STAT 542: MULTIVARIATE STATISTICAL ANALYSIS

1. Random Vectors and Covariance Matrices.

1.1. Review of vectors and matrices. (The results are stated for vectors and matrices with real entries but also hold for complex entries.)

An $m \times n$ matrix $A \equiv \{a_{ij}\}$ is an array of mn numbers:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

This matrix represents the *linear mapping* (≡ *linear transformation)*

$$\begin{aligned} A : \mathcal{R}^n &\to \mathcal{R}^m \\ x &\mapsto Ax, \end{aligned} \tag{1.1}$$

where $x \in \mathcal{R}^n$ is written as an $n \times 1$ column vector and

$$Ax \equiv \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \equiv \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix} \in \mathcal{R}^m.$$

Thus the jth column vector of A is the image $A\mathbf{u}_j$ of the jth coordinate column vector $\mathbf{u}_j$ (see (1.19)). The mapping (1.1) is clearly *linear:*

$$A(ax + by) = aAx + bAy.$$

Matrix addition: If $A \equiv \{a_{ij}\}$ and $B \equiv \{b_{ij}\}$ are $m \times n$ matrices, then

$$(A + B)_{ij} = a_{ij} + b_{ij}.$$

Matrix multiplication: If A is $m \times n$ and B is $n \times p$, then the *matrix product* AB is the $m \times p$ matrix AB whose ij-th element is

$$(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj}. \tag{1.2}$$

Then AB is the matrix of the composition $\mathcal{R}^p \overset{B}{\to} \mathcal{R}^n \overset{A}{\to} \mathcal{R}^m$ of the two linear mappings determined by A and B [verify]:

$$(AB)x = A(Bx) \qquad \forall x \in \mathcal{R}^p.$$

Rank of a matrix: The *row (column) rank* of a matrix $S : m \times n$ is the dimension of the linear space spanned by its rows (columns). The *rank* of A is the dimension r of the largest nonzero *minor* ($= r \times r$ subdeterminant) of A. Then [verify]

$$\begin{aligned}
\text{row rank}(A) &\leq \min(m, n), \\
\text{column rank}(A) &\leq \min(m, n), \\
\text{rank}(A) &\leq \min(m, n), \\
\text{row rank}(A) &= \text{column rank}(A) \\
= \text{rank}(A) &= \text{rank}(A') \\
= \text{rank}(AA') &= \text{rank}(A'A).
\end{aligned}$$

Furthermore, for $A : m \times n$ and $B : n \times p$,

$$\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B)).$$

Inverse matrix: If $A : n \times n$ is a square matrix, its *inverse* A^{-1} (if it exists) is the unique matrix that satisfies

$$AA^{-1} = A^{-1}A = I,$$

where $I \equiv I_n$ is the $n \times n$ identity matrix $\text{diag}(1, \ldots, 1)$. If A^{-1} exists then A is called *nonsingular* (or *regular*). The following are equivalent:

(a) A is nonsingular.

(b) The n columns of A are linearly independent (i.e., column $\text{rank}(A) = n$). Equivalently, $Ax \neq 0$ for every nonzero $x \in \mathcal{R}^n$.

(c) The n rows of A are linearly independent (i.e., row $\text{rank}(A) = n$). Equivalently, $x'A \neq 0$ for every nonzero $x \in \mathcal{R}^n$.

(d) The determinant $|A| \neq 0$ (i.e., $\text{rank}(A) = n$). [Define det geometrically.]

- If A is nonsingular then A^{-1} is nonsingular and $(A^{-1})^{-1} = A$.

- If $A : m \times m$ and $C : n \times n$ are nonsingular and B is $m \times n$, then [verify]

$$\text{rank}(AB) = \text{rank}(B) = \text{rank}(BC).$$

If $A : n \times n$ and $B : n \times n$ are nonsingular then so is AB, and [verify]

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{1.3}$$

If $A \equiv \mathrm{diag}(d_1, \ldots, d_n)$ with all $d_i \neq 0$ then $A^{-1} = \mathrm{diag}(d_1^{-1}, \ldots, d_n^{-1})$.

Transpose matrix: If $A \equiv \{a_{ij}\}$ is $m \times n$, its *transpose* is the $n \times m$ matrix A' (sometimes denoted by A') whose ij-th element is a_{ji}. That is, the m row vectors (n column vectors) of A are the m column vectors (n row vectors) of A'. Note that [verify]

$$(A + B)' = A' + B'; \tag{1.4}$$

$$(AB)' = B'A' \qquad (A : m \times n,\ B : n \times p); \tag{1.5}$$

$$(A^{-1})' = (A')^{-1} \qquad (A : n \times n,\ \text{nonsingular}). \tag{1.6}$$

Trace: For a square matrix $A \equiv \{a_{ij}\} : n \times n$, the *trace* of A is

$$\mathrm{tr}(A) = \sum\nolimits_{i=1}^{n} a_{ii}, \tag{1.7}$$

the sum of the diagonal entries of A. Then

$$\mathrm{tr}(aA + bB) = a\,\mathrm{tr}(A) + b\,\mathrm{tr}(B); \tag{1.8}$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA); \qquad (\textit{Note} : A : m \times n,\ B : n \times m) \tag{1.9}$$

$$\mathrm{tr}(A') = \mathrm{tr}(A). \qquad (A : n \times n) \tag{1.10}$$

Proof of (1.9):

$$\begin{aligned}\mathrm{tr}(AB) &= \sum\nolimits_{i=1}^{m} (AB)_{ii} = \sum\nolimits_{i=1}^{m} \left(\sum\nolimits_{k=1}^{n} a_{ik} b_{ki} \right) \\ &= \sum\nolimits_{k=1}^{n} \left(\sum\nolimits_{i=1}^{m} b_{ki} a_{ik} \right) = \sum\nolimits_{k=1}^{n} (BA)_{kk} = \mathrm{tr}(BA).\end{aligned}$$

Determinant: For a square matrix $A \equiv \{a_{ij}\} : n \times n$, its *determinant* is

$$|A| = \sum_{\pi} \epsilon(\pi) \prod_{i=1}^{n} a_{i\pi(i)}$$

$$= \pm\text{Volume}(A([0,1]^n)),$$

where π ranges over all $n!$ permutations of $1, \ldots, n$ and $\epsilon(\pi) = \pm 1$ according to whether π is an even or odd permutation. Then

(1.11) $\quad |AB| = |A| \cdot |B| \qquad (A, B : n \times n);$

(1.12) $\quad |A^{-1}| = |A|^{-1}$

(1.13) $\quad |A'| = |A|;$

(1.14) $\quad |A| = \prod_{i=1}^{n} a_{ii}$ if A is triangular (or diagonal).

Orthogonal matrix. An $n \times n$ matrix Γ is *orthogonal* if

$$\Gamma\Gamma' = I. \tag{1.15}$$

This is equivalent to the fact that the n row vectors of Γ form an orthonormal basis for $\mathcal{R}^n$. Note that (1.15) implies that $\Gamma' = \Gamma^{-1}$, hence also $\Gamma'\Gamma = I$, which is equivalent to the fact that the n column vectors of Γ also form an orthonormal basis for $\mathcal{R}^n$.

Note that Γ preserves angles and lengths, i.e., preserves the usual inner product and norm in $\mathcal{R}^n$: for $x, y \in \mathcal{R}^n$,

$$(\Gamma x,\, \Gamma y) \equiv (\Gamma x)'(\Gamma y) = x'\Gamma'\Gamma y = x'y \equiv (x,\, y),$$

so

$$\|\Gamma x\|^2 \equiv (\Gamma x,\, \Gamma x) = (x,\, x) \equiv \|x\|^2.$$

In fact, any orthogonal transformation is a product of rotations and reflections. Also, from (1.13) and (1.15), $|\Gamma|^2 = 1$, so $|\Gamma| = \pm 1$.

Complex numbers and matrices. For a complex number $c \equiv a+ib \in \mathbf{C}$, let $\bar{c} \equiv a - ib$ denote the *complex conjugate* of c. Note that $\bar{\bar{c}} = c$ and

$$c\bar{c} = a^2 + b^2 \equiv |c|^2,$$
$$\overline{cd} = \bar{c}\bar{d}.$$

For any complex matrix $C \equiv \{c_{ij}\}$, let $\bar{C} = \{\bar{c}_{ij}\}$ and define $C^* = \bar{C}'$. Note that

$$(CD)^* = D^* C^*. \tag{1.16}$$

The characteristic roots $\equiv$ eigenvalues of the $n \times n$ matrix A are the n roots $l_1, \ldots, l_n$ of the polynomial equation

$$|A - l\, I| = 0. \tag{1.17}$$

These roots may be real or complex; the complex roots occur in conjugate pairs. Note that the eigenvalues of a triangular or diagonal matrix are just its diagonal elements.

By the equivalence of (b) and (d) for the (possibly complex) matrix $A - lI$, for each eigenvalue l there exists some nonzero (possibly complex) vector $u \in \mathbf{C}^n$ s.t.

$$(A - l\, I)u = 0,$$

equivalently,

$$Au = lu. \tag{1.18}$$

The vector u is called a *characteristic vector* $\equiv$ *eigenvector* for the eigenvalue l. Since any nonzero multiple cu is also an eigenvector for l, we will usually normalize u to be a unit vector, i.e., $\|u\|^2 \equiv u^* u = 1$.

For example, if A is a diagonal matrix, say

$$A = \mathrm{diag}(d_1, \ldots, d_n) \equiv \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix},$$

then its eigenvalues are just $d_1, \ldots, d_n$, with corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$, where

$$\mathbf{u}_i \equiv (0, \ldots, 0, \underbrace{1}_{i}, 0, \ldots, 0)' \tag{1.19}$$

is the i-th unit coordinate vector.

Note, however, that in general, eigenvalues need not be distinct and eigenvectors need not be unique. For example, if A is the identity matrix I, then its eigenvalues are $1, \ldots, 1$ and *every* unit vector $u \in \mathcal{R}^n$ is an eigenvector for the eigenvalue 1: $Iu = 1 \cdot u$.

However, eigenvectors u, v associated with two *distinct* eigenvalues l, m cannot be colinear: if $u = cv$ then

$$lu = Au = cAv = cmv = mu,$$

which contradicts the assumption that $l \neq m$.

Symmetric matrix. An $n \times n$ matrix $S \equiv \{s_{ij}\}$ is *symmetric* if $S = S'$, i.e., if $s_{ij} = s_{ji}\ \forall i, j$.

Lemma 8.1. *Let S be a real symmetric $n \times n$ matrix.*

(a) Each eigenvalue l of S is real and has a real eigenvector $u \in \mathcal{R}^n$.

(b) If $l \neq m$ are distinct eigenvalues of S with corresponding real eigenvectors u and v, then $u \perp v$, i.e., $u'v = 0$. Thus if all the eigenvalues of S are distinct, each eigenvalue l has exactly one real eigenvector u.

(c) If S^{-1} exists, it is also symmetric.

Proof. (a) Let l be an eigenvalue of S with eigenvector $u \neq 0$. Then

$$Su = lu \quad \Rightarrow \quad u^*Su = lu^*u = l.$$

But S is real and symmetric, so $S^* = S$, hence

$$\overline{u^*Su} = (u^*Su)^* = u^*S^*(u^*)^* = u^*Su.$$

Thus u^*Su is real, hence l is real. Since $S - l\,I$ is real, the existence of a real eigenvector u for l now follows from (b), p.2.

(b) We have $Su = lu$ and $Sv = mv$, hence

$$l(u'v) = (lu)'v = (Su)'v = u'Sv = u'(mv) = m(u'v),$$

so $u'v = 0$ since $l \neq m$.

(c) $I = SS^{-1} = (SS^{-1})' = (S^{-1})'S'$, so $(S^{-1})' = (S')^{-1} = S^{-1}$. □

Proposition 1.2. Spectral decomposition of a real symmetric matrix. *Let S be a real symmetric $n \times n$ matrix with eigenvalues $l_1, \ldots, l_n$ (necessarily real). Then there exists a real orthogonal matrix Γ such that*

$$S = \Gamma D_l \Gamma', \tag{1.20}$$

where $D_l = \mathrm{diag}(l_1, \ldots, l_n)$. Since $S\Gamma = \Gamma D_l$, the i-th column vector γ_i of Γ is a real eigenvector for l_i.

Proof. For simplicity suppose that $l_1, \ldots, l_n$ are distinct. Let $\gamma_1, \ldots, \gamma_n$ be the corresponding unique real unit eigenvectors (apply Lemma 1.1b). Since $\gamma_1, \ldots, \gamma_n$ is an orthonormal basis for $\mathcal{R}^n$, the matrix

$$\Gamma \equiv (\gamma_1, \ldots, \gamma_n) \quad : n \times n \tag{1.21}$$

satisfies $\Gamma'\Gamma = I$, i.e., Γ is an orthogonal matrix. Since each γ_i is an eigenvector for l_i, $S\Gamma = \Gamma D_l$ [verify], which is equivalent to (1.20).

[The case where the eigenvalues are not distinct can be established by a "perturbation" argument. Perturb S slightly so that its eigenvalues become distinct (non-trivial) and apply the first case. Now use a limiting argument based on the compactness of the set of all $n \times n$ orthogonal matrices.] □

Lemma 1.3. *If S is a real symmetric matrix with eigenvalues $l_1, \ldots, l_n$,,*

$$\mathrm{tr}(S) = \sum\nolimits_{i=1}^{n} l_i \ ; \tag{1.22}$$

$$|S| = \prod\nolimits_{i=1}^{n} l_i \ . \tag{1.23}$$

Proof. This is immediate from the spectral decomposition (1.20) of S. □

Positive definite matrix. An $n \times n$ matrix S is *positive semi-definite (psd)* (also written as $S \geq 0$) if it is symmetric and its quadratic form is nonnegative:

$$x'Sx \geq 0 \qquad \forall\, x \in \mathcal{R}^n; \tag{1.24}$$

S is *positive definite (pd)* (also written as $S > 0$) if it is symmetric and its quadratic form is positive:

$$x'Sx > 0 \qquad \forall \text{ nonzero } x \in \mathcal{R}^n. \tag{1.25}$$

- The identity matrix is pd: $x'Ix = \|x\|^2 > 0$ if $x \neq 0$.
- A diagonal matrix $\mathrm{diag}(d_1, \ldots, d_n)$ is psd (pd) iff each $d_i \;\geq 0$ (> 0).
- If $S : n \times n$ is psd, then ASA' is psd for any $A : m \times n$.
- If $S : n \times n$ is pd, then ASA' is pd for any $A : m \times n$ of full rank $m \leq n$.
- AA' is psd for any $A : m \times n$.
- AA' is pd for any $A : m \times n$ of full rank $m \leq n$.

Note: This shows that the proper way to "square" a matrix A is to form AA' (or $A'A$), not A^2, which need not even be symmetric.

- S pd $\Rightarrow$ S has full rank $\Rightarrow$ S^{-1} exists $\Rightarrow$ $S^{-1} \equiv (S^{-1})S(S^{-1})'$ is pd.

Lemma 1.4. *(a) A real symmetric $n \times n$ matrix S with eigenvalues $l_1, \ldots, l_n$ is psd (pd) iff each $l_i \;\geq 0$ (> 0). In particular, $|S| \geq 0$ (> 0) if S is psd (pd), so a pd matrix is nonsingular.*

(b) Suppose S is pd with distinct eigenvalues $l_1 > \cdots > l_n > 0$ and corresponding unique real unit eigenvectors $\gamma_1, \ldots, \gamma_n$. Then the set

$$\mathcal{E} \equiv \{x \in \mathcal{R}^n \mid x'S^{-1}x = 1\} \tag{1.26}$$

is the ellipsoid with principle axes $\sqrt{l_1}\gamma_1, \ldots, \sqrt{l_n}\gamma_n$.

Proof. (a) Apply the above results and the spectral decomposition (1.20).

(b) From (1.20), $S = \Gamma D_l \Gamma'$ with $\Gamma = (\gamma_1 \cdots \gamma_n)$, so $S^{-1} = \Gamma D_l^{-1} \Gamma$ and,

$$\begin{aligned}
\mathcal{E} &= \{x \in \mathcal{R}^n \mid (\Gamma' x)' D_l^{-1} (\Gamma' x) = 1\} \\
&= \Gamma \{y \in \mathcal{R}^n \mid y' D_l^{-1} y = 1\} \qquad\qquad (y = \Gamma' x) \\
&= \Gamma \left\{ y \equiv (y_1, \ldots, y_n)' \;\middle|\; \frac{y_1^2}{l_1} + \cdots + \frac{y_n^2}{l_n} = 1 \right\} \\
&\equiv \Gamma \mathcal{E}_0.
\end{aligned}$$

But $\mathcal{E}_0$ is the ellipsoid with principal axes $\sqrt{l_1}\mathbf{u}_1, \ldots, \sqrt{l_n}\mathbf{u}_n$ (recall (1.19)) and $\Gamma \mathbf{u}_i = \gamma_i$, so $\mathcal{E}$ is the ellipsoid with principle axes $\sqrt{l_1}\gamma_1, \ldots, \sqrt{l_n}\gamma_n$. □

Square root of a pd matrix. Let S be an $n \times n$ pd matrix. Any $n \times n$ matrix A such that $AA' = S$ is called a *square root* of S, denoted by $S^{\frac{1}{2}}$. From the spectral decomposition $S = \Gamma D_l \Gamma'$, one version of $S^{\frac{1}{2}}$ is

$$S^{\frac{1}{2}} = \Gamma \operatorname{diag}(l_1^{\frac{1}{2}}, \ldots, l_n^{\frac{1}{2}})\, \Gamma' \equiv \Gamma D_l^{\frac{1}{2}} \Gamma'; \tag{1.27}$$

this is a *symmetric square root* of S. Any square root $S^{\frac{1}{2}}$ is nonsingular, for

$$|S^{\frac{1}{2}}| = |S|^{\frac{1}{2}} \neq 0. \tag{1.28}$$

Partitioned pd matrix. Partition the pd matrix $S : n \times n$ as

$$S = \begin{matrix} & \begin{matrix} n_1 & n_2 \end{matrix} \\ \begin{matrix} n_1 \\ n_2 \end{matrix} & \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \end{matrix}, \tag{1.29}$$

where $n_1 + n_2 = n$. Then both S_{11} and S_{22} are symmetric pd [why?], $S_{12} = S_{21}'$, and [verify!]

$$\begin{pmatrix} I_{n_1} & -S_{12} S_{22}^{-1} \\ 0 & I_{n_2} \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} I_{n_1} & 0 \\ -S_{22}^{-1} S_{21} & I_{n_2} \end{pmatrix} = \begin{pmatrix} S_{11\cdot 2} & 0 \\ 0 & S_{22} \end{pmatrix}, \tag{1.30}$$

where

$$S_{11\cdot 2} \equiv S_{11} - S_{12} S_{22}^{-1} S_{21} \tag{1.31}$$

is necessarily pd [why?] This in turn implies the two fundamental identities

$$(1.32)\qquad \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} I_{n_1} & S_{12}S_{22}^{-1} \\ 0 & I_{n_2} \end{pmatrix} \begin{pmatrix} S_{11\cdot 2} & 0 \\ 0 & S_{22} \end{pmatrix} \begin{pmatrix} I_{n_1} & 0 \\ S_{22}^{-1}S_{21} & I_{n_2} \end{pmatrix},$$

$$(1.33)\quad \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I_{n_1} & 0 \\ -S_{22}^{-1}S_{21} & I_{n_2} \end{pmatrix} \begin{pmatrix} S_{11\cdot 2}^{-1} & 0 \\ 0 & S_{22}^{-1} \end{pmatrix} \begin{pmatrix} I_{n_1} & -S_{12}S_{22}^{-1} \\ 0 & I_{n_2} \end{pmatrix},$$

The following three consequences of (1.32) and (1.33) are immediate:

(1.34) *S is pd $\iff$ $S_{11\cdot 2}$ and S_{22} are pd $\iff$ $S_{22\cdot 1}$ and S_{11} are pd.*

$$(1.35)\qquad |S| = |S_{11\cdot 2}| \cdot |S_{22}| = |S_{22\cdot 1}| \cdot |S_{11}| \,.$$

For $x \equiv \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathcal{R}^n$, the quadratic form $x'S^{-1}x$ can be decomposed as

$$(1.36)\qquad x'S^{-1}x = (x_1 - S_{12}S_{22}^{-1}x_2)'S_{11\cdot 2}^{-1}(x_1 - S_{12}S_{22}^{-1}x_2) + x_2'S_{22}^{-1}x_2.$$

Exercise 1.5. Cholesky decompositions of a pd matrix. Use (1.32) and induction on n to obtain an *upper triangular* square root U of S, i.e., $S = UU'$. Similarly, S has a *lower triangular* square root L, i.e. $S = LL'$.

Note: Both $U \equiv \{u_{ij}\}$ and $L \equiv \{l_{ij}\}$ are *unique* if the positivity conditions $u_{ii} > 0\ \forall i$ and $l_{ii} > 0\ \forall i$ are imposed on their diagonal elements. To see this for U, suppose that $UU' = VV'$ where V is also an upper triangular matrix with each $v_{ii} > 0$. Then $U^{-1}V(U^{-1}V)' = I$, so $\Gamma \equiv U^{-1}V$ is both upper triangular and orthogonal, hence $\Gamma = \text{diag}(\pm 1, \ldots, \pm 1) =: D$ [why?] Thus $V = UD$, and the positivity conditions imply that $D = I$. □

Projection matrix. An $n \times n$ matrix P is a *projection* matrix if it is symmetric and *idempotent:* $P^2 = P$.

Lemma 1.6. *P is a projection matrix iff it has the form*

$$(1.37)\qquad P = \Gamma \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \Gamma'$$

for some orthogonal matrix $\Gamma : n \times n$ *and some* $m \leq n$. *In this case, rank*$(P) = m = \mathrm{tr}(P)$.

Proof. Since P is symmetric, $P = \Gamma D_l \Gamma'$ by its spectral decomposition. But the idempotence of P implies that each $l_i = 0$ or 1. (A permutation of the rows and columns, which is also an orthogonal transformation, may be necessary to obtain the form (1.37).) □

Interpretation of (1.37): Partition Γ as

$$\Gamma = n \begin{pmatrix} \overset{m}{\Gamma_1} & \overset{n-m}{\Gamma_2} \end{pmatrix}, \tag{1.38}$$

so (1.37) becomes

$$P = \Gamma_1 \Gamma_1'. \tag{1.39}$$

But Γ is orthogonal so $\Gamma'\Gamma = I_n$, hence

$$\Gamma'\Gamma \equiv \begin{pmatrix} \Gamma_1'\Gamma_1 & \Gamma_1'\Gamma_2 \\ \Gamma_2'\Gamma_1 & \Gamma_2'\Gamma_2 \end{pmatrix} = \begin{pmatrix} I_m & 0 \\ 0 & I_{n-m} \end{pmatrix}. \tag{1.40}$$

Thus from (1.39) and (1.40),

$$\begin{aligned} P\Gamma_1 &= (\Gamma_1\Gamma_1')\,\Gamma_1 = \Gamma_1, \\ P\Gamma_2 &= (\Gamma_1\Gamma_1')\,\Gamma_2 = 0. \end{aligned}$$

This shows that P *represents the linear transformation that projects* $\mathcal{R}^n$ *orthogonally onto the column space of* Γ_1, *which has dimension* $m = \mathrm{tr}(P)$.

Furthermore, $I_n - P$ is also symmetric and idempotent [verify] with $\mathrm{rank}(I_n - P) = n - m$. In fact,

$$I_n - P = \Gamma\Gamma' - P = (\Gamma_1\Gamma_1' + \Gamma_2\Gamma_2') - \Gamma_1\Gamma_1' = \Gamma_2\Gamma_2',$$

so $I_n - P$ *represents the linear transformation that projects* $\mathcal{R}^n$ *orthogonally onto the column space of* Γ_2, *which has dimension* $n - m = \mathrm{tr}(I_n - P)$.

Note that the column spaces of Γ_1 and Γ_2 are perpendicular, since $\Gamma_1'\Gamma_2 = 0$. Equivalently, $P(I_n - P) = (I_n - P)P = 0$, i.e., applying P and $I_n - P$ successively sends any $x \in \mathcal{R}^n$ to 0.

1.2. Matrix exercises.

1. For $U : p \times q$ and $S : p \times p$ with $S > 0$ (positive definite), show that

$$|UU' + S| = |S| \cdot |U'S^{-1}U + I_q|,$$

where $|\cdot|$ denotes the determinant and I_q is the $q \times q$ identity matrix.

2. For $a : p \times 1$ and $S : p \times p$ with $S > 0$, show that

$$a'(aa' + S)^{-1}a = \frac{a'S^{-1}a}{a'S^{-1}a + 1}.$$

3. For $S : p \times p$ and $T : p \times p$ with $S > 0$ and $T \geq 0$, show that

$$l_i(T(S+T)^{-1}) = \frac{l_i(TS^{-1})}{1 + l_i(TS^{-1})}, \qquad i = 1, \ldots, p,$$

where $l_1(\cdot) \geq \cdots \geq l_p(\cdot)$ denote the ordered eigenvalues.

4. Let $A > 0$ and $B > 0$ be $p \times p$ matrices with $A \geq B$. Partition A as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

and let $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$. Partition B in the same way and similarly define $B_{11.2}$. Show:

(i) $A_{11} \geq B_{11}$.

(ii) $B^{-1} \geq A^{-1}$.

(iii) $A_{11\cdot 2} \geq B_{11\cdot 2}$.

5. For $S : p \times p$ with $S > 0$, partition S and S^{-1} as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \qquad S^{-1} = \begin{pmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{pmatrix},$$

respectively. Show that $S^{11} \geq S_{11}^{-1}$, and equality holds iff $S_{12} = 0$, or equivalently, iff $S^{12} = 0$.

6. Now partition S and S^{-1} as

$$S = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \equiv \begin{pmatrix} S_{(12)} & S_{(12)3} \\ S_{3(12)} & S_{33} \end{pmatrix},$$

$$S^{-1} = \begin{pmatrix} S^{11} & S^{12} & S^{13} \\ S^{21} & S^{22} & S^{23} \\ S^{31} & S^{32} & S^{33} \end{pmatrix} \equiv \begin{pmatrix} S^{(12)} & S^{(12)3} \\ S^{3(12)} & S^{33} \end{pmatrix}.$$

Then

$$\begin{aligned} S_{(12)\cdot 3} &\equiv S_{(12)} - S_{(12)3} S_{33}^{-1} S_{3(12)} \\ &= \begin{pmatrix} S_{11} - S_{13} S_{33}^{-1} S_{31} & S_{12} - S_{13} S_{33}^{-1} S_{32} \\ S_{21} - S_{23} S_{33}^{-1} S_{31} & S_{22} - S_{23} S_{33}^{-1} S_{32} \end{pmatrix} \\ &\equiv \begin{pmatrix} S_{11\cdot 3} & S_{12\cdot 3} \\ S_{21\cdot 3} & S_{22\cdot 3} \end{pmatrix}, \end{aligned}$$

with similar relations holding for $S^{(12)\cdot 3}$. Note that

$$S^{(12)} = (S_{(12)\cdot 3})^{-1}, \qquad S_{(12)} = (S^{(12)\cdot 3})^{-1},$$

but in general

$$S^{11} \neq (S_{11\cdot 2})^{-1}, \qquad S_{11} = (S^{11\cdot 2})^{-1};$$

instead,

$$S^{11} = (S_{11\cdot (23)})^{-1}, \qquad S_{11} \neq (S^{11\cdot (23)})^{-1}.$$

Show:

(i) $(S_{(12)\cdot 3})_{11\cdot 2} = S_{11\cdot (23)}$.

(ii) $S_{11\cdot 2} = (S^{11\cdot 3})^{-1}$.

(iii) $S_{12\cdot 3}(S_{22\cdot 3})^{-1} = -(S^{11})^{-1} S^{12}$.

(iv) $S_{11} \geq S_{11\cdot 2} \geq S_{11\cdot (23)}$. When do the inequalities become equalities?

(v) $S_{12\cdot 3}(S_{22\cdot 3})^{-1} = -(S^{11\cdot 4})^{-1} S^{12\cdot 4}$ (for a 4×4 partitioning.)

1.3. Random vectors and covariance matrices. Let $X \equiv \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ be a rvtr in $\mathcal{R}^n$. The *expected value* of X is the vector

$$\mathrm{E}(X) \equiv \begin{pmatrix} \mathrm{E}(X_1) \\ \vdots \\ \mathrm{E}(X_n) \end{pmatrix},$$

which is the center of gravity of the probability distribution of X in $\mathcal{R}^n$. Note that expectation is linear: for rvtrs X, Y and constant matrices A, B,

$$\mathrm{E}(AX + BY) = A\,\mathrm{E}(X) + B\,\mathrm{E}(Y). \tag{1.41}$$

Similarly, if $Z \equiv \begin{pmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & & \vdots \\ Z_{m1} & \cdots & Z_{mn} \end{pmatrix}$ is a random matrix in $\mathcal{R}^{m\times n}$, $\mathrm{E}(Z)$ is also defined component-wise:

$$\mathrm{E}(Z) = \begin{pmatrix} \mathrm{E}(Z_{11}) & \cdots & \mathrm{E}(Z_{1n}) \\ \vdots & & \vdots \\ \mathrm{E}(Z_{m1}) & \cdots & \mathrm{E}(Z_{mn}) \end{pmatrix}.$$

Then for constant matrices $A : k \times m$ and $B : n \times p$,

$$\mathrm{E}(AZB) = A\,\mathrm{E}(Z)\,B. \tag{1.42}$$

The covariance matrix of X ($\equiv$ the *variance-covariance matrix*), is

$$\mathrm{Cov}(X) = \mathrm{E}\,[(X - \mathrm{E}X)(X - \mathrm{E}X)']$$

$$= \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{pmatrix}.$$

The following formulas are essential: for $X : n \times 1$, $A : m \times n$, $a : n \times 1$,

$$\text{Cov}(X) = \text{E}(XX') - (\text{E}X)(\text{E}X)'; \tag{1.43}$$

$$\text{Cov}(AX + b) = A\,\text{Cov}(X)\,A'; \tag{1.44}$$

$$\text{Var}(a'X + b) = a'\,\text{Cov}(X)\,a. \tag{1.45}$$

Lemma 1.7. *Let $X \equiv (X_1, \ldots, X_n)'$ be a random vector in $\mathcal{R}^n$.*

(a) $\text{Cov}(X)$ *is psd.*

(b) $\text{Cov}(X)$ *is pd unless $\exists$ a nonzero $a \equiv (a_1, \ldots, a_n)' \in \mathcal{R}^n$ s.t. the linear combination*

$$a'X \equiv a_1X_1 + \cdots + a_nX_n = \textit{constant},$$

i.e., the support of X is contained in some hyperplane of dimension $\leq n-1$.

Proof. (a) This follows immediately from (1.45).

(b) If $\text{Cov}(X)$ is not pd, then $\exists$ a nonzero $a \in \mathcal{R}^n$ s.t.

$$0 = a'\,\text{Cov}(X)\,a = \text{Var}(a'X).$$

But this implies that $a'X = \text{const}$. □

For rvtrs $X : m \times 1$ and $Y : n \times 1$, define

$$\text{Cov}(X, Y) = \text{E}\left[(X - \text{E}X)(Y - \text{E}Y)'\right]$$

$$= \begin{pmatrix} \text{Cov}(X_1, Y_1) & \text{Cov}(X_1, Y_2) & \cdots & \text{Cov}(X_1, Y_n) \\ \text{Cov}(X_2, Y_1) & \text{Cov}(X_2, Y_2) & \cdots & \text{Cov}(X_2, Y_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(X_m, Y_1) & \text{Cov}(X_m, Y_2) & \cdots & \text{Cov}(X_m, Y_n) \end{pmatrix}.$$

Clearly $\text{Cov}(X, Y) = [\text{Cov}(Y, X)]'$. Then [verify]

$$\text{Cov}(X \pm Y) = \text{Cov}(X) + \text{Cov}(Y) \pm \text{Cov}(X, Y) \pm \text{Cov}(Y, X). \tag{1.46}$$

and [verify]

$$\begin{aligned} X \perp\!\!\!\perp Y &\Rightarrow \text{Cov}(X, Y) = 0 \\ &\Rightarrow \text{Cov}(X \pm Y) = \text{Cov}(X) + \text{Cov}(Y). \end{aligned} \tag{1.47}$$

Variance of sample average (sample mean) of rvtrs: Let $X_1, \ldots, X_n$ be i.i.d. rvtrs in $\mathcal{R}^p$, each with mean vector μ and covariance matrix Σ. Set

$$\bar{X}_n = \tfrac{1}{n}(X_1 + \cdots + X_n).$$

Then $\mathrm{E}(\bar{X}_n) = \mu$ and, by (1.47),

$$\mathrm{Cov}(\bar{X}_n) = \tfrac{1}{n^2}\mathrm{Cov}(X_1 + \cdots + X_n) = \tfrac{1}{n}\Sigma. \tag{1.48}$$

Exercise 1.8. Verify the *Weak Law of Large Numbers (WLLN)* for rvtrs: $\bar{X}_n$ converges to μ in probability ($X_n \xrightarrow{p} \mu$), that is, for each $\epsilon > 0$,

$$P[\|\bar{X}_n - \mu\| \le \epsilon] \to 1 \quad \textit{as } n \to \infty.$$

Example 1.9a. Equicorrelated random variables. Let $X_1, \ldots, X_n$ be rvs with common mean μ and common variance σ^2. Suppose they are *equicorrelated*, i.e., $\mathrm{Corr}(X_i, X_j) = \rho\ \forall i \ne j$. Let

$$\bar{X}_n = \tfrac{1}{n}(X_1 + \cdots + X_n), \qquad s_n^2 = \tfrac{1}{n-1}\textstyle\sum_{i=1}^n (X_i - \bar{X}_n)^2, \tag{1.49}$$

the *sample mean* and *sample variance*, respectively. Then

$$\mathrm{E}(\bar{X}_n) = \mu \qquad\qquad (\text{so } \bar{X}_n \text{ is } \textit{unbiased} \text{ for } \mu);. \tag{1.50}$$

$$\begin{aligned}\mathrm{Var}(\bar{X}_n) &= \tfrac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n)\\ &= \tfrac{1}{n^2}[n\sigma^2 + n(n-1)\rho\sigma^2] \qquad [\text{why?}]\\ &= \tfrac{\sigma^2}{n}[1 + (n-1)\rho]. \qquad (1.51)\end{aligned}$$

When $X_1, \ldots, X_n$ are uncorrelated ($\rho = 0$), in particular when they are independent, then (1.51) reduces to $\frac{\sigma^2}{n}$, which $\to 0$ as $n \to \infty$. When $\rho \ne 0$, however, $\mathrm{Var}(\bar{X}_n) \to \sigma^2\rho \ne 0$ so *the WLLN fails for equicorrelated i.d. rvs* Also, (1.51) imposes the constraint

$$-\tfrac{1}{n-1} \le \rho \le 1. \tag{1.52}$$

Next, using (1.51),

$$\begin{aligned}\mathrm{E}(s_n^2) &= \tfrac{1}{n-1}\mathrm{E}\left(\textstyle\sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2\right)\\ &= \left(\tfrac{1}{n-1}\right)\left[n(\sigma^2 + \mu^2) - n\left(\tfrac{\sigma^2}{n}[1 + (n-1)\rho] + \mu^2\right)\right]\\ &= (1-\rho)\sigma^2. \qquad (1.53)\end{aligned}$$

Thus s_n^2 is unbiased for σ_n^2 if $\rho = 0$ but not otherwise. $\bar{\square}$

Example 1.9b. We now re-derive (1.51) and (1.53) via covariance matrices, using properties (1.44) and (1.45). Set $X = (X_1, \ldots, X_n)'$, so

$$\text{(1.54)} \qquad \mathrm{E}(X) = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} \equiv \mu\, \mathbf{e}_n, \qquad \text{where } \mathbf{e}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} : n \times 1,$$

$$\mathrm{Cov}(X) = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

$$\text{(1.55)} \qquad \equiv \sigma^2[(1-\rho)I_n + \rho\, \mathbf{e}_n \mathbf{e}_n'].$$

Then $\bar{X}_n = \frac{1}{n}\mathbf{e}_n' X$, so by (1.45),

$$\begin{aligned} \mathrm{Var}(\bar{X}_n) &= \tfrac{\sigma^2}{n^2}\mathbf{e}_n'[(1-\rho)I_n + \rho\, \mathbf{e}_n \mathbf{e}_n']\mathbf{e}_n \\ &= \tfrac{\sigma^2}{n^2}[(1-\rho)n + \rho n^2] \qquad\qquad [\text{since } \mathbf{e}_n'\mathbf{e}_n = n] \\ &= \tfrac{\sigma^2}{n}[1 + (n-1)\rho], \end{aligned}$$

which agrees with (1.51).

To find $\mathrm{E}(s_n^2)$, write

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X}_n)^2 &= \sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2 \\ &= X'X - \tfrac{1}{n}(\mathbf{e}_n' X)^2 \\ &= X'X - \tfrac{1}{n}(X'\mathbf{e}_n)(\mathbf{e}_n' X) \\ &\equiv X'\left(I_n - \vec{\mathbf{e}}_n \vec{\mathbf{e}_n}'\right) X \\ \text{(1.56)} \qquad &\equiv X'QX, \end{aligned}$$

where $\vec{\mathbf{e}}_n \equiv \left(\frac{\mathbf{e}_n}{\sqrt{n}}\right)$ is a unit vector, $P \equiv \vec{\mathbf{e}}_n \vec{\mathbf{e}_n}'$ is the projection matrix of rank $1 \equiv \mathrm{tr}(\vec{\mathbf{e}}_n \vec{\mathbf{e}_n}')$ that projects $\mathcal{R}^n$ orthogonally onto the 1-dimensional subspace spanned by $\mathbf{e}_n$, and $Q \equiv I_n - \vec{\mathbf{e}}_n \vec{\mathbf{e}_n}'$ is the projection matrix of rank $n-1 \equiv \operatorname{tr} Q$ that projects $\mathcal{R}^n$ orthogonally onto the $(n-1)$-dimensional subspace $\mathbf{e}_n^\perp$ [draw figure]. Now complete the following exercise:

Exercise 1.10. Prove Lemma 1.11 below, and use it to show that

$$\mathrm{E}(X'QX) = (n-1)(1-\rho)\sigma^2, \tag{1.57}$$

which is equivalent to (1.53). $\bar{\square}$

Lemma 1.11. Let $X : n \times 1$ be a rvtr with $\mathrm{E}(X) = \mu$ and $\mathrm{Cov}(X) = \Sigma$. Then for any $n \times n$ symmetric matrix A,

$$\mathrm{E}(X'AX) = \mathrm{tr}(A\Sigma) + \mu' A \mu. \tag{1.58}$$

(This generalizes the relation $\mathrm{E}(X^2) = \mathrm{Var}(X) + (\mathrm{E}\, X)^2$.)

Example 1.9c. Eqn. (1.53) also can be obtained from the properties of the projection matrix Q. First note that [verify]

$$Q\mathbf{e}_n = \sqrt{n} Q \vec{\mathbf{e}}_n = 0. \tag{1.59}$$

Define

$$Y \equiv \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = QX \quad : n \times 1, \tag{1.60}$$

so

$$\mathrm{E}(Y) = Q\,\mathrm{E}(X) = \mu\, Q\mathbf{e}_n = 0, \tag{1.61}$$

$$\begin{aligned} \mathrm{E}(YY') = \mathrm{Cov}(Y) &= \sigma^2 Q[(1-\rho)I_n + \rho \mathbf{e}_n \mathbf{e}_n']Q' \\ &= \sigma^2(1-\rho)Q. \end{aligned} \tag{1.62}$$

Thus, since Q is idempotent ($Q^2 = Q$),

$$\begin{aligned} \mathrm{E}(X'QX) = \mathrm{E}(Y'Y) &= \mathrm{E}[\,\mathrm{tr}\,(YY')] \\ &= \mathrm{tr}[\,\mathrm{E}\,(YY')] \\ &= \sigma^2(1-\rho)\,\mathrm{tr}\,(Q) \\ &= \sigma^2(1-\rho)(n-1), \end{aligned}$$

which again is equivalent to (1.53). □

Exercise 1.12. Show that $\text{Cov}(X) \equiv \sigma^2[(1-\rho)I_n + \rho\mathbf{e}_n\mathbf{e}_n']$ in (1.55) has one eigenvalue $= \sigma^2[1+(n-1)\rho]$ with eigenvector $\mathbf{e}_n$, and $n-1$ eigenvalues $= \sigma^2(1-\rho)$. □

Exercise 1.13. Suppose that $\Sigma = \text{Cov}(X) : n \times n$ and let $\lambda_1 \geq \cdots \geq \lambda_n$ denote the eigenvalues of Σ. Show that the extreme eigenvalues satisfy

$$\lambda_1 = \max_{\|a\|=1} \text{Var}(a'X),$$
$$\lambda_n = \min_{\|a\|=1} \text{Var}(a'X).$$ □

The Courant-Fischer Minmax Theorem. Denote a k-dimensional subspace of $\mathcal{R}^n$ by L_k. Then

$$\lambda_k = \min_{L_k} \max_{a \in L_k,\, \|a\|=1} \text{Var}(a'X),$$
$$\lambda_{n-k+1} = \max_{L_k} \min_{a \in L_k,\, \|a\|=1} \text{Var}(a'X).$$ □

2. The Multivariate Normal Distribution (MVND).

2.1. Definition and basic properties.

Consider a random vector $Z \equiv (Z_1, \ldots, Z_p)' \in \mathcal{R}^p$, where $Z_1, \ldots, Z_p$ are i.i.d. standard normal random variables, i.e., $Z_i \sim N(0,1)$, so $\mathrm{E}(Z) = 0$ and $\mathrm{Cov}(Z) = I_p$. The pdf of Z (i.e., the joint pdf of $Z_1, \ldots, Z_p$) is

$$\begin{aligned} f(z) &= (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}(z_1^2+\cdots+z_p^2)} \\ &= (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}z'z}, \qquad z \in \mathcal{R}^p. \end{aligned} \tag{2.1}$$

For any nonsingular matrix $A : p \times p$ and any $\mu : p \times 1 \in \mathcal{R}^p$, consider the random vector $X := AZ + \mu$. Since the Jacobian of this linear (actually, affine) mapping is $|\frac{\partial X}{\partial Z}| = |A| > 0$ and $Z = A^{-1}(X - \mu)$, the pdf of X is

$$\begin{aligned} f(x) &= (2\pi)^{-\frac{p}{2}} |A|^{-1} e^{-\frac{1}{2}(A^{-1}(x-\mu))' A^{-1}(x-\mu)} \\ &= (2\pi)^{-\frac{p}{2}} |AA'|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)'(AA')^{-1}(x-\mu)} \\ &= (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \qquad x \in \mathcal{R}^p, \end{aligned} \tag{2.2}$$

where

$$\begin{aligned} \mathrm{E}(X) &= A\,\mathrm{E}(Z) + \mu = \mu, \\ \mathrm{Cov}(X) &= A\,\mathrm{Cov}(Z)\,A' = AA' \equiv \Sigma > 0. \end{aligned}$$

Since the distribution of X depends only on μ and Σ, we denote this distribution by $N_p(\mu, \Sigma)$, the multivariate normal distribution (MVND) on $\mathcal{R}^p$ with mean vector μ and covariance matrix Σ.

Exercise 2.1. (a) Show that the moment generating function of Z is

$$m_Z(t) \equiv \mathrm{E}(e^{t'Z}) = e^{\frac{1}{2}t't}. \tag{2.3}$$

(b) Let $X = AZ + \mu$ where now $A : q \times p$ and $\mu \in \mathcal{R}^q$. Show that the mgf of X is

$$m_X(t) \equiv \mathrm{E}(e^{t'X}) = e^{t'\mu + \frac{1}{2}t'\Sigma t} \tag{2.4}$$

where $\Sigma \equiv AA' = \mathrm{Cov}(X)$. Thus the distribution of $X \equiv AZ + \mu$ depends only on μ and Σ even when A is singular and/or a non-square matrix, so we may again write $X \sim N_q(\mu, \Sigma)$.

Lemma 2.1. Affine transformations preserve normality.

If $X \sim N_q(\mu, \Sigma)$, *then for* $C : r \times q$ *and* $d : r \times 1$,

$$Y \equiv CX + d \sim N_r(C\mu + d,\, C\Sigma C'). \tag{2.5}$$

Proof. Represent X as $AZ + \mu$, so $Y = (CA)Z + (C\mu + d)$ is also an affine transformation of Z, hence also has an MVND with $\mathrm{E}(Y) = C\mu + d$ and $\mathrm{Cov}(Y) = (CA)(CA)' = C\Sigma C'$. □

Lemma 2.2. Independence $\iff$ zero covariance.

Suppose that $X \sim N_p(\mu, \Sigma)$ *and partition* X, μ, *and* Σ *as*

$$X = \begin{matrix} p_1 \\ p_2 \end{matrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{matrix} p_1 \\ p_2 \end{matrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{matrix} \\ p_1 \\ p_2 \end{matrix} \begin{matrix} \begin{matrix} p_1 & \quad p_2 \end{matrix} \\ \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{matrix}, \tag{2.6}$$

where $p_1 + p_2 = p$. *Then* $X_1 \perp\!\!\!\perp X_2 \iff \Sigma_{12} = 0$.

Proof. This follows from the pdf (2.2) or the mgf (2.4). □

Proposition 2.3. Marginal & conditional distributions are normal.

If $X \sim N_p(\mu, \Sigma)$ *and* Σ_{22} *is pd then*

$$X_1 \mid X_2 \sim N_{p_1}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2),\, \Sigma_{11\cdot 2}\right), \tag{2.8}$$

$$X_2 \sim N_{p_2}(\mu_2, \Sigma_{22}). \tag{2.9}$$

Proof. Method 1: Assume also that Σ *is nonsingular.* By the quadratic identity (1.33) applied with μ, x, and Σ partitioned as in (2.6),

$$\begin{aligned} &(x - \mu)'\Sigma^{-1}(x - \mu) \\ &= \left(x_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)' \Sigma_{11\cdot 2}^{-1}(\cdots) + (x_2 - \mu_2)'\Sigma_{22}^{-1}(\cdots). \end{aligned} \tag{2.10}$$

Since also $|\Sigma| = |\Sigma_{11\cdot 2}|\,|\Sigma_{22}|$, the result follows from the pdf (2.2).

Method 2. By Lemma 2.1 and the quadratic identity (1.32),

$$\begin{pmatrix} X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \\ X_2 \end{pmatrix} = \begin{pmatrix} I_{p_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p_2} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\sim N_{p_1+p_2}\left(\begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right). \tag{2.11}$$

Thus by Lemma 2.1 for $C = (\, I_{p_1} \quad 0_{p_1 \times p_2} \,)$ and $(\, 0_{p_2 \times p_1} \quad I_{p_2} \,)$, respectively,

$$X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \sim N_{p_1}\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\, \Sigma_{11\cdot 2}\right),$$
$$X_2 \sim N_{p_2}\left(\mu_2, \Sigma_{22}\right),$$

so (2.9) holds. Also $X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \perp\!\!\!\perp X_2$ by (2.11) and Lemma 2.2, so

$$X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \mid X_2 \sim N_{p_1}\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2,\, \Sigma_{11\cdot 2}\right)$$

which yields (2.8). □

Exercise 2.2. Conditional independence $\iff$ zero precision. Let $X \sim N_p(\mu, \Sigma)$ and partition X and Σ^{-1} (the *precision matrix*) as

$$X = \begin{matrix} p_1 \\ p_2 \\ p_3 \end{matrix}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad \Sigma^{-1} = \begin{matrix} \\ p_1 \\ p_2 \\ p_3 \end{matrix}\begin{matrix} \begin{matrix} p_1 & p_2 & p_3 \end{matrix} \\ \begin{pmatrix} \Sigma^{11} & \Sigma^{12} & \Sigma^{13} \\ \Sigma^{21} & \Sigma^{22} & \Sigma^{23} \\ \Sigma^{31} & \Sigma^{32} & \Sigma^{33} \end{pmatrix} \end{matrix}, \tag{2.12}$$

where $p_1 + p_2 + p_3 = p$. Then $X_1 \perp\!\!\!\perp X_2 \mid X_3 \iff \Sigma^{12} = 0$. □

2.2. The MVND and the chi-square distribution.

The *chi-square distribution* χ^2_n *with* n *degrees of freedom (df)* can be defined as the distribution of

$$Z_1^2 + \cdots + Z_n^2 \equiv Z'Z \equiv \|Z\|^2,$$

where $Z \equiv (Z_1, \ldots, Z_n)' \sim N_n(0, I_n)$. (That is, $Z_1, \ldots, Z_n$ are i.i.d. standard $N(0,1)$ rvs.) Recall that

$$\chi^2_n \sim \text{Gamma}\left(\alpha = \tfrac{n}{2},\, \lambda = \tfrac{1}{2}\right), \tag{2.13}$$
$$\mathrm{E}(\chi^2_n) = n, \tag{2.14}$$
$$\mathrm{Var}(\chi^2_n) = 2n. \tag{2.15}$$

Now consider $X \sim N_n(\mu, \Sigma)$ with Σ pd. Then

$$Z \equiv \Sigma^{-1/2}(X-\mu) \sim N_n(0, I_n), \tag{2.16}$$

$$Z'Z = (X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_n. \tag{2.17}$$

Suppose, however, that we omit Σ^{-1} in (2.17) and seek the distribution of $(X-\mu)'(X-\mu)$. Then this will *not* have a chi-square distribution in general. Instead, by the spectral decomposition $\Sigma = \Gamma D_\lambda \Gamma'$, (2.16) yields

$$\begin{aligned}(X-\mu)'(X-\mu) = Z'\Sigma Z &= (\Gamma' Z)' D_\lambda (\Gamma' Z)\\ &\equiv V' D_\lambda V = \lambda_1 V_1^2 + \cdots \lambda_n V_n^2,\end{aligned} \tag{2.18}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of Σ and $V \equiv \Gamma' Z \sim N_n(0, I_n)$. Thus the distribution of $(X-\mu)'(X-\mu)$ is a *positive linear combination of independent* χ^2_1 *rvs,* which is not (proportional to) a χ^2_n rv. [Check via mgfs!]

Lemma 2.5. Quadratic forms and projection matrices.
Let $X \sim N_n(\xi, \sigma^2 I_n)$ *and let* P *be an* $n \times n$ *projection matrix with rank*$(P) =$ $\mathrm{tr}(P) \equiv m$. *Then the quadratic form determined by* $X - \xi$ *and* P *satisfies*

$$(X-\xi)'P(X-\xi) \sim \sigma^2 \chi^2_m. \tag{2.23}$$

Proof. By Lemma 1.6, there exists an orthogonal matrix $\Gamma : n \times n$ s.t.

$$P = \Gamma \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \Gamma'.$$

Then $Y \equiv \Gamma'(X-\xi) \sim N_n(0, \sigma^2 I_n)$, so with $Y = (Y_1, \ldots, Y_n)'$,

$$(X-\xi)'P(X-\xi) = Y' \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} Y = Y_1^2 + \cdots Y_m^2 \sim \sigma^2 \chi^2_m. \qquad \square$$

2.3. The noncentral chi-square distribution.

Extend the result (2.17) to (2.30) as follows: First let $Z \equiv (Z_1, \ldots, Z_n)' \sim N_n(\xi, I_n)$, where $\xi \equiv (\xi_1, \ldots, \xi_n)' \in \mathcal{R}^n$. The distribution of

$$Z_1^2 + \cdots + Z_n^2 \equiv Z'Z \equiv \|Z\|^2$$

is called the *noncentral chi-square distribution with n degrees of freedom (df) and noncentrality parameter* $\|\xi\|^2$, denoted by $\chi_n^2(\|\xi\|^2)$. Note that $Z_1, \ldots, Z_n$ are independent, each with variance = 1, but now $\mathrm{E}(Z_i) = \xi_i$.

To show that the distribution of $\|Z\|^2$ depends on ξ only through its (squared) length $\|\xi\|^2$, choose[1] an orthogonal (rotation) matrix $\Gamma : n \times n$ such that $\Gamma\xi = (\|\xi\|, 0, \ldots, 0)'$, i.e., Γ rotates ξ into $(\|\xi\|, 0, \ldots, 0)'$, and set

$$Y = \Gamma Z \sim N_n(\Gamma\xi,\, \Gamma\Gamma') = N_n\left((\|\xi\|, 0, \ldots, 0)',\, I_n\right).$$

Then the desired result follows since

$$\begin{aligned} \|Z\|^2 = \|Y\|^2 &\equiv Y_1^2 + Y_2^2 + \cdots + Y_n^2 \\ &\sim [N_1(\|\xi\|,\, 1)]^2 + [N_1(0,\, 1)]^2 + \cdots + [N_1(0,\, 1)]^2 \\ &\equiv \chi_1^2(\|\xi\|^2) + \chi_1^2 + \cdots + \chi_1^2 \\ &\equiv \chi_1^2(\|\xi\|^2) + \chi_{n-1}^2, \end{aligned} \tag{2.24}$$

where the chi-square variates in each line are mutually independent.

Let $V \equiv Y_1^2 \sim \chi_1^2(\delta) \sim [N_1(\sqrt{\delta},\, 1)]^2$, where $\delta = \|\xi\|^2$. We find the pdf of V as follows:

$$\begin{aligned} f_V(v) &= \frac{d}{dv} P[-\sqrt{v} \le Y_1 \le \sqrt{v}] \\ &= \frac{d}{dv} \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{v}}^{\sqrt{v}} e^{-\frac{1}{2}(t-\sqrt{\delta})^2} dt \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta}{2}} \frac{d}{dv} \int_{-\sqrt{v}}^{\sqrt{v}} e^{t\sqrt{\delta}} e^{-\frac{t^2}{2}} dt \end{aligned}$$

[1] Let the first row of Γ be $\vec{\xi} \equiv \frac{\xi}{\|\xi\|}$ and let the remaining $n-1$ rows be any orthonormal basis for $L^{\perp}$.

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta}{2}} \frac{d}{dv} \int_{-\sqrt{v}}^{\sqrt{v}} \left[\sum_{k=0}^{\infty} \frac{t^k \delta^{\frac{k}{2}}}{k!} \right] e^{-\frac{t^2}{2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta}{2}} \sum_{k=0}^{\infty} \frac{\delta^{\frac{k}{2}}}{k!} \frac{d}{dv} \int_{-\sqrt{v}}^{\sqrt{v}} t^k e^{-\frac{t^2}{2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta}{2}} \sum_{k=0}^{\infty} \frac{\delta^k}{(2k)!} \frac{d}{dv} \int_{-\sqrt{v}}^{\sqrt{v}} t^{2k} e^{-\frac{t^2}{2}} dt \qquad \text{[why?]}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta}{2}} \sum_{k=0}^{\infty} \frac{\delta^k}{(2k)!} v^{k-\frac{1}{2}} e^{-\frac{v}{2}} \qquad \text{[verify]}$$

$$(2.25) \qquad = \underbrace{e^{-\frac{\delta}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^k}{k!}}_{\text{Poisson}\left(\frac{\delta}{2}\right) \text{ weights}} \underbrace{\left[\frac{v^{\frac{1+2k}{2}-1} e^{-\frac{v}{2}}}{2^{\frac{1+2k}{2}} \Gamma\left(\frac{1+2k}{2}\right)} \right]}_{\text{pdf of } \chi^2_{1+2k}} \cdot c_k,$$

where

$$c_k = \frac{2^k\, k!\, 2^{\frac{1+2k}{2}}\, \Gamma\left(\frac{1+2k}{2}\right)}{(2k)!\, \sqrt{2\pi}} = 1$$

by the Legendre duplication formula for the Gamma function. Thus we have represented the pdf of a $\chi_1^2(\delta)$ rv as a mixture (weighted average) of central chi-square pdfs with Poisson weights. This can be written as follows:

$$(2.26) \qquad \chi_1^2(\delta) \,\big|\, K \;\sim\; \chi^2_{1+2K} \qquad \text{where} \qquad K \sim \text{Poisson}\,(\delta/2)\,.$$

Thus by (2.24) this implies that $Z'Z \equiv \|Z\|^2 \sim \chi_n^2(\delta)$ satisfies

$$(2.27) \qquad \chi_n^2(\delta) \,\big|\, K \;\sim\; \chi^2_{n+2K} \qquad \text{where} \qquad K \sim \text{Poisson}\,(\delta/2)\,.$$

That is, *the pdf of a noncentral chi-square rv $\chi_n^2(\delta)$ is a Poisson$(\delta/2)$-mixture of the pdfs of central chi-square rvs with $n+2k$ df, $k = 0, 1, \ldots$.*

The representation (2.27) can be used to obtain the mean and variance of $\chi_n^2(\delta)$:

$$\begin{aligned}
\mathrm{E}[\chi^2_n(\delta)] &= \mathrm{E}\{\mathrm{E}[\chi^2_{n+2K} \mid K]\} \\
&= \mathrm{E}(n+2K) \\
&= n + 2(\delta/2) \\
&= n + \delta; \qquad (2.28) \\
\mathrm{Var}[\chi^2_n(\delta)] &= \mathrm{E}[\mathrm{Var}(\chi^2_{n+2K} \mid K)] + \mathrm{Var}[\mathrm{E}(\chi^2_{n+2K} \mid K)] \\
&= \mathrm{E}[2(n+2K)] + \mathrm{Var}(n+2K) \\
&= [2n + 4(\delta/2)] + 4(\delta/2) \\
&= 2n + 4\delta. \qquad (2.29)
\end{aligned}$$

Exercise 2.6. Show that the noncentral chi-square distribution $\chi^2_n(\delta)$ is stochastically increasing in both n and δ. □

Next, consider $X \sim N_n(\mu, \Sigma)$ with a general pd Σ. Then

$$X'\Sigma^{-1}X = (\Sigma^{-\frac{1}{2}}X)'(\Sigma^{-\frac{1}{2}}X) \sim \chi^2_n(\mu'\Sigma^{-1}\mu), \qquad (2.30)$$

since

$$Z \equiv \Sigma^{-\frac{1}{2}}X \sim N_n(\Sigma^{-\frac{1}{2}}\mu,\, I_n)$$

and

$$\|\Sigma^{-\frac{1}{2}}\mu\|^2 = \mu'\Sigma^{-1}\mu.$$

Note that by Exercise 2.6, the distribution of $X'\Sigma^{-1}X$ in (2.30) is stochastically increasing in n and $\mu'\Sigma^{-1}\mu$.

Finally, let $X \sim N_n(\xi, \sigma^2 I_n)$ and let P be a projection matrix with $\mathrm{rank}(P) = m$. Then $P = \Gamma_1\Gamma_1'$ where $\Gamma_1'\Gamma_1 = I_m$ (cf. (2.20) - (2.22)), so

$$\|PX\|^2 = \|\Gamma_1\Gamma_1'X\|^2 = (\Gamma_1\Gamma_1'X)'(\Gamma_1\Gamma_1'X) = X'\Gamma_1\Gamma_1'X = \|\Gamma_1'X\|^2.$$

But

$$\Gamma_1'X \sim N_m(\Gamma_1'\xi,\, \sigma^2\Gamma_1'\Gamma_1) = N_m(\Gamma_1'\xi,\, \sigma^2 I_m),$$

so by (2.30) with $X = \Gamma_1' X$, $\mu = \Gamma_1' \xi$, and $\Sigma = \sigma^2 I_m$,

$$\frac{\|PX\|^2}{\sigma^2} = \frac{(\Gamma_1' X)'(\Gamma_1' X)}{\sigma^2} \sim \chi^2_m\left(\frac{\xi'\Gamma_1\Gamma_1'\xi}{\sigma^2}\right) = \chi^2_m\left(\frac{\|P\xi\|^2}{\sigma^2}\right).$$

Thus

$$\|PX\|^2 \sim \sigma^2\chi^2_m\left(\frac{\|P\xi\|^2}{\sigma^2}\right). \tag{2.31}$$

2.4. Joint pdf of a random sample from the MVND $N_p(\mu, \Sigma)$.

Let $X_1, \dots, X_n$ be an i.i.d random sample from $N_p(\mu, \Sigma)$. Assume that Σ is positive definite (pd) so that each X_i has pdf given by (2.2). Thus the joint pdf of $X_1, \dots, X_n$ is

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu)'\Sigma^{-1}(x_i-\mu)} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left(\sum_{i=1}^n (x_i-\mu)(x_i-\mu)'\right)\right]} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{-\frac{n}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}s\right)}, \end{aligned} \tag{2.32}$$

or alternatively,

$$= \frac{e^{-\frac{n}{2}\mu'\Sigma^{-1}\mu}}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{n\bar{x}'\Sigma^{-1}\mu - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}v\right)}, \tag{2.33}$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \qquad S = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})', \qquad V = \sum_{i=1}^n X_i X_i'.$$

It follows from (2.32) and (2.33) that $(\bar{X}, S)$ *and* $(\bar{X}, V)$ *are equivalent representations of the minimal sufficient statistic for* (μ, Σ). Also from (2.33), *with no further restrictions on* (μ, Σ), *this MVN statistical model constitutes a* $p + \frac{p(p+1)}{2}$*-dimensional full exponential family with natural parameter* $(\Sigma^{-1}\mu, \Sigma^{-1})$.

3. The Wishart Distribution.

3.1. Definition and basic properties.

Let $X_1, \ldots, X_n$ be an i.i.d. random sample from $N_p(0, \Sigma)$ and set

$$X = (X_1, \ldots, X_n) : p \times n,$$
$$S = XX' = \sum_{i=1}^{n} X_i X_i' : p \times p.$$

The distribution of S is called the p-variate (central) *Wishart distribution* with n degrees of freedom and scale matrix Σ, denoted by $W_p(n, \Sigma)$. $\bar{\square}$.

Clearly S is a random symmetric positive semi-definite matrix with $\mathrm{E}(S) = n\Sigma$. When $p = 1$ and $\Sigma = \sigma^2$, $W_1(n, \sigma^2) = \sigma^2 \chi^2_n$.

Lemma 3.1. Preservation under linear transformation. *For $A : q \times p$,*

$$ASA' \sim W_q(n, A\Sigma A'). \tag{3.1}$$

In particular, for $a : p \times 1$,

$$a'Sa \sim (a'\Sigma a) \cdot \chi^2_n. \tag{3.2}$$

Lemma 3.2. Nonsingularity $\equiv$ positive-definiteness of $S \sim W_p(n, \Sigma)$. *S is positive definite with probability one $\iff$ Σ is pd and $n \geq p$.*

Proof. ($\Rightarrow$): Recall that $S \sim XX'$ with $X : p \times n$. If $n < p$ then

$$\mathrm{rank}(S) = \mathrm{rank}(X) \leq \min(p, n) = n < p,$$

so S is singular with probability one, hence not positive definite. If Σ is not pd then $\exists\, a : p \times 1$, $a \neq 0$, s.t. $a'\Sigma a = 0$. Thus by (3.2),

$$a'Sa \sim (a'\Sigma a) \cdot \chi^2_n = 0,$$

so S is singular w.pr.1.

($\Leftarrow$) *Method 1 (Stein; Eaton-Perlman (1973) Ann. Statist.)* Assume that Σ is pd and $n \geq p$. Since

$$S = XX' = \sum_{i=1}^{p} X_i X_i' + \sum_{i=p+1}^{n} X_i X_i',$$

it suffices to show that $\sum_{i=1}^{p} X_i X_i'$ is pd w. pr. 1. Thus we can take $n = p$, so $X : p \times p$ is a square matrix. Then $|S| = |X|^2$, so it suffices to show that X itself is nonsingular w.pr.1. But

$$\{X \text{ singular}\} = \bigcup_{i=1}^{p} \{X_i \in \mathcal{S}_i \equiv \text{span}\,\{X_j \mid j \neq i\}\},$$

so

$$\begin{aligned}\Pr[X \text{ singular}] &\leq \sum_{i=1}^{p} \Pr[X_i \in \mathcal{S}_i] \\ &= \sum_{i=1}^{p} \mathrm{E}\,\{\Pr[X_i \in \mathcal{S}_i \mid X_j,\, j \neq i]\} = 0,\end{aligned}$$

since $\dim(\mathcal{S}_i) < p$ and the distribution of $X_i \sim N_p(0, \Sigma)$ is absolutely continuous w.r.to Lebesgue measure on $\mathcal{R}^p$. Thus $\Pr[X \text{ nonsingular}] = 1$.

($\Leftarrow$) *Method 2 (Okamoto (1973) Ann. Statist.)* Apply:

Lemma 3.3 *(Okamoto). Let $Z \equiv (Z_1, \ldots, Z_k) \in \mathcal{R}^k$ be a random vector with a pdf that is absolutely continuous w.r.to Lebesgue measure on $\mathcal{R}^k$. Let $g(z) \equiv g(z_1, \ldots, z_k)$ be a nontrivial polynomial (i.e., $g \not\equiv 0$). Then*

$$\Pr[\,g(Z) = 0\,] = 0. \tag{3.3}$$

Proof. *(sketch)* Use induction on k. The result is true for $k = 1$ since g can have only finitely many roots. Now assume the result is true for $k - 1$ and extend to k by Fubini's Theorem (equivalently, by conditioning on $Z_1, \ldots, Z_{k-1}$). □

Proposition 3.4. *Let $X : p \times n$ be a random matrix with a pdf that is absolutely continuous w.r.to Lebesgue measure on $\mathcal{R}^{p\times n}$. Then*

$$\Pr[\,X \text{ has full rank}\,] = 1. \tag{3.4}$$

If $p \leq n$, this implies that

$$\Pr[\, S \equiv XX' \textit{ is positive definite} \,] = 1. \tag{3.5}$$

Proof. Without loss of generality (wlog) assume that $p \leq n$ and partition X as (X_1, X_2) with $X_1 : p \times p$. Since $\text{rank}(X_1) < p$ iff $|X_1| = 0$, and since the determinant $|X_1| \equiv g(X_1)$ is a nontrivial polynomial,

$$\Pr[\,\text{rank}(X_1) = p\,] = 1$$

by Lemma 3.3. But $\text{rank}(X_1) = p \Rightarrow \text{rank}(X) = p$, so (3.4) holds. □

Okamoto's Lemma also yields the following important result:

Proposition 3.5. *Let $l_1(S) \geq \cdots \geq l_p(S)$ denote the eigenvalues (necessarily real) of $S \equiv XX'$. Under the assumptions of* Proposition 3.4,

$$\Pr[\, l_1(S) > \cdots > l_p(S) > 0\,] = 1. \tag{3.6}$$

Proof. *(sketch)* The eigenvalues of $S \equiv XX'$ are the roots of the nontrivial polynomial $h(l) \equiv |XX' - l\, I_p|$. These roots are distinct iff the discriminant of h is nonzero. Since the discriminant is itself a nontrivial polynomial of the coefficients of the polynomial h, hence a nontrivial polynomial of the elements of X, (3.6) follows from Okamoto's Lemma. □

Lemma 3.6. Additivity: *If $S_1 \perp\!\!\!\perp S_2$ with $S_i \sim W_p(n_i, \Sigma)$, then*

$$S_1 + S_2 \sim W_p(n_1 + n_2, \Sigma). \tag{3.7}$$

3.2. Covariance matrices of Kronecker product form.

If $X_1, \ldots, X_n$ are independent rvtrs each with covariance matrix $\Sigma : p \times p$, then $\text{Cov}(X) = \Sigma \otimes I_n$, a *Kronecker product*. We now determine how a covariance matrix of the general Kronecker product form $\text{Cov}(X) = \Sigma \otimes \Lambda$ transforms under a linear transformation AXB (see Proposition 3.9).

The **Kronecker product** of the $p \times q$ matrix A and the $m \times n$ matrix B is the $pm \times qn$ matrix

$$A \otimes B := \begin{pmatrix} Ab_{11} & \cdots & Ab_{1n} \\ \vdots & \ddots & \vdots \\ Ab_{m1} & \cdots & Ab_{mn} \end{pmatrix}.$$

(i) $A \otimes B$ is **bilinear:**

$$(\alpha_1 A_1 + \alpha_2 A_2) \otimes B = \alpha_1 (A_1 \otimes B) + \alpha_2 (A_2 \otimes B)$$
$$A \otimes (\beta_1 B_1 + \beta_2 B_2) = \beta_1 (A \otimes B_1) + \beta_2 (A \otimes B_2).$$

(ii)
$$(\underbrace{A}_{p\times q} \otimes \underbrace{B}_{m\times n})(\underbrace{C}_{q\times r} \otimes \underbrace{D}_{n\times s}) = (\underbrace{AC}_{p\times r} \otimes \underbrace{BD}_{m\times s}).$$

(iii)
$$(A \otimes B)' = A' \otimes B',$$
$$A = A',\ B = B' \implies A \otimes B = (A \otimes B)'.$$

(iv) If $\Gamma : p \times p$ and $\Psi : n \times n$ are orthogonal matrices, then $\Gamma \otimes \Psi : pn \times pn$ is orthogonal. [apply (ii) and (iii)]

(v) If $A : p \times p$ and $B : n \times n$ are real symmetric matrices with eigenvalues $\alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_n$, respectively, then $A \otimes B : pn \times pn$ is also real and symmetric with eigenvalues $\{\alpha_i \beta_j \mid i = 1, \ldots, p,\ j = 1, \ldots, n\}$.

Proof. Write the spectral decompositions of A and B as

$$A = \Gamma D_\alpha \Gamma', \qquad B = \Psi D_\beta \Psi',$$

respectively, where $D_\alpha = \text{diag}(\alpha_1, \ldots, \alpha_p)$ and $D_\beta = \text{diag}(\beta_1, \ldots, \beta_n)$. Then

$$\begin{aligned} A \otimes B &= (\Gamma D_\alpha \Gamma') \otimes (\Psi D_\beta \Psi') \\ &= (\Gamma \otimes \Psi)\,(D_\alpha \otimes D_\beta)\,(\Gamma \otimes \Psi)' \end{aligned} \tag{3.8}$$

by (ii) and(iii). Since $\Gamma \otimes \Psi$ is orthogonal and $D\alpha \otimes D_\beta$ is diagonal with diagonal entries $\{\alpha_i\beta_j \mid i = 1, \ldots, p,\ j = 1, \ldots, n\}$, (3.8) is a spectral decomposition of the real symmetric matrix $A \otimes B$, so the result follows.□

(vi)
$$\begin{aligned} A \text{ psd},\ B \text{ psd} &\implies A \otimes B \text{ psd}, \\ A \text{ pd},\ B \text{ pd} &\implies A \otimes B \text{ pd}. \qquad \text{[apply (3.8)]} \end{aligned}$$

Let $X \equiv (X_1, \ldots, X_n) : p \times n$ be a random matrix. By convention we shall define the covariance matrix $\text{Cov}(X)$ to be the covariance matrix of the $pn \times 1$ column vector $\widetilde{X}$ formed by "stacking" the column vectors of X:

$$\text{Cov}(X) := \text{Cov}(\widetilde{X}) \equiv \text{Cov}\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \equiv \begin{pmatrix} \text{Cov}(X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n) \end{pmatrix}.$$

Lemma 3.7. *Let $X = \{X_{ij}\}$, $\Sigma = \{\sigma_{ii'}\}$, $\Lambda = \{\lambda_{jj'}\}$. Then*

$$\text{Cov}(X) = \Sigma \otimes \Lambda \iff \text{Cov}(X_{ij},\, X_{i'j'}) = \sigma_{ii'}\lambda_{jj'}$$

for all $i, i' = 1, \ldots, p$ and all $j, j' = 1, \ldots, n$. [straightforward - verify] □

Lemma 3.8. $\text{Cov}(X) = \Sigma \otimes \Lambda \iff \text{Cov}(X') = \Lambda \otimes \Sigma$.

Proof. Set $U = X'$, so $U_{ij} = X_{ji}$. Then

$$\text{Cov}(U_{ij},\, U_{i'j'}) = \text{Cov}(X_{ji},\, X_{j'i'}) = \sigma_{jj'}\lambda_{ii'},$$

hence $\text{Cov}(X') = \text{Cov}(U) = \Lambda \otimes \Sigma$ by Lemma 3.7. □

Proposition 3.9. *If* $\mathrm{Cov}(X) = \Sigma \otimes \Lambda$ *then*

$$\mathrm{Cov}(\underbrace{A}_{q\times p}\underbrace{X}_{p\times n}\underbrace{B}_{n\times m}) = \underbrace{(A\Sigma A')}_{q\times q} \otimes \underbrace{(B'\Lambda B)}_{m\times m}. \tag{3.9}$$

Thus if $X \sim N_{p\times n}(\zeta,\, \Sigma \otimes \Lambda)$ *then*

$$AXB \sim N_{q\times m}\,(A\zeta B,\, (A\Sigma A')\,\otimes\,(B'\Lambda B)) \tag{3.10}$$

Proof. (a) Because $AX = (AX_1, \ldots, AX_n)$ it follows that

$$\widetilde{AX} = (A \otimes I_n)\widetilde{X},$$

so

$$\begin{aligned}\mathrm{Cov}(AX) \equiv \mathrm{Cov}(\widetilde{AX}) &= (A \otimes I_n)\,\mathrm{Cov}(\widetilde{X})\,(A \otimes I_n)' \\ &= (A \otimes I_n)\,(\Sigma \otimes \Lambda)\,(A \otimes I_n)' \\ &= (A\Sigma A')\,\otimes\,\Lambda \qquad \text{[by (ii)].}\end{aligned}$$

(b) Next,

$$\mathrm{Cov}(X') = \Lambda \otimes \Sigma \qquad \text{[Lemma 3.8],}$$

so

$$\mathrm{Cov}(B'X') = (B'\Lambda B) \otimes \Sigma \qquad \text{[(b)],}$$

hence

$$\mathrm{Cov}(XB) \equiv \mathrm{Cov}\,((B'X')') = \Sigma \otimes (B'\Lambda B) \quad \text{[Lemma 3.8].}$$

Looking ahead: Our goal will be to determine the joint distribution of the matrices $(S_{11\cdot 2},\, S_{12},\, S_{22})$ *that arise from a partitioned Wishart matrix* S. *In §3.4 we will see that the conditional distribution of* $S_{12} \mid S_{22}$ *follows a normal multivariate linear model (NMLM) of the form (3.14) in §3.3, whose covariance structure has Kronecker product form. Therefore we will first study this NMLM and determine the joint distribution of its MLEs* $(\hat{\beta},\, \hat{\Sigma})$ *given by (3.15) and(3.16). This will readily yield the joint distribution of* $(S_{11\cdot 2},\, S_{12},\, S_{22})$, *which in turn will have several interesting consequences, including the evaluation of* $\mathrm{E}(S^{-1})$ *and the distribution of Hotelling's* T^2 *statistic* $\bar{X}_n' S^{-1} \bar{X}_n$.

3.3. The multivariate linear model.

The standard *univariate linear model* consists of a set $X \equiv (X_1, \ldots, X_n)$ of uncorrelated univariate observations with common variance $\sigma^2 > 0$ such that $\mathrm{E}(X)$ lies in a specified linear subspace $L \subset \mathcal{R}^n$ with $\dim(L) = q < n$. If $Z : q \times n$ (the *design matrix*) is any fixed matrix whose rows span L then

$$L = \{\beta Z \mid \beta : 1 \times q \in \mathcal{R}^q\}, \tag{3.11}$$

so this linear model can be expressed as follows:

$$\begin{aligned} \mathrm{E}(X) &= \beta Z, \qquad && \beta : 1 \times q, \\ \mathrm{Cov}(X) &= \sigma^2 I_n, \qquad && \sigma^2 > 0. \end{aligned} \tag{3.12}$$

In the standard *multivariate linear model*, $X \equiv (X_1, \ldots, X_n) : p \times n$ is a series of uncorrelated p-variate observations with common covariance matrix $\Sigma > 0$ such that *each row* of $\mathrm{E}(X)$ lies in the specified linear subspace $L \subset \mathcal{R}^n$. This linear model can be expressed as follows:

$$\begin{aligned} \mathrm{E}(X) &= \beta Z, \qquad && \beta : p \times q, \\ \mathrm{Cov}(X) &= \Sigma \otimes I_n, \qquad && \Sigma > 0. \end{aligned} \tag{3.13}$$

If in addition we assume that $X_1, \ldots, X_n$ are normally distributed, then (3.13) can be expressed as the *normal multivariate linear model (NMLM)*

$$X \sim N_{p\times n}(\beta Z, \Sigma \otimes I_n), \qquad \beta : p \times q, \quad \Sigma > 0. \tag{3.14}$$

Often Z is called a *design matrix* for the linear model. We now assume that Z is of rank $q \leq n$, so ZZ' is nonsingular and β is identifiable:

$$\beta = (\mathrm{E}(X))\, Z'(ZZ')^{-1}.$$

The maximum likelihood estimator $(\hat\beta, \hat\Sigma)$. We now show that the MLE $(\hat\beta, \hat\Sigma)$ exists w. pr. 1 iff $n - q \geq p$ and is given by

$$\hat\beta = XZ'(ZZ')^{-1}, \tag{3.15}$$

$$\hat\Sigma = \tfrac{1}{n} X \left(I_n - Z'(ZZ')^{-1}Z\right) X' \equiv \tfrac{1}{n} XQX'. \tag{3.16}$$

Because the observation vectors $X_1, \dots, X_n$ are independent under the NMLM (3.14), the joint pdf of $X \equiv (X_1, \dots, X_n)$ is given by

$$
\begin{aligned}
f_{\beta,\Sigma}(x) &= \frac{c_1}{|\Sigma|^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\sum_{i=1}^n (x_i - \beta Z_i)'\Sigma^{-1}(x_i - \beta Z_i)} \\
&= \frac{c_1}{|\Sigma|^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}\left(\sum_{i=1}^n (x_i - \beta Z_i)(x_i - \beta Z_i)'\right)\right]} \\
&= \frac{c_1}{|\Sigma|^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}(x - \beta Z)(x - \beta Z)'\right]}, \qquad (3.17)
\end{aligned}
$$

where $c_1 = (2\pi)^{-\frac{np}{2}}$ and $Z_1, \dots, Z_n$ are the columns of Z. To find the MLEs $\hat{\beta}$, $\hat{\Sigma}$, first fix Σ and maximize (3.17) w.r.to β. This can be accomplished by "minimizing" the matrix-valued quadratic form

$$\Delta(\beta) := (X - \beta Z)(X - \beta Z)' \qquad (3.18)$$

w.r.to the *Loewner ordering*[2], which *a fortiori* minimizes $\operatorname{tr}[\Sigma^{-1}\Delta(\beta)]$ [verify]. Since each row of βZ lies in $L \equiv \text{row space}\,(Z) \subset \mathcal{R}^n$, this suggests that the minimizing $\hat{\beta}$ be chosen such that each row of $\hat{\beta} Z$ is the orthogonal projection of the corresponding row of X onto L. But the matrix of this orthogonal projection is

$$P \equiv Z'(ZZ')^{-1}Z \quad : n \times n$$

so we should choose $\hat{\beta}$ such that $\hat{\beta} Z = XZ'(ZZ')^{-1}Z$, or equivalently,

$$\hat{\beta} = XZ'(ZZ')^{-1}. \qquad (3.19)$$

To verify that $\hat{\beta}$ minimizes $\Delta(\beta)$, write $X - \beta Z = (X - \hat{\beta} Z) + (\hat{\beta} - \beta)Z$, so

$$
\begin{aligned}
\Delta(\beta) &= (X - \hat{\beta} Z)(X - \hat{\beta} Z)' + (\hat{\beta} - \beta)ZZ'(\hat{\beta} - \beta)' \\
&\quad + \underbrace{(X - \hat{\beta} Z)Z'}_{=0}(\hat{\beta} - \beta)' + (\hat{\beta} - \beta)\underbrace{Z(X - \hat{\beta} Z)'}_{=0}.
\end{aligned}
$$

[2] $T \geq S$ iff $T - S$ is psd.

Since ZZ' is pd, $\Delta(\beta)$ is uniquely minimized w.r. to the Loewner ordering when $\beta = \hat{\beta}$. Since $\hat{\beta}$ does not depend on Σ, this establishes (3.15). Thus

$$\begin{aligned}
(3.20)\quad \min_{\beta} \Delta(\beta) &= (X - \hat{\beta}Z)(X - \hat{\beta}Z)' \\
&= X\left(I_n - Z'(ZZ')^{-1}Z)(I_n - Z'(ZZ')^{-1}Z\right)' X' \\
&\equiv X\,(I_n - P)(I_n - P)'\,X' \\
&\equiv XQQ'X' \qquad [\text{set } Q = I_n - P] \\
&= XQX' \qquad [\,Q, \text{ like } P, \text{ is a projection matrix}]
\end{aligned}$$

Furthermore, it follows from (3.17) and (3.20) that for fixed $\Sigma > 0$,

$$\max_{\beta} f_{\beta,\Sigma}(x) = \frac{c_1}{|\Sigma|^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}xQx')}. \tag{3.21}$$

To maximize (3.21) w. r. to Σ we apply the following lemma:

Lemma 3.10. *If W is pd then*

$$\max_{\Sigma>0} \frac{1}{|\Sigma|^{\frac{n}{2}}}\, e^{-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}W)} = \frac{1}{|\hat{\Sigma}|^{\frac{n}{2}}} \cdot e^{-\frac{np}{2}}, \tag{3.22}$$

where $\hat{\Sigma} \equiv \frac{1}{n}W$ is the unique maximizing value of Σ.

Proof. Since the mappings

$$\begin{aligned}
\Sigma &\mapsto \Sigma^{-1} &&:= \Lambda \\
\Lambda &\mapsto (W^{\frac{1}{2}})'\Lambda W^{\frac{1}{2}} &&:= \Omega
\end{aligned}$$

are both bijections of $\mathcal{S}_p^+$ onto itself, the maximum in (3.22) is given by

$$\begin{aligned}
(3.23)\quad \max_{\Lambda>0} |\Lambda|^{\frac{n}{2}}\; e^{-\frac{1}{2}\operatorname{tr}(\Lambda W)} &= \frac{1}{|W|^{\frac{n}{2}}} \max_{\Omega>0} |\Omega|^{\frac{n}{2}}\; e^{-\frac{1}{2}\operatorname{tr}\Omega} \\
&= \frac{1}{|W|^{\frac{n}{2}}} \max_{\omega_1\ge\cdots\ge\omega_p>0} \prod_{i=1}^{p} \omega_i^{\frac{n}{2}}\, e^{-\frac{1}{2}\omega_i},
\end{aligned}$$

where $\omega_1, \ldots, \omega_p$ are the eigenvalues of Ω. Since $n \log \omega - \omega$ is strictly concave in ω, its maximum value n is uniquely attained at $\hat{\omega} = n$, hence the maximizing values of $\omega_1, \ldots, \omega_p$ are $\hat{\omega}_1 = \cdots = \hat{\omega}_p = n$. Thus the unique maximizing value of Ω is $\hat{\Omega} = nI_p$, hence $\hat{\Lambda} = nW^{-1}$ and $\hat{\Sigma} = \frac{1}{n}W$. □

If W is psd but singular, then the maximum in (3.23) is $+\infty$ [verify]. Thus the MLE $\hat{\Sigma}$ for the NMLM (3.14) exists and is given by $\hat{\Sigma} = \frac{1}{n}XQX'$ iff XQX' is pd. We now derive the distribution of XQX' and show that

$$XQX' \text{ is pd w. pr. } 1 \iff n - q \geq p. \tag{3.24}$$

Thus the condition $n - q \geq p$ is necessary and sufficient for the existence and uniqueness of the MLE $\hat{\Sigma}$ as stated in (3.16).

First we find the joint distn of $(\hat{\beta}, \hat{\Sigma})$. From (3.14) and (3.10),

$$\begin{aligned} X(Z', Q) &\sim N_{p\times(q+n)}\left(\beta Z(Z', Q),\, \Sigma \otimes \left(\begin{pmatrix} Z \\ Q \end{pmatrix} (Z', Q)\right)\right) \\ &= N_{p\times(q+n)}\left((\beta ZZ', 0),\, \Sigma \otimes \begin{pmatrix} ZZ' & 0 \\ 0 & Q \end{pmatrix}\right) \quad [ZQ = 0], \end{aligned}$$

from which it follows that

$$XZ' \sim N_{p\times q}\left(\beta ZZ',\, \Sigma \otimes (ZZ')\right), \tag{3.25}$$

$$XQ \sim N_{p\times n}\left(0,\, \Sigma \otimes Q\right), \tag{3.26}$$

$$XZ' \perp\!\!\!\perp XQ. \tag{3.27}$$

Because $Q \equiv I_n - Z'(ZZ')^{-1}Z$ is a projection matrix with

$$\mathrm{rank}(Q) = \mathrm{tr}(Q) = n - q,$$

its spectral decomposition is (recall (1.37))

$$Q = \Gamma \begin{pmatrix} I_{n-q} & 0 \\ 0 & 0 \end{pmatrix} \Gamma' \tag{3.28}$$

for some $p \times p$ orthogonal matrix Γ. Set $V = XQ\Gamma$, so from (3.26),

$$V \sim N_{p\times n}\left(0,\, \Sigma \otimes \begin{pmatrix} I_{n-q} & 0 \\ 0 & 0 \end{pmatrix}\right).$$

This shows that [verify]

$$XQX' \equiv VV' \sim W_p(n-q,\, \Sigma), \tag{3.29}$$

hence (3.24) follows from Lemma 3.2. Lastly, by (3.25), (3.29), and (3.27),

$$\hat{\beta} \equiv XZ'(ZZ')^{-1} \sim N_{p\times q}\left(\beta,\, \Sigma \otimes (ZZ')^{-1}\right), \tag{3.30}$$

$$n\hat{\Sigma} \equiv XQX' \sim W_p(n-q,\, \Sigma), \tag{3.31}$$

$$\hat{\beta} \perp\!\!\!\perp \hat{\Sigma}. \tag{3.32}$$

Remark 3.11. From (3.31), the MLE $\hat{\Sigma}$ is a biased estimator of Σ:

$$\mathrm{E}(\hat{\Sigma}) = \left(1 - \frac{q}{n}\right)\Sigma.$$

Instead, the adjusted MLE $\breve{\Sigma} := \frac{1}{n-q} XQX'$ is unbiased. □

Special case of the NMLM: a random sample from $N_p(\mu,\, \Sigma)$.

If $X_1, \ldots, X_n$ is an i.i.d. sample from $N_p(\mu,\, \Sigma)$ then the joint distribution of $X \equiv (X_1, \ldots, X_n)$ is a special case of the NMLM (3.13):

$$X \sim N_{p\times n}(\mu \mathbf{e}_n',\, \Sigma \otimes I_n), \qquad \mu : p \times 1, \quad \Sigma > 0. \tag{3.33}$$

Here $q = 1$, $Z = \mathbf{e}_n'$, and $Q = I_n - \mathbf{e}_n(\mathbf{e}_n'\mathbf{e}_n)^{-1}\mathbf{e}_n'$, so from (3.30) - (3.32),

$$\hat{\mu} = X\mathbf{e}_n(\mathbf{e}_n'\mathbf{e}_n)^{-1} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \sim N_p\left(\mu,\, \tfrac{1}{n}\Sigma\right), \tag{3.34}$$

$$n\hat{\Sigma} = XQX' = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)\left(X_i - \bar{X}_n\right)' \sim W_p(n-1,\, \Sigma), \tag{3.35}$$

$$\bar{X}_n \perp\!\!\!\perp \hat{\Sigma}\,. \tag{3.36}$$

3.4. Distribution of a partitioned Wishart matrix.

Let $\mathcal{S}_p^+$ denote the convex cone of real positive definite $p \times p$ matrices and let $\mathcal{M}_{m\times n}$ denote the vector space of all real $m \times n$ matrices. Partition the pd matrix $S : p \times p \in \mathcal{S}_p^+$ as

$$S = \begin{matrix} & p_1 & p_2 \\ p_1 & S_{11} & S_{12} \\ p_2 & S_{21} & S_{22} \end{matrix}, \tag{3.37}$$

where $p_1 + p_2 = p$. Now apply (1.32) and (1.34) to obtain:

Lemma 3.12. *The following correspondence is bijective:*

$$\begin{aligned} \mathcal{S}_p^+ &\leftrightarrow \mathcal{S}_{p_1}^+ \times \mathcal{M}_{p_1\times p_2} \times \mathcal{S}_{p_2}^+ \\ S &\leftrightarrow (S_{11\cdot 2},\, S_{12},\, S_{22}). \end{aligned} \tag{3.38}$$

Note that we cannot replace $S_{11\cdot 2}$ by S_{11} in (3.38) because of the constraints imposed on S itself by the pd condition. That is, the range of $(S_{11},\, S_{12},\, S_{22})$ is not the Cartesian product of the three ranges.

Proposition 3.13.*** *Let $S \sim W_p(n,\, \Sigma)$ be partitioned as in (3.37) with $n \geq p_2$ and $\Sigma_{22} > 0$. Then the joint distribution of $(S_{11\cdot 2},\, S_{12},\, S_{22})$ can be specified as follows:*

$$S_{12} \,\big|\, S_{22} \sim N_{p_1\times p_2}\left(\Sigma_{12}\Sigma_{22}^{-1}S_{22},\, \Sigma_{11\cdot 2} \otimes S_{22}\right), \tag{3.39}$$

$$S_{22} \sim W_{p_2}(n,\, \Sigma_{22}), \tag{3.40}$$

$$S_{11\cdot 2} \sim W_{p_1}(n - p_2,\, \Sigma_{11\cdot 2}), \tag{3.41}$$

$$S_{11\cdot 2} \perp\!\!\!\perp (S_{12},\, S_{22}). \tag{3.42}$$

Proof. Represent S as XX' with $X \equiv \begin{matrix} p_1 \\ p_2 \end{matrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{p\times n}(0,\, \Sigma \otimes I_n)$, so

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} X_1X_1' & X_1X_2' \\ X_2X_1' & X_2X_2' \end{pmatrix}. \tag{3.43}$$

By Proposition 3.4, the conditions $n \geq p_2$ and $\Sigma_{22} > 0$ imply that $\operatorname{rank}(X_2) = p_2$ w. pr. 1, hence $S_{22} \equiv X_2X_2'$ is pd w. pr. 1. Thus $S_{11\cdot 2}$ is well defined and is given by

$$S_{11\cdot 2} = X_1\left(I_n - X_2'(X_2X_2')^{-1}X_2\right)X_1' \equiv X_1QX_1'. \tag{3.44}$$

From (2.8) the conditional distribution of $X_1 \mid X_2$ is given by

$$X_1 \bigm| X_2 \sim N_{p_1 \times n}\left(\Sigma_{12}\Sigma_{22}^{-1}X_2,\, \Sigma_{11\cdot 2} \otimes I_n\right), \tag{3.45}$$

which is a NMLM (3.14) with the following correspondences:

$$\begin{array}{lll} X \leftrightarrow X_1, & \beta \leftrightarrow \Sigma_{12}\Sigma_{22}^{-1}, & p \leftrightarrow p_1, \\ Z \leftrightarrow X_2, & \Sigma \leftrightarrow \Sigma_{11\cdot 2}, & q \leftrightarrow p_2. \end{array}$$

Thus from (3.25), (3.31), (3.32), (3.43), and (3.44), conditionally on X_2,

$$S_{12} \bigm| X_2 \sim N_{p_1 \times p_2}\left(\Sigma_{12}\Sigma_{22}^{-1}S_{22},\, \Sigma_{11\cdot 2} \otimes S_{22}\right), \tag{3.46}$$

$$S_{11\cdot 2} \bigm| X_2 \sim W_{p_1}(n - p_2,\, \Sigma_{11\cdot 2}), \tag{3.47}$$

$$S_{12} \perp\!\!\!\perp S_{11\cdot 2} \mid X_2. \tag{3.48}$$

Clearly (3.46) $\Rightarrow$ (3.39), while (3.40) follows from Lemma 3.1 with $A = (0_{p_2 \times p_1}, I_{p_2})$. Also, (3.47) $\Rightarrow$ (3.41) and (3.47) $\Rightarrow$ $S_{11\cdot 2} \perp\!\!\!\perp X_2$, which combines with (3.48) to yield $S_{11\cdot 2} \perp\!\!\!\perp (S_{12}, X_2)$,[3] which implies (3.42). $\square$

Note that (3.39) can be restated in two equivalent forms:

$$S_{12}S_{22}^{-1} \bigm| S_{22} \sim N_{p_1 \times p_2}\left(\Sigma_{12}\Sigma_{22}^{-1},\, \Sigma_{11\cdot 2} \otimes S_{22}^{-1}\right), \tag{3.49}$$

$$S_{12}S_{22}^{-\frac{1}{2}\prime} \bigm| S_{22} \sim N_{p_1 \times p_2}\left(\Sigma_{12}\Sigma_{22}^{-1}S_{22}^{\frac{1}{2}},\, \Sigma_{11\cdot 2} \otimes I_{p_2}\right), \tag{3.50}$$

where $S_{22}^{\frac{1}{2}}$ can be any (Borel-measurable) square root of S_{22} It follows from (3.50) and (3.42) that

$$\Sigma_{12} = 0 \implies S_{12}S_{22}^{-\frac{1}{2}\prime} \perp\!\!\!\perp S_{22} \perp\!\!\!\perp S_{11\cdot 2}. \tag{3.51}$$

We remark that Proposition 3.13 can also be derived directly from the pdf of the Wishart distribution, the existence of which requires the stronger conditions $n \geq p$ and $\Sigma > 0$. We shall derive the Wishart pdf in §8.4.

Proposition 3.13 yields many useful results – some examples follow.

---

[3] Because $A \perp\!\!\!\perp B \mid C$ and $B \perp\!\!\!\perp C \Rightarrow B \perp\!\!\!\perp (A, C)$ [verify].

Example 3.14. Distribution of the generalized variance.

If $S \sim W_p(n, \Sigma)$ with $n \geq p$ and $\Sigma > 0$ then

$$|S| \sim |\Sigma| \cdot \prod_{i=1}^{p} \chi^2_{n-p+i}, \tag{3.52}$$

a product of independent chi-square variates.

Proof. Partition S as in (3.37) with $p_1 = 1$, $p_2 = p - 1$. Then

$$\begin{aligned} |S| = |S_{11\cdot 2}| \cdot |S_{22}| &\sim |W_1(n-p+1,\, \Sigma_{11\cdot 2})| \cdot |W_{p-1}(n,\, \Sigma_{22})| \\ &\sim \left(\Sigma_{11\cdot 2}\, \chi^2_{n-p+1}\right) \cdot |W_{p-1}(n,\, \Sigma_{22})| \end{aligned}$$

with the two factors independent. The result follows by induction on p. □

Note that (3.52) implies that although $\frac{1}{n}S$ is an unbiased estimator of Σ, $\left|\frac{1}{n}S\right|$ is a biased estimator of $|\Sigma|$:

$$\mathrm{E}\left|\tfrac{1}{n}S\right| = |\Sigma| \cdot \textstyle\prod_{i=1}^{p} \left(\frac{n-p+i}{n}\right) < |\Sigma| \, . \tag{3.53}$$

Proposition 3.15. *Let $S \sim W_p(n, \Sigma)$ with $n \geq p$ and $\Sigma > 0$. If $A : q \times p$ has rank $q \leq p$ then*

$$\left(AS^{-1}A'\right)^{-1} \sim W_q\left(n - p + q,\, \left(A\Sigma^{-1}A'\right)^{-1}\right). \tag{3.54}$$

When $A = a' : 1 \times p$ this becomes

$$\frac{1}{a'S^{-1}a} \sim \frac{1}{a'\Sigma^{-1}a} \cdot \chi^2_{n-p+1}. \tag{3.55}$$

Note: Compare (3.54) to (3.1): $ASA' \sim W_q(n,\, A\Sigma A')$, which holds with no restrictions on n, p, Σ, A, or q.

Our proof of (3.54) requires the **singular value decomposition** of A:

Lemma 3.16. *If $A : q \times p$ has rank $q \le p$ then there exist an orthogonal matrix $\Gamma : q \times q$ and a row-orthogonal matrix $\Psi_1 : q \times p$ such that*

$$A = \Gamma D_a \Psi_1, \tag{3.56}$$

where $D_a = \mathrm{diag}(a_1, \ldots, a_q)$ and $a_1^2 \ge \cdots \ge a_q^2 > 0$ are the ordered eigenvalues of AA'.[4] By extending Ψ_1 to a $p \times p$ orthogonal matrix $\Psi \equiv \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}$, we have the alternative representations

$$A = \Gamma \left(D_a,\, 0_{q\times(p-q)}\right) \Psi, \tag{3.57}$$

$$= C \left(I_q,\, 0_{q\times(p-q)}\right) \Psi, \tag{3.58}$$

where $C \equiv \Gamma D_a : q \times q$ is nonsingular.

Proof. Let $AA' = \Gamma D_a^2 \Gamma'$ be the spectral decomposition of the pd $q \times q$ matrix AA'. Thus

$$D_a^{-1}\Gamma' AA' \Gamma D_a^{-1} = I_q,$$

so $\Psi_1 := D_a^{-1}\Gamma' A : q \times p$ satisfies $\Psi_1 \Psi_1' = I_q$, i.e., the rows of Ψ_1 are orthonormal. Thus (3.56) holds, then (3.57) and (3.58) are immediate. □

Proof of Proposition 3.15. It follows from (3.58) that [verify]

$$\left(AS^{-1}A'\right)^{-1} = C'^{-1}\check{S}_{11\cdot 2}C^{-1},$$
$$\left(A\Sigma^{-1}A'\right)^{-1} = C'^{-1}\check{\Sigma}_{11\cdot 2}C^{-1},$$

where $\check{S} = \Psi S \Psi'$ and $\check{\Sigma} = \Psi \Sigma \Psi'$ are partitioned as in (3.37) with $p_1 = q$ and $p_2 = p - q$. Since $\check{S} \sim W_p(n,\, \check{\Sigma})$, it follows from Proposition 3.13 that

$$\check{S}_{11\cdot 2} \sim W_q\left(n - (p-q),\, \check{\Sigma}_{11\cdot 2}\right),$$

so

$$C'^{-1}\check{S}_{11\cdot 2}C^{-1} \sim W_q\left(n - (p-q),\, C'^{-1}\check{\Sigma}_{11\cdot 2}C^{-1}\right),$$

which gives (3.54). □

[4] $a_1 \ge \cdots \ge a_q > 0$ are called the *singular values* of A.

Example 3.17. Distribution of Hotelling's T^2 statistic.

Let $X \sim N_p(\mu, \Sigma)$ and $S \sim W_p(n, \Sigma)$ be independent, $n \geq p$, $\Sigma > 0$. Define

$$T^2 = X'S^{-1}X.$$

Then

$$T^2 \sim \frac{\chi^2_p(\mu'\Sigma^{-1}\mu)}{\chi^2_{n-p+1}} \equiv F_{p,\,n-p+1}(\mu'\Sigma^{-1}\mu), \tag{3.59}$$

a (nonnormalized) noncentral F distribution. (The two chi-square variates are independent.)

Proof. Decompose T^2 as $\left(\frac{X'S^{-1}X}{X'\Sigma^{-1}X}\right) \cdot X'\Sigma^{-1}X$. By (3.55) and the independence of X and S,

$$X'S^{-1}X \;\Big|\; X \sim X'\Sigma^{-1}X \cdot \frac{1}{\chi^2_{n-p+1}},$$

so

$$\frac{X'S^{-1}X}{X'\Sigma^{-1}X} \;\Bigg|\; X \sim \frac{1}{\chi^2_{n-p+1}},$$

independent of X. Since $X'\Sigma^{-1}X \sim \chi^2_p(\mu'\Sigma^{-1}\mu)$ by (2.30), (3.59) holds. □

For any fixed $\mu_0 \in \mathcal{R}^p$, replace X and μ in Example 3.17 by $X - \mu_0$ and $\mu - \mu_0$, respectively, to obtain the following generalization of (3.59):

$$\begin{aligned} T^2 &\equiv (X-\mu_0)'S^{-1}(X-\mu_0) \\ &\sim \frac{\chi^2_p\left((\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)\right)}{\chi^2_{n-p+1}} \equiv F_{p,\,n-p+1}\left((\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)\right). \end{aligned} \tag{3.60}$$

Note: In Example 6.11 and Exercise 6.12 it will be shown that T^2 is the UMP invariant test statistic and the LRT statistic for testing $\mu = \mu_0$ vs. $\mu \neq \mu_0$ with Σ unknown. When $\mu = \mu_0$,

$$T^2 \sim F_{p,\,n-p+1}, \tag{3.61}$$

which determines the significance level of the test. □

Example 3.18. Expected value of S^{-1}.

Suppose that $S \sim W_p(n, \Sigma)$ with $n \geq p$ and $\Sigma > 0$, so S^{-1} exists with pr. 1. When does $\mathrm{E}(S^{-1})$ exist, and what is its value? We answer this by combining Proposition 3.13 with an invariance argument.

First consider the case $\Sigma = I$. Partition S and S^{-1} as

$$S = \begin{pmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \qquad S^{-1} = \begin{pmatrix} s^{11} & S^{12} \\ S^{21} & S^{22} \end{pmatrix},$$

respectively, with $p_1 = 1$ and $p_2 = p - 1$. Then by (3.41),

$$s^{11} = \frac{1}{s_{11\cdot 2}} \sim \frac{1}{\chi^2_{n-p+1}},$$

so

$$\mathrm{E}(s^{11}) = \tfrac{1}{n-p-1} < \infty \quad \text{iff} \quad n \geq p+2. \tag{3.62}$$

Similarly for the other diagonal elements of S^{-1}: $\mathrm{E}(s^{ii}) < \infty$ iff $n \geq p + 2$. Because each off-diagonal element s^{ij} of S^{-1} satisfies

$$|s^{ij}| \leq \sqrt{s^{ii}s^{jj}} \leq \tfrac{1}{2}(s^{ii} + s^{jj}),$$

we see that $\mathrm{E}(S^{-1}) =: \Delta$ exists iff $n \geq p + 2$. Furthermore, because $\Sigma = I$, $S \sim \Gamma S \Gamma'$ for every $p \times p$ orthogonal matrix Γ, hence

$$\Gamma \Delta \Gamma' = \Gamma\, \mathrm{E}(S^{-1})\, \Gamma' = \mathrm{E}\left((\Gamma S \Gamma')^{-1}\right) = \mathrm{E}(S^{-1}) = \Delta \quad \forall\, \Gamma.$$

Exercise 3.19. Show that $\Gamma\Delta\Gamma' = \Delta \;\forall\,\Gamma \Rightarrow \Delta = \delta\, I$ for some scalar δ. □

Thus $\mathrm{E}(S^{-1}) = \delta\, I$, and $\delta = \frac{1}{n-p-1}$ by (3.62). Therefore when $\Sigma = I$,

$$\mathrm{E}(S^{-1}) = \tfrac{1}{n-p-1}\, I \qquad\qquad (n \geq p+2).$$

Now consider the general case $\Sigma > 0$. Since

$$S \sim \Sigma^{\frac{1}{2}} \check{S} \Sigma^{\frac{1}{2}} \quad \text{with } \check{S} \sim W_p(n, I),$$

we conclude that

$$
\begin{aligned}
\mathrm{E}(S^{-1}) &= \mathrm{E}\left(\left(\Sigma^{\frac{1}{2}}\check{S}\Sigma^{\frac{1}{2}}\right)^{-1}\right) \\
&= \Sigma^{-\frac{1}{2}}\mathrm{E}(\check{S}^{-1})\Sigma^{-\frac{1}{2}} \\
&= \tfrac{1}{n-p-1}\Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}} \\
&= \tfrac{1}{n-p-1}\Sigma^{-1} \qquad\qquad (n \geq p+2). \qquad (3.63)
\end{aligned}
$$

Proposition 3.20. Bartlett's decomposition.

Let $S \sim W_p(n,\, I)$ with $n \geq p$. Set $S = TT'$ where $T \equiv \{t_{ij} \mid 1 \leq j \leq i \leq p\}$ is the unique lower triangular square root of S with $t_{ii} > 0$, $i = 1, \ldots, p$ (see Exercise 1.5). Then the $\{t_{ij}\}$ are mutually independent rvs with

$$
(3.64) \qquad \begin{cases} t_{ii}^2 \sim \chi^2_{n-i+1}, & 1 = 1, \ldots, p, \\ t_{ij} \sim N_1(0,\, 1), & 1 \leq j < i \leq p. \end{cases}
$$

Proof. Use induction on p. The result is obvious for $p = 1$. Partition S as in (3.37) with $p_1 = p-1$ and $p_2 = 1$ so by the induction hypothesis, $S_{11} = T_1 T_1'$ for a lower triangular matrix T_1 that satisfies (3.64) with p replaced by $p-1$. Then

$$
S \equiv \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} T_1 & 0 \\ S_{21}T_1^{-1'} & s_{22\cdot 1}^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} T_1' & T_1^{-1}S_{12} \\ 0 & s_{22\cdot 1}^{\frac{1}{2}} \end{pmatrix} \equiv TT',
$$

where $T : p \times p$ is lower triangular with $t_{ii} > 0$, $i = 1, \ldots, p$. Since $T_1 = S_{11}^{\frac{1}{2}}$ and $\Sigma = I$, it follows from (3.51), (3.50), and (3.41) (with the indices "1" and "2" interchanged) that

$$
\begin{aligned}
&S_{21}T_1^{-1'} \perp\!\!\!\perp T_1 \perp\!\!\!\perp s_{22\cdot 1} \\
&S_{21}T_1^{-1'} \sim N_{1\times(p-1)}\left(0,\, 1 \otimes I_{p-1}\right), \\
&\quad s_{22\cdot 1} \sim \chi^2_{n-p+1}\ ,
\end{aligned}
$$

from which the induction step follows. $\square$

Example 3.21. Distribution of the sample multiple correlation coefficient R^2.

Let $S \sim W_p(n, \Sigma)$ with $n \geq p$ and $\Sigma > 0$. Partition S and Σ as

$$S = \begin{matrix} & \begin{matrix} 1 & p-1 \end{matrix} \\ \begin{matrix} 1 \\ p-1 \end{matrix} & \begin{pmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \end{matrix}, \qquad \Sigma = \begin{matrix} & \begin{matrix} 1 & p-1 \end{matrix} \\ \begin{matrix} 1 \\ p-1 \end{matrix} & \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{matrix}, \tag{3.65}$$

and define

$$R^2 = \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}}, \qquad \rho^2 = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\sigma_{11}},$$

$$\begin{aligned} U &= \frac{R^2}{1-R^2} = \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11\cdot 2}} = \frac{\left(S_{12}S_{22}^{-1/2}\right)\left(S_{12}S_{22}^{-1/2}\right)'}{s_{11\cdot 2}}, \\ \zeta &= \frac{\rho^2}{1-\rho^2} = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\sigma_{11\cdot 2}} \\ V \equiv V(S_{22}, \Sigma) &= \frac{\Sigma_{12}\Sigma_{22}^{-1}S_{22}\Sigma_{22}^{-1}\Sigma_{21}}{\sigma_{11\cdot 2}} \ . \end{aligned} \tag{3.66}$$

From Proposition 3.13 and (3.50) we have

$$\begin{aligned} S_{12}S_{22}^{-1/2'} \mid S_{22} &\sim N_{1\times(p-1)}\left(\Sigma_{12}\Sigma_{22}^{-1}S_{22}^{1/2},\, \sigma_{11.2} \otimes I_{p-1}\right), \\ s_{11\cdot 2} &\sim \sigma_{11.2} \cdot \chi^2_{n-p+1}, \\ S_{22} &\sim W_{p-1}(n, \Sigma_{22}), \\ s_{11\cdot 2} &\perp\!\!\!\perp (S_{12}, S_{22}), \end{aligned}$$

so [verify]

$$\begin{aligned} U \bigm| S_{22} &\sim \frac{\chi^2_{p-1}(V)}{\chi^2_{n-p+1}} \stackrel{\text{distn}}{=} F_{p-1,\, n-p+1}(V), \\ V &\sim \zeta \cdot \chi^2_n. \end{aligned}$$

Therefore the joint distribution of $(U, V) \equiv (U, V(S_{22}, \Sigma))$ is given by

$$\begin{aligned} U \mid V &\sim F_{p-1,\, n-p+1}(V), \\ V &\sim \zeta \cdot \chi^2_n. \end{aligned} \tag{3.67}$$

Equivalently, if we set $Z := V/\zeta$ so Z is ancillary (but unobservable), then

$$\begin{aligned} U \,\big|\, Z &\sim F_{p-1,\, n-p+1}(\zeta Z), \\ Z &\sim \chi^2_n, \end{aligned} \tag{3.68}$$

from which the unconditional distribution of U can be obtained by averaging over Z (see Exercise 3.22 and Example A.18 in Appendix A). □

Exercise 3.22. From (A.7) in Appendix A, the conditional distribution $F_{p-1,n-p+1}(\zeta Z)$ of $U \mid Z$ can be represented as a Poisson mixture of central F distributions:

$$F_{p-1,\, n-p+1}(\zeta Z) \,\big|\, Z, K \sim F_{p-1+2K,\, n-p+1}, \quad K|Z \sim \text{Poisson}\,(\zeta Z/2)\,. \tag{3.69}$$

Use (3.68), (3.69), and (A.8) to show that the unconditional distribution of U (resp., R^2) can be represented as a negative binomial mixture of central F (resp., Beta) rvs:

$$U \,\big|\, K \;\sim\; F_{p-1+2K,\, n-p+1}\,, \tag{3.70}$$

$$R^2 \equiv \frac{U}{U+1} \,\bigg|\, K \;\sim\; B\big(\tfrac{p-1}{2} + K,\, \tfrac{n-p+1}{2}\big)\,, \tag{3.71}$$

$$K \sim \text{Negative binomial}\,(\rho^2), \tag{3.72}$$

that is,

$$\Pr[\,K = k\,] = \frac{\Gamma\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{n}{2}\right)\, k!} \left(1-\rho^2\right)^{\frac{n}{2}} \left(\rho^2\right)^k, \quad k = 0, 1, \ldots. \quad □$$

Note: In Example 6.26 and Exercise 6.27 it will be shown that R^2 is the LRT statistic and the UMP invariant test statistic for testing $\rho^2 = 0$ vs. $\rho^2 > 0$. When $\rho^2 = 0$ ($\iff \Sigma_{12} = 0 \iff \zeta = 0$), $U \perp\!\!\!\perp Z$ by (3.68) and

$$U \sim F_{p-1,\, n-p+1}, \tag{3.73}$$

$$R^2 \sim B\big(\tfrac{p-1}{2},\, \tfrac{n-p+1}{2}\big), \tag{3.74}$$

either of which determines the significance level of the test. □

4. The Wishart Density; Jacobians of Matrix Transformations.

We have deduced properties of a Wishart random matrix $S \sim W_p(n, \Sigma)$ by using its representation $S = XX'$ in terms of a multivariate normal random matrix $X \sim N_{p\times n}(0, \Sigma \otimes I_n)$. We have not required the density of the Wishart distribution on $\mathcal{S}_p^+$ (the cone of $p\times p$ positive definite symmetric matrices). In this section we derive this density, a multivariate extension of the (central) chi-square density. Throughout it is assumed that $n \geq p$.

Assume first that $\Sigma = I$. From Bartlett's decomposition $S = TT'$ in Proposition 3.20, the joint pdf of $T \equiv \{T_{ij}\}$ is given by [verify!]

$$
\begin{aligned}
f(T) &= \prod_{1\leq j<i\leq p} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t_{ij}^2} \cdot \prod_{i=1}^{p} \frac{1}{2^{\frac{n-i-1}{2}}\Gamma\left(\frac{n-i+1}{2}\right)} t_{ii}^{n-i} e^{-\frac{1}{2}t_{ii}^2} \\
&= \frac{1}{2^{\frac{pn}{2}-p}\pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p}\Gamma\left(\frac{n-i+1}{2}\right)} \prod_{i=1}^{p} t_{ii}^{n-i} \cdot \exp\Big(-\frac{1}{2}\sum_{1\leq j\leq i\leq p} t_{ij}^2\Big) \qquad (4.1) \\
&=: c'_{p,n} \cdot \prod_{i=1}^{p} t_{ii}^{n-i} \cdot \exp\Big(-\frac{1}{2}\mathrm{tr}\, TT'\Big).
\end{aligned}
$$

Since the pdf of S is given by $f(S) = f(T)\left|\frac{\partial T}{\partial S}\right|$, we first must find the Jacobian $\left|\frac{\partial S}{\partial T}\right| \equiv 1/\left|\frac{\partial T}{\partial S}\right|$ of the mapping $S = TT'$. [This derivation of the Wishart pdf will resume in §4.4.]

4.1. Jacobians of vector/matrix transformations.

Consider a smooth bijective mapping ($\equiv$ diffeomorphism)

$$
\begin{aligned}
&A \to B \\
&x \equiv (x_1, \ldots, x_n) \mapsto y \equiv (y_1, \ldots, y_n),
\end{aligned}
\qquad (4.2)
$$

where A and B are open subsets of $\mathcal{R}^n$. The *Jacobian matrix* of this mapping is given by

$$
\left(\frac{\partial y}{\partial x}\right) := \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_n}{\partial x_n} \end{pmatrix}, \qquad (4.3)
$$

and the *Jacobian* of the mapping is given by $\left|\frac{\partial y}{\partial x}\right| := [\det\left(\frac{\partial y}{\partial x}\right)]^+$. Jacobians obey several elementary properties.

Chain rule: Suppose that $x \mapsto y$ and $y \mapsto z$ are diffeomorphisms. Then $x \mapsto z$ is a diffeomorphism and

$$\left|\frac{\partial z}{\partial x}\right| = \left|\frac{\partial z}{\partial y}\right|_{y=y(x)} \cdot \left|\frac{\partial y}{\partial x}\right|. \tag{4.4}$$

Proof. This follows from the chain rule for partial derivatives:

$$\frac{\partial z_i(y_1(x_1,\ldots,x_n),\ldots,y_n(x_1,\ldots,x_n))}{\partial x_j} = \sum_k \frac{\partial z_i}{\partial y_k}\frac{\partial y_k}{\partial x_j} = \left[\left(\frac{\partial z}{\partial y}\right)\left(\frac{\partial y}{\partial x}\right)\right]_{ij}.$$

Therefore $\left(\frac{\partial z}{\partial x}\right) = \left(\frac{\partial z}{\partial y}\right)\left(\frac{\partial y}{\partial x}\right)$; now take determinants. □

Inverse rule: Suppose that $x \mapsto y$ is a diffeomorphism. Then

$$\left|\frac{\partial x}{\partial y}\right|_{y=y(x)} = \left|\frac{\partial y}{\partial x}\right|^{-1}. \tag{4.5}$$

Proof. Apply the chain rule with $z = x$. □

Combination rule: Suppose that $x \mapsto u$ and $y \mapsto v$ are (unrelated) diffeomorphisms. Then

$$\left|\frac{\partial(u,v)}{\partial(x,y)}\right| = \left|\frac{\partial u}{\partial x}\right| \cdot \left|\frac{\partial v}{\partial y}\right|. \tag{4.6}$$

Proof. The Jacobian matrix is given by

$$\left(\frac{\partial(u,v)}{\partial(x,y)}\right) = \begin{pmatrix} \frac{\partial u}{\partial x} & 0 \\ 0 & \frac{\partial v}{\partial y} \end{pmatrix}.$$

Extended combination rule: Suppose that $(x,y) \mapsto (u,v)$ is a diffeomorphism of the form $u = u(x)$, $v = v(x,y)$. Then (4.6) continues to hold.

Proof. The Jacobian matrix is given by

$$\left(\frac{\partial(u,v)}{\partial(x,y)}\right) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ 0 & \frac{\partial v}{\partial y} \end{pmatrix}.$$

4.2. Jacobians of linear mappings. Let

$A\colon p\times p$ and $B\colon n\times n$ be nonsingular matrices,
$L\colon p\times p$ and $M\colon p\times p$ be nonsingular lower triangular matrices,
$U\colon p\times p$ and $V\colon p\times p$ be nonsingular upper triangular matrices,
c a nonzero scalar.

($A, B, L, M, U, V,$ c are non-random.) Then (4.4) – (4.6) imply the following facts:

(a) *vectors.* $y = cx$, $x, y\colon 1\times n$: $\left|\frac{\partial y}{\partial x}\right| = |c|^n$. [combination rule]

(b) *matrices.* $Y = cX$, $X, Y : p\times n$: $\left|\frac{\partial Y}{\partial X}\right| = |c|^{pn}$. [comb. rule]

(c) *symmetric matrices.* $Y = cX$, $X, Y : p\times p$, symmetric: $\left|\frac{\partial Y}{\partial X}\right| = |c|^{\frac{p(p+1)}{2}}$. [comb. rule]

(d) *matrices.* $Y = AX$, $X, Y : p\times n$: $\left|\frac{\partial Y}{\partial X}\right| = |A|^n$. [comb. rule]
$Y = XB$, $X, Y : p\times n$: $\left|\frac{\partial Y}{\partial X}\right| = |B|^p$. [comb. rule]
$Y = AXB$, $X, Y : p\times n$: $\left|\frac{\partial Y}{\partial X}\right| = |A|^n|B|^p$. [chain rule]

(e) *symmetric matrices.* $Y = AXA'$, $X, Y : p\times p$, symmetric:

$$\left|\frac{\partial Y}{\partial X}\right| = |A|^{p+1}.$$

Proof. Use the fact that A can be written as the product of elementary matrices of the forms

$$M_i(c) := \text{Diag}(1, \ldots, 1, c, 1, \ldots, 1),$$
$$E_{ij} := \begin{pmatrix} 1 & & & 0 \\ & \ddots & 1_{ij} & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}.$$

Verify the result when $A = M_i(c)$ and $A = E_{ij}$, then apply the chain rule. □

(f) *triangular matrices:*

- $Y = LX,\ \ X, Y: p \times p$ lower triangular:

$$\left|\frac{\partial Y}{\partial X}\right| = \prod_{i=1}^{p} |l_{ii}|^{i}.$$

Proof. Since $y_{ij} = \sum_{k=j}^{i} l_{ik} x_{kj}$ $(i \geq j)$, the Jacobian matrix is

$$\left(\frac{\partial Y}{\partial X}\right) = \begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{22}}{\partial x_{11}} & \cdots & \frac{\partial y_{pp}}{\partial x_{11}} \\ \frac{\partial y_{11}}{\partial x_{21}} & \frac{\partial y_{21}}{\partial x_{21}} & \frac{\partial y_{22}}{\partial x_{21}} & \cdots & \frac{\partial y_{pp}}{\partial x_{21}} \\ \frac{\partial y_{11}}{\partial x_{22}} & \frac{\partial y_{21}}{\partial x_{22}} & \frac{\partial y_{22}}{\partial x_{22}} & \cdots & \frac{\partial y_{pp}}{\partial x_{22}} \\ \vdots & & & & \vdots \\ \frac{\partial y_{11}}{\partial x_{pp}} & \frac{\partial y_{21}}{\partial x_{pp}} & \frac{\partial y_{22}}{\partial x_{pp}} & \cdots & \frac{\partial y_{pp}}{\partial x_{pp}} \end{pmatrix} = \begin{pmatrix} l_{11} & l_{21} & \cdots & & l_{pp} \\ 0 & l_{22} & & & \\ 0 & 0 & l_{22} & & \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & l_{pp} \end{pmatrix}.$$

This is a $p(p+1)/2 \times p(p+1)/2$ upper triangular matrix whose determinant is $\prod_{i=1}^{p} l_{ii}^{i}$.[5] □

- $Y = UX,\ \ X, Y: p \times p$ upper triangular:

$$\left|\frac{\partial Y}{\partial X}\right| = \prod_{i=1}^{p} |u_{ii}|^{p-i+1}.$$

- $Y = XL,\ \ X, Y: p \times p$ lower triangular:

$$\left|\frac{\partial Y}{\partial X}\right| = \prod_{i=1}^{p} |l_{ii}|^{p-i+1}.$$

Proof. Write $Y' = L'X'$ and apply the preceding case with $U = L'$. □

---

[5] A more revealing proof follows by noting that $Y = LX$ can be written column-by column as $Y_1 = L_1 X_1, , \ldots, Y_p = L_p X_p$, where X_i and Y_i are the $(p-i+1)\times 1$ non-zero parts of the columns of X and Y and where L_i is the lower $(p-i+1)\times(p-i+1)$ principal submatrix of L. Since $Y_i = L_i X_i$ has Jacobian $|L_i|^{+} = \prod_{j=i}^{p} |l_{jj}|$, the result follows from the composition rule.

- $Y = XU,\ \ X, Y : p \times p$ upper triangular:

$$\left|\frac{\partial Y}{\partial X}\right| = \prod_{i=1}^{p} |u_{ii}|^{i}.$$

Proof. Write $Y' = U'X'$ and apply the first case with $L = U'$. □

- $Y = LXM,\ \ X, Y : p \times p$ lower triangular:

$$\left|\frac{\partial Y}{\partial X}\right| = \prod_{i=1}^{p} |l_{ii}|^{i} \cdot \prod_{i=1}^{p} |m_{ii}|^{p-i+1}$$

Proof. Apply the chain rule. □

- $Y = UXV,\ \ X, Y : p \times p$ upper triangular:

$$\left|\frac{\partial Y}{\partial X}\right| = \prod_{i=1}^{p} |u_{ii}|^{p-i+1} \cdot \prod_{i=1}^{p} |v_{ii}|^{i}.$$

Proof. Write $Y' = V'X'U'$ and apply the last case with $L = V'$ and $M = U'$. □

(g) *triangular/symmetric matrices:*

- $Y = X + X'$, $X : p \times p$ lower (or upper) triangular, $Y : p \times p$ symmetric:

$$\left|\frac{\partial Y}{\partial X}\right| = 2^{p}.$$

Proof. Since $y_{ii} = 2x_{ii}, 1 \le i \le p$, while $y_{ij} = x_{ij}, 1 \le j < i \le p$. □

- $Y = L'X + X'L,\ \ X : p \times p$ lower triangular, $\ Y : p \times p$ symmetric:

$$\left|\frac{\partial Y}{\partial X}\right| = 2^{p} \prod_{i=1}^{p} |l_{ii}|^{i}.$$

Proof. Clearly $X \mapsto Y$ is a linear mapping. To show that it is 1-1:

$$\begin{aligned} & L'X_1 + X_1'L = L'X_2 + X_2'L \\ \Longrightarrow\ \ & L'(X_1 - X_2) = -(X_1 - X_2)'L \\ \Longrightarrow\ & (X_1 - X_2)L^{-1} = -[(X_1 - X_2)L^{-1}]'. \end{aligned}$$

Thus $(X_1 - X_2)L^{-1}$ is both lower triangular and skew-symmetric, hence is 0, so $X_1 = X_2$. Next, to find the required Jacobian, apply the chain rule to the sequence of mappings

$$X \mapsto XL^{-1} \mapsto XL^{-1}+(XL^{-1})' \mapsto L'[XL^{-1}+(XL^{-1})']L \equiv L'X+X'L.$$

Therefore the Jacobian is given by [verify!]

$$\left|\frac{\partial Y}{\partial X}\right| = \prod_{i=1}^{p} |l_{ii}^{-1}|^{p-i+1} \cdot 2^p \cdot \prod_{i=1}^{p} |l_{ii}|^{p+1} \; = \; 2^p \prod_{i=1}^{p} |l_{ii}|^{i}.$$

- $Y = U'X + X'U,\;\; X : p \times p$ upper triangular, $Y : p \times p$ symmetric:

$$\left|\frac{\partial Y}{\partial X}\right| = 2^p \prod_{i=1}^{p} |u_{ii}|^{p-i+1}.$$

- $Y = XL' + LX',\;\; X : p \times p$ lower triangular, $Y : p \times p$ symmetric:

$$\left|\frac{\partial Y}{\partial X}\right| = 2^p \prod_{i=1}^{p} |l_{ii}|^{p-i+1}.$$

Proof. Apply the preceding case with $U = L'$ and X replaced by $\tilde{X} := X'$. □

- $Y = XU' + UX',\;\; X : p \times p$ upper triangular, $Y : p \times p$ symmetric:

$$\left|\frac{\partial Y}{\partial X}\right| = 2^p \prod_{i=1}^{p} |u_{ii}|^{i}.$$

Proof. Apply the first case with $L = U'$ and X replaced by $\tilde{X} := X'$. □

4.3. Jacobians of nonlinear mappings.

(*) *The Jacobian of a nonlinear diffeomorphism $x \mapsto y$ is the same as the Jacobian of the linearized differential mapping $dx \mapsto dy$.* Here,

$$dx := (dx_1, \ldots, dx_n) \quad \text{and} \quad dy := (dy_1, \ldots, dy_n).$$

For $n = 1$, (*) is immediate from the linear relation between dx and dy given by the formal differential identity $dy = \left(\frac{dy}{dx}\right)dx$, where $\frac{dy}{dx}$ is treated as a scalar constant c. For $n \geq 2$, the equations for total differentials

$$\begin{aligned} dy_1 &= \Big(\frac{\partial y_1}{\partial x_1}\Big)dx_1 + \cdots + \Big(\frac{\partial y_1}{\partial x_n}\Big)dx_n, \\ &\ \vdots \qquad\quad \vdots \qquad\qquad\qquad \vdots \\ dy_n &= \Big(\frac{\partial y_n}{\partial x_1}\Big)dx_1 + \cdots + \Big(\frac{\partial y_n}{\partial x_1}\Big)dx_n, \end{aligned} \tag{4.7}$$

can be expressed in vector-matrix notation as the single linear relation

$$dy = \Big(\frac{\partial y}{\partial x}\Big)dx \tag{4.8}$$

with $\left(\frac{\partial y}{\partial x}\right)$ treated as a constant matrix, which again implies (*).

The following elementary rules for matrix differentials will combine with (*) to allow calculation of Jacobians for apparently complicated nonlinear diffeomorphisms. Here, if $X \equiv (x_{ij})$ is a matrix variable, dX denotes the matrix of differentials (dx_{ij}). If X is a structured matrix (e.g., symmetric or triangular) then dX has the same structure.

(1) *sum:* $d(X + Y) = dX + dY$. [verify]

(2) *product:* $d(XY) = (dX)Y + X(dY)$. [verify]

(3) *inverse:* $d(X^{-1}) = -X^{-1}(dX)X^{-1}$. [apply (2) with $Y = X^{-1}$]

(4) *determinant:* $d|X| = |X| \operatorname{tr}[X^{-1}(dX)]$. [Use Laplace expansion]

Four examples of nonlinear Jacobians:

(a) *matrix inversion:* if $Y = X^{-1}$ with $X, Y : p \times p$ (unstructured) then

$$\left|\frac{\partial Y}{\partial X}\right| = |X|^{-2p}.$$

Proof. Apply (3) and §4.2(d). □

(b) *matrix inversion:* if $Y = X^{-1}$ with $X, Y : p \times p$ symmetric, then

$$\left|\frac{\partial Y}{\partial X}\right| = |X|^{-p-1}.$$

Proof. Apply (3) and §4.2(e). □

(c) *lower triangular decomposition:* if $S = TT'$ with $S : p \times p$ symmetric pd and $T : p \times p$ lower triangular with $t_{11} > 0, \ldots, t_{pp} > 0$ (Cholesky), then

$$\left|\frac{\partial S}{\partial T}\right| = 2^p \prod_{i=1}^{p} t_{ii}^{p-i+1}.$$

Proof. By (2), $dS = (dT)T' + T(dT)'$; now apply §4.2(g). □

(d) *upper triangular decomposition:* if $S = UU'$ with $S : p \times p$ symmetric pd and $U : p \times p$ upper triangular with $u_{11} > 0, \ldots, u_{pp} > 0$ (Cholesky), then

$$\left|\frac{\partial S}{\partial U}\right| = 2^p \prod_{i=1}^{p} u_{ii}^{i}.$$

Proof. By (2), $dS = (dU)U' + U(dU)'$; again apply §4.2(g). □

4.4. The Wishart density.

We continue the discussion following (4.1). When $\Sigma = I_p$ and $n \geq p$, the pdf $f(T)$ of T (recall that $S = TT'$ with T lower triangular) is given by (4.1). Thus by the inverse rule and §4.3(c) the pdf of S is given by

$$
\begin{aligned}
f(S) =& f(T(S)) \cdot \frac{1}{\left|\frac{\partial S}{\partial T}\right|_{T=T(S)}} \\
=& c'_{p,n} \cdot \prod_{i=1}^{p} t_{ii}^{n-i} \cdot \exp\left(-\tfrac{1}{2}\mathrm{tr}\, TT'\right) \cdot 2^{-p} \prod_{i=1}^{p} t_{ii}^{-p+i-1} \\
=& 2^{-p} c'_{p,n} \cdot \left(\prod_{i=1}^{p} t_{ii}\right)^{n-p-1} \cdot \exp\left(-\tfrac{1}{2}\mathrm{tr}\, TT'\right) \\
=& c_{p,n} \cdot |S|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\mathrm{tr}\, S}, \qquad S \in \mathcal{S}_p^+,
\end{aligned}
\tag{4.9}
$$

where

$$
c_{p,\,n}^{-1} \equiv 2^{\frac{pn}{2}} \cdot \pi^{\frac{p(p-1)}{4}} \prod\nolimits_{i=1}^{p} \Gamma\left(\tfrac{n-i+1}{2}\right) \equiv 2^{\frac{pn}{2}} \cdot \Gamma_p\left(\tfrac{n}{2}\right). \tag{4.10}
$$

Finally, for $\Sigma > 0$ the Jacobian of the mapping $S \mapsto \Sigma^{1/2} S \Sigma^{1/2}$ is $|\Sigma|^{\frac{p+1}{2}}$ (apply §4.2(e)), so the general Wishart pdf for $S \sim W_p(n, \Sigma)$ is given by

$$
\frac{c_{p,\,n}}{|\Sigma|^{\frac{n}{2}}} \cdot |S|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\mathrm{tr}\, \Sigma^{-1} S}, \qquad S \in \mathcal{S}_p^+, \tag{4.11}
$$

a multivariate extension of the density of $\sigma^2 \chi_n^2$. □

Exercise 4.1. Moments of the determinant of a Wishart random matrix. Use (4.11) to show that

$$
\mathrm{E}(|S|^k) = |\Sigma|^k \cdot \frac{2^{pk}\, \Gamma_p\left(\frac{n}{2} + k\right)}{\Gamma_p\left(\frac{n}{2}\right)}, \quad k = 1, 2, \ldots\ . \tag{4.12}
$$

Exercise 4.2. Matrix-variate beta distribution.

Let S and T and be independent with $S \sim W_p(r, \Sigma)$, $T \sim W_p(n, \Sigma)$, $r \geq p$, $n \geq p$, and $\Sigma > 0$, so $S > 0$ and $T > 0$ w. pr. 1. Define

$$\begin{aligned} U &= (S+T)^{-\frac{1}{2}} S \left((S+T)^{-\frac{1}{2}}\right)', \\ V &= S+T. \end{aligned} \tag{4.13}$$

Show that the range of (U, V) is given by $\{0 < U < I\} \times \{V > 0\}$ and verify that (4.13) is a bijection. Show that the joint pdf of $(U,\, V)$ is given by

$$\begin{aligned} f(U, V) &= \frac{c_{p,\,r} c_{p,\,n}}{c_{p,\,r+n}} |U|^{\frac{r-p-1}{2}} |I-U|^{\frac{n-p-1}{2}} \\ &\quad \cdot \frac{c_{p,\,r+n}}{|\Sigma|^{\frac{r+n}{2}}} |V|^{\frac{r+n-p-1}{2}} e^{-\frac{1}{2} \operatorname{tr} \Sigma^{-1} V} \\ &= f(U) \cdot f(V), \end{aligned} \tag{4.14}$$

so U and V are independent and the distribution of U does not depend on Σ. (Note that the distribution of U is a matrix generalization of the beta distribution.) Therefore

$$\mathrm{E}(|S|^k) = \mathrm{E}(|U|^k |V|^k) = \mathrm{E}(|U|^k)\, \mathrm{E}(|V|^k), \tag{4.15}$$

so the moments of $|U|$ can be expressed in terms of the moments of determinants of the two Wishart matrices S and V via (4.12) as follows:

$$\mathrm{E}(|U|^k) = \frac{\mathrm{E}(|S|^k)}{\mathrm{E}(|V|^k)} = \frac{\Gamma_p\left(\frac{r+n}{2}\right)\, \Gamma_p\left(\frac{r}{2} + k\right)}{\Gamma_p\left(\frac{r}{2}\right)\, \Gamma_p\left(\frac{r+n}{2} + k\right)}. \tag{4.16}$$

Hint: To find the Jacobian of (4.13), apply the chain rule to the sequence of mappings

$$(S,\, T) \mapsto (S,\, V) \mapsto (U,\, V).$$

Use the extended combination rule to find the two intermediate Jacobians.

Exercise 4.3. Distribution of the sample correlation matrix when Σ is diagonal.

Let $S \sim W_p(n, D_\sigma)$ $(n \geq p)$, where $D_\sigma := \operatorname{diag}(\sigma_1, \ldots, \sigma_p) > 0$. Define the sample correlation matrix $R \equiv \{r_{ij}\}$ by

$$r_{ij} = s_{ii}^{-1/2} s_{ij} s_{jj}^{-1/2},$$

where $S \equiv \{s_{ij}\}$. Find the joint pdf of $R, s_{11}, \ldots, s_{pp}$. Show that they are mutually independent and find $f(R)$.

Hint: First determine the range of $(R, s_{11}, \ldots, s_{pp})$. Next, the joint pdf of R, s_{11}, $\ldots$, s_{pp} is given by

$$\begin{aligned} f(R, s_{11}, \ldots, s_{pp}) =& f(S) \cdot \left| \frac{\partial(S)}{\partial(R, s_{11}, \ldots, s_{pp})} \right| \\ =& f(S) \cdot \left| \frac{\partial(s_{12}, \ldots, s_{p-1,p}, s_{11}, \ldots, s_{pp})}{\partial(R, s_{11}, \ldots, s_{pp})} \right| \\ =& \frac{c_{p,n}}{|D_\sigma|^{\frac{n}{2}}} \cdot |S|^{\frac{n-p-1}{2}} e^{-\frac{1}{2} \operatorname{tr} D_\sigma^{-1} S} \cdot \left| \frac{\partial(s_{12}, \ldots, s_{p-1,p})}{\partial R} \right| \\ =& \frac{c_{p,n}}{\prod_{i=1}^p \sigma_i^{\frac{n}{2}}} \cdot |R|^{\frac{n-p-1}{2}} \cdot \prod_{i=1}^p s_{ii}^{\frac{n-p-1}{2}} e^{-\frac{s_{ii}}{2\sigma_i}} \cdot \prod_{i=1}^p s_{ii}^{\frac{p-1}{2}} \\ =& c_{p,n} \cdot |R|^{\frac{n-p-1}{2}} \cdot \prod_{i=1}^p \Big(\frac{1}{\sigma_i}\Big) \Big(\frac{s_{ii}}{\sigma_i}\Big)^{\frac{n}{2}-1} e^{-\frac{s_{ii}}{2\sigma_i}}, \end{aligned}$$

where $f(S)$ is given by (4.11) with $\Sigma = D_\sigma$ and the Jacobian is calculated using the extended combination rule and the relation $s_{ij} = s_{ii}^{1/2} r_{ij} s_{jj}^{1/2}$. This establishes the mutual independence, and will yield the marginal pdf of R. (The mutual independence also can be established by means of Basu's Lemma.) $\square$

Exercise 4.4. Inverse Wishart distribution. Let $S \sim W_p(n, \Sigma)$ with $n \geq p$ and $\Sigma > 0$. Show that the pdf of $W \equiv S^{-1}$ is

$$c_{p,n} \frac{|\Omega|^{\frac{n}{2}}}{|W|^{\frac{n+p+1}{2}}} e^{-\frac{1}{2} \operatorname{tr} \Omega W^{-1}}, \qquad W \in \mathcal{S}_p^+. \tag{4.17}$$

where $\Omega = \Sigma^{-1}$. $\square$

5. Estimating a Covariance Matrix.

Consider the problem of estimating Σ based on a Wishart random matrix $S \sim W_p(n, \Sigma)$ with $\Sigma \in \mathcal{S}_p^+$. Assume that $n \geq p$ so that S is nonsingular[6] w. pr. 1. The loss incurred by an estimate $\hat{\Sigma}$ is measured by a *loss function* $L(\hat{\Sigma}, \Sigma)$ such that $L \geq 0$ and $L = 0$ iff $\hat{\Sigma} = \Sigma$. An estimator $\hat{\Sigma} \equiv \hat{\Sigma}(S)$ is evaluated in terms of its *risk function* $\equiv$ *expected loss:*

$$R(\hat{\Sigma}, \Sigma) = \mathrm{E}_\Sigma\,[L(\hat{\Sigma}, \Sigma)].$$

We shall consider two specific loss functions (each convex in $\hat{\Sigma}$):

$$\begin{aligned} \textit{Quadratic loss}: \qquad & L_1(\hat{\Sigma}, \Sigma) = \mathrm{tr}\big(\hat{\Sigma}\Sigma^{-1} - I\big)^2, \\ \textit{Stein's loss}: \qquad & L_2(\hat{\Sigma}, \Sigma) = \mathrm{tr}\big(\hat{\Sigma}\Sigma^{-1}\big) - \log|\hat{\Sigma}\Sigma^{-1}| - p. \end{aligned}$$

We prefer L_2 over L_1 because L_1 penalizes overestimates more than underestimates, unlike L_2:

$$\begin{aligned} & L_1(\hat{\Sigma}, I) \to p \;\; as \;\; \hat{\Sigma} \to 0, \qquad L_1(\hat{\Sigma}, I) \to \infty \;\; as \;\; \hat{\Sigma} \to \infty; \\ & L_2(\hat{\Sigma}, I) \to \infty \;\; as \;\; \hat{\Sigma} \to 0 \;\; or \;\; \infty. \end{aligned}$$

5.1. Equivariant estimators of Σ.

Let G be a subgroup of $GL \equiv GL(p)$, the *general linear group* of all $p \times p$ nonsingular real matrices. Each $A \in G$ acts on $\mathcal{S}_p^+$ according to the mapping

$$\begin{aligned} \mathcal{S}_p^+ &\to \mathcal{S}_p^+ \\ \Sigma &\mapsto A\Sigma A'. \end{aligned} \tag{5.1}$$

A loss function L is *G-invariant* if

$$L(A\hat{\Sigma}A', A\Sigma A') = L(\hat{\Sigma}, \Sigma) \qquad \forall\, A \in G. \tag{5.2}$$

[6] If $n < p$ it would seem impossible to estimate Σ. However several proposals recently been put forth to address this case, which occurs for example with microarray data where $p \approx 10^5$ but $n \approx 10^3$. [References?]

Note that both L_1 and L_2 are *fully invariant*, i.e., are GL-invariant. If L is G-invariant then the risk function of any estimator $\hat{\Sigma} \equiv \hat{\Sigma}(S)$ transforms as follows: for $A \in G$,

$$
\begin{aligned}
R(\hat{\Sigma}(S),\, A\Sigma A') &= \mathrm{E}_{A\Sigma A'}\left[L(\hat{\Sigma}(S),\, A\Sigma A')\right] \\
&= \mathrm{E}_{\Sigma}\left[L(\hat{\Sigma}(ASA'),\, A\Sigma A')\right] \\
&= \mathrm{E}_{\Sigma}\left[L\left(A^{-1}\hat{\Sigma}(ASA')A^{-1\prime},\, \Sigma\right)\right] \\
&= R\left(A^{-1}\hat{\Sigma}(ASA')A^{-1\prime},\, \Sigma\right).
\end{aligned}
\tag{5.3}
$$

An estimator $\hat{\Sigma} \equiv \hat{\Sigma}(S)$ is G-*equivariant* if

$$
\hat{\Sigma}(ASA') = A\,\hat{\Sigma}(S)\,A' \qquad \forall\, A \in G,\ \forall\, S \in \mathcal{S}_p^+. \tag{5.4}
$$

If L is G-invariant and $\hat{\Sigma}$ is G-equivariant then by (5.3) the risk function is also G-invariant:

$$
R(\hat{\Sigma},\, \Sigma) = R(\hat{\Sigma},\, A\Sigma A') \qquad \forall\ A \in G, \tag{5.5}
$$

that is, $R(\hat{\Sigma},\, \Sigma)$ is constant on G-*orbits* of $\mathcal{S}_p^+$ (see Definition 6.1).

We say that G acts *transitively* on $\mathcal{S}_p^+$ if $\mathcal{S}_p^+$ has only one G-orbit under the action of G. Note that G acts transitively on $\mathcal{S}_p^+$ iff every $\Sigma \in \mathcal{S}_p^+$ has a square root $\Sigma_G \in G$, i.e., $\Sigma = \Sigma_G \Sigma_G'$. Thus both GL and $GT \equiv GT(p)$ (the subgroup of all $p \times p$ nonsingular lower triangular matrices) act transitively on $\mathcal{S}_p^+$.[7] If L is G-invariant, $\hat{\Sigma}$ is G-equivariant, and G acts transitively on $\mathcal{S}_p^+$, then the risk function is constant on $\mathcal{S}_p^+$:

$$
R(\hat{\Sigma},\, \Sigma) = R(\hat{\Sigma},\, I) \quad \forall\, \Sigma \in \mathcal{S}_p^+ \qquad [\text{set } A = \Sigma_G^{-1} \text{ in (5.5)}] \tag{5.6}
$$

5.2. The best fully equivariant estimator of Σ.

Lemma 5.1. *An estimator* $\hat{\Sigma}(S)$ *is* GL*-equivariant iff* $\hat{\Sigma}(S) = \delta\, S$ *for some scalar* $\delta > 0$.

[7] For the latter, apply the Cholesky decomposition, Exercise 1.5.

Proof. Set $G = GL$ and $A = S_{GL}^{-1}$ in (5.4) to obtain

$$\hat{\Sigma}(I) = S_{GL}^{-1}\,\hat{\Sigma}(S)\,S_{GL}^{-1\,\prime},$$

so

$$\hat{\Sigma}(S) = S_{GL}\,\hat{\Sigma}(I)\,S_{GL}'. \tag{5.7}$$

Next set $A = \Gamma \in \mathcal{O}_p$ and $S = I$ in (5.4) to obtain

$$\hat{\Sigma}(I) = \Gamma\,\hat{\Sigma}(I)\,\Gamma' \qquad \forall\,\Gamma \in \mathcal{O}_p, \tag{5.8}$$

where $\mathcal{O}_p$ is the subgroup of all $p \times p$ orthogonal matrices. By Exercise 3.19, (5.8) implies that $\hat{\Sigma}(I) = \delta\,I$, so $\hat{\Sigma}(S) = \delta\,S$ by (5.7), as stated. □

We now find the optimal fully equivariant estimators $\hat{\Sigma}(S) \equiv \hat{\delta}\,S$ w. r. to the loss function L_1 and L_2, respectively.

Proposition 5.2. *(a) The best fully equivariant estimator w. r. to the loss function L_1 is the biased estimator $\frac{1}{n+p+1}S$.*
(b) The best fully equivariant estimator w. r. to the loss function L_2 is the unbiased estimator $\frac{1}{n}S$.

Proof. (a) Let $S = \{s_{ij} \mid i, j = 1\ldots, p\}$. Because GL acts transitively on $\mathcal{S}_p^+$ and L_1 is GL-invariant, $\delta\,S$ has constant risk given by

$$\begin{aligned}
\mathrm{E}_I[L_1(\delta\,S,\,I)] &= \mathrm{E}_I[\mathrm{tr}(\delta\,S - I)^2] \\
&= \delta^2\mathrm{E}_I(\mathrm{tr}\,S^2) - 2\delta\mathrm{E}_I(\mathrm{tr}\,S) + \mathrm{tr}\,I^2 \\
&= \delta^2\mathrm{E}_I\Big(\sum\sum\nolimits_{i,j} s_{ij}^2\Big) - 2\delta\,\mathrm{E}\Big(\sum\nolimits_i s_{ii}\Big) + p \\
&= \delta^2\Big[\mathrm{E}_I\Big(\sum\nolimits_i s_{ii}^2\Big) + E_I\Big(\sum\sum\nolimits_{i\neq j} s_{ij}^2\Big)\Big] - 2\delta np + p \\
&\overset{*}{=} \delta^2\Big[(2n+n^2)p + p(p-1)n\Big] - 2\delta np + p \qquad (5.9)\\
&= \{n[\delta^2(n+p+1) - 2\delta] + 1\}p. \qquad (5.10)
\end{aligned}$$

The quadratic function of δ in (5.10) is minimized by $\hat{\delta} = \frac{1}{n+p+1}$.

*To verify (5.9), first note that when $\Sigma = I$, $s_{ii}^2 \sim \chi_n^2$ so

$$\mathrm{E}_I\big(s_{ii}^2\big) = \mathrm{Var}_I\big(\chi_n^2\big) + \big(\mathrm{E}_I(\chi_n^2)\big) = 2n + n^2.$$

Next, $s_{ij} \sim s_{12}$ since

$$\Pi\, S\, \Pi' \sim W_p(n,\, \Pi\, \Pi') = W_p(n,\, I) \sim S$$

for any permutation matrix Π. Also $s_{12}s_{22}^{-1/2} \perp\!\!\!\perp s_{22}$ and $s_{12}s_{22}^{-1/2} \sim N(0,\, 1)$ by (3.50) and (3.51), so

$$\mathrm{E}_I\big(s_{12}^2\big) = \mathrm{E}_I\Big(\frac{s_{12}^2}{s_{22}} \cdot s_{22}\Big) = \mathrm{E}_I\Big(\frac{s_{12}^2}{s_{22}}\Big) \cdot \mathrm{E}_I(s_{22}) = 1 \cdot n = n.$$

(b) Because GL acts transitively on $\mathcal{S}_p^+$ and L_2 is GL-invariant, δS has constant risk given by

$$\begin{aligned} \mathrm{E}_I[L_2(\delta S,\, I)] &= \mathrm{E}_I[\mathrm{tr}(\delta S) - \log|\delta S| - p] \\ &= \delta\, \mathrm{E}_I(\mathrm{tr}\, S) - p \log \delta - \mathrm{E}_I(\log|S|\,) - p \\ &= \delta n p - p \log \delta - \mathrm{E}_I(\log|S|\,) - p. \qquad (5.11) \end{aligned}$$

This is minimized by $\hat{\delta} = \frac{1}{n}$. □

5.3. The best GT-equivariant estimator of Σ.

Lemma 5.3. *Let $S_T = S_{GT}$. An estimator $\hat{\Sigma}(S)$ is GT-equivariant iff*

$$\hat{\Sigma}(S) = S_T\, \Delta\, S_T{}' \qquad (5.12)$$

for a fixed diagonal matrix $\Delta \equiv \mathrm{diag}(\delta_1, \ldots, \delta_p)$ with each $\delta_i > 0$.

Proof. Set $G = GT$ and $A = S_T^{-1}$ in (5.4) to obtain

$$\hat{\Sigma}(I) = S_T^{-1}\, \hat{\Sigma}(S)\, S_T^{-1\prime},$$

so

$$\hat{\Sigma}(S) = S_T\, \hat{\Sigma}(I)\, S_T{}'. \qquad (5.13)$$

Next set $A = D_\pm \equiv \mathrm{diag}(\pm 1, \ldots, \pm 1) \in GT$ and $S = I$ in (5.4) to obtain

$$\hat{\Sigma}(I) = D_\pm\, \hat{\Sigma}(I)\, D_\pm \qquad \forall\, D_\pm. \qquad (5.14)$$

But (5.14) implies that $\hat{\Sigma}(I) = \Delta$ for some diagonal matrix $\Delta \in \mathcal{S}_p^+$, [verify], hence (5.12) follows from (5.13). □

We now present Charles Stein's derivation of the optimal GT-equivariant estimator $\hat{\Sigma}_T(S) := S_T\,\hat{\Delta}_T\,S_T{}'$ w. r. to the loss function L_2. Remarkably, $\hat{\Sigma}_T(S)$ is not of the form $\delta\,S$, hence is not GL-equivariant. Because GT is a proper subgroup of GL, the class of GT-equivariant estimators properly contains the class of GL-equivariant estimators, hence $\hat{\Sigma}_T$ dominates the best fully equivariant estimator $\frac{1}{n}S$. Thus the latter, which is also the best unbiased estimator and the MLE, is neither admissible nor minimax. (Similar results hold for the quadratic loss function L_1.)

Proposition 5.4.[8] *The best GT-equivariant estimator w. r. to the loss function L_2 is*

$$\hat{\Sigma}_T(S) = S_T\,\hat{\Delta}_T\,S_T{}', \tag{5.15}$$

where

$$\hat{\Delta}_T = \mathrm{diag}(\hat{\delta}_{T,1},\ldots,\hat{\delta}_{T,p}) \tag{5.16}$$

and

$$\hat{\delta}_{T,i} = \tfrac{1}{n+p+1-2i}\;. \tag{5.17}$$

The minimum risk function is the constant risk function

$$R_{T,\min}(\Sigma) = \sum\nolimits_{i=1}^{p} \big[\log(n+p+1-2i) - \mathrm{E}(\log\chi^2_{n-i+1})\big]. \tag{5.18}$$

Proof. Let $S_T = \{t_{ij} \mid 1 \le j \le i \le p\}$. Because GT acts transitively on $\mathcal{S}_p^+$ and L_2 is GT-invariant, each GT-equivariant estimator $S_T\,\Delta\,S_T{}'$ has constant risk $R_2(S_T\,\Delta\,S_T{}',\,\Sigma)$ given by

$$\begin{aligned}
&E_I\left[L_2(S_T\Delta S_T{}',\,I)\right]\\
&= \mathrm{E}_I\left[\mathrm{tr}\big(S_T\Delta S_T{}'\big) - \log|S_T\Delta S_T{}'| - p\right]\\
&= \mathrm{E}_I\left[\mathrm{tr}\big(S_T{}'S_T\Delta\big)\right] - \sum\nolimits_{i=1}^{p}\log\delta_i - \mathrm{E}_I\left[\log\big|S_TS_T{}'\big|\right] - p\\
&= \mathrm{E}_I\left[\sum\nolimits_{i=1}^{p}(t_{ii}^2 + t_{(i+1)i}^2 + \cdots t_{pi}^2)\delta_i\right] - \sum\nolimits_{i=1}^{p}\log\delta_i - d_{p,n}\\
&\overset{(3.64)}{=} \sum\nolimits_{i=1}^{p}\left[((n-i+1)+(p-i))\,\delta_i - \log\delta_i\right] - d_{p,n}\\
&= \sum\nolimits_{i=1}^{p}\left[(n+p+1-2i)\delta_i - \log\delta_i\right] - d_{p,n},
\end{aligned} \tag{5.19}$$

where

$$d_{p,n} = \mathrm{E}\Big(\sum\nolimits_{i=1}^{p}\log\chi^2_{n-i+1}\Big) + p. \tag{5.20}$$

[8] James and Stein (1962), *Proc. 4th Berkeley Symp. Math. Statist. Prob. V.1.*

The ith term in (5.19) is minimized by $\hat{\delta}_i = \frac{1}{n+p+1-2i}$, thus (5.17)-(5.18).□

For the loss function L_2, the improvement in risk offered by Stein's estimator $\hat{\Sigma}_T(S) = S_T\,\hat{\Delta}_T\,S_T{}'$ compared to the unbiased estimator $\frac{1}{n}S$ is $\approx$ 5-20% for moderate values of p.[9] However, this estimator is itself inadmissible and can be improved upon readily as follows:

Replace the lower triangular group GT with the upper triangular group GU to obtain the alternative version of Stein's estimator given by

$$\hat{\Sigma}_U(S) = S_U\,\hat{\Delta}_U\,S_U{}', \tag{5.21}$$

where $S_U \equiv S_{GU}$ is the unique upper triangular square root of S and $\hat{\Delta}_U = \mathrm{diag}(\hat{\delta}_{U,1},\dots,\hat{\delta}_{U,p})$ with

$$\hat{\delta}_{U,i} = \hat{\delta}_{T,\,p-i+1} = \tfrac{1}{n-p-1+2i}.$$

Because GU also acts transitively on $\mathcal{S}_p^+$, the risk function of $\hat{\Sigma}_U$ is also constant on $\mathcal{S}_p^+$ with the same constant value as the risk function of $\hat{\Sigma}_T$ [why?[10]] Since $L_2(\hat{\Sigma},\Sigma)$ is strictly convex in $\hat{\Sigma}$ [**(4)** p.54], so is $R_2(\hat{\Sigma},\Sigma)$

[9] S. Lin and M. Perlman (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis – VI, P. R. Krishnaiah, ed.,* pp. 411-429.

[10] Use an invariance argument: Let Π denote the $p \times p$ permutation matrix corresponding to the permutation $(1,\dots,p) \to (p,\dots,1)$. Then

$$\tilde{S} := \Pi S \Pi' = \Pi S_U S_U' \Pi' = (\Pi S_U \Pi')\,(\Pi S_U' \Pi')'$$

and $\Pi S_U \Pi'$ is lower triangular, so $\Pi S_U \Pi' = \tilde{S}_T$ by uniqueness. Also $\hat{\Delta}_U = \Pi' \hat{\Delta}_T \Pi$, so from(5.21),

$$\begin{aligned}\hat{\Sigma}_U(S) &= (\Pi'\tilde{S}_T\Pi)(\Pi'\hat{\Delta}_T\Pi)(\Pi'\tilde{S}_T\Pi)' = \Pi'(\tilde{S}_T\hat{\Delta}_T\tilde{S}_T)\Pi \\ &= \Pi'\hat{\Sigma}_T(\tilde{S})\,\Pi = \Pi'\hat{\Sigma}_T(\Pi S\Pi')\,\Pi.\end{aligned}$$

Now apply (5.3) with $A = \Pi$ to obtain

$$R_2\left(\hat{\Sigma}_U(S),\,\Sigma\right) \equiv R_2\left(\Pi'\hat{\Sigma}_T(\Pi S\Pi')\,\Pi,\,\Sigma\right) = R_2\left(\hat{\Sigma}_T(S),\,\Pi\Sigma\Pi'\right),$$

so $\hat{\Sigma}_U$ and $\hat{\Sigma}_T$ must have the same (constant) risk function, as asserted. □

[verify], hence

$$R_2\left(\tfrac{1}{2}\big(\hat{\Sigma}_T + \hat{\Sigma}_U\big),\, \Sigma\right) < \tfrac{1}{2}R_2(\hat{\Sigma}_T,\, \Sigma) + \tfrac{1}{2}R_2(\hat{\Sigma}_U,\, \Sigma) = R_2(\hat{\Sigma}_T,\, \Sigma).$$

Therefore the estimator $\frac{1}{2}\big(\hat{\Sigma}_T + \hat{\Sigma}_U\big)$ strictly dominates $\hat{\Sigma}_T$ (and $\hat{\Sigma}_U$).

The preceding discussion suggests another estimator that strictly dominates $\frac{1}{2}\big(\hat{\Sigma}_T + \hat{\Sigma}_U\big)$, namely

$$\hat{\Sigma}_P(S) := \frac{1}{p!} \sum_{\Pi \in \mathcal{P}(p)} \Pi'\, \hat{\Sigma}_T(\Pi S \Pi')\, \Pi,$$

where $\mathcal{P} \equiv \mathcal{P}(p)$ is the subgroup of all $p \times p$ permutation matrices. Again the strict convexity of L_2 implies that $\hat{\Sigma}_P$ dominates $\hat{\Sigma}_T$, in fact [verify!]

$$R_2(\hat{\Sigma}_P,\, \Sigma) < R_2\left(\tfrac{1}{2}\big(\hat{\Sigma}_T + \hat{\Sigma}_U\big),\, \Sigma\right) < R_2(\hat{\Sigma}_T,\, \Sigma).$$

Remark 5.5. By Proposition 5.4, $\hat{\Sigma}_T$ is minimax w.r.to L_2 among all GT-equivariant estimators. Because GT-is an *amenable* group, the Hunt-Stein Theorem (cf. Lehmann and Casella, *Theory of Point Estimation, 2nd Ed., Theorem 9.2, p.422*) implies that GT is minimax among *all* estimators, hence so are $\hat{\Sigma}_U$, $\hat{\Sigma}_P$, and $\hat{\Sigma}_O$ in (5.23) below. □

5.4. Orthogonally equivariant estimators of Σ.

The estimator $\hat{\Sigma}_P(S)$ is the average over $\mathcal{P}$ of the transformed estimators $\Pi'\, \hat{\Sigma}_T(\Pi S \Pi')\, \Pi$ and is itself permutation-equivariant [verify]:

$$\hat{\Sigma}_P(\Pi S \Pi') = \Pi\, \hat{\Sigma}_P(S)\, \Pi' \qquad \forall\, \Pi \in \mathcal{P}. \tag{5.22}$$

Because $\mathcal{P}$ is a proper subgroup of the orthogonal group $\mathcal{O}$, the preceding discussion suggests the following estimator, obtained by averaging over $\mathcal{O}$:

$$\hat{\Sigma}_O(S) = \int_{\mathcal{O}} \Psi'\, \hat{\Sigma}_T(\Psi S \Psi')\, \Psi\, d\nu(\Psi), \tag{5.23}$$

where ν is the Haar probability measure on $\mathcal{O}$, i.e. the unique (left $\equiv$ right) orthogonally invariant probability measure on $\mathcal{O}$. Since [verify!]

$$\hat{\Sigma}_O(S) = \int_{\mathcal{O}} \Psi'\, \hat{\Sigma}_P(\Psi S \Psi')\, \Psi\, d\nu(\Psi), \tag{5.24}$$

the strict convexity of L_2 implies that $\hat{\Sigma}_O$ in turn dominates $\hat{\Sigma}_P$ [verify!]:

$$R_2(\hat{\Sigma}_O, \Sigma) < R_2(\hat{\Sigma}_P, \Sigma).$$

The estimator $\hat{\Sigma}_O$, studied by Akimichi Takemura,[11] is *orthogonally equivariant:* for any $\Gamma \in \mathcal{O}$,

$$\begin{aligned}
\hat{\Sigma}_O(\Gamma S \Gamma') &= \int_{\mathcal{O}} \Psi' \, \hat{\Sigma}_T \left(\Psi(\Gamma S \Gamma')\Psi' \right) \Psi \, d\nu(\Psi) \\
&= \int_{\mathcal{O}} \Gamma(\Psi\Gamma)' \, \hat{\Sigma}_T \left((\Psi\Gamma) S (\Psi\Gamma)' \right) (\Psi\Gamma)\Gamma' \, d\nu(\Psi) \\
&\overset{*}{=} \Gamma \left(\int_{\mathcal{O}} \Phi' \, \hat{\Sigma}_T \left(\Phi S \Phi' \right) \Phi \, d\nu(\Phi) \right) \Gamma' \\
&= \Gamma \, \hat{\Sigma}_O(S) \, \Gamma', \qquad (5.25)
\end{aligned}$$

where $*$ follows from the substitution $\Psi \to \Phi \equiv \Psi\Gamma$ and the orthogonal invariance of ν: $d\nu(\Psi) = d\nu(\Psi\Gamma) \equiv d\nu(\Phi)$. The estimator $\hat{\Sigma}_O$ offers greater improvement over $\frac{1}{n}S$ than does $\hat{\Sigma}_T(S)$, often a reduction in risk of 20-30%.

Clearly the unbiased estimator $\frac{1}{n}S$ is orthogonally equivariant [verify]. The class of orthogonally equivariant estimators is characterized as follows:

Lemma 5.6. *For any* $S \in \mathcal{S}_p^+$ *let* $S = \Gamma_S \, D_{l(S)} \, \Gamma_S'$ *be its spectral decomposition. Here* $l(S) = (l_1(S), \ldots, l_p(S))$ *where* $l_1 > \cdots > l_p$ (> 0) *are the ordered eigenvalues of* S*, the columns of* Γ_S *are the corresponding eigenvectors, and* $D_{l(S)} = \mathrm{diag}(l_1(S), \ldots, l_p(S))$*. An estimator* $\hat{\Sigma} \equiv \hat{\Sigma}(S)$ *is* $\mathcal{O}$*-equivariant iff*

$$\hat{\Sigma}(S) = \Gamma_S \, D_{\phi(l(S))} \, \Gamma_S' \qquad (5.26)$$

where $D_{\phi(l)} = \mathrm{diag}(\phi_1(l_1, \ldots, l_p), \ldots, \phi_p(l_1, \ldots, l_p))$ *with* $\phi_1 \geq \cdots \geq \phi_p > 0$.

Proof. For any $\Psi \in \mathcal{O}$ and $S \in \mathcal{S}_p^+$,

$$\Psi \, S \, \Psi' = (\Psi\Gamma_S) \, D_{l(S)} \, (\Psi\Gamma_S)',$$

[11] An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.* (1984) **8** 367-376.

hence $\Gamma_{\Psi S\Psi'} = \Psi\Psi_S$ and $l(\Psi S\Psi') = l(S)$. Thus if $\hat\Sigma(S)$ satisfies (5.26) then

$$\hat\Sigma(\Psi S\Psi') = \Gamma_{\Psi S\Psi'}\, D_{\phi(l(\Psi S\Psi'))}\, \Gamma'_{\Psi S\Psi'} = \Psi\Gamma_S D_{\phi(l(S))}\Gamma'_S\Psi = \Psi\,\hat\Sigma(S)\,\Psi',$$

so $\hat\Sigma$ is $\mathcal{O}$-equivariant.

Conversely, if $\hat\Sigma$ is $\mathcal{O}$-equivariant then

$$\hat\Sigma(S) = \Gamma_S\,\hat\Sigma(\Gamma'_S\,S\,\Gamma_S)\,\Gamma'_S = \Gamma_S\,\hat\Sigma(D_{l(S)})\,\Gamma'_S. \tag{5.27}$$

But

$$\hat\Sigma(D_{l(S)}) = \hat\Sigma(D_\pm D_{l(S)} D_\pm) = D_\pm\,\hat\Sigma(D_{l(S)})\,D_\pm \;\forall\, D_\pm \equiv \mathrm{diag}(\pm1,\dots,\pm) \in \mathcal{O},$$

hence (recall (5.14)) $\hat\Sigma(D_{l(S)})$ must be a diagonal matrix whose entries depend on S only through $l(S)$. That is,

$$\hat\Sigma(D_{l(S)}) = D_{\phi(l(S))}$$

for some $\phi(l(S) \equiv (\phi_1(l(S)),\dots,\phi_p(l(S)))$, so (5.27) yields (5.26). □

By (5.5), the risk function $R_2(\hat\Sigma,\,\Sigma)$ of an $\mathcal{O}$-equivariant estimator $\hat\Sigma$ is constant on $\mathcal{O}$-orbits of $\mathcal{S}_p^+$, hence satisfies

$$R_2(\hat\Sigma,\,\Sigma) = R(\hat\Sigma,\,D_{\lambda(\Sigma)}), \tag{5.28}$$

where $\lambda(\Sigma) \equiv (\lambda_1(\Sigma) \geq \dots \geq \lambda_p(\Sigma)\ (> 0))$ is the vector of the ordered eigenvalues of Σ. Thus, by restricting consideration to orthogonally equivariant estimators, the problem of estimating Σ reduces to that of estimating the population eigenvalues $\lambda(\Sigma)$ based on the sample eigenvalues $l(S)$.

The need to shrink/expand the sample eigenvalues. Stein gave three facts to show that the largest (smallest) sample eigenvalues are overestimates (underestimates) of the largest (smallest) population eigenvalues. He sought improved orthogonally equivariant estimators that shrink (expand) the largest (smallest) sample eigenvalues. First, the extremal representations

$$l_1(S) = \max_{x'x=1} x'Sx, \tag{5.29}$$

$$l_p(S) = \min_{x'x=1} x'Sx, \tag{5.30}$$

show that $l_1(S)$ and $l_p(S)$ are, respectively, convex and concave functions of S [verify]. Thus by Jensen's inequality,

$$\text{E}_\Sigma\left[l_1(S)\right] \geq l_1\left[\text{E}(S)\right] = l_1(n\Sigma) \equiv n\,\lambda_1(\Sigma), \tag{5.31}$$

$$\text{E}_\Sigma\left[l_p(S)\right] \leq l_p\left[\text{E}(S)\right] = l_p(n\Sigma) \equiv n\,\lambda_p(\Sigma). \tag{5.32}$$

Thus $\frac{1}{n}l_1$ *tends to overestimate* λ_1 *and should be shrunk, while* $\frac{1}{n}l_p$ *underestimate* λ_p *and should be expanded.* This holds for the other eigenvalues also: $\frac{1}{n}l_2,\ \frac{1}{n}l_3,\ldots$ should be shrunk while $\frac{1}{n}l_{p-1},\ \frac{1}{n}l_{p-2},\ldots$ should be expanded.

Next from (3.53) and the strict concavity of $\log x$,

$$\begin{aligned} \text{E}\left[\prod\nolimits_{i=1}^{p} \tfrac{1}{n}l_i\,(S)\right] = \text{E}[|\tfrac{1}{n}S|] &= |\Sigma| \cdot \prod\nolimits_{i=1}^{p} \left(1 - \tfrac{p-i}{n}\right) \\ &< |\Sigma| \cdot \left(1 - \tfrac{p-1}{2n}\right)^p \\ &< \prod\nolimits_{i=1}^{n} \lambda_i(\Sigma) \cdot e^{-\frac{p(p-1)}{2n}}. \end{aligned} \tag{5.33}$$

Thus $\prod_{i=1}^{p} \frac{1}{n}l_i\,(S)$ will tend to *underestimate* $\prod_{i=1}^{n} \lambda_i(\Sigma)$ unless $n \gg p^2$, which does not usually hold in applications. This suggests that the shrinkage/expansion of the sample eigenvalues should not be done in a linear manner: *the smaller sample eigenvalues should be expanded proportionately more than the larger sample eigenvalues should be shrunk.*

A more precise justification is based on the limit of the empiriccal distribution of the sample eigenvalues as $p \to \infty$, obtained by Marchenko and Pasteur (1967). A strong consequence of their result is that when $\Sigma = \lambda\, I_p$ (equivalently, $\lambda_1(\Sigma) = \cdots = \lambda_p(\Sigma) = \lambda$) and both $n,\, p \to \infty$ while $\frac{p}{n} \to \eta$ for some fixed $\eta \in (0,\, 1]$, then

$$\tfrac{1}{n}l_1\,(S) \overset{a.s.}{\to} \lambda\,(1 + \sqrt{\eta})^2, \tag{5.34}$$

$$\tfrac{1}{n}l_p\,(S) \overset{a.s.}{\to} \lambda\,(1 - \sqrt{\eta})^2. \tag{5.35}$$

Thus if it were known that $\Sigma = \lambda\, I_p$ then $\frac{1}{n}l_1\,(S)$ should be shrunk by the factor $1/(1 + \sqrt{\eta})^2$ while $\frac{1}{n}l_p\,(S)$ should be expanded by the factor $1/(1 - \sqrt{\eta})^2$. Furthermore, *the expansion is proportionately greater than the shrinkage since*

$$\tfrac{1}{(1+\sqrt{\eta})^2} \cdot \tfrac{1}{(1-\sqrt{\eta})^2} = \tfrac{1}{(1-\eta)^2} > 1.$$

The shrinkage and expansion factors in (5.34) and (5.35) are derived only for the case $\Sigma = \lambda I_p$ (the "worst case" in that the most shrinkage/expansion is required). In general the appropriate shrinkage/expansion factors (equivalently, the functions $\phi_1, \ldots, \phi_p$ in (5.26)) depend on $\lambda_1(\Sigma), \ldots,$ $\lambda_p(\Sigma)$, so must themselves be estimated adaptively.

Exercise 5.7. *(Takemura).* When $p = 2$, show that $\hat{\Sigma}_O(S)$ has the form (5.26) with

$$\phi_1(l_1, l_2) = \left(\frac{\sqrt{l_1}\, \hat{\delta}_{T,1}}{\sqrt{l_1} + \sqrt{l_2}} + \frac{\sqrt{l_2}\, \hat{\delta}_{T,2}}{\sqrt{l_1} + \sqrt{l_2}} \right) l_1, \tag{5.36}$$

$$\phi_2(l_1, l_2) = \left(\frac{\sqrt{l_2}\, \hat{\delta}_{T,1}}{\sqrt{l_1} + \sqrt{l_2}} + \frac{\sqrt{l_1}\, \hat{\delta}_{T,2}}{\sqrt{l_1} + \sqrt{l_2}} \right) l_2.$$

where $\hat{\delta}_{T,1} = \frac{1}{n+1}$ and $\hat{\delta}_{T,2} = \frac{1}{n-1}$ (set $p = 2$ in (5.17)). Verify that $\phi_1 > (=)\ \phi_2$ when $l_1 > (=)\ l_2$, $\phi_2 > \frac{l_2}{n}$, while $\phi_1 < \frac{l_1}{n}$ when $\sqrt{\frac{l_1}{l_2}} > \frac{n+1}{n-1}$, so $\hat{\Sigma}_O$ "expands" the smallest eigenvalue of $\frac{1}{n}S$ and usually "shrinks" its largest eigenvalue when $p = 2$. (Takemura showed that this holds for $\hat{\Sigma}_O$ for all $p \geq 2$.) $\square$

Stein's unbiased estimate of the risk of an orthogonally equivariant estimator: Based on the following expression for the risk of an orthogonally invariant estimator, Stein argued that the shrinkage/expansion should be stronger than that given by $\hat{\Sigma}_O$. For $\hat{\Sigma}(S) = \Gamma_S D_{\phi(l(S))} \Gamma'_S$ (see (5.26)), if $\phi_i \equiv \phi_i(l)$ is smooth then

$$\begin{aligned} &\mathrm{E}_\Sigma[L_2(\hat{\Sigma}, \Sigma)] + d_{p,n} \\ &= \mathrm{E}_\Sigma \left\{ \sum_{i=1}^p \left[(n-p-1)\frac{\phi_i}{l_i} + 2\phi_i \sum_{j \neq i} \frac{1}{l_i - l_j} + 2\frac{\partial \phi_i}{\partial l_i} - \log\left(\frac{\phi_i}{l_i}\right] \right) \right\} \\ &= \mathrm{E}_\Sigma \left\{ \sum_{i=1}^p \left[\left(n - p + 1 + 2l_i \sum_{j \neq i} \frac{1}{l_i - l_j} \right) \psi_i - \log \psi_i + 2l_i \frac{\partial \psi_i}{\partial l_i} \right] \right\}, \end{aligned} \tag{5.37}$$

where $\psi_i = \phi_i / l_i$ and $d_{p,n}$ is given by (5.20). Stein's proof[12] is essentially based on multivariate integration by parts; also see L. R. Haff (c. 1980-91).

[12] I first learned of this result at Stein's 1975 IMS Rietz Lecture in Atlanta, which

The term inside the bracket $\{\cdots\}$ is called an *unbiased estimate of the risk*, and does not depend on the unknown parameter Σ. Thus in principle, the optimal $\phi_i(l)$'s or $\psi_i(l)$'s can be found by the calculus of variations.

First consider othogonally invariant estimators (5.26) having the simple form $\phi_i = c_i\, l_i$ (i.e., $\psi_i = c_i$), $i = 1, \ldots, p$, for constants $0 < c_1 \leq \cdots \leq c_p$. From (5.37) the risk function under L_2 can be written as

$$\begin{aligned} &\mathrm{E}_\Sigma\Big\{\Big[\big(n-p+1\big)\sum\nolimits_i c_i + 2\sum\nolimits_i\sum\nolimits_{j>i}\Big(\frac{c_i l_i - c_j l_j}{l_i - l_j}\Big) - \log c_i\Big]\Big\} - d_{p,n} \\ &\quad \leq \big(n-p+1\big)\sum\nolimits_i c_i + 2\sum\nolimits_i\sum\nolimits_{j>i} c_i - \log c_i - d_{p,n} \\ &\quad = \sum\nolimits_i \big[\big(n+p+1-2i\big)c_i - \log c_i\big] - d_{p,n}, \end{aligned}$$

identical to (5.19). Thus this upper bound is minimized at (recall (5.17))

$$\tilde{c}_i = \hat{\delta}_{T,i} = \tfrac{1}{n+p+1-2i}, \quad i = 1, \ldots, p, \tag{5.38}$$

and the minimum value of the upper bound is (cf. (5.18))

$$\sum\nolimits_i \log(n+p+1-2i) - d_{p,n} \equiv R_{T,\min}(\Sigma),$$

the constant-risk function of the minimax estimator $\hat{\Sigma}_T$, cf. Remark 5.5. The resulting estimator $\hat{\Sigma}_{\tilde{c}} \equiv \Gamma_S D_{\tilde{c}\, l(S)} \Gamma_S'$ is therefore also minimax. (If the adjusted eigenvalue estimates $\tilde{c}_1 l_1, \ldots, \tilde{c}_p l_p$ are not monotone decreasing, isotonization should be used.)

However, the shrinkage factors for $l_1(S)$ and $l_p(S)$ suggested by (5.34)-(5.35) are more extreme than those from $\tilde{c}_1 \equiv \frac{1}{n+p-1}$ and $\tilde{c}_p \equiv \frac{1}{n-p+1}$:

$$1 > n\tilde{c}_1 \equiv \tfrac{n}{n+p-1} \approx \tfrac{1}{1+\eta} > \tfrac{1}{(1+\sqrt{\eta})^2}, \tag{5.39}$$

$$1 < n\tilde{c}_p \equiv \tfrac{n}{n-p+1} \approx \tfrac{1}{1-\eta} < \tfrac{1}{(1-\sqrt{\eta})^2}. \tag{5.40}$$

Stein obtained the following adaptive eigenvalue estimators by ignoring the partial derivatives $\partial\psi_i/\partial l_i$ in (5.37) and minimizing over the ψ_i:

$$\phi_i^*(l_1, \ldots, l_p) = \left(\frac{1}{n-p+1+2l_i\sum_{j\neq i}\frac{1}{l_i - l_j}}\right)^+ l_i. \tag{5.41}$$

---

remains unpublished in English; he published his results in a Russian journal in 1977. He also wrote lecture notes for his multivariate courses at Stanford and U. Washington.

The term inside the parentheses can be negative hence its positive part is taken. Also the required ordering $\phi_1^* \geq \ldots \geq \phi_p^*$ need not hold, in which case the ordering is achieved by an isotonization algorithm – see Lin and Perlman (1985) for details. Despite these complications, Stein's estimator offers substantial improvement over the other estimators considered thus far – the reduction in risk can be 70-80% when $\Sigma \approx \lambda I_p$.

If the population eigenvalues are widely dispersed, i.e.,

$$\lambda_1(\Sigma) \gg \cdots \gg \lambda_p(\Sigma), \tag{5.42}$$

then the sample eigenvalues $\{l_i\}$ will also be widely dispersed, so

$$l_i \sum_{j \neq i} \frac{1}{l_i - l_j} = \sum_{j>i} \frac{l_i}{l_i - l_j} + \sum_{j<i} \frac{l_i}{l_i - l_j} \approx (p-i) + 0,$$

in which case (5.41) reduces to [verify]

$$\phi_i^*(l_1, \ldots, l_p) = \left(\frac{1}{n+p-1+2i}\right) l_i \equiv \tilde{c}_i l_i \tag{5.43}$$

(recall (5.38)). On the other hand, if two or more $\lambda_i(\Sigma)$'s are nearly equal then the same approximately will be true for the corresponding l_i's, in which case the shrinkage/expansion offered by the ϕ_i^*'s will be more pronounced than in (5.43), a desirable feature as indicated by (5.39) and (5.40).

* * *

When $p \geq 3$, it is difficult to evaluate the integral for Takemura's estimator $\hat{\Sigma}_O(S)$ in (5.23). However, the integral can be approximated by Monte Carlo simulation from the Haar probability distribution over $\mathcal{O}$. This can be accomplished as follows:

Lemma 5.8. Let $Z \sim N_{p \times p}(0, I_p \otimes I_p)$. The distribution of the random orthogonal matrix $\Gamma \equiv (ZZ')^{-1/2} Z$ is the Haar measure on $\mathcal{O}$, i.e., the unique left $\equiv$ right orthogonally invariant probability measure on the compact topological group $\mathcal{O}$.

Proof. It suffices to show that the distribution is right orthogonally invariant, i.e., that $\Gamma \sim \Gamma\,\Psi$ for all $\Psi \in \mathcal{O}$. But $Z \sim Z\Psi$, hence

$$\Gamma\,\Psi = [(Z\Psi)(Z\Psi)']^{-1/2} Z\,\Psi \sim (ZZ')^{-1/2} Z = \Gamma. \qquad \square$$

6. Invariant Tests of Hypotheses. (See Lehmann TSH Ch. 6, 8.)

Motivation for invariant tests (and equivariant estimators):

(a) Respect the symmetries of a statistical problem.

(b) Unbiasedness fails to yield a UMPU test when testing more than one parameter. Restricting to invariant tests *sometimes* leads to a UMPI test, but at least reduces the class of tests to be compared.

6.1. Invariant statistical models and maximal invariant statistics.

A *statistical model* is a family $\mathcal{P}$ of probability distributions defined on a sample space $(\mathcal{X}, \mathcal{A})$, where $\mathcal{A}$ is the sigma-field of measurable subsets of $\mathcal{X}$. Often $\mathcal{P}$ has a parametric representation: $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$. (The parameterization is assumed to be identifiable.)

Let G be a group of measurable mappings of $\mathcal{X}$ into itself. Then G *acts on* $\mathcal{X}$ if

(1) $(g_1 g_2)x = g_1(g_2 x)\ \forall g_1, g_2 \in G, \forall x \in \mathcal{X}$.

(2) $1_G x = x\ \forall x \in \mathcal{X}$. ($1_G$ denotes the identity element in G.)

Here (1) and (2) imply that the mapping $g : \mathcal{X} \to g\mathcal{X}$ is a bijection $\forall\, g \in G$.

Definition 6.1. Suppose that G acts on $\mathcal{X}$. For $x \in \mathcal{X}$, the G-*orbit* of x is the subset $Gx := \{gx \mid g \in G\} \subseteq \mathcal{X}$, i.e., the set of all images of x under the actions in G. The *orbit space*

$$\mathcal{X}/G := \{Gx \mid x \in \mathcal{X}\}$$

is the set of all G-orbits. The *orbit projection* π is the mapping

$$\begin{aligned} \pi : \mathcal{X} &\to \mathcal{X}/\mathcal{G} \\ x &\mapsto Gx. \end{aligned}$$

Trivially, π is a G-*invariant* function, that is, π is constant on G-orbits:

$$\pi(x) = \pi(gx)\ \forall x, g.$$

[Since G itself is invariant under group multiplication: $\{g'g \mid g' \in G\} = G$.]

Definition 6.2. A function $t : \mathcal{X} \to \mathcal{T}$ is a *maximal invariant statistic* (MIS) if it is equivalent to the orbit projection π, i.e., if t is constant on G-orbits and distinguishes G-orbits (takes different values on different orbits.)

Lemma 6.3. Suppose that $t : \mathcal{X} \to \mathcal{T}$ satisfies

(3) t is G-invariant;

(4) if $u : \mathcal{X} \to \mathcal{U}$ is G-invariant, i.e., satisfies $u(x) = u(gx)\ \forall x, g$, then u depends on x only through the value of $t(x)$, i.e., $u(x) = w(t(x))$ for some function $w : \mathcal{T} \to \mathcal{U}$.

Then t is a maximal invariant statistic.

Proof. We need only show that t distinguishes G-orbits. This follows from (4) with $u = \pi$. □

If G acts on $\mathcal{X}$ and P is a probability measure on $\mathcal{X}$, let $gP \equiv P \circ g^{-1}$:

$$(gP)(A) := P(g^{-1}(A))\ \forall A \in \mathcal{A}.$$

Equivalently, if $X \sim P$ then $gX \sim gP$.

Definition 6.4. The statistical model $\mathcal{P}$ is *G-invariant* if $g\mathcal{P} \subseteq \mathcal{P}\ \forall g \in G$.

If $\mathcal{P}$ is G-invariant then by (1) and (2),

(5) $(g_1g_2)P = g_1(g_2P)\ \forall g_1, g_2 \in G, \forall P \in \mathcal{P}$. [since $(g_1g_2)^{-1} = g_2^{-1}g_1^{-1}$]

(6) $1_G P = P\ \forall P \in \mathcal{P}$.

Then (5) implies that

$$\mathcal{P} = g(g^{-1}\mathcal{P}) \subseteq g\mathcal{P}\ \ \forall g \in G,$$

so $g\mathcal{P} = \mathcal{P}\ \forall g$ and the mapping $g : \mathcal{P} \to g\mathcal{P}$ is a bijection for each $g \in G$. Furthermore, if $\mathcal{P}$ has a parametric representation $\{P_\theta \mid \theta \in \Theta\}$ then, equivalently, G acts on Θ according to

$$P_{g\theta} := gP_\theta \equiv P_\theta \circ g^{-1}.$$

Also equivalently, if $X \sim P_\theta$ then $gX \sim P_{g\theta}$. In this case, (5) and (6) become

(7) $(g_1 g_2)\theta = g_1(g_2\theta)\ \forall g_1, g_2 \in G, \forall \theta \in \Theta$.

(8) $1_G\theta = \theta\ \forall \theta \in \Theta$. (Thus, $G\Theta = \Theta$.)

Again, (7) and (8) imply that $g\Theta = \Theta\ \forall g$ and the mapping $g : \Theta \to g\Theta$ is a bijection for each $g \in G$. Note that if $dP_\theta(x) = f_\theta(x)dx$ then the G-invariance of $\mathcal{P}$ is equivalent to

(9) $f_\theta(x) = f_{g\theta}(gx)\left|\frac{\partial(gx)}{\partial x}\right|$ [verify].

Definition 6.5. Assume that $\mathcal{P} \equiv \{P_\theta \mid \theta \in \Theta\}$ is G-invariant. For $\theta \in \Theta$, the G-*orbit* of θ is the subset $G\theta := \{g\theta \mid g \in G\} \subseteq \Theta$. A function $\tau : \Theta \to \Xi$ is a *maximal invariant parameter* (MIP) if it is constant on G-orbits and distinguishes G-orbits. $\bar{\square}$

As in Lemma 6.3, τ is a maximal invariant parameter iff τ is G-invariant and any G-invariant parameter $\sigma(\theta)$ depends on θ only through the value of $\tau \equiv \tau(\theta)$.

Lemma 6.6. Assume that $u : \mathcal{X} \to \mathcal{U}$ is G-invariant. Then the distribution of u depends on θ only through the value of the maximal invariant parameter τ. (In particular, the distribution of a maximal invariant statistic t depends only on τ.)

Proof. We need only show that the distribution of u is G-invariant. But this is immediate, since for any measurable subset $B \subseteq \mathcal{U}$,

$$P_{g\theta}[\,u(X) \in B] = P_\theta[\,u(gX) \in B] = P_\theta[\,u(X) \in B].$$

6.2. Invariant hypothesis testing problems.

Suppose that $\mathcal{P} \equiv \{P_\theta \mid \theta \in \Theta\}$ is G-invariant and we wish to test

$$H_0 : \theta \in \Theta_0 \qquad \text{vs.} \qquad H : \theta \in \Theta \setminus \Theta_0 \tag{6.1}$$

based on X, where Θ_0 is a proper subset of Θ such that $\mathcal{P}_0 \equiv \{P_\theta \mid \theta \in \Theta_0\}$ is also G-invariant. Then (6.1) is called a G-*invariant testing problem.* A sensible approach to such a testing problem is to respect the symmetry of the problem (i.e., its G-invariance) and restrict attention to test statistics that are G-invariant. Equivalently, this leads us to consider the "invariance-reduced" problem where we test H_0 vs. H based on the value of a MIS

$t \equiv t(X)$ rather than on the value of X itself. In general this may entail a loss of information, but optimal invariant tests often (but not always) remain admissible among all possible tests.

Because $\mathcal{P}_0$ and $\mathcal{P}$ are G-invariant, the invariance-reduced testing problem can be restated equivalently as that of testing

$$H_0 : \tau \in \Xi_0 \qquad \text{vs.} \qquad H : \tau \in \Xi \setminus \Xi_0 \tag{6.2}$$

based on a MIS t, for appropriate sets Ξ_0 and Ξ in the range of the MIP τ. Our goal will be to determine the distribution of the MIS t and apply the principles of hypothesis testing to (6.2). In particular, if a UMP test exists for (6.2), it is called *UMP invariant (UMPI) with respect to* G for (6.1).

In cases where the class of invariant tests is still so large that no UMPI test exists, the likelihood ratio test (LRT) for (6.1), which rejects H_0 for *large* values of the LRT statistic

$$\Lambda(x) := \frac{\max_\Theta f_\theta(x)}{\max_{\Theta_0} f_\theta(x)},$$

is often a satisfactory G-invariant test. By Wilks' Theorem,

$$2 \log \Lambda \xrightarrow{d} \chi^2_{d-d_0} \quad d := \dim(\Theta),\ d_0 := \dim(\Theta_0), \tag{6.3}$$

under H_0 as the underlying i.i.d. sample size $\to \infty$ (see Examples).

Lemma 6.7. The LRT statistic is G-invariant:

$$\Lambda(gx) = \Lambda(x) \;\; \forall\, g \in G.$$

Proof. Apply property (9) in §6.1. $\square$

Example 6.8. Testing a mean vector with known covariance matrix: one observation.

Consider the problem of testing

$$\mu = 0 \quad \text{vs.} \quad \mu \neq 0 \quad \text{based on} \quad X \sim N_p(\mu, I_p). \tag{6.3}$$

Here $\mathcal{X} = \Theta = \mathcal{R}^p$ and $\Theta_0 = \{0\}$. Let $G = \mathcal{O} \equiv$ the group of all $p \times p$ orthogonal matrices g acting on $\mathcal{X}$ and Θ via

$$x \mapsto gx \quad \text{and} \quad \mu \mapsto g\mu,$$

respectively. Because

$$gX \sim N_p(g\mu,\, gg' \equiv I),$$

Θ and Θ_0 are G-invariant. For $x,\ \mu \in \mathcal{X} \equiv \mathcal{R}^p$, the G-orbits of x and μ are the spheres

$$\{x' \in \mathcal{R}^p \mid \|x'\| = \|x\|\} \quad \text{and} \quad \{\mu' \in \mathcal{R}^p \mid \|\mu'\| = \|\mu\|\},$$

respectively, so

$$t \equiv t(X) = \|X\|^2 \quad \text{and} \quad \tau \equiv \tau(\mu) = \|\mu\|^2$$

represent the MIS and MIP, resp. The distribution of t is $\chi^2_p(\tau)$, the noncentral chi-square distribution with noncentrality parameter τ. Any G-invariant statistic depends on X only through $\|X\|^2$, and its distribution depends on μ only through $\|\mu\|^2$. The invariance-reduced problem (6.2) becomes that of testing

$$\tau = 0 \quad \text{vs.} \quad \tau > 0 \quad \text{based on} \quad \|X\|^2 \sim \chi^2_p(\tau).$$

Since $\chi^2_p(\tau)$ has monotone likelihood ratio (MLR) in τ (see Appendix A on MLR, Example A.14), by the Neyman-Pearson (NP) Lemma the uniformly most powerful (UMP) level α test for this problem rejects $\|\mu\|^2 = 0$ if

$$\|X\|^2 > \chi^2_{p;\,\alpha},$$

the upper α quantile of the χ^2_p distribution, and is unbiased. Thus this test is UMPI level α for (6.3) and is unbiased for (6.3). $\square$

Exercise 6.9. (a) In Example 6.8 show that the UMP invariant level α test is the level α LRT based on X for (6.3).

(b) The power function of this LRT is given by

$$\beta_p(\tau) := \Pr{}_\tau[\,\|X\|^2 > \chi^2_{p;\,\alpha}] \equiv \Pr[\,\chi^2_p(\tau) > \chi^2_{p;\,\alpha}].$$

It follows from Remark A.13 and the MLR property (or the log concavity of the normal pdf) that $\beta_p(\tau)$ is increasing in τ, hence this test is unbiased. Show that for fixed τ, $\beta_p(\tau)$ is *decreasing* in p. *Hint:* apply the NP Lemma.

(c) (Kiefer and Schwartz (1965) *Ann. Math. Statist.*) Show that the LRT is a proper Bayes test for (6.3), and therefore is admissible among *all* tests for (6.3).

Hint: consider the following prior distribution:

$$\begin{aligned}
&\Pr[\,\mu = 0] = \gamma,\\
&\Pr[\,\mu \neq 0] = 1-\gamma,\\
&\mu \,\big|\, \mu \neq 0 \;\sim\; N_p(0,\,\lambda I), \quad (0<\gamma<1,\ \lambda>0).
\end{aligned}$$

Example 6.10. Testing a mean vector with unknown covariance matrix: one observation.

Consider the problem of testing

$$\mu = 0 \quad \text{vs.} \quad \mu \neq 0 \quad \text{based on} \quad X \sim N_p(\mu,\,\Sigma)$$

with $\Sigma > 0$ unknown. Here

$$\mathcal{X} = \mathcal{R}^p, \qquad \Theta = \mathcal{R}^p \times \mathcal{S}_p^+, \qquad \Theta_0 = \{0\} \times \mathcal{S}_p^+.$$

Now we may take $G = GL(p)$, the group of all $p \times p$ nonsingular matrices g, acting on $\mathcal{X}$ and Θ via

$$x \mapsto gx \quad \text{and} \quad (\mu, \Sigma) \mapsto (g\mu,\, g\Sigma g')$$

respectively. Again Θ and Θ_0 are G-invariant. Now there are only two G-orbits in $\mathcal{X}$: $\{0\}$ and $\mathcal{R}^p \setminus \{0\}$ [verify], so any G-invariant statistic is constant on $\mathcal{R}^p \setminus \{0\}$, hence its distribution does not depend on μ. Thus no G-invariant test can distinguish between the hypotheses $\mu = 0$ and $\mu \neq 0$ on the basis of a single observation X when Σ is unknown. □

Example 6.11. Testing a mean vector with unknown covariance matrix: repeated observations.
Consider the problem of testing

$$(6.5) \qquad \mu = 0 \;\text{ vs. }\; \mu \neq 0 \quad \text{based on} \;\; (X, T) \sim N_p(\mu, \Sigma) \otimes W_p(n, \Sigma)$$

with $\Sigma > 0$ unknown and $n \geq p$. Here

$$\mathcal{X} = \Theta = \mathcal{R}^p \times \mathcal{S}_p^+, \qquad \Theta_0 = \{0\} \times \mathcal{S}_p^+.$$

Let $G = GL$ act on $\mathcal{X}$ and Θ via

$$(x, t) \mapsto (gx, \, gtg') \quad \text{and} \quad (\mu, \Sigma) \mapsto (g\mu, g\Sigma g'),$$

respectively. Because

$$(gX, gTg') \sim N_p(g\mu, \, g\Sigma g') \otimes W_p(n, \, g\Sigma g'),$$

Θ and Θ_0 are G-invariant. It follows from Lemma 6.3 that

$$T^2 \equiv T^2(X, T) := X'T^{-1}X \quad \text{and} \quad \tau \equiv \tau(\mu, \Sigma) := \mu'\Sigma^{-1}\mu$$

represent the MIS and MIP, respectively [verify!]. We have seen that

$$\text{Hotelling's } T^2 \equiv X'T^{-1}X \sim \frac{\chi_p^2(\tau)}{\chi_{n-p+1}^2},$$

the ratio of two independent chisquare variates, the first noncentral. (This ratio has the (nonnormalized) noncentral F distribution $F_{p,\,n-p+1}(\tau)$.) The invariance-reduced problem (6.2) becomes that of testing

$$(6.6) \qquad \tau = 0 \;\text{ vs. }\; \tau > 0 \quad \text{based on} \quad T^2 \sim F_{p,\,n-p+1}(\tau).$$

Because $F_{p,\,n-p+1}(\tau)$ has MLR in τ (see Example A.15), the UMP level α test for (6.6) rejects $\tau = 0$ if $T^2 > F_{p,\,n-p+1;\,\alpha}$ and is unbiased. Thus this test is UMPI level α for (6.5), and is unbiased for (6.5). □

Exercise 6.12. (a) In Example 6.11, show that the UMP invariant level α test ($\equiv$ the T^2 test) is the level α LRT based on (X, T) for (6.5).

(b) The power function of this LRT is given by

$$\beta_{p,\,n-p+1}(\tau) := \Pr{}_\tau[\,T^2 > F_{p,\,n-p+1;\alpha}] \equiv \Pr[\,F_{p,\,n-p+1}(\tau) > F_{p,\,n-p+1;\alpha}].$$

It follows from MLR that $\beta_{p,\,n-p+1}(\tau)$ is increasing in τ, hence this test is unbiased. Show that for fixed τ and p, $\beta_{p,\,n-p+1}(\tau)$ is *increasing in* n.

(c)* (Kiefer and Schwartz (1965) *Ann. Math. Statist.*). Show that the LRT is a proper Bayes test for testing (6.5) based on (X, T), and thus is admissible among *all* tests for (6.5).

Hint: consider the prior probability distribution on $\Theta_0 \cup \Theta$ given by

$$\begin{aligned}
\Pr[\Theta_0] &= \gamma,\\
\Pr[\Theta] &= 1-\gamma,\ \ (0<\gamma<1);\\
(\mu,\Sigma)\ \big|\ \Theta_0 &\sim \pi_0,\\
(\mu,\Sigma)\ \big|\ \Theta &\sim \pi,
\end{aligned}$$

where π_0 and π are measures on $\Theta_0 \equiv \{0\} \times \mathcal{S}_p^+$ and $\Theta \equiv \mathcal{R}^p \times \mathcal{S}_p^+$ respectively, defined as follows: π_0 assigns all its mass to points of the form

$$(\mu,\Sigma) = (0,\,(\eta\eta' + I)^{-1}), \quad \eta \in \mathcal{R}^p,$$

where η has pdf proportional to $|\eta\eta' + I|^{-(n+1)/2}$; π assigns all its mass to points of the form

$$(\mu,\Sigma) = ((\eta\eta' + I)^{-1}\eta,\,(\eta\eta' + I)^{-1}) \quad \eta \in \mathcal{R}^p,$$

where η has pdf proportional to

$$|\eta\eta' + I|^{-(n+1)/2} \exp\left(\tfrac{1}{2}\eta'(\eta\eta' + I)^{-1}\eta\right).$$

Verify that π_0 and π are *proper* measures, i.e., verify that the corresponding pdfs of η have finite total mass. Show that the T^2 test is the Bayes test for this prior distribution. $\bar{\square}$

Note: An entirely different method for showing the admissibility of the T^2 test among *all* tests for (6.5) was given by Stein (*Ann. Math. Statist.* 1956), based on the exponential structure of the distribution of (X, S).

Example 6.13. Testing a mean vector with covariates and unknown covariance matrix.

Similar to Example 6.11, but with the following changes. Partition X, T, μ, and Σ as

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

respectively, where X_i and μ_i are $p_i \times 1$, T_{ij} and Σ_{ij} are $p_i \times p_j$, $i, j = 1, 2$, with $p_1 + p_2 = p$. Suppose it is known that $\mu_2 = 0$, that is, the second group of p_2 variables are covariates. Consider the problem of testing

$$(6.7) \qquad \begin{array}{c} \mu_1 = 0 \;\text{ vs. }\; \mu_1 \neq 0 \\ \text{based on } \;(X, T) \;\sim\; N_p(\mu, \Sigma) \otimes W_p(n, \Sigma) \end{array}$$

with $\Sigma > 0$ unknown and $n \geq p$. Again $\mathcal{X} = \mathcal{R}^p \times \mathcal{S}_p^+$, but now

$$\Theta = \mathcal{R}^{p_1} \times \mathcal{S}_p^+, \qquad \Theta_0 = \{0\} \times \mathcal{S}_p^+.$$

Let G_1 be the group of all nonsingular block-triangular $p \times p$ matrices g:

$$g = \begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix},$$

so G_1 is a *subgroup* of the invariance group GL in Example 6.11. Here $G_1 \equiv \{g\}$ acts on $\mathcal{X}$ and Θ via the actions

$$(x, t) \mapsto (gx, gtg') \quad \text{and} \quad (\mu_1, \Sigma) \mapsto (g_{11}\mu_1, g\Sigma g'),$$

respectively. Then Θ and Θ_0 are G_1-invariant [verify]. $\square$

Exercise 6.14. (a) In Example 6.13, apply Lemma 6.3 and (1.30) to show that

$$\begin{aligned} (L, M) &\equiv (L(X,T),\, M(X,T)) \\ &:= \left(\tfrac{(X_1 - T_{12}T_{22}^{-1}X_2)' T_{11\cdot 2}^{-1} (X_1 - T_{12}T_{22}^{-1}X_2)}{1 + X_2' T_{22}^{-1} X_2},\; X_2' T_{22}^{-1} X_2 \right) \end{aligned}$$

is a (two-dimensional!) MIS, while

$$\tau_1 \equiv \tau_1(\mu_1, \Sigma) := \mu_1' \Sigma_{11\cdot 2}^{-1} \mu_1$$

is a (one-dimensional!) MIP. Thus the invariance-reduced problem (6.2) becomes that of testing

$$\tau_1 = 0 \ \text{ vs. } \ \tau_1 > 0 \quad \text{based on} \quad (L, M). \tag{6.8}$$

(b) Show that the joint distribution of $(L, M) \equiv (L(X, T),\ M(X, T))$ can be described as follows:

$$\begin{aligned} L \mid M &\sim \frac{\chi^2_{p_1}\big(\frac{\tau_1}{1+M}\big)}{\chi^2_{n-p+1}} \equiv F_{p_1,\, n-p+1}\big(\tfrac{\tau_1}{1+M}\big), \\ M &\sim \frac{\chi^2_{p_2}}{\chi^2_{n-p_2+1}} \equiv F_{p_2,\, n-p_2+1}. \end{aligned} \tag{6.9}$$

Hint: Begin by finding the conditional distribution of $X_1 - T_{12}T_{22}^{-1}X_2$ given (X_2, T_{22}).

(c) Show that the level α LRT based on (X, T) for (6.7) is the test that rejects $(\mu_1, \mu_2) = (0, 0)$ if

$$L > F_{p_1,\, n-p+1;\, \alpha}.$$

This test is the *conditionally* UMP level α test for (6.8) given the ancillary statistic M and is conditionally unbiased for (6.8), therefore unconditionally unbiased for (6.7) $\equiv$ (6.8).

(d)** Show that no UMP size α test exists for (6.8), so no UMPI test exists for (6.7). *Therefore the LRT is not UMPI!* (See Remark 6.16).

(e)* In Exercise 6.12b, show $\beta_{p,m}(\tau)$ is *decreasing* in p for fixed τ and m.

Hint: Apply the results (6.9) concerning the joint distribution of (L, M) $\bar{\square}$

Remark 6.15. Since $T^2 \equiv X'T^{-1}X = L(1 + M) + M$, the overall T^2 test in Example 6.11 is also G_1-invariant in Example 6.13, so it is of interest to compare its power function to that of the LRT in Example 6.13. Given M, the conditional power function of the LRT is given by

$$\Pr_\tau\big[\,F_{p_1,\,n-p+1}\big(\tfrac{\tau_1}{1+M}\big) > F_{p_1,\,n-p+1;\alpha}\;\big|\;M\big] \equiv \beta_{p_1,\,n-p+1}\big(\tfrac{\tau_1}{1+M}\big),$$

while the (unconditional) power of the size-α T^2 test is $\beta_{p,\,n-p+1}(\tau_1)$ because $\tau = \tau_1$ when $\mu_2 = 0$. Since $\beta_{p,m}(\delta)$ is *decreasing* in p but increasing in δ (recall Exercises 6.12b, 6.14e), neither power function dominates the other.

Another possible test in Example 6.13 rejects $(\mu_1, \mu_2) = (0, 0)$ iff

$$T_1^2 := X_1' T_{11}^{-1} X_1 > F_{p_1,\,n-p_1+1,\,\alpha},$$

a test that ignores the covariates and is *not* G_1-invariant [verify]. Since

$$T_1^2 \sim F_{p_1,\,n-p_1+1}(\tilde{\tau}_1)$$

where $\tilde{\tau}_1 := \mu_1' \Sigma_{11}^{-1} \mu_1$, the power function of the level α test based on T_1^2 is $\beta_{p_1,\,n-p_1+1}(\tilde{\tau}_1)$. Because $\tilde{\tau}_1 \le \tau_1$ but $\beta_{p,m}(\delta)$ is decreasing in p and increasing in m, the power function of T_1^2 neither dominates nor is dominated by that of the LRT or of T^2. □

Remark 6.16. Despite their apparent similarity, the invariant testing problems (6.6) and (6.8) are fundamentally different, due to the fact that in (6.8) the dimensionality of the MIS (L, M) exceeds that of the MIP τ_1. Marden and Perlman (1980) (*Ann. Statist.*) show that in Example 6.13, no UMP invariant test exists, and the level α LRT is *inadmissible* for large (but therefore uninteresting) α values, e.g, $\alpha \succeq 0.25$. However, for small (typical) α values it is admissible among invariant tests. Furthermore, it is the UMP conditional test given M, it is G_1-invariant, its power function compares well numerically to those of T^2, T_1^2, and other competing tests, and it is easy to apply. Thus its use is recommended. □

Exercise 6.17. Let (X, T) be as in Examples 6.11 and 6.13. Consider the problem of testing $\mu_1 = 0$ vs. $\mu_1 \neq 0$ with μ_2 and Σ unknown. Find a natural invariance group G_2 such that the test that rejects $\mu_1 = 0$ if

$$T_1^2 := X_1' T_{11}^{-1} X_1 > F_{p_1,\,n-p_1+1;\,\alpha}$$

is UMP among all G_2-invariant level α tests. $\bar{\square}$

Example 6.18. Testing a covariance matrix.
Consider the problem of testing

$$(6.10) \qquad \Sigma = I \quad \text{vs.} \quad \Sigma \neq I \quad \text{based on } S \sim W_p(r, \Sigma) \qquad (r \geq p).$$

Here $\mathcal{X} = \Theta = \mathcal{S}_p^+$ and $\Theta_0 = \{I_p\}$. This problem is invariant under the actions of $G \equiv \mathcal{O}_p$ on $\mathcal{S}_p^+$ given by $S \mapsto gSg'$ and $\Sigma \mapsto g\Sigma g'$. It follows from Lemma 6.3 and the spectral decompositions of $S, \Sigma \in \mathcal{S}_p^+$ that the MIS and MIP are represented by, respectively,

$$l(S) \equiv (l_1(S) \geq \cdots \geq l_p(S)) \quad := \text{the set of (ordered) eigenvalues of } S,$$
$$\lambda(\Sigma) \equiv (\lambda_1(\Sigma) \geq \cdots \geq \lambda_p(\Sigma)) := \text{the set of (ordered) eigenvalues of } \Sigma;$$

[verify]. By Lemma 6.6, the distribution of $l(S)$ depends on Σ only through $\lambda(\Sigma)$; this distribution is complicated when Σ is not of the form κI_p for some $\kappa > 0$. The invariance-reduced problem is that of testing

$$(6.11) \qquad \lambda(\Sigma) = (1, \ldots, 1) \quad \text{vs.} \quad \lambda(\Sigma) \neq (1, \ldots, 1) \quad \text{based on } l(S).$$

Here, unlike Examples 6.8, 6.11, and 6.13, when $p \geq 2$ the alternative hypothesis remains multi-dimensional even after reduction by invariance, so it is not to be expected that a UMPI test exists (it does not). $\bar{\square}$

Exercise 6.19. In Example 6.18 derive the LRT for (6.10). Express the test statistic in terms of $l(S)$.

Answer: The LRT rejects $\Sigma = I$ for large values of $\frac{e^{\operatorname{tr} S}}{|S|}$, or equivalently, for large values of

$$\sum\nolimits_{i=1}^{p} [l_i(S/r) - \log l_i(S/r) - 1]. \qquad \bar{\square}$$

Note. Recall the Courant-Fischer Minmax Theorem: if $\Sigma = \operatorname{Cov}(X)$ then

$$\lambda_k = \min_{L_k} \max_{a \in L_k,\, \|a\|=1} \operatorname{Var}(a'X),$$
$$\lambda_{n-k+1} = \max_{L_k} \min_{a \in L_k,\, \|a\|=1} \operatorname{Var}(a'X).$$

The normalized linear combination $a'X$ such that $\lambda_k = \mathrm{Var}(a'X)$ is the *k-th principal component* of X. $\bar{\square}$

Exercise 6.20. Testing sphericity. Change (6.10) as follows: test

$$\Sigma = \kappa I,\ 0 < \kappa < \infty \quad \text{vs.} \quad \Sigma \neq \kappa I \quad \text{based on } S \sim W_p(r, \Sigma). \tag{6.12}$$

Show that this problem remains invariant under the extended group

$$\bar{G} := \{\bar{g} = ag \mid a > 0,\ g \in \mathcal{O}_p\}.$$

Express a MIS and MIP for this problem in terms of $l(S)$ and $\lambda(\Sigma)$ respectively. Find the LRT for this problem and express it in terms of $l(S)$.
(The hypothesis $\Sigma = \kappa I, 0 < \sigma < \infty$ is called the hypothesis of *sphericity*.)

Answer: The LRT rejects the sphericity hypothesis for large values of $\frac{\frac{1}{p}\mathrm{tr} S}{|S|^{\frac{1}{p}}}$, or equivalently, for large values of

$$\frac{\frac{1}{p}\sum_{i=1}^{p} l_i(S)}{\prod_{i=1}^{p} l_i(S)^{\frac{1}{p}}},$$

the ratio of the arithmetic and geometric means of $l_1(S), \ldots, l_p(S)$. $\bar{\square}$

Exercise 6.21. If the identity matrix I is replaced by any fixed matrix $\Sigma_0 \in \mathcal{S}_p^+$, show that the results in Exercises 6.19 and 6.20 can be applied after the linear transformations $S \mapsto \Sigma_0^{-1/2} S \Sigma_0^{-1/2'}$ and $\Sigma \mapsto \Sigma_0^{-1/2} \Sigma \Sigma_0^{-1/2'}$.

Example 6.22. Testing independence of two sets of variates.
Let $S \sim W_p(n, \Sigma)$ with $n \geq p$. Partition S and Σ as

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

respectively. Here, S_{ij} and Σ_{ij} are $p_i \times p_j$ matrices, $i, j = 1, 2$, where $p_1 + p_2 = p$. Let $\Theta = \mathcal{S}_p^+$ as before, but now take

$$\Theta_0 = \{\Sigma \in \mathcal{S}_p^+ \mid \Sigma_{12} = 0\},$$

so (6.1) becomes the problem of testing

$$(6.13) \qquad \Sigma_{12} = 0 \quad \text{vs.} \quad \Sigma_{12} \neq 0 \quad \text{based on } S \sim W_p(n, \Sigma) \qquad (n \geq p).$$

If G is the group of non-singular block-diagonal $p \times p$ matrices of the form

$$g = \begin{pmatrix} g_{11} & 0 \\ 0 & g_{22} \end{pmatrix}$$

(so $G = GL(p_1) \times GL(p_2)$), then (6.13) is invariant under the action of G on $\mathcal{S}_p^+$ given by $s \mapsto gsg'$ [verify]. Let $q = \min\{p_1, p_2\}$. It follows from Lemma 6.3 and the singular value decomposition that a MIS is [verify!]

$$r^2(S) \equiv (r_1^2(S) \geq \cdots \geq r_q^2(S)) := \text{the nonzero eigenvalues of } S_{12}S_{22}^{-1}S_{21}S_{11}^{-1}.$$

The $r_i(S)$ are the *canonical correlation coefficients of* S. A MIP is [verify!]

$$\rho^2(\Sigma) \equiv (\rho_1^2(\Sigma) \geq \cdots \geq \rho_q^2(\Sigma)) := \text{the nonzero eigenvalues of } \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}.$$

The $\rho_i(\Sigma)$ are the *canonical correlation coefficients of* Σ (see Exercise 6.25).

The distribution of $r(S)$ depends on Σ only through $\rho(\Sigma)$; it is complicated when $\Sigma_{12} \neq 0$. The invariance-reduced problem is that of testing

$$(6.14) \qquad \rho(\Sigma) = (0, \ldots, 0) \quad \text{vs.} \quad \rho(\Sigma) \geq (0, \ldots, 0) \quad \text{based on } r(S).$$

When $p \geq 2$ the alternative hypothesis remains multi-dimensional even after reduction by invariance, so a UMPI test for (6.13) does not exist. □

Remark 6.23. This model and testing problem can be reduced to the multivariate linear model and MANOVA testing problem (see Remark 8.5) by conditioning on S_{22}:

$$(6.15) \qquad X := S_{12}S_{22}^{-1/2'} \mid S_{22} \sim N_{p_1 \times p_2}\big(\beta S_{22}^{1/2}, \, \Sigma_{11\cdot 2} \otimes I_{p_2}\big),$$

where $\beta = \Sigma_{12}\Sigma_{22}^{-1}$. Since $\Sigma_{12} = 0$ iff $\beta = 0$, the conditional testing problem (6.13) is equivalent to that of testing $\beta = 0$ vs. $\beta \neq 0$ based on $(X, \, S_{11\cdot 2})$, a MANOVA testing problem under the conditional distribution of X. □

Exercise 6.24. In Example 6.22 find the LRT for (6.13). Express the test statistic in terms of $r(S)$. Show this LRT statistic is equivalent to the conditional LRT statistic for testing $\beta = 0$ vs. $\beta \neq 0$ based on the conditional distribution of $(S_{12}S_{22}^{-1/2'},\, S_{11\cdot 2})$ given S_{22} (see Exercise 6.37a). Show that when $\Sigma_{12} = 0$, the conditional and unconditional distributions of the LRT statistic are identical. [This distribution can be expressed in terms of Wilks' distribution $U(p_1,\, p_2,\, n - p_2)$ – see Exercises 6.37c, d, e.]

Partial answer: The (unconditional and conditional) LRT rejects $\Sigma_{12} = 0$ for *large* values of

$$\frac{|S_{11}||S_{22}|}{|S|},$$

or equivalently, for *small* values of

$$\prod_{i=1}^{q}(1 - r_i^2(S)). \tag{6.16}$$

Exercise 6.25. Suppose that $\Sigma = \operatorname{Cov}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$. Show that

$$\rho_1(\Sigma) = \max_{a_1 \neq 0,\, a_2 \neq 0} \operatorname{Corr}(a_1' X_1,\, a_2' X_2) \equiv \max_{a_1 \neq 0,\, a_2 \neq 0} \frac{a_1' \Sigma_{12} a_2}{\sqrt{a_1' \Sigma_{11} a_1}\sqrt{a_2' \Sigma_{22} a_2}}.$$

Hint: Apply the Cauchy-Schwartz inequality. □

Example 6.26. Testing a multiple correlation coefficient.
In Example 6.22 set $p_1 = 1$, so $p_2 = p - 1$ and $q \equiv \min(1,\, p - 1) = 1$. Now the MIS $r_1(S) \geq 0$ and the MIP $\equiv \rho_1(\Sigma) \geq 0$ are one-dimensional and can be expressed explicitly as follows (recall Example 3.21):

$$r_1^2(S) = \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}} =: R^2, \qquad \rho_1^2(\Sigma) = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\sigma_{11}} =: \rho^2.$$

The invariance-reduced problem (6.14) becomes that of testing

$$\rho^2 = 0 \quad \text{vs.} \quad \rho^2 > 0 \qquad \text{based on } R^2. \tag{6.17}$$

By normality, the hypotheses

$$\Sigma_{12} = 0, \quad \rho^2 = 0, \quad \text{and} \quad X_1 \perp\!\!\!\perp X_2 \tag{6.18}$$

are mutually equivalent. By (6.16) and (3.74) the size α LRT for testing $\Sigma_{12} = 0$ vs. $\Sigma_{12} \neq 0$ rejects $\Sigma_{12} = 0$ if $R^2 > B\left(\frac{p}{2}, \frac{n-p+1}{2}; \alpha\right)$.

Note: ρ and R are called the *population (resp., sample) multiple correlation coefficients* for the following reason: if

$$\begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \mathrm{Cov}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

then

$$\rho = \max_{a_2 \neq 0} \mathrm{Corr}(X_1, a_2' X_2) = \max_{a_2 \neq 0} \frac{\Sigma_{12} a_2}{\sqrt{\sigma_{11}}\sqrt{a_2' \Sigma_{22} a_2}},$$

with equality attained at $\hat{a}_2 = \Sigma_{22}^{-1}\Sigma_{21}$. [Verify; apply the Cauchy-Schwartz inequality – this implies that $\Sigma_{12}\Sigma_{22}^{-1}X_2$ is the *best linear predictor* of X_1 based on X_2 (when $\mathrm{E}X = 0$).] $\bar{\square}$

Exercise 6.27. Show that the R^2-test is UMPI and unbiased.

Solution: In Example A.18 of Appendix A it is shown that the pdf of R^2 has MLR in ρ^2. Thus this R^2-test is the UMP size α test for the invariance-reduced problem (6.17), hence is the UMPI size α test for $\Sigma_{12} = 0$ vs. $\Sigma_{12} \neq 0$, and is unbiased. $\bar{\square}$

Remark 6.28. (Kiefer and Schwartz (1965) *Ann. Math. Statist.*) By an argument similar to that in Exercise 6.12c*, the R^2-test is a proper Bayes test for testing $\Sigma_{12} = 0$ vs. $\Sigma_{12} \neq 0$ based on S, and thus is admissible among *all* tests for this problem. $\bar{\square}$

Remark 6.29. When $\Sigma_{12} = 0$, $R^2 \equiv \frac{U}{1+U} \sim B\left(\frac{p-1}{2}, \frac{n-p+1}{2}\right)$ (see (3.74)), so

$$\mathrm{E}(R^2) = \frac{p-1}{n} > 0 = \rho^2.$$

Thus, under the null hypothesis of independence, R^2 is an *overestimate* of $\rho_1^2(\Sigma) \equiv 0$ (unless $n \gg p$), hence might naively suggest dependence of X_1 on X_2. $\bar{\square}$

Example 6.30. Testing independence of $k \geq 3$ sets of variates. In the framework of Example 6.22, partition S and Σ as

$$S = \begin{pmatrix} S_{11} & \cdots & S_{1k} \\ \vdots & & \vdots \\ S_{k1} & \cdots & S_{kk} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} \\ \vdots & & \vdots \\ \Sigma_{k1} & \cdots & \Sigma_{kk} \end{pmatrix},$$

respectively, where $k \geq 3$. Again S_{ij} and Σ_{ij} are $p_i \times p_j$ matrices, $i, j = 1, \ldots, k$, where $p_1 + \cdots + p_k = p$. Take

$$\Theta_0 = \{\Sigma \mid \Sigma_{ij} = 0,\, i \neq j\},$$

so (6.1) becomes the problem of testing

$$(6.19) \quad \Sigma_{ij} = 0,\, i \neq j \text{ vs. } \Sigma_{ij} \neq 0 \text{ for some } i \neq j \quad \text{based on } S \sim W_p(n, \Sigma)$$

with $n \geq p$. If G is the set of all non-singular block-diagonal $p \times p$ matrices

$$g \equiv \begin{pmatrix} g_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & g_{kk} \end{pmatrix},$$

so $G = GL(p_1) \times \cdots \times GL(p_k)$ and $\mathcal{P}$, then (6.19) is G-invariant. Now no explicit representation of the MIS and MIP is known (probably none exists). Again the alternative hypothesis remains multi-dimensional even after reduction by invariance, so a UMPI test does not exist. $\bar{\Box}$

Exercise 6.31. In Example 6.30, derive the LRT for (6.19).

Answer: The LRT rejects $\Sigma_{ij} = 0$, $i \neq j$, for large values of $\frac{\prod_{i=1}^{k} |S_{ii}|}{|S|}$.

Note: This LRT is proper Bayes and admissible among all tests for (6.19) (Kiefer and Schwartz (1965) *Ann. Math. Statist.*) and is unbiased. $\bar{\Box}$

Example 6.32. Testing equality of two covariance matrices. Consider the problem of testing

$$(6.20)\qquad \begin{array}{c}\Sigma_1 = \Sigma_2 \text{ vs. } \Sigma_1 \neq \Sigma_2 \\ \text{based on } \ (S_1, S_2) \sim W_p(n_1, \Sigma_1) \times W_p(n_2, \Sigma_2)\end{array}$$

with n_1, $n_2 \geq p$. Here

$$\mathcal{X} = \Theta = \mathcal{S}_p^+ \times \mathcal{S}_p^+, \qquad \Theta_0 = \mathcal{S}_p^+.$$

This problem is invariant under the action of GL on $\mathcal{S}_p^+ \times \mathcal{S}_p^+$ given by

$$(6.21)\qquad (s_1, s_2) \mapsto (gs_1g',\, gs_2g')$$

It follows from Lemma 6.3 and the simultaneous diagonalizibility of two positive definite matrices that the MIS and MIP are represented by

$$\begin{aligned} f(S_1, S_2) &\equiv (f_1(S_1, S_2) \geq \cdots \geq f_p(S_1, S_2)) \;\; := \text{the eigenvalues of } S_1S_2^{-1}, \\ \phi(\Sigma_1, \Sigma_2) &\equiv (\phi_1(\Sigma_1, \Sigma_2) \geq \cdots \geq \phi_p(\Sigma_1, \Sigma_2)) := \text{the eigenvalues of } \Sigma_1\Sigma_2^{-1}, \end{aligned}$$

respectively [verify!]. By Lemma 6.6 the distribution of $f(S_1, S_2)$ depends on (Σ_1, Σ_2) only through $\phi(\Sigma_1, \Sigma_2)$; this distribution is complicated when $\Sigma_1 \neq \kappa\Sigma_2$. The invariance-reduced problem becomes that of testing

$$(6.22)\quad \phi(\Sigma_1, \Sigma_2) = (1, \ldots, 1) \text{ vs. } \phi(\Sigma_1, \Sigma_2) \neq (1, \ldots, 1) \;\text{ based on } f(S_1, S_2).$$

When $p \geq 2$ the alternative hypothesis remains multi-dimensional even after reduction by invariance, so a UMPI test for (6.20) does not exist. □

Exercise 6.33. In Example 6.32, derive the LRT for (6.20) and express the test statistic in terms of $f(S_1, S_2)$. Show that the LRT statistic is minimized when $\frac{1}{n_1}S_1 = \frac{1}{n_2}S_2$.

Answer: The LRT rejects $\Sigma_1 = \Sigma_2$ for large values of

$$\frac{|S_1 + S_2|^{n_1+n_2}}{|S_1|^{n_1}|S_2|^{n_2}},$$

or equivalently, for large values of

$$\prod_{i=1}^{p} (1 + f_i^{-1})^{n_1} (1 + f_i)^{n_2}, \qquad (f_i \equiv f_i(S_1, S_2)).$$

The ith term in the product is minimized when $f_i = n_1/n_2$. ◻

Example 6.34. Testing equality of $k \geq 3$ covariance matrices. Consider the problem of testing

$$\text{(6.23)} \qquad \begin{gathered} \Sigma_1 = \cdots = \Sigma_k \quad \text{vs.} \quad \Sigma_i \neq \Sigma_j \quad \text{for some } i \neq j \\ \text{based on } (S_1, \ldots, S_k) \sim W_p(n_1, \Sigma_1) \times \cdots \times W_p(n_k, \Sigma_k). \end{gathered}$$

with $n_1 \geq p, \ldots, n_k \geq p$. Here

$$\mathcal{X} = \Theta = \mathcal{S}_p^+ \times \cdots \times \mathcal{S}_p^+ \ (k \text{ times}), \qquad \Theta_0 = \mathcal{S}_p^+.$$

This problem is invariant under the action of GL on $\mathcal{S}_p^+ \times \cdots \times \mathcal{S}_p^+$ given by

$$(s_1, \ldots, s_k) \mapsto (g s_1 g', \ldots, g s_k g').$$

As in Example 6.30, no explicit representation of the MIS and MIP are known (probably none exists). The alternative hypothesis is multidimensional after reduction by invariance; no UMPI test for (6.23) exists. ◻

Exercise 6.35. In Example 6.34, derive the LRT for (6.23). Show that the LRT statistic is minimized when $\frac{1}{n_1} S_1 = \cdots = \frac{1}{n_k} S_k$.

Answer: The LRT rejects $\Sigma_1 = \cdots = \Sigma_k$ for large values of

$$\frac{\left| \sum_{i=1}^{k} S_i \right|^{\sum n_i}}{\prod_{i=1}^{k} |S_i|^{n_i}}.$$

To minimize this, apply the case $k = 2$ repeatedly.

Note: This LRT, also called *Bartlett's test,* is unbiased when $k \geq 2$. (Perlman (1980) *Ann. Statist.*) ◻

Example 6.36. The canonical MANOVA testing problem.
Consider the problem of testing

$$\begin{aligned} &\mu = 0 \quad \text{vs.} \quad \mu \neq 0 \quad (\Sigma \text{ unknown}) \\ &\text{based on} \quad (X,\, T) \sim N_{p\times r}(\mu, \Sigma \otimes I_r) \otimes W_p(n, \Sigma) \end{aligned} \tag{6.24}$$

with $\Sigma > 0$ unknown and $n \geq p$. (Example 6.11 is the special case where $r = 1$.) Here

$$\mathcal{X} = \Theta = \mathcal{R}^{p\times r} \times \mathcal{S}_p^+, \qquad \Theta_0 = \{0\} \times \mathcal{S}_p^+.$$

This problem is invariant under the action of the group $GL \times \mathcal{O}_r \equiv \{(g, \gamma)\}$ acting on $\mathcal{X}$ and Θ via

$$\begin{aligned} (x, t) &\mapsto (gx\gamma',\, gtg'), \\ (\mu, \Sigma) &\mapsto (g\mu\gamma',\, g\Sigma g'), \end{aligned} \tag{6.25}$$

respectively. It follows from Lemma 6.3 and the singular value decomposition that a MIS is [verify!]

$$\begin{aligned} f(X,T) &\equiv (f_1(X,T) \geq \cdots \geq f_q(X,T)) \\ &:= \text{the nontrivial eigenvalues of } X'T^{-1}X, \end{aligned} \tag{6.26}$$

where $q := \min(p, r)$ (equivalently, the nontrivial eigenvalues of $XX'T^{-1}$), and a MIP is [verify!]

$$\begin{aligned} \phi(\mu,\Sigma) &\equiv (\phi_1(\mu,\Sigma) \geq \cdots \geq \phi_q(\mu,\Sigma)) \\ &:= \text{the nontrivial eigenvalues of } \mu'\Sigma^{-1}\mu, \end{aligned} \tag{6.27}$$

(equivalently, the nontrivial eigenvalues of $\mu\mu'\Sigma^{-1}$). The distribution of $f(X,T)$ depends on (μ, Σ) only through $\phi(\mu, \Sigma)$; it is complicated when $\mu \neq 0$. The invariance-reduced problem (6.2) becomes that of testing

$$\phi(\mu,\Sigma) = (0,\ldots,0) \quad \text{vs.} \quad \phi(\mu,\Sigma) \geq (0,\ldots,0) \quad \text{based on } f(X,T). \tag{6.28}$$

Here the MIS and MIP have the same dimension, namely q, and a UMP invariant test will not exist when $q \equiv \min(p, r) \geq 2$.

Note that $f(X,T)$ reduces to the T^2 statistic when $r = 1$, so in the general case the distribution of $f(X,T)$ is a generalization of the (central and noncentral) F distribution. The distribution of $(f_1(X,T),\ldots,f_q(X,T))$ when $\mu = 0$ is given in Exercise 7.2.

(The reduction of the general MANOVA testing problem to this canonical form will be presented in §8.2.) $\bar{\square}$

Exercise 6.37a. In Example 6.36, derive the LRT for testing $\mu = 0$ vs. $\mu \neq 0$ based on (X,T). Express the test statistic in terms of $f(X,T)$. Show that when $\mu = 0$, $XX'+T$ is independent of $X'T^{-1}X$, hence is independent of $f(X,T)$ and therefore independent of the LRT statistic.

Partial solution: The LRT rejects $\mu = 0$ for *large* values of

$$\text{(6.29)} \qquad \frac{|XX'+T|}{|T|} \stackrel{*}{=} |X'T^{-1}X + I_r| \equiv \prod\nolimits_{i=1}^{q}(f_i(X,T)+1).$$

(*recall Exercise 1 on p.12). When $\mu = 0$, $XX'+T$ is a complete and sufficient statistic for Σ, and $X'T^{-1}X$ is an ancillary statistic, hence they are independent by Basu's Lemma. (Also see §7.1.) $\bar{\square}$

Exercise 6.37b. Let U be the matrix-variate Beta rv (recall Exercise 4.2) defined as

$$\text{(6.30)} \qquad U := (XX'+T)^{-1/2}\,T\,(XX'+T)^{-1/2\,\prime} \quad (n \geq p).$$

Derive the moments of $|U| \equiv \frac{|T|}{|XX'+T|}$ under the null hypothesis $\mu = 0$.

Solution: By Exercise 6.37a and (6.29), $|XX'+T| \perp\!\!\!\perp |U|$, so

$$\text{(6.31)} \qquad \mathrm{E}(|T|^k) = \mathrm{E}(|XX'+T|^k|U|^k) = \mathrm{E}(|XX'+T|^k)\,\mathrm{E}(|U|^k),$$

hence (compare to (4.16))

$$\text{(6.32)} \qquad \mathrm{E}(|U|^k) = \frac{\Gamma_p\left(\frac{r+n}{2}\right)\Gamma_p\left(\frac{n}{2}+k\right)}{\Gamma_p\left(\frac{n}{2}\right)\Gamma_p\left(\frac{r+n}{2}+k\right)}. \qquad \bar{\square}$$

Exercise 6.37c. Let $U(p,\, r,\, n)$ denote the null ($\mu = 0$) distribution of $|U|$. ($U(p,\, r,\, n)$ is called *Wilks' distribution.*) Show that this distribution can be represented as the product of independent Beta distributions:

$$U(p, r, n) \sim \prod_{i=1}^{r} B\big(\tfrac{n-p+i}{2},\, \tfrac{p}{2}\big), \tag{6.33}$$

where the Beta variates are mutually independent.

Note: The moments of $|U|$ given in (6.32) or obtained directly from (6.33) can be used to obtain the *Box approximation*, an improvement of the chi-square approximation to the Wilks' distribution $U(p,\, r,\, n)$. (See T. W. Anderson's book, §8.5; also see Ledet-Jensen (1991), JASA.) $\bar{\square}$

Exercise 6.37d. In Exercise 6.24 it was found that the LRT for testing

$$\Sigma_{12} = 0 \quad \text{vs.} \quad \Sigma_{12} \neq 0 \tag{6.34}$$

(i.e., testing independence of two sets of variates) rejects $\Sigma_{12} = 0$ for *small* values of $\frac{|S|}{|S_{11}||S_{22}|} = \frac{|S_{11\cdot 2}|}{|S_{12}S_{22}^{-1}S_{21}+S_{11\cdot 2}|}$. Show that the null ($\Sigma_{12} = 0$) distribution of this LRT statistic is $U(p_1,\, p_2,\, n - p_2)$ – see Exercise 6.24. $\bar{\square}$

Exercise 6.37e. Show that $U(p,\, r,\, n) \sim U(r,\, p,\, r + n - p)$, hence

$$U(p,\, r,\, n) \sim \prod_{i=1}^{p} B\big(\tfrac{n-p+i}{2},\, \tfrac{r}{2}\big). \tag{6.35}$$ $\bar{\square}$

Remark 6.38. Perlman and Olkin (*Annals of Statistics* 1980) applied the FKG inequality to show that the LRTs in Exercises 6.24 and 6.37a are unbiased. $\bar{\square}$

Example 6.39. The canonical GMANOVA Model.

(Example 6.13 is a special case.) [To be completed] $\bar{\square}$

Example 6.40. An inadmissible UMPI test. *(C. Stein – see Lehmann TSH Example 11 p.305 and Example 9 p.522.)*

Consider Example 6.32 (testing $\Sigma_1 = \Sigma_2$) with $p > 1$ but with $n_1 = n_2 = 1$, so S_1 and S_2 are each singular of rank 1. This problem again remains invariant under the action of GL on (S_1, S_2) given by (6.21):

$$(S_1, S_2) \mapsto (gS_1g',\, gS_2g').$$

Here, however, GL acts transitively [verify] on this sample space since S_1, S_2 each have rank 1, so the MIS is trivial: $t(S_1, S_2) \equiv$ const. This implies that the only size α invariant test is $\phi(S_1, S_2) \equiv \alpha$, so its power is identically α. However, there exist more powerful non-invariant tests. For any nonzero $a : p \times 1$, let

$$V_a = \frac{a'S_1a}{a'S_2a} \sim \frac{a'\Sigma_1 a}{a'\Sigma_2 a} \cdot F_{1,1} \equiv \delta_a \cdot F_{1,1} \tag{6.36}$$

and let ϕ_a denote the UMPU size α test for testing $\delta_a = 1$ vs. $\delta_a \neq 1$ based on V_a (cf. TSH Ch.5 §3). Then [verify]: ϕ_a is unbiased size α for testing $\Sigma_1 = \Sigma_2$, with power $> \alpha$ when $\delta_a \neq 1$, so ϕ_a dominates the UMPI test ϕ.

Note: This failure of invariance to yield a nontrivial UMPI test is usually attributed to the group GL being "too large", i.e., not "amenable".[13] However, this example is somewhat artificial in that the sample sizes are too small ($n_1 = n_2 = 1$) to permit estimation of Σ_1 and Σ_2. It would be of interest to find (if possible?) an example of a trivial UMPI test in a less contrived model. □

Exercise 6.41. Another inadmissible UMPI test. *(see Lehmann TSH Problem 11 p.532.)*

Consider Example 6.10 (testing $\mu = 0$ with Σ unknown) with $n > 1$ observations but $n < p$. As in Example 6.40, show that the UMPI GL-invariant test is trivial but there exists more powerful non-invariant tests. □

[13] See Bondar and Milnes (1981) *Zeit. f. Wahr.* **57**, pp. 103-128.

7. Distribution of Eigenvalues. (See T.W.Anderson book, Ch. 13.)

In the invariant testing problems of Examples 6.22 (testing $\Sigma_{12} = 0$), 6.32 (testing $\Sigma_1 = \Sigma_2$), and 6.36 (the canonical MANOVA testing problem), the maximal invariant statistic (MIS) can be represented as the set of nontrivial eigenvalues of a matrix of one of the equivalent forms

$$ST^{-1} \quad \text{or} \quad S(S+T)^{-1} \quad \text{or} \quad T(S+T)^{-1},$$

where S and T are independent Wishart matrices. Under the null hypothesis, both are central Wishart matrices with common Σ. Because the LRT statistic is invariant (Lemma 6.7), it is necessarily a function of these eigenvalues. When the dimensionality of the invariance-reduced alternative hypothesis is ≥ 2,[14] however, no single invariant test is UMPI, and other reasonable invariant test statistics[15] have been proposed: for example,

$$r_1^2 \ \text{(Roy)} \ \text{ and } \ \sum\nolimits_{i=1}^{q} \frac{r_i^2}{1-r_i^2} \ \text{(Lawley} - \text{Hotelling)}$$

in Example 6.22, where $(S, T) = (S_{12}S_{22}^{-1}S_{21},\, S_{11\cdot 2})$, and

$$f_1 \ \text{(Roy)} \ \text{ and } \ \sum\nolimits_{i=1}^{q} f_i \ \text{(Lawley} - \text{Hotelling)}$$

in Example 6.36, where $(S,\, T) = (XX',\, T)$.

Thus, to determine the distribution of such invariant test statistics it is necessary to determine the distribution of the eigenvalues of ST^{-1}, or equivalently, of $S(S+T)^{-1}$ or $T(S+T)^{-1}$.

7.1. The central distribution of the eigenvalues of $S(S+T)^{-1}$.

Let S and T be independent with $S \sim W_p(r, \Sigma)$ and $T \sim W_p(n, \Sigma)$, $\Sigma > 0$. Assume further that $n \geq p$, so $T > 0$ w. pr. 1. Let

$$1 \geq \ b_1 \geq \cdots \geq b_q \ > 0 \quad \text{and} \quad f_1 \geq \cdots \geq f_q \ > 0$$

[14] For example, see (6.14), (6.22), and (6.26).

[15] Schwartz (*Ann. Math. Statist.* (1967) 698-710), presents a sufficient condition and a (weaker) necessary condition for an invariant test to be admissible among all tests.

denote the $q \equiv \min(p, r)$ ordered nontrivial[16] eigenvalues of $S(S+T)^{-1}$ (the Beta form) and ST^{-1} (the F form), respectively. Set

$$\begin{aligned} b &\equiv (b_1, \ldots, b_q) \equiv \{b_i(S,T)\}, \\ f &\equiv (f_1, \cdots, f_q) \equiv \{f_i(S,T)\}. \end{aligned} \tag{7.1}$$

First we shall derive the pdf of b, then obtain the pdf of f using the relation

$$f_i = \frac{b_i}{1-b_i}. \tag{7.2}$$

Because b is GL-invariant, i.e.,

$$b_i(S,\, T) = b_i(ASA',\, ATA') \qquad \forall\, A \in GL,$$

the distribution of b does not depend on Σ [verify], so we may set $\Sigma = I_p$. Denote this distribution by $b(p, r, n)$ and the corresponding distribution of f by $f(p, r, n)$.

Exercise 7.1. Show that (compare to Exercise 6.37e)

$$\begin{aligned} b(p,\, r,\, n) &= b(r,\, p,\, n+r-p), \\ f(p,\, r,\, n) &= f(r,\, p,\, n+r-p). \end{aligned} \tag{7.3}$$

Outline of solution: Let W be a partitioned Wishart random matrix:

$$W \equiv \begin{matrix} \\ p \\ r \end{matrix} \begin{matrix} \quad p \qquad\quad r \\ \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \end{matrix} \sim W_{p+r}(m,\, I_{p+r}).$$

Assume that $m \geq \max(p, r)$, so $W_{11} > 0$ and $W_{22} > 0$ w. pr. 1. By the properties of the distribution of a partitioned Wishart matrix (Proposition 3.13),

(a) the distribution of the nontrivial eigenvalues of $W_{12}W_{22}^{-1}W_{21}W_{11}^{-1}$ is $b(p,\, r,\, m-r)$ [verify!]

[16] If $p > r$ then $q = r$ and $p - r$ of the eigenvalues of $T(T+S)^{-1}$ are trivially $\equiv 1$. By Okamoto's Lemma the nontrivial eigenvalues are distinct w. pr. 1.

(b) the distribution of the nontrivial eigenvalues of $W_{21}W_{11}^{-1}W_{12}W_{22}^{-1}$ is $b(r,\, p,\, m-p)$ [verify!].

But these two sets of nontrivial eigenvalues are identical[17] so the result follows by setting $n = m - r$. $\square$

By Exercise 7.1 it suffices to derive the distribution $b(p, r, n)$ when $r \geq p$, where $q = p$. Because $r \geq p$, also $S > 0$ w. pr. 1, so by (4.11) the joint pdf of (S, T) is

$$c_{p,r}\, c_{p,n} \cdot |S|^{\frac{r-p-1}{2}} |T|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\mathrm{tr}(S+T)}, \qquad S > 0,\ T > 0.$$

Make the transformation

$$(S,\, T) \mapsto (S,\ V \equiv S + T).$$

By the extended combination rule, the Jacobian is 1, so the joint pdf of (S, V) is

$$c_{p,r}\, c_{p,n} \cdot |S|^{\frac{r-p-1}{2}} |V - S|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\mathrm{tr} V}, \qquad V > S > 0.$$

By E.3, there exists a unique [verify] nonsingular $p \times p$ matrix $E \equiv \{e_{ij}\}$ with $e_{1j} > 0$, $j = 1, \ldots, p$, such that

$$\begin{aligned} S &= E\, D_b\, E', \\ V &= E\, E', \end{aligned} \tag{7.4}$$

where $D_b := \mathrm{Diag}(b_1, \ldots, b_p)$. Thus the joint pdf of (b, E) is given by

$$f(b, E) = c_{p,r} c_{p,n} \cdot |D_b|^{\frac{r-p-1}{2}} |I_p - D_b|^{\frac{n-p-1}{2}} |EE'|^{\frac{r+n-2p-2}{2}} e^{-\frac{1}{2}\mathrm{tr}\, EE'} \left| \frac{\partial(S, V)}{\partial(b, E)} \right|,$$

where the range $\mathcal{R}_{b,E}$ is the Cartesian product $\mathcal{R}_b \times \mathcal{R}_E$ with

$$\begin{aligned} \mathcal{R}_b &:= \{b \mid 1 > b_1 > \cdots > b_p > 0\}, \\ \mathcal{R}_E &:= \{E \mid e_{1j} > 0,\ -\infty < e_{ij} < \infty,\ i = 2, \ldots, p,\ j = 1, \ldots, p\}. \end{aligned}$$

[17] Because $|\lambda I_p - AB| = \begin{vmatrix} \lambda I_p & A \\ B & I_r \end{vmatrix} = |\lambda I_p| \cdot |\frac{1}{\lambda}(\lambda I_r - BA)|$.

We will show that

$$(7.5)\qquad \Big|\frac{\partial(S,V)}{\partial(b,E)}\Big| = 2^p \cdot |EE'|^{\frac{p+2}{2}} \cdot \prod_{i<j}(b_i - b_j),$$

hence

$$\begin{aligned} f(b,E) =& 2^p c_{p,r}\, c_{p,n} \cdot \prod_{i=1}^p b_i^{\frac{r-p-1}{2}} \prod_{i=1}^p (1-b_i)^{\frac{n-p-1}{2}} \prod_{1\le i<j\le p} (b_i - b_j) \\ &\cdot |EE'|^{\frac{r+n-p}{2}} e^{-\frac{1}{2}\operatorname{tr} EE'}. \end{aligned}$$

Because $\mathcal{R}_{b,E} = \mathcal{R}_b \times \mathcal{R}_E$, this implies that b *and* E *are independent* with marginal pdfs given by

$$(7.6)\qquad f(b) = c_b \cdot \prod_{i=1}^p b_i^{\frac{r-p-1}{2}} (1-b_i)^{\frac{n-p-1}{2}} \cdot \prod_{1\le i<j\le p} (b_i - b_j),\ \ b \in \mathcal{R}_b,$$

$$(7.7)\qquad f(E) = c_E \cdot |EE'|^{\frac{r+n-p}{2}} e^{-\frac{1}{2}\operatorname{tr} EE'}, \qquad\qquad E \in \mathcal{R}_E,$$

where

$$c_b\, c_E = 2^p\, c_{p,r}\, c_{p,n}.$$

Thus, to determine c_b it suffices to determine c_E. This is accomplished as follows:

$$\begin{aligned} c_E^{-1} &= \int_{\mathcal{R}_E} |EE'|^{\frac{r+n-p}{2}} e^{-\frac{1}{2}\operatorname{tr} EE'} dE \\ &= 2^{-p} \int_{\mathcal{R}^{p^2}} |EE'|^{\frac{r+n-p}{2}} e^{-\frac{1}{2}\operatorname{tr} EE'} dE && \text{[by symmetry]} \\ &= 2^{-p} (2\pi)^{\frac{p^2}{2}} \int_{\mathcal{R}^{p^2}} |EE'|^{\frac{r+n-p}{2}} \prod_{i,j=1}^p \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}e_{ij}^2} de_{ij} \\ &= 2^{-p} (2\pi)^{\frac{p^2}{2}} \cdot \mathrm{E}\Big(|W_p(p, I_p)|^{\frac{r+n-p}{2}}\Big) && \text{[why?]} \\ &= 2^{-p} (2\pi)^{\frac{p^2}{2}} \cdot \frac{c_{p,p}}{c_{p,r+n}}. && \text{[by (4.12)]} \end{aligned}$$

Therefore (recall (4.10))

$$c_b \equiv c_b(p, r, n) = (2\pi)^{\frac{p^2}{2}} \cdot \frac{c_{p,p}\, c_{p,r}\, c_{p,n}}{c_{p,r+n}}. \tag{7.8}$$

This completes the derivation of the pdf $f(b)$ in (7.6), hence determines the distribution $b(p, r, n)$ when $r \geq p$. Note that this can be viewed as another generalization of the Beta distribution.

Verification of the Jacobian (7.5):

By the linearization method (*) in §4.3,

$$\left|\frac{\partial(S, V)}{\partial(b, E)}\right| = \left|\frac{\partial(dS, dV)}{\partial(db, dE)}\right|. \tag{7.9}$$

From (7.4),

$$\begin{aligned} dS &= (dE)D_b E' + E\, D_{db}\, E' + ED_b(dE)', \\ dV &= (dE)E' + E(dE)', \end{aligned}$$

hence, defining

$$\begin{aligned} dF &= E^{-1}(dE), \\ dG &= E^{-1}(dS)(E^{-1})', \\ dH &= E^{-1}(dV)(E^{-1})', \end{aligned}$$

we have

$$dG = (dF)D_b + D_{db} + D_b(dF)', \tag{7.10}$$

$$dH = (dF) + (dF)'. \tag{7.11}$$

To evaluate $\left|\frac{\partial(dS,dV)}{\partial(db,dE)}\right|$, apply the chain rule to the sequence

$$(db, dE) \mapsto (db, dF) \mapsto (dG, dH) \mapsto (dS, dV)$$

to obtain

$$\left|\frac{\partial(dS, dV)}{\partial(db, dE)}\right| = \left|\frac{\partial(db, dF)}{\partial(db, dE)}\right| \cdot \left|\frac{\partial(dG, dH)}{\partial(db, dF)}\right| \cdot \left|\frac{\partial(dS, dV)}{\partial(dG, dH)}\right|.$$

By 4.2(d), 4.2(e), and the combination rule in §4.1,

$$(7.12)\qquad \left|\frac{\partial(db,dF)}{\partial(db,dE)}\right| = \left|\frac{\partial(dF)}{\partial(dE)}\right| = |E|^{-p},$$

$$(7.13)\qquad \left|\frac{\partial(dS,dV)}{\partial(dG,dH)}\right| = \left|\frac{\partial(dS)}{\partial(dG)}\right| \cdot \left|\frac{\partial(dV)}{\partial(dH)}\right| = |E|^{p+1}|E|^{p+1} = |E|^{2(p+1)}.$$

Lastly we evaluate $\left|\frac{\partial(dG,dH)}{\partial(db,dF)}\right| =: J$. Here $dG \equiv \{dg_{ij}\}$ and $dH \equiv \{dh_{ij}\}$ are $p \times p$ symmetric matrices, $dF \equiv \{df_{ij}\}$ is a $p \times p$ unconstrained matrix, and db is a vector of dimension p. From (7.10) and (7.11),

$$\begin{aligned} dg_{ii} &= 2(df_{ii})b_i + db_i, && i = 1, \dots, p, \\ dh_{ii} &= 2df_{ii}, && i = 1, \dots, p, \\ dg_{ij} &= (df_{ij})b_j + b_i(df_{ji}), && 1 \le i < j \le p, \\ dh_{ij} &= df_{ij} + df_{ji}, && 1 \le i < j \le p. \end{aligned}$$

Therefore

$$J = \left|\frac{\partial\big((dg_{ii}), (dh_{ii}), (dg_{ij}), (dh_{ij})\big)}{\partial\big((db_i), (df_{ii}), (df_{ij}), (df_{ji})\big)}\right|$$

$$= \begin{vmatrix} I_p & 0 & 0 & 0 \\ 2D_b & 2I_p & 0 & 0 \\ 0 & 0 & D_1 & I_{p(p-1)/2} \\ 0 & 0 & D_2 & I_{p(p-1)/2} \end{vmatrix},$$

where

$$\begin{aligned} D_1 &:= \mathrm{Diag}(b_2, \dots, b_p, b_3, \dots, b_p, \dots, b_{p-1}, b_p, b_p) \\ D_2 &:= \mathrm{Diag}(b_1, \dots, b_1, b_2, \dots, b_2, \dots, b_{p-2}, b_{p-2}, b_{p-1}), \end{aligned}$$

hence [verify!]

$$(7.14)\qquad J = 2^p\,|D_1 - D_2| = 2^p \prod_{1 \le i < j \le p} (b_i - b_j).$$

The desired Jacobian (7.5) follows from (7.12), (7.13), and (7.14). ◻

Exercise 7.2. Use (7.2) to show that if $r \geq p$, the pdf of $(f_1, \cdots, f_p)$ is given by

$$(7.15) \qquad c_b(p,r,n) \prod_{i=1}^{p} f_i^{\frac{r-p-1}{2}} (1+f_i)^{-\frac{r+n}{2}} \prod_{1\leq i<j\leq p} (f_i - f_j),$$

where $c_b(p,r,n)$ is given by (7.8). If $r < p$ then the pdf of $(f_1, \cdots, f_r)$ follows from $f(p,\, r,\, n) = f(r,\, p,\, n+r-p)$ in (7.3). $\square$

Exercise 7.3. Under the weaker assumption that $r+n \geq p$, show that the distribution of $b \equiv \{b_i(S,T)\}$ does not depend on Σ and that b and V are independent. (Note that $f \equiv \{f_i(S,T)\}$ is not defined unless $n \geq p$.)

Hint: Apply the GL-invariance of $\{b_i(S,T)\}$ and Basu's Lemma. If $r \geq p$ and $n \geq p$ the result also follows from Exercise 4.2. $\square$

7.2. The eigenvalues and eigenvectors of $S \sim W_p(r, I)$ when $r \geq p$.

In the invariant testing problems of Examples 6.18 (testing $\Sigma = I$) and Exercise 6.20 (testing $\Sigma = \kappa I$), the maximal invariant statistic (MIS) can be represented in terms of the set of ordered eigenvalues

$$\{l_1 \geq \cdots \geq l_p\} \equiv \{l_i(S)\}$$

of a single Wishart matrix $S \sim W_p(r,\, \Sigma)$ $(r \geq p,\ \Sigma > 0)$. Again the LRT statistic is invariant so is necessarily a function of these eigenvalues.

As in §7.1, when the dimensionality of the invariance-reduced alternative hypothesis is ≥ 2 (e.g. (6.11)), no single invariant test is UMPI – other reasonable invariant test statistics include

$$1 - I_{\{a<l_p<l_1<b\}} \ \text{(Roy)} \ \text{ and } \ \sum\nolimits_{i=1}^{p} \left(\tfrac{1}{r} l_i - 1\right)^2 \ \text{(Nagao)}$$

in Example 6.18 and

$$\tfrac{l_1}{l_p} \ \text{(Roy)} \ \text{ and } \ \sum\nolimits_{i=1}^{p} l_i \cdot \sum\nolimits_{i=1}^{p} \tfrac{1}{l_i} \ \text{(Nagao)}$$

in Example 6.20. To determine the null distribution of such invariant test statistics we need to find the distribution of $(l_1, \ldots, l_p)$ when $\Sigma = I$.

Proposition 7.4. (the "Eigenlemma"). Let U be a random $p \times p$ positive definite matrix with spectral decomposition $U = \Gamma_U D_{l(U)} \Gamma'_U$. Here the columns of the random orthogonal matrix $\Gamma \equiv \Gamma_U \in \mathcal{O}$ are the eigenvectors of U and $l \equiv l(U) = (l_1 > \cdots > l_p)$ are the ordered eigenvalues of U. Suppose that the pdf of U depends only on its eigenvalues, that is, its pdf is $g(U) = \phi(l)$ for some ϕ. Then l has pdf

$$\phi^*(l) = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})}\, \phi(l) \prod_{1 \le i < j \le p} (l_i - l_j) \tag{7.16}$$

on $\mathcal{R}_l := \{l \mid \infty > l_1 > \cdots > l_p > 0\}$. Furthermore l and Γ are independent, and Γ has the orthogonally invariant Haar probability distribution on $\mathcal{O}$.

Proof. Begin with the spectral decomposition

$$\begin{aligned} U &= \Gamma\, D_l\, \Gamma', \\ I &= \Gamma\, \Gamma', \end{aligned} \tag{7.17}$$

where $D_l = \mathrm{Diag}(l_1, \ldots, l_p)$. The joint pdf of (Γ, l) is given by

$$g(\Gamma, l) = \left| \frac{\partial U}{\partial(\Gamma, l)} \right| \cdot \phi(l), \qquad (\Gamma, l) \in \mathcal{O} \times \mathcal{R}_l, \tag{7.18}$$

for an arbitrary but smooth parameterization of $\Gamma \in \mathcal{O}$. From (7.17),

$$\begin{aligned} dU &= (d\Gamma)\, D_l\, \Gamma' + \Gamma\, D_{dl}\, \Gamma' + \Gamma\, D_l\, (d\Gamma)', \\ 0 &= (d\Gamma)\, \Gamma' + \Gamma\, (d\Gamma)', \end{aligned}$$

hence, defining $dF = \Gamma^{-1}(d\Gamma)$,

$$\begin{aligned} dG :&= \Gamma^{-1}(dU)(\Gamma^{-1})' = (dF)D_l + D_{dl} + D_l(dF)', \\ 0 \ &= (dF) + (dF)'. \end{aligned}$$

Thus $dG \equiv \{dg_{ij}\}$ is symmetric, $dF \equiv \{df_{ij}\}$ is skew-symmetric, and

$$dG = (dF)D_l + D_{dl} - D_l(dF). \tag{7.19}$$

To evaluate $\left|\frac{\partial U}{\partial(\Gamma,l)}\right| = \left|\frac{\partial(dU)}{\partial(d\Gamma,dl)}\right|$, apply the chain rule to the sequence

$$(d\Gamma, dl) \mapsto (dF, dl) \mapsto dG \mapsto dU$$

to obtain

$$\left|\frac{\partial(dU)}{\partial(d\Gamma, dl)}\right| = \underbrace{\left|\frac{\partial(dF, dl)}{\partial(d\Gamma, dl)}\right|}_{=\eta(\Gamma)} \cdot \underbrace{\left|\frac{\partial(dG)}{\partial(dF, dl)}\right|}_{\equiv J} \cdot \underbrace{\left|\frac{\partial(dU)}{\partial(dG)}\right|}_{=1} \qquad \text{[verify]}$$

for some function $\eta(\Gamma)$. From (7.19),

$$\begin{aligned} dg_{ii} &= dl_i, & i &= 1, \ldots, p, \\ dg_{ij} &= (df_{ij})(l_j - l_i), & 1 &\le i < j \le p, \end{aligned}$$

(note that $df_{ii} = 0$ by skew-symmetry), so

$$J = \left|\frac{\partial\big((dg_{ij}), (dg_{ii})\big)}{\partial\big((df_{ij}, (dl_i))\big)}\right| = \begin{vmatrix} D & 0 \\ * & I_p \end{vmatrix} = |D| = \prod_{1\le i<j\le p} (l_i - l_j),$$

where D is the $\frac{p(p-1)}{2} \times \frac{p(p-1)}{2}$ diagonal matrix

$$\mathrm{Diag}(l_2 - l_1, \ldots, l_p - l_1, l_3 - l_2, \ldots, l_p - l_2, \ldots, l_p - l_{p-1}).$$

Thus from (7.18), the joint pdf of (Γ, l) is

$$g(l, \Gamma) = \eta(\Gamma)\, \phi(l) \prod_{1\le i<j\le p} (l_i - l_j), \qquad (\Gamma, l) \in \mathcal{O} \times \mathcal{R}_l, \tag{7.20}$$

hence Γ and l are independent. Furthermore, the pdf of l is

$$\phi^*(l) = \left(\int_{\mathcal{O}_p} \eta(\Gamma) d\Gamma\right) \phi(l) \prod_{1\le i<j\le p} (l_i - l_j). \tag{7.21}$$

Denote the integral by K. It can be evaluated by the theory of differential forms on smooth manifolds[18] but we will take a different approach.

[18] This approach is followed in the books by R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (1982) and R. H. Farrell, *Multivariate Calculation* (1985).

Consider the special case where U has the matrix-variate beta pdf given in (4.14): for $0 < U < I$,

$$\begin{aligned} g(U) &= \frac{c_{p,\,r}c_{p,\,n}}{c_{p,\,r+n}}|U|^{\frac{r-p-1}{2}}|I-U|^{\frac{n-p-1}{2}} \\ &= \frac{c_{p,\,r}c_{p,\,n}}{c_{p,\,r+n}}\prod_{i=1}^{p} l_i^{\frac{r-p-1}{2}}(1-l_i)^{\frac{n-p-1}{2}} \equiv \phi(l), \end{aligned}$$

where $r \geq p$ and $n \geq p$. Then by (7.21),

$$\phi^*(l) = (K)\,\frac{c_{p,\,r}c_{p,\,n}}{c_{p,\,r+n}}\prod_{i=1}^{p} l_i^{\frac{r-p-1}{2}}(1-l_i)^{\frac{n-p-1}{2}}\prod_{1\leq i<j\leq p}(l_i-l_j).$$

By comparing this to (7.6) and applying (7.8) we see that

$$K\frac{c_{p,\,r}c_{p,\,n}}{c_{p,\,r+n}} = c_b = (2\pi)^{\frac{p^2}{2}}\,\frac{c_{p,\,p}c_{p,\,r}c_{p,\,n}}{c_{p,\,r+n}},$$

hence

$$K = (2\pi)^{\frac{p^2}{2}}c_{p,\,p} = \frac{(2\pi)^{\frac{p^2}{2}}}{2^{\frac{p^2}{2}}\Gamma_p(\frac{p}{2})} = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})}.$$

which proves (7.16).

Lastly, because the invariant Haar probability measure on $\mathcal{O}$ is unique, it suffices to show that $\Gamma_U \sim \Psi\Gamma_U$ for any fixed $\Psi \in \mathcal{O}$. If we define

$$\tilde{U} = \Psi\,U\,\Psi' \sim W_p(r,\,I),$$

then

$$\tilde{U} = (\Psi\Gamma_U)D_{l(U)}(\Psi\Gamma_U)',$$

so $\Gamma_{\tilde{U}} = \Psi\,\Gamma_U$ by the uniqueness[20] of the spectral decomposition. Because $\tilde{U} \sim U$, necessarily $\Gamma_{\tilde{U}} \sim \Gamma_U$, which implies that $\Psi\Gamma_U \sim \Gamma_U$, as required.□

[20] It must be noted Γ_U is uniquely defined only if the signs of a specified row vector are fixed, say all positive. Lemma 13.3.2 of T.W. Anderson's book deals with this unpleasant complication. A more elegant approach is to define the eigenvectors without regard to direction, that is, to identify each unit vector with its antipode on the unit sphere.

We now apply the Eigenlemma to obtain the distribution of the ordered eigenvalues $l \equiv l(S)$ of $S \sim W_p(r, I)$ with $r \geq p$. Here S has pdf

$$g(S) = c_{p,\,r}|S|^{\frac{r-p-1}{2}} e^{-\frac{1}{2}\mathrm{tr}S} = c_{p,\,r}\textstyle\prod_{i=1}^p l_i^{\frac{r-p-1}{2}} e^{-\frac{1}{2}l_i} \equiv \phi(l),$$

so by (7.16) l has pdf

$$\phi^*(l) = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})}\, c_{p,\,r} \prod_{i=1}^p l_i^{\frac{r-p-1}{2}} e^{-\frac{1}{2}l_i} \prod_{1\leq i<j\leq p} (l_i - l_j)$$

$$= \frac{\pi^{\frac{p^2}{2}}}{2^{\frac{pr}{2}}\Gamma_p\left(\frac{p}{2}\right)\Gamma_p\left(\frac{r}{2}\right)} \prod_{i=1}^p l_i^{\frac{r-p-1}{2}} e^{-\frac{1}{2}l_i} \prod_{1\leq i<j\leq p} (l_i - l_j). \tag{7.22}$$

Exercise 7.5*.** (*A second possible approach to (7.22).*) Use the limit representation

$$l_i(S) = \lim_{n\to\infty} f_i\left(S,\, \tfrac{1}{n}T\right) = \lim_{n\to\infty} nf_i\left(S,\, T\right). \tag{7.23}$$

Let $l_{i,n} = nf_i(S,T)$, $i = 1,\dots,p$, and derive the pdf ϕ_n of $(l_{1,n},\dots,l_{p,n})$ from the pdf of $(f_1,\dots,f_p)$ in (7.15). Use Stirling's approximation for the gamma function to find $\lim_{n\to\infty}\phi_n \equiv \phi^*$. Show that ϕ^* is a pdf by the Dominated Convergence Theorem or otherwise, then apply Scheffe's Theorem to conclude that ϕ^* is the pdf of l. □

7.3. Stein's integral representation of the density of a maximal invariant statistic with applications to noncentral distributions.

Proposition 7.6. Suppose that the distribution of X is given by a pdf $f(x)$ w. r. to a measure μ on the sample space $\mathcal{X}$. Assume that μ is invariant under the action of a compact topological group G acting on $\mathcal{X}$, i.e. μ is G-invariant: $\mu(gB) = \mu(B)$ for all events $B \subseteq \mathcal{X}$ and all $g \in G$. If

$$t : \mathcal{X} \to \mathcal{T}$$
$$x \mapsto t(x)$$

is a maximal invariant statistic then the pdf of t w.r. to the induced measure $\tilde{\mu} = \mu(t^{-1})$ on $\mathcal{T}$ is given by

$$\bar{f}(x) = \int_G f(gx)\, d\nu(g), \tag{7.26}$$

where ν is the Haar probability measure on G.

Proof: First we show that $\bar{f}(x)$ is actually a function of the MIS t. The integral is simply the average of $f(\cdot)$ over all members gx in the G-orbit of x. By the left G-invariance of ν, $\bar{f}(\cdot)$ is also G-invariant:

$$\bar{f}(g_1 x) = \int_G f(g g_1 x)\, d\nu(g) = \int_G f(g' x)\, d\nu(g_1^{-1} g') = \bar{f}(x) \quad \forall\, g_1 \in G,$$

hence $\bar{f}(x) = h(t(x))$ for some function $h(t)$.

Next, for any event $A \subseteq \mathcal{T}$ and any $g \in G$,

$$\begin{aligned} P[\,t(X) \in A\,] &= \int_{\mathcal{X}} I_A(t(x))\, f(x)\, d\mu(x) \\ &= \int_{\mathcal{X}} I_A(t(g^{-1}x))\, f(x)\, d\mu(x) \\ &= \int_{\mathcal{X}} I_A(t(z))\, f(gz)\, d\mu(z), \end{aligned}$$

by the G-invariance of t and μ, so

$$\begin{aligned} P[\,t(X) \in A\,] &= \int_G \left[\int_{\mathcal{X}} I_A(t(z))\, f(gz)\, d\mu(z) \right] d\nu(g) \\ &= \int_{\mathcal{X}} I_A(t(z)) \left[\int_G f(gz)\, d\nu(g) \right] d\mu(z) \\ &= \int_{\mathcal{X}} I_A(t(z))\, h(t(z))\, d\mu(z) \\ &= \int_{\mathcal{T}} I_A(t)\, h(t)\, d\tilde{\mu}(t). \end{aligned}$$

Thus $h(t) \equiv \bar{f}(x)$ is the pdf of $t \equiv t(X)$ w. r. to $d\tilde{\mu}(t)$. $\bar{\square}$

Example 7.7. Let $\mathcal{X} = \mathcal{R}^p$, $G = \mathcal{O}_p$, and $\mu =$ Lebesgue measure on $\mathcal{R}^p$, an $\mathcal{O}_p$-invariant measure. Here $\gamma \in \mathcal{O}_p$ acts on $\mathcal{R}^p$ via $x \mapsto \gamma x$. A maximal invariant statistic is $t(X) = \|X\|^2$. If X has pdf $f(x)$ w. r. to μ then the integral representation (7.26) states that $t(X) \equiv \|X\|^2$ has pdf

$$h(t) = \int_{\mathcal{O}_p} f(\gamma x)\, d\nu_p(\gamma) \tag{7.27}$$

w. r. to $d\tilde{\mu}(t)$ on $(0,\infty)$, where ν_p is the Haar probability measure on $\mathcal{O}_p$. In particular, if $f(x)$ is also $\mathcal{O}_p$-invariant, i.e., if

$$f(x) = k(\|x\|^2)$$

for some $k(\cdot)$ on $(0,\,\infty)$, then the pdf of $t(X)$ w.r.to $d\tilde{\mu}(t)$ is simply

$$h(t) = k(t), \qquad t \in (0,\infty). \tag{7.28}$$

The induced measure $d\tilde{\mu}(t)$ can be found by considering a special case: If $X \sim N_p(0,\, I_p)$ then $t \equiv \|X\|^2 \sim \chi^2_p$. Here

$$f(x) = \tfrac{1}{(2\pi)^{\frac{p}{2}}}\, e^{-\frac{1}{2}\|x\|^2} \equiv k(\|x\|^2) \qquad \text{w.r. to } d\mu(x),$$

so t has pdf

$$h(t) = k(t) = \tfrac{1}{(2\pi)^{\frac{p}{2}}}\, e^{-\frac{1}{2}t} \qquad\qquad \text{w.r. to } d\tilde{\mu}(t).$$

We also know, however, that t has the χ^2_p pdf

$$w(t) \equiv \tfrac{1}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)}\, t^{\frac{p}{2}-1}e^{-\frac{1}{2}t} \qquad \text{w.r. to } dt \ (\equiv \text{Lebesgue measure}).$$

Therefore $d\tilde{\mu}(t)$ is determined as follows:

$$d\tilde{\mu}(t) = \frac{w(t)}{k(t)}\,dt = \frac{\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)}\, t^{\frac{p}{2}-1}\,dt. \tag{7.29}$$

Application: We can use Stein's representation (7.27) to give an alternative derivation of the noncentral chi-square pdf in (2.25) – (2.27). Suppose that $X \sim N_p(\xi,\, I_p)$ with $\xi \neq 0$, so

$$t \equiv \|X\|^2 \sim \chi^2_p(\delta) \quad \text{with } \delta = \|\xi\|^2.$$

Here

$$f(x) = \tfrac{1}{(2\pi)^{\frac{p}{2}}}\, e^{-\frac{1}{2}\|x-\xi\|^2},$$

so by (7.27), t has pdf w. r. to $d\tilde{\mu}(t)$ given by

$$\begin{aligned}
h(t) &= \frac{1}{(2\pi)^{\frac{p}{2}}} \int_{\mathcal{O}_p} e^{-\frac{1}{2}\|\gamma x - \xi\|^2}\, d\nu_p(\gamma) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}\|\xi\|^2} e^{-\frac{1}{2}\|x\|^2} \int_{\mathcal{O}_p} e^{x'\gamma\xi}\, d\nu_p(\gamma) \\
&\overset{(1)}{=} \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{\delta}{2}} e^{-\frac{t}{2}} \int_{\mathcal{O}_p} e^{t^{\frac{1}{2}}\delta^{\frac{1}{2}}\gamma_{11}}\, d\nu_p(\gamma) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{\delta}{2}} e^{-\frac{t}{2}} \sum_{k=0}^{\infty} \frac{(t\delta)^{\frac{k}{2}}}{k!} \int_{\mathcal{O}_p} \gamma_{11}^k\, d\nu_p(\gamma) \qquad [\,\gamma = \{\gamma_{ij}\}\,] \\
&\overset{(2)}{=} \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{\delta}{2}} e^{-\frac{t}{2}} \sum_{k=0}^{\infty} \frac{(t\delta)^{k}}{(2k)!} \int_{\mathcal{O}_p} \gamma_{11}^{2k}\, d\nu_p(\gamma) \\
&\overset{(3)}{=} \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{\delta}{2}} e^{-\frac{t}{2}} \sum_{k=0}^{\infty} \frac{(t\delta)^{k}}{(2k)!} \mathrm{E}\left[\mathrm{Beta}\left(\tfrac{1}{2},\ \tfrac{p-1}{2}\right)\right]^k \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{\delta}{2}} e^{-\frac{t}{2}} \sum_{k=0}^{\infty} \frac{(t\delta)^{k}}{(2k)!} \frac{\Gamma\left(\frac{1}{2}+k\right)\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p}{2}+k\right)\Gamma\left(\frac{1}{2}\right)}.
\end{aligned}$$

(1) This follows from the left and right invariance of the Haar measure ν_p.

(2) By the invariance of ν_p the distribution of γ_{11} is even, i.e., $\gamma_{11} \sim -\gamma_{11}$, so its odd moments vanish.

(3) By left invariance, the first column of γ is uniformly distributed on the unit sphere in $\mathcal{R}^p$, hence (*) $\gamma_{11}^2 \sim \mathrm{Beta}\left(\frac{1}{2},\ \frac{p-1}{2}\right)$.

Thus from (7.29) and Legendre's duplication formula, t has pdf w. r. to dt given by

$$h(t)\frac{d\tilde{\mu}(t)}{dt} = \frac{1}{2^{\frac{p}{2}}} e^{-\frac{\delta}{2}} e^{-\frac{t}{2}} \sum_{k=0}^{\infty} \frac{t^{\frac{p}{2}-1}(t\delta)^k}{(2k)!} \frac{\Gamma\left(\frac{1}{2}+k\right)}{\Gamma\left(\frac{p}{2}+k\right)\Gamma\left(\frac{1}{2}\right)}$$

$$(7.30) \qquad = \underbrace{e^{-\frac{\delta}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^k}{k!}}_{\text{Poisson}\left(\frac{\delta}{2}\right)\ \text{weights}} \underbrace{\left[\frac{t^{\frac{p+2k}{2}-1} e^{-\frac{t}{2}}}{2^{\frac{p+2k}{2}}\Gamma\left(\frac{p+2k}{2}\right)}\right]}_{\text{pdf of } \chi^2_{p+2k}},$$

as also found in (2.27). □

Exercise 7.8. Derive (*). □

Example 7.9. Extend Example 7.7 as follows. Let

$$\mathcal{X} = \mathcal{R}^{p\times r}, \quad G = \mathcal{O}_p \times \mathcal{O}_r, \quad \mu = \text{Lebesgue measure on } \mathcal{R}^{p\times r},$$

so μ is $(\mathcal{O}_p \times \mathcal{O}_r)$-invariant. Here $(\gamma, \psi) \in \mathcal{O}_p \times \mathcal{O}_r$ acts on $\mathcal{R}^{p\times r}$ via

$$x \mapsto \gamma x \psi'.$$

Assume first that $r \geq p$. A maximal invariant statistic is [verify!]

$$t(X) = (\, l_1(XX') \geq \cdots \geq l_p(XX')\,) \equiv l(XX'), \tag{7.31}$$

the ordered nontrivial eigenvalues of XX' [verify]. If X has pdf $f(x)$ w. r. to μ then by Stein's integral representation (7.26), $l \equiv l(XX')$ has pdf

$$h(l) = \int_{\mathcal{O}_p}\int_{\mathcal{O}_r} f(\gamma x \psi')\, d\nu_p(\gamma)\, d\nu_r(\psi) \tag{7.32}$$

w. r. to $d\tilde{\mu}(l)$ on $\mathcal{R}_l$. If $f(x)$ is also $(\mathcal{O}_p \times \mathcal{O}_r)$-invariant, i.e., if

$$f(x) = k(\, l(xx')\,)$$

for some $k(\cdot)$ on $\mathcal{R}_l$, then the pdf of $l(XX')$ w. r. to $d\tilde{\mu}(l)$ is simply

$$h(l) = k(l), \qquad l \in \mathcal{R}_l. \tag{7.33}$$

The induced measure $d\tilde{\mu}(l)$ can be found by considering a special case:

$$X \sim N_{p\times r}(0,\, I_p \otimes I_r) \implies XX' \sim W_p(r,\, I_p).$$

Here

$$\begin{aligned} f(x) &= \tfrac{1}{(2\pi)^{\frac{pr}{2}}}\, e^{-\frac{1}{2}\operatorname{tr} xx'} \\ &= \tfrac{1}{(2\pi)^{\frac{pr}{2}}}\, e^{-\frac{1}{2}\sum_{i=1}^p l_i} \equiv k(l) \qquad \text{w.r. to } d\mu(x) \text{ on } \mathcal{R}^{p\times r}, \end{aligned}$$

so l has pdf

$$h(l) = k(l) = \tfrac{1}{(2\pi)^{\frac{pr}{2}}}\, e^{-\frac{1}{2}\sum_{i=1}^p l_i} \qquad \text{w.r. to } d\tilde{\mu}(l) \text{ on } \mathcal{R}_l.$$

We also know from (7.16) that l has the pdf

$$w(l) \equiv \frac{\pi^{\frac{p}{2}}}{2^{\frac{pr}{2}}\Gamma_p\left(\frac{p}{2}\right)\Gamma_p\left(\frac{r}{2}\right)} \prod_{i=1}^{p} l_i^{\frac{r-p-1}{2}} e^{-\frac{1}{2}l_i} \prod_{1\le i<j\le p}(l_i - l_j)$$

w.r. to dl ($\equiv$ Lebesgue measure) on $\mathcal{R}_l$. Therefore $d\tilde{\mu}(l)$ is determined as follows:

$$(7.34)\qquad d\tilde{\mu}(l) = \frac{w(l)}{k(l)}\, dl = \frac{\pi^{\frac{p(r+1)}{2}}}{\Gamma_p\left(\frac{p}{2}\right)\Gamma_p\left(\frac{r}{2}\right)} \prod_{i=1}^{p} l_i^{\frac{r-p-1}{2}} \prod_{1\le i<j\le p}(l_i - l_j)\, dl.$$

Finally, the case $r < p$ follows from (7.33) by interchanging p and r, since XX' and $X'X$ have the same nontrivial eigenvalues.

Application: Stein's representation (7.32) provides an integral representation for the pdf of the eigenvalues of a noncentral Wishart matrix. If

$$X \sim N_{p\times r}(\xi,\, I_p \otimes I_r)$$

with $\xi \neq 0$, the distribution of XX' depends on ξ only through $\xi\xi'$ [verify], hence is designated the *noncentral Wishart distribution* $W_p(r,\, I_p;\, \xi\xi')$.

Assume first that $r \ge p$. The distribution of the ordered eigenvalues

$$l \equiv l(XX') \equiv (\, l_1(XX') \ge \cdots \ge l_p(XX')\,)$$

of S depends on $\xi\xi'$ only through the ordered eigenvalues

$$\lambda \equiv \lambda(\xi\xi') \equiv (\, \lambda_1(\xi\xi') \ge \cdots \ge \lambda_p(\xi\xi')\,)$$

of $\xi\xi'$, hence is designated by $l(p,\, r;\, \lambda)$. Here

$$f(x) = \tfrac{1}{(2\pi)^{\frac{pr}{2}}}\, e^{-\frac{1}{2}\mathrm{tr}\,(x-\xi)(x-\xi)'},$$

so by (7.32), l has pdf $h(l)$ w. r. to $d\tilde{\mu}(l)$ given by

$h(l)$

$$= \frac{1}{(2\pi)^{\frac{pr}{2}}} \int_{\mathcal{O}_p} \int_{\mathcal{O}_r} e^{-\frac{1}{2}\mathrm{tr}\,(\gamma x\psi' - \xi)(\gamma x\psi' - \xi)'}\, d\nu_p(\gamma)\, d\nu_r(\psi)$$

$$= \frac{1}{(2\pi)^{\frac{pr}{2}}} e^{-\frac{1}{2}\mathrm{tr}\,\xi\xi'} e^{-\frac{1}{2}\mathrm{tr}\,xx'} \int_{\mathcal{O}_p} \int_{\mathcal{O}_r} e^{\mathrm{tr}\,\gamma x\psi'\xi'}\, d\nu_p(\gamma)\, d\nu_r(\psi)$$

$$\overset{(1)}{=} \frac{1}{(2\pi)^{\frac{pr}{2}}} e^{-\frac{1}{2}\sum\lambda_i} e^{-\frac{1}{2}\sum l_i} \int_{\mathcal{O}_p} \int_{\mathcal{O}_r} e^{\mathrm{tr}\,\gamma D_l^{\frac{1}{2}}\tilde{\psi}' D_\lambda^{\frac{1}{2}}}\, d\nu_p(\gamma)\, d\nu_r(\psi)$$

$$= \frac{1}{(2\pi)^{\frac{pr}{2}}} e^{-\frac{1}{2}\sum\lambda_i} e^{-\frac{1}{2}\sum l_i} \int_{\mathcal{O}_p} \int_{\mathcal{O}_r} e^{\sum_{i=1}^p \sum_{j=1}^p l_i^{\frac{1}{2}} \lambda_j^{\frac{1}{2}} \gamma_{ji}\psi_{ji}} d\nu_p(\gamma)\, d\nu_r(\psi)$$

$$\overset{(2)}{=} \frac{1}{(2\pi)^{\frac{pr}{2}}} e^{-\frac{1}{2}\sum\lambda_i} e^{-\frac{1}{2}\sum l_i} \int_{\mathcal{O}_p} \int_{\mathcal{O}_r} \prod_{i=1}^p \left[\sum_{k=0}^\infty \frac{1}{(2k)!} l_i^k \left(\sum_{j=1}^p \lambda_j^{\frac{1}{2}} \gamma_{ji}\psi_{ji} \right)^{2k} \right] d\nu_p(\gamma)\, d\nu_r(\psi).$$

(1) Here $D_\lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$, $D_l = \mathrm{diag}(l_1, \dots, l_p)$, and $\tilde{\psi}$ is the leading $p \times p$ submatrix of ψ. The equality follows from the left and right invariance of the Haar measures ν_p and ν_r and from the singular value decompositions of ξ and x. The representation (1) is due to A. James *Ann. Math. Statist.* (1961, 1964). Note that the double integral in (1) is a convex and symmetric ($\equiv$ permutation-invariant) function of $l_1^{\frac{1}{2}}, \dots, l_p^{\frac{1}{2}}$ on the *unordered* positive orthant $\mathcal{R}_+^p$ [explain].

(2) By the invariance of ν_p the distribution of $\gamma_i \equiv (\gamma_{1i}, \dots, \gamma_{pi})'$, the ith column of γ, is even, i.e., $\gamma_i \sim -\gamma_i$. Apply this for $i = 1, \dots, p$, using the following expansion at each step:

$$\frac{1}{2}(e^x + e^{-x}) = \sum_{k=0}^\infty \frac{x^{2k}}{(2k)!}.$$

Thus from (7.34), l has pdf $f_\lambda(l)$ w. r. to dl given by

$$f_\lambda(l) = h(l) \frac{d\tilde{\mu}(l)}{dl}$$

$$= \frac{\pi^{\frac{p}{2}} e^{-\frac{1}{2}\sum\lambda_i}}{2^{\frac{pr}{2}} \Gamma_p\left(\frac{p}{2}\right) \Gamma_p\left(\frac{r}{2}\right)} \prod_{i=1}^p l_i^{\frac{r-p-1}{2}} e^{-\frac{1}{2}l_i} \prod_{1 \le i < j \le p} (l_i - l_j)$$

$$(7.35) \qquad \cdot \int_{\mathcal{O}_p} \int_{\mathcal{O}_r} \prod_{i=1}^p \left[\sum_{k=0}^\infty \frac{1}{(2k)!} l_i^k \left(\sum_{j=1}^p \lambda_j^{\frac{1}{2}} \gamma_{ji}\psi_{ji} \right)^{2k} \right] d\nu_p(\gamma)\, d\nu_r(\psi)$$

on the range $\mathcal{R}_l$. The case $r < p$ now follows by interchanging p and r in (7.35), since XX' and $X'X$ have the same nontrivial eigenvalues. $\square$

Remark 7.10. The integrand in (7.35) is a multiple power series in $\{l_i\}$, and similarly in $\{\lambda_j\}$ – this can be expanded and integrated term-by-term, leading to an extension of the Poisson mixture representation (7.30) for the noncentral chi-square pdf. However, important information already can be obtained from the integral representation (7.35). By comparing the noncentral pdf $f_\lambda(l)$ in (7.35) to the central pdf $f_0(l) = \phi^*(l)$ in (7.22), the likelihood ratio is

$$\text{(7.36)} \qquad \frac{f_\lambda(l)}{f_0(l)} = e^{-\frac{1}{2}\sum \lambda_i} \int_{\mathcal{O}_p} \int_{\mathcal{O}_r} [\cdots],$$

the double integral given in (7.35). From this representation it is immediate that $f_\lambda(l)/f_0(l)$ is strictly increasing in each l_i, hence in each $l_i^{\frac{1}{2}}$, and as already noted in (1) its extension to the positive orthant $\mathcal{R}_+^p$ is convex and symmetric in the $\{l_i^{\frac{1}{2}}\}$. Thus the symmetric extension to $\mathcal{R}_+^p$ of the acceptance region $A \subseteq \mathcal{R}_l$ of any proper Bayes test for testing $\lambda = 0$ vs. $\lambda > 0$ based on l must be convex and decreasing in $\{l_i^{\frac{1}{2}}\}$ [explain and verify!].

Wald's fundamental theorem of decision theory states that the closure in the weak* topology of the set of all proper Bayes acceptance regions determines an essentially complete class of tests. Because convexity and monotonicity are preserved under weak* limits, this implies that the symmetric extension to $\mathcal{R}_+^p$ of *any* admissible acceptance region $A \subseteq \mathcal{R}_l$ must be convex and decreasing in $\{l_i^{\frac{1}{2}}\}$. This shows, for example, that the test which *rejects* $\lambda = 0$ for large values of the *minimum* eigenvalue $l_p(XX')$ is inadmissible among invariant tests [verify!], hence among all tests.

Furthermore, Perlman and Olkin (*Ann. Statist.* (1980) pp.1326-41) used the monotonicity of the likelihood ratio (7.36) and the FKG inequality to establish the unbiasedness of *all monotone invariant tests,* i.e., all tests with acceptance regions of the form $\{g(l_1, \ldots, g_p) \le c\}$ with g nondecreasing in each l_i. $\square$

Exercise 7.10. Eigenvalues of $S \sim W_p(r, \Sigma)$ when $\Sigma \neq I_p$.

(a) Assume that $r \geq p$ and $\Sigma > 0$. Show that the pdf of $l \equiv (l_1, \cdots, l_p)$ is

$$
\begin{aligned}
f_\lambda(l) = {} & \frac{\pi^{\frac{p}{2}}}{2^{\frac{pr}{2}} \Gamma_p\left(\frac{p}{2}\right) \Gamma_p\left(\frac{r}{2}\right) \left(\prod_{i=1}^p \lambda_i\right)^{\frac{r}{2}}} \prod_{i=1}^p l_i^{\frac{r-p-1}{2}} \prod_{1 \leq i < j \leq p} (l_i - l_j) \\
& \cdot \int_{\mathcal{O}_p} e^{-\frac{1}{2} \operatorname{tr} D_\lambda^{-1} \gamma D_l \gamma'} \, d\nu_p(\gamma), \qquad l \in \mathcal{R}_l, \qquad (7.37)
\end{aligned}
$$

where $l \equiv (l_1, \cdots, l_p)$ and $\lambda \equiv (\lambda_1, \ldots, \lambda_p)$ are the ordered eigenvalues of S and Σ, respectively. (Note that (7.37) reduces to (7.22) when $\Sigma = I$.) □

(b) Consider the problem of testing $\Sigma = I_p$ vs. $\Sigma \geq I_p$. Show that a necessary condition for the admissibility of an invariant test is that the symmetric extension to $\mathcal{R}_+^p$ of its acceptance region $A \subseteq \mathcal{R}_l$ must be convex and decreasing in $\{l_i\}$. (Thus the test based on $l_p(S)$ is inadmissible.)) □

Remark 7.12. [Expand:] Stein's integral formula (7.26) for the pdf of a maximal invariant statistic under the action of a compact topological group G can be partially extended to the case where G is locally compact. Important examples include the general linear group GL and the triangular groups GT and GU, which occur in the MANOVA testing problem (cf. §8.2). In this case, however, the integral representation does not provide the normalizing constant for the pdf of the MIS, but still provides a useful expression for the likelihood ratio, e.g. (7.36). References include:

S. A. Andersson (1982). Distributions of maximal invariants using quotient measures. *Ann. Statist.* **10** 955-961.

M. L. Eaton (1989). *Group Invariance Applications in Statistics.* Regional Conference Series in Probability and Statistics Vol. 1, Institute of Mathematical Statistics.

R. A. Wijsman (1990). *Invariant Measures on Groups and their Use in Statistics.* Lecture Notes – Monograph Series Vol. 14, Institute of Mathematical Statistics.

8. The MANOVA Model and Testing Problem. (Lehmann *TSH* Ch.8.)

8.1. Characterization of a MANOVA subspace.

In Section 3.3 the *multivariate linear model* was defined as follows: $X_1, \dots, X_m$ (note m not n) are independent $p \times 1$ vector observations having common unknown pd covariance matrix Σ. Let $X_j \equiv (X_{1j}, \dots, X_{pj})'$, $j = 1, \dots, m$. We assume that each of the p variates satisfies the *same* univariate linear model, that is,

$$\text{(8.1)} \qquad \mathrm{E}(X_{i1}, \dots, X_{im}) = \beta_i Z, \qquad i = 1, \dots, p,$$

where $Z : l \times m$ is the design matrix, $\mathrm{rank}(Z) = l \leq m$, and $\beta_i : 1 \times l$ is a vector of unknown regression coefficients. Equivalently, (8.1) can be expressed geometrically as

$$\text{(8.2)} \quad \mathrm{E}(X_{i1}, \dots, X_{im}) \in L(Z) \equiv \text{row space of } Z \subseteq \mathcal{R}^m, \qquad i = 1, \dots, p.$$

In matrix form, (8.1) and 8.2) can be written as

$$\text{(8.3)} \qquad \mathrm{E}(\tilde{X}) \in \{\beta Z \mid \beta \in \mathcal{M}(p,\, l)\} =: L_p(Z),$$

where $\mathcal{M}(p,\, l)$ denotes the vector space of all real $p \times l$ matrices,

$$\tilde{X} \equiv (X_1, \dots, X_m) \in \mathcal{M}(p,\, m),$$
$$\beta \equiv \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}.$$

Note that $L_p(Z)$ is a linear subspace of $\mathcal{M}(p, m)$ with

$$\dim(L_p(Z)) = p \cdot \dim(L(Z)) = p\, l,$$

a multiple of p. Then (8.2) can be expressed equivalently as[20]

$$\text{(8.4)} \qquad \mathrm{E}(\tilde{X}) \in \oplus_{i=1}^{p} L(Z) \equiv \left\{ \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix} \;\middle|\; v_1, \dots, v_p \in L(Z) \right\}.$$

[20] Note: (8.3) and (8.4) can also be written as $E(\tilde{X}) \in \mathcal{R}^p \otimes L(Z)$.

The forms (8.1)–(8.4) are all *extrinsic*, in that they require specification of the design matrix Z, which in turn is specified only after a choice of coordinate system. We seek to express these equivalent forms in an *intrinsic algebraic form* that will allow us to determine when a specified linear subspace $L \subseteq \mathcal{M}(p,m)$ can be written as $L_p(Z)$ for some Z. This is accomplished by means of an invariant ≡ coordinate-free definition of a *MANOVA subspace.*

Definition 8.1. A linear subspace $L \subseteq \mathcal{M}(p,m)$ is called a *MANOVA subspace* if

$$\mathcal{M}(p,p)\, L \subseteq L. \tag{8.5}$$

Because $\mathcal{M}(p,p)$ is in fact a matrix *algebra* (i.e., closed under matrix multiplication as well as matrix addition) that contains the identity matrix I_p, (8.5) is equivalent to the condition $\mathcal{M}(p,p)\, L = L$. □

Proposition 8.2. *Suppose that L is a linear subspace of $\mathcal{M}(p,m)$. The following are equivalent:*

(a) L is a MANOVA subspace.

(b) $L = L_p(Z)$ for some $Z \in \mathcal{M}(l,m)$ of rank $l \leq m$ (so dim$(L) = p\,l$).

(c) There exists a unique $m \times m$ projection matrix P such that

$$L = \{x \in \mathcal{M}(p,m) \mid x = xP\}. \tag{8.7}$$

Note: if $L = L_p(Z)$ then $P = Z'(ZZ')^{-1}Z$ and $l = \mathrm{tr}(P)$ [verify].

(d) There exists an orthogonal matrix $\Gamma : m \times m$ such that

$$L\Gamma = \left\{(\mu,\, 0_{p\times(m-l)}) \;\middle|\; \mu \in \mathcal{M}(p,l)\right\} \quad (l \leq m). \tag{8.6}$$

Γ is obtained from the spectral decomposition $P = \Gamma\,\mathrm{diag}(I_l,\, 0_{m-l})\,\Gamma'$.

Note: (8.6) is the **canonical form of a MANOVA subspace**.

Proof. The equivalence of (b), (c), and (d) is proved exactly as for univariate linear models [reference - Seber?] It is straightforward to show that (b) ⇒ (a). We now show that (a) ⇒ (c).

Let $\epsilon_i \equiv (0, \ldots, 0, 1, 0, \ldots, 0)'$ denote the i-th coordinate vector in $\mathcal{R}^p \equiv \mathcal{M}(1,p)$ and define $L_i := \epsilon_i' L \subseteq \mathcal{M}(1,n)$, $i = 1, \ldots, p$. Then for every pair i, j, it follows from (a) that

$$L_j = \epsilon_j' L = \epsilon_i' \Pi_{ij} L \subseteq \epsilon_i' L = L_i,$$

where $\Pi_{ij} \in \mathcal{M}(p,p)$ is the i, j-permutation matrix, so

$$L_1 = \cdots = L_p =: \tilde{L} \subseteq \mathcal{M}(1,m).$$

Let $P : m \times m$ be the unique projection matrix onto $\tilde{L}$. Then for $x \in L$,

$$\begin{aligned} xP &= I_p xP = \Big(\sum\nolimits_{i=1}^p \epsilon_i \epsilon_i'\Big) xP \\ &= \Big(\sum\nolimits_{i=1}^p \epsilon_i \epsilon_i' xP\Big) = \Big(\sum\nolimits_{i=1}^p \epsilon_i \epsilon_i' x\Big) = \Big(\sum\nolimits_{i=1}^p \epsilon_i \epsilon_i'\Big) x = x, \end{aligned}$$

where the fourth equality holds since $\epsilon_i' x \in L_i \equiv \tilde{L}$, hence

$$L \subseteq \{x \in \mathcal{M}(p,m) \mid x = xP\}.$$

Conversely, for $x \in \mathcal{M}(p,m)$,

$$\begin{aligned} xP = x &\implies \epsilon_i' xP = \epsilon_i' x, \qquad\qquad i = 1, \ldots, p, \\ &\implies \epsilon_i' x \in \tilde{L} \equiv L_i \\ &\implies \epsilon_i' x = \epsilon_i' x_i \quad \text{for some } x_i \in L \\ &\implies x \equiv \Big(\sum\nolimits_{i=1}^p \epsilon_i \epsilon_i'\Big) x = \sum\nolimits_{i=1}^p (\epsilon_i \epsilon_i') x_i \in L, \end{aligned}$$

where the final membership follows from (a) and the assumption that L is a linear subspace. Thus

$$L \supseteq \{x \in \mathcal{M}(p,m) \mid x = xP\},$$

which completes the proof. $\bar{\square}$

Remark 8.3. In the statistical literature, multivariate linear models often occur in the form

$$L_p(Z,C) := \{\beta Z \mid \beta \in \mathcal{M}(p,l),\ \beta C = 0\}, \tag{8.8}$$

where $C : l \times s$ (with $\text{rank}(C) = s \le l$) determines s linear constraints on β. To see that $L_p(Z,C)$ is in fact a MANOVA subspace and thus can be re-expressed in the form $L_p(Z_0)$ for some design matrix Z_0, by Proposition 8.2 it suffices to verify that

$$\mathcal{M}(p,p)\, L_p(Z,C) \subseteq L_p(Z,C),$$

which is immediately evident. □

8.2. Reduction of a MANOVA testing problem to canonical form.

A *normal MANOVA model* is simply a normal multivariate linear model (3.14), i.e., one observes

$$\tilde{X} \equiv (X_1, \dots, X_m) \sim N_{p\times m}(\eta,\, \Sigma \otimes I_m) \quad \text{with} \quad \eta \in L \subseteq \mathcal{R}^{p\times m}, \tag{8.9}$$

where L is a MANOVA subspace of $\mathcal{R}^{p\times m}$ and $\Sigma > 0$ is unknown.

The *MANOVA testing problem* is that of testing

$$\eta \in L_0 \quad \text{vs.} \quad \eta \in L \qquad \text{based on } \tilde{X}, \tag{8.10}$$

for two MANOVA subspaces $L_0 \subset L \subset \mathcal{R}^{p\times m}$ with

$$\dim(L_0) \equiv p\, l_0 < p\, l \equiv \dim(L).$$ □

Proposition 8.4. *(extension of Proposition 8.2(d)). Let $r = l - l_0$, $n = m - l$. There exists an $m \times m$ orthogonal matrix Γ^* such that*

$$\begin{aligned} L\,\Gamma^* &= \{(\xi,\, \mu,\, 0_{p\times n}) \mid \xi \in \mathcal{M}(p,\, l_0),\ \mu \in \mathcal{M}(p,\, r)\,\}, \\ L_0\,\Gamma^* &= \{(\xi,\, 0_{p\times r},\, 0_{p\times n}) \mid \xi \in \mathcal{M}(p,\, l_0)\,\}. \end{aligned} \tag{8.11}$$

Proof. Again this is proved exactly as for univariate linear subspaces: From (8.6), choose $\Gamma : n \times n$ orthogonal such that

$$L\Gamma = \{(\xi,\, \mu,\, 0_{p\times n}) \mid \xi \in \mathcal{M}(p,\, l_0),\ \mu \in \mathcal{M}(p,\, r)\}.$$

By (8.5), $L_0\Gamma\begin{pmatrix} I_l \\ 0_{n\times l}\end{pmatrix}$ is a MANOVA subspace of $\mathcal{R}^{pl}$, so we can find $\Gamma_0 : l \times l$ orthogonal so that

$$L_0\Gamma\begin{pmatrix} I_l \\ 0_{n\times l}\end{pmatrix}\Gamma_0 = \{(\xi,\, 0_{p\times r}) \mid \xi \in \mathcal{M}(p,\, l_0)\}.$$

Now take $\Gamma^* = \Gamma\begin{pmatrix} \Gamma_0 & 0_{l\times n} \\ 0_{n\times l} & I_n\end{pmatrix}$ and verify that (8.11) holds. □

Now set $X^* = \tilde{X}\Gamma^* \equiv (U,\, X,\, Y) : p \times (l_0 + r + n)$. From (8.11) the MANOVA testing problem (8.10) is transformed to the following:

$$\begin{aligned} &\textit{based on } (U,\, X,\, Y) \sim N_{p\times(l_0+r+n)}\big((\xi,\, \mu,\, 0_{p\times n}),\; \Sigma\otimes I\big), \\ &\textit{test } \mu = 0 \;\; \textit{vs.} \;\; \mu \neq 0. \end{aligned} \tag{8.12}$$

Here $\xi \in \mathcal{M}(p,\, l_0)$, $\mu \in \mathcal{M}(p,\, r)$, and $\Sigma \in \mathcal{S}_p^+$. This testing problem is invariant under $G^* := \mathcal{M}(p,\, l_0)$ acting as a translation group on U (and ξ):

$$\begin{aligned} (U,\, X,\, Y) &\mapsto (U + b,\, Y,\, X), \\ (\xi,\, \mu,\, \Sigma) &\mapsto (\xi + b,\, \mu,\, \Sigma). \end{aligned} \tag{8.13}$$

Since $\mathcal{M}(p,\, l_0)$ acts transitively on itself, the MIS and MIP are (Y, X) and (μ, Σ), resp., and the invariance-reduced problem becomes the following:

$$\begin{aligned} &\textit{based on } (X, Y) \sim N_{p\times(r+n)}\big((\mu,\, 0_{p\times n}),\, \Sigma\otimes I\big), \\ &\textit{test } \mu = 0 \;\; \textit{vs.} \;\; \mu \neq 0. \end{aligned} \tag{8.14}$$

For this problem, $(X,\, S) \equiv (X,\, YY')$ is a sufficient statistic [verify], so (8.14) is reduced by sufficiency to the canonical MANOVA testing problem (6.24). As in Example 6.36, (6.24) is reduced by invariance under (6.25) to the testing problem (6.28) based on the nontrivial eigenvalues of $X'S^{-1}X$.

(The condition $n \geq p$, needed for the existence of the MLE $\hat{\Sigma}$ in (6.24) and (8.14), is equivalent to $m \geq l + p$ in (8.9) and (8.10).)

[Add: Wijsman's integral representation for the likelihood ratio; Schwartz.]

Remark 8.5. By Proposition 8.2(b) and Remark 8.3, $L_p(Z)$ and $L_p(Z,C)$ are MANOVA subspaces of $\mathcal{R}^{p\times m}$ such that $L_p(Z,C) \subset L_p(Z)$. Thus the general MANOVA testing problem (8.10) is often stated as that of testing

$$\eta \in L_p(Z,C) \quad \text{vs.} \quad \eta \in L_p(Z). \tag{8.15}$$

[Add Examples] □̄

Exercise 8.6. Derive the LRT for (8.15).

Hint: The LRT already has been derived for the canonical MANOVA testing problem in Exercise 6.37a. Now express the LRT statistic in terms of the observation matrix Y, the design matrix Z, and the constraint matrix C. □̄

8.3. Related topics.

8.3.1. Seemingly unrelated regressions (SUR).

If the p variates follow *different* univariate linear models, i.e., if (8.1) is extended to

$$\mathrm{E}(X_{i1},\dots,X_{im}) = \beta_i Z_i \in L(Z_i), \qquad i = 1,\dots,p, \tag{8.16}$$

where $Z_1 : l_1 \times m$, ..., $Z_p : l_p \times m$ are design matrices with *different* row spaces, then the model (8.16) is called a *seemingly unrelated regression (SUR) model*. The p univariate models are only "seemingly" unrelated because they are correlated if Σ is not diagonal. Under the assumption of normality, explicit likelihood inference (i.e., MLEs and LRTs) is not possible unless the row spaces $L(Z_1),\dots,L(Z_p)$ are nested. (But see Remark 8.9.)□̄

8.3.2. Invariant formulation of block-triangular matrices.

The invariant algebraic definition of a MANOVA subspace in Definition 8.1 suggests an invariant algebraic definition of generalized block-triangular matrices. First, for any increasing sequence of integers

$$0 \equiv p_0 < p_1 < p_2 < \cdots < p_r < p_{r+1} \equiv p \quad (1 < r < p)$$

define the sequence

$$\{0\} \subset V_1 \subset V_2 \subset \cdots \subset V_r \subset \mathcal{R}^p \tag{8.17}$$

of proper linear subspaces of $\mathcal{R}^p$ as follows:

$$V_i = \text{span}\{\epsilon_1, \epsilon_2, \ldots, \epsilon_{p_i}\}, \qquad i = 1, \ldots, r. \tag{8.18}$$

Consider a partitioned matrix

$$A \equiv (A_{ij} \mid 1 \le i, j \le r) \in \mathcal{M}(p, p), \tag{8.19}$$

where $A_{ij} \in \mathcal{M}(p_i - p_{i-1}, p_j - p_{j-1})$. Then A is upper block triangular, i.e., $A_{ij} = 0$ for $1 \le j < i \le r$, if and only if [verify!]

$$AV_i \subseteq V_i, \quad i = 1, \ldots, r$$

Thus the set of $\mathcal{A}$ of upper block-triangular matrices can be defined in the following algebraic way:

$$\mathcal{A} \equiv \mathcal{A}(p_1, \ldots, p_r) := \{A \in \mathcal{M}(p, p) \mid AV_i \subseteq V_i,\, i = 1, \ldots, r\}. \tag{8.20}$$

Exercise 8.7. Give an algebraic definition of the set of *lower* block triangular matrices. $\bar{\square}$

More generally, let

$$\{0\} \subset V_1 \subset V_2 \subset \cdots \subset V_r \subset \mathcal{R}^p \tag{8.21}$$

be a general increasing sequence of proper linear subspaces of $\mathcal{R}^p$ and define

$$\mathcal{A} \equiv \mathcal{A}(V_1, \ldots, V_r) := \{A \in \mathcal{M}(p, p) \mid AV_i \subseteq V_i,\, i = 1, \ldots, p\}. \tag{8.22}$$

Note that this is a completely invariant $\equiv$ coordinate-free algebraic definition, and immediately implies that A is a matrix *algebra*, i.e., is closed under matrix addition and multiplication [verify], and $I_p \in \mathcal{A}$. The algebra A is called the algebra of *block-triangular matrices with respect to* $V_1, \ldots, V_r$. The proper subset $\mathcal{A}^* \subset \mathcal{A}$ consisting of all nonsingular matrices in $\mathcal{A}$ is a matrix *group*, i.e., it contains the identity matrix and is closed under matrix inversion [verify]. Finally, it is readily seen that $\mathcal{A}(V_1, \ldots, V_r)$ is isomorphic to $\mathcal{A}(p_1, \ldots, p_r)$ under a similarity transformation, where $p_i := \dim(V_i)$. $\bar{\square}$

Remark 8.8. Suppose that $V_1, \ldots, V_r$ is an *arbitrary (i.e., non-nested)* finite collection of proper linear supspaces of $\mathcal{R}^p$. Define $\mathcal{A} \equiv \mathcal{A}(V_1, \ldots, V_r)$ as in (8.22). Then $\mathcal{A}$ is a *generalized block-triangular matrix algebra* [verify!] and $\mathcal{A}^*$ is a *generalized block-triangular matrix group.* Note too that

$$\mathcal{A}(V_1, \ldots, V_r) = \mathcal{A}(\mathcal{L}(V_1, \ldots, V_r)), \tag{8.23}$$

where $\mathcal{L}(V_1, \ldots, V_r)$ is the *lattice of linear subspaces generated from* $(V_1, \ldots, V_r)$ by all possible finite unions and intersections. □

Remark 8.9. The algebra $\mathcal{A} \equiv \mathcal{A}(\mathcal{L}(V_1, \ldots, V_r))$ plays an important role in the theory of normal *lattice conditional independence (LCI) models* (Andersson and Perlman (1993) *Annals of Statistics*). A subspace $L \subseteq \mathcal{M}(p, n)$ is called an $\mathcal{A}$-*subspace* if $\mathcal{A}L \subseteq L$. It is shown by A&P (*IMS Lecture Notes Vol. 24,* 1994) that if the linear model subspace L of a normal multivariate linear model is an $\mathcal{A}$-subspace and if the covariance structure satisfies a corresponding set of LCI constraints, then the MLE and LRT statistics can be obtained explicitly. This was extended to ADG covariance models by A&P (*J. Multivariate Analysis* 1998), and to SUR models and non-nested missing data models with conforming LCI covariance structure by Drton, Andersson, and Perlman (*J. Multivariate Analysis* 2006). (See §9.1-9.2.) □

8.3.3. The GMANOVA model and testing problem.

(Recall Example 6.39.) [To be completed] □

9. Testing and Estimation with Missing/Incomplete Data.

Let $X_1, \ldots, X_m$ be an i.i.d. random sample from $N_p(\mu, \Sigma)$ with μ and Σ unknown. Partition X_k, μ, and Σ as

$$X_k = \begin{matrix} p_1 \\ p_2 \end{matrix} \begin{pmatrix} X_{1k} \\ X_{2k} \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{matrix} \\ p_1 \\ p_2 \end{matrix} \begin{matrix} p_1 \quad\; p_2 \\ \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{matrix}.$$

Consider n additional i.i.d. observations $V_1, \ldots, V_n$ from $N_{p_2}(\mu_2, \Sigma_{22})$, independent of $X_1, \ldots, X_m$. Here $V_1, \ldots, V_n$ can be viewed as *incomplete* observations from the original distribution $N_p(\mu, \Sigma)$. We shall find the MLEs $\hat{\mu}, \hat{\Sigma}$ based on $X_1, \ldots, X_m, V_1, \ldots, V_n$.

Because

$$\begin{aligned} X_{1k} \mid X_{2k} &\sim N_{p_1}(\alpha + \beta X_{2k}, \Sigma_{11\cdot 2}), \\ \beta &= \Sigma_{12}\Sigma_{22}^{-1}, \\ \alpha &= \mu_1 - \beta\mu_2, \end{aligned}$$

the likelihood function (LF $\equiv$ joint pdf of $X_1, \ldots, X_m$, $V_1, \ldots, V_n$) can be written in the form

$$\begin{aligned} (9.1)\ &\prod_{k=1}^{m} f^{(1)}_{\alpha,\beta,\Sigma_{11\cdot 2}}(x_{1k} \mid x_{2k}) \prod_{k=1}^{m} f^{(2)}_{\mu_2,\Sigma_{22}}(x_{2k}) \prod_{k=1}^{n} f^{(2)}_{\mu_2,\Sigma_{22}}(v_k) \\ &= c \cdot |\Sigma_{11\cdot 2}|^{-m/2} \exp\big(-\tfrac{1}{2} \mathrm{tr}\, \Sigma_{11\cdot 2}^{-1} \textstyle\sum_{k=1}^{m} (x_{1k} - \alpha - \beta x_{2k})^{*2} \big) \\ &\quad \cdot |\Sigma_{22}|^{-(m+n)/2} \exp\big(-\tfrac{1}{2} \mathrm{tr}\, \Sigma_{22}^{-1} \big[\sum_{k=1}^{m} (x_{2k} - \mu_2)^{*2} + \sum_{k=1}^{n} (v_k - \mu_2)^{*2} \big] \big), \end{aligned}$$

where $(x)^{*2} := xx'$ and the parameters $\alpha, \beta, \Sigma_{11\cdot 2}, \mu_2, \Sigma_{22}$ vary independently over their respective ranges. Thus we see that the LF is the product of two LFs, the first that of a normal multivariate linear regression model

$$N_{p_1 m}\left((\alpha, \beta) \begin{pmatrix} e' \\ Z \end{pmatrix}, \Sigma_{11\cdot 2} \right)$$

with $e' = (1, \ldots, 1) : 1 \times m$ and $Z = (X_{21}, \ldots, X_{2m})$, and the second that of $m + n$ i.i.d. observations from $N_p(\mu_2, \Sigma_{22})$.

The MLEs for these models are given in (3.15), (3.16), (3.34), and (3.35). To assure the existence of the MLE, the *single* condition $m \geq p+1$ is necessary and sufficient [verify!]. (This is the same condition required for existence of the MLE based on the complete observations $X_1, \ldots, X_m$ only.) If this condition holds, then the MLEs of α, β, $\Sigma_{11\cdot 2}$, μ_2, Σ_{22} are as follows:

$$
\begin{aligned}
&\hat{\alpha} = \bar{X}_1 - \hat{\beta}\bar{X}_2, && \hat{\mu}_2 = \frac{m\bar{X}_2 + n\bar{V}}{m+n}, \\
(9.2)\qquad &\hat{\beta} = S_{12}S_{22}^{-1}, && \hat{\Sigma}_{22} = \tfrac{1}{m+n}\Big[S_{22} + W + \tfrac{mn}{m+n}(\bar{X}_2 - \bar{V})^{*2}\Big], \\
&\hat{\Sigma}_{11\cdot 2} = \tfrac{1}{m}S_{11\cdot 2},
\end{aligned}
$$

[verify!], where

$$S = \sum\nolimits_{k=1}^{m}(X_k - \bar{X})^{*2}, \qquad W = \sum\nolimits_{k=1}^{n}(V_k - \bar{V})^{*2}.$$

Verify that $\frac{m+n}{m+n-1}\hat{\Sigma}_{22}$ is the sample covariance matrix based on the *combined* sample $X_{21}, \ldots, X_{2m}, V_1, \ldots, V_n$. Furthermore, the maximum value of the LF is given by

$$(9.3)\qquad c \cdot |\hat{\Sigma}_{11\cdot 2}|^{-m/2}|\hat{\Sigma}_{22}|^{-(m+n)/2}\exp\big(-\tfrac{1}{2}(mp + np_2)\big).$$

Remark 9.1. The pairs $(\bar{X}, S)$ and $(\bar{V}, W)$ together form a complete and sufficient statistic for the above incomplete data model. □

Remark 9.2. This analysis can be extended to the case of a *monotone* $\equiv$ *nested* incomplete data model. The observed data consists of independent observations of the forms

$$(9.4)\qquad \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{pmatrix}, \quad \begin{pmatrix} \\ X_2 \\ \vdots \\ X_r \end{pmatrix}, \quad \ldots, \quad \begin{pmatrix} \\ \\ \\ X_r \end{pmatrix},$$

where a complete observation $X \sim N_p(\mu, \Sigma)$. The MLEs are obtained by factoring the joint pdf of $X_1, \ldots, X_r$ as

$$(9.5)\quad f(x_1,\ldots,x_r) = f(x_1|x_2,\ldots,x_r)f(x_2|x_3,\ldots,x_r)\cdots f(x_{r-1}|x_r)f(x_r)$$

and noting that each conditional pdf is the LF of a normal linear regression model. □

Exercise 9.4. Find the LRTs based on $X_1,\ldots,X_m,V_1,\ldots,V_n$ for testing problems (i) and (ii) below. Argue that no explicit expression is available for the LRT statistic in (iii). (Eaton and Kariya (1983) *Ann. Statist.*)

(i) $H_1 : \mu_2 = 0$ vs. $H : \mu_2 \neq 0$ (μ_1 and Σ unspecified).

(ii) $H_2 : \mu_1 = 0,\ \mu_2 = 0$ vs. $H : \mu_1 \neq 0,\ \mu_2 \neq 0$ (Σ unspecified).

(iii) $H_3 : \mu_1 = 0$ vs. $H : \mu_1 \neq 0$ (μ_2 and Σ unspecified).

Partial solutions: First, for each testing problem, the LF is given by (9.1) and its maximum under H given by (9.3).

(i) Because $\alpha = \mu_1$ when $\mu_2 = 0$, it follows from (9.1) that the LF under H_1 is given by

$$(9.6)\quad \begin{aligned} &c \cdot|\Sigma_{11\cdot 2}|^{-m/2} \exp\Big(- \tfrac{1}{2}\mathrm{tr}\, \Sigma_{11\cdot 2}^{-1} \sum_{k=1}^{m}(x_{1k} - \mu_1 - \beta x_{2k})^{*2}\Big) \\ &\cdot |\Sigma_{22}|^{-(m+n)/2} \exp\Big(- \tfrac{1}{2}\mathrm{tr}\, \Sigma_{22}^{-1}\Big[\sum_{k=1}^{m}(x_{2k})^{*2} + \sum_{k=1}^{n}(v_k)^{*2}\Big]\Big), \end{aligned}$$

Thus the maximum of the LF under H_1 is given by

$$(9.7)\quad c \cdot |\hat{\Sigma}_{11\cdot 2}|^{-m/2}|\tilde{\Sigma}_{22}|^{-(m+n)/2} \exp\big(-\tfrac{1}{2}(mp + np_2)\big),$$

where

$$\tilde{\Sigma}_{22} = \tfrac{1}{m+n}\Big[\sum_{k=1}^{m}(X_{2k})^{*2} + \sum_{k=1}^{n}(V_k)^{*2}\Big]$$

[verify!]. Thus, by (9.3) and (9.7) the LRT rejects H_2 in favor of H for large values of [verify!]

$$\begin{aligned}\frac{|\tilde{\Sigma}_{22}|}{|\hat{\Sigma}_{22}|} &= \frac{\left|\hat{\Sigma}_{22} + \left(\frac{m\bar{X}_2+n\bar{V}}{m+n}\right)^{*2}\right|}{|\hat{\Sigma}_{22}|} \\ &= 1 + \left(\frac{m\bar{X}_2+n\bar{V}}{m+n}\right)' \hat{\Sigma}_{22}^{-1} \left(\frac{m\bar{X}_2+n\bar{V}}{m+n}\right) \\ &\equiv 1 + T_2^2.\end{aligned}$$

Note that T_2^2 is exactly the T^2 statistic for testing $\mu_2 = 0$ vs. $\mu_2 \neq 0$ based on the combined sample $X_{21}, \ldots, X_{2m}, V_1, \ldots, V_n$, so the LRT ignores the observations $X_{11}, \ldots, X_{1m}$.

(ii) The LRT statistic is the product of the LRT statistics for problem (i) and for the problem of testing $\mu_1 = 0,\ \mu_2 = 0$ vs. $\mu_1 \neq 0,\ \mu_2 = 0$ (see Exercise 6.14). Both LRTs can be obtained explicitly, but the distribution of their product is not simple. (See Eaton and Kariya (1983).)

(iii) Under $H_3 : \mu_1 = 0,\ \mu_2$ appears in different forms in the two exponentials on the right-hand side of (9.1), hence maximization over μ_2 cannot be done explicitly. $\bar{\square}$

Exercise 9.5. For simplicity, assume μ is known, say $\mu = 0$. Find the LRT based on $X_1, \ldots, X_m,\ V_1, \ldots, V_n$ for testing

$$H_0 : \Sigma_{12} = 0 \quad \text{vs.} \quad H : \Sigma_{12} \neq 0 \ \ (\Sigma_{11} \text{ and } \Sigma_{22} \text{ unspecified}).$$

Solution: The LRT statistic for this problem is the same as if the additional observations $V_1, \ldots, V_n$ were not present (cf. Exercise 6.24), namely $\frac{|S_{11}||S_{22}|}{|S|}$. This can be seen by examining the LF factorization in (9.1) when $\mu = 0$ (so $\alpha = 0$ and $\mu_2 = 0$). The null hypothesis $H_0 : \Sigma_{12} = 0$ is equivalent to $\beta = 0$, so the second exponential on the right-hand side of (9.1) is the same under H_0 and H, hence has the same maximum value under H_0 and H. Thus this second factor cancels when forming the LRT statistic, hence the LRT does not involve $V_1, \ldots, V_n$. $\bar{\square}$

9.1. Lattice conditional independence (LCI) models for non-monotone missing/incomplete data.

If the incomplete data pattern is *non-monotone* $\equiv$ *non-nested*, then no explicit expressions exist for the MLEs. Instead, an iterative procedure such as the EM algorithm must be used to compute the MLEs. (Caution: convergence to the MLE is not always guaranteed, and the choice of starting point may affect the convergence properties.)

An example of a non-monotone incomplete data pattern is

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad \begin{pmatrix} X_1 \\ \\ X_3 \end{pmatrix}, \quad \begin{pmatrix} \\ X_2 \\ X_3 \end{pmatrix}, \quad \begin{pmatrix} \\ \\ X_3 \end{pmatrix}. \tag{9.8}$$

Here no compatible factorization of the joint pdf such as (9.5) is possible. However, Rubin (*Multiple Imputation*, 1987) and Andersson and Perlman (*Statist. Prob. Letters,* 1991) have pointed out that a compatible factorization is possible if a parsimonious set of *lattice conditional independence* (LCI) restrictions determined by the incomplete data pattern is imposed on the (unknown) covariance matrix Σ. In the present example, these restrictions reduce to the single condition $X_1 \perp\!\!\!\perp X_2 \mid X_3$, in which case the joint pdf of X_1, X_2, X_3 factors as

$$f(x_1, x_2, x_3) = f(x_1|x_3) f(x_2|x_3) f(x_3). \tag{9.9}$$

Here again each conditional pdf is the LF of a normal linear regression model, so the MLEs of the corresponding regression parameters can be obtained explicitly.

Of course, the LCI restriction may not be defensible, but it can be tested. If it is rejected, at least the MLEs obtained under the LCI restriction may serve as a reasonable starting value for the EM algorithm. (See L. Wu and M. D. Perlman (2000) *Communications in Statistics - Simulation and Computation* **29** 481-509.)

[Add notes on LCI models.]

9.2. LCI models for seemingly unrelated regressions (SUR).

[Add notes on SUR models.]

10. Group-symmetry Covariance Models.

[Hand out notes.]

11. James-Stein Shrinkage Estimators.

[Hand out notes.]

12. Unbiasedness of Multivariate Tests.

[Hand out notes.]

13. Admissibility of Multivariate Tests.

[Hand out notes.]

Appendix A. Monotone Likelihood Ratio and Total Positivity.

In Section 6 we study multivariate hypothesis testing problems which remain invariant under a group of symmetry transformations. In order to respect these symmetries, we shall restrict consideration to test functions that possess the same invariance properties and seek a uniformly most powerful invariant (UMPI) test. Under multivariate normality, the distribution of a UMPI test statistic is often a noncentral chi-square or related noncentral distribution. To verify the UMPI property it is necessary to establish that the noncentral distribution has *monotone likelihood ratio (MLR)* with respect to the noncentrality parameter. For this we will rely on the relation between the MLR property and total positivity of order 2.

Definition A.1. Let $f(x,y) \geq 0$ be defined on $A \times B$, a Cartesian product of intervals in $\mathcal{R}^1$. We say that f is **totally positive of order 2 (TP2)** if

$$\begin{vmatrix} f(x_1,y_1) & f(x_1,y_2) \\ f(x_2,y_1) & f(x_2,y_2) \end{vmatrix} \geq 0 \qquad \forall\, x_1 < x_2,\ y_1 < y_2,$$

i.e., *if*

$$f(x_1,y_1)f(x_2,y_2) \geq f(x_1,y_2)f(x_2,y_1). \tag{A.1}$$

If $f > 0$ on $A \times B$ then (5.1) is equivalent to the following condition:

$$\frac{f(x_2,y)}{f(x_1,y)} \text{ is nondecreasing in } y \ \ \forall\, x_1 < x_2. \tag{A.2}$$

Note that $f(x,y)$ is TP2 on $A \times B$ iff $g(y,x) \equiv f(x,y)$ is TP2 on $B \times A$. □

Fact A.2. *If f and g are TP2 on $A \times B$ then $f \cdot g$ is TP2 on $A \times B$. In particular, $a(x)b(y)f(x,y)$ is TP2 for any $a(\cdot) \geq 0$ and $b(\cdot) \geq 0$.* □

Fact A.3. *If f is TP2 on $A' \times B'$ and $\phi : A \to A'$ and $\psi : B \to B'$ are both increasing or both decreasing, then $f(\phi(x), \psi(y))$ is TP2 on $A \times B$.* □

Fact A.4. *If $f(x,y) > 0$ and $\frac{\partial^2 \log f}{\partial x \partial y} \geq 0$ on $A \times B$ then f is TP2.* □

Fact A.5. *If $f(x, y) = g(x - y)$ and $g : \mathcal{R}^1 \to [0, \infty)$ is log-concave, then f is TP2 on $\mathcal{R}^2$.*

Proof. Let $h(x) = \log g(x)$. For $x_1 < x_2,\ y_1 < y_2$ set

$$\begin{aligned} s &= x_1 - y_1, \quad u = x_1 - y_2, \\ t &= x_2 - y_2, \quad v = x_2 - y_1. \end{aligned}$$

Then [verify]

$$\begin{gathered} u \le \min(s, t) \le \max(s, t) \le v, \\ s + t = u + v, \end{gathered}$$

so, since h is concave,

$$h(s) + h(t) \ge h(u) + h(v),$$

which is equivalent to the TP2 condition $(A.1)$ for $f(x, y) \equiv g(x - y)$. □

These Facts yield the following examples of TP2 functions $f(x, y)$:

Example A.6. *Exponential kernel:* $f(x, y) = e^{xy}$ is TP2 on $\mathcal{R}^1 \times \mathcal{R}^1$.

Example A.7. *Exponential family:* $f(x, y) = a(x)b(y)e^{\phi(x)\psi(y)}$ is TP2 on $A \times B$ if $a(\cdot) \ge 0$ on A, $b(\cdot) \ge 0$ on B, $\phi(\cdot)$ is increasing on A, and $\psi(\cdot)$ is increasing on B. In particular, $f(x, y) = x^y$ is TP2 on $(0, \infty) \times \mathcal{R}^1$.

Example A.8. *Order kernel:* $f(x, y) = (x - y)_+^\alpha$ and $f(x, y) = (x - y)_-^\alpha$ are TP2 on $\mathcal{R}^1 \times \mathcal{R}^1$ for $\alpha \ge 0$. [$I_{(0,\infty)}$ and $I_{(-\infty,0)}$ are log concave on $\mathcal{R}^1$.]

The following is a celebrated result in the theory of total positivity.

Proposition A.9. Composition Lemma ≡ Karlin's Lemma *(due to Polya and Szego). If $g(x, y)$ is TP2 on $A \times B$ and $h(x, y)$ is TP2 on $B \times C$, then for any σ-finite measure μ,*

$$(A.3) \qquad f(x, z) := \int_B g(x, y)h(y, z)d\mu(y)$$

is TP2 on $A \times C$.

Proof. For $x_1 \le x_2$ and $z_1 \le z_2$,

$$\begin{aligned}
&f(x_1, z_1)f(x_2, z_2) - f(x_1, z_2)f(x_2, z_1) \\
&= \iint g(x_1, y)g(x_2, u)[h(y, z_1)h(u, z_2) - h(y, z_2)h(u, z_1)]d\mu(y)d\mu(u) \\
&= \iint_{\{y<u\}} + \iint_{\{y>u\}} + \underbrace{\iint_{\{y=u\}}}_{=0}.
\end{aligned}$$

By interchanging the dummy variables y and u, however, we see that

$$\begin{aligned}
&\iint_{\{y>u\}} g(x_1, y)g(x_2, u)[h(y, z_1)h(u, z_2) - h(y, z_2)h(u, z_1)]d\mu(y)d\mu(u) \\
=&\iint_{\{y<u\}} g(x_1, u)g(x_2, y)[h(u, z_1)h(y, z_2) - h(u, z_2)h(y, z_1)]d\mu(y)d\mu(u)
\end{aligned}$$

so

$$\begin{aligned}
&\iint_{\{y<u\}} + \iint_{\{y>u\}} \\
=&\iint_{\{y<u\}} [g(x_1, y)g(x_2, u) - g(x_1, u)g(x_2, y)] \\
&\qquad \cdot [h(y, z_1)h(u, z_2) - h(y, z_2)h(u, z_1)]d\mu(y)d\mu(u) \ge 0
\end{aligned}$$

since g and h are TP2. Thus h is TP2. $\square$

Example A.10. *Power series:* $f(x, y) = \sum_{k=0}^{\infty} c_k x^k y^k$ is TP2 on $(0, \infty) \times (0, \infty)$ if $c_k \ge 0 \ \forall k$.

Proof. Apply the Composition Lemma with $g(x, k) = x^k$, $h(k, y) = y^k$, and μ the measure that assigns mass c_k to $k = 0, 1, \ldots$. $\square$

Definition A.11. Let $\{f(x|\lambda) \mid \lambda \in \Lambda\}$ be a 1-parameter family of pdfs (discrete or continuous) for a real random vector X with range $\mathcal{X}$, where Λ is an interval in $\mathcal{R}^1$. We say that $f(x|\lambda)$ has *monotone likelihood ratio (MLR)* in $T(x)$ for a real-valued statistic $T(x) \in \mathcal{T}$ (an interval in $\mathcal{R}^1$) if $f(x|\lambda) = \phi(T(x)|\lambda)$ where $\phi(t|\lambda)$ is TP2 on $\mathcal{T} \times \Lambda$. $\square$

Proposition A.12. MLR preserves monotonicity. *If $f(x|\lambda)$ has MLR in $T(X)$ and $g(t)$ is nondecreasing in t, then*

$$\mathrm{E}_\lambda[g(T(X))] \equiv \int_{\mathcal{X}} g(T(x))\, f(x|\lambda) d\nu(x)$$

is nondecreasing in λ (ν is either counting measure or Lebesgue measure).

Proof. Set $h(\lambda) = \mathrm{E}_\lambda[g(T(X))]$. Then for any $\lambda_1 \le \lambda_2$ in Λ,

$$\begin{aligned} &h(\lambda_2) - h(\lambda_1) \\ &= \int g(T(x))[\phi(T(x)|\lambda_2) - \phi(T(x)|\lambda_1)]d\nu(x) \\ &= \tfrac{1}{2} \iint [g(T(x)) - g(T(y))]\, [\phi(T(x)|\lambda_2)\phi(T(y)|\lambda_1) - \phi(T(y)|\lambda_2)\phi(T(x)|\lambda_1)] \\ &\qquad\qquad d\nu(x)d\nu(y) \ge 0, \end{aligned}$$

since the two $[\cdots]$ terms are both ≥ 0 if $T(x) \ge T(y)$ or both ≤ 0 if $T(x) \le T(y)$. □

Remark A.13. If $\{f(x|\lambda)\}$ has MLR in $T(x)$, then for each $t \in \mathcal{T}$,

$$\mathrm{Pr}_\lambda[T(X) > t] \equiv \mathrm{E}_\lambda\left[I_{(t,\infty)}T(X)\right]$$

is nondecreasing in λ, hence $T(X)$ is stochastically increasing in λ. □

Example A.14. The noncentral chi-square distribution $\chi^2_n(\delta)$ has MLR w.r.to δ.

From (2.27), a noncentral chi-square rv $\chi^2_n(\delta)$ with n df and noncentrality parameter δ is a Poisson($\delta/2$)-mixture of central chi-square rvs:

$$\chi^2_n(\delta) \;\big|\; K = k \;\sim\; \chi^2_{n+2k}, \qquad K \sim \mathrm{Poisson}(\delta/2). \tag{A.4}$$

Thus if $f_n(x|\delta)$ and $f_n(x)$ denote the pdfs of $\chi^2_n(\delta)$ and χ^2_n, then

$$\begin{aligned} f_n(x|\delta) &= \sum_{k=0}^{\infty} f_{n+2k}(x)\ \mathrm{Pr}[K=k] \\ &= \sum_{k=0}^{\infty} \left[\frac{x^{\frac{n}{2}+k-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}+1}\Gamma\left(\frac{n}{2}+k\right)}\right] \cdot \left[\frac{e^{-\frac{\delta}{2}}\left(\frac{\delta}{2}\right)^k}{k!}\right] \\ &= 2^{-\frac{n}{2}-1}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} \cdot e^{-\frac{\delta}{2}} \cdot \sum_{k=0}^{\infty} \frac{x^k\delta^k}{\Gamma\left(\frac{n}{2}+k\right)}. \end{aligned} \tag{A.5}$$

Thus by A.2, A.3, and A.10, $f_n(x|\delta)$ is TP2 in (x, δ). $\bar{\square}$

Example A.15. The noncentral F distribution $F_{m,n}(\delta)$ has MLR w.r.to δ. Let

$$(A.6) \qquad F_{m,n}(\delta) \stackrel{\text{distn}}{=} \frac{\chi^2_m(\delta)}{\chi^2_n},$$

the ratio of two independent chi-square rvs with $\chi^2_m(\delta)$ noncentral and χ^2_n central. From (A.4), $F_{m,n}(\delta)$ can be represented as a Poisson mixture of central F distributions:

$$(A.7) \qquad F_{m,n}(\delta) \,\big|\, K = k \;\sim\; F_{m+2k,n}, \qquad K \sim \text{Poisson}\,(\delta/2)\,,$$

so if $f_{m,n}(x|\delta)$ and $f_{m,n}(x)$ now denote the pdfs of $F_{m,n}(\delta)$ and $F_{m,n}$, then

$$\begin{aligned} f_{m,n}(x|\delta) &= \sum_{k=0}^{\infty} f_{m+2k,n}(x)\ \Pr[K = k] \\ &= \sum_{k=0}^{\infty} \left[\frac{\Gamma\left(\frac{m+n}{2} + k\right)}{\Gamma\left(\frac{m}{2} + k\right)\Gamma\left(\frac{n}{2}\right)} \cdot \frac{x^{\frac{m}{2}+k-1}}{(x+1)^{\frac{m+n}{2}+k-1}} \right] \cdot \left[\frac{e^{-\frac{\delta}{2}}\left(\frac{\delta}{2}\right)^k}{k!} \right] \\ (A.8) \qquad &= \frac{x^{\frac{m}{2}-1}}{\Gamma\left(\frac{n}{2}\right)(x+1)^{\frac{m+n}{2}-1}} \cdot e^{-\frac{\delta}{2}} \cdot \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{m+n}{2} + k\right)}{\Gamma\left(\frac{m}{2} + k\right) k!\, 2^k} \left(\frac{x}{x+1}\right)^k \delta^k, \end{aligned}$$

Thus by A.2 and A.10, $f_{m,n}(x|\delta)$ is TP2 in (x, δ). $\bar{\square}$

Question A.16. Does $\chi^2_n(\delta)$ have MLR w.r.to n? (δ fixed) Does $F_{m,n}(\delta)$ have MLR w.r.to n? (m, δ fixed). w.r.to m? (n, δ fixed) $\bar{\square}$

Proposition A.17. Scale mixture of a TP2 kernel. *Let $g(x, y)$ be TP2 on $\mathcal{R}^1 \times (0, \infty)$ and let h be a nonnegative function on $(0, \infty)$ such that $h\,(y/\zeta)$ is TP2 for $(y, \zeta) \in (0, \infty) \times (0, \infty)$. Then*

$$(A.9) \qquad f(x, \zeta) := \int_0^{\infty} g(x, \zeta z) h(z) dz$$

is TP2 on $\mathcal{R}^1 \times (0, \infty)$.

Proof. Set $y = \zeta z$, so

$$f(x,\zeta) = \frac{1}{\zeta}\int_0^\infty g(x,y)\, h\left(\frac{y}{\zeta}\right) dy,$$

hence the result follows from the Composition Lemma. □

Example A.18. The distribution of the multiple correlation coefficient R^2 has MLR w.r. to ρ^2.

Let R^2, ρ^2, U, ζ, and Z be as defined in Example 3.21 (also see Example 6.26 and Exercise 6.27). From (3.68),

$$\begin{aligned} U \,\big|\, Z &\sim F_{p-1,\,n-p+1}(\zeta Z), \\ Z &\sim \chi^2_n, \end{aligned} \tag{A.10}$$

so the unconditional pdf of u with parameter ζ is given by

$$f(u|\zeta) = \int_0^\infty f_{p-1,\,n-p+1}(u|\,\zeta z)\, f_n(z) dz$$

where $f_{p-1,\,n-p+1}(\cdot|\,\zeta z)$ and $f_n(\cdot)$ are the pdfs for $F_{p-1,\,n-p+1}(\zeta z)$ and χ^2_n, respectively. Then $f_{p-1,\,n-p+1}(u|y)$ is TP2 in (u,y) by Example A.15, while

$$f_n\left(\tfrac{y}{\zeta}\right) = c\cdot\left(\tfrac{y}{\zeta}\right)^{\frac{n}{2}-1} e^{-\frac{y}{2\zeta}}$$

is TP2 in (y,ζ) by Example A.7, so $f(u|\zeta)$ is TP2 in (u,ζ) by Proposition A.17. Finally, because U and ζ are increasing functions of R^2 and ρ^2, respectively, it follows by Fact A.3 that the distribution of R^2 has MLR w.r.to ρ^2. □

Exercise A.19. Prove Fact A.4.

Appendix B. Some Useful Results from Matrix Theory.

[See handout]