CHAPTER 7

First Applications of Invariance

We now begin to reap some of the benefits of the labor of Chapter 6. The one unifying notion throughout this chapter is that of a group of transformations acting on a space. Within this framework independence and distributional properties of random vectors are discussed and a variety of structural problems are considered. In particular, invariant probability models are introduced and the invariance of likelihood ratio tests and maximum likelihood estimators is established. Further, maximal invariant statistics are discussed in detail.

7.1. LEFT $\mathcal{O}_n$ INVARIANT DISTRIBUTIONS ON $n \times p$ MATRICES

The main concern of this section is conditions under which the two matrices Ψ and U in the decomposition $X = \Psi U$ (see Example 6.20) are stochastically independent when X is a random $n \times p$ matrix. Before discussing this problem, a useful construction of the uniform distribution on $\mathcal{F}_{p,n}$ is presented. Throughout this section, $\mathcal{X}$ denotes the space of $n \times p$ matrices of rank p so $n \geqslant p$. First, a technical result.

Proposition 7.1. Let $X \in \mathcal{L}_{p,n}$ have a normal distribution with mean zero and $\mathrm{Cov}(X) = I_n \otimes I_p$. Then $P\{X \in \mathcal{X}\} = 1$ and the complement of $\mathcal{X}$ in $\mathcal{L}_{p,n}$ has Lebesgue measure zero.

Proof. Let $X_1, \ldots, X_p$ denote the p columns of X. Thus $X_1, \ldots, X_p$ are independent random vectors in R^n and $\mathcal{L}(X_i) = N(0, I_n)$, $i = 1, \ldots, p$. It is shown that $P\{X \in \mathcal{X}^c\} = 0$. To say that $X \in \mathcal{X}^c$ is to say that, for some index i,

$$X_i \in \operatorname{span}\{X_j | j \neq i\}.$$

Therefore,

$$P\{X \in \mathfrak{X}^c\} = P\left\{\bigcup_{i=1}^{p}\left[X_i \in \operatorname{span}\{X_j | j \neq i\}\right]\right\}$$

$$\leqslant \sum_1^p P\{X_i \in \operatorname{span}\{X_j | j \neq i\}\}.$$

However, X_i is independent of the set of random vectors $\{X_j | j \neq i\}$ and the probability of any subspace M of dimension less than n is zero. Since $p \leqslant n$, the subspace span $\{X_j | j \neq i\}$ has dimension less than n. Thus conditioning on X_j for $j \neq i$, we have

$$P\{X_i \in \operatorname{span}\{X_j | j \neq i\}\} = \mathcal{E}P\{X_i \in \operatorname{span}\{X_j | j \neq i\} | X_j,\ j \neq i\} = 0.$$

Hence $P\{X \in \mathfrak{X}^c\} = 0$. Since $\mathfrak{X}^c$ has probability zero under the normal distribution on $\mathcal{L}_{p,n}$ and since the normal density function with respect to Lebesgue measure is strictly positive on $\mathcal{L}_{p,n}$, it follows the $\mathfrak{X}^c$ has Lebesgue measure zero. □

If $X \in \mathcal{L}_{p,n}$ is a random vector that has a density with respect to Lebesgue measure, the previous result shows that $P\{X \in \mathfrak{X}\} = 1$ since $\mathfrak{X}^c$ has Lebesgue measure zero. In particular, if $X \in \mathcal{L}_{p,n}$ has a normal distribution with a nonsingular covariance, then $P\{X \in \mathfrak{X}\} = 1$, and we often restrict such normal distributions to $\mathfrak{X}$ in order to insure that X has rank p. For many of the results below, it is assumed that X is a random vector in $\mathfrak{X}$, and in applications X is a random vector in $\mathcal{L}_{p,n}$, which has been restricted to $\mathfrak{X}$ after it has been verified that $\mathfrak{X}^c$ has probability zero under the distribution of X.

Proposition 7.2. Suppose $X \in \mathfrak{X}$ has a normal distribution with $\mathcal{L}(X) = N(0, I_n \otimes I_p)$. Let $X_1, \ldots, X_p$ be the columns of X and let $\Psi \in \mathfrak{F}_{p,n}$ be the random matrix whose p columns are obtained by applying the Gram–Schmidt orthogonalization procedure to $X_1, \ldots, X_p$. Then Ψ has the uniform distribution on $\mathfrak{F}_{p,n}$, that is, the distribution of Ψ is the unique probability measure on $\mathfrak{F}_{p,n}$ that is invariant under the action of $\mathcal{O}_n$ on $\mathfrak{F}_{p,n}$ (see Example 6.16).

Proof. Let Q be the probability distribution of Ψ of $\mathcal{F}_{p,n}$. It must be verified that

$$Q(\Gamma B) = Q(B), \qquad \Gamma \in \mathcal{O}_n$$

for all Borel sets B of $\mathcal{F}_{p,n}$. If $\Gamma \in \mathcal{O}_n$, it is clear that $\mathcal{L}(\Gamma X) = \mathcal{L}(X)$. Also, it is not difficult to verify that Ψ, which we now write as a function of X, say $\Psi(X)$, satisfies

$$\Psi(\Gamma X) = \Gamma\Psi(X), \qquad \Gamma \in \mathcal{O}_n.$$

This follows by looking at the Gram–Schmidt Procedure, which defined the columns of Ψ. Thus

$$Q(B) = P\{\Psi(X) \in B\} = P\{\Psi(\Gamma X) \in B\} = P\{\Gamma\Psi(X) \in B\}$$

$$= P\{\Psi(X) \in \Gamma' B\} = Q(\Gamma' B)$$

for all $\Gamma \in \mathcal{O}_n$. The second equality above follows from the observation that $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$. Hence Q is an $\mathcal{O}_n$-invariant probability measure on $\mathcal{F}_{p,n}$ and the uniqueness of such a measure shows that Q is what was called the uniform distribution on $\mathcal{F}_{p,n}$. □

Now, consider the two spaces $\mathcal{X}$ and $\mathcal{F}_{p,n} \times G_U^+$. Let ϕ be the function on $\mathcal{X}$ to $\mathcal{F}_{p,n} \times G_U^+$ that maps X into the unique pair (Ψ, U) such that $X = \Psi U$. Obviously, $\phi^{-1}(\Psi, U) = \Psi U \in \mathcal{X}$.

Definition 7.1. If $X \in \mathcal{X}$ is a random vector with a distribution P, then P is *left invariant under* $\mathcal{O}_n$ if $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ for all $\Gamma \in \mathcal{O}_n$.

The remainder of this section is devoted to a characterization of the $\mathcal{O}_n$-left invariant distributions on $\mathcal{X}$. It is shown that, if $X \in \mathcal{X}$ has an $\mathcal{O}_n$-left invariant distribution, then for $\phi(X) = (\Psi, U) \in \mathcal{F}_{p,n} \times G_U^+$, Ψ and U are stochastically independent and Ψ has a uniform distribution on $\mathcal{F}_{p,n}$. This assertion and its converse are given in the following proposition.

Proposition 7.3. Suppose $X \in \mathcal{X}$ is a random vector with an $\mathcal{O}_n$-left invariant distribution P and write $(\Psi, U) = \phi(X)$. Then Ψ and U are stochastically independent and Ψ has a uniform distribution on $\mathcal{F}_{p,n}$. Conversely, if $\Psi \in \mathcal{F}_{p,n}$ and $U \in G_U^+$ are independent and if Ψ has a uniform distribution on $\mathcal{F}_{p,n}$, then $X = \Psi U$ has an $\mathcal{O}_n$-left invariant distribution on $\mathcal{X}$.

Proof. The joint distribution Q of (Ψ, U) is determined by

$$Q(B_1 \times B_2) = P(\phi^{-1}(B_1 \times B_2))$$

where B_1 is a Borel subset of $\mathcal{F}_{p,n}$ and B_2 is a Borel subset of G_U^+. Also,

$$\iint f(\Psi, U) Q(d\Psi, dU) = \int f(\phi(X)) P(dX)$$

for any Borel measurable function that is integrable. The group $\mathcal{O}_n$ acts on the left of $\mathcal{F}_{p,n} \times G_U^+$ by

$$\Gamma(\Psi, U) = (\Gamma\Psi, U)$$

and it is clear that

$$\phi(\Gamma X) = \Gamma\phi(X) \quad \text{for } X \in \mathcal{X}, \Gamma \in \mathcal{O}_n.$$

We now show that Q is invariant under this group action and apply Proposition 6.10. For $\Gamma \in \mathcal{O}_n$,

$$\begin{aligned} \iint f(\Gamma(\Psi, U)) Q(d\Psi, dU) &= \int f(\Gamma\phi(X)) P(dX) = \int f(\phi(\Gamma X)) P(dX) \\ &= \int f(\phi(X)) P(dX) \\ &= \iint f(\Psi, U) Q(d\Psi, dU). \end{aligned}$$

Therefore, Q is $\mathcal{O}_n$-invariant and, by Proposition 6.10, Q is a product measure $Q_1 \times Q_2$ where Q_1 is taken to be the uniform distribution on $\mathcal{F}_{p,n}$. That Q_2 is a probability measure is clear since Q is a probability measure. The first assertion has been established. For the converse, let Q_1 and Q_2 be the distributions of Ψ and U so Q_1 is the uniform distribution on $\mathcal{F}_{p,n}$ and $Q_1 \times Q_2$ is the joint distribution of (Ψ, U) in $\mathcal{F}_{p,n} \times G_U^+$. The distribution P of $X = \Psi U = \phi^{-1}(\Psi, U)$ is determined by the equation

$$\int f(X) P(dX) = \iint f(\phi^{-1}(\Psi, U)) Q_1(d\Psi) Q_2(dU),$$

for all integrable f. To show P is $\mathcal{O}_n$-left invariant, it must be verified that

$$\int f(\Gamma X) P(dX) = \int f(X) P(dX)$$

for all integrable f and $\Gamma \in \mathcal{O}_n$. But

$$\int f(\Gamma X)P(dX) = \iint f(\Gamma\phi^{-1}(\Psi, U))Q_1(d\Psi)Q_2(dU)$$

$$= \iint f(\phi^{-1}(\Gamma\Psi, U))Q_1(d\Psi)Q_2(dU)$$

$$= \iint f(\phi^{-1}(\Psi, U))Q_1(d\Psi)Q_2(dU) = \int f(X)P(dX)$$

where the next to the last equality follows from the $\mathcal{O}_n$-invariance of Q_1. Thus P is $\mathcal{O}_n$-left invariant. □

When $p = 1$, Proposition 7.3 is interesting. In this case $\mathcal{X} = R^n - \{0\}$ and the $\mathcal{O}_n$-left invariant distributions on $\mathcal{X}$ are exactly the orthogonally invariant distributions on R^n that have no probability at $0 \in R^n$. If $X \in R^n - \{0\}$ has an orthogonally invariant distribution, then $\Psi = X/\|X\| \in \mathcal{F}_{1,n}$ is independent of $U = \|X\|$ and Ψ has a uniform distribution on $\mathcal{F}_{1,n}$.

There is an analogue of Proposition 7.3 for the decomposition of $X \in \mathcal{X}$ into (Ψ, A) where $\Psi \in \mathcal{F}_{p,n}$ and $A \in \mathcal{S}_p^+$ (see Proposition 5.5).

Proposition 7.4. Suppose $X \in \mathcal{X}$ is a random vector with an $\mathcal{O}_n$-left invariant distribution and write $\phi(X) = (\Psi, A)$ where $\Psi \in \mathcal{F}_{p,n}$ and $A \in \mathcal{S}_p^+$ are the unique matrices such that $X = \Psi A$. Then Ψ and A are independent and Ψ has a uniform distribution on $\mathcal{F}_{p,n}$. Conversely, if $\Psi \in \mathcal{F}_{p,n}$ and $A \in \mathcal{S}_p^+$ are independent and if Ψ has a uniform distribution on $\mathcal{F}_{p,n}$, then $X = \Psi A$ has an $\mathcal{O}_n$-left invariant distribution on $\mathcal{X}$.

Proof. The proof is essentially the same as that of Proposition 7.3 and is left to the reader. □

Thus far, it has been shown that if $X \in \mathcal{X}$ has an $\mathcal{O}_n$-left invariant distribution for $X = \Psi U$, Ψ and U are independent and Ψ has a uniform distribution. However, nothing has been said about the distribution of $U \in G_U^+$. The next result gives the density function of U with respect to the right invariant measure

$$\nu_r(dU) = \frac{dU}{\prod_1^p u_{ii}^i}$$

in the case that X has a density of a special form.

Proposition 7.5. Suppose $X \in \mathfrak{X}$ has a distribution P given by a density function

$$f_0(X'X), \qquad X \in \mathfrak{X}$$

with respect to the measure

$$\mu(dX) = \frac{dX}{|X'X|^{n/2}}$$

on $\mathfrak{X}$. Then the density function of U (with respect to ν_r) in the representation $X = \Psi U$ is

$$g_0(U) = 2^p(\sqrt{2\pi})^{np}\omega(n, p) f_0(U'U).$$

Proof. If $X \in \mathfrak{X}$, $U(X)$ denotes the unique element of G_U^+ such that $X = \Psi U(X)$ for some $\Psi \in \mathfrak{F}_{p,n}$. To show g_0 is the density function of U, it is sufficient to verify that

$$\int h(U(X)) f_0(X'X)\mu(dX) = \int h(U) g_0(U)\nu_r(dU)$$

for all integrable functions h. Since $X'X = U'(X)U(X)$, the results of Example 6.20 show that

$$\int h(U(X)) f_0(U'(X)U(X))\mu(dX) = c\int h(U) f_0(U'U)\nu_r(dU)$$

where $c = 2^p(\sqrt{2\pi})^{np}\omega(n, p)$. Since $g_0(U) = cf_0(U'U)$, g_0 is the density of U. □

A similar argument gives the density of $S = X'X$.

Proposition 7.6. Suppose $X \in \mathfrak{X}$ has distribution P given by a density function

$$f_0(X'X), \qquad X \in \mathfrak{X}$$

with respect to the measure μ. Then the density of $S = X'X$ is

$$g_0(S) = (\sqrt{2\pi})^{np}\omega(n, p) f_0(S)$$

with respect to the measure

$$\nu(dS) = \frac{dS}{|S|^{(p+1)/2}}.$$

Proof. With the notation $S(X) = X'X$, it is sufficient to verify that

$$\int h(S(X)) f_0(X'X)\mu(dX) = \int h(S) g_0(S)\nu(dS)$$

for all integrable functions h. Combining the identities (6.4) and (6.5), we have

$$\int h(S(X)) f_0(X'X)\mu(dX) = c\int h(S) f_0(S)\nu(dS)$$

where $c = (\sqrt{2\pi})^{np}\omega(n, p)$. Since $g_0 = cf_0$, the proof is complete. □

When $X \in \mathfrak{X}$ has the density assumed in Propositions 7.5 and 7.6, it is clear that the distribution of X is $\mathcal{O}_n$-left invariant. In this case, for $X = \Psi U$, Ψ and U are independent, Ψ has a uniform distribution on $\mathfrak{F}_{p,n}$, and U has the density given in Proposition 7.5. Thus the joint distribution of Ψ and U has been completely described. Similar remarks apply to the situation treated in Proposition 7.6. The reader has probably noticed that the distribution of $S = X'X$ was derived rather than the distribution of A in the representation $X = \Psi A$ for $\Psi \in \mathfrak{F}_{p,n}$ and $A \in \mathfrak{S}_p^+$. Of course, $S = A^2$ so A is the unique positive definite square root of S. The reason for giving the distribution of S rather than that of A is quite simple—the distribution of A is substantially more complicated than that of S and harder to derive.

In the following example, we derive the distributions of U and S when $X \in \mathfrak{X}$ has a nonsingular $\mathcal{O}_n$-left invariant normal distribution.

◆ **Example 7.1.** Suppose $X \in \mathfrak{X}$ has a normal distribution with a nonsingular covariance and also assume that $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ for all $\Gamma \in \mathcal{O}_n$. Thus $\mathcal{E}X = \Gamma\mathcal{E}X$ for all $\Gamma \in \mathcal{O}_n$, which implies that $\mathcal{E}X = 0$. Also, $\operatorname{Cov}(X)$ must satisfy $\operatorname{Cov}((\Gamma \otimes I_p)X) = \operatorname{Cov}(X)$ since $\mathcal{L}(X) = \mathcal{L}((\Gamma \otimes I_p)X)$. From Proposition 2.19, this implies that

$$\operatorname{Cov}(X) = I_n \otimes \Sigma$$

for some positive definite Σ as $\operatorname{Cov}(X)$ is assumed to be nonsingular. In summary, if X has a normal distribution in $\mathfrak{X}$ that is $\mathcal{O}_n$-left invariant, then

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

Conversely, if X is normal with mean zero and $\mathrm{Cov}(X) = I_n \otimes \Sigma$, then $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ for all $\Gamma \in \mathcal{O}_n$. Now that the $\mathcal{O}_n$-left invariant normal distributions on $\mathfrak{X}$ have been described, we turn to the distribution of $S = X'X$ and U described in Propositions 7.5 and 7.6. When $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$, the density function of X with respect to the measure $\mu(dX) = dX/|X'X|^{n/2}$ is

$$f_0(X) = (\sqrt{2\pi})^{-np}|\Sigma^{-1}X'X|^{n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}X'X\right].$$

Therefore, the density of S with respect to the measure

$$\nu(dS) = \frac{dS}{|S|^{(p+1)/2}}, \qquad S \in \mathcal{S}_p^+$$

is given by

$$g_0(S) = \omega(n, p)|\Sigma^{-1}S|^{n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S\right]$$

according to Proposition 7.6. This density is called the *Wishart density* with parameters Σ, p, and n. Here, p is the dimension of S and n is called the degrees of freedom. When S has such a density function, we write $\mathcal{L}(S) = W(\Sigma, p, n)$, which is read "the distribution of S is Wishart with parameters Σ, p, and n." A slightly more general definition of the Wishart distribution is given in the next chapter, where a thorough discussion of the Wishart distribution is presented. A direct application of Proposition 7.5 yields the density

$$g_1(U) = 2^p\omega(n, p)|\Sigma^{-1}U'U|^{n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}U'U\right]$$

with respect to measure

$$\nu_r(dU) = \frac{dU}{\prod_1^p u_{ii}^i}$$

when $X = \Psi U$, $\Psi \in \mathcal{F}_{p,n}$, and $U \in G_U^+$. Here, the nonzero elements of U are u_{ij}, $1 \leqslant i \leqslant j \leqslant p$. When $\Sigma = I_p$, g_1 becomes

$$\begin{aligned} g_1(U)\nu_r(dU) &= 2^p\omega(n, p)\prod_{i=1}^p u_{ii}^i\exp\left[-\tfrac{1}{2}\operatorname{tr}U'U\right]\nu_r(dU) \\ &= 2^p\omega(n, p)\prod_{i=1}^p u_{ii}^{n-i}\exp\left[-\tfrac{1}{2}\sum_{1\leqslant j} u_{ij}^2\right]dU. \end{aligned}$$

In G_U^+, the diagonal elements of U range between 0 and ∞ and the elements above the diagonal range between $-\infty$ and $+\infty$. Writing the density above as

$$g_1(U)\nu_r(dU) = 2^p\omega(n, p)\prod_1^p \left(u_{ii}^{n-i}\exp\left[-\tfrac{1}{2}u_{ii}^2\right] du_{ii}\right)$$

$$\times \prod_{i<j}\left(\exp\left[-\tfrac{1}{2}u_{ij}^2\right] du_{ij}\right),$$

we see that this density factors into a product of functions that are, when normalized by a constant, density functions. It is clear by inspection that

$$\mathcal{L}(u_{ij}) = N(0,1) \quad \text{for } i < j.$$

Further, a simple change of variable shows that

$$\mathcal{L}\left(u_{ii}^2\right) = \chi^2_{n-i+1}, \qquad i = 1,\dots, p.$$

Thus when $\Sigma = I_p$, the nonzero elements of U are independent, the elements above the diagonal are all $N(0,1)$, and the square of the ith diagonal element has a chi-square distribution with $n - i + 1$ degrees of freedom. This result is sometimes useful for deriving the distribution of functions of $S = U'U$. ♦

7.2. GROUPS ACTING ON SETS

Suppose $\mathcal{X}$ is a set and G is a group that acts on the left of $\mathcal{X}$ according to Definition 6.3. The group G defines a natural equivalence relation between elements of $\mathcal{X}$—namely, write $x_1 \simeq x_2$ if there exists a $g \in G$ such that $x_1 = gx_2$. It is easy to check that $\simeq$ is in fact an equivalence relation. Thus the group G partitions the set $\mathcal{X}$ into disjoint equivalence classes, say

$$\mathcal{X} = \bigcup_{\alpha\in A} \mathcal{X}_\alpha,$$

where A is an index set and the equivalence classes $\mathcal{X}_\alpha$ are disjoint. For each $x \in \mathcal{X}$, the set $\{gx|g \in G\}$ is the *orbit* of x under the action of G. From the definition of the equivalence relation, it is clear that, if $x \in \mathcal{X}_\alpha$, then $\mathcal{X}_\alpha$ is just the orbit of x. Thus the decomposition of $\mathcal{X}$ into equivalence classes is simply a decomposition of $\mathcal{X}$ into disjoint orbits and two points are equivalent iff they are in the same orbit.

Definition 7.2. Suppose G acts on the left of $\mathcal{X}$. A function f on $\mathcal{X}$ to $\mathcal{Y}$ is *invariant* if $f(x) = f(gx)$ for all $x \in \mathcal{X}$ and $g \in G$. The function f is *maximal invariant* if f is invariant and $f(x_1) = f(x_2)$ implies that $x_1 = gx_2$ for some $g \in G$.

Obviously, f is invariant iff f is constant on each orbit in X. Also, f is maximal invariant iff it is constant on each orbit and takes different values on different orbits.

Proposition 7.7. Suppose f maps $\mathcal{X}$ onto $\mathcal{Y}$ and f is maximal invariant. Then h, mapping $\mathcal{X}$ into $\mathcal{Z}$, is invariant iff $h(x) = k(f(x))$ for some function k mapping $\mathcal{Y}$ into $\mathcal{Z}$.

Proof. If $h(x) = k(f(x))$, then h is invariant as f is invariant. Conversely, suppose h is invariant. Given $y \in \mathcal{Y}$, the set $\{x | f(x) = y\}$ is exactly one orbit in $\mathcal{X}$ since f is maximal invariant. Let $z \in \mathcal{Z}$ be the value of h on this orbit (h is invariant), and define $k(y) = z$. Obviously, k is well defined and $k(f(x)) = h(x)$. □

Proposition 7.7 is ordinarily paraphrased by saying that a function is invariant iff it is a function of a maximal invariant. Once a maximal invariant function has been constructed, then all the invariant functions are known—namely, they are functions of the maximal invariant function. If the group G acts transitively on $\mathcal{X}$, then there is just one orbit and the only invariant functions are the constants. We now turn to some examples.

◆ **Example 7.2.** Let $\mathcal{X} = R^n - \{0\}$ and let $G = \mathcal{O}_n$ act on $\mathcal{X}$ as a group of matrices acts on a vector space. Given $x \in \mathcal{X}$, it is clear that the orbit of x is $\{y | \|y\| = \|x\|\}$. Let $S_r = \{x | \|x\| = r\}$, so

$$\mathcal{X} = \bigcup_{r>0} S_r$$

is the decomposition of $\mathcal{X}$ into equivalence classes. The real number $r > 0$ indexes the orbits. That $f(x) = \|x\|$ is a maximal invariant function follows from the invariance of f and the fact that f takes a different value on each orbit. Thus a function is invariant under the action of G on $\mathcal{X}$ iff it is a function of $\|x\|$. Now, consider the space $S_1 \times (0, \infty)$ and define the function ϕ on $\mathcal{X}$ to $S_1 \times (0, \infty)$ by

$\phi(x) = (x/\|x\|, \|x\|)$. Obviously, ϕ is one-to-one, onto, and $\phi^{-1}(u, r) = ru$ for $(u, r) \in S_1 \times (0, \infty)$. Further, the group action on $\mathfrak{X}$ corresponds to the group action on $S_1 \times (0, \infty)$ given by

$$\Gamma(u, r) \equiv (\Gamma u, r), \qquad \Gamma \in \mathcal{O}_n.$$

In other words, $\phi(\Gamma x) = \Gamma\phi(x)$ so ϕ is an equivariant function (see Definition 6.14). Since $\mathcal{O}_n$ acts transitively on S_1, a function h on $S_1 \times (0, \infty)$ is invariant iff $h(u, r)$ does not depend on u. For this example, the space $\mathfrak{X}$ has been mapped onto $S_1 \times (0, \infty)$ by ϕ so that the group action on $\mathfrak{X}$ corresponds to a special group action on $S_1 \times (0, \infty)$—namely, $\mathcal{O}_n$ acts transitively on S_1 and is the identity on $(0, \infty)$. The whole point of introducing $S_1 \times (0, \infty)$ is that the function $h_0(u, r) = r$ is obviously a maximal invariant function due to the special way in which $\mathcal{O}_n$ acts on $S_1 \times (0, \infty)$. To say it another way, the orbits in $S_1 \times (0, \infty)$ are $S_1 \times \{r\}$, $r > 0$, so the product space structure provides a convenient way to index the orbits and hence to give a maximal invariant function. This type of product space structure occurs in many other examples. ◆

The following example provides a useful generalization of the example above.

◆ **Example 7.3.** Suppose $\mathfrak{X}$ is the space of all $n \times p$ matrices of rank $p, p \leqslant n$. Then $\mathcal{O}_n$ acts on the left of $\mathfrak{X}$ by matrix multiplication. The first claim is that $f_0(X) = X'X$ is a maximal invariant function. That f_0 is invariant is clear, so assume that $f_0(X_1) = f_0(X_2)$. Thus $X_1'X_1 = X_2'X_2$ and, by Proposition 1.31, there exists a $\Gamma \in \mathcal{O}_n$ such that $\Gamma X_1 = X_2$. This proves that f_0 is a maximal invariant. Now, the question is: where did f_0 come from? To answer this question, recall that each $X \in \mathfrak{X}$ has a unique representation as $X = \Psi A$ where $\Psi \in \mathfrak{F}_{p,n}$ and $A \in \mathbb{S}_p^+$. Let ϕ denote the map that sends X into the pair $(\Psi, A) \in \mathfrak{F}_{p,n} \times \mathbb{S}_p^+$ such that $X = \Psi A$. The group $\mathcal{O}_n$ acts on $\mathfrak{F}_{p,n} \times \mathbb{S}_p^+$ by

$$\Gamma(\Psi, A) = (\Gamma\Psi, A)$$

and ϕ satisfies

$$\phi(\Gamma X) = \Gamma\phi(X).$$

It is clear that $h_0(\Psi A) \equiv A$ is a maximal invariant function on $\mathfrak{F}_{p,n} \times \mathbb{S}_p^+$ under the action of $\mathcal{O}_n$ since $\mathcal{O}_n$ acts transitively on $\mathfrak{F}_{p,n}$.

Also, the orbits in $\mathcal{F}_{p,n} \times \mathcal{S}_p^+$ are $\mathcal{F}_{p,n} \times \{A\}$ for $A \in \mathcal{S}_p^+$. It follows immediately from the equivariance of ϕ that

$$\phi^{-1}\left(\mathcal{F}_{p,n} \times \{A\}\right) = \left\{X | X = \Psi A \text{ for some } \Psi \in \mathcal{F}_{p,n}\right\}$$

are the orbits in $\mathcal{X}$ under the action of $\mathcal{O}_n$. Thus we have a convenient indexing of the orbits in $\mathcal{X}$ given by A. A maximal invariant function on $\mathcal{X}$ must be a one-to-one function of an orbit index—namely, $A \in \mathcal{S}_p^+$. However, $f_0(X) = X'X = A^2$ when

$$X \in \left\{X | X = \Psi A, \quad \text{for some } \Psi \in \mathcal{F}_{p,n}\right\}.$$

Since A is the unique positive definite square root of $A^2 = X'X$, we have explicitly shown why f_0 is a one-to-one function of the orbit index A. A similar orbit indexing in $\mathcal{X}$ can be given by elements $U \in G_U^+$ by representing each $X \in \mathcal{X}$ as $X = \Psi U$, $\Psi \in \mathcal{F}_{p,n}$, and $U \in G_U^+$. The details of this are left to the reader. ◆

◆ **Example 7.4.** In this example, the set $\mathcal{X}$ is $(R^p - \{0\}) \times \mathcal{S}_p^+$. The group Gl_p acts on the left of $\mathcal{X}$ in the following manner:

$$A(y, S) \equiv (Ay, ASA')$$

for $(y, S) \in \mathcal{X}$ and $A \in Gl_p$. A useful method for finding a maximal invariant function is to consider a point $(y, S) \in \mathcal{X}$ and then "reduce" (y, S) to a convenient representative in the orbit of (y, S). The orbit of (y, S) is $\{A(y, S) | A \in Gl_p\}$. To reduce a given point (y, S) by $A \in Gl_p$, first choose $A = \Gamma S^{-1/2}$ where $\Gamma \in \mathcal{O}_p$ and $S^{-1/2}$ is the inverse of the positive definite square root of S. Then

$$ASA' = \Gamma S^{-1/2} S S^{-1/2} \Gamma' = \Gamma\Gamma' = I_p$$

and

$$A(y, S) = \left(\Gamma S^{-1/2} y, I\right),$$

which is in the orbit of (y, S). Since $S^{-1/2}y$ and $\|S^{-1/2}y\|\varepsilon_1$ have the same length ($\varepsilon_1' = (1, 0, \ldots, 0)$), we can choose $\Gamma \in \mathcal{O}_p$ such that

$$\Gamma S^{-1/2} y = \|S^{-1/2} y\| \varepsilon_1.$$

Therefore, for each $(y, S) \in \mathcal{X}$, the point

$$\left(\|S^{-1/2} y\| \varepsilon_1, I_p\right)$$

is in the orbit of (y, S). Let

$$f_0(y, S) = y'S^{-1}y = \|S^{-1/2}y\|^2.$$

The above reduction argument suggests, but does not prove, that f_0 is maximal invariant. However, the reduction argument does provide a method for checking that f_0 is maximal invariant. First, f_0 is invariant. To show f_0 is maximal invariant, if $f_0(y_1, S_1) = f_0(y_2, S_2)$, we must show there exists an $A \in Gl_p$ such that $A(y_1, S_1) = (y_2, S_2)$. From the reduction argument, there exists $A_i \in Gl_p$ such that

$$A_i(y_i, S_i) = \left(\|S_i^{-1/2}y_i\|\varepsilon_1, I_p\right), \qquad i = 1, 2.$$

Since $f_0(y_1, S_1) = f_0(y_2, S_2)$,

$$\|S_1^{-1/2}y_1\| = \|S_2^{-1/2}y_2\|$$

and this shows that

$$A_1(y_1, S_1) = A_2(y_2, S_2).$$

Setting $A = A_2^{-1}A_1$, we see that $A(y_1, S_1) = (y_2, S_2)$ so f_0 is maximal invariant. As in the previous two examples, it is possible to represent $\mathcal{X}$ as a product space where a maximal invariant is obvious. Let

$$\mathcal{Y} = \left\{(u, S) | u \in R^p, S \in \mathbb{S}_p^+, u'S^{-1}u = 1\right\}.$$

Then Gl_p acts on the left of $\mathcal{Y}$ by

$$A(u, S) \equiv (Au, ASA').$$

The reduction argument used above shows that the action of Gl_p is transitive on $\mathcal{Y}$. Consider the map ϕ from $\mathcal{X}$ to $\mathcal{Y} \times (0, \infty)$ given by

$$\phi(x, S) = \left(\left(\frac{x}{(x'S^{-1}x)^{1/2}}, S\right), x'S^{-1}x\right).$$

The group action of Gl_p on $\mathcal{Y} \times (0, \infty)$ is

$$A((u, S), r) \equiv (A(u, S), r)$$

and a maximal invariant function is

$$f_1((u, S), r) = r$$

since Gl_p is transitive on $\mathcal{Y}$. Clearly, ϕ is a one-to-one onto function and satisfies

$$\phi(A(x, S)) = A\phi(x, S).$$

Thus $f_1(\phi(x, S)) = x'S^{-1}x$ is maximal invariant. ◆

In the three examples above, the space $\mathcal{X}$ has been represented as a product space $\mathcal{Y} \times \mathcal{Z}$ in such a way that the group action on $\mathcal{X}$ corresponds to a group action on $\mathcal{Y} \times \mathcal{Z}$—namely,

$$g(y, z) = (gy, z)$$

and G acts transitively on $\mathcal{Y}$. Thus it is obvious that

$$f_1(y, z) = z$$

is maximal invariant for G acting on $\mathcal{Y} \times \mathcal{Z}$. However, the correspondence ϕ, a one-to-one onto mapping, satisfies

$$\phi(gx) = g\phi(x) \quad \text{for } g \in G, x \in \mathcal{X}.$$

The conclusion is that $f_1(\phi(x))$ is a maximal invariant function on $\mathcal{X}$. A direct proof in the present generality is easy. Since

$$f_1(\phi(gx)) = f_1(g\phi(x)) = f_1(\phi(x)),$$

$f_1(\phi(x))$ is invariant. If $f_1(\phi(x_1)) = f_1(\phi(x_2))$, then there is a $g \in G$ such that $g\phi(x_1) = \phi(x_2)$ since f_1 is maximal invariant on $\mathcal{Y} \times \mathcal{Z}$. But $g\phi(x_1) = \phi(gx_1) = \phi(x_2)$, so $gx_1 = x_2$ as ϕ is one-to-one. Thus $f_1(\phi(x))$ is maximal invariant. In the next example, a maximal invariant function is easily found but the product space representation in the form just discussed is not available.

◆ **Example 7.5.** The group $\mathcal{O}_p$ acts on $\mathcal{S}_p^+$ by

$$\Gamma(S) = \Gamma S \Gamma', \qquad \Gamma \in \mathcal{O}_p.$$

A maximal invariant function is easily found using a reduction argument similar to that given in Example 7.4. From the spectral theorem for matrices, every $S \in \mathcal{S}_p^+$ can be written in the form $S = \Gamma_1 D \Gamma_1'$ where $\Gamma_1 \in \mathcal{O}_p$ and D is a diagonal matrix whose diagonal elements are the ordered eigenvalues of S, say

$$\lambda_1(S) \geqslant \lambda_2(S) \geqslant \cdots \geqslant \lambda_p(S).$$

Thus $\Gamma_1' S \Gamma_1 = D$, which shows that D is in the orbit of S. Let f_0 on $\mathcal{S}_p^+$ to R^p be defined by: $f_0(S)$ is the vector of ordered eigenvalues of S. Obviously, f_0 is $\mathcal{O}_p$-invariant and, to show f_0 is maximal invariant, suppose $f_0(S_1) = f_0(S_2)$. Then S_1 and S_2 have the same eigenvalues and we have

$$S_i = \Gamma_i D \Gamma_i', \qquad i = 1,2$$

where D is the diagonal matrix of eigenvalues of S_i, $i = 1,2$. Thus $\Gamma_2 \Gamma_1' S_1 (\Gamma_2 \Gamma_1')' = S_2$, so f_0 is maximal invariant. To describe the technical difficulty when we try to write $\mathcal{S}_p^+$ as a product space, first consider the case $p = 2$. Then $\mathcal{S}_2^+ = \mathfrak{X}_1 \cup \mathfrak{X}_2$ where

$$\mathfrak{X}_1 = \{S | S \in \mathcal{S}_2^+, \lambda_1(S) = \lambda_2(S)\}$$

and

$$\mathfrak{X}_2 = \{S | S \in \mathcal{S}_2^+, \lambda_1(S) > \lambda_2(S)\}.$$

That $\mathcal{O}_2$ acts on both $\mathfrak{X}_1$ and $\mathfrak{X}_2$ is clear. The function ϕ_1 defined on $\mathfrak{X}_1$ by $\phi_1(S) = \lambda_1(S) \in (0, \infty)$ is maximal invariant and ϕ_1 establishes a one-to-one correspondence between $\mathfrak{X}_1$ and $(0, \infty)$. For $\mathfrak{X}_2$, define ϕ_2 by

$$\phi_2(S) = \begin{pmatrix} \lambda_1(S) & 0 \\ 0 & \lambda_2(S) \end{pmatrix}$$

so ϕ_2 is a maximal invariant function and takes values in the set $\mathfrak{Y}$ of all 2×2 diagonal matrices with diagonal elements y_1 and y_2, $y_1 > y_2 > 0$. Let $\mathfrak{D}_2$ be the subgroup of $\mathcal{O}_2$ consisting of those diagonal matrices with ± 1 for each diagonal element. The argument given in Example 6.21 shows that the mapping constructed there establishes a one-to-one onto correspondence between $\mathfrak{X}_2$ and $(\mathcal{O}_2/\mathfrak{D}_2) \times \mathfrak{Y}$, and $\mathcal{O}_2$ acts on $(\mathcal{O}_2/\mathfrak{D}_2) \times \mathfrak{Y}$ by

$$\Gamma(z, y) = (\Gamma z, y); \qquad (z, y) \in (\mathcal{O}_2/\mathfrak{D}_2) \times \mathfrak{Y}.$$

Further, ϕ satisfies

$$\phi(\Gamma(z, y)) = \Gamma(\phi(z, y)).$$

Thus for $p = 2$, $\mathcal{S}_2^+$ has been decomposed into $\mathfrak{X}_1$ and $\mathfrak{X}_2$, which are both invariant under $\mathcal{O}_2$. The action of $\mathcal{O}_2$ on $\mathfrak{X}_1$ is trivial in that $\Gamma x = x$ for all $x \in \mathfrak{X}_1$ and a maximal invariant function on $\mathfrak{X}_1$ is the identity function. Also, $\mathfrak{X}_2$ was decomposed into a product space where $\mathcal{O}_2$ acted transitively on the first component of the product space and trivially on the second component. From this decomposition, a maximal invariant function was obvious. Similar decompositions for $p > 2$ can be given for $\mathcal{S}_p^+$, but the number of component spaces increases. For example, when $p = 3$, let $\lambda_1(S) \geqslant \lambda_2(S) \geqslant \lambda_3(S)$ denote the ordered eigenvalues of $S \in \mathcal{S}_3^+$. The relevant decomposition for $\mathcal{S}_3^+$ is

$$\mathcal{S}_3^+ = \mathfrak{X}_1 \cup \mathfrak{X}_2 \cup \mathfrak{X}_3 \cup \mathfrak{X}_4$$

where

$$\mathfrak{X}_1 = \{S|\lambda_1(S) = \lambda_2(S) = \lambda_3(S)\}$$

$$\mathfrak{X}_2 = \{S|\lambda_1(S) = \lambda_2(S) > \lambda_3(S)\}$$

$$\mathfrak{X}_3 = \{S|\lambda_1(S) > \lambda_2(S) = \lambda_3(S)\}$$

$$\mathfrak{X}_4 = \{S|\lambda_1(S) > \lambda_2(S) > \lambda_3(S)\}.$$

Each of the four components is acted on by $\mathcal{O}_3$ and can be written as a product space with the structure described previously. The details of this are left to the reader. In some situations, it is sufficient to consider the subset $\mathfrak{Z}$ of $\mathcal{S}_p^+$ where

$$\mathfrak{Z} = \{S|\lambda_1(S) > \lambda_2(S) > \cdots > \lambda_p(S)\}.$$

The argument given in Example 6.21 shows how to write $\mathfrak{Z}$ as a product space so that a maximal invariant function is obvious under the action of $\mathcal{O}_p$ on $\mathfrak{Z}$. ◆

Further examples of maximal invariants are given as the need arises. We end this section with a brief discussion of equivariant functions. Recall (see Definition 6.14) that a function ϕ on $\mathfrak{X}$ onto $\mathfrak{Y}$ is called equivariant if $\phi(gx) = \bar{g}\phi(x)$ where G acts on $\mathfrak{X}$, $\bar{G}$ acts on $\mathfrak{Y}$, and $\bar{G}$ is a homomorphic image of G. If $\bar{G} = \{\bar{e}\}$ consists only of the identity, then equivariant functions are invariant under G. In this case, we have a complete description of all the equivariant functions—namely, a function is equivariant iff it is a function of a maximal invariant function on $\mathfrak{X}$. In the general case when $\bar{G}$ is not the trivial group, a useful description of all the equivariant functions appears to be rather difficult. However, there is one special case when the equivariant functions can be characterized.

Assume that G acts transitively on $\mathfrak{X}$ and $\bar{G}$ acts transitively on $\mathfrak{Y}$, where $\bar{G}$ is a homomorphic image of G. Fix $x_0 \in \mathfrak{X}$ and let

$$H_0 = \{g \mid gx_0 = x_0\}.$$

The subgroup H_0 of G is called the *isotropy subgroup* of x_0. Also, fix $y_0 \in \mathfrak{Y}$ and let

$$K_0 = \{\bar{g} \mid \bar{g}y_0 = y_0\}$$

be the isotropy subgroup of y_0.

Proposition 7.8. In order that there exist an equivariant function ϕ on $\mathfrak{X}$ to $\mathfrak{Y}$ such that $\phi(x_0) = y_0$, it is necessary and sufficient that $\bar{H}_0 \subseteq K_0$. Here $\bar{H}_0 \subseteq \bar{G}$ is the image of H_0 under the given homomorphism.

Proof. First, suppose that ϕ is equivariant and satisfies $\phi(x_0) = y_0$. Then, for $g \in H_0$,

$$\phi(x_0) = \phi(gx_0) = \bar{g}\phi(x_0) = \bar{g}y_0 = y_0$$

so $\bar{g} \in K_0$. Thus $\bar{H}_0 \subseteq K_0$. Conversely, suppose that $\bar{H}_0 \subseteq K_0$. For $x \in \mathfrak{X}$, the transitivity of G on $\mathfrak{X}$ implies that $x = gx_0$ for some g. Define ϕ on $\mathfrak{X}$ to $\mathfrak{Y}$ by

$$\phi(x) = \bar{g}y_0 \quad \text{where } x = gx_0.$$

It must be shown that ϕ is well defined and is equivariant. If $x = g_1x_0 = g_2x_0$, then $g_2^{-1}g_1 \in H_0$ so $\overline{g_2^{-1}g_1} \in K_0$. Thus

$$\phi(x) = \bar{g}_1 y_0 = \bar{g}_2 y_0$$

since

$$\bar{g}_2^{-1}\bar{g}_1 y_0 = \overline{g_2^{-1}g_1}\, y_0 = y_0.$$

Therefore ϕ is well defined and is onto $\mathcal{Y}$ since $\bar{G}$ acts transitively on $\mathcal{Y}$. That ϕ is equivariant is easily checked. □

The proof of Proposition 7.8 shows that an equivariant function is determined by its value at one point when G acts transitively on $\mathcal{X}$. More precisely, if ϕ_1 and ϕ_2 are equivariant functions on $\mathcal{X}$ such that $\phi_1(x_0) = \phi_2(x_0)$ for some $x_0 \in \mathcal{X}$, then $\phi_1(x) = \phi_2(x)$ for all x. To see this, write $x = gx_0$ so

$$\phi_1(x) = \phi_1(gx_0) = \bar{g}\phi_1(x_0) = \bar{g}\phi_2(x_0) = \phi_2(gx_0) = \phi_2(x).$$

Thus to characterize all the equivariant functions, it is sufficient to determine the possible values of $\phi(x_0)$ for some fixed $x_0 \in \mathcal{X}$. The following example illustrates these ideas.

◆ **Example 7.6.** Suppose $\mathcal{X} = \mathcal{Y} = \mathbb{S}_p^+$ and $G = \bar{G} = Gl_p$ where the homomorphism is the identity. The action of Gl_p on $\mathbb{S}_p^+$ is

$$A(S) = ASA'; \qquad A \in Gl_p,\ S \in \mathbb{S}_p^+.$$

To characterize the equivariant functions, pick $x_0 = I_p \in \mathbb{S}_p^+$. An equivariant function ϕ must satisfy

$$\phi(I_p) = \phi(\Gamma\Gamma') = \Gamma\phi(I_p)\Gamma'$$

for all $\Gamma \in \mathcal{O}_p$. By Proposition 2.13, a matrix $\phi(I_p)$ satisfies this equation iff $\phi(I_p) = kI_p$ for some real constant k. Since $\phi(I_p) \in \mathbb{S}_p^+$, $k > 0$. Thus

$$\phi(I_p) = kI_p, \qquad k > 0$$

and for $S \in \mathbb{S}_p^+$,

$$\phi(S) = \phi(S^{1/2}S^{1/2}) = S^{1/2}\phi(I_p)S^{1/2} = kS.$$

Therefore, every equivariant function has the form $\phi(S) = kS$ for some $k > 0$. ◆

Further applications of the above ideas occur in the following sections after it is shown that, under certain conditions, maximum likelihood estimators are equivariant functions.

7.3. INVARIANT PROBABILITY MODELS

Invariant probability models provide the mathematical framework in which the connection between statistical problems and invariance can be studied. Suppose $(\mathcal{X}, \mathcal{B})$ is a measurable space and G is a group of transformations acting on $\mathcal{X}$ such that each $g \in G$ is a one-to-one onto measurable function from $\mathcal{X}$ to $\mathcal{X}$. If P is a probability measure on $(\mathcal{X}, \mathcal{B})$ and $g \in G$, the probability measure gP on $(\mathcal{X}, \mathcal{B})$ is defined by

$$(gP)(B) \equiv P(g^{-1}B); \qquad g \in G, \quad B \in \mathcal{B}.$$

It is easily verified that $(g_1 g_2)P = g_1(g_2 P)$ so the group G acts on the space of all probability measures defined on $(\mathcal{X}, \mathcal{B})$.

Definition 7.3. Let $\mathcal{P}$ be a set of probability measures defined on $(\mathcal{X}, \mathcal{B})$. The set $\mathcal{P}$ is *invariant* under G if for each $P \in \mathcal{P}$, $gP \in \mathcal{P}$ for all $g \in G$. Sets of probability measures $\mathcal{P}$ are called *probability models*, and when $\mathcal{P}$ is invariant under G, we speak of a *G-invariant probability model.*

If $X \in \mathcal{X}$ is a random vector with $\mathcal{L}(X) = P$, then $\mathcal{L}(gX) = gP$ for $g \in G$ since

$$\Pr\{gX \in B\} = \Pr\{X \in g^{-1}B\} = P\{g^{-1}B\} = (gP)(B).$$

Thus $\mathcal{P}$ is invariant under G iff whenever $\mathcal{L}(X) \in \mathcal{P}$, $\mathcal{L}(gX) \in \mathcal{P}$ for all $g \in G$.

There are a variety of ways to construct invariant probability models from other invariant probability models. For example, if $\mathcal{P}_\alpha$, $\alpha \in A$, are G-invariant probability models, it is clear that

$$\bigcup_{\alpha \in A} \mathcal{P}_\alpha \quad \text{and} \quad \bigcap_{\alpha \in A} \mathcal{P}_\alpha$$

are both G-invariant. Now, given $(\mathcal{X}, \mathcal{B})$ and a G-invariant probability model $\mathcal{P}$, form the product space

$$\mathcal{X}^{(n)} = \mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}$$

and the product σ-algebra $\mathcal{B}^{(n)}$ on $\mathcal{X}^{(n)}$. For $P \in \mathcal{P}$, define $P^{(n)}$ on $\mathcal{B}^{(n)}$ by first defining

$$P^{(n)}(B_1 \times B_2 \times \cdots \times B_n) = \prod_{i=1}^{n} P(B_i)$$

where $B_i \in \mathcal{B}$. Once $P^{(n)}$ is defined on sets of the form $B_1 \times \cdots \times B_n$, its extension to $\mathcal{B}^{(n)}$ is unique. Also, define G acting on $\mathcal{X}^{(n)}$ by

$$g(x_1, \ldots, x_n) \equiv (gx_1, \ldots, gx_n)$$

for $x = (x_1, \ldots, x_n) \in \mathcal{X}^{(n)}$.

Proposition 7.9. Let $\mathcal{P}^{(n)} = \{P^{(n)} | P \in \mathcal{P}\}$. Then $\mathcal{P}^{(n)}$ is a G-invariant probability model on $(\mathcal{X}^{(n)}, \mathcal{B}^{(n)})$ when $\mathcal{P}$ is G-invariant.

Proof. It must be shown that $gP^{(n)} \in \mathcal{P}^{(n)}$ for $g \in G$ and $P^{(n)} \in \mathcal{P}^{(n)}$. However, $P^{(n)}$ is the product measure

$$P^{(n)} = P \times P \times \cdots \times P; \qquad P \in \mathcal{P}$$

and $P^{(n)}$ is determined by its values on sets of the form $B_1 \times \cdots \times B_n$. But

$$\begin{aligned} (gP)^{(n)}(B_1 \times \cdots \times B_n) &= P^{(n)}(g^{-1}B_1 \times g^{-1}B_2 \times \cdots \times g^{-1}B_n) \\ &= \prod_1^n P(g^{-1}B_i) = \prod_1^n (gP)(B_i) \end{aligned}$$

where the first equality follows from the definition of the action of G on $\mathcal{X}^{(n)}$. Then $gP^{(n)}$ is the product measure

$$gP^{(n)} = (gP) \times (gP) \times \cdots \times (gP),$$

which is in $\mathcal{P}^{(n)}$ as $gP \in \mathcal{P}$. □

For an application of Proposition 7.9, suppose X is a random vector with $\mathcal{L}(X) \in \mathcal{P}$ where $\mathcal{P}$ is a G-invariant probability model on $\mathcal{X}$. If $X_1, \ldots, X_n$ are independent and identically distributed with $\mathcal{L}(X_i) \in \mathcal{P}$, then the random vector

$$Y = (X_1, \ldots, X_n) \in \mathcal{X}^{(n)}$$

has distribution $P^{(n)} \in \mathcal{P}^{(n)}$ when $\mathcal{L}(X_i) = P$, $i = 1, \ldots, n$. Thus $\mathcal{P}^{(n)}$ is a G-invariant probability model for Y.

In most applications, probability models $\mathcal{P}$ are described in the form $\mathcal{P} = \{P_\theta | \theta \in \theta\}$ where θ is a *parameter* and Θ is the *parameter space*. When discussing indexed families of probability measures, the term "parameter space" is used only in the case that the indexing is one-to-one—that is,

$P_{\theta_1} = P_{\theta_2}$ implies that $\theta_1 = \theta_2$. Now, suppose $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ is G-invariant. Then for each $g \in G$ and $\theta \in \Theta$, $gP_\theta \in \mathcal{P}$, so $gP_\theta = P_{\theta'}$ for some unique $\theta' \in \Theta$. Define a function $\bar{g}$ on Θ to Θ by

$$gP_\theta = P_{\bar{g}\theta}, \quad \theta \in \Theta.$$

In other words, $\bar{g}\theta$ is the unique point in Θ that satisfies the above equation.

Proposition 7.10. Each $\bar{g}$ is a one-to-one onto function from Θ to Θ. Let $\bar{G} = \{\bar{g} | g \in G\}$. Then $\bar{G}$ is a group under the group operation of function composition and the mapping $g \to \bar{g}$ is a group homomorphism from G to $\bar{G}$, that is:

(i) $\overline{g_1 g_2} = \bar{g}_1 \bar{g}_2$.
(ii) $\overline{g^{-1}} = \bar{g}^{-1}$.

Proof. To show that $\bar{g}$ is one-to-one, suppose $\bar{g}\theta_1 = \bar{g}\theta_2$. Then

$$gP_{\theta_1} = P_{\bar{g}\theta_1} = P_{\bar{g}\theta_2} = gP_{\theta_2},$$

which implies that $P_{\theta_1} = P_{\theta_2}$ so $\theta_1 = \theta_2$. The verification that $\bar{g}$ is onto goes as follows. If $\theta \in \Theta$, let $\theta' = \overline{g^{-1}}\theta$. Then

$$P_{\bar{g}\theta'} = gP_{\theta'} = g\left(g^{-1}P_\theta\right) = \left(gg^{-1}\right)P_\theta = P_\theta,$$

so $\bar{g}\theta' = \theta$. Equations (i) and (ii) follow by calculations similar to those above. This shows that $\bar{G}$ is the homomorphic image of G and $\bar{G}$ is a group. □

An important special case of a G-invariant parametric model is the following. Suppose G acts on $(\mathcal{X}, \mathcal{B})$ and assume that ν is a σ-finite measure on $(\mathcal{X}, \mathcal{B})$ that is relatively invariant with multiplier χ, that is,

$$\int f(g^{-1}x)\nu(dx) = \chi(g)\int f(x)\nu(dx), \qquad g \in G$$

for all integrable functions f. Assume that $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ is a parametric model and

$$P_\theta(B) = \int I_B(x)p(x|\theta)\nu(dx)$$

for all measurable sets B. Thus $p(\cdot|\theta)$ is a density for P_θ with respect to ν. If

$\mathcal{P}$ is G-invariant, then

$$gP_\theta = P_{\bar{g}\theta} \qquad \text{for } g \in G, \theta \in \Theta.$$

Therefore,

$$\begin{aligned} gP_\theta(B) = P_\theta(g^{-1}B) &= \int I_B(gx)p(x|\theta)\nu(dx) \\ &= \int I_B(gx)p(g^{-1}gx|\theta)\nu(dx) \\ &= \chi(g^{-1})\int I_B(x)p(g^{-1}x|\theta)\nu(dx) \\ &= P_{\bar{g}\theta}(B) = \int I_B(x)p(x|\bar{g}\theta)\nu(dx) \end{aligned}$$

for all measurable sets B. Thus the density p must satisfy

$$\chi(g^{-1})p(g^{-1}x|\theta) = p(x|\bar{g}\theta) \qquad \text{a.e. } (\nu)$$

or, equivalently,

$$p(x|\theta) = p(gx|\bar{g}\theta)\chi(g) \qquad \text{a.e. } (\nu).$$

It should be noted that the null set where the above equality does not hold may depend on both θ and g. However, in most applications, a version of the density is available so the above equality is valid everywhere. This leads to the following definition.

Definition 7.4. The family of densities $\{p(\cdot|\theta)|\theta \in \Theta\}$ with respect to the relatively invariant measure ν with multiplier χ is $(G - \bar{G})$-invariant if

$$p(x|\theta) = p(gx|\bar{g}\theta)\chi(g)$$

for all x, θ, and g.

It is clear that if a family of densities is $(G - \bar{G})$-invariant where $\bar{G}$ is a homomorphic image of G that acts on Θ, then the family of probability measures defined by these densities is a G-invariant probability model. A few examples illustrate these notions.

◆ **Example 7.7.** Let $\mathfrak{X} = R^n$ and suppose $f(\|x\|^2)$ is a density with respect to Lebesgue measure on R^n. For $\mu \in R^n$ and $\Sigma \in \mathbb{S}_p^+$, set

$$p(x|\mu, \Sigma) = |\Sigma|^{-1/2} f((x - \mu)'\Sigma^{-1}(x - \mu)).$$

For each μ and Σ, $p(\cdot|\mu, \Sigma)$ is a density on R^n. The affine group Al_n acts on R^n by $(A, b)x = Ax + b$ and Lebesgue measure is relatively invariant with multiplier

$$\chi(A, b) = |\det(A)|$$

where $(A, b) \in Al_n$. Consider the parameter space $R^n \times \mathbb{S}_p^+$ and the family of densities

$$\{p(\cdot|\mu, \Sigma)|(\mu, \Sigma) \in R^n \times \mathbb{S}_p^+\}.$$

The group Al_n acts on the parameter space $R^n \times \mathbb{S}_p^+$ by

$$(A, b)(\mu, \Sigma) = (A\mu + b, A\Sigma A').$$

It is now verified that the family of densities above is $(G - \overline{G})$-invariant where $G = \overline{G} = Al_n$. For $(A, b) \in Al_n$,

$$\begin{aligned} p((A, b)x|(A, b)(\mu, \Sigma)) &= p(Ax + b|(A\mu + b, A\Sigma A')) \\ &= |A\Sigma A'|^{-1/2} f\big((Ax + b - A\mu - b)' \\ &\quad \times (A\Sigma A')^{-1}(Ax + b - A\mu - b)\big) \\ &= |\det A|^{-1}|\Sigma|^{-1/2} f((x - \mu)'\Sigma^{-1}(x - \mu)) \\ &= \frac{1}{\chi(A, b)} p(x|(\mu, \Sigma)). \end{aligned}$$

Therefore, the parametric model determined by the family of densities is Al_n-invariant. ◆

A useful method for generating a G-invariant probability model on a measurable space $(\mathfrak{X}, \mathfrak{B})$ is to consider a fixed probability measure P_0 on $(\mathfrak{X}, \mathfrak{B})$ and set

$$\mathfrak{P} = \{gP_0|g \in G\}.$$

Obviously, $\mathfrak{P}$ is G-invariant. However, in many situations, the group G does not serve as a parameter space for $\mathcal{P}$ since $g_1 P_0 = g_2 P_0$ does not necessarily imply that $g_1 = g_2$. For example, consider $\mathfrak{X} = R^n$ and let P_0 be given by

$$P_0(B) = \int_{R^n} I_B(x) f(\|x\|^2)\, dx$$

where $f(\|x\|^2)$ is the density on R^n of Example 7.7. Also, let $G = Al_n$. To obtain the density of gP_0, suppose X is a random vector with $\mathcal{L}(X) = P_0$. For $g = (A, b) \in Al_n$, $(A, b)X = AX + b$ has a density given by

$$p(x|b, AA') = |\det(AA')|^{-1/2} f\big((x-b)'(AA')^{-1}(x-b)\big)$$

and this is the density of $(A, b)P_0$. Thus the parameter space for

$$\mathcal{P} = \{(A, b)P_0 | (A, b) \in Al_n\}$$

is $R^n \times \mathcal{S}_n^+$. Of course, the reason that Al_n is not a parameter space for $\mathcal{P}$ is that

$$(\Gamma, 0)P_0 = P_0$$

for all $n \times n$ orthogonal matrices Γ. In other words, P_0 is an orthogonally invariant probability on R^n.

Some of the linear models introduced in Chapter 4 provide interesting examples of parametric models that are generated by groups of transformations.

◆ **Example 7.8.** Consider an inner product space $(V, [\cdot, \cdot])$ and let P_0 be a probability measure on V so that if $\mathcal{L}(X) = P_0$, then $\mathcal{E}X = 0$ and $\text{Cov}(X) = I$. Given a subspace M of V, form the group G whose elements consist of pairs (a, x) with $a > 0$ and $x \in M$. The group operation is

$$(a_1, x_1)(a_2, x_2) \equiv (a_1 a_2, a_1 x_2 + x_1).$$

The probability model $\mathcal{P} = \{gP_0 | g \in G\}$ consists of all the distributions of $(a, x)X = aX + x$ where $\mathcal{L}(X) = P_0$. Clearly,

$$\mathcal{E}(aX + x) = x \quad \text{and} \quad \text{Cov}(aX + x) = a^2 I.$$

Therefore, if $\mathcal{L}(Y) \in \mathcal{P}$, then $\mathcal{E}Y \in M$ and $\text{Cov}(Y) = \sigma^2 I$ for some $\sigma^2 > 0$, so $\mathcal{P}$ is a linear model for Y. For this particular example the group G is a parameter space for $\mathcal{P}$. This linear model is generated by G in the sense that $\mathcal{P}$ is obtained by transforming a fixed probability measure P_0 by elements of G. ◆

An argument similar to that in Example 7.8 shows that the multivariate linear model introduced in Example 4.4 is also generated by a group of transformations.

◆ **Example 7.9.** Let $\mathcal{L}_{p,n}$ be the linear space of real $n \times p$ matrices with the usual inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{L}_{p,n}$. Assume that P_0 is a probability measure on $\mathcal{L}_{p,n}$ so that, if $\mathcal{L}(X) = P_0$, then $\mathcal{E}X = 0$ and $\operatorname{Cov}(X) = I_n \otimes I_p$. To define a regression subspace M, let Z be a fixed $n \times k$ real matrix and set

$$M = \{y \mid y = ZB,\ B \in \mathcal{L}_{p,k}\}.$$

Obviously, M is a subspace of $\mathcal{L}_{p,n}$. Consider the group G whose elements are pairs (A, y) with $A \in Gl_p$ and $y \in M$. Then G acts on $\mathcal{L}_{p,n}$ by

$$(A, y)x = xA' + y = (I_n \otimes A)x + y,$$

and the group operation is

$$(A_1, y_1)(A_2, y_2) = (A_1 A_2,\ y_2 A_1' + y_1).$$

The probability model $\mathcal{P} = \{gP_0 \mid g \in G\}$ consists of the distributions of $(A, y)X = (I_n \otimes A)X + y$ where $\mathcal{L}(X) = P_0$. Since

$$\mathcal{E}((I_n \otimes A)X + y) = y \in M$$

and

$$\operatorname{Cov}((I_n \otimes A)X + y) = I_n \otimes AA',$$

if $\mathcal{L}(Y) \in \mathcal{P}$, then $\mathcal{E}Y \in M$ and $\operatorname{Cov}(Y) = I_n \otimes \Sigma$ for some $p \times p$ positive definite matrix Σ. Thus $\mathcal{P}$ is a multivariate linear model as described in Example 4.4. If $p > 1$, the group G is not a parameter space for $\mathcal{P}$, but G does generate $\mathcal{P}$. ◆

Most of the probability models discussed in later chapters are examples of probability models generated by groups of transformations. Thus these models are G-invariant and this invariance can be used in a variety of ways.

First, invariance can be used to give easy derivations of maximum likelihood estimators and to suggest test statistics in some situations. In addition, distributional and independence properties of certain statistics are often best explained in terms of invariance.

7.4. THE INVARIANCE OF LIKELIHOOD METHODS

In this section, it is shown that under certain conditions maximum likelihood estimators are equivariant functions and likelihood ratio tests are invariant functions. Throughout this section, G is a group of transformations that act measurably on $(\mathfrak{X}, \mathfrak{B})$ and ν is a σ-finite relatively invariant measure on $(\mathfrak{X}, \mathfrak{B})$ with multiplier χ. Suppose that $\mathfrak{P} = \{P_\theta | \theta \in \Theta\}$ is a G-invariant parametric model such that each P_θ has a density $p(\cdot|\theta)$, which satisfies

$$p(x|\theta) = p(gx|\bar{g}\theta)\chi(g)$$

for all $x \in \mathfrak{X}$, $\theta \in \Theta$, and $g \in G$. The group $\bar{G} = \{\bar{g} | g \in G\}$ is the homomorphic image of G described in Proposition 7.10. In the present context, a point estimator of θ, say t, mapping $\mathfrak{X}$ into Θ, is equivariant (see Definition 6.14) if

$$t(gx) = \bar{g}t(x), \qquad g \in G, \quad x \in \mathfrak{X}.$$

Proposition 7.11. Given the $(G - \bar{G})$-invariant family of densities $\{p(\cdot|\theta) | \theta \in \Theta\}$, assume there exists a unique function $\hat{\theta}$ mapping $\mathfrak{X}$ into Θ that satisfies

$$\sup_{\theta \in \Theta} p(x|\theta) = p(x|\hat{\theta}(x)).$$

Then $\hat{\theta}$ is an equivariant function—that is,

$$\hat{\theta}(gx) = \bar{g}\hat{\theta}(x), \qquad x \in \mathfrak{X}, \quad g \in G.$$

Proof. By assumption, $\hat{\theta}(gx)$ is the unique point in Θ that satisfies

$$\sup_{\theta \in \Theta} p(gx|\theta) = p(gx|\hat{\theta}(gx)).$$

But

$$p(gx|\theta) = \chi(g^{-1})p(x|\bar{g}^{-1}\theta)$$

so

$$\sup_\theta p(gx|\theta) = \chi(g^{-1}) \sup_\theta p(x|\bar{g}^{-1}\theta) = \chi(g^{-1}) \sup_\theta p(x|\theta)$$

$$= \chi(g^{-1}) p(x|\hat{\theta}(x)) = p(gx|\bar{g}\hat{\theta}(x)).$$

Thus

$$p(gx|\hat{\theta}(gx)) = p(gx|\bar{g}\hat{\theta}(x))$$

and, by the uniqueness assumption,

$$\hat{\theta}(gx) = \bar{g}\hat{\theta}(x). \qquad \square$$

Of course, the estimator $\hat{\theta}(x)$ whose existence and uniqueness is assumed in Proposition 7.11 is the maximum likelihood estimator of θ. That $\hat{\theta}$ is an equivariant function is useful information about the maximum likelihood estimator, but the above result does not indicate how to use invariance to find the maximum likelihood estimator. The next result rectifies this situation.

Proposition 7.12. Let $\{p(\cdot|\theta)|\theta \in \Theta\}$ be a $(G - \bar{G})$-invariant family of densities on $(\mathfrak{X}, \mathfrak{B})$. Fix a point $x_0 \in \mathfrak{X}$ and let $\mathcal{O}_{x_0}$ be the orbit of x_0. Assume that

$$\sup_{\theta \in \Theta} p(x_0|\theta) = p(x_0|\theta_0)$$

and that θ_0 is unique. For $x \in \mathcal{O}_{x_0}$, define $\hat{\theta}(x)$ by

$$\hat{\theta}(x) = \bar{g}_x \theta_0 \qquad \text{where } x = g_x x_0.$$

Then $\hat{\theta}$ is well defined on $\mathcal{O}_{x_0}$ and satisfies

(i) $\hat{\theta}(gx) = \bar{g}\hat{\theta}(x)$, $x \in \mathcal{O}_{x_0}$.
(ii) $\sup_{\theta \in \Theta} p(x|\theta) = p(x|\hat{\theta}(x))$, $x \in \mathcal{O}_{x_0}$.

Furthermore, $\hat{\theta}$ is unique.

Proof. The density $p(\cdot|\theta)$ satisfies

$$p(y|\theta) = p(gy|\bar{g}\theta)\chi(g)$$

where χ is a multiplier on G. To show $\hat{\theta}$ is well defined on $\mathcal{O}_{x_0}$, it must be verified that if $x = g_x x_0 = h_x x_0$, then $\bar{g}_x \theta_0 = \bar{h}_x \theta_0$. Set $k = h_x^{-1} g_x$ so $k x_0 = x_0$ and we need to show that $\bar{k}\theta_0 = \theta_0$. But

$$p(x_0|\theta_0) = \sup_{\theta \in \Theta} p(kx_0|\theta) = \chi(k^{-1}) \sup_\theta p(x_0|\bar{k}^{-1}\theta)$$

$$= \chi(k^{-1}) \sup_\theta p(x_0|\theta) = \chi(k^{-1}) p(x_0|\theta_0)$$

$$= p(kx_0|\bar{k}\theta_0) = p(x_0|\bar{k}\theta_0).$$

By the uniqueness assumption, $\bar{k}\theta_0 = \theta_0$ so $\hat{\theta}$ is well defined on $\mathcal{O}_{x_0}$. To establish (i), if $x = g_x x_0$, then $gx = (gg_x)x_0$ so

$$\hat{\theta}(gx) = \overline{gg_x}\theta_0 = \bar{g}(\bar{g}_x\theta_0) = \bar{g}\hat{\theta}(x).$$

For (ii), $x = g_x x_0$ so

$$\sup_\theta p(x|\theta) = \sup_\theta p(g_x x_0|\theta) = \chi(g_x^{-1}) \sup_\theta p(x_0|\bar{g}_x^{-1}\theta_0)$$

$$= \chi(g_x^{-1}) \sup_\theta p(x_0|\theta) = \chi(g_x^{-1}) p(x_0|\theta_0)$$

$$= p(g_x x_0|\bar{g}_x\theta_0) = p(x|\hat{\theta}(x)).$$

To establish the uniqueness of $\hat{\theta}$, fix $x \in \mathcal{O}_{x_0}$ and consider $\theta_1 \neq \bar{g}_x\theta_0$. Then

$$p(x|\theta_1) = p(g_x x_0|\bar{g}_x \bar{g}_x^{-1}\theta_1) = \chi(g_x^{-1}) p(x_0|\bar{g}_x^{-1}\theta_1)$$

$$< \chi(g_x^{-1}) p(x_0|\theta_0) = p(x|\hat{\theta}(x)).$$

The strict inequality follows from the uniqueness assumption concerning θ_0. □

In applications, Proposition 7.12 is used as follows. From each orbit in the sample space $\mathcal{X}$, we pick a convenient point x_0 and show that $p(x_0|\theta)$ is uniquely maximized at θ_0. Then for other points x in this orbit, write $x = g_x x_0$ and set $\hat{\theta}(x) = \bar{g}_x\theta_0$. The function $\hat{\theta}$ is then the maximum likelihood estimator of θ and is equivariant. In some situations, there is only one orbit in $\mathfrak{X}$ so this method is relatively easy to apply.

◆ **Example 7.10.** Consider $\mathfrak{X} = \Theta = \mathbb{S}_p^+$ and let

$$p(S|\Sigma) = \omega(n, p)|\Sigma^{-1}S|^{n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S\right]$$

for $S \in \mathbb{S}_p^+$ and $\Sigma \in \mathbb{S}_p^+$. The constant $\omega(n, p)$, $n \geqslant p$, was defined in Example 5.1. That $p(\cdot|\Sigma)$ is a density with respect to the measure

$$\nu(dS) = \frac{dS}{|S|^{(p+1)/2}}$$

follows from Example 5.1. The group Gl_p acts on $\mathbb{S}_p^+$ by

$$A(S) \equiv ASA'$$

for $A \in Gl_p$ and $S \in \mathbb{S}_p^+$ and the measure ν is invariant. Also, it is clear that the density $p(\cdot|\Sigma)$ satisfies

$$p(ASA'|A\Sigma A') = p(S|\Sigma).$$

To find the maximum likelihood estimator of $\Sigma \in \mathbb{S}_p^+$, we apply the technique described above. Consider the point $I_p \in \mathbb{S}_p^+$ and note that the orbit of I_p under the action of Gl_p is $\mathbb{S}_p^+$ so in this case there is only one orbit. Thus to apply Proposition 7.12, it must be verified that

$$\sup_{\Sigma \in \mathbb{S}_p^+} p(I_p|\Sigma) = p(I_p|\Sigma_0)$$

where Σ_0 is unique. Taking the logarithm of $p(I_p|\Sigma)$ and ignoring the constant term, we have

$$\sup_{\Sigma \in \mathbb{S}_p^+}\left[\frac{n}{2}\log|\Sigma^{-1}| - \frac{1}{2}\operatorname{tr}\Sigma^{-1}\right] = \sup_{B \in \mathbb{S}_p^+}\left[\frac{n}{2}\log|B| - \frac{1}{2}\operatorname{tr}B\right]$$

$$= \sup_{\lambda_i > 0}\left[\frac{n}{2}\sum_1^p \log\lambda_i - \frac{1}{2}\sum_1^p \lambda_i\right]$$

where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $B = \Sigma^{-1} \in \mathbb{S}_p^+$. However, for $\lambda > 0$, $n\log\lambda - \lambda$ is a strictly concave function of λ and is uniquely maximized at $\lambda = n$. Thus the function

$$\frac{n}{2}\sum_1^n \log \lambda_i - \frac{1}{2}\sum_1^n \lambda_i$$

is uniquely maximized at $\lambda_1 = \cdots = \lambda_n = n$, which means that

$$\frac{n}{2}\log|B| - \frac{1}{2}\operatorname{tr} B$$

is uniquely maximized at $B = nI$. Therefore,

$$\sup_{\Sigma \in \mathbb{S}_p} p(I_p|\Sigma) = p\left(I_p|\frac{1}{n}I_p\right)$$

and $(1/n)I_p$ is the unique point in $\mathbb{S}_p^+$ that achieves this supremum. To find the maximum likelihood estimator of Σ, say $\hat{\Sigma}(S)$, write $S = AA'$ for $A \in Gl_p$. Then

$$\hat{\Sigma}(S) = A\left(\frac{1}{n}I_p\right) = \frac{1}{n}AI_pA' = \frac{1}{n}S.$$

In summary,

$$\hat{\Sigma} = \frac{1}{n}S$$

is the unique maximum likelihood estimator of Σ and

$$\begin{aligned}
\sup_{\Sigma \in \mathbb{S}_p^+} p(S|\Sigma) &= p\left(S|\frac{1}{n}S\right) \\
&= \omega(n, p)\left|\left(\frac{1}{n}S\right)^{-1}S\right|^{n/2}\exp\left[-\frac{1}{2}\operatorname{tr}\left(\frac{1}{n}S\right)^{-1}S\right] \\
&= \omega(n, p)n^{np/2}\exp\left[-\frac{np}{2}\right].
\end{aligned}$$

The results of this example are used later to derive the maximum likelihood estimator of a covariance matrix in a variety of multivariate normal models. ◆

We now turn to the invariance of likelihood ratio tests. First, invariant testing problems need to be defined. Let $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ be a parametric probability model on $(\mathfrak{X}, \mathfrak{B})$ and suppose that G acts measurably on $\mathfrak{X}$. Let Θ_0 and Θ_1 be two disjoint subsets of Θ. On the basis of an observation vector $X \in \mathfrak{X}$ with $\mathcal{L}(X) \in \mathcal{P}_0 \cup \mathcal{P}_1$ where

$$\mathcal{P}_i = \{P_\theta | \theta \in \Theta_i\}, \qquad i = 0, 1,$$

suppose it is desired to test the hypothesis

$$H_0 : \mathcal{L}(X) \in \mathcal{P}_0$$

against the alternative

$$H_1 : \mathcal{L}(X) \in \mathcal{P}_1.$$

Definition 7.5. The above hypothesis testing problem is invariant under G if $\mathcal{P}_0$ and $\mathcal{P}_1$ are both G-invariant probability models.

Now suppose that $\mathcal{P}_0 = \{P_\theta | \theta \in \Theta_0\}$ and $\mathcal{P}_1 = \{P_\theta | \theta \in \Theta_1\}$ are disjoint families of probability measures on $(\mathfrak{X}, \mathfrak{B})$ such that each P has a density $p(\cdot|\theta)$ with respect to a σ-finite measure ν. Consider

$$\Lambda(x) = \frac{\sup_{\theta \in \Theta_0} p(x|\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} p(x|\theta)}.$$

For testing the null hypothesis that $\mathcal{L}(X) \in \mathcal{P}_0$ versus the alternative that $\mathcal{L}(X) \in \mathcal{P}_1$, the test that rejects the null hypothesis iff $\Lambda(x) < k$, where k is chosen to control the level of the test, is commonly called the *likelihood ratio test*.

Proposition 7.13. Given the family of densities $\{p(\cdot|\theta) | \theta \in \Theta_0 \cup \Theta_1\}$, assume the testing problem for $\mathcal{L}(X) \in \mathcal{P}_0$ versus $\mathcal{L}(X) \in \mathcal{P}_1$ is invariant under a group G and suppose that

$$p(x|\theta) = p(gx|\bar{g}\theta)\chi(g)$$

for some multiplier χ. Then the likelihood ratio

$$\Lambda(x) \equiv \frac{\sup_{\theta \in \Theta_0} p(x|\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} p(x|\theta)}$$

is an invariant function.

Proof. It must be shown that $\Lambda(x) = \Lambda(gx)$ for $x \in \mathfrak{X}$ and $g \in G$. For $g \in G$,

$$\Lambda(gx) = \frac{\sup_{\theta \in \Theta_0} p(gx|\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} p(gx|\theta)} = \frac{\sup_{\theta \in \Theta_0} \chi(g^{-1}) p(x|\bar{g}^{-1}\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} \chi(g^{-1}) p(x|\bar{g}^{-1}\theta)}$$

$$= \frac{\sup_{\theta \in \Theta_0} p(x|\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} p(x|\theta)} = \Lambda(x).$$

The next to the last equality follows from the positivity of χ and the invariance of Θ_0 and $\Theta_0 \cup \Theta_1$. □

For invariant testing problems, Proposition 7.13 shows that that test function determined by Λ, namely

$$\phi_0(x) = \begin{cases} 1 & \text{if } \Lambda(x) < k \\ 0 & \text{if } \Lambda(x) \geqslant k, \end{cases}$$

is an invariant function. More generally, any test function ϕ is invariant if $\phi(x) = \phi(gx)$ for all $x \in \mathfrak{X}$ and $g \in G$. The whole point of the above discussion is to show that, when attention is restricted to invariant tests for invariant testing problems, the likelihood ratio test is never excluded from consideration. Furthermore, if a particular invariant test has been shown to have an optimal property among invariant tests, then this test has been compared to the likelihood ratio test. Illustrations of these comments are given later in this section when we consider testing problems for the multivariate normal distribution.

Comments similar to those above apply to equivariant estimators. Suppose $\{p(\cdot|\theta)|\theta \in \Theta\}$ is a $(G - \bar{G})$-invariant family of densities and satisfies

$$p(x|\theta) = p(gx|\bar{g}\theta)\chi(g)$$

for some multiplier χ. If the conditions of Proposition 7.12 hold, then an equivariant maximum likelihood estimator exists. Thus if an equivariant estimator t with some optimal property (relative to the class of all equivariant estimators) has been found, then this property holds when t is compared to the maximum likelihood estimator. The Pitman estimator, derived in the next example, is an illustration of this situation.

◆ **Example 7.11.** Let f be a density on R^p with respect to Lebesgue measure and consider the translation family of densities $\{p(\cdot|\theta)|\theta \in R^p\}$ defined by

$$p(x|\theta) = f(x - \theta), \qquad x, \theta \in R^p.$$

For this example, $\mathfrak{X} = \Theta = G = R^p$ and the group action is

$$g(x) = x + g, \qquad x, g \in R^p.$$

It is clear that

$$p(gx|g\theta) = p(x|\theta),$$

so the family of densities is invariant and the multiplier is unity. It is assumed that

$$\int_{R^p} xf(x)\,dx = 0 \quad \text{and} \quad \int \|x\|^2 f(x)\,dx < +\infty.$$

Initially, assume we have one observation X with $\mathcal{L}(X) \in \{p(\cdot|\theta)|\theta \in R^p\}$. The problem is to estimate the parameter θ. As a measure of how well an estimator t performs, consider

$$R(t, \theta) \equiv \mathcal{E}_\theta \|t(X) - \theta\|^2.$$

If $t(X)$ is close to θ on the average, then $R(t, \theta)$ should be small. We now want to show that, if t is an equivariant estimator of θ, then

$$R(t, \theta) = R(t, 0)$$

and the equivariant estimator $t_0(X) = X$ minimizes $R(t, 0)$ over all equivariant estimators. If t is an equivariant estimator, then

$$t(x + g) = t(x) + g$$

so, with $g = -x$,

$$t(x) = x + t(0).$$

Therefore, every equivariant estimator has the form $t(x) = x + c$ where $c \in R^p$ is a constant. Conversely, any such estimator $t(x) =$

$x + c$ is equivariant. For $t(x) = x + c$,

$$R(t, \theta) = \mathcal{E}_\theta \|t(X) - \theta\|^2 = \int \|x + c - \theta\|^2 f(x - \theta)\, dx$$

$$= \int \|x + c\|^2 f(x)\, dx = R(t, 0).$$

To minimize $R(t, 0)$ over all equivariant t, the integral

$$\int \|x + c\|^2 f(x)\, dx$$

must be minimized by an appropriate choice of c. But

$$\mathcal{E}\|X + c\|^2 = \mathcal{E}\|X - \mathcal{E}(X)\|^2 + \|\mathcal{E}(X) + c\|^2$$

so

$$c = -\mathcal{E}(X) = \int x f(x)\, dx = 0$$

minimizes the above integral. Hence $t_0(X) = X$ minimizes $R(t, 0)$ over all equivariant estimators. Now, we want to generalize this result to the case when $X_1, \ldots, X_n$ are independent and identically distributed with $\mathcal{L}(X_i) \in \{p(\cdot|\theta)|\theta \in R^p\}$, $i = 1, \ldots, p$. The argument is essentially the same as when $n = 1$. An estimator t is equivariant if

$$t(x_1 + g, \ldots, x_n + g) = t(x_1, \ldots, x_n) + g$$

so, setting $g = -x_1$,

$$t(x_1, \ldots, x_n) = x_1 + t(0, x_2 - x_1, \ldots, x_n - x_1).$$

Conversely, if

$$t(x_1, \ldots, x_n) = x_1 + \Psi(x_2 - x_1, \ldots, x_n - x_1)$$

then t is equivariant. Here, Ψ is some measurable function taking values in R^p. Thus a complete description of the equivariant estimators has been given. For such an estimator,

$$R(t,\theta) = \mathcal{E}_\theta \|t(X_1,\ldots,X_n) - \theta\|^2 = \mathcal{E}_\theta \|t(X_1 - \theta,\ldots,X_n - \theta)\|^2$$
$$= \mathcal{E}_0 \|t(X_1,\ldots,X_n)\|^2 = R(t,0).$$

To minimize $R(t,0)$, we need to choose the function Ψ to minimize

$$R(t,0) = \mathcal{E}_0 \|X_1 + \Psi(X_2 - X_1,\ldots,X_n - X_1)\|^2.$$

Let $U_i = X_i - X_1$, $i = 2,\ldots,n$. Then

$$R(t,0) = \mathcal{E}_0 \|X_1 + \Psi(U_2,\ldots,U_n)\|^2$$
$$= \mathcal{E}\{\mathcal{E}\{\|X_1 + \Psi(U_2,\ldots,U_n)\|^2 | U_2,\ldots,U_n\}\}.$$

However, conditional on $(U_2,\ldots,U_n) \equiv U$,

$$\mathcal{E}(\|X_1 + \Psi(U)\|^2|U) = \mathcal{E}(\|X_1 - \mathcal{E}(X_1|U) + \mathcal{E}(X_1|U) + \Psi(U)\|^2|U)$$
$$= \mathcal{E}(\|X_1 - \mathcal{E}(X_1|U)\|^2|U) + \|\mathcal{E}(X_1|U) + \Psi(U)\|^2.$$

Thus it is clear that

$$\Psi_0(U) \equiv -\mathcal{E}(X_1|U)$$

minimizes $R(t,0)$. Hence the equivariant estimator

$$t_0(X_1,\ldots,X_n) = X_1 - \mathcal{E}_0(X_1|X_2 - X_1,\ldots,X_n - X_1)$$

satisfies

$$R(t_0,\theta) = R(t_0,0) \leqslant R(t,0) = R(t,\theta)$$

for all $\theta \in R^p$ and all equivariant estimators t. The estimator t_0 is commonly called the Pitman estimator. ◆

7.5. DISTRIBUTION THEORY AND INVARIANCE

When a family of distributions is invariant under a group of transformations, useful information can often be obtained about the distribution of invariant functions by using the invariance. For example, some of the results in Section 7.1 are generalized here.

The first result shows that the distribution of an invariant function depends invariantly on a parameter. Suppose $(\mathfrak{X}, \mathfrak{B})$ is a measurable space acted on measurably by a group G. Consider an invariant probability model $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ and let $\overline{G}$ be the induced group of transformations on Θ. Thus

$$gP_\theta = P_{\bar{g}\theta}, \qquad \theta \in \Theta, \quad g \in G.$$

A measurable mapping τ on $(\mathfrak{X}, \mathfrak{B})$ to $(\mathfrak{Y}, \mathcal{C})$ induces a family of distributions on $(\mathfrak{Y}, \mathcal{C})$, $\{Q_\theta | \theta \in \Theta\}$ given by

$$Q_\theta(C) \equiv P_\theta(\tau^{-1}(C)), \qquad C \in \mathcal{C}, \quad \theta \in \Theta.$$

Proposition 7.14. If τ is G-invariant, then $Q_\theta = Q_{\bar{g}\theta}$ for $\theta \in \Theta$ and $\bar{g} \in \overline{G}$.

Proof. For each $C \in \mathcal{C}$, it must be shown that

$$Q_\theta(C) = Q_{\bar{g}\theta}(C)$$

or, equivalently, that

$$P_\theta(\tau^{-1}(C)) = P_{\bar{g}\theta}(\tau^{-1}(C)).$$

But

$$P_{\bar{g}\theta}(\tau^{-1}(C)) = (gP_\theta)(\tau^{-1}(C)) = P_\theta(g^{-1}\tau^{-1}C)) = P_\theta\big((\tau g)^{-1}(C)\big).$$

Since $\tau g = \tau$ as τ is invariant,

$$Q_{\bar{g}\theta}(C) = P_\theta\big((\tau g)^{-1}(C)\big) = P_\theta(\tau^{-1}(C)) = Q_\theta(C). \qquad \square$$

An alternative formulation of Proposition 7.14 is useful. If $\mathcal{L}(X) \in \{P_\theta | \theta \in \Theta\}$ and if τ is G-invariant, then the induced distribution of $Y = \tau(X)$, which is Q_θ, satisfies $Q_\theta = Q_{\bar{g}\theta}$. In other words, the distribution of an invariant function depends only on a maximal invariant parameter. By definition, a maximal invariant parameter is any function defined on Θ that is maximal invariant under the action of $\overline{G}$ on Θ. Of course, Θ is usually not a parameter space for the family $\{Q_\theta | \theta \in \Theta\}$ as $Q_\theta = Q_{\bar{g}\theta}$, but any maximal $\overline{G}$-invariant function on Θ often serves as a parameter index for the distribution of $Y = \tau(X)$.

◆ **Example 7.12.** In this example, we establish a property of the distribution of the bivariate sample correlation coefficient. Consider a family of densities $p(\cdot|\mu, \Sigma)$ on R^2 given by

$$p(x|\mu, \Sigma) = |\Sigma|^{-1/2} f_0\big((x-\mu)'\Sigma^{-1}(x-\mu)\big)$$

where $\mu \in R^2$ and $\Sigma \in \mathbb{S}_2^+$. Hence

$$\int_{R^2} f_0\big(\|x\|^2\big)\, dx = 1$$

and it is assumed that

$$\int_{R^2} \|x\|^2 f_0\big(\|x\|^2\big)\, dx < +\infty.$$

Since the distribution on R^2 determined by f_0 is orthogonally invariant, if $Z \in R^2$ has density $f_0(\|x\|^2)$, then

$$\mathcal{E}Z = 0 \quad \text{and} \quad \operatorname{Cov}(Z) = cI_2$$

for some $c > 0$ (see Proposition 2.13). Also, $Z_1 = \Sigma^{1/2}Z + \mu$ has density $p(\cdot|\mu, \Sigma)$ when Z has density $f_0(\|x\|^2)$. Thus

$$\mathcal{E}Z_1 = \mu \quad \text{and} \quad \operatorname{Cov}(Z_1) = c\Sigma.$$

The group Al_2 acts on R^2 by

$$(A, b)x = Ax + b$$

and it is clear that the family of distributions, say $\mathcal{P} = \{P_{\mu,\Sigma} | (\mu, \Sigma) \in \mathcal{R}^2 \times \mathbb{S}_2^+\}$, having the densities $p(\cdot|\mu, \Sigma)$, $\mu \in R^2$, $\Sigma \in \mathbb{S}_2^+$, is invariant under this group action. Lebesgue measure on R^2 is relatively invariant with multiplier

$$\chi(A, b) = |\det(A)|$$

and

$$p(x|\mu, \Sigma) = p((A, b)x | A\mu + b, A\Sigma A')\chi(A, b).$$

Obviously, the group action on the parameter space is

$$(A, b)(\mu, \Sigma) = (A\mu + b, A\Sigma A')$$

and

$$(A, b)P_{\mu,\Sigma} = P_{(A, B)(\mu, \Sigma)}; \qquad P_{\mu,\Sigma} \in \mathcal{P}.$$

Now, let $X_1,\ldots, X_n$, $n \geqslant 3$, be a random sample with $\mathcal{L}(X_i) \in \mathcal{P}$ so the probability model for the random sample is Al_2-invariant by Proposition 7.9. Consider $\bar{X} = (1/n)\Sigma_1^n X_i$ and $S = \Sigma_1^n(X_i - \bar{X})(X_i - \bar{X})'$ so $\bar{X}$ is the sample mean and S is the sample covariance matrix (not normalized). Obviously, $S = S(X_1,\ldots, X_n)$ is a function of $X_1,\ldots, X_n$ and

$$S(AX_1 + b,\ldots, AX_n + b) = AS(X_1,\ldots, X_n)A'.$$

That is, S is an equivariant function on $(R^2)^n$ to $\mathbb{S}_2^+$ where the group action on $\mathbb{S}_2^+$ is

$$(A, b)(S) = ASA'.$$

Writing $S \in \mathbb{S}_2^+$ as

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

the sample correlation coefficient is

$$r = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}.$$

Also, the population correlation coefficient is

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

when the distribution under consideration is $P_{\mu,\Sigma}$, and

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

Now, given that the random sample is from $P_{\mu,\Sigma}$, the question is: how does the distribution of r depend on (μ, Σ)? To show that the distribution of r depends only on ρ, we use an invariance argument. Let G be the subgroup of Al_2 defined by

$$G = \left\{(A, b) | (A, b) \in Al_2, A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, a_{ii} > 0, i = 1,2\right\}.$$

For $(A, b) \in G$, a bit of calculation shows that $r = r(X_1,\ldots, X_n) =$

$r(AX_1 + b, \ldots, AX_n + b)$ so r is a G-invariant function of $X_1, \ldots, X_n$. By Proposition 7.14, the distribution of r, say $Q_{\mu,\Sigma}$, satisfies

$$Q_{\mu,\Sigma} = Q_{(A,b)(\mu,\Sigma)}, \qquad (A,b) \in G.$$

Thus $Q_{\mu,\Sigma}$ depends on μ, Σ only through a maximal invariant function on the parameter space $R^2 \times \mathbb{S}_2^+$ under the action of G. Of course, the action of G is

$$(A,b)(\mu,\Sigma) = (A\mu + b, A\Sigma A'), \qquad (A,b) \in G.$$

We now claim that

$$\rho = \rho(\mu,\Sigma) = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

is a maximal G-invariant function. To see this, consider $(\mu, \Sigma) \in R^2 \times \mathbb{S}_2^+$. By choosing

$$A = \begin{pmatrix} \sigma_{11}^{-1/2} & 0 \\ 0 & \sigma_{22}^{-1/2} \end{pmatrix}$$

and $b = -A\mu$, $(A,b) \in G$ and

$$(A,b)(\mu,\Sigma) = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

so this point is in the orbit of (μ, Σ) and an orbit index is ρ. Thus ρ is maximal invariant and the distribution of r depends only on (μ, Σ) through the maximal invariant function ρ. Obviously, the distribution of r also depends on the function f_0, but f_0 was considered fixed in this discussion. ◆

Proposition 7.14 asserts that the distribution of an invariant function depends only on a maximal invariant parameter, but this result is not especially useful if the exact distribution of an invariant function is desired. The remainder of the section is concerned with using invariance arguments, when G is compact, to derive distributions of maximal invariants and to characterize the G-invariant distributions.

First, we consider the distribution of a maximal invariant function when a compact topological group G acts measurably on a space $(\mathfrak{X}, \mathfrak{B})$. Suppose that μ_0 is a σ-finite G-invariant measure on $(\mathfrak{X}, \mathfrak{B})$ and f is a density with respect to μ_0. Let τ be a measurable mapping from $(\mathcal{X}, \mathcal{B})$ onto $(\mathcal{Y}, \mathcal{C})$. Then τ induces a measure on $(\mathcal{Y}, \mathcal{C})$, say ν_0, given by

$$\nu_0(C) = \mu_0(\tau^{-1}(C))$$

and the equation

$$\int_{\mathcal{X}} h(\tau(x))\mu_0(dx) = \int_{\mathcal{Y}} h(y)\nu_0(dy)$$

holds for all integrable functions h on $(\mathcal{Y}, \mathcal{C})$. Since the group G is compact, there exists a unique probability measure, say δ, that is left and right invariant.

Proposition 7.15. Suppose the mapping τ from $(\mathcal{X}, \mathcal{B})$ onto $(\mathcal{Y}, \mathcal{C})$ is maximal invariant under the action of G on $\mathcal{X}$. If $X \in \mathcal{X}$ has density f with respect to μ_0, then the density of $Y = \tau(X)$ with respect to ν_0 is given by

$$q(\tau(x)) = \int_G f(gx)\delta(dg).$$

Proof. First, the integral

$$\int f(gx)\delta(dg)$$

is a G-invariant function of x and thus can be written as a function of the maximal invariant τ. This defines the function q on $\mathcal{Y}$. To show that q is the density of Y, it suffices to show that

$$\mathcal{E}k(Y) = \int_{\mathcal{Y}} k(y)q(y)\nu_0(dy)$$

for all bounded measurable functions k. But

$$\mathcal{E}k(Y) = \mathcal{E}k(\tau(X)) = \int_{\mathcal{X}} k(\tau(x))f(x)\mu_0(dx)$$

$$= \int_{\mathcal{X}} k(\tau(x))f(gx)\mu_0(dx).$$

The last equality holds since μ_0 is G-invariant and τ is G-invariant. Since δ is a probability measure

$$\mathcal{E}k(Y) = \int_G \int_{\mathcal{X}} k(\tau(x)) f(gx) \mu_0(dx) \delta(dg).$$

Using Fubini's Theorem, the definition of q and the relationship between μ_0 and ν_0, we have

$$\mathcal{E}k(Y) = \int_{\mathcal{X}} k(\tau(x)) q(\tau(x)) \mu_0(dx) = \int_{\mathcal{Y}} k(y) q(y) \nu_0(dy). \qquad \square$$

In most situations, the compact group G will be the orthogonal group or some subgroup of the orthogonal group. Concrete applications of Proposition 7.15 involve two separate steps. First, the function q must be calculated by evaluating

$$\int_G f(gx) \delta(dg).$$

Also, given μ_0 and the maximal invariant τ, the measure ν_0 must be found.

◆ **Example 7.13.** Take $\mathcal{X} = R^n$ and let μ_0 be Lebesgue measure. The orthogonal group $\mathcal{O}_n$ acts on R^n and a maximal invariant function is $\tau(x) = \|x\|^2$ so $\mathcal{Y} = [0, \infty)$. If a random vector $X \in R^n$ has a density f with respect to Lebesgue measure, Proposition 7.15 tells us how to find the density of $Y = \|X\|^2$ with respect to the measure ν_0. To find ν_0, consider the particular density

$$f_0(x) = (\sqrt{2\pi})^{-n} \exp[-\tfrac{1}{2}\|x\|^2].$$

Thus $\mathcal{L}(X) = N(0, I_n)$, so $\mathcal{L}(Y) = \chi^2_n$ and the density of Y with respect to Lebesgue measure dy on $[0, \infty)$ is

$$p(y) = \frac{y^{n/2-1} \exp[-\tfrac{1}{2}y]}{2^{n/2}\Gamma(n/2)}.$$

Therefore,

$$p(y)\, dy = q_0(y) \nu_0(dy)$$

where

$$q_0(\tau(x)) = \int_{\mathcal{O}_n} f_0(\Gamma x) \delta(d\Gamma).$$

Since $f_0(\Gamma x) = f_0(x)$, the integration of f_0 over $\mathcal{O}_n$ is trivial and

$$q_0(y) = (\sqrt{2\pi})^{-n}\exp[-\tfrac{1}{2}y].$$

Solving for $\nu_0(dy)$, we have

$$\nu_0(dy) = \frac{(2\pi)^{n/2}}{2^{n/2}\Gamma(n/2)}y^{n/2-1}\,dy = \frac{[\Gamma(\frac{1}{2})]^n}{\Gamma(n/2)}y^{n/2-1}\,dy$$

since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Now that ν_0 has been found, consider a general density f on R^n. Then

$$q(\tau(x)) = \int_{\mathcal{O}_n} f(\Gamma x)\delta(d\Gamma)$$

and $q(y)$ is the density of $Y = \|X\|^2$ with respect to ν_0. When the density f is given by

$$f(x) = h(\|x\|^2),$$

then it is clear that

$$q(y) = h(y), \qquad y \in [0, \infty)$$

so the distribution of Y has been found in this case. The noncentral chi-square distribution of $Y = \|X\|^2$ provides an interesting example where the integration over $\mathcal{O}_n$ is not trivial. Suppose $\mathcal{L}(X) = N(\mu, I_n)$ so

$$\begin{aligned} f(x) &= (\sqrt{2\pi})^{-n}\exp[-\tfrac{1}{2}\|x - \mu\|^2] \\ &= (\sqrt{2\pi})^{-n}\exp[-\tfrac{1}{2}(\|x\|^2 - 2x'\mu + \|\mu\|^2)]. \end{aligned}$$

Thus

$$\begin{aligned} q(\tau(x)) = (\sqrt{2\pi})^{-n}&\exp[-\tfrac{1}{2}\|\mu\|^2]\exp[-\tfrac{1}{2}\|x\|^2] \\ &\times \int_{\mathcal{O}_n}\exp[(\Gamma x)'\mu]\delta(d\Gamma). \end{aligned}$$

Since x and $\|x\|\varepsilon_1$ have the same length, $x = \|x\|\Gamma_1\varepsilon_1$ for some $\Gamma_1 \in \mathcal{O}_n$ where ε_1 is the first standard unit vector in R^n. Similarly,

$\mu = \|\mu\|\Gamma_2\varepsilon_1$ for some $\Gamma_2 \in \mathcal{O}_n$. Setting $\lambda = \|\mu\|^2$,

$$q(y) = (\sqrt{2\pi})^{-n}\exp[-\tfrac{1}{2}\lambda]\exp[-\tfrac{1}{2}y]$$
$$\times \int_{\mathcal{O}_n} \exp\left[\sqrt{y\lambda}\,(\Gamma\Gamma_1\varepsilon_1)'\Gamma_2\varepsilon_1\right]\delta(d\Gamma).$$

Thus to evaluate q, we need to calculate

$$H(u) = \int_{\mathcal{O}_n} \exp[u\varepsilon_1'\Gamma_1'\Gamma'\Gamma_2\varepsilon_1]\delta(d\Gamma).$$

Since δ is left and right invariant,

$$H(u) = \int_{\mathcal{O}_n} \exp[u\varepsilon_1'\Gamma'\varepsilon_1]\delta(d\Gamma) = \int_{\mathcal{O}_n} \exp[u\gamma_{11}]\delta(d\Gamma)$$

where γ_{11} is the $(1,1)$ element of Γ. The representation of the uniform distribution on $\mathcal{O}_n$ given in Proposition 7.2 shows that when Γ is uniform on $\mathcal{O}_n$, then

$$\mathcal{L}(\gamma_{11}) = \mathcal{L}\left(\frac{Z_1}{\|Z\|}\right)$$

where $\mathcal{L}(Z) = N(0, I_n)$ and Z_1 is the first coordinate of Z. Expanding the exponential in a power series, we have

$$H(u) = \sum_{j=0}^{\infty} \frac{1}{j!}\int_{\mathcal{O}_n} u^j\gamma_{11}^j\delta(d\Gamma) = \sum_{j=0}^{\infty} \frac{u^j}{j!}\mathcal{E}\left(\frac{Z_1}{\|Z\|}\right)^j.$$

Thus the moments of $U_1 \equiv Z_1/\|Z\|$ need to be found. Obviously, $\mathcal{L}(U_1) = \mathcal{L}(-U_1)$, so all odd moments of U_1 are zero. Also, $U_1^2 = Z_1^2/(Z_1^2 + \Sigma_2^n Z_i^2)$, which has a beta distribution with parameters $\frac{1}{2}$ and $(n-1)/2$. Therefore,

$$a_j \equiv \mathcal{E}(U_1^2)^j = \frac{\Gamma(n/2)\Gamma(j+\frac{1}{2})}{\Gamma(n/2+j)\Gamma(\frac{1}{2})}$$

so

$$H(u) = \sum_{j=0}^{\infty} \frac{a_j u^{2j}}{(2j)!}.$$

Hence

$$q(y) = (\sqrt{2\pi})^{-n}\exp\left[-\frac{1}{2}\lambda\right]\exp\left[-\frac{1}{2}y\right]\sum_{j=0}^{\infty}\lambda^j y^j\frac{a_j}{(2j)!}$$

is the density of Y with respect to the measure ν_0. A bit of algebra and some manipulation with the gamma function shows that

$$q(y)\nu_0(dy) = \left\{\sum_{j=0}^{\infty}\frac{\exp[-\frac{1}{2}\lambda]}{j!}(\lambda/2)^j h_{n+2j}(y)\right\}dy$$

where

$$h_m(y) = \frac{y^{m/2-1}\exp[-\frac{1}{2}y]}{2^{m/2}\Gamma(m/2)}$$

is the density of a χ_m^2 distribution. This is the expression for the density of the noncentral chi-square distribution discussed in Chapter 3. ◆

◆ **Example 7.14.** In this example, we derive the density function of the order statistic of a random vector $X \in R^n$. Suppose X has a density f with respect to Lebesgue measure and let $X_1, \ldots, X_n$ be the coordinates of X. Consider the space $\mathcal{Y} \subseteq R^n$ defined by

$$\mathcal{Y} = \{y | y \in R^n, y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n\}.$$

The order statistic of X is the random vector $Y \in \mathcal{Y}$ consisting of the ordered values of the coordinates of X. More precisely, Y_1 is the smallest coordinate of X, Y_2 is the next smallest coordinate of X, and so on. Thus $Y = \tau(X)$ where τ maps each $x \in R^n$ into the ordered coordinates of x—say $\tau(x) \in \mathcal{Y}$. To derive the density function of Y, we show that Y is a maximal invariant under a compact group operating on R^n and then apply Proposition 7.15. Let G be the group of all one-to-one onto functions from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$—that is, G is the permutation group of $\{1, 2, \ldots, n\}$. Of course, the group operation is function composition, the group inverse is function inverse, and G has $n!$ elements. The group G acts on the left of R^n in the following way. For $x \in R^n$ and $\pi \in G$, define $\pi x \in R^n$ to have ith coordinate $x(\pi^{-1}(i))$. Thus the ith coordinate of πx is the $\pi^{-1}(i)$ coordinate of x, so

$$(\pi x)(i) \equiv x(\pi^{-1}(i)).$$

The reason for the inverse on π in this definition is so that G acts on the left of R^n—that is,

$$(\pi_1\pi_2)x = \pi_1(\pi_2 x).$$

It is routine to verify that the function τ on $\mathcal{X}$ to $\mathcal{Y}$ is a maximal invariant under the action of G on R^n. Also, Lebesgue measure, say l, is invariant so Proposition 7.15 is applicable as G is a finite group and hence compact. Obviously, the density q of $Y = \tau(X)$ is

$$q(\tau(x)) = \frac{1}{n!}\sum_{\pi} f(\pi x) = \frac{1}{n!}\sum_{\pi} f(\pi\tau(x))$$

so

$$q(y) = \frac{1}{n!}\sum_{\pi} f(\pi y)$$

for $y \in \mathcal{Y}$. To derive the measure ν_0 on $\mathcal{Y}$, consider a measurable subset $C \subseteq \mathcal{Y}$. Then

$$\tau^{-1}(C) = \bigcup_{\pi\in G}(\pi C)$$

and

$$\nu_0(C) = l(\tau^{-1}(C)) = l\Big(\bigcup_{\pi\in G}(\pi C)\Big) = \sum_{\pi\in G} l(\pi(C)) = n!\,l(C).$$

The third equality follows since $(\pi_1 C) \cap (\pi_2 C)$ has Lebesgue measure zero for $\pi_1 \neq \pi_2$ as the boundary of $\mathcal{Y}$ in R^n has Lebesgue measure zero. Thus ν_0 is just $n!$ times l restricted to $\mathcal{Y}$. Therefore, the density of the order statistic Y, with respect to ν_0 restricted to $\mathcal{Y}$, is

$$q(y) = \frac{1}{n!}\sum_{\pi} f(\pi y).$$

When f is invariant under permutations, as is the case when $X_1, \ldots, X_n$ are independent and identically distributed, we have

$$q(y) = f(y), \qquad y \in \mathcal{Y}. \qquad \blacklozenge$$

The next example is an extension of Example 7.13 and is related to the results in Proposition 7.6.

◆ **Example 7.15.** Suppose X is a random vector in $\mathcal{L}_{p,n}$, $n \geqslant p$, which has a density f with respect to Lebesgue measure dx on $\mathcal{L}_{p,n}$. Let τ map $\mathcal{L}_{p,n}$ onto the space of $p \times p$ positive semidefinite matrices, say $\bar{\mathcal{S}}_p^+$, by $\tau(x) = x'x$. The problem in this example is to derive the density of $S = \tau(X) = X'X$. The compact group $\mathcal{O}_n$ acts on $\mathcal{L}_{p,n}$ and a group element $\Gamma \in \mathcal{O}_n$ sends x into Γx. It follows immediately from Proposition 1.31 that τ is a maximal invariant function under the action of $\mathcal{O}_n$ on $\mathcal{L}_{p,n}$. Since dx is invariant under $\mathcal{O}_n$, Proposition 7.15 shows that the density of S is

$$q(\tau(x)) = \int_{\mathcal{O}_n} f(\Gamma x)\mu(d\Gamma)$$

with respect to the measure ν_0 on $\bar{\mathcal{S}}_p^+$ induced by dx and τ. To find the measure ν_0, we argue as in Example 7.13. Consider the particular density

$$f_0(x) = (\sqrt{2\pi})^{-np}\exp[-\tfrac{1}{2}\operatorname{tr}(x'x)]$$

on $\mathcal{L}_{p,n}$ so $\mathcal{L}(X) = N(0, I_n \otimes I_p)$. For this f_0, the density of S is

$$q_0(S) = q_0(\tau(x)) = \int_{\mathcal{O}_n} f_0(\Gamma x)\mu(d\Gamma) = (\sqrt{2\pi})^{-np}\exp[-\tfrac{1}{2}\operatorname{tr}(S)]$$

with respect to ν_0. However, by Propostion 7.6, the density of S with respect to $dS/|S|^{(p+1)/2}$ is

$$q_1(S) = \omega(n, p)|S|^{n/2}\exp[-\tfrac{1}{2}\operatorname{tr}(S)].$$

Therefore,

$$q_1(S)\frac{dS}{|S|^{(p+1)/2}} = q_0(S)\nu_0(dS)$$

so

$$\begin{aligned}\omega(n, p)|S|^{(n-p-1)/2}&\exp[-\tfrac{1}{2}\operatorname{tr}(S)]\, dS \\ &= (\sqrt{2\pi})^{-np}\exp[-\tfrac{1}{2}\operatorname{tr}(S)]\nu_0(dS),\end{aligned}$$

which shows that

$$\nu_0(dS) = (\sqrt{2\pi})^{np}\omega(n, p)|S|^{(n-p-1)/2}\, dS.$$

In the above argument, we have ignored the set of Lebesgue measure zero where $x \in \mathcal{L}_{p,n}$ has rank less than p. The justification for this is left to the reader. Now that ν_0 has been found, the density of S for a general density f is obtained by calculating

$$q(\tau(x)) = \int_{\mathcal{O}_n} f(\Gamma x)\mu(d\Gamma).$$

When $f(x) = h(x'x)$, then $f(\Gamma x) = h(x'x) = h(\tau(x))$ and $q(S) = h(S)$. In this case, the integration over $\mathcal{O}_n$ is trivial. Another example where the integration over $\mathcal{O}_n$ is not trivial is given in the next chapter when we discuss the noncentral Wishart distribution. ◆

As motivation for the next result of this section, consider the situation discussed in Proposition 7.3. This result gives a characterization of the $\mathcal{O}_n$-left invariant distributions by representing each of these distributions as a product measure where one measure is a fixed $\mathcal{O}_n$-invariant distribution and the other measure is arbitrary. The decomposition of the space $\mathcal{X}$ into the product space $\mathcal{F}_{p,n} \times G_U^+$ provided the framework in which to state this representation of $\mathcal{O}_n$-left invariant distributions. In some situations, this product space structure is not available (see Example 7.5) but a product measure representation for $\mathcal{O}_n$-invariant distributions can be obtained. It is established below that, under some mild regularity conditions, such a representation can be given for probability measures that are invariant under any compact topological group that acts on the sample space. We now turn to the technical details.

In what follows, G is a compact topological group that acts measurably on a measure space $(\mathcal{X}, \mathcal{B})$ and P is a G-invariant probability measure on $(\mathcal{X}, \mathcal{B})$. The unique invariant probability measure on G is denoted by μ and the symbol $U \in G$ denotes a random variable with values in G and distribution μ. The σ-algebra for G is the Borel σ-algebra of open sets so U is a measurable function defined on some probability space with induced distribution μ. Since G acts on $\mathcal{X}$, $\mathcal{X}$ can be written as a disjoint union of orbits, say

$$\mathcal{X} = \bigcup_{\alpha \in \mathcal{A}} \mathcal{X}_\alpha$$

where $\mathcal{A}$ is an index set for the orbits and $\mathcal{X}_\alpha \cap \mathcal{X}_{\alpha'} = \phi$ if $\alpha \neq \alpha'$. Let x_α be a fixed element of $\mathcal{X}_\alpha = \{gx_\alpha | g \in G\}$. Also, set

$$\mathcal{Y} = \{x_\alpha | \alpha \in A\} \subseteq \mathcal{X}$$

and assume that $\mathcal{Y}$ is a measurable subset of $\mathcal{X}$. The function τ defined on $\mathcal{X}$ to $\mathcal{Y}$ by

$$\tau(x) = x_\alpha \qquad \text{if } x \in \mathcal{X}_\alpha$$

is obviously a maximal invariant function under the action of G on $\mathcal{X}$. It is assumed that τ is a measurable function from $\mathcal{X}$ to $\mathcal{Y}$ where $\mathcal{Y}$ has the σ-algebra inherited from $\mathcal{X}$. A subset $B_1 \subseteq \mathcal{Y}$ is measurable iff $B_1 = \mathcal{Y} \cap B$ for some $B \in \mathcal{B}$. If $X \in \mathcal{X}$ has distribution P, then the maximal invariant $Y = \tau(X)$ has the induced distribution Q defined by

$$Q(B_1) = P(\tau^{-1}(B_1))$$

for measurable subsets $B_1 \subseteq \mathcal{Y}$. What we would like to show is that P is represented by the product measure $\mu \times Q$ on $G \times \mathcal{Y}$ in the following sense. If $Y \in \mathcal{Y}$ has the distribution Q and is independent of $U \in G$, then the random variable $Z = UY \in \mathcal{X}$ has the distribution P. In other words, $\mathcal{L}(X) = \mathcal{L}(UY)$ where U and Y are independent. Here, UY means the group element U operating on the point $Y \in \mathcal{X}$. The intuitive argument that suggests this representation is the following. The distribution of X, conditional on $\tau(X) = x_\alpha$, should be G-invariant on $\mathcal{X}_\alpha$ as the distribution of X is G-invariant. But G acts transitively on $\mathcal{X}_\alpha$ and, since G is compact, there should be a unique invariant probability distribution on $\mathcal{X}_\alpha$ that is induced by μ on G. In other words, conditional on $\tau(X) = x_\alpha$, X should have the same distribution as Ux_α where U is "uniform" on G. The next result makes all of this precise.

Proposition 7.16. Consider $\mathcal{X}$, $\mathcal{Y}$, and G to be as above with their respective σ-algebras. Assume that the mapping h on $G \times \mathcal{Y}$ to $\mathcal{X}$ given by $h(g, y) = gy$ is measurable.

(i) If $U \in G$ and $Y \in \mathcal{Y}$ are independent with $\mathcal{L}(U) = \mu$ and $\mathcal{L}(Y) = Q$, then the distribution of $X = UY$ is a G-invariant distribution on $\mathcal{X}$.

(ii) If $X \in \mathcal{X}$ has a G-invariant distribution, say P, let the maximal invariant $Y = \tau(X)$ have an induced distribution Q on $\mathcal{Y}$. Let $U \in G$ have the distribution μ and be independent of X. Then $\mathcal{L}(X) = \mathcal{L}(UY)$.

Proof. For the proof of (i), it suffices to show that

$$\mathcal{E}f(X) = \mathcal{E}f(gX)$$

for all integrable functions f and all $g \in G$. But

$$\mathcal{E}f(gX) = \mathcal{E}f(g(UY)) = \mathcal{E}f((gU)Y) = \mathcal{E}\mathcal{E}[f((gU)Y)|Y]$$
$$= \mathcal{E}\mathcal{E}[f(UY)|Y] = \mathcal{E}f(UY) = \mathcal{E}f(X).$$

In the above calculation, we have used the assumption that U and Y are independent, so conditional on Y, $\mathcal{L}(U) = \mathcal{L}(gU)$ for $g \in G$.

To prove (ii) it suffices to show that

$$\mathcal{E}f(X) = \mathcal{E}f(UY)$$

for all integrable f. Since the distribution of X is G-invariant

$$\mathcal{E}f(X) = \mathcal{E}f(gX), \quad g \in G.$$

Therefore,

$$\mathcal{E}f(X) = \mathcal{E}_U\mathcal{E}_X f(UX),$$

as U and X are independent. Thus

$$\int_{\mathfrak{X}} f(x)P(dx) = \int_G\int_{\mathfrak{X}} f(gx)P(dx)\mu(dg) = \int_{\mathfrak{X}}\int_G f(gx)\mu(dg)P(dx).$$

However, for $x \in \mathfrak{X}_\alpha$, there exists an element $k \in G$ such that $x = kx_\alpha$. Using the definition of τ and the right invariance of μ, we have

$$\int_G f(gx)\mu(dg) = \int_G f(gkx_\alpha)\mu(dg) = \int_G f(gx_\alpha)\mu(dg)$$
$$= \int_G f(g\tau(x))\mu(dg).$$

Hence

$$\int_{\mathfrak{X}} f(x)P(dx) = \int_{\mathfrak{X}}\int_G f(g\tau(x))\mu(dg)P(dx) = \int_{\mathfrak{Y}}\int_G f(gy)\mu(dg)Q(dy)$$

where the second equality follows from the definition of the induced measure Q. In terms of the random variables,

$$\mathcal{E}f(X) = \mathcal{E}_U\mathcal{E}_Y f(UY)$$

where U and Y are independent as U and X are independent. $\square$

The technical advantage of Proposition 7.16 over the method discussed in Section 7.1 is that the space $\mathfrak{X}$ is not assumed to be in one-to-one correspondence with the product space $G \times \mathfrak{Y}$. Obviously, the mapping h on $G \times \mathfrak{Y}$ to $\mathfrak{X}$ is onto, but h will ordinarily not be one-to-one.

◆ **Example 7.16.** In this example, take $\mathfrak{X} = \mathbb{S}_p$, the set of all $p \times p$ symmetric matrices. The group $G = \mathcal{O}_p$ acts on $\mathbb{S}_p$ by

$$\Gamma(S) = \Gamma S \Gamma', \quad S \in \mathbb{S}_p, \quad \Gamma \in \mathcal{O}_p.$$

For $S \in \mathbb{S}_p$, let

$$\tau(S) = Y = \begin{pmatrix} y_1 & & & \\ & y_2 & & \\ & & \ddots & \\ & & & y_p \end{pmatrix} \in \mathbb{S}_p$$

where $y_1 \geqslant \cdots \geqslant y_p$ are ordered eigenvalues of S and the off-diagonal elements of Y are zero. Also, let $\mathfrak{Y} = \{Y | Y = \tau(S),\ S \in \mathbb{S}_p\}$. The spectral theorem shows that τ is a maximal invariant function under the action of $\mathcal{O}_p$ and the elements of $\mathfrak{Y}$ index the orbits in $\mathbb{S}_p$. The measurability assumptions of Proposition 7.16 are easily verified, so every $\mathcal{O}_n$-invariant distribution on $\mathbb{S}_p$, say P, has the representation given by

$$\int_{\mathbb{S}_p} f(S) P(dS) = \int_{\mathcal{O}_p} \int_{\mathfrak{Y}} f(\Gamma Y \Gamma') Q(dY) \mu(d\Gamma)$$

where μ is the uniform distribution on $\mathcal{O}_p$ and Q is the induced distribution of Y. In terms of random variables, if $\mathcal{L}(S) = P$ and $\mathcal{L}(\Gamma S \Gamma') = \mathcal{L}(S)$ for all $\Gamma \in \mathcal{O}_p$, then

$$\mathcal{L}(S) = \mathcal{L}(\Psi \tau(S) \Psi')$$

where Ψ is uniform on $\mathcal{O}_p$ and is independent of the matrix of eigenvalues of S. As a particular case, consider the probability measure P_0 on $\mathbb{S}_p^+ \subseteq \mathbb{S}_p$ with the Wishart density

$$p_0(S) = \omega(p, n)|S|^{(n-p-1)/2} \exp[-\tfrac{1}{2} \operatorname{tr} S] I(S)$$

where $n \geqslant p$, $I(S) = 1$ if $S \in \mathbb{S}_p^+$ and is zero otherwise. That p_0 is a density on $\mathbb{S}_p$ with respect to Lebesgue measure dS on $\mathbb{S}_p$ follows from Example 5.1. Also, p_0 is $\mathcal{O}_p$-invariant since dS is $\mathcal{O}_p$-invariant and $p_0(\Gamma S\Gamma') = p_0(S)$ for all $S \in \mathbb{S}_p$ and $\Gamma \in \mathcal{O}_p$. Thus the above results are applicable to this particular Wishart distribution. ◆

The final example of this section deals with the singular value decomposition of a random $n \times p$ matrix.

◆ **Example 7.17.** The compact group $\mathcal{O}_n \times \mathcal{O}_p$ acts on the space $\mathcal{L}_{p,n}$ by

$$(\Gamma, \Delta)X \equiv \Gamma X\Delta'; \qquad (\Gamma, \Delta) \in \mathcal{O}_n \times \mathcal{O}_p, \quad X \in \mathcal{L}_{p,n}.$$

For definiteness, we take $p \leqslant n$. Define τ on $\mathcal{L}_{p,n}$ by

$$\tau(X) = \left(\begin{array}{cccc} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \\ \hline & & 0 & \end{array} \right)$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_p \geqslant 0$ and $\lambda_1^2, \ldots, \lambda_p^2$ are the ordered eigenvalues of $X'X$. Let $\mathcal{Y} \subseteq \mathcal{L}_{p,n}$ be the range of τ so $\mathcal{Y}$ is a closed subset of $\mathcal{L}_{p,n}$. It is clear that $\tau(\Gamma X\Delta') = \tau(X)$ for $\Gamma \in \mathcal{O}_n$ and $\Delta \in \mathcal{O}_p$ so τ is invariant. That τ is a maximal invariant follows easily from the singular value decomposition theorem. Thus the elements of $\mathcal{Y}$ index the orbits in $\mathcal{L}_{p,n}$ and every $X \in \mathcal{L}_{p,n}$ can be written as

$$X = \Gamma y\Delta' = (\Gamma, \Delta)y$$

for some $y \in \mathcal{Y}$ and $(\Gamma, \Delta) \in \mathcal{O}_n \times \mathcal{O}_p$. The measurability assumptions of Proposition 7.16 are easily checked. Thus if P is an $(\mathcal{O}_n \times \mathcal{O}_p)$-invariant probability measure on $\mathcal{L}_{p,n}$ and $\mathcal{L}(X) = P$, then

$$\mathcal{L}(X) = \mathcal{L}(\Gamma Y\Delta')$$

where (Γ, Δ) has a uniform distribution on $\mathcal{O}_n \times \mathcal{O}_p$, Y has a distribution Q induced by τ and P, and Y and (Γ, Δ) are independent. However, we can say a bit more. Since $\mathcal{O}_n \times \mathcal{O}_p$ is a product group, the unique invariant probability measure on $\mathcal{O}_n \times \mathcal{O}_p$ is the product measure $\mu_1 \times \mu_2$ where $\mu_1(\mu_2)$ is the unique invariant probability measure on $\mathcal{O}_n(\mathcal{O}_p)$. Thus Γ and Δ are independent and each is uniform in its respective group. In summary,

$$\mathcal{L}(X) = \mathcal{L}(\Gamma Y \Delta').$$

where Γ, Y, and Δ are mutually independent with the distributions given above. As a particular case, consider the density

$$f_0(X) = (\sqrt{2\pi})^{-np} \exp[-\tfrac{1}{2}\operatorname{tr}(X'X)]$$

with respect to Lebesgue measure on $\mathcal{L}_{p,n}$. Since $f_0(\Gamma X \Delta') = f_0(X)$ and Lebesgue measure is $(\mathcal{O}_n \times \mathcal{O}_p)$-invariant, the probability measure defined by f_0 is $(\mathcal{O}_n \times \mathcal{O}_p)$-invariant. Therefore, when $\mathcal{L}(X) = N(0, I_n \times I_p)$, X has the same distribution as $\Gamma Y \Delta'$ where Γ and Δ are uniform and Y has the induced distribution Q on $\mathcal{Y}$. ◆

7.6. INDEPENDENCE AND INVARIANCE

Considerations that imply the stochastic independence of an invariant function and an equivariant function are the subject of this section. To motivate the abstract discussion to follow, we begin with the familiar random sample from a univariate normal distribution. Consider $X \in \mathcal{X}$ with $\mathcal{L}(X) = N(\mu e, \sigma^2 I_n)$ where $\mu \in R$, $\sigma^2 > 0$, and e is the vector of ones in R^n. The set $\mathcal{X}$ is $R^n - \operatorname{span}\{e\}$ and the reason for choosing this as the sample space is to guarantee that $\Sigma_1^n (x_i - \bar{x})^2 > 0$ for $x \in \mathcal{X}$. The coordinates of X, say $X_1, \ldots, X_n$, are independent and $\mathcal{L}(X_i) = N(\mu, \sigma^2)$ for $i = 1, \ldots, n$. When μ and σ^2 are unknown parameters, the statistic $t(X) = (s, \bar{X})$ where

$$\bar{X} = \frac{1}{n} \sum_1^n X_i, \qquad s^2 = \sum_1^n (X_i - \bar{X})^2$$

is minimal sufficient and complete. The reason for using s rather than s^2 in the definition of $t(X)$ is based on invariance considerations. The affine group Al_1 acts on $\mathcal{X}$ by

$$(a, b)x \equiv ax + be$$

for $(a, b) \in Al_1$. Let G be the subgroup of Al_1 given by $G = \{(a, b) | (a, b) \in Al_1,\ a > 0\}$ so G also acts on $\mathcal{X}$.

The probability model for $X \in \mathcal{X}$ is generated by G in the sense that if $Z \in \mathcal{X}$ and $\mathcal{L}(Z) = N(0, I_n)$,

$$\mathcal{L}((a, b)Z) = \mathcal{L}(aZ + be) = N(be, a^2 I_n).$$

Thus the set of distributions $\mathcal{P} = \{N(\mu e, \sigma^2 I_n) | \mu \in R, \sigma^2 > 0\}$ is obtained from an $N(0, I_n)$ distribution by a group operation. For this example, the group G serves as a parameter space for $\mathcal{P}$. Further, the statistic t takes its values in G and satisfies

$$t((a, b)X) = (a, b)(s, \overline{X}),$$

that is, t evaluated at $(a, b)X = aX + be$ is the same as the group element (a, b) composed with the group element $(s, \overline{X})$. Thus t is an equivariant function defined on $\mathcal{X}$ to G and G acts on both $\mathcal{X}$ and G. Now, which functions of X, say $h(X)$, might be independent of $t(X)$? Intuitively, since $t(X)$ is sufficient, $t(X)$ "contains all the information in X about the parameters." Thus if $h(X)$ has a distribution that does not depend on the parameter value (such an $h(X)$ will be called *ancillary*), there is some reason to believe that $h(X)$ and $t(X)$ might be independent. However, the group structure given above provides a method for constructing ancillary statistics. If h is an invariant function of X, then the distribution of h is an invariant function of the parameter (μ, σ^2). But the group G acts transitively on the parameter space (i.e., G), so any invariant function will be ancillary. Also, h is invariant iff h is a function of a maximal invariant statistic. This suggests that a maximal invariant statistic will be independent of $t(X)$. Consider the statistic

$$Z(X) = (t(X))^{-1}X = \frac{X - \overline{X}e}{s},$$

where the inverse on $t(X)$ denotes the group inverse in G. The verification that $Z(X)$ is maximal invariant partially justifies choosing t to have values in G. For $(a, b) \in G$,

$$\begin{aligned} Z((a, b)X) &= (t((a, b)X))^{-1}(a, b)X = ((a, b)t(X))^{-1}(a, b)X \\ &= (t(X))^{-1}(a, b)^{-1}(a, b)X = (t(X))^{-1}X = Z(X), \end{aligned}$$

so Z is invariant. Also, if

$$(t(X))^{-1}X = Z(X) = Z(Y) = (t(Y))^{-1}Y,$$

then

$$Y = \left[t(Y)(t(X))^{-1}\right]X,$$

so X and Y are in the same orbit. Thus Z is maximal invariant and is an ancillary statistic. That $Z(X)$ and $t(X)$ are stochastically independent for each value of μ and σ^2 follows from Basu's Theorem given in the Appendix. The whole purpose of this discussion was to show that sufficiency coupled with the invariance suggested the independence of $Z(X)$ and $t(X)$. The role of the equivariance of t is not completely clear, but it is essential in the more abstract treatment that follows.

Let P_0 be a fixed probability on $(\mathfrak{X}, \mathfrak{B})$ and suppose that G is a group that acts measurably on $(\mathfrak{X}, \mathfrak{B})$. Consider a measurable function t on $(\mathfrak{X}, \mathfrak{B})$ to $(\mathfrak{Y}, \mathcal{C}_1)$ and assume that $\bar{G}$ is a homomorphic image of G that acts transitively on $(\mathfrak{Y}, \mathcal{C}_1)$ and that

$$t(gx) = \bar{g}t(x); \qquad x \in \mathfrak{X}, \quad g \in G.$$

Thus t is an equivariant function. For technical reasons that become apparent later, it is assumed that $\bar{G}$ is a locally compact and σ-compact topological group endowed with the Borel σ-algebra. Also, the mapping $(\bar{g}, y) \to \bar{g}y$ from $\bar{G} \times \mathfrak{Y}$ to $\mathfrak{Y}$ is assumed to be jointly measurable.

Now, let h be a measurable function on $(\mathfrak{X}, \mathfrak{B})$ to $(\mathfrak{Z}, \mathcal{C}_2)$, which is G-invariant. If $X \in \mathfrak{X}$ and $\mathcal{L}(X) = P_0$, we want to find conditions under which $Y \equiv t(X)$ and $Z \equiv h(X)$ are stochastically independent. The following informal argument, which is made precise later, suggests the conditions needed. To show that Y and Z are independent, it is sufficient to verify that, for all bounded measurable functions f on $(\mathfrak{Z}, \mathcal{C}_2)$,

$$H(y) = \mathcal{E}_{P_0}(f(h(X))|t(X) = y)$$

is constant for $y \in \mathfrak{Y}$. That this condition is sufficient follows by integrating H with respect to the induced distribution of Y, say Q_0. More precisely, if k is a bounded function on $(\mathfrak{Y}, \mathcal{C}_1)$ and $H(y) = H(y_0)$ for $y \in \mathfrak{Y}$, then

$$\begin{aligned}
\mathcal{E}_{P_0}[k(t(X))f(h(X))] &= \int \mathcal{E}_{P_0}[k(t(X))f(h(X))|t(X) = y]Q_0(dy) \\
&= \int k(y)\mathcal{E}_{P_0}[f(h(X))|t(X) = y]Q_0(dy) \\
&= \int k(y)H(y)Q_0(dy) = H(y_0)\int k(y)Q_0(dy) \\
&= \mathcal{E}_{P_0}f(h(X))\mathcal{E}_{P_0}k(t(X)),
\end{aligned}$$

and this implies independence. The assumption that H is constant justifies the next to the last equality while the last equality follows from

$$H(y_0) = \int H(y) Q_0(dy) = \mathcal{E}_{P_0} f(h(X))$$

when H is constant. Thus under what conditions will H be constant? Since $\bar{G}$ acts transitively on $\mathcal{Y}$, if H is $\bar{G}$-invariant, then H must be constant and conversely. However,

$$\begin{aligned} H(\bar{g}^{-1}y) &= \mathcal{E}_{P_0}[f(h(X))|t(X) = \bar{g}^{-1}y] = \mathcal{E}_{P_0}[f(h(X))|\bar{g}t(X) = y] \\ &= \mathcal{E}_{P_0}[f(h(X))|t(gX) = y] = \mathcal{E}_{P_0}[f(h(gX))|t(gX) = y] \\ &= \mathcal{E}_{gP_0}[f(h(X))|t(X) = y]. \end{aligned}$$

The equivariance of t and the invariance of h justify the third and fourth equalities while the last equality is a consequence of $\mathcal{L}(gX) = gP_0$ when $\mathcal{L}(X) = P_0$. Now, if $t(X)$ is a sufficient statistic for the family $\mathcal{P} = \{gP_0 | g \in G\}$, then the last member of the above string of equalities is just $H(y)$. Under this sufficiency assumption, $H(y) = H(\bar{g}^{-1}y)$ so H is invariant and hence is a constant. The technical problem with this argument is caused by the nonuniqueness of conditional expectations. The conclusion that $H(y) = H(\bar{g}^{-1}y)$ should really be $H(y) = H(\bar{g}^{-1}y)$ except for $y \in N_g$ where N_g is a set of Q_0 measure zero. Since this null set can depend on g, even the conclusion that H is a constant a.e. (Q_0) is not justified without some further work. Once these technical problems are overcome, we prove that, if $t(X)$ is sufficient for $\{gP_0 | g \in G\}$, then for each $g \in G$, $h(X)$ and $t(X)$ are stochastically independent when $\mathcal{L}(X) = gP_0$.

The first gap to fill concerns almost invariant functions.

Definition 7.6. Let $(\mathcal{X}_1, \mathcal{B}_1)$ be a measurable space that is acted on measurably by a group G_1. If μ is a σ-finite measure on $(\mathcal{X}_1, \mathcal{B}_1)$ and f is a real-valued Borel measurable function, f is *almost G_1-invariant* if for each $g \in G_1$, the set $N_g = \{x | f(x) \neq f(gx)\}$ has μ measure zero.

The following result shows that under certain conditions, an almost G_1-invariant function is equal a.e. (μ) to a G_1-invariant function.

Proposition 7.17. Suppose that G_1 acts measurably on $(\mathcal{X}_1, \mathcal{B}_1)$ and that G_1 is a locally compact and σ-compact topological group with the Borel σ-algebra. Assume that the mapping $(g, x) \to gx$ from $G_1 \times \mathcal{X}_1$ to $\mathcal{X}_1$ is measurable. If μ is a σ-finite measure on $(\mathcal{X}_1, \mathcal{B}_1)$ and f is a bounded almost G_1-invariant function, then there exists a measurable invariant function f_1 such that $f = f_1$ a.e. (μ).

Proof. This follows from Theorem 4, p. 227 of Lehmann (1959) and the proof is not repeated here. □

The next technical problem has to do with conditional expectations.

Proposition 7.18. In the notation introduced earlier, suppose $(\mathcal{X}, \mathcal{B})$ and $(\mathcal{Y}, \mathcal{C}_1)$ are measurable spaces acted on by groups G and $\bar{G}$ where $\bar{G}$ is a homomorphic image of G. Assume that τ is an equivariant function from $\mathcal{X}$ to $\mathcal{Y}$. Let P_0 be a probability measure on $(\mathcal{X}, \mathcal{B})$ and let Q_0 be the induced distribution of $\tau(X)$ when $\mathcal{L}(X) = P_0$. If f is a bounded G-invariant function on $\mathcal{X}$, let

$$H(y) \equiv \mathcal{E}_{P_0}(f(X)|\tau(X) = y),$$

and

$$H_1(y) = \mathcal{E}_{gP_0}(f(X)|\tau(X) = y).$$

Then $H_1(\bar{g}y) = H(y)$ a.e. (Q_0) for each fixed $\bar{g} \in G$.

Proof. The conditional expectations are well defined since f is bounded. $H(y)$ is the unique a.e. (Q_0) function that satisfies the equation

$$\int_{\mathcal{Y}} k(y)H(y)Q_0(dy) = \int_{\mathcal{X}} k(\tau(x))f(x)P_0(dx)$$

for all bounded measurable k. The probability measure gP_0 satisfies the equation

$$\int_{\mathcal{X}} f_1(x)(gP_0)(dx) = \int_{\mathcal{X}} f_1(gx)P_0(dx)$$

for all bounded f_1. Since τ is equivariant, this implies that if $\mathcal{L}(X) = gP_0$, then $\mathcal{L}(\tau(X)) = \bar{g}Q_0$. Using this, the invariance of f, and the characterizing property of conditional expectation, we have for all bounded k,

$$\int_{\mathcal{Y}} H(y)k(y)Q_0(dy) = \int_{\mathcal{X}} f(x)k(\tau(x))P_0(dx)$$
$$= \int f(gx)k(\bar{g}^{-1}\tau(gx))P_0(dx)$$
$$= \int f(x)k(\bar{g}^{-1}\tau(x))(gP_0)(dx)$$
$$= \int H_1(y)k(\bar{g}^{-1}y)(\bar{g}Q_0)(dy)$$
$$= \int H_1(\bar{g}y)k(\bar{g}^{-1}\bar{g}y)Q_0(dy)$$
$$= \int H_1(\bar{g}y)k(y)Q_0(dy).$$

Since the first and the last terms in this equality are equal for all bounded k, we have that $H(y) = H_1(\bar{g}y)$ a.e. (Q_0). □

With the technical problems out of the way, the main result of this section can be proved.

Proposition 7.19. Consider measurable spaces $(\mathcal{X}, \mathcal{B})$ and $(\mathcal{Y}, \mathcal{C}_1)$, which are acted on measurably by groups G and $\bar{G}$ where $\bar{G}$ is a homomorphic image of G. It is assumed that the conditions of Proposition 7.17 hold for the group $\bar{G}$ and the space $(\mathcal{Y}, \mathcal{C}_1)$, and that $\bar{G}$ acts transitively on $\mathcal{Y}$. Let τ on $\mathcal{X}$ to $\mathcal{Y}$ be measurable and equivariant. Also let $(\mathcal{Z}, \mathcal{C}_2)$ be a measurable space and let h be a G-invariant measurable function from $\mathcal{X}$ to $\mathcal{Z}$. For a random variable $X \in \mathcal{X}$ with $\mathcal{L}(X) = P_0$, set $Y = \tau(X)$ and $Z = h(X)$ and assume that $\tau(X)$ is a sufficient statistic for the family $\{gP_0 | g \in G\}$ of distributions on $(\mathcal{X}, \mathcal{B})$. Under these assumptions, Y and Z are independent when $\mathcal{L}(X) = gP_0$, $g \in G$.

Proof. First we prove that Y and Z are independent when $\mathcal{L}(X) = P_0$. Fix a bounded measurable function f on Z and let

$$H_g(y) = \mathcal{E}_{gP_0}(f(h(X)) | \tau(X) = y).$$

Since $\tau(X)$ is a sufficient statistic, there is a measurable function H on $\mathcal{Y}$ such that

$$H_g(y) = H(y) \qquad \text{for } y \notin N_g$$

where N_g is a set of $\bar{g}Q_0$-measure zero. Thus $(\bar{g}Q_0)(N_g) = Q_0(\bar{g}^{-1}N_g) = 0$.

Let e denote the identity in G. We now claim that H is a Q_0 almost G-invariant function. By Proposition 7.18, $H_e(y) = H_g(\bar{g}y)$ a.e. (Q_0). However, $H(y) = H_e(y)$ a.e. Q_0 and $H_g(\bar{g}y) = H(\bar{g}y)$ for $\bar{g}y \notin N_g$, where $Q_0(\bar{g}^{-1}N_g) = 0$. Thus $H_g(\bar{g}y) = H(\bar{g}y)$ a.e. Q_0, and this implies that $H(y) = H(\bar{g}y)$ a.e. Q_0. Therefore, there exists a $\bar{G}$-invariant measurable function, say $\tilde{H}$, such that $H = \tilde{H}$ a.e. Q_0. Since $\bar{G}$ is transitive on $\mathcal{Y}$, $\tilde{H}$ must be a constant, so H is a constant a.e. Q_0. Therefore,

$$H_e(y) = \mathcal{E}_{P_0}(f(h(X))|\tau(X) = y)$$

is a constant a.e. Q_0 and, as noted earlier, this implies that $Z = h(X)$ and $Y = \tau(X)$ are independent when $\mathcal{L}(X) = P_0$. When $\mathcal{L}(X) = g_1 P_0$, let $\tilde{P}_0 = g_1 P_0$ and note that $\{gP_0 | g \in G\} = \{g\tilde{P}_0 | g \in G\}$ so $\tau(X)$ is sufficient for $\{g\tilde{P}_0 | g \in G\}$. The argument given for P_0 now applies for $\tilde{P}_0$. Thus Z and Y are independent when $\mathcal{L}(X) = g_1 P_0$. □

A few comments concerning this result are in order. Since G acts transitively on $\{gP_0 | g \in G\}$ and $Z = h(X)$ is G-invariant, the distribution of Z is the same under each gP_0, $g \in G$. In other words, Z is an ancillary statistic. Basu's Theorem, given in the Appendix, asserts that a sufficient statistic, whose induced family of distributions is complete, is independent of an ancillary statistic. Although no assumptions concerning invariance are made in the statement of Basu's Theorem, most applications are to problems where invariance is used to show a statistic is ancillary. In Proposition 7.19, the completeness assumption of Basu's Theorem has been replaced by the invariance assumptions and, most particularly, by the assumption that the group $\bar{G}$ acts transitively on the space $\mathcal{Y}$.

◆ **Example 7.18.** The normal distribution example at the beginning of this section provided a situation where the sample mean and sample variance are independent of a scale and translation invariant statistic. We now consider a generalization of that situation. Let $\mathcal{X} = R^n - (\text{span}\{e\})$ where e is the vector of ones in R^n and suppose that a random vector $X \in \mathcal{X}$ has a density $f(\|x\|^2)$ with respect to Lebesgue measure dx on $\mathcal{X}$. The group G in the example at the beginning of this section acts on $\mathcal{X}$ by

$$(a, b)x = ax + be, \qquad (a, b) \in G.$$

Consider the statistic $t(X) = (s, \bar{X})$ where

$$\bar{X} = \frac{1}{n}\sum_1^n X_i \qquad \text{and} \qquad s^2 = \sum_1^n (X_i - \bar{X})^2.$$

Then t takes values in G and satisfies

$$t((a, b)X) = (a, b)t(X)$$

for $(a, b) \in G$. It is shown that $t(X)$ and the G-invariant statistic

$$Z(X) \equiv (t(X))^{-1}X = \frac{X - \bar{X}e}{s}$$

are independent. The verification that $Z(X)$ is invariant goes as follows:

$$\begin{aligned} Z((a, b)X) &= (t((a, b)X))^{-1}(a, b)X \\ &= ((a, b)t(X))^{-1}(a, b)X = (t(X))^{-1}X = Z(X). \end{aligned}$$

To apply Proposition 7.19, let P_0 be the probability measure with density $f(\|x\|^2)$ on $\mathfrak{X}$ and let $G = \bar{G} = \mathfrak{Y}$. Thus $t(X)$ is equivariant and $Z(X)$ is invariant. The sufficiency of $t(X)$ for the parametric family $\{gP_0 | g \in G\}$ is established by using the factorization theorem. For $(a, b) \in G$, it is not difficult to show that $(a, b)P_0$ has a density $k(x|a, b)$ with respect to dx given by

$$k(x|a, b) = \frac{1}{a^n} f\left(\left\|\frac{x - be}{a}\right\|^2\right), \qquad x \in \mathfrak{X}.$$

Since

$$\left\|\frac{x - be}{a}\right\|^2 = \frac{1}{a^2}\left(\sum_1^n x_i^2 - 2b\sum_1^n x_i + nb^2\right),$$

the density $k(x|a, b)$ is a function of Σx_i^2 and Σx_i so the pair $(\Sigma X_i^2, \Sigma X_i)$ is a sufficient statistic for the family $\{gP_0 | g \in G\}$. However, $t(X) = (s, \bar{X})$ is a one-to-one function of $(\Sigma X_i^2, \Sigma X_i)$ so $t(X)$ is a sufficient statistic. The remaining assumptions of Proposition 7.19 are easily verified. Therefore, $t(X)$ and $Z(X)$ are independent under each of the measures $(a, b)P_0$ for (a, b) in G. ◆

Before proceeding with the next example, an extension of Proposition 7.1 is needed.

Proposition 7.20. Consider the space $\mathcal{L}_{p,n}$, $n \geqslant p$, and let Q be an $n \times n$ rank k orthogonal projection. If $k \geqslant p$, then the set

$$B = \{X | X \in \mathcal{L}_{p,n}, \operatorname{rank}(QX) < p\}$$

has Lebesgue measure zero.

Proof. Let $X \in \mathcal{L}_{p,n}$ be a random vector with $\mathcal{L}(X) = N(0, I_n \otimes I_p) = P_0$. It suffices to show that $P_0(B) = 0$ since P_0 and Lebesgue measure are absolutely continuous with respect to each other. Also, write Q as

$$Q = \Gamma' D \Gamma, \qquad \Gamma \in \mathcal{O}_n$$

where

$$D = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Since

$$\operatorname{rank}(\Gamma' D \Gamma X) = \operatorname{rank}(D \Gamma X)$$

and $\mathcal{L}(\Gamma X) = \mathcal{L}(X)$, it suffices to show that

$$P_0(\operatorname{rank}(DX) < p) = 0.$$

Now, partition X as

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \qquad X_1 : k \times p$$

so

$$DX = \begin{pmatrix} X_1 \\ 0 \end{pmatrix} \in \mathcal{L}_{p,n}.$$

Thus $\operatorname{rank}(DX) = \operatorname{rank}(X_1)$. Since $k \geqslant p$ and $\mathcal{L}(X_1) = N(0, I_k \otimes I_p)$, Proposition 7.1 implies that X_1 has rank p with probability one. Thus $P_0(B) = 0$ so B has Lebesgue measure zero. □

◆ **Example 7.19.** This is generalization of Example 7.18 and deals with the general multivariate linear model discussed in Chapter 4.

As in Example 4.4, let M be a linear subspace of $\mathcal{L}_{p,n}$ defined by

$$M = \{x | x \in \mathcal{L}_{p,n}, x = ZB, B \in \mathcal{L}_{p,k}\}$$

where Z is a fixed $n \times k$ matrix of rank k. For reasons that are apparent in a moment, it is assumed that $n - k \geqslant p$. The orthogonal projection onto M relative to the natural inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{L}_{p,n}$ is $P_M = P_z \otimes I_p$ where

$$P_z = Z(Z'Z)^{-1}Z'$$

is a rank k orthogonal projection on R^n. Also, $Q_M \equiv Q_z \otimes I_p$, where $Q_z = I_n - P_z$ is the orthogonal projection onto $M^\perp$ and Q_z is a rank $n - k$ orthogonal projection on R^n. For this example, the sample space $\mathfrak{X}$ is

$$\mathfrak{X} = \{x | x \in \mathcal{L}_{p,n}, \operatorname{rank}(Q_z x) = p\}.$$

Since $n - k \geqslant p$, Proposition 7.20 implies that the complement of $\mathfrak{X}$ has Lebesgue measure zero in $\mathcal{L}_{p,n}$. In this example, the group G has elements that are pairs (T, u) with $T \in G_T^+$ where T is $p \times p$ and $u \in M$. The group operation is

$$(T_1, u_1)(T_2, u_2) = (T_1 T_2, u_1 + u_2 T_1')$$

and the action of G on $\mathfrak{X}$ is

$$(T, u)x = xT' + u.$$

For this example, $\mathcal{Y} = \overline{G} = G$ and t on $\mathfrak{X}$ to G is defined by

$$t(x) = (T(x), P_M x) \in G$$

where $T(x)$ is the unique element in G_T^+ such that $x'Q_z x = T(x)T'(x)$. The assumption that $n - k \geqslant p$ insures that $x'Q_z x$ has rank p. It is now routine to verify that

$$t((T, u)x) = (T, u)t(x)$$

for $x \in \mathfrak{X}$ and $(T, u) \in G$. Using this relationship, the function

$$h(x) \equiv (t(x))^{-1}x$$

is easily shown to be a maximal invariant under the action of G on $\mathfrak{X}$. Now consider a random vector $X \in \mathfrak{X}$ with $\mathcal{L}(X) = P_0$ where P_0 has a density $f(\langle x, x \rangle)$ with respect to Lebesgue measure on $\mathfrak{X}$. We apply Proposition 7.19 to show that $t(X)$ and $h(X)$ are independent under gP_0 for each $g \in G$. Since $\mathfrak{Y} = \overline{G} = G$, t is an equivariant function and $\overline{G}$ acts transitively on $\mathfrak{Y}$. The measurability assumptions of Proposition 7.19 are easily checked. It remains to show that $t(X)$ is a sufficient statistic for the family $\{gP_0 | g \in G\}$. For $g = (T, \mu) \in G$, gP_0 has a density given by

$$p(x|(T, \mu)) = |TT'|^{-n/2} f\big(\langle (x - \mu)(TT')^{-1}, x - \mu \rangle\big).$$

Letting $\Sigma = TT'$ and arguing as in Example 4.4, it follows that

$$\begin{aligned}\langle (x - \mu)\Sigma^{-1}, x - \mu \rangle &= \langle (P_M x - \mu)\Sigma^{-1}, P_M x - \mu \rangle \\ &\quad + \operatorname{tr}\big(\Sigma^{-1} x' Q_z x\big)\end{aligned}$$

since $\mu \in M$. Therefore, the density $p(x|(T, \mu))$ is a function of the pair $(x'Q_z x, P_M x)$ so this pair is a sufficient statistic for the family $\{gP_0 | g \in G\}$. However, $T(x)$ is a one-to-one function of $x'Q_z x$ so

$$t(x) = \big(T(x), P_M x\big)$$

is also a sufficient statistic. Thus Proposition 7.19 implies that $t(X)$ and $h(X)$ are stochastically independent under each probability measure gP_0 for $g \in G$. Of course, the choice of f that motivated this example is

$$f(w) = (\sqrt{2\pi})^{-np} \exp\big[-\tfrac{1}{2}w\big]$$

so that P_0 is the probability measure of a $N(0, I_n \otimes I_p)$ distribution on $\mathfrak{X}$.

One consequence of Proposition 7.19 is that the statistic $h(X)$ is ancillary. But for the case at hand, we now describe the distribution of $h(X)$ and show that its distribution does not even depend on the particular density f used to define P_0. Recall that

$$\begin{aligned}h(x) &= (t(x))^{-1} x = (x - P_M x)(T'(x))^{-1} = (Q_M x)(T'(x))^{-1} \\ &= (Q_z x)(T'(x))^{-1}\end{aligned}$$

where $T(x)T'(x) = x'Q_z x$ and $T(x) \in G_T^+$. Fix $x \in \mathfrak{X}$ and set

$\Psi = (Q_z x)(T'(x))^{-1}$. First note that

$$\Psi'\Psi = (T(x))^{-1} x' Q_z x (T'(x))^{-1} = I_p$$

so Ψ is a linear isometry. Let N be the orthogonal complement in R^n of the linear subspace spanned by the columns of the matrix Z. Clearly, $\dim(N) = n - k$ and the range of Ψ is contained in N since Q_z is the orthogonal projection onto N. Therefore, Ψ is an element of the space

$$\mathfrak{F}_p(N) = \{\Psi | \Psi'\Psi = I_p, \text{range}(\Psi) \subseteq N\}.$$

Further, the group

$$H = \{\Gamma | \Gamma \in \mathcal{O}_n, \Gamma(N) = N\}$$

is compact and acts transitively on $\mathfrak{F}_p(N)$ under the group action

$$\Psi \to \Gamma\Psi, \qquad \Psi \in \mathfrak{F}_p(N), \qquad \Gamma \in H.$$

Now, we return to the original problem of describing the distribution of $W = h(X)$ when $\mathcal{L}(X) = P_0$. The above argument shows that $W \in \mathfrak{F}_p(N)$. Since the compact group H acts transitively on $\mathfrak{F}_p(N)$, there is a unique invariant probability measure ν on $\mathfrak{F}_p(N)$. It will be shown that $\mathcal{L}(W) = \nu$ by proving $\mathcal{L}(\Gamma W) = \mathcal{L}(W)$ for all $\Gamma \in H$. It is not difficult to verify that $\Gamma Q_z = Q_z \Gamma$ for $\Gamma \in H$. Since $\mathcal{L}(\Gamma X) = \mathcal{L}(X)$ and $T(\Gamma X) = T(X)$, we have

$$\begin{aligned} \mathcal{L}(\Gamma W) &= \mathcal{L}(\Gamma h(X)) = \mathcal{L}\left(\Gamma Q_z X (T'(X))^{-1}\right) \\ &= \mathcal{L}\left(Q_z \Gamma X (T'(\Gamma X))^{-1}\right) = \mathcal{L}\left(Q_z X (T'(X))^{-1}\right) \\ &= \mathcal{L}(h(X)) = \mathcal{L}(W). \end{aligned}$$

Therefore, the distribution of W is H-invariant so $\mathcal{L}(W) = \nu$. ◆

Further applications of Proposition 7.19 occur in the next three chapters. In particular, this result is used to derive the distribution of the determinant of a certain matrix that arises in testing problems for the general linear model.

PROBLEMS

1. Suppose the random $n \times p$ matrix $X \in \mathcal{X}$ ($\mathcal{X}$ as in Section 7.1) has a density given by $f(x) = k|x'x|^{\gamma}\exp[-\frac{1}{2}\operatorname{tr} x'x]$ with respect to dx. The constant k depends on n, p, and γ (see Problem 6.10). Derive the density of $S = X'X$ and the density of U in the representation $X = \Psi U$ with $U \in G_T^+$ and $\Psi \in \mathcal{F}_{p,n}$.

2. Suppose $X \in \mathcal{X}$ has an $\mathcal{O}_n$-left invariant distribution. Let $P(X) = X(X'X)^{-1}X'$ and $S(X) = X'X$. Prove that $P(X)$ and $S(X)$ are independent.

3. Let Q be an $n \times n$ non-negative definite matrix of rank r and set $A = \{x | x \in \mathcal{L}_{p,n}, x'Qx \text{ has rank } p\}$. Show that, if $r \geqslant p$, then A^c has Lebesgue measure zero.

4. With $\mathcal{X}$ as in Section 7.1, $\mathcal{O}_n \times Gl_p$ acts on $\mathcal{X}$ by $x \to \Gamma x A'$ for $\Gamma \in \mathcal{O}_n$ and $A \in Gl_p$. Also, $\mathcal{O}_n \times Gl_p$ acts on $\mathcal{S}_p^+$ by $S \to ASA'$. Show that $\phi(x) = kx'x$ is equivariant for each constant $k > 0$. Are these the only equivariant functions?

5. The permutation group $\mathcal{P}_n$ acts on R^n via matrix multiplication $x \to gx$, $g \in \mathcal{P}_n$. Let $\mathcal{Y} = \{y | y \in R^n, y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n\}$. Define $f: R^n \to \mathcal{Y}$ by $f(x)$ is the vector of ordered values of the set $\{x_1, \ldots, x_n\}$ with multiple values listed.
   (i) Show f is a maximal invariant.
   (ii) Set $I_0(u) = 1$ if $u \geqslant 0$ and 0 if $u < 0$. Define $F_n(t) = n^{-1}\Sigma_1^n I_0(t - x_i)$ for $t \in R^1$. Show F_n is also a maximal invariant.

6. Let M be a proper subspace of the inner product space $(V, (\cdot, \cdot))$. Let A_0 be defined by $A_0 x = -x$ for $x \in M$ and $A_0 x = x$ for $x \in M^{\perp}$.
   (i) Verify that the set of pairs (B, y), with $y \in M$ and B either A_0 or A_0^2, forms a subgroup of the affine group $Al(V)$. Let G be this group.
   (ii) Show that G acts on M and on V.
   (iii) Suppose $t: V \to M$ is equivariant ($t(Bx + y) = Bt(x) + y$ for $(B, y) \in G$ and $x \in V$). Prove that $t(x) = P_M x$.

7. Let M be a subspace of R^n ($M \neq R^n$) so the complement of $\mathcal{X} = R^n \cap M^c$ has Lebesgue measure zero. Suppose $X \in \mathcal{X}$ has a density given by

$$p(x|\mu, \sigma) = \frac{1}{\sigma^n} f_0\left(\frac{\|x - \mu\|^2}{\sigma^2}\right)$$

where $\mu \in M$ and $\sigma > 0$. Assume that $\int \|x\|^2 f_0(\|x\|^2)\, dx < +\infty$. For $a > 0$, $\Gamma \in \mathcal{O}_n(M)$, and $b \in M$, the affine transformation $(a, \Gamma, b)x = a\Gamma x + b$ acts on $\mathfrak{X}$.

(i) Show that the probability model for X ($\mu \in M$, $\sigma > 0$) is invariant under the above affine transformations. What is the induced group action on (μ, σ^2)?

(ii) Show that the only equivariant estimator of μ is $P_M x$. Show that any equivariant estimator of σ^2 has the form $k\|Qx\|^2$ for some $k > 0$.

8. With $\mathfrak{X}$ as in Section 7.1, suppose f is a function defined on Gl_p to $[0, \infty)$, which satisfies $f(AB) = f(BA)$ and

$$\int_{\mathfrak{X}} f(x'x) \frac{dx}{|x'x|^{n/2}} = 1.$$

(i) Show that $f(x'x\Sigma^{-1})$, $\Sigma \in \mathcal{S}_p^+$, is a density on $\mathfrak{X}$ with respect to $dx/|x'x|^{n/2}$ and under this density, the covariance (assuming it exists) is $cI_n \otimes \Sigma$ where $c > 0$.

(ii) Show that the family of distributions of (i) indexed by $\Sigma \in \mathcal{S}_p^+$ is invariant under the group $\mathcal{O}_n \times Gl_p$ acting on $\mathfrak{X}$ by $(\Gamma, A)x = \Gamma x A'$. Also, show that $\overline{(\Gamma, A)}\Sigma = A\Sigma A'$.

(iii) Show that the equivariant estimators of Σ all have the form $kX'X$, $k > 0$.

Now, assume that

$$\sup_{C \in \mathcal{S}_p^+} f(C) = f(C_0)$$

where $C_0 \in \mathcal{S}_p^+$ is unique.

(iv) Show $C_0 = \alpha I_p$ for some $\alpha > 0$.

(v) Find the maximum likelihood estimator of Σ (expressed in terms of X and α in (iv)).

9. In an inner product space $(V, (\cdot, \cdot))$, suppose X has a distribution P_0.

(i) Show that $\mathcal{L}(\|X\|) = \mathcal{L}(\|Y\|)$ whenever $\mathcal{L}(Y) = gP_0$, $g \in \mathcal{O}(V)$.

(ii) In the special case that $\mathcal{L}(X) = \mathcal{L}(\mu + Z)$ where μ is a fixed vector and Z has an $\mathcal{O}(V)$-invariant distribution, how does the distribution of $\|X\|$ depend on μ?

10. Under the assumptions of Problem 4.5, use an invariance argument to show that the distribution of F depends on (μ, σ^2) only through the parameter $(\|\mu\|^2 - \|P_\omega \mu\|^2)/\sigma^2$. What happens when $\mu \in \omega$?

11. Suppose $X_1, \dots, X_n$ is a random sample from a distribution on R^p $(n > p)$ with density $p(x|\mu, \Sigma) = |\Sigma|^{-1/2} f((x-\mu)'\Sigma^{-1}(x-\mu))$ where $\mu \in R^p$ and $\Sigma \in \mathcal{S}_p^+$. The parameter $\theta = \det(\Sigma)$ is sometimes called the *population generalized variance*. The *sample generalized variance* is $V = \det((1/n)S)$ where $S = \Sigma_1^n (x_i - \bar{x})(x_i - \bar{x})'$. Show that the distribution of V depends on (μ, Σ) only through θ.

12. Assume the conditions under which Proposition 7.16 was proved. Given a probability Q on $\mathcal{Y}$, let $\bar{Q}$ denote the extension of Q to $\mathcal{X}$—that is, $\bar{Q}(B) = Q(B \cap \mathcal{Y})$ for $B \in \mathcal{B}$. For $g \in G$, $g\bar{Q}$ is defined in the usual way—$(g\bar{Q})(B) = \bar{Q}(g^{-1}B)$.

(i) Assume that P is a probability measure on $\mathcal{X}$ and

$$P = \int_G g\bar{Q}\mu(dg), \tag{7.1}$$

that is,

$$P(B) = \int_G (g\bar{Q})(B)\mu(dg); \qquad B \in \mathcal{B}.$$

Show that P is G-invariant.

(ii) If P is G-invariant, show that (7.1) holds for some Q.

13. Under the assumptions used to prove Proposition 7.16, let $\mathcal{P}$ be all the G-invariant distributions. Prove that $\tau(X)$ is a sufficient statistic for the family $\mathcal{P}$.

14. Suppose $X \in R^n$ has coordinates $X_1, \dots, X_n$ that are i.i.d. $N(\mu, 1)$, $\mu \in R^1$. Thus the parameter space for the distributions of X is the additive group $G = R^1$. The function $t: R^n \to G$ given by $t(x) = \bar{x}$ gives a complete sufficient statistic for the model for X. Also, G acts on R^n by $gx = x + ge$ where $e \in R^n$ is the vector of ones.

(i) Show that $t(gx) = gt(x)$ and that $Z(X) = (t(X))^{-1}X$ is an ancillary statistic. Here, $(t(X))^{-1}$ means the group inverse of $t(X) \in G$ so $(t(X))^{-1}X = X - \bar{X}e$. What is the distribution of $Z(X)$?

(ii) Suppose we want to find a minimum variance unbiased estimator (MVUE) of $h(\mu) = \mathcal{E}_\mu f(X_1)$ where f is a given function. The Rao–Blackwell Theorem asserts that the MVUE is $\mathcal{E}(f(X_1)|t(X) = t)$. Show that this conditional expectation is

$$\int_{-\infty}^{\infty} f(z + t)\frac{1}{\sqrt{2\pi\delta}} \exp\left[-\frac{z^2}{2\delta^2}\right] dz$$

where $\delta^2 = \text{var}(X_1 - \bar{X}) = (n-1)/n$. Evaluate this for $f(x) = 1$ if $x \leqslant u_0$ and $f(x) = 0$ if $x > u_0$.

(iii) What is the MVUE of the parametric function $(\sqrt{2\pi})^{-1}\exp[-\frac{1}{2}(x_0 - \mu)^2]$ where x_0 is a fixed number?

15. Using the notation, results, and assumptions of Example 7.18, find an unbiased estimator based on $t(X)$ of the parametric function $h(a, b) = ((a, b)P_0)(X_1 \leqslant u_0)$ where u_0 is a fixed number and X_1 is the first coordinate of X. Express the answer in terms of the distribution of Z_1—the first coordinate of Z. What is this distribution? In the case that P_0 is the $N(0, I_n)$ distribution, show this gives a MVUE for $h(a, b)$.

16. This problem contains an abstraction of the technique developed in Problems 14 and 15. Under the conditions used to prove Proposition 7.19, assume the space $(\mathcal{Y}, \mathcal{C}_1)$ is $(G, \mathcal{B}_G)$ and $\bar{G} = G$. The equivariance assumption on τ then becomes $\tau(gx) = g \circ \tau(x)$ since $\tau(x) \in G$. Of course, $\tau(X)$ is assumed to be a sufficient statistic for $\{gP_0 | g \in G\}$.

(i) Let $Z(X) = (\tau(X))^{-1}X$ where $(\tau(X))^{-1}$ is the group inverse of $\tau(X)$. Show that $Z(X)$ is a maximal invariant and $Z(X)$ is ancillary. Hence Proposition 7.19 applies.

(ii) Let Q_0 denote the distribution of Z when $\mathcal{L}(X)$ is one of the distributions gP_0, $g \in G$. Show that a version of the conditional expectation $\mathcal{E}(f(X)|\tau(X) = g)$ is $\mathcal{E}_{Q_0} f(gZ)$ for any bounded measurable f.

(iii) Apply the above to the case when P_0 is $N(0, I_n \otimes I_p)$ on $\mathcal{X}$ (as in Section 7.1) and take $G = G_T^+$. The group action is $x \to xT'$ for $x \in \mathcal{X}$ and $T \in G_T^+$. The map τ is $\tau(X) = T$ in the representation $X = \Psi T'$ with $\Psi \in \mathcal{F}_{p,n}$ and $T \in G_T^+$. What is Q_0?

(iv) When $X \in \mathcal{X}$ is $N(0, I_n \otimes \Sigma)$ with $\Sigma \in \mathcal{S}_p^+$, use (iii) to find a MVUE of the parametric function

$$(\sqrt{2\pi})^{-p}|\Sigma|^{-1/2}\exp\left[-\tfrac{1}{2}u_0'\Sigma^{-1}u_0\right]$$

where u_0 is a fixed vector in R^p.

NOTES AND REFERENCES

1. For some material related to Proposition 7.3, see Dawid (1978). The extension of Proposition 7.3 to arbitrary compact groups (Proposition 7.16) is due to Farrell (1962). A related paper is Das Gupta (1979).

2. If G acts on $\mathfrak{X}$ and t is a function from $\mathfrak{X}$ onto $\mathfrak{Y}$, it is natural to ask if we can define a group action on $\mathfrak{Y}$ (using t and G) so that t becomes equivariant. The obvious thing to do is to pick $y \in \mathfrak{Y}$, write $y = t(x)$, and then define gy to be $t(gx)$. In order that this definition make sense it is necessary (and sufficient) that whenever $t(x) = t(\tilde{x})$, then $t(gx) = t(g\tilde{x})$ for all $g \in G$. When this condition holds, it is easy to show that G then acts on $\mathfrak{Y}$ via the above definition and t is equivariant. For some further discussion, see Hall, Wijsman, and Ghosh (1965).

3. Some of the early work on invariance by Stein, and Hunt and Stein, first appeared in print in the work of other authors. For example, the famous Hunt–Stein Theorem given in Lehmann (1959) was established in 1946 but was never published. This early work laid the foundation for much of the material in this chapter. Other early invariance works include Hotelling (1931), Pitman (1939), and Peisakoff (1950). The paper by Kiefer (1957) contains a generalization of the Hunt–Stein Theorem. For some additional discussion on the development of invariance arguments, see Hall, Wijsman, and Ghosh (1965).

4. Proposition 7.15 is probably due to Stein, but I do not know a reference.

5. Make the assumptions on $\mathfrak{X}$, $\mathfrak{Y}$, and G that lead to Proposition 7.16, and note that $\mathfrak{Y}$ is just a particular representation of the quotient space $\mathfrak{X}/G$. If ν is any σ-finite G-invariant measure on $\mathfrak{X}$, let δ be the measure on $\mathfrak{Y}$ defined by

$$\delta(C) = \nu\big(\tau^{-1}(C)\big), \qquad C \subseteq \mathfrak{Y}.$$

Then (see Lehmann, 1959, p. 39),

$$\int_{\mathfrak{X}} h(\tau(x))\nu(dx) = \int_{\mathfrak{Y}} h(y)\delta(dy)$$

for all measurable functions h. The proof of Proposition 7.16 shows that for any ν-integrable function f, the equation

$$\int_{\mathfrak{X}} f(x)\nu(dx) = \int_{\mathfrak{Y}}\int_G f(gy)\mu(dg)\delta(dy) \tag{7.2}$$

holds. In an attempt to make sense of (7.2) when G is not compact, let

μ_r denote a right invariant measure on G. For $f \in \mathcal{K}(\mathcal{X})$, set

$$\hat{f}(x) = \int_G f(gx)\mu_r(dg).$$

Assuming this integral is well defined (it may not be in certain examples —e.g., $\mathcal{X} = R^n - \{0\}$ and $G = Gl_n$), it follows that $\hat{f}(hx) = \hat{f}(x)$ for $h \in G$. Thus $\hat{f}$ is invariant and can be regarded as a function on $\mathcal{Y} = \mathcal{X}/G$. For any measure δ on $\mathcal{Y}$, write $\int \hat{f}\, d\delta$ to mean the integral of $\hat{f}$, expressed as a function of y, with respect to the measure δ. In this case, the right-hand side of (7.2) becomes

$$J(f) = \int\left(\int_G f(gx)\mu_r(dg)\right) d\delta = \int \hat{f}\, d\delta.$$

However, for $h \in G$, it is easy to show

$$J(hf) = \Delta_r^{-1}(h)J(f)$$

so J is a relatively invariant integral. As usual, Δ_r is the right-hand modulus of G. Thus the left-hand side of (7.2) must also be relatively invariant with multiplier Δ_r^{-1}. The argument thus far shows that when μ in (7.2) is replaced by μ_r (this choice looks correct so that the inside integral defines an invariant function), the resulting integral J is relatively invariant with multiplier Δ_r^{-1}. Hence the only possible measures ν for which (7.2) can hold must be relatively invariant with multiplier Δ_r^{-1}. However, given such a ν, further assumptions are needed in order that (7.2) hold for some δ (when G is not compact and μ is replaced by μ_r). Some examples where (7.2) is valid for noncompact groups are given in Stein (1956), but the first systematic account of such a result is Wijsman (1966), who uses some Lie group theory. A different approach due to Schwarz is reported in Farrell (1976). The description here follows Andersson (1982) most closely.

6. Proposition 7.19 is a special case of a result in Hall, Wijsman, and Ghosh (1965). Some version of this result was known to Stein but never published by him. The development here is a modification of that which I learned from Bondesson (1977).