CHAPTER 8

The Wishart Distribution

The Wishart distribution arises in a natural way as a matrix generalization of the chi-square distribution. If $X_1, \ldots, X_n$ are independent with $\mathcal{L}(X_i) = N(0, 1)$, then $\Sigma_1^n X_i^2$ has a chi-square distribution with n degrees of freedom. When the X_i are random vectors rather than real-valued random variables say $X_i \in R^p$ with $\mathcal{L}(X_i) = N(0, I_p)$, one possible way to generalize the above sum of squares is to form the $p \times p$ positive semidefinite matrix $S = \Sigma_1^n X_i X_i'$. Essentially, this representation of S is used to define a Wishart distribution. As with the definition of the multivariate normal distribution, our definition of the Wishart distribution is not in terms of a density function and allows for Wishart distributions that are singular. In fact, most of the properties of the Wishart distribution are derived without reference to densities by exploiting the representation of the Wishart in terms of normal random vectors. For example, the distribution of a partitioned Wishart matrix is obtained by using properties of conditioned normal random vectors.

After formally defining the Wishart distribution, the characteristic function and convolution properties of the Wishart are derived. Certain generalized quadratic forms in normal random vectors are shown to have Wishart distributions and the basic decomposition of the Wishart into submatrices is given. The remainder of the chapter is concerned with the noncentral Wishart distribution in the rank one case and certain distributions that arise in connection with likelihood ratio tests.

8.1. BASIC PROPERTIES

The Wishart distribution, or more precisely, the family of Wishart distributions, is indexed by a $p \times p$ positive semidefinite symmetric matrix Σ, by a dimension parameter p, and by a degrees of freedom parameter n. Formally, we have the following definition.

Definition 8.1. A random $p \times p$ symmetric matrix S has a Wishart distribution with parameters Σ, p, and n if there exist independent random vectors $X_1, \ldots, X_n$ in R^p such that $\mathcal{L}(X_i) = N(0, \Sigma)$, $i = 1, \ldots, n$ and

$$\mathcal{L}(S) = \mathcal{L}\left(\sum_1^n X_i X_i'\right).$$

In this case, we write $\mathcal{L}(S) = W(\Sigma, p, n)$.

In the above definition, p and n are positive integers and Σ is a $p \times p$ positive semidefinite matrix. When $p = 1$, it is clear that the Wishart distribution is just a chi-square distribution with n degrees of freedom and scale parameter $\Sigma \geqslant 0$. When $\Sigma = 0$, then $X_i = 0$ with probability one, so $S = 0$ with probability one. Since $\Sigma_1^n X_i X_i'$ is positive semidefinite, the Wishart distribution has all of its mass on the set of positive semidefinite matrices. In an abuse of notation, we often write

$$S = \sum_1^n X_i X_i'$$

when $\mathcal{L}(S) = W(\Sigma, p, n)$. As distributional questions are the primary concern in this chapter, this abuse causes no technical problems. If $X \in \mathcal{L}_{p,n}$ has rows $X_1', \ldots, X_n'$, it is clear that $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$ and $X'X = \Sigma_1^n X_i X_i'$. Thus if $\mathcal{L}(S) = W(\Sigma, p, n)$, then $\mathcal{L}(S) = \mathcal{L}(X'X)$ where $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$ in $\mathcal{L}_{p,n}$. Also, the converse statement is clear. Some further properties of the Wishart distribution follow.

Proposition 8.1. If $\mathcal{L}(S) = W(\Sigma, p, n)$ and A is an $r \times p$ matrix, then $\mathcal{L}(ASA') = W(A\Sigma A', r, n)$.

Proof. Since $\mathcal{L}(S) = W(\Sigma, p, n)$,

$$\mathcal{L}(S) = \mathcal{L}(X'X)$$

where $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$ in $\mathcal{L}_{p,n}$. Thus $\mathcal{L}(ASA') = \mathcal{L}(AX'XA') = \mathcal{L}[((I_n \otimes A)X)'(I_n \otimes A)X]$. But $Y = (I_n \otimes A)X$ satisfies $\mathcal{L}(Y) = N(0, I_n \otimes (A\Sigma A'))$ in $\mathcal{L}_{r,n}$ and $\mathcal{L}(Y'Y) = \mathcal{L}(ASA')$. The conclusion follows from the definition of the Wishart distribution. $\square$

One consequence of Proposition 8.1 is that, for fixed p and n, the family of distributions $\{W(\Sigma, p, n)|\Sigma \geqslant 0\}$ can be generated from the $W(I_p, p, n)$ distribution and $p \times p$ matrices. Here, the notation $\Sigma \geqslant 0$ ($\Sigma > 0$) means that Σ is positive semidefinite (positive definite). To see this, if $\mathcal{L}(S) = W(I_p, p, n)$ and $\Sigma = AA'$, then

$$\mathcal{L}(ASA') = W(AA', p, n) = W(\Sigma, p, n).$$

In particular, the family $\{W(\Sigma, p, n)|\Sigma > 0\}$ is generated by the $W(I_p, p, n)$ distribution and the group Gl_p acting on $\mathcal{S}_p$ by $A(S) \equiv ASA'$. Many proofs are simplified by using the above representation of the Wishart distribution. The question of the nonsingularity of the Wishart distribution is a good example. If $\mathcal{L}(S) = W(\Sigma, p, n)$, then S has a *nonsingular Wishart distribution* if S is positive definite with probability one.

Proposition 8.2. Suppose $\mathcal{L}(S) = W(\Sigma, p, n)$. Then S has a nonsingular Wishart distribution iff $n \geqslant p$ and $\Sigma > 0$. If S has a nonsingular Wishart distribution, then S has a density with respect to the measure $\nu(dS) = dS/|S|^{(p+1)/2}$ given by

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

Here, $\omega(p, n)$ is the Wishart constant defined in Example 5.1.

Proof. Represent the $W(\Sigma, p, n)$ distribution as $\mathcal{L}(AS_1A')$ where $\mathcal{L}(S_1) = W(I_p, p, n)$ and $AA' = \Sigma$. Obviously, the rank of A is the rank of Σ and $\Sigma > 0$ iff rank of Σ is p. If $n < p$, then by Proposition 7.1, if $\mathcal{L}(X_i) = N(0, I_p)$, $i = 1, \ldots, n$, the rank of $\Sigma_1^n X_i X_i'$ is n with probability one. Thus $S_1 = \Sigma_1^n X_i X_i'$ has rank n, which is less than p, and $S = AS_1A'$ has rank less than p with probability one. Also, if the rank of Σ is $r < p$, then A has rank r so AS_1A' has rank at most r no matter what n happens to be. Therefore, if $n < p$ or if Σ is singular, then S is singular with probability one. Now, consider the case when $n \geqslant p$ and Σ is positive definite. Then $S_1 = \Sigma_1^n X_i X_i'$ has rank p with probability one by Proposition 7.1, and A has rank p. Therefore, $S = AS_1A'$ has rank p with probability one.

When $\Sigma > 0$, the density of $X \in \mathcal{L}_{p,n}$ is

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

when $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$. When $n \geqslant p$, it follows from Proposition 7.6 that the density of S with respect to $\nu(dS)$ is $p(S|\Sigma)$. □

Recall that the natural inner product on $\mathcal{S}_p$, when $\mathcal{S}_p$ is regarded as a subspace of $\mathcal{L}_{p,p}$, is

$$\langle S_1, S_2 \rangle = \operatorname{tr} S_1 S_2, \qquad S_i \in \mathcal{S}_p, \quad i = 1,2.$$

The mean vector, covariance, and characteristic function of a Wishart distribution on the inner product space $(\mathcal{S}_p, \langle \cdot, \cdot \rangle)$ are given next.

Proposition 8.3. Suppose $\mathcal{L}(S) = W(\Sigma, p, n)$ on $(\mathcal{S}_p, \langle \cdot, \cdot \rangle)$. Then

(i) $\mathcal{E}S = n\Sigma$.
(ii) $\operatorname{Cov}(S) = 2n\Sigma \otimes \Sigma$.
(iii) $\phi(A) \equiv \mathcal{E}\exp[i\langle A, S\rangle] = |I_p - 2i\Sigma A|^{-n/2}$.

Proof. To prove (i) write $S = \Sigma_1^n X_i X_i'$ where $\mathcal{L}(X_i) = N(0, \Sigma)$, and $X_1, \ldots, X_n$ are independent. Since $\mathcal{E}X_i X_i' = \Sigma$, it is clear that $\mathcal{E}S = n\Sigma$. For (ii), the independence of $X_1, \ldots, X_n$ implies that

$$\operatorname{Cov}(S) = \operatorname{Cov}\left(\sum_1^n X_i X_i'\right) = \sum_1^n \operatorname{Cov}(X_i X_i') = n \operatorname{Cov}(X_1 X_1')$$

$$= n \operatorname{Cov}(X_1 \Box X_1)$$

where $X_1 \Box X_1$ is the outer product of X_1 relative to the standard inner product on R^p. Since $\mathcal{L}(X_1) = \mathcal{L}(CZ)$ where $\mathcal{L}(Z) = N(0, I_p)$ and $CC' = \Sigma$, it follows from Proposition 2.24 that $\operatorname{Cov}(X_1 \Box X_1) = 2\Sigma \otimes \Sigma$. Thus (ii) holds. To establish (iii), first write $C'AC = \Gamma D \Gamma'$ where $A \in \mathcal{S}_p$, $CC' = \Sigma$, $\Gamma \in \mathcal{O}_n$, and D is a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_p$. Then

$$\phi(A) = \mathcal{E}\exp[i \operatorname{tr}(AS)] = \mathcal{E}\exp\left[i \operatorname{tr} A\left(\sum_1^n X_j X_j'\right)\right]$$

$$= \mathcal{E}\prod_{j=1}^n \exp\left[i \operatorname{tr} A X_j X_j'\right] = \prod_{j=1}^n \mathcal{E}\exp\left[i \operatorname{tr} A X_j X_j'\right]$$

$$[\mathcal{E}\exp[i \operatorname{tr} A X_1 X_1']]^n = [\mathcal{E}\exp[iX_1' A X_1]]^n \equiv (\xi(A))^n.$$

Again, $\mathcal{L}(X_1) = \mathcal{L}(CZ)$ where $\mathcal{L}(Z) = N(0, I_p)$. Also, $\mathcal{L}(\Gamma Z) = \mathcal{L}(Z)$ for

$\Gamma \in \mathcal{O}_p$. Therefore,

$$\begin{aligned}\xi(A) &= \mathcal{E}\exp[iX_1'AX_1] = \mathcal{E}\exp[iZ'C'ACZ]\\ &= \mathcal{E}\exp[iZ'DZ] = \mathcal{E}\exp\left[i\sum_{j=1}^{p}\lambda_j Z_j^2\right]\end{aligned}$$

where $Z_1, \ldots, Z_p$ are the coordinates of Z. Since $Z_1, \ldots, Z_p$ are independent with $\mathcal{L}(Z_j) = N(0, 1)$, Z_j^2 has a χ_1^2 distribution and we have

$$\begin{aligned}\xi(A) &= \prod_{j=1}^{p}\mathcal{E}\exp\left[i\lambda_j Z_j^2\right] = \prod_{j=1}^{p}\left(1 - 2i\lambda_j\right)^{-1/2}\\ &= |I_p - 2iD|^{-1/2} = |I_p - 2i\Gamma D\Gamma'|^{-1/2} = |I_p - 2iC'AC|^{-1/2}\\ &= |I_p - 2iCC'A|^{-1/2} = |I_p - 2i\Sigma A|^{-1/2}.\end{aligned}$$

The next to the last equality is a consequence of Proposition 1.35. Thus (iii) holds. □

Proposition 8.4. If $\mathcal{L}(S_i) = W(\Sigma, p, n_i)$ for $i = 1, 2$ and if S_1 and S_2 are independent, then $\mathcal{L}(S_1 + S_2) = W(\Sigma, p, n_1 + n_2)$.

Proof. An application of (iii) of Proposition 8.3 yields this convolution result. Specifically,

$$\begin{aligned}\phi(A) &= \mathcal{E}\exp[i\langle A, S_1 + S_2\rangle] = \prod_{j=1}^{2}\mathcal{E}\exp i\langle A, S_j\rangle\\ &= \prod_{j=1}^{2}|I_p - 2i\Sigma A|^{-n_j/2} = |I_p - 2i\Sigma A|^{-(n_1+n_2)/2}.\end{aligned}$$

The uniqueness of characteristic functions shows that $\mathcal{L}(S_1 + S_2) = W(\Sigma, p, n_1 + n_2)$. □

It should be emphasized that $\langle \cdot, \cdot \rangle$ is not what we might call the standard inner product on $\mathcal{S}_p$ when $\mathcal{S}_p$ is regarded as a $[p(p+1)/2]$-dimensional coordinate space. For example, if $p = 2$, and $S, T \in \mathcal{S}_p$, then

$$\langle S, T\rangle = \operatorname{tr} ST = s_{11}t_{11} + s_{22}t_{22} + 2s_{12}t_{12}$$

while the three-dimensional coordinate space inner product between S and

T would be $s_{11}t_{11} + s_{22}t_{22} + s_{12}t_{12}$. In this connection, equation (ii) of Proposition 8.3 means that

$$\begin{aligned}\operatorname{cov}(\langle A, S\rangle, \langle B, S\rangle) &= 2n\langle A, (\Sigma \otimes \Sigma)B\rangle \\ &= 2n\langle A, \Sigma B \Sigma\rangle = 2n \operatorname{tr}(A\Sigma B\Sigma),\end{aligned}$$

that is, (ii) depends on the inner product $\langle \cdot , \cdot\rangle$ on $\mathbb{S}_p$ and is not valid for other inner products.

In Chapter 3, quadratic forms in normal random vectors were shown to have chi-square distributions under certain conditions. Similar results are available for generalized quadratic forms and the Wishart distribution. The following proposition is not the most general possible, but suffices in most situations.

Proposition 8.5. Consider $X \in \mathcal{L}_{p,n}$ where $\mathcal{L}(X) = N(\mu, Q \otimes \Sigma)$. Let $S = X'PX$ where P is $n \times n$ and positive semidefinite, and write $P = A^2$ with A positive semidefinite. If AQA is a rank k orthogonal projection and if $P\mu = 0$, then

$$\mathcal{L}(S) = W(\Sigma, p, k).$$

Proof. With $Y = AX$, it is clear that $S = Y'Y$ and

$$\mathcal{L}(Y) = N(A\mu, (AQA) \otimes \Sigma).$$

Since $\mathfrak{N}(A) = \mathfrak{N}(P)$ and $P\mu = 0$, $A\mu = 0$ so

$$\mathcal{L}(Y) = N(0, (AQA) \otimes \Sigma).$$

By assumption, $B = AQA$ is a rank k orthogonal projection. Also, $S = Y'Y = Y'BY + Y'(I - B)Y$, and $\mathcal{L}((I - B)Y) = N(0, 0 \otimes \Sigma)$ so $Y'(I - B)Y$ is zero with probability one. Thus it remains to show that if $\mathcal{L}(Y) = N(0, B \otimes \Sigma)$ where B is a rank k orthogonal projection, then $S = Y'BY$ has a $W(\Sigma, p, k)$ distribution. Without loss of generality (make an orthogonal transformation),

$$B = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} : n \times n.$$

Partitioning Y into $Y_1 : k \times p$ and $Y_2 : (n - k) \times p$, it follows that $S = Y_1'Y_1$

and

$$\mathcal{L}(Y_1) = N(0, I_k \otimes \Sigma).$$

Thus $\mathcal{L}(S) = W(\Sigma, p, k)$. □

◆ **Example 8.1.** We again return to the multivariate normal linear model introduced in Example 4.4. Consider $X \in \mathcal{L}_{p,n}$ with

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

where μ is an element of the subspace $M \subseteq \mathcal{L}_{p,n}$ defined by

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

Here, Z is an $n \times k$ matrix of rank k and it is assumed that $n - k \geqslant p$. With $P_z = Z(Z'Z)^{-1}Z'$, $P_M = P_z \otimes I_p$ is the orthogonal projection onto M and $Q_M = Q_z \otimes I_p$, $Q_z = I - P_z$, is the orthogonal projection onto $M^{\perp}$. We know that

$$\hat{\mu} = P_M X = (P_z \otimes I_p)X = P_z X$$

is the maximum likelihood estimator of μ. As demonstrated in Example 4.4, the maximum likelihood estimator of Σ is found by maximizing

$$p(x|\hat{\mu}, \Sigma) = |\Sigma|^{-n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}x'Q_z x\right].$$

Since $n - k \geqslant p$, $x'Q_z x$ has rank p with probability one. When $X'Q_z X$ has rank p, Example 7.10 shows that

$$\hat{\Sigma} = \frac{1}{n}X'Q_z X$$

is the maximum likelihood estimator of Σ. The conditions of Proposition 8.5 are easily checked to verify that $S \equiv X'Q_z X$ has a $W(\Sigma, p, n-k)$ distribution. In summary, for the multivariate linear model, $\hat{\mu} = P_M X$ and $\hat{\Sigma} = n^{-1}X'Q_z X$ are the maximum likelihood estimators of μ and Σ. Further, $\hat{\mu}$ and $\hat{\Sigma}$ are independent and

$$\mathcal{L}(\hat{\mu}) = N(\mu, P_z \otimes \Sigma)$$

$$\mathcal{L}(n\hat{\Sigma}) = W(\Sigma, p, n-k).$$ ◆

8.2. PARTITIONING A WISHART MATRIX

The partitioning of the Wishart distribution considered here is motivated partly by the transformation described in Proposition 5.8. If $\mathcal{L}(S) = W(\Sigma, p, n)$ where $n \geqslant p$, partition S as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where $S_{21} = S'_{12}$ and let

$$S_{11\cdot 2} = S_{11} - S_{12}S_{22}^{-1}S_{21}.$$

Here, S_{ij} is $p_i \times p_j$ for $i, j = 1, 2$ so $p_1 + p_2 = p$. The primary result of this section describes the joint distribution of $(S_{11\cdot 2}, S_{21}, S_{22})$ when Σ is nonsingular. This joint distribution is derived by representing the Wishart distribution in terms of the normal distribution. Since $\mathcal{L}(S) = W(\Sigma, p, n)$, $S = X'X$ where $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$. Discarding a set of Lebesgue measure zero, X is assumed to take values in $\mathcal{X}$, the set of all $n \times p$ matrices of rank p. With

$$X = (X_1, X_2), \qquad X_i : n \times p_i, \quad i = 1, 2,$$

it is clear that

$$S_{ij} = X'_i X_j \qquad \text{for } i, j = 1, 2.$$

Thus

$$S_{11\cdot 2} = X'_1 X_1 - X'_1 X_2 (X'_2 X_2)^{-1} X'_2 X_1 = X'_1 Q X_1$$

where

$$Q = I_n - X_2 (X'_2 X_2)^{-1} X'_2 \equiv I_n - P$$

is an orthogonal projection of rank $n - p_2$ for each value of X_2 when $X \in \mathcal{X}$. To obtain the desired result for the Wishart distribution, it is useful to first give the joint distribution of (QX_1, PX_1, X_2).

Proposition 8.6. The joint distribution of (QX_1, PX_1, X_2) can be described as follows. Conditional on X_2, QX_1 and PX_1 are independent with

$$\mathcal{L}(QX_1 | X_2) = N(0, Q \otimes \Sigma_{11\cdot 2})$$

and

$$\mathcal{L}(PX_1|X_2) = N\left(X_2\Sigma_{22}^{-1}\Sigma_{21}, P \otimes \Sigma_{11\cdot 2}\right).$$

Also,

$$\mathcal{L}(X_2) = N(0, I_n \otimes \Sigma_{22}).$$

Proof. From Example 3.1, the conditional distribution of X_1 given X_2, say $\mathcal{L}(X_1|X_2)$, is

$$\mathcal{L}(X_1|X_2) = N\left(X_2\Sigma_{22}^{-1}\Sigma_{21}, I_n \otimes \Sigma_{11\cdot 2}\right).$$

Thus conditional on X_2, the random vector

$$W \equiv \begin{pmatrix} Q \otimes I_{p_1} \\ P \otimes I_{p_1} \end{pmatrix} X_1 = \begin{pmatrix} QX_1 \\ PX_1 \end{pmatrix} : (2n) \times p_1$$

is a linear transformation of X_1. Thus W has a normal distribution with mean vector

$$\begin{pmatrix} Q \otimes I_{p_1} \\ P \otimes I_{p_1} \end{pmatrix} X_2\Sigma_{22}^{-1}\Sigma_{21} = \begin{pmatrix} 0 \\ X_2\Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix}$$

since $QX_2 = 0$ and $PX_2 = X_2$. Also, using the calculational rules for partitioned linear transformations, the covariance of W is

$$\begin{pmatrix} Q \otimes I_{p_1} \\ P \otimes I_{p_1} \end{pmatrix} (I_n \otimes \Sigma_{11\cdot 2})(Q \otimes I_{p_1}, P \otimes I_{p_1}) = \begin{pmatrix} Q & 0 \\ 0 & P \end{pmatrix} \otimes \Sigma_{11\cdot 2}$$

since $QP = 0$. The conditional independence and conditional distribution of QX_1 and PX_1 follow immediately. That X_2 has the claimed marginal distribution is obvious. □

Proposition 8.7. Suppose $\mathcal{L}(S) = W(\Sigma, p, n)$ with $n \geqslant p$ and $\Sigma > 0$. Partition S into S_{ij}, $i, j = 1, 2$, where S_{ij} is $p_i \times p_j$, $p_1 + p_2 = p$, and partition Σ similarly. With $S_{11\cdot 2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$, $S_{11\cdot 2}$ and (S_{21}, S_{22}) are stochastically independent. Further,

$$\mathcal{L}(S_{11\cdot 2}) = W(\Sigma_{11\cdot 2}, p_1, n - p_2)$$

and conditional on S_{22},

$$\mathcal{L}(S_{21}|S_{22}) = N\left(S_{22}\Sigma_{22}^{-1}\Sigma_{21}, S_{22} \otimes \Sigma_{11\cdot 2}\right).$$

The marginal distribution of S_{22} is $W(\Sigma_{22}, p_2, n)$.

Proof. In the notation of Proposition 8.6, consider $X \in \mathfrak{X}$ with $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$ and $S = X'X$. Then $S_{ij} = X_i'X_j$ for $i, j = 1, 2$ and $S_{11\cdot 2} = X_1'QX_1$. Since $PX_2 = X_2$ and $S_{21} = X_2'X_1$, we see that $S_{21} = (PX_2)'X_1 = X_2'PX_1$, and conditional on X_2,

$$\mathcal{L}(S_{21}|X_2) = N\left(X_2'X_2\Sigma_{22}^{-1}\Sigma_{21}, (X_2'X_2) \otimes \Sigma_{11\cdot 2}\right).$$

To show that $S_{11\cdot 2}$ and (S_{21}, S_{22}) are independent, it suffices to show that

$$\mathcal{E}f(S_{11\cdot 2})h(S_{21}, S_{22}) = \mathcal{E}f(S_{11\cdot 2})\mathcal{E}h(S_{21}, S_{22})$$

for bounded measurable functions f and h with the appropriate domains of definition. Using Proposition 8.6, we argue as follows. For fixed X_2, QX_1 and PX_1 are independent so $S_{11\cdot 2} = X_1'QQX_1$ and $S_{21} = X_2'PX_1$ are conditionally independent. Also,

$$\mathcal{L}(QX_1|X_2) = N(0, Q \otimes \Sigma_{11\cdot 2})$$

and Q is a rank $n - p_2$ orthogonal projection. By Proposition 8.5, $\mathcal{L}(X_1'QX_1|X_2) = W(\Sigma_{11\cdot 2}, p_1, n - p_2)$ for each X_2 so $X_1'QX_1$ and X_2 are independent. Conditioning on X_2, we have

$$\begin{aligned}
\mathcal{E}f(S_{11\cdot 2})h(S_{21}, S_{22}) &= \mathcal{E}\mathcal{E}\left[f(X_1'QX_1)h(X_2'PX_1, X_2'X_2)|X_2\right] \\
&= \mathcal{E}\left[\mathcal{E}(f(X_1'QX_1)|X_2)\mathcal{E}(h(X_2'PX_1, X_2'X_2)|X_2)\right] \\
&= \mathcal{E}\left[\mathcal{E}f(X_1'QX_1)\mathcal{E}(h(X_2'PX_1, X_2'X_2)|X_2)\right] \\
&= \mathcal{E}f(X_1'QX_1)\mathcal{E}\left[\mathcal{E}(h(X_2'PX_1, X_2'X_2)|X_2)\right] \\
&= \mathcal{E}f(X_1'QX_1)\mathcal{E}h(X_2'PX_1, X_2'X_2) \\
&= \mathcal{E}f(S_{11\cdot 2})\mathcal{E}h(S_{21}, S_{22}).
\end{aligned}$$

Therefore, $S_{11\cdot 2}$ and (S_{21}, S_{22}) are stochastically independent. To describe the joint distribution of S_{21} and S_{22}, again condition on X_2. Then

$$\mathcal{L}(S_{21}|X_2) = N\left(X_2'X_2\Sigma_{22}^{-1}\Sigma_{21}, (X_2'X_2) \otimes \Sigma_{11\cdot 2}\right)$$

and this conditional distribution depends on X_2 only through $S_{22} = X_2'X_2$. Thus

$$\mathcal{L}(S_{21}|S_{22}) = N\left(S_{22}\Sigma_{22}^{-1}\Sigma_{21}, S_{22} \otimes \Sigma_{11\cdot 2}\right).$$

That S_{22} has the claimed marginal distribution is obvious. □

By simply permuting the indices in Proposition 8.7, we obtain the following proposition.

Proposition 8.8. With the notation and assumptions of Proposition 8.7, let $S_{22\cdot 1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$. Then $S_{22\cdot 1}$ and (S_{11}, S_{12}) are stochastically independent and

$$\mathcal{L}(S_{22\cdot 1}) = W(\Sigma_{22\cdot 1}, p_2, n - p_1).$$

Conditional on S_{11},

$$\mathcal{L}(S_{12}|S_{11}) = N\left(S_{11}\Sigma_{11}^{-1}\Sigma_{12}, S_{11} \otimes \Sigma_{22\cdot 1}\right)$$

and the marginal distribution of S_{11} is $W(\Sigma_{11}, p_1, n)$.

Proposition 8.7 is one of the most useful results for deriving distributions of functions of Wishart matrices. Applications occur in this and the remaining chapters. For example, the following assertion provides a simple proof of the distribution of Hotelling's-T^2, discussed in the next chapter.

Proposition 8.9. Suppose S_0 has a nonsingular Wishart distribution, say $W(\Sigma, p, n)$, and let A be an $r \times p$ matrix of rank r. Then

$$\mathcal{L}\left(\left(AS_0^{-1}A'\right)^{-1}\right) = W\left(\left(A\Sigma^{-1}A'\right)^{-1}, r, n - p + r\right).$$

Proof. First, an invariance argument shows that it is sufficient to consider the case when $\Sigma = I$. More precisely, write $\Sigma = B^2$ with $B > 0$ and let $C = AB^{-1}$. With $S = B^{-1}S_0B^{-1}$, $\mathcal{L}(S) = W(I, p, n)$ and the assertion is that

$$\mathcal{L}\left((CS^{-1}C')^{-1}\right) = W\left((CC')^{-1}, r, n - p + r\right).$$

Now, let $\Psi = C'(CC')^{-1/2}$, so the assertion becomes

$$\mathcal{L}\left((\Psi' S^{-1}\Psi)^{-1}\right) = W(I_r, r, n-p+r).$$

However, Ψ is $p \times r$ and satisfies $\Psi'\Psi = I_r$—that is, Ψ is a linear isometry. Since $\mathcal{L}(\Gamma' S \Gamma) = \mathcal{L}(S)$ for all $\Gamma \in \mathcal{O}_p$,

$$\mathcal{L}\left((\Psi'\Gamma' S^{-1}\Gamma\Psi)^{-1}\right) = \mathcal{L}\left((\Psi' S^{-1}\Psi)^{-1}\right).$$

Choose Γ so that

$$\Gamma\Psi = \begin{pmatrix} I_r \\ 0 \end{pmatrix} : p \times r.$$

For this choice of Γ, the matrix $(\Psi'\Gamma' S^{-1}\Gamma\Psi)^{-1}$ is just the inverse of the $r \times r$ upper left corner of S^{-1}, and this matrix is

$$S_{11} - S_{12}S_{22}^{-1}S_{21} \equiv V$$

where V is $r \times r$. By Proposition 8.7,

$$\mathcal{L}(V) = W(I_r, r, n-p+r)$$

since $\mathcal{L}(S) = W(I, p, n)$. This establishes the assertion of the proposition. □

When $r = 1$ in Proposition 8.9, the matrix A' is nonzero vector, say $A' = a \in R^p$. In this case,

$$\mathcal{L}\left(\frac{a'\Sigma^{-1}a}{a'S^{-1}a}\right) = \chi^2_{n-p+1}$$

when $\mathcal{L}(S) = W(\Sigma, p, n)$. Another decomposition result for the Wishart distribution, which is sometimes useful, follows.

Lemma 8.10. Suppose S has a nonsingular Wishart distribution, say $\mathcal{L}(S) = W(\Sigma, p, n)$, and let $S = TT'$ where $T \in G_T^+$. Then the density of T with respect to the left invariant measure $\nu(dT) = dT/\prod t_{ii}^i$ is

$$p(T|\Sigma) = 2^p\omega(n, p)|\Sigma^{-1}TT'|^{n/2}\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}TT'\right].$$

If S and T are partitioned as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \qquad T = \begin{pmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{pmatrix}$$

where S_{ij} is $p_i \times p_j$, $p_1 + p_2 = p$, then $S_{11} = T_{11}T_{11}'$, $S_{12} = T_{11}T_{21}'$, and $S_{22\cdot 1} = T_{22}T_{22}'$. Further, the pair (T_{11}, T_{21}) is independent of T_{22} and

$$\mathcal{L}(T_{21}'|T_{11}) = N\big(T_{11}'\Sigma_{11}^{-1}\Sigma_{12}, I_{p_1} \otimes \Sigma_{22\cdot 1}\big).$$

Proof. The expression for the density of T is a consequence of Proposition 7.5, and a bit of algebra shows that $S_{11} = T_{11}T_{11}'$, $S_{12} = T_{11}T_{21}'$, and $S_{22\cdot 1} = T_{22}T_{22}'$. The independence of (T_{11}, T_{21}) and T_{22} follows from Proposition 8.8 and the fact that the mapping between S and T is one-to-one and onto. Also,

$$\mathcal{L}(S_{12}|S_{11}) = N\big(S_{11}\Sigma_{11}^{-1}\Sigma_{12}, S_{11} \otimes \Sigma_{22\cdot 1}\big).$$

Since S_{11} and T_{11} are one-to-one functions of each other and $S_{12} = T_{11}T_{21}'$,

$$\mathcal{L}(T_{11}T_{21}'|T_{11}) = N\big(T_{11}T_{11}'\Sigma_{11}^{-1}\Sigma_{12}, T_{11}T_{11}' \otimes \Sigma_{22\cdot 1}\big).$$

Thus

$$\mathcal{L}(T_{21}'|T_{11}) = N\big(T_{11}'\Sigma_{11}^{-1}\Sigma_{12}, I_{p_1} \otimes \Sigma_{22\cdot 1}\big),$$

as

$$T_{21}' = \big(T_{11}^{-1} \otimes I_{p_2}\big)(T_{11}T_{21}')$$

and T_{11} is fixed. □

Proposition 8.11. Suppose S has a nonsingular Wishart distribution with $\mathcal{L}(S) = W(\Sigma, p, n)$ and assume that Σ is diagonal with diagonal elements $\sigma_{11}, \dots, \sigma_{pp}$. If $S = TT'$ with $T \in G_T^+$, then the random variables $\{t_{ij}|i \geqslant j\}$ are mutually independent and

$$\mathcal{L}(t_{ij}) = N(0, \sigma_{ii}) \qquad \text{for } i > j$$

and

$$\mathcal{L}(t_{ii}^2) = \sigma_{ii}\chi_{n-i+1}^2, \qquad i = 1, \dots, p.$$

Proof. First, partition S, Σ, and T as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}, \qquad T = \begin{pmatrix} t_{11} & 0 \\ T_{21} & T_{22} \end{pmatrix}$$

where S_{11} is 1×1. Since $\Sigma_{12} = 0$, the conditional distribution of T_{21} given T_{11} does not depend on T_{11} and Σ_{22} has diagonal elements $\sigma_{22}, \ldots, \sigma_{pp}$. It follows from Proposition 8.10 that t_{11}, T'_{21}, and T_{22} are mutually independent and

$$\mathcal{L}(T'_{21}) = N(0, \Sigma_{22}).$$

The elements of T_{21} are $t_{21}, t_{31}, \ldots, t_{p1}$, and since Σ_{22} is diagonal, these are independent with

$$\mathcal{L}(t_{i1}) = N(0, \sigma_{ii}), \qquad i = 2, \ldots, p.$$

Also,

$$\mathcal{L}(t_{11}^2) = \sigma_{11}\chi_n^2$$

and

$$\mathcal{L}(S_{22\cdot 1}) = \mathcal{L}(T_{22}T'_{22}) = W(\Sigma_{22}, p-1, n-1).$$

The conclusion of the proposition follows by an induction argument on the dimension parameter p. □

When $\mathcal{L}(S) = W(\Sigma, p, n)$ is a nonsingular Wishart distribution, the random variable $|S|$ is called the *generalized variance*. The distribution of $|S|$ is easily derived using Proposition 8.11. First, write $\Sigma = B^2$ with $B > 0$ and let $S_1 = B^{-1}SB^{-1}$. Then $\mathcal{L}(S_1) = W(I, p, n)$ and $|S| = |\Sigma||S_1|$. Also, if $TT' = S_1$, $T \in G_T^+$, then $\mathcal{L}(t_{ii}^2) = \chi^2_{n-i+1}$ for $i = 1, \ldots, p$, and $t_{11}, \ldots, t_{pp}$ are mutually independent. Thus

$$\mathcal{L}(|S|) = \mathcal{L}(|\Sigma||S_1|) = \mathcal{L}(|\Sigma||TT'|) = \mathcal{L}\left(|\Sigma|\prod_{i=1}^{p} t_{ii}^2\right).$$

Therefore, the distribution of $|S|$ is the same as the constant $|\Sigma|$ times a product of p independent chi-square random variables with $n - i + 1$ degrees of freedom for $i = 1, \ldots, p$.

8.3. THE NONCENTRAL WISHART DISTRIBUTION

Just as the Wishart distribution is a matrix generalization of the chi-square distribution, the noncentral Wishart distribution is a matrix analog of the noncentral chi-square distribution. Also, the noncentral Wishart distribution arises in a natural way in the study of distributional properties of test statistics in multivariate analysis.

Definition 8.2. Let $X \in \mathcal{L}_{p,n}$ have a normal distribution $N(\mu, I_n \otimes \Sigma)$. A random matrix $S \in \mathcal{S}_p$ has a noncentral Wishart distribution with parameters Σ, p, n, and $\Delta \equiv \mu'\mu$ if $\mathcal{L}(S) = \mathcal{L}(X'X)$. In this case, we write $\mathcal{L}(S) = W(\Sigma, p, n; \Delta)$.

In this definition, it is not obvious that the distribution of $X'X$ depends on μ only through $\Delta = \mu'\mu$. However, an invariance argument establishes this. The group $\mathcal{O}_n$ acts on $\mathcal{L}_{p,n}$ by sending x into Γx for $x \in \mathcal{L}_{p,n}$ and $\Gamma \in \mathcal{O}_n$. A maximal invariant under this action is $x'x$. When $\mathcal{L}(X) = N(\mu, I_n \otimes \Sigma)$, $\mathcal{L}(\Gamma X) = N(\Gamma\mu, I_n \otimes \Sigma)$ and we know the distribution of $X'X$ depends only on a maximal invariant parameter. But the group action on the parameter space is $(\mu, \Sigma) \to (\Gamma\mu, \Sigma)$ and a maximal invariant is obviously $(\mu'\mu, \Sigma)$. Thus the distribution of $X'X$ depends only on $(\mu'\mu, \Sigma)$.

When $\Delta = 0$, the noncentral Wishart distribution is just the $W(\Sigma, p, n)$ distribution. Let $X_1', \ldots, X_n'$ be the rows of X in the above definition so $X_1, \ldots, X_n$ are independent and $\mathcal{L}(X_i) = N(\mu_i, \Sigma)$ where $\mu_1', \ldots, \mu_n'$ are the rows of μ. Obviously,

$$\mathcal{L}(X_i X_i') = W(\Sigma, p, 1; \Delta_i)$$

where $\Delta_i = \mu_i \mu_i'$. Thus $S_i = X_i X_i'$, $i = 1, \ldots, n$, are independent and it is clear that, if $S = X'X$, then

$$\mathcal{L}(S) = \mathcal{L}\left(\sum_1^n S_i\right).$$

In other words, the noncentral Wishart distribution with n degrees of freedom can be represented as the convolution of n noncentral Wishart distributions each with one degree of freedom. This argument shows that, if $\mathcal{L}(S_i) = W(\Sigma, p, n_i; \Delta_i)$ for $i = 1, 2$ and if S_1 and S_2 are independent, then $\mathcal{L}(S_1 + S_2) = W(\Sigma, p, n_1 + n_2, \Delta_1 + \Delta_2)$. Since

$$\mathcal{E} X_i X_i' = \Sigma + \mu_i \mu_i',$$

it follows that

$$\mathcal{E}S = n\Sigma + \Delta$$

when $\mathcal{L}(S) = W(\Sigma, p, n; \Delta)$. Also,

$$\operatorname{Cov}(S) = \sum_1^n \operatorname{Cov}(S_i)$$

but an explicit expression for $\operatorname{Cov}(S_i)$ is not needed here. As with the central Wishart distribution, it is not difficult to prove that, when $\mathcal{L}(S) = W(\Sigma, p, n; \Delta)$, then S is positive definite with probability one iff $n \geqslant p$ and $\Sigma > 0$. Further, it is clear that if $\mathcal{L}(S) = W(\Sigma, p, n; \Delta)$ and A is an $r \times p$ matrix, then $\mathcal{L}(ASA') = W(A\Sigma A', r, n; A\Delta A')$. The next result provides an expression for the density function of S in a special case.

Proposition 8.12. Suppose $\mathcal{L}(S) = W(\Sigma, p, n; \Delta)$ where $n \geqslant p$ and $\Sigma > 0$, and assume that Δ has rank one, say $\Delta = \eta\eta'$ with $\eta \in R^p$. The density of S with respect to $\nu(dS) = dS/|S|^{(p+1)/2}$ is given by

$$p_1(S|\Sigma, \Delta) = p(S|\Sigma)\exp\left[-\tfrac{1}{2}\eta'\Sigma^{-1}\eta\right] H\left(\left(\eta'\Sigma^{-1}S\Sigma^{-1}\eta\right)^{1/2}\right)$$

where $p(S|\Sigma)$ is the density of a $W(\Sigma, p, n)$ distribution given in Proposition 8.2 and the function H is defined in Example 7.13.

Proof. Consider $X \in \mathcal{L}_{p,n}$ with $\mathcal{L}(X) = N(\mu, I_n \otimes \Sigma)$ where $\mu \in \mathcal{L}_{p,n}$ and $\mu'\mu = \Delta$. Since $S = X'X$ is a maximal invariant under the action of $\mathcal{O}_n$ on $\mathcal{L}_{p,n}$, the results of Example 7.15 show that the density of S with respect to the measure $\nu_0(dS) = (\sqrt{2\pi})^{np}\omega(n, p)|S|^{(n-p-1)/2}\, dS$ is

$$h(S) = \int_{\mathcal{O}_n} f(\Gamma X)\mu_0(d\Gamma).$$

Here, f is the density of X and μ_0 is the unique invariant probability measure on $\mathcal{O}_n$. The density of X is

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

Substituting this into the expression for $h(S)$ and doing a bit of algebra shows that the density $p_1(S|\Sigma, \Delta)$ with respect to ν is

$$p_1(S|\Sigma, \Delta) = p(S|\Sigma)\exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\Delta\right]\int_{\mathcal{O}_n}\exp\left[\operatorname{tr}\Gamma X\Sigma^{-1}\mu'\right]\mu_0(d\Gamma).$$

The problem is now to evaluate the integral over $\mathcal{O}_n$. It is here where we use the assumption that Δ has rank one. Since $\Delta = \mu'\mu$, μ must have rank one so $\mu = \xi\eta'$ where $\xi \in R^n$, $\|\xi\| = 1$, and $\eta \in R^p$, $\Delta = \eta\eta'$. Since $\|\xi\| = 1$, $\xi = \Gamma_1\varepsilon_1$ for some $\Gamma_1 \in \mathcal{O}_n$ where $\varepsilon_1 \in R^n$ is the first unit vector. Setting $u = (\eta'\Sigma^{-1}S\Sigma^{-1}\eta)^{1/2}$, $X\Sigma^{-1}\eta = u\Gamma_2\varepsilon_1$ for some $\Gamma_2 \in \mathcal{O}_n$ as $u\varepsilon_1$ and $X\Sigma^{-1}\eta$ have the same length. Therefore,

$$\int_{\mathcal{O}_n} \exp[\operatorname{tr} \Gamma X\Sigma^{-1}\mu']\mu_0(d\Gamma) = \int_{\mathcal{O}_n} \exp[\operatorname{tr} \Gamma X\Sigma^{-1}\eta\xi']\mu_0(d\Gamma)$$

$$= \int_{\mathcal{O}_n} \exp[\xi'\Gamma X\Sigma^{-1}\eta]\mu_0(d\Gamma) = \int_{\mathcal{O}_n} \exp[u\varepsilon_1'\Gamma_1'\Gamma\Gamma_2\varepsilon_1]\mu_0(d\Gamma)$$

$$= \int_{\mathcal{O}_n} \exp[u\varepsilon_1'\Gamma\varepsilon_1]\mu_0(d\Gamma) = \int_{\mathcal{O}_n} \exp[u\gamma_{11}]\mu_0(d\Gamma) \equiv H(u).$$

The right and left invariance of μ_0 was used in the third to the last equality and γ_{11} is the (1, 1) element of Γ. The function H was evaluated in Example 7.13. Therefore, when $\Delta = \eta\eta'$,

$$p_1(S|\Sigma, \Delta) = p(S|\Sigma)\exp\left[-\tfrac{1}{2}\eta'\Sigma^{-1}\eta\right] \times H\left(\left(\eta'\Sigma^{-1}S\Sigma^{-1}\eta\right)^{1/2}\right). \qquad \square$$

The final result of this section is the analog of Proposition 8.5 for the noncentral Wishart distribution.

Proposition 8.13. Consider $X \in \mathcal{L}_{p,n}$ where $\mathcal{L}(X) = N(\mu, Q \otimes \Sigma)$ and let $S = X'PX$ where $P \geqslant 0$ is $n \times n$. Write $P = A^2$ with $A \geqslant 0$. If $B \equiv AQA$ is a rank k orthogonal projection and if $AQP\mu = A\mu$, then

$$\mathcal{L}(S) = W(\Sigma, p, k; \mu'P\mu).$$

Proof. The proof of this result is quite similar to that of Proposition 8.5 and is left to the reader. $\square$

It should be noted that there is not an analog of Proposition 8.7 for the noncentral Wishart distribution, at least as far as I know. Certainly, Proposition 8.7 is false as stated when S is noncentral Wishart.

8.4. DISTRIBUTIONS RELATED TO LIKELIHOOD RATIO TESTS

In the next two chapters, statistics that are the ratio of determinants of Wishart matrices arise as tests statistics related to likelihood ratio tests.

Since the techniques for deriving the distributions of these statistics are intimately connected with properties of the Wishart distribution, we have chosen to treat this topic here rather than interrupt the flow of the succeeding chapters with such considerations.

Let $X \in \mathcal{L}_{p,m}$ and $S \in \mathcal{S}_p^+$ be independent and suppose that $\mathcal{L}(X) = N(\mu, I_m \otimes \Sigma)$ and $\mathcal{L}(S) = W(\Sigma, p, n)$ where $n \geqslant p$ and $\Sigma > 0$. We are interested in deriving the distribution of the random variable

$$U = \frac{|S|}{|S + X'X|}$$

for some special values of the mean matrix μ of X. The argument below shows that the distribution of U depends on (μ, Σ) only through $\Sigma^{-1/2}\mu'\mu\Sigma^{-1/2}$ where $\Sigma^{1/2}$ is the positive definite square root of Σ. Let $S = \Sigma^{1/2}S_1\Sigma^{1/2}$ and $Y = X\Sigma^{-1/2}$. Then S_1 and Y are independent, $\mathcal{L}(S_1) = W(I, p, n)$, and $\mathcal{L}(Y) = N(\mu\Sigma^{-1/2}, I_m \otimes I_p)$. Also,

$$U = \frac{|S|}{|S + X'X|} = \frac{|S_1|}{|S_1 + Y'Y|}.$$

However, the discussion of the previous section shows that $Y'Y$ has a noncentral Wishart distribution, say $\mathcal{L}(Y'Y) = W(I, p, n; \Delta)$ where $\Delta = \Sigma^{-1/2}\mu'\mu\Sigma^{-1/2}$. In the following discussion we take $\Sigma = I_p$ and denote the distribution of U by

$$\mathcal{L}(U) = U(n, m, p; \Delta)$$

where $\Delta = \mu'\mu$. When $\mu = 0$, the notation

$$\mathcal{L}(U) = U(n, m, p)$$

is used. In the case that $p = 1$,

$$U = \frac{S}{S + X'X}$$

where $\mathcal{L}(S) = \chi_n^2$. Since $\mathcal{L}(X) = N(\mu, I_m)$, $\mathcal{L}(X'X) = \chi_m^2(\Delta)$ where $\Delta = \mu'\mu \geqslant 0$. Thus

$$U = \frac{1}{1 + \chi_m^2(\Delta)/\chi_n^2}.$$

When $\chi_m^2(\Delta)$ and χ_n^2 are independent, the distribution of the ratio

$$F(m, n; \Delta) \equiv \frac{\chi_m^2(\Delta)}{\chi_n^2}$$

is called a noncentral F distribution with parameters m, n, and Δ. When $\Delta = 0$, the distribution of $F(m, n; 0)$ is denoted by $F_{m,n}$ and is simply called an F distribution with (m, n) degrees of freedom. It should be noted that this usage is not standard as the above ratio has not been normalized by the constant n/m. At times, the relationship between the F distribution and the beta distribution is useful. It is not difficult to show that, when χ_m^2 and χ_n^2 are independent, the random variable

$$V = \frac{\chi_n^2}{\chi_n^2 + \chi_m^2}$$

has a beta distribution with parameters $n/2$ and $m/2$, and this is written as $\mathcal{L}(V) = \mathcal{B}(n/2, m/2)$. In other words, V has a density on $(0, 1)$ given by

$$p(v) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} v^{\alpha-1}(1 - v)^{\beta-1}$$

where $\alpha = n/2$ and $\beta = m/2$. More generally, the distribution of the random variable

$$V(\Delta) = \frac{\chi_n^2}{\chi_n^2 + \chi_m^2(\Delta)}$$

is called a noncentral beta distribution and the notation $\mathcal{L}(V(\Delta)) = \mathcal{B}(n/2, m/2; \Delta)$ is used. In summary, when $p = 1$,

$$\mathcal{L}(U) = \mathcal{B}\left(\frac{n}{2}, \frac{m}{2}; \Delta\right)$$

where $\Delta = \mu'\mu \geqslant 0$.

Now, we consider the distribution of U when $m = 1$. In this case, $\mathcal{L}(X') = N(\mu', I_p)$ where $X' \in R^p$ and

$$U = \frac{|S|}{|S + X'X|} = |I_p + S^{-1}X'X|^{-1} = (1 + XS^{-1}X')^{-1}.$$

The last equality follows from Proposition 1.35.

Proposition 8.14. When $m = 1$,

$$\mathcal{L}(U) = \mathcal{B}\left(\frac{n-p+1}{2}, \frac{p}{2}; \delta\right)$$

where $\delta = \mu\mu' \geqslant 0$.

Proof. It must be shown that

$$\mathcal{L}(XS^{-1}X') = F(p, n-p+1, \delta).$$

For X fixed, $X \neq 0$, Proposition 8.10 shows that

$$\mathcal{L}\left(\frac{XX'}{XS^{-1}X'}\right) = \chi^2_{n-p+1}$$

when $\mathcal{L}(S) = W(I, p, n)$. Since this distribution does not depend on X, we have that $(XX')/XS^{-1}X'$ and XX' are independent. Further,

$$\mathcal{L}(XX') = \chi^2_p(\delta)$$

since $\mathcal{L}(X') = N(\mu', I_p)$. Thus

$$\mathcal{L}(XS^{-1}X') = \mathcal{L}\left(\frac{XS^{-1}X'}{XX'}XX'\right) = F(n-p+1, p; \delta). \qquad \square$$

The next step in studying $\mathcal{L}(U)$ is the case when $m > 1$, $p > 1$, but $\mu = 0$.

Proposition 8.15. Suppose X and S are independent where $\mathcal{L}(S) = W(I, p, n)$ and $\mathcal{L}(X) = N(0, I_m \otimes I_p)$. Then

$$\mathcal{L}(U) = \mathcal{L}\left(\prod_1^m U_i\right)$$

where $U_1, \ldots, U_m$ are independent and $\mathcal{L}(U_i) = \mathcal{B}((n-p+i)/2, p/2)$.

Proof. The proof is by induction on m and, when $m = 1$, we know

$$\mathcal{L}(U) = \mathcal{B}((n-p+1)/2, p/2).$$

Since $X'X = \Sigma_1^m X_i X_i'$ where X has rows $X_1', \dots, X_m'$,

$$U = \frac{|S|}{|S + X'X|} = \frac{|S|}{|S + X_1 X_1'|} \times \frac{|S + X_1 X_1'|}{|S + X_1 X_1' + \Sigma_2^m X_i X_i'|}.$$

The first claim is that

$$U_1 \equiv \frac{|S|}{|S + X_1 X_1'|}$$

and

$$W = \frac{|S + X_1 X_1'|}{|S + X_1 X_1' + \Sigma_2^m X_i X_i'|}$$

are independent random variables. Since $X_1, \dots, X_m$ are independent and independent of S, to show U_1 and W are independent, it suffices to show that U_1 and $S + X_1 X_1'$ are independent. To do this, Proposition 7.19 is applicable. The group Gl_p acts on (S, X_1) by

$$A(S, X_1) = (ASA', AX_1)$$

and the induced group action on $T = S + X_1 X_1'$ sends T into ATA'. The induced group action is clearly transitive. Obviously, T is an equivariant function and also U_1 is an invariant function under the group action on (S, X_1). That T is a sufficient statistic for the parametric family generated by Gl_p and the fixed joint distribution of (S, X_1) is easily checked via the factorization criterion. By Proposition 7.19, U_1 and $S + X_1 X_1'$ are independent. Therefore,

$$\mathcal{L}(U) = \mathcal{L}(U_1 W)$$

where U_1 and W are independent and

$$\mathcal{L}(U_1) = \mathcal{B}\left(\frac{n - p + 1}{2}, \frac{p}{2}\right).$$

However, $\mathcal{L}(S + X_1 X_1') = W(I, p, n + 1)$ and the induction hypothesis applied to W yields

$$\mathcal{L}(W) = \mathcal{L}\left(\prod_{i=1}^{m-1} W_i\right)$$

where $W_1, \ldots, W_{m-1}$ are independent with

$$\mathcal{L}(W_i) = \mathcal{L}\left(\frac{n+1-p+i}{2}, \frac{p}{2}\right).$$

Setting $U_i = W_{i-1}$, $i = 2, \ldots, m$, we have

$$\mathcal{L}(U) = \mathcal{L}\left(\prod_{i=1}^{m} U_i\right)$$

where $U_1, \ldots, U_m$ are independent and

$$\mathcal{L}(U_i) = \mathfrak{B}\left(\frac{n-p+i}{2}, \frac{p}{2}\right).$$ □

The above proof shows that U_i's are given by

$$U_i = \frac{|S + \Sigma_{j=1}^{i-1} X_j X_j'|}{|S + \Sigma_{j=1}^{i} X_j X_j'|}, \qquad i = 1, \ldots, m$$

and that these random variables are independent. Since $\mathcal{L}(S + \Sigma_1^{i-1} X_j X_j') = W(I, p, n+i-1)$, Proposition 8.14 yields

$$\mathcal{L}(U_i) = \mathfrak{B}\left(\frac{n-p+i}{2}, \frac{p}{2}\right).$$

In the special case that Δ has rank one, the distribution of U can be derived by an argument similar to that in the proof of Proposition 8.15.

Proposition 8.16. Suppose X and S are independent where $\mathcal{L}(S) = W(I, p, n)$ and $\mathcal{L}(X) = N(\mu, I_m \otimes I_p)$. Assume that $\mu = \xi\eta'$ with $\xi \in R^m$, $\|\xi\| = 1$, and $\eta \in R^p$. Then

$$\mathcal{L}(U) = \mathcal{L}\left(\prod_1^m U_i\right)$$

where $U_1, \ldots, U_m$ are independent,

$$\mathcal{L}(U_i) = \mathfrak{B}\left(\frac{n-p+i}{2}, \frac{p}{2}\right), \qquad i = 1, \ldots, m-1$$

and

$$\mathcal{L}(U_m) = \mathfrak{B}\left(\frac{n-p+m}{2}, \frac{p}{2}; \eta'\eta\right).$$

Proof. Let ε_m be the mth standard unit in R^m. Then $\Gamma\xi = \varepsilon_m$ for some $\Gamma \in \mathcal{O}_m$ as $\|\xi\| = \|\varepsilon_m\|$. Since

$$U = \frac{|S|}{|S + X'X|} = \frac{|S|}{|S + X'\Gamma'\Gamma X|}$$

and $\mathcal{L}(\Gamma X) = N(\varepsilon_m \eta', I_m \otimes I_p)$, we can take $\xi = \varepsilon_m$ without loss of generality. As in the proof of Proposition 8.15, $X'X = \Sigma_1^m X_i X_i'$ where $X_1, \dots, X_m$ are independent. Obviously, $\mathcal{L}(X_i) = N(0, I_p)$, $i = 1, \dots, m-1$, and $\mathcal{L}(X_m) = N(\eta, I_p)$. Now, write $U = \prod_1^m U_i$ where

$$U_i = \frac{|S + \Sigma_{j=1}^{i-1} X_j X_j'|}{|S + \Sigma_{j=1}^{i} X_j X_j'|}, \qquad i = 1, \dots, m.$$

The argument given in the proof of Proposition 8.15 shows that

$$U_1 = \frac{|S|}{|S + X_1 X_1'|}$$

and $\{S + X_1 X_1', X_2, \dots, X_m\}$ are independent. The assumption that X_1 has mean zero is essential here in order to verify the sufficiency condition necessary to apply Proposition 7.19. Since $U_2, \dots, U_m$ are functions of $\{S + X_1 X_1', X_2, \dots, X_m\}$, U_1 is independent of $\{U_2, \dots, U_m\}$. Now, we simply repeat this argument $m - 1$ times to conclude that $U_1, \dots, U_m$ are independent, keeping in mind that $X_1, \dots, X_{m-1}$ all have mean zero, but X_m need not have mean zero. As noted earlier,

$$\mathcal{L}(U_i) = \mathcal{B}\left(\frac{n - p + i}{2}, \frac{p}{2}\right); \qquad i = 1, \dots, m - 1.$$

By Proposition 8.14,

$$\mathcal{L}(U_m) = \mathcal{L}\left(\frac{|S + \Sigma_1^{m-1} X_i X_i'|}{|S + \Sigma_1^m X_i X_i'|}\right) = \mathcal{B}\left(\frac{n - p + m}{2}, \frac{p}{2}; \eta'\eta\right). \qquad \square$$

Now, we return to the case when $\mu = 0$. In terms of the notation $\mathcal{L}(U) = U(n, m, p)$, Proposition 8.14 asserts that

$$U(n, 1, p) = \mathcal{B}\left(\frac{n - p + 1}{2}, \frac{p}{2}\right).$$

Further, Proposition 8.15 can be written

$$U(n, m, p) = \prod_{i=1}^{m} U(n + i - 1, 1, p)$$

where this equation means that the distribution $U(n, m, p)$ can be represented as the distribution of the product of m independent random variables with distribution $U(n + i - 1, 1, p)$ for $i = 1, \ldots, m$. An alternative representation of $U(n, m, p)$ in terms of p independent random variables when $m \geqslant p$ follows. If $m \geqslant p$ and

$$U = \frac{|S|}{|S + X'X|}$$

with $\mathcal{L}(S) = W(I, p, n)$ and $\mathcal{L}(X) = N(0, I_m \otimes I_p)$, the matrix $T = X'X$ has a nonsingular Wishart distribution, $\mathcal{L}(T) = W(I, p, m)$. The following technical result provides the basic step for decomposing $U(n, m, p)$ into a product of p independent factors.

Proposition 8.17. Partition S into S_{ij} where S_{ij} is $p_i \times p_j$, $i, j = 1, 2$, and $p_1 + p_2 = p$. Partition T similarly and let

$$Z = S_{12}'S_{11}^{-1}S_{12} + T_{12}'T_{11}^{-1}T_{12} - (S_{12} + T_{12})'(S_{11} + T_{11})^{-1}(S_{12} + T_{12}).$$

Then the five random vectors S_{11}, T_{11}, $S_{22\cdot 1}$, $T_{22\cdot 1}$, and Z are mutually independent. Further,

$$\mathcal{L}(Z) = W(I, p_2, p_1).$$

Proof. Since S and T are independent by assumption, $(S_{11}, S_{12}, S_{22\cdot 1})$ and $(T_{11}, T_{12}, T_{22\cdot 1})$ are independent. Also, Proposition 8.8 shows that (S_{11}, S_{12}) and $S_{22\cdot 1}$ are independent with

$$\mathcal{L}(S_{22\cdot 1}) = W(I, p_2, n - p_1),$$

$$\mathcal{L}(S_{12}|S_{11}) = N(0, S_{11} \otimes I_{p_2}),$$

and

$$\mathcal{L}(S_{11}) = W(I, p_1, n).$$

Similar remarks hold for (T_{11}, T_{12}) and $T_{22\cdot 1}$ with n replaced by m. Thus the four random vectors (S_{11}, S_{12}), $S_{22\cdot 1}$, (T_{11}, T_{12}), and $T_{22\cdot 1}$ are mutually independent. Since Z is a function of (S_{11}, S_{12}) and (T_{11}, T_{12}), the proposition follows if we show that Z is independent of the vector (S_{11}, T_{11}). Conditional on (S_{11}, T_{11}),

$$\mathcal{L}\left(\left(\begin{matrix} S_{12} \\ T_{12} \end{matrix}\right)\Big|(S_{11}, T_{11})\right) = N\left(0, \left(\begin{matrix} S_{11} & 0 \\ 0 & T_{11} \end{matrix}\right) \otimes I_{p_2}\right).$$

Let $A(B)$ be the positive definite square root of $S_{11}(T_{11})$. With $V = A^{-1}S_{12}$ and $W = B^{-1}T_{12}$,

$$\mathcal{L}\left(\left[\begin{matrix} V \\ W \end{matrix}\right]\Big|(S_{11}, T_{11})\right) = N(0, I_{2p_1} \otimes I_{p_2}).$$

Also,

$$\begin{aligned} Z &= S_{12}'S_{11}^{-1}S_{12} + T_{12}'T_{11}^{-1}T_{12} - (S_{12} + T_{12})'(S_{11} + T_{11})^{-1}(S_{12} + T_{12}) \\ &= \left[\begin{matrix} V \\ W \end{matrix}\right]'\left[\begin{matrix} V \\ W \end{matrix}\right] - \left[\begin{matrix} V \\ W \end{matrix}\right]'\left[\begin{matrix} A \\ B \end{matrix}\right](A^2 + B^2)^{-1}\left[\begin{matrix} A \\ B \end{matrix}\right]'\left[\begin{matrix} V \\ W \end{matrix}\right] = \left[\begin{matrix} V \\ W \end{matrix}\right]'Q\left[\begin{matrix} V \\ W \end{matrix}\right] \end{aligned}$$

where

$$Q = I_{2p_1} - \left[\begin{matrix} A \\ B \end{matrix}\right](A^2 + B^2)^{-1}\left[\begin{matrix} A \\ B \end{matrix}\right]'.$$

However, Q is easily shown to be an orthogonal projection of rank p_1. By Proposition 8.5,

$$\mathcal{L}(Z|(S_{11}, T_{11})) = W(I, p_2, p_1)$$

for each value of (S_{11}, T_{11}). Therefore, Z is independent of (S_{11}, T_{11}) and the proof is complete. □

Proposition 8.18. If $m \geqslant p$, then

$$U(n, m, p) = \prod_{i=1}^{p} U(n - p + i, m, 1).$$

Proof. By definition,

$$U(n, m, p) = \mathcal{L}\left(\frac{|S|}{|S + T|}\right)$$

where S and T are independent, $\mathcal{L}(T) = W(I, p, m)$ and $\mathcal{L}(S) = W(I, p, n)$

with $n \geqslant p$. In the notation of Proposition 8.17, partition S and T with $p_1 = 1$ and $p_2 = p - 1$. Then S_{11}, T_{11}, $S_{22\cdot1}$, $T_{22\cdot1}$, and

$$Z = S_{12}'S_{22}^{-1}S_{12} + T_{12}'T_{11}^{-1}T_{12} - (S_{12} + T_{12})'(S_{11} + T_{11})^{-1}(S_{12} + T_{12})$$

are mutually independent. However,

$$|S| = |S_{11}||S_{22\cdot1}|$$

and

$$|S + T| = |S_{11} + T_{11}||(S + T)_{22\cdot1}| = |S_{11} + T_{11}||S_{22\cdot1} + T_{22\cdot1} + Z|.$$

Thus

$$\frac{|S|}{|S + T|} = \frac{|S_{11}|}{|S_{11} + T_{11}|} \times \frac{|S_{22\cdot1}|}{|S_{22\cdot1} + T_{22\cdot1} + Z|}$$

and the two factors on the right side of this equality are independent by Proposition 8.17. Obviously,

$$\mathcal{L}\left(\frac{|S_{11}|}{|S_{11} + T_{11}|}\right) = U(n, m, 1).$$

Since $\mathcal{L}(T_{22\cdot1}) = W(I, p - 1, m - 1)$, $\mathcal{L}(Z) = W(I, p - 1, 1)$, and $T_{22\cdot1}$ and Z are independent, it follows that

$$\mathcal{L}(T_{22\cdot1} + Z) = W(I, p - 1, m).$$

Therefore,

$$\mathcal{L}\left(\frac{|S_{22\cdot1}|}{|S_{22\cdot1} + T_{22\cdot1} + Z|}\right) = U(n - 1, m, p - 1),$$

which implies the relation

$$U(n, m, p) = U(n, m, 1)U(n - 1, m, p - 1).$$

Now, an easy induction argument establishes

$$U(n, m, p) = \prod_{i=1}^{p} U(n - i + 1, m, 1),$$

which implies that

$$U(n, m, p) = \prod_{i=1}^{p} U(n - p + i, m, 1)$$

and this completes the proof. □

Combining Propositions 8.15 and 8.18 leads to the following.

Proposition 8.19. If $m \geqslant p$, then

$$U(n, m, p) = U(n - p + m, p, m).$$

Proof. For arbitrary m, Proposition 8.15 yields

$$U(n, m, p) = \prod_{i=1}^{m} \mathfrak{B}\left(\frac{n - p + i}{2}, \frac{p}{2}\right)$$

where this notation means that the distribution $U(n, m, p)$ can be represented as the product of m independent beta-random variables with the factors in the product having a $\mathfrak{B}((n - p + i)/2, p/2)$ distribution. Since

$$U(n - p + i, m, 1) = \mathfrak{B}\left(\frac{n - p + i}{2}, \frac{m}{2}\right),$$

Proposition 8.18 implies that

$$U(n, m, p) = \prod_{i=1}^{p} U(n - p + i, m, 1) = \prod_{i=1}^{p} \mathfrak{B}\left(\frac{n - p + i}{2}, \frac{m}{2}\right).$$

Applying Proposition 8.15 to $U(n - p + m, p, m)$ yields

$$U(n - p + m, p, m) = \prod_{i=1}^{p} \mathfrak{B}\left(\frac{n - p + m - m + i}{2}, \frac{m}{2}\right)$$

$$= \prod_{i=1}^{p} \mathfrak{B}\left(\frac{n - p + i}{2}, \frac{m}{2}\right),$$

which is the distribution $U(n, m, p)$. □

In practice, the relationship $U(n, m, p) = U(n - p + m, p, m)$ shows that it is sufficient to deal with the case that $m \leqslant p$ when tabulating the distribution $U(n, m, p)$. Rather accurate approximations to the percentage points of the distribution $U(n, m, p)$ are available and these are discussed in detail in Anderson (1958, Chapter 8). This topic is not pursued further here.

PROBLEMS

1. Suppose S is $W(\Sigma, 2, n)$, $n \geqslant 2$, $\Sigma > 0$. Show that the density of $r = s_{12}/\sqrt{s_{11}s_{22}}$ can be written as

$$p(r|\rho) = \Gamma^2\left(\frac{n}{2}\right)2^n\omega(2, n)(1 - \rho^2)^{n/2}(1 - r^2)^{(n-1)/2}\psi(\rho r)$$

where $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ and ψ is defined as follows. Let X_1 and X_2 be independent chi-square random variables each with n degrees of freedom. Then $\psi(t) = \mathcal{E}\exp[t(X_1X_2)^{1/2}]$ for $|t| \leqslant 1$. Using this representation, prove that $p(r|\rho)$ has a monotone likelihood ratio.

2. The gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$, denoted by $G(\alpha, \lambda)$, has the density

$$p(x|\alpha, \lambda) = \frac{x^{\alpha-1}}{\lambda^\alpha\Gamma(\alpha)}\exp\left[\frac{-x}{\lambda}\right], \qquad x > 0$$

with respect to Lebesgue measure on $(0, \infty)$.

(i) Show the characteristic function of this distribution is $(1 - i\lambda t)^{-\alpha}$.

(ii) Show that a $G(n/2, 2)$ distribution is that of a χ^2_n distribution.

3. The above problem suggests that it is natural to view the gamma family as an extension of the chi-squared family by allowing nonintegral degrees of freedom. Since the $W(\Sigma, p, n)$ distribution is a generalization of the chi-squared distribution, it is reasonable to ask if we can define a Wishart distribution for nonintegral degrees of freedom. One way to pose this question is to ask for what values of α is $\phi_\alpha(A) = |I_p - 2iA|^\alpha$, $A \in \mathcal{S}_p$, a characteristic function. (We have taken $\Sigma = I_p$ for convenience).

(i) Using Proposition 8.3 and Problem 7.1, show that ϕ_α is a characteristic function for $\alpha = 1/2, \ldots, (p-1)/2$ and all real $\alpha > (p-1)/2$. Give the density that corresponds to ϕ_α for $\alpha > (p-1)/2$. $W(I_p, p, 2\alpha)$ denotes such a distribution.

(ii) For any $\Sigma \geqslant 0$ and the values of α given in (i), show that $\phi_\alpha(\Sigma A)$, $A \in \mathcal{S}_p$, is a characteristic function.

4. Let S be a random element of the inner product space $(\mathcal{S}_p, \langle \cdot, \cdot \rangle)$ where $\langle \cdot, \cdot \rangle$ is the usual trace inner product on $\mathcal{S}_p$. Say that S has an $\mathcal{O}_p$-invariant distribution if $\mathcal{L}(S) = \mathcal{L}(\Gamma S \Gamma')$ for each $\Gamma \in \mathcal{O}_p$. Assume S has an $\mathcal{O}_p$-invariant distribution.

(i) Assuming $\mathcal{E}S$ exists, show that $\mathcal{E}S = cI_p$ where $c = \mathcal{E}s_{11}$ and s_{ij} is the i, j element of S.

(ii) Let $D \in \mathcal{S}_p$ be diagonal with diagonal elements $d_1, \ldots, d_p$. Show that $\operatorname{var}(\langle D, S \rangle) = (\gamma - \beta)\Sigma d_i^2 + \beta(\Sigma_1^p d_i)^2$ where $\gamma = \operatorname{var}(s_{11})$ and $\beta = \operatorname{cov}\{s_{11}, s_{22}\}$.

(iii) For $A \in \mathcal{S}_p$, show that $\operatorname{var}(\langle A, S \rangle) = (\gamma - \beta)\langle A, A \rangle + \beta(I_p, A)^2$. From this conclude that $\operatorname{Cov}(S) = (\gamma - \beta)I_p \otimes I_p + \beta I_p \square I_p$.

5. Suppose $S \in \mathcal{S}_p^+$ has a density f with respect to Lebesgue measure dS restricted to $\mathcal{S}_p^+$. For each $n \geqslant p$, show there exists a random matrix $X \in \mathcal{L}_{p,n}$ that has a density with respect to Lebesgue measure on $\mathcal{L}_{p,n}$ and $\mathcal{L}(X'X) = \mathcal{L}(S)$.

6. Show that Proposition 8.4 holds for all n_1, n_2 equal to $1, 2, \ldots, p - 1$ or any real number greater than $p - 1$.

7. (The inverse Wishart distribution.) Say that a positive definite $S \in \mathcal{S}_p^+$ has an inverse Wishart distribution with parameters Λ, p, and ν if $\mathcal{L}(S^{-1}) = W(\Lambda^{-1}, p, \nu + p - 1)$. Here $\Lambda \in \mathcal{S}_p^+$ and ν is a positive integer. The notation $\mathcal{L}(S) = IW(\Lambda, p, \nu)$ signifies that $\mathcal{L}(S^{-1}) = W(\Lambda^{-1}, p, \nu + p - 1)$.

(i) If $\mathcal{L}(S) = IW(\Lambda, p, \nu)$ and A is $r \times p$ of rank r, show that $\mathcal{L}(ASA') = IW(A\Lambda A', r, \nu)$.

(ii) If $\mathcal{L}(S) = IW(I_p, p, \nu)$ and $\Gamma \in \mathcal{O}_p$, show that $\mathcal{L}(\Gamma S \Gamma') = \mathcal{L}(S)$.

(iii) If $\mathcal{L}(S) = IW(\Lambda, p, \nu)$, show that $\mathcal{E}(S) = (\nu - 2)^{-1}\Lambda$. Show that $\operatorname{Cov}(S)$ has the form $c_1 \Lambda \otimes \Lambda + c_2 \Lambda \square \Lambda$—what are c_1 and c_2?

(iv) Now, partition S into $S_{11} : q \times q$, $S_{12} : q \times r$, and $S_{22} : r \times r$ with S as in (iii). Show that $\mathcal{L}(S_{11}) = IW(\Lambda_{11}, q, \nu)$. Also show that $\mathcal{L}(S_{22 \cdot 1}) = IW(\Lambda_{22 \cdot 1}, r, \nu + q)$.

8. (The matric t distribution.) Suppose X is $N(0, I_r \otimes I_p)$ and S is $W(I_p, p, m)$, $m \geqslant p$. Let $S^{-1/2}$ denote the inverse of the positive definite square root of S. When S and X are independent, the matrix $T = XS^{-1/2}$ is said to have a matric t distribution and is denoted by $\mathcal{L}(T) = T(m - p + 1, I_r, I_p)$.

(i) Show that the density of T with respect to Lebesgue measure on $\mathcal{L}_{p,r}$ is given by

$$p(T) = \frac{\omega(m, p)}{(\sqrt{2\pi})^{rp}\omega(m+r, p)} \frac{1}{|I_p + T'T|^{(m+r)/2}}.$$

Also, show that $\mathcal{L}(T) = \mathcal{L}(\Gamma T \Delta')$ for $\Gamma \in \mathcal{O}_r$ and $\Delta \in \mathcal{O}_p$. Using this, show $\mathcal{E}T = 0$ and $\text{Cov}(T) = c_1 I_r \otimes I_p$ when these exist. Here, c_1 is a constant equal to the variance of any element of T.

(ii) Suppose V is $IW(I_p, p, \nu)$ and that T given V is $N(0, I_r \otimes V)$. Show that the unconditional distribution of T is $T(\nu, I_r, I_p)$.

(iii) Using Problem 7 and (ii), show that if T is $T(\nu, I_r, I_p)$, and T_{11} is the $k \times q$ upper left-hand corner of T, then T_{11} is $T(\nu, I_k, I_q)$.

9. (Multivariate F distribution.) Suppose S_1 is $W(I_p, p, m)$ (for $m = 1, 2, \ldots$) and is independent of S_2, which is $W(I_p, p, \nu + p - 1)$ (for $\nu = 1, 2, \ldots$). The matrix $F = S_2^{-1/2} S_1 S_2^{-1/2}$ has a matric F distribution that is denoted by $F(m, \nu, I_p)$.

(i) If S is $IW(I_p, p, \nu)$ and V given S is $W(S, p, m)$, show that the unconditional distribution of V is $F(m, \nu, I_p)$.

(ii) Suppose T is $T(\nu, I_r, I_p)$. Show that $T'T$ is $F(r, \nu, I_p)$.

(iii) When $r \geqslant p$, show that the $F(r, \nu, I_p)$ distribution has a density with respect to $dF/|F|^{(p+1)/2}$ given by

$$p(F) = \frac{\omega(r, p)\omega(\nu + p - 1, p)}{\omega(r + \nu + p - 1, p)} \frac{|F|^{r/2}}{|I_p + F|^{(\nu+p+r-1)/2}}.$$

(iv) For $r \geqslant p$, show that, if F is $F(r, \nu, I_p)$, then F^{-1} is $F(\nu + p - 1, r - p + 1, I_p)$.

(v) If F is $F(r, \nu, I_p)$ and F_{11} is the $q \times q$ upper left block of F, use (ii) to show that F_{11} is $F(r, \nu, I_q)$.

(vi) Suppose X is $N(0, I_r \otimes I_p)$ with $r \leqslant p$ and S is $W(I_p, p, m)$ with $m \geqslant p$, X and S independent. Show that $XS^{-1}X'$ is $F(p, m - p + 1, I_r)$.

10. (Multivariate beta distribution.) Let S_1 and S_2 be independent and suppose $\mathcal{L}(S_i) = W(I_p, p, m_i)$, $i = 1, 2$, with $m_1 + m_2 \geqslant p$. The random matrix $B = (S_1 + S_2)^{-1/2} S_1 (S_1 + S_2)^{-1/2}$ has a p-dimensional multivariate beta distribution with parameters m_1 and m_2. This is written $\mathcal{L}(B) = B(m_1, m_2, I_p)$ (when $p = 1$, this is the univariate beta distribution with parameters $m_1/2$ and $m_2/2$).

(i) If B is $B(m_1, m_2, I_p)$ show that $\mathcal{L}(\Gamma B \Gamma') = \mathcal{L}(B)$ for all $\Gamma \in \mathcal{O}_p$. Use Example 7.16 to conclude that $\mathcal{L}(B) = \mathcal{L}(\Psi D \Psi')$ where $\Psi \in \mathcal{O}_p$ is uniform and is independent of the diagonal matrix D with elements $\lambda_1 \geqslant \cdots \geqslant \lambda_p \geqslant 0$. The distribution of D is determined by specifying the distribution of $\lambda_1, \ldots, \lambda_p$ and this is the distribution of the ordered roots of $(S_1 + S_2)^{-1/2} S_1 (S_1 + S_2)^{-1/2}$.

(ii) With S_1 and S_2 as in the definition of B, show that $S_1^{1/2}(S_1 + S_2)^{-1} S_1^{1/2}$ is $B(m_1, m_2, I_p)$.

(iii) Suppose F is $F(m, \nu, I_p)$. Use (i) and (ii) to show that $(I + F)^{-1}$ is $B(p + \nu - 1, m, I_p)$ and $F(I + F)^{-1}$ is $B(m, p + \nu - 1, I_p)$.

(iv) Suppose that X is $N(0, I_r \otimes I_p)$ and that it is independent of S, which is $W(I_p, p, m)$. When $r \leqslant p$ and $m \geqslant p$, show that $X(S + X'X)^{-1}X'$ is $B(p, r + m - p, I_r)$.

(v) If B is $B(m_1, m_2, I_p)$ and $m_1 \geqslant p$, show that $\det(B)$ is distributed as $U(m_1, m_2, p)$ in the notation of Section 7.4.

NOTES AND REFERENCES

1. The Wishart distribution was first derived in Wishart (1928).

2. For some alternative discussions of the Wishart distribution, see Anderson (1958), Dempster (1969), Rao (1973), and Muirhead (1982).

3. The density function of the noncentral Wishart distribution in the general case is obtained by "evaluating"

$$\int_{\mathcal{O}_n} \exp[\operatorname{tr} \Gamma X \Sigma^{-1} \mu'] \mu_0(d\Gamma) \tag{8.1}$$

(see the proof of Proposition 8.12). The problem of evaluating

$$\psi(A) \equiv \int_{\mathcal{O}_n} \exp[\operatorname{tr} \Gamma A] \mu_0(d\Gamma)$$

for $A \in \mathcal{L}_{n,n}$ has received much attention since the paper of James (1954). Anderson (1946) first gave the noncentral Wishart density when

μ has rank 1 or rank 2. Much of the theory surrounding the evaluation of ψ and series expansions for ψ can be found in Muirhead (1982).

4. Wilks (1932) first proved Proposition 8.15 by calculating all the moments of U and showing these matched the moments of $\prod U_i$. Anderson (1958) also uses the moment method to find the distribution of U. This method was used by Box (1949) to provide asymptotic expansions for the distribution of U (see Anderson, 1958, Chapter 8).