) or Eaton (1985)]. To describe the implications of Theorem 4.4, first note that the uniform distribution on $O_n \times O_p$ is just product Haar measure because $O_n \times O_p$ is a direct product. Thus, $U = (U_1, U_2)$ is uniform on $O_n \times O_p$ when U_1 is uniform on O_n, U_2 is uniform on O_p and U_1 and U_2 are independent. Therefore, when X has an invariant distribution on $\mathscr{L}_{p,n}$,

$$\mathscr{L}(X) = \mathscr{L}(U_1 Y U_2'),$$

where U_1, Y and U_2 are mutually independent and Y has an arbitrary distribution on $\mathbf{Y}$. □

4.3. Null robustness applications. The material in this section comes mainly from Das Gupta (1979) and from Eaton and Kariya (1984). The problem discussed here is motivated by the following example. Consider a random vector

$X \in R^n$ and set

$$T(X) = \frac{e'X}{\|X\|},$$

where e is the vector of 1's in R^n and $\|\cdot\|$ denotes the usual norm on R^n. Student's t statistic is a one-to-one function of $T(X)$ so the distribution of $T(X)$ determines the distribution of Student's t statistic and conversely. In fact, a bit of algebra shows that

$$t_{n-1} = \frac{(n-1)^{1/2}T}{n^{1/2}(1 - n^{-1}T^2)^{1/2}},$$

where t_{n-1} denotes the usual t statistic. When the coordinates of X are iid $N(0,1)$, then t_{n-1} has the Student t_{n-1} distribution and hence the distribution of T is fixed, say Q_0. Now, we ask: Under what conditions on $\mathscr{L}(X)$ does $\mathscr{L}(T)$ remain fixed at Q_0 (and hence t_{n-1} will still have Student's t_{n-1} distribution)? Fisher observed that if $\mathscr{L}(X)$ is O_n-invariant, as it is when the coordinates of X are iid $N(0,1)$, then $X/\|X\|$ is uniform on $\{x | x \in R^n, \ \|x\| = 1\}$. Thus, the distribution of $T(X)$ must remain the same when the coordinates of X are iid $N(0,1)$ as when $\mathscr{L}(X)$ is O_n-invariant. What makes this argument tick is:

(i) $T(X) = T(cX)$, $c > 0$.
(ii) $X/\|X\|$ is uniform when $\mathscr{L}(X)$ is O_n-invariant.

Condition (i) implies T is a function of $X/\|X\|$ while (ii) fixes the distribution of $X/\|X\|$.

There are a number of other examples where arguments similar to the one above can be used to show that distributions of statistics of interest can be derived using invariance rather than distributional assumptions. In such cases, the conditions under which the statistic has the given distribution can sometimes be substantially weakened. It is this general problem to which we now turn.

Here is one way to describe the problem. A random variable $X \in \mathbf{X}$ is given as is a statistic $T(X)$. A compact group K acts measurably on $\mathbf{X}$. Let $\mathscr{P}_K$ denote the set of all probability measures on $\mathbf{X}$ which are K invariant. In what follows, $\mathscr{L}(T(X)|P)$ denotes the distribution of $T(X)$ when $\mathscr{L}(X) = P$. The problem addressed below is the following:

Under what conditions is it the case that

$$\mathscr{L}(T(X)|P) = \mathscr{L}(T(X)|P') \quad \text{for all } P, P' \in \mathscr{P}_K? \tag{4.10}$$

Under some regularity, Das Gupta (1979) provided some sufficient conditions for (4.10) to hold. To describe these, first assume that

$$T(x_1) = T(x_2) \quad \text{implies} \quad T(kx_1) = T(kx_2) \quad \text{for all } k \in K. \tag{4.11}$$

Then, by Theorem 2.3, the group action of K can be moved to the range space of T, say $(\mathbf{Y}, \mathscr{B}_1)$. Assume that K acts measurably on $(\mathbf{Y}, \mathscr{B}_1)$.

THEOREM 4.5. *If K acts transitively on* $\mathbf{Y}$, *then* (4.10) *holds.*

PROOF. The transitivity of K on $(\mathbf{Y}, \mathscr{B}_1)$ implies there is exactly one K invariant probability measure on $(\mathbf{Y}, \mathscr{B}_1)$, say Q_0. But, when $\mathscr{L}(X) \in \mathscr{P}_K$, then $\mathscr{L}(X) = \mathscr{L}(kX)$ for all $k \in K$. Using the definition of the induced group action, we see that $\mathscr{L}(X) \in \mathscr{P}_K$ implies

$$\mathscr{L}(T(X)) = \mathscr{L}(T(kX)) = \mathscr{L}(kT(X)).$$

Thus, the induced distribution of T is K invariant so $\mathscr{L}(T(X)) = Q_0$. □

EXAMPLE 4.7. Suppose $X_1, \ldots, X_n$ is a random sample from a p-dimensional $N(\mu, \Sigma)$ with μ and Σ unknown. The usual Hotelling statistic for testing $\mu = 0$ is easily shown to be a function of

$$T(X) = X(X'X)^{-1}X',$$

where X: $n \times p$ has rows $X_1', \ldots, X_n'$. When $\mu = 0$,

$$\mathscr{L}(X) = N(0, I_n \otimes \Sigma),$$

so $\mathscr{L}(\gamma X) = \mathscr{L}(X)$ for $\gamma \in O_n$. The sample space $\mathbf{X}$ for this example is taken to be the set of $n \times p$ real matrices of rank p (a set of Lebesgue measure 0 has been removed). Thus, the range of T is the set $S_{n,p}$ of $n \times n$ rank p orthogonal projections.

With $K = O_n$, assumption (4.11) is valid as is the transitivity of the induced group action on $S_{n,p}$. The conclusion is that $\mathscr{L}(T)$ is equal to the unique invariant probability on $S_{n,p}$ which was introduced in Example 2.10. Hence the null distribution of Hotelling's statistic is the same when $\mathscr{L}(X) = N(0, I_n \otimes \Sigma)$ as when $\mathscr{L}(X) = \mathscr{L}(\gamma X)$ for $\gamma \in O_n$. □

In some cases of interest, assumption (4.11) does not hold but (4.10) is still valid. An alternative set of assumptions which yields (4.10) is given in Eaton and Kariya (1984). To describe these, assume that G is a topological group which acts measurably and transitively on $\mathbf{X}$, K is a compact subgroup of G and H is a subgroup of G such that

$$G = K \cdot H = \{kh | k \in K, h \in H\}.$$

Therefore, the subgroups K and H generate G.

THEOREM 4.6. *Assume that $T(X)$ is an H-invariant function. If either H or K is a normal subgroup of G, then* (4.10) *holds.*

PROOF. First assume that K is normal in G. To establish the theorem, it suffices to show that for any bounded measurable function f,

$$\int f(T(x))P(dx) = \int f(T(x))P'(dx) \quad \text{for } P, P' \in \mathscr{P}_K. \tag{4.12}$$

Because $kP = P$, we have

$$\int f(T(x))P(dx) = \int f(T(kx))P(dx).$$

Integrating both sides of this equality with respect to the invariant probability measure on K yields

$$\int f(T(x))P(dx) = \int\int f(T(kx))\nu(dk)P(dx).$$

Fix $x_0 \in \mathbf{X}$ and use the transitivity of $G = K \cdot H$ to write $x = k_1 h x_0$. Using the invariance of T under H and the invariance of ν, we have

$$\int f(T(kx))\nu(dx) = \int f(T(kk_1hx_0))\nu(dk) = \int f\big(T\big(h^{-1}khx_0\big)\big)\nu(dk).$$

Since the map $k \to h^{-1}kh$ is a continuous isomorphism of K, the uniqueness of ν implies that ν is invariant under this map. Therefore

$$\int f(T(kx))\nu(dk) = \int f(T(kx_0))\nu(dk),$$

which yields the equation

$$\int f(T(x))P(dx) = \int f(T(kx_0))\nu(dk).$$

Because this equation holds for each $P \in \mathscr{P}_K$, obviously (4.12) holds. Thus the theorem is proved when K is normal in G. The proof when H is normal in G is similar and the details are left to the reader. □

EXAMPLE 4.8. In this example where canonical correlations are discussed, Theorem 4.6 is applicable but Theorem 4.5 is not. Consider a random matrix Z: $n \times p$ which has rank p and partition Z as $Z = (Z_1 Z_2)$, where Z_i is $n \times p_i$, $i = 1, 2$. Without essential loss of generality, the mean 0 case is treated here. The random rank p_i orthogonal projection

$$Q_i = Z_i(Z_i'Z_i)^{-1}Z_i', \qquad i = 1, 2,$$

takes values in the space $S_{n,\,p_i}$ of the last example. The *squared canonical correlations* are defined to be the $r = \min\{p_1, p_2\}$ largest eigenvalues of Q_1Q_2. That this definition agrees with more traditional definitions is easily checked.

Given Z, let $T(Z)$ be the vector of the r largest eigenvalues (arranged in order) of Q_1Q_2. When Z is $N(0, I_n \otimes I_p)$, the density of $T(Z)$ is given in Anderson [(1958), Chapter 13]. To describe a large class of distributions of Z for which the distribution of $T(Z)$ is that when Z is $N(0, I_n \otimes I_p)$, consider the group G whose elements are (γ, ψ, A, B) with $\gamma, \psi \in O_n$, $A \in \mathrm{Gl}_{p_1}$ and $B \in \mathrm{Gl}_{p_2}$. The action of G on (z_1, z_2) is

$$(z_1, z_2) \to (\gamma z_1 A', \psi z_2 B')$$

and the group operation is

$$(\gamma_1, \psi_1, A_1, B_1)(\gamma_2, \psi_2, A_2, B_2) = (\gamma_1\gamma_2, \psi_1\psi_2, A_1A_2, B_1B_2).$$

That G is transitive on the set $\mathbf{X}$ of $n \times p$ matrices of rank p is easily checked.

To apply Theorem 4.6, let

$$H = \{\gamma, \psi, A, B)|\gamma = \psi\}$$

and

$$K = \{(\gamma, \psi, A, B) | \psi = I_n,\ A = I_{p_1},\ B = I_{p_2}\}.$$

Then $G = K \cdot H$, K is compact and K is normal in G. Since $T(Z)$ is H-invariant, (4.10) holds. In other words,

$$\mathscr{L}((Z_1, Z_2)) = \mathscr{L}((\gamma Z_1, Z_2))$$

implies that the distribution of $T(Z)$ is the same as when $\mathscr{L}(Z) = N(0, I_n \otimes I_p)$. □

Other examples and references can be found in Das Gupta (1979) and Eaton and Kariya (1984).