CHAPTER 9

Application to Type I Problems: Special Group Structure

9.1. Preliminaries. For convenience, but with abuse of notation, we shall in this and subsequent chapters not distinguish in notation between a random matrix and its possible values. Furthermore, the distribution of a random matrix, say X, will usually be written $P(dX)$ without any label on P, even though this P may represent a different distribution from one example to the next. The examples that will be exhibited below have been chosen because they deal with transformations that are basic and have important applications, and/or because they provide a good illustration of the method developed in Chapter 8. No attempt at exhaustiveness has been made. In particular, the treatment of GMANOVA has not been included. However, it can be found in Wijsman (1986), Section 6.5. There are several examples in this and the next chapter where the method could be applied from scratch, but where it is also possible to combine the results of two other, simpler, examples. In such cases we shall always follow the latter route. Concerning notation, from now on we shall not insist on restricting the use of symbols X and T to their meaning in Chapters 1 and 8, i.e., to the random variables with values in $\mathcal{X}$, $\mathcal{T}$, respectively. For instance, T will often be used to denote a triangular matrix. For future use we rewrite equation (8.20):

$$x = ghx_0. \tag{9.1.1}$$

It should be kept in mind that the right-hand side of (9.1.1) depends on g and h only through $[g] = y$ and $[h] = t$.

All examples will be of the following general nature: $\mathfrak{X}$ is an invariant (under the group K) subset of some Euclidean space and the relatively invariant measure λ of Theorem 8.14 is Lebesgue measure. Then the value of the multiplier $\chi(k)$ for $k \in K$ is needed. If the transformation of $\mathfrak{X}$ by k is a translation, then $\chi(k) = 1$, of course. If the action of k is a nonsingular linear transformation of $\mathfrak{X}$, then by (4.3.7) $\chi(k)$ equals the absolute value of the determinant of the linear transformation. It follows that for an orthogonal transformation the value of χ is 1. Other transformations are often of the following types (for notation see Section 7.7): if C is nonsingular, then

$$X \in M(m,n), \quad X \to CX, \quad \chi(C) = |C|^n, \tag{9.1.2}$$

$$X \in M(m,n), \quad X \to XC, \quad \chi(C) = |C|^m, \tag{9.1.3}$$

$$S \in PD(n), \quad S \to CSC', \quad \chi(C) = |C|^{n+1}, \tag{9.1.4}$$

$$A \in AS(n), \quad A \to CAC', \quad \chi(C) = |C|^{n-1}. \tag{9.1.5}$$

Here (9.1.2) follows from (4.3.7) by applying it to the n-fold product of R^m with itself, and (9.1.3) follows similarly. For (9.1.4) see (5.3.13); (9.1.5) is obtained in a similar fashion.

9.2. Distribution of $X'X$ and Cholesky decomposition. Let X be a random $q \times p$ matrix with distribution $P(dX) = p(X)(dX)$ (note that (dX) is Lebesgue measure). We may assume that X is of maximal rank since the P-probability that this is not so is 0. Therefore, put $s = \min(p,q)$ and define $\mathfrak{X} = \{X \in M(q,p) : \operatorname{rank}(X) = s\}$. Let $G = O(q)$ with action on $\mathfrak{X}$ defined by $X \to \Gamma X$, $\Gamma \in G$. The multiplier is $\chi(g) = 1$. A maximal invariant under G is $X'X$. There are two cases to be considered that have to be dealt with slightly differently: $q \geq p$ and $q < p$. The first case has immediate relevance to the Wishart matrix (central or noncentral) if the rows of X are independent and p-variate normal with common covariance matrix. Also of interest is the Cholesky decomposition

$$X'X = TT', \quad T \in LT(p). \tag{9.2.1}$$

The result for general q is needed in Section 9.4.

Case 1: $q \geq p$. Take $H = LT(p)$ with action $X \to XT'$, $T \in H$, and multiplier $\chi(h) = \chi(T) = |T|^q$ by (9.1.3). Then $K = GH$ is a transitive group (coinciding with $G \times H$) in which G and H commute. Choose $x_0 = [I_p, 0]' : q \times p$ (the block of zeros is absent if $q = p$). Then $H_0 = \{e\}$, and the validity of the hypotheses of Theorem 8.14 is easily checked. Partition $\Gamma = [\Gamma_1, \Gamma_2]$ with $\Gamma_1 : q \times p$, then in terms of the present symbols the equation (9.1.1) becomes

$$X = \Gamma_1 T'. \tag{9.2.2}$$

Note that T is the unique matrix $T \in LT(p)$ in the Cholesky decomposition (9.2.1). Since $H_0 = \{e\}$, $\mathcal{T} \equiv H/H_0 = H$, so that T is a maximal invariant. We shall now find the form that the factorization (8.22) takes in this example. The left-hand side of (8.22) is (dX). In the function $\beta(h)$ of (8.21), since G and H commute, we have $\Delta^K(h) = \Delta^H(h)$ so that $\beta(h) = \chi(h) = |T|^q$, and note that in (8.22) $t = h$. Furthermore, $\chi(y) = 1$, $\mu_{\mathcal{T}} = \mu_H = \mu_{LT(p)}$ is given by (7.7.2) (with n replaced by p) and $\mu_{\mathcal{Y}}$ by (7.7.15) (with n replaced by q, s by $q - p$). In order to find c of (8.22) write down (8.26) in terms of differential forms:

$$(dX) = cc_{q-p}(d\Gamma_1)(dT). \tag{9.2.3}$$

Then differentiate (9.2.2) at x_0, i.e., $\Gamma = I_q$, $T = I_p$: $dX = d\Gamma_1 + dT'$. Taking the wedge product of all dx_{ij} gives $(dX) = (d\Gamma_1)(dT)$. Comparing this with (9.2.3) yields $c = c_{q-p}^{-1}$. Thus, (8.2.2) becomes

$$(dX) = c_{q-p}^{-1}\mu_{\mathcal{Y}}(dy)|T|^q\mu_{LT(p)}(dT). \tag{9.2.4}$$

It was tacitly assumed in this derivation that $q > p$. However, the result remains valid if $q = p$, with $\mathcal{Y} = G$, thanks to the convention $c_0 = 1$. The distribution of the maximal invariant T, given by (8.23), now becomes

$$P(dT) = c_{q-p}^{-1} \prod_{i=1}^{p} t_{ii}^{q-i}(dT) \int p(\Gamma_1 T')\mu_{O(q)}(d\Gamma), \tag{9.2.5}$$

where c_{q-p} was defined in (7.7.9). If p is taken as the joint density of q iid $N(0, I_p)$ variables, then $p(\Gamma_1 T')$ is proportional to $\exp(-\frac{1}{2}\operatorname{tr} TT')$ and (9.2.5) reduces to

$$P(dT) = c \exp\left(-\tfrac{1}{2}\operatorname{tr} TT'\right) \prod_{i=1}^{p} t_{ii}^{q-i} (dT) \tag{9.2.6}$$

(with some $c > 0$) which reveals the well-known fact that in the Cholesky decomposition of the standard Wishart matrix the elements of T are independent, with standard normals off, and χ-variables on the diagonal.

In (9.2.2) the $q \times p$ matrices Γ_1 take their values in the Stiefel manifold $V_{p,q}$ of p-frames in q-space (James (1954), Muirhead (1982); usually q is denoted n in this context). Thus, $\mathcal{Y} = G/G_0$ can be identified with $V_{p,q}$ and the measure $\mu_{\mathcal{Y}}$ in (9.2.4) is a version of the invariant measure on $V_{p,q}$ under the (left) action of G. The factorization (9.2.4) is implicit in Eaton (1983), Example 6.18, and is stated in Muirhead (1982) as Equation (19) of Theorem 2.1.13. An explicit expression for the invariant measure on $V_{p,n}$ can be found in James (1954), Section 4.7, or Muirhead (1982), Equation (20) of Theorem 2.1.13. However, it should be noted that this explicit expression is not needed for (9.2.4); only the expression for $\mu_{\mathcal{Y}}$ at $y = [e]$ is used. The factorization (9.2.4) also shows that Γ_1 and T are independent if $p(\Gamma_1 T')$ factors; for instance, if $p(X)$ is invariant under $X \to \Gamma X$ ($\Gamma \in O(q)$) as is the case when the rows of X are iid $N(0, \Sigma)$.

The distribution (9.2.5) of T leads directly to the distribution of $S = X'X = TT'$ (by (9.2.1)) via (5.3.16) after replacing in the latter n by p, with the result

$$P(dS) = 2^{-p} c_{q-p}^{-1} |S|^{\frac{1}{2}(q-p-1)} (dS) \int p(\Gamma X) \mu_{O(q)}(d\Gamma), \tag{9.2.7}$$

in which X is any $q \times p$ matrix such that $X'X = S$ (this is justified by the invariance of $\mu_{O(q)}$ and the fact that there is a fixed $\Gamma_0 \in O(q)$ such that $\Gamma_0 X = [T, 0]'$, which after absorbing Γ_0 into Γ reduces the integral in (9.2.7) to the one in (9.2.5)).

The $W(n,\Sigma)$ distribution emerges from (9.2.7) (after replacing q by the more customary n) if the n rows of X are iid $N(0,\Sigma)$. In that case $p(X) = (2\pi)^{-\frac{1}{2}pn}|\Sigma|^{-\frac{1}{2}n}\exp(-\frac{1}{2}\operatorname{tr} X\Sigma^{-1}X')$. Then $p(\Gamma X) = p(X)$ does not depend on Γ and can be taken outside the integral in (9.2.7). The remaining integration produces the constant c_n given by (7.7.9). Furthermore, $\operatorname{tr} X\Sigma^{-1}X' = \operatorname{tr}\Sigma^{-1}S$. Thus, the distribution of $S \sim W(n,\Sigma)$ is

$$P(dS) = (2\pi)^{-\frac{1}{2}pn}2^{-p}c_n c_{n-p}^{-1}|\Sigma|^{-\frac{1}{2}(n-p-1)}e^{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}S}(dS), \tag{9.2.8}$$

provided $n \geq p$. This derivation of the well-known distribution of the central Wishart matrix (see, e.g., Anderson, 1984, Sect. 7.2) shows that the contributions to the numerical constant come from three sources: the factor $(2\pi)^{-\frac{1}{2}pn}$ from the standard normal distribution; the factor 2^{-p} from the relation between left invariant measure on $LT(p)$ and on $PD(p)$; and the factor $c_n c_{n-p}^{-1}$ from integration of nonnormalized Haar measures on orthogonal groups. The same phenomenon seems to occur in other typical multivariate distributions. For instance, see (10.1.3).

Equations (9.2.5) and (9.2.7) can be used for the distribution of T or $X'X$ when X has arbitrary density. Equation (9.2.7) will also be used in Section 10.2 for the distribution of singular values of a matrix.

Case 2: $q < p$. Partition $X = [X_1, X_2]$, where we may assume $X_1 \in GL(q)$ by removing from R^{pq} a null set. Take $H = LT(q) \times M(q, p-q)$ with action $X_1 \to X_1T'$, $X_2 \to X_2 + A$, $T \in LT(q)$, $A \in M(q,p-q)$. Then the combined actions of G and H amount to the transitive action of a group $K = GH$, such that H is normal in K. If $h = (T,A) \in H$, then $\chi(h) = |T|^q$, by (9.1.3). Choose $x_0 = [I_q, 0]$, then $G_0 = H_0 = \{e\}$ and Theorem 8.14 applies. Since $G_0 = \{e\}$, $\mu_{\mathcal{Y}}(dy) = \mu_G(dg) = (d\Gamma)$ at $g = e$ (equation (7.7.7)). Since $H_0 = \{e\}$, $\mu_{\mathcal{T}}(dt) = \mu_H(dh) = (dT)(dA)$ at $h = e$ (use (7.7.2)). Thus, (8.26) reads

$$(dX) = c(d\Gamma)(dT)(dA) \quad \text{at} \quad x = x_0. \tag{9.2.9}$$

On the other hand, (9.1.1) reads

$$X = \Gamma[T', A] \tag{9.2.10}$$

and differentiation of (9.2.10) at $\Gamma = T = I_q$, $A = 0$, yields $dX = [d\Gamma + dT', dA]$. Equating the wedge products on both sides gives $(dX) = (d\Gamma)(dT)(dA)$ at $x = x_0$. Comparison with (9.2.9) shows that $c = 1$. Therefore, (8.23) becomes

$$P(dT, dA) = \prod_{i=1}^{q} t_{ii}^{q-i}(dT)(dA) \int p(\Gamma[T', A])\mu_{O(q)}(d\Gamma). \tag{9.2.11}$$

This result can be put in a form analogous to (9.2.7). However, the matrix $S = X'X$ is now singular so it does not have a density. Instead, partition $S = ((S_{ij}))$, $i, j = 1, 2$, with $S_{11} \in PD(q)$, then (S_{11}, S_{21}) has a density and $S_{22} = S_{21}S_{11}^{-1}S_{21}'$. With (dS) we shall now mean $(dS_{11})(dS_{21})$, and with $P(dS)$ the distribution of (S_{11}, S_{21}). With help of (9.2.10) we find $S_{11} = TT'$, $S_{21} = A'T'$. Differentiation of these equations, together with (5.3.16) (with n replaced by q) and (9.1.3), yields

$$\prod_{i=1}^{q} t_{ii}^{-i}(dT)(dA) = 2^{-q}|S_{11}|^{-\frac{1}{2}(p+1)}(dS), \tag{9.2.12}$$

and substitution into (9.2.11) gives

$$P(dS) = 2^{-q}|S_{11}|^{\frac{1}{2}(q-p-1)}(dS) \int p(\Gamma X)\mu_{O(q)}(d\Gamma) \tag{9.2.13}$$

with X any $q \times p$ matrix for which $X'X = S$. For the sequel it is convenient to present (9.2.7) and (9.2.13) as one formula. To this end define, for integer r,

$$c(r) = \begin{cases} 1 & \text{if } r \le 0 \\ 2^r c_r^{-1} & \text{if } r > 0 \end{cases} \tag{9.2.14}$$

with c_r defined in (7.7.9). Then for any positive integers p, q, and $s = \min(p, q)$, (9.2.7) and (9.2.13) can both be written as

$$P(dS) = 2^{-q}c(q-p)|S_{11}|^{\frac{1}{2}(q-p-1)}(dS) \int p(\Gamma X)\mu_{O(q)}(d\Gamma) \tag{9.2.15}$$

in which X is any $q \times p$ matrix such that $X'X = S$, S_{11} is the $s \times s$ upper left submatrix of S, and (dS) is the wedge product of all ds_{ij} with $i \ge j \le s$.

9.3. MANOVA under triangular group. Let $\mathfrak{X}$ be the space of all (U,S), $U \in M(p,q)$, $S \in PD(p)$, and $G = LT(p)$ with action

$$U \to UT', \quad S \to TST', \quad T \in G. \tag{9.3.1}$$

This problem has been considered by Schwartz (1967) in the multivariate analysis of variance (MANOVA) case, where the rows of U are independent p-variate normal with common covariance matrix $\Sigma \in PD(p)$ and S is an independent $W(n,\Sigma)$ matrix. When, in addition, $q = 1$, the problem has been treated by Giri, Kiefer, and Stein (1963), and by Farrell (1985), Section 9.2. The problem for arbitrary q has application to MANOVA if for some reason it is decided not to reduce by the full invariance group but only by the lower triangular group (as in Schwartz, 1967). The result can also be used to great advantage as one of the steps in the full reduction of the MANOVA problem; see Sections 9.4 and 10.3.

Take $H = M(q,p)$ with action $U \to U+X$, $X \in H$, S unchanged. Then H is normal in the transitive group GH. Take $x_0 = (0, I_p) \in \mathfrak{X}$ so that (9.1.1) reads

$$U = XT', \quad S = TT', \quad T \in G, \quad X \in H. \tag{9.3.2}$$

Then $G \cap H$, G_0, and H_0 are all trivial so that $\mathfrak{Y} = G$, $\mathfrak{T} = H$, and $X \in H$ is therefore a maximal invariant. In terms of (U,S) it follows from (9.3.2) that this maximal invariant is

$$X = UT_S'^{-1} \tag{9.3.3}$$

where

$$S = T_S T_S', \quad T_S \in LT(p), \tag{9.3.4}$$

is the Cholesky decomposition of S. Theorem 8.14 applies and $\beta = \chi$ since H is normal. Furthermore, $\chi(t) = \chi(X) = 1$, and $\chi(y) =$

$\chi(T) = |T|^{q+p+1}$ by (9.3.1), (9.1.3), and (9.1.4). In order to evaluate c in (8.23) write down (8.26) in terms of the present symbols:

$$(9.3.5) \qquad (dU)(dS) = c(dT)(dX), \quad \text{at } x = x_0.$$

Differentiate (9.3.2) at $T = I_p$, $X = 0$ and obtain $(dU)(dS) = 2^P(dX)(dT)$ at $x = x_0$ (the result $(dS) = 2^p(dT)$ was obtained in the computation that led to (5.3.16)). Comparison with (9.3.5) shows $c = 2^p$. Hence, (8.23) reads

$$(9.3.6) \qquad P(dX) = 2^p(dX) \int p(XT', TT') \prod_{i=1}^{p} t_{ii}^{q+p+1-i}(dT),$$

when (U, S) has density $p(U, S)$.

9.4. Distribution of $US^{-1}U'$; MANOVA under general linear group. As in Section 9.3, $x = (U, S)$ with $U : q \times p$ and $S \in PD(p)$, but now $G = GL(p)$ with action

$$(9.4.1) \qquad U \to UC', \quad S \to CSC', \quad C \in G.$$

This is part of the usual invariance reduction in MANOVA. (The remaining reduction by an orthogonal group will be treated in Section 10.3.) A maximal invariant of the action (9.4.1) is

$$(9.4.2) \qquad Q = US^{-1}U'.$$

Its distribution can be obtained by combining the results of Section 9.2 and 9.3. Take X of (9.3.3) with T_S of (9.3.4), then $Q = XX'$. The distribution of X is given by (9.3.6), and then the distribution of Q is given by (9.2.15) in which S is to be replaced by Q, X' by X, and p and q interchanged. In the resulting double integral the function $p(X\Gamma'T', TT')|T|^{p+q+1}$ is integrated with respect to $\mu_{LT(p)}(dT)\mu_{O(p)}(d\Gamma)$. This can be contracted, using (7.7.10) and (7.7.1), with the result

$$(9.4.3) \quad P(dQ) = c(p-q)|Q_{11}|^{\frac{1}{2}(p-q-1)}(dQ) \cdot \int p(XC', CC')|C|^{q+1}(dC),$$

in which X is any $q \times p$ matrix such that $XX' = Q$, $p(U,S)$ is the density of (U,S), and Q_{11} is the $s \times s$ upper left submatrix of Q, with $s = \min(p,q)$. The wedge product (dQ) is to be understood in the same sense as (dS) in (9.2.15), and the constant $c(\cdot)$ is defined in (9.2.14).

In the central MANOVA case the rows of U are iid $N(0,\Sigma)$. The distribution of Q does not depend on Σ so that we may choose $\Sigma = I_p$. Then $p(U,S) = c\exp(-\frac{1}{2}\operatorname{tr}U'U - \frac{1}{2}\operatorname{tr}S)|S|^{\frac{1}{2}(n-p-1)}$ (we shall not keep track of the values of the constant, generically denoted c). Substitute this into (9.4.3), then the integral becomes

$$\int \exp\left(-\frac{1}{2}\operatorname{tr}C(I_p + X'X)C'\right)|C|^{n+q}|C|^{-p}(dC).$$

Make a change of variable $C(I_p + X'X)^{1/2} = B$ and use the right invariance of the Haar measure $|C|^{-p}(dC)$, then the value of the integral is proportional to $|I_p + X'X|^{-\frac{1}{2}(n+q)} = |I_q + XX'|^{-\frac{1}{2}(n+q)} = |I_q + Q|^{-\frac{1}{2}(n+q)}$ and (9.4.3) becomes

$$P(dQ) = c|Q_{11}|^{\frac{1}{2}(p-q-1)}|I_q + Q|^{-\frac{1}{2}(n+q)}(dQ). \tag{9.4.4}$$

This distribution was obtained for the case $q \leq p$ by Olkin and Rubin (1964) who termed it the *central Studentized Wishart distribution.* Khatri (1965) obtained for arbitrary q the analogous distribution for the complex noncentral case.

9.5. Multivariate beta. This was treated by Olkin and Rubin (1964), Theorem 3.2 in the central case. The noncentral case is treated in Farrell (1985), Section 10.3 for two Wishart matrices. Let $S_0, S_1, \ldots, S_k \in PD(p)$ and let $G = LT(p)$ with action $S_j \to TS_jT'$, $j = 0,\ldots,k$, $T \in G$. Choose $H = LT(p) \times \cdots \times LT(p)$ (k factors), with action $S_j \to Z_jS_jZ_j'$, $Z_j \in LT(p)$, $j = 1,\ldots,k$. Then $K = GH$ is transitive over $\mathcal{X} = PD(p) \times \cdots \times PD(p)$ ($k+1$ factors), and H is normal in K. Choose x_0 as the point where $S_j = I_p$, $j = 0,\ldots,k$, then $G \cap H = G_0 = H_0 = \{e\}$ so that $\mathcal{Y} = G$, $\mathcal{T} = H$. It follows that

$(Z_1, \ldots, Z_k)$ is a maximal invariant, or, equivalently, $(U_1, \ldots, U_k)$, where $U_j = Z_j Z_j'$, $j = 1, \ldots, k$. Easy computations of a similar nature as in the previous sections, plus use of (5.3.16), transform (8.23) into

$$P(dU_1, \ldots, dU_k) = 2^p (dU_1) \ldots (dU_k) \cdot \int p(TT', TU_1T', \ldots, TU_kT') \prod_{i=1}^{p} t_{ii}^{(p+1)(k+1)-i} (dT), \tag{9.5.1}$$

in which $p(S_0, \ldots, S_k)$ is the density of $(S_0, \ldots, S_k)$. If one takes for the latter $k+1$ independent Wishart matrices with the same covariance matrix, then (9.5.1) reduces easily to Equation (3.6) in Olkin and Rubin (1964).