Generalizations of the Wishart Distributions Arising From Monotone Incomplete Multivariate Normal Data

Wan-Ying Chang (Washington Dept. of Fish and Wildlife)
D.R. (Penn State University and SAMSI)
Background

We have a population of “patients”

We draw a random sample of $N$ patients, and measure $m$ variables on each patient:

1. Visual acuity
2. LDL
3. Systolic blood pressure
4. Glucose intolerance
5. Insulin response to oral glucose
6. Actual weight ÷ Expected weight
... ... $m$
$m$ Red blood cell count
We obtain data:

<table>
<thead>
<tr>
<th>Patient</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>\cdots</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>\begin{pmatrix} v_{1,1} \ v_{1,2} \ \vdots \ v_{1,m} \end{pmatrix}</td>
<td>\begin{pmatrix} v_{2,1} \ v_{2,2} \ \vdots \ v_{2,m} \end{pmatrix}</td>
<td>\begin{pmatrix} v_{3,1} \ v_{3,2} \ \vdots \ v_{3,m} \end{pmatrix}</td>
<td>\cdots</td>
<td>\begin{pmatrix} v_{N,1} \ v_{N,2} \ \vdots \ v_{N,m} \end{pmatrix}</td>
</tr>
</tbody>
</table>

Vector notation for the data: \( V_1, V_2, \ldots, V_N \)

\( V_1 \): The \( m \) measurements on patient 1, stacked into a column
Classical multivariate analysis

Statistical analysis of data consisting of $N$ vectors, each containing $m$ entries

Common assumption: The population has a multivariate normal distribution

$V$: The vector of measurements on a randomly chosen patient

Multivariate normal populations are characterized by:

$\mu$: The population mean vector
$\Sigma$: The population covariance matrix
For a given data set, $\mu$ and $\Sigma$ are unknown

We wish to perform inference about $\mu$ and $\Sigma$

Construct confidence regions for, and test hypotheses about, $\mu$ and $\Sigma$


Johnson and Wichern (2002). *Applied Multivariate Statistical Analysis*

Muirhead (1982). *Aspects of Multivariate Statistical Theory*
Standard notation: \( V \sim N_p(\mu, \Sigma) \)

The usual formula for the density function of \( V \)

\[
f(v) = (2\pi)^{-m/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(v - \mu)'\Sigma^{-1}(v - \mu)\right), \quad v \in \mathbb{R}^m
\]

\( V_1, V_2, \ldots, V_N \): Measurements on \( N \) randomly chosen patients

Estimate \( \mu \) and \( \Sigma \) using Fisher’s maximum likelihood principle

Likelihood function: \( L(\mu, \Sigma) = \prod_{j=1}^{N} f(v_j) \)

Maximize the likelihood function \( L \) w.r.t. \( \mu \) and \( \Sigma \)

Maximum likelihood estimator: The value of \( \mu, \Sigma \) which maximizes \( L \)
\[ \hat{\mu} = \frac{1}{N} \sum_{j=1}^{N} V_j: \text{ The sample mean} \]

\[ \hat{\Sigma} = \frac{1}{N-1} \sum_{j=1}^{n} (V_j - \bar{V})(V_j - \bar{V})': \text{ The sample covariance matrix} \]

What are the probability distributions of \( \hat{\mu} \) and \( \hat{\Sigma} \)?

\( \hat{\mu} \sim N_p(\mu, \frac{1}{N} \Sigma) \)

As \( N \to \infty, \frac{1}{N} \Sigma \to 0 \), so \( \hat{\mu} \to \mu \)

\( \hat{\Sigma} \) has a “Wishart” distribution, a generalization of the \( \chi^2 \)
Monotone incomplete data

Some patients were not measured completely (or the dog ate some of the data, or a measuring machine broke down)

The resulting data set, with ∗ denoting missing data

\[
\begin{pmatrix}
\mathbf{v}_{1,1} \\
\mathbf{v}_{1,2} \\
\mathbf{v}_{1,3} \\
\vdots \\
\mathbf{v}_{1,m}
\end{pmatrix}
\begin{pmatrix}
\mathbf{v}_{2,1} \\
\mathbf{v}_{2,2} \\
\mathbf{v}_{2,3} \\
\vdots \\
\mathbf{v}_{2,m}
\end{pmatrix}
\begin{pmatrix}
\ast \\
\ast \\
\mathbf{v}_{3,2} \\
\vdots \\
\mathbf{v}_{3,m}
\end{pmatrix}
\ldots
\begin{pmatrix}
\ast \\
\ast \\
\ast \\
\vdots \\
\mathbf{v}_{N,2} \\
\vdots \\
\mathbf{v}_{N,m}
\end{pmatrix}
\]

Monotone data: Each ∗ is followed by ∗’s all the way to the end

We may need to renumber patients to see if data are monotone
Physical Fitness Data

Patients: Men taking a physical fitness course at NCSU

Three variables were measured

Oxygen intake rate (ml per kg body weight per minute)

RunTime (time taken, in minutes, to run 1.5 miles)

RunPulse (heart rate while running)
<table>
<thead>
<tr>
<th>Oxygen</th>
<th>RunTime</th>
<th>RunPulse</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.609</td>
<td>11.37</td>
<td>178</td>
</tr>
<tr>
<td>45.313</td>
<td>10.07</td>
<td>185</td>
</tr>
<tr>
<td>54.297</td>
<td>8.65</td>
<td>156</td>
</tr>
<tr>
<td>51.855</td>
<td>10.33</td>
<td>166</td>
</tr>
<tr>
<td>49.156</td>
<td>8.95</td>
<td>180</td>
</tr>
<tr>
<td>40.836</td>
<td>10.95</td>
<td>168</td>
</tr>
<tr>
<td>44.811</td>
<td>11.63</td>
<td>176</td>
</tr>
<tr>
<td>45.681</td>
<td>11.95</td>
<td>176</td>
</tr>
<tr>
<td>39.203</td>
<td>12.88</td>
<td>168</td>
</tr>
<tr>
<td>45.790</td>
<td>10.47</td>
<td>186</td>
</tr>
<tr>
<td>50.545</td>
<td>9.93</td>
<td>148</td>
</tr>
<tr>
<td>48.673</td>
<td>9.40</td>
<td>186</td>
</tr>
<tr>
<td>47.920</td>
<td>11.50</td>
<td>170</td>
</tr>
<tr>
<td>47.467</td>
<td>10.50</td>
<td>170</td>
</tr>
<tr>
<td>50.388</td>
<td>10.08</td>
<td>168</td>
</tr>
<tr>
<td>47.273</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Monotone data have a staircase pattern; we will consider the two-step pattern

Partition $V$ into an incomplete part of dimension $p$ and a complete part of dimension $q$

\[
\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} , \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}, \begin{pmatrix} * \\ Y_{n+1} \end{pmatrix}, \begin{pmatrix} * \\ Y_{n+2} \end{pmatrix}, \ldots, \begin{pmatrix} * \\ Y_N \end{pmatrix}
\]

Assume that the individual vectors are independent and are drawn from $N_m(\mu, \Sigma)$

Goal: Carry out maximum likelihood inference for $\mu$ and $\Sigma$

Obtain formulas as explicit as in the classical context
Where do monotone incomplete data arise?

Panel survey data (Census Bureau, Bureau of Labor Statistics)
Astronomy
Early detection of diseases
Wildlife survey research
Covert communications
Mental health research
Climate and atmospheric studies

Sometimes we are given \( n \) independent observations on \( \left( \frac{X}{Y} \right) \) in addition to \( N - n \) independent observations on \( Y \)
Difficulty: The likelihood function is more complicated

$$L = \prod_{i=1}^{n} f_{X,Y}(x_i, y_i) \cdot \prod_{i=n+1}^{N} f_Y(y_i)$$

$$= \prod_{i=1}^{n} f_Y(y_i) f_{X|Y}(x_i) \cdot \prod_{i=n+1}^{N} f_Y(y_i)$$

$$= \prod_{i=1}^{N} f_Y(y_i) \cdot \prod_{i=1}^{n} f_{X|Y}(x_i)$$
Partition $\mu$ and $\Sigma$ similarly:

$$
\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}
$$

Define

$$
\mu_{1.2} = \mu_1 - \Sigma_{12} \Sigma_{22}^{-1} \mu_2
$$

$$
\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}
$$

$$
Y \sim N_q(\mu_2, \Sigma_{22}), \quad X|Y \sim N_p(\mu_{1.2}, \Sigma_{11.2})
$$

Wilks, Anderson, Morrison, Olkin, Jinadasa, D. Tracy, … : Their hard work produced $\hat{\mu}$ and $\hat{\Sigma}$
Sample means:

\[
\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j, \quad \bar{Y}_1 = \frac{1}{n} \sum_{j=1}^{n} Y_j.
\]

\[
\bar{Y}_2 = \frac{1}{N-n} \sum_{j=n+1}^{N} Y_j, \quad \bar{Y} = \frac{1}{N} \sum_{j=1}^{N} Y_j.
\]

Sample covariance matrices:

\[
A_{11} = \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})', \quad A_{12} = \sum_{j=1}^{n} (X_j - \bar{X})(Y_j - \bar{Y}_1)'
\]

\[
A_{22,n} = \sum_{j=1}^{n} (Y_j - \bar{Y}_1)(Y_j - \bar{Y}_1)', \quad A_{22,N} = \sum_{j=1}^{N} (Y_j - \bar{Y})(Y_j - \bar{Y})'
\]
The MLEs of \( \mu \) and \( \Sigma \)

Notation: \( \tau = n/N, \quad \bar{\tau} = 1 - \tau \)

\[
\hat{\mu}_1 = \bar{X} - \bar{\tau} A_{12} A_{22,n}^{-1} (\bar{Y}_1 - \bar{Y}_2), \quad \hat{\mu}_2 = \bar{Y}
\]

\( \hat{\mu}_1 \) is called the \textit{regression estimator} of \( \mu_1 \)

In sample surveys, extra observations on a subset of variables are used to improve estimation of a parameter

\( \hat{\Sigma} \) is more complicated:

\[
\hat{\Sigma}_{11} = \frac{1}{n} (A_{11} - A_{12} A_{22,n}^{-1} A_{21}) + \frac{1}{N} A_{12} A_{22,n}^{-1} A_{22,N} A_{22,n}^{-1} A_{21}
\]

\[
\hat{\Sigma}_{12} = \frac{1}{N} A_{12} A_{22,n}^{-1} A_{22,N}
\]

\[
\hat{\Sigma}_{22} = \frac{1}{N} A_{22,N}
\]
Problems which have been unsolved for up to seventy years:

Find the *exact* distributions of $\hat{\mu}$ and $\hat{\Sigma}$

Explicit confidence levels for elliptical confidence regions for $\mu$

In testing hypotheses on $\mu$ or $\Sigma$, determine whether the likelihood ratio test statistics are unbiased

Describe the moments (means, variances, correlations) of the components of $\hat{\mu}$

How does $\hat{\mu}$ behave as $n$ and/or $N \to \infty$?

The crucial obstacle: The distribution of $\hat{\mu}$ for fixed $n, N$
The exact distribution of $\hat{\mu}$

For $n > p + q$,

$$\hat{\mu} \overset{\mathcal{L}}{=} \mu + V_1 + (n^{-1} - N^{-1}) \sqrt{Q_1/Q_2} \begin{pmatrix} V_2 \\ 0 \end{pmatrix}$$

with $V_1 \sim N_{p+q}(0, \Omega)$, $V_2 \sim N_p(0, \Sigma_{11.2})$, $Q_1 \sim \chi^2_q$, $Q_2 \sim \chi^2_{n-q}$,

$$\Omega = N^{-1}\Sigma + (n^{-1} - N^{-1}) \begin{pmatrix} \Sigma_{11.2} & 0 \\ 0 & 0 \end{pmatrix}$$

and $V_1$, $V_2$, $Q_1$, and $Q_2$ are independent

$n = N$: $\hat{\mu} \equiv$ the sample mean; it is $N_m(\mu, N^{-1}\Sigma)$-dist’d.

If $\Sigma_{12} = 0$ then $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent
\( \hat{\mu} \) is unbiased: \( E(\hat{\mu}) = \mu \)

The covariance matrix of \( \hat{\mu} \)

\[
\text{Cov}(\hat{\mu}) = \frac{1}{N} \Sigma + \frac{(n - 2)\bar{r}}{n(n - q - 2)} \begin{pmatrix} \Sigma_{11.2} & 0 \\ 0 & 0 \end{pmatrix}
\]

Higher moments of \( \hat{\mu} \) now are straightforward, however ...

Because of the term \( 1/Q_2 \) in the distribution, no even moment of order \( n - q \) or higher is finite.
The asymptotic distribution of $\hat{\mu}$

Let $n, N \to \infty$ with $n/N \to \delta$, where $0 < \delta \leq 1$. Then

$$\sqrt{N} (\hat{\mu} - \mu) \xrightarrow{\mathcal{L}} N_{p+q} \left( 0, \Sigma + (\delta^{-1} - 1) \begin{pmatrix} \Sigma_{11} & 2 \\ 0 & 0 \end{pmatrix} \right)$$

Many other asymptotic results can be obtained from the exact distribution of $\hat{\mu}$

If $n$ and $N \to \infty$ with $n/N \to 0$ then $\hat{\mu}_2 \to \mu_2$, almost surely, and the distribution of $\hat{\mu}_1$ is . . .
The analog of Hotelling’s statistic:

\[
T^2 = (\hat{\mu} - \mu)' \hat{\text{Cov}}(\hat{\mu})^{-1} (\hat{\mu} - \mu)
\]

where

\[
\hat{\text{Cov}}(\hat{\mu}) = \frac{1}{N} \sum + \frac{(n - 2)\bar{\tau}}{n(n - q - 2)} \begin{pmatrix}
\hat{\Sigma}_{11.2} & 0 \\
0 & 0
\end{pmatrix}
\]

An obvious ellipsoidal confidence region for \(\mu\) is

\[
\left\{ \nu \in \mathbb{R}^{p+q} : (\hat{\mu} - \nu)' \hat{\text{Cov}}(\hat{\mu})^{-1} (\hat{\mu} - \nu) \leq c \right\}
\]

What is the level of confidence associated with this region?
Reminder

\[ \bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j, \quad \bar{Y}_1 = \frac{1}{n} \sum_{j=1}^{n} Y_j \]

\[ \bar{Y}_2 = \frac{1}{N - n} \sum_{j=n+1}^{N} Y_j, \quad \bar{Y} = \frac{1}{N} \sum_{j=1}^{N} Y_j \]

\[ A_{11} = \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})', \quad A_{12} = \sum_{j=1}^{n} (X_j - \bar{X})(Y_j - \bar{Y}_1)' \]

\[ A_{22,n} = \sum_{j=1}^{n} (Y_j - \bar{Y}_1)(Y_j - \bar{Y}_1)', \quad A_{22,N} = \sum_{j=1}^{N} (Y_j - \bar{Y})(Y_j - \bar{Y})' \]
\( \tau = \frac{n}{N}, \quad \bar{\tau} = 1 - \tau \)

\[
\hat{\Sigma}_{11} = \frac{1}{n} (A_{11} - A_{12}A_{22,n}^{-1}A_{21}) + \frac{1}{N} A_{12,n} A_{22}^{-1}A_{22,n}A_{21}^{-1}A_{12,n}A_{22}^{-1}
\]

\[
\hat{\Sigma}_{12} = \frac{1}{N} A_{12,n} A_{22,n} A_{22,21}
\]

\[
\hat{\Sigma}_{22} = \frac{1}{N} A_{22,n}
\]

Anderson and Olkin (1985) gave an elegant derivation of \( \hat{\Sigma} \)
A decomposition of $\hat{\Sigma}$

Notation: $A_{11.2,n} := A_{11} - A_{12}A_{22,n}^{-1}A_{21}$

$$n\hat{\Sigma} = \tau \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22,n} \end{pmatrix} + \bar{\tau} \begin{pmatrix} A_{11.2,n} & 0 \\ 0 & 0 \end{pmatrix}$$

$$+ \tau \begin{pmatrix} A_{12}A_{22,n}^{-1} & 0 \\ 0 & I_q \end{pmatrix} \begin{pmatrix} B & B \\ B & B \end{pmatrix} \begin{pmatrix} A_{22,n}^{-1} & A_{21} & 0 \\ 0 & I_q \end{pmatrix}$$

where

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22,n} \end{pmatrix} \sim W_{p+q}(n - 1, \Sigma) \text{ and } B \sim W_q(N - n, \Sigma_{22})$$

are independent. Also, $N\hat{\Sigma}_{22} \sim W_q(N - 1, \Sigma_{22})$
\[ A_{22,N} = \sum_{j=1}^{n} (Y_j - \bar{Y}_1 + \bar{Y}_1 - \bar{Y})(Y_j - \bar{Y}_1 + \bar{Y}_1 - \bar{Y})' \]

\[ + \sum_{j=n+1}^{N} (Y_j - \bar{Y}_2 + \bar{Y}_2 - \bar{Y})(Y_j - \bar{Y}_2 + \bar{Y}_2 - \bar{Y})' \]

\[ A_{22,N} = A_{22,n} + B \]

\[ B = \sum_{j=n+1}^{N} (Y_j - \bar{Y}_2)(Y_j - \bar{Y}_2)' + \frac{n(N-n)}{N}(\bar{Y}_1 - \bar{Y}_2)(\bar{Y}_1 - \bar{Y}_2)' \]

Verify that the terms in the decomposition of \( \hat{\Sigma} \) are independent
Even the distribution of $\hat{\Sigma}_{11}$ is non-trivial

If $\Sigma_{12} = 0$ then $A_{22,n}, B, A_{11.2,n}, A_{12}A_{22,n}^{-1}A_{21}, \bar{X}, \bar{Y}_1$, and $\bar{Y}_2$ are independent

Matrix $F$-distribution: $F_{a,b}^{(q)} = W_2^{-1/2} W_1 W_2^{-1/2}$

where $W_1 \sim W_q(a, \Sigma_{22})$ and $W_2 \sim W_q(b, \Sigma_{22})$
**Theorem:** Suppose that $\Sigma_{12} = 0$. Then

$$
\Sigma_{11}^{-1/2} \hat{\Sigma}_{11} \Sigma_{11}^{-1/2} \overset{\mathcal{L}}{=} \frac{1}{n} W_1 + \frac{1}{N} W_2^{1/2} (I_p + F') W_2^{1/2}
$$

where $W_1 \sim W_p(n - q - 1, I_p)$, $W_2 \sim W_p(q, I_p)$, $F' \sim F_{N-n,n-q+p-1}^{(p)}$, and $W_1$, $W_2$, and $F'$ are independent

$$
N \Sigma_{11}^{-1/2} \hat{\Sigma}_{11} \Sigma_{11}^{-1/2} \overset{\mathcal{L}}{=} \frac{N}{n} \Sigma_{11}^{-1/2} A_{11,2,n} \Sigma_{11}^{-1/2}
$$

$$
+ \Sigma_{11}^{-1/2} A_{12} A_{22,n}^{-1} (A_{22,n} + B) A_{22,n}^{-1} A_{21} \Sigma_{11}^{-1/2}
$$
With no assumptions on $\Sigma_{12}$:

$$\hat{\Sigma}_{12} \hat{\Sigma}^{-1}_{22} \equiv \Sigma_{12} \Sigma^{-1}_{22} + \Sigma_{11,2}^{1/2} W^{-1/2} K \Sigma^{-1/2}_{22}$$

where $W \sim W_p(n - q + p - 1, I_p)$, $K \sim N_{pq}(0, I_p \otimes I_q)$, and $W$ and $K$ are independent.

In particular, $\hat{\Sigma}_{12} \hat{\Sigma}^{-1}_{22}$ is an unbiased estimator of $\Sigma_{12} \Sigma^{-1}_{22}$.
Define

\[ \Delta_{11} = \Sigma_{11.2}, \quad \Delta_{12} = \Sigma_{12} \Sigma_{22}^{-1}, \quad \Delta_{22} = \Sigma_{22} \]

The matrix LDU decomposition of any positive definite matrix

\[ \Sigma = \begin{pmatrix} I_p & \Delta_{12} \\ 0 & I_q \end{pmatrix} \begin{pmatrix} \Delta_{11} & 0 \\ 0 & \Delta_{22} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ \Delta_{21} & I_q \end{pmatrix} \]

There is a 1:1 correspondence between \( \Sigma \) and

\[ \Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \]
The inverse transformation from $\Delta$ to $\Sigma$:

\[
\begin{align*}
\Sigma_{11} &= \Delta_{11} + \Delta_{12} \Delta_{22} \Delta_{21} \\
\Sigma_{12} &= \Delta_{12} \Delta_{22} \\
\Sigma_{22} &= \Delta_{22}
\end{align*}
\]

Starting with $\hat{\Sigma}$, we construct $\hat{\Delta}$ exactly as above
A hypergeometric function of matrix argument

Herz (1955), Muirhead (1982)

Multivariate gamma function: For \( a > (q - 1)/2 \),

\[
\Gamma_q(a) = \pi^{q(q-1)/4} \prod_{j=1}^{q} \Gamma(a - \frac{1}{2}(j - 1))
\]

Multivariate beta function: For \( a, b > (q - 1)/2 \),

\[
B_q(a, b) = \frac{\Gamma_q(a)\Gamma_q(b)}{\Gamma_q(a + b)}
\]
\( M \) is \( q \times q \) and symmetric; \( a, b - a > (q - 1)/2 \)

\[
\begin{align*}
1F_1^{(q)}(a; b; M) & = \frac{1}{B_q(a, b-a)} \int_{0<u<I_q} |u|^{a-\frac{1}{2}(q+1)} |I_q - u|^{b-a-\frac{1}{2}(q+1)} e^{\text{tr} M u} du,
\end{align*}
\]

\[1F_1(KMK') = 1F_1(M) \text{ for all } q \times q \text{ orthogonal matrices } K\]

Conclude: \( 1F_1(M) \) depends only on the eigenvalues of \( M \)

Herz (1955): If \( \text{rank}(M) = r \) then \( 1F_1^{(q)}(a; b; M) = 1F_1^{(r)}(a; b; M_0) \)

where \( M_0 \) is any \( r \times r \) matrix whose non-zero eigenvalues coincide with those of \( M \)
The distribution of $\hat{\Delta}$

Start with

$$
\begin{pmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22,n}
\end{pmatrix} = \sum_{j=1}^{n} \begin{pmatrix}
X_j - \bar{X} \\
Y_j - \bar{Y}_1
\end{pmatrix} \begin{pmatrix}
X_j - \bar{X} \\
Y_j - \bar{Y}_1
\end{pmatrix}'

\sim W_{p+q}(n - 1, \Sigma)

A well-known property of the Wishart distribution:

$A_{11,2,n}$ and $\{A_{12}, A_{22,n}\}$ are mutually independent

Conclude: $A_{11,2,n}$ and $\{A_{12}, A_{22,N}\}$ are mutually independent
Express the $\hat{\Delta}$’s in terms of the $A$’s

\[ n\hat{\Delta}_{11} = A_{11\cdot2,n} \sim W_p(n - q - 1, \Delta_{11}) \]

\[ N\hat{\Delta}_{22} = A_{22,N} \sim W_q(N - 1, \Delta_{22}) \]

\[ \hat{\Delta}_{12} = A_{12}A_{22,n}^{-1} \]

\[ \hat{\Delta}_{12}|_{A_{22,n}} = A_{12}A_{22,n}^{-1}|_{A_{22,n}} \sim N_{pq}(\Delta_{12}, \Delta_{11} \otimes A_{22,n}^{-1}) \]
We shall evaluate the p.d.f. of $\hat{\Delta}$ at

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$$

$\hat{\Delta}_{11}$ and $\{\hat{\Delta}_{12}, \hat{\Delta}_{22}\}$ are independent

All that remains to be done is to evaluate the conditional p.d.f. of $\hat{\Delta}_{12} | \hat{\Delta}_{22}$ at $T_{12}$.
$\Xi_1$, $\Xi_2$, and $\Xi_3$: Random matrices of the same dimension such that $(\Xi_1, \Xi_2)$ and $\Xi_3$ are independent. Then the conditional p.d.f. of $\Xi_1$ given $\Xi_2 + \Xi_3 = \xi$, is

$$f_{\Xi_1|\Xi_2+\Xi_3=\xi}(\xi_1) = \frac{1}{f_{\Xi_2+\Xi_3}(\xi)} \int f_{\Xi_1|\Xi_2=\xi_2}(\xi_1) f_{\Xi_2}(\xi_2) f_{\Xi_3}(\xi - \xi_2) \, d\xi_2$$

Set $\Xi_1 = A_{12} A_{22,n}^{-1} \equiv \hat{\Delta}_{12}$, $\Xi_2 = N^{-1} A_{22,n}$, $\Xi_3 = N^{-1} B$

$\Xi_2 + \Xi_3 = N^{-1}(A_{22,n} + B) \equiv \hat{\Delta}_{22}$

The lemma gives us the distribution of $\hat{\Delta}_{12}|\hat{\Delta}_{22}$
Substitute the densities of \( \Xi_1 | \Xi_2 \) and \( \Xi_2 \) into the integral and evaluate:

\[
\begin{align*}
    f_{\Delta_{12} | \Delta_{22} = T_{22}}(T_{12}) &= c |\Delta_{11}|^{-q/2} |\Delta_{22}|^{-(N-1)/2} \\
    &\times \exp\left(-\frac{1}{2} \text{tr} \ N \Delta_{22}^{-1} T_{22}\right) \\
    &\times |T_{22}|^{\frac{1}{2}(N+p-1) - \frac{1}{2}(q+1)} \\
    &\times {}_1F_1\left(q + 1 \left| \frac{1}{2}(n + p - 1); -\frac{1}{2} N (T_{12} - \Delta_{12})' \Delta_{11}^{-1} (T_{12} - \Delta_{12}) T_{22} \right.\right)
\end{align*}
\]
Conclude:

\[ f_{\Delta}(T) = f_{\Delta_{11}}(T_{11}) f_{\Delta_{22}}(T_{22}) f_{\Delta_{12}|\Delta_{22}=T_{22}}(T_{12}) \]

To transform back to \( \hat{\Sigma} \), the Jacobians are

\[ J(\Delta_{11} \rightarrow \hat{\Sigma}_{11}) J(\Delta_{12} \rightarrow \hat{\Sigma}_{12}) J(\Delta_{22} \rightarrow \hat{\Sigma}_{22}) = 1 \cdot |\hat{\Sigma}_{22}^{-1}|^p \cdot 1 \]

The density function of \( \hat{\Sigma} \) is

\[ f_{\hat{\Sigma}}(T) = f_{\Delta_{11},\Delta_{12},\Delta_{22}}(T_{11} - T_{12}T_{22}^{-1}T_{21}, T_{12}T_{22}^{-1}, T_{22}) |T_{22}|^{-p} \]
What can we do with, or obtain from, this expression for the p.d.f.?

Saddlepoint approximations: Apply Butler-Wood formulas

Can we let $q \to \infty$? Free probability ...

$\hat{\Delta}_{11}$ and $\hat{\Delta}_{22}$ are Wishart matrices, so we know their eigenvalue distributions

We can apply interlacing theorems to bound the distribution of the eigenvalues of $\hat{\Delta}$

Open problem: Find the distribution of the eigenvalues of $\hat{\Sigma}$
The distribution of $\hat{\Sigma}$ is non-trivial but the distribution of $|\hat{\Sigma}|$ is simple

$$|\hat{\Sigma}| = |\hat{\Sigma}_{11.2}| \cdot |\hat{\Sigma}_{22}| = |\hat{\Delta}_{11}| \cdot |\hat{\Delta}_{22}|$$

These two are independent and each is a product of $\chi^2$ random variables

Hao and Krishnamoorthy (2001): The sample generalized variance $|\hat{\Sigma}|$ is distributed as

$$n^{-p} N^{-q} |\Sigma| \cdot \chi^2_{n-q-1} \chi^2_{n-q-2} \cdots \chi^2_{n-q-1-p} \chi^2_{N-1} \chi^2_{N-2} \cdots \chi^2_{N-q}$$

This result raises the possibility of good results in hypothesis testing on $\Sigma$
Testing that $\Sigma = \Sigma_0$

Sample: Same as before, 2-step monotone incomplete

$\Sigma_0$: A given, positive definite matrix

Test $H_0: \Sigma = \Sigma_0$ against $H_a: \Sigma \neq \Sigma_0$

Hao and Krishnamoorthy (2001): The LRT statistic for testing $H_0$ against $H_a$ is

$$\lambda_1 \propto |A_{22,N}|^{N/2} \exp \left( - \frac{1}{2} \text{tr} A_{22,N} \right) \times |A_{11\cdot2,n}|^{n/2} \exp \left( - \frac{1}{2} \text{tr} A_{11\cdot2,n} \right) \times \exp \left( - \frac{1}{2} \text{tr} A_{12} A_{22,n}^{-1} A_{21} \right).$$
Is the LRT unbiased? If $C$ is a critical region of size $\alpha$, is

$$P(\lambda_1 \in C|H_a) \geq P(\lambda_1 \in C|H_0)?$$

In the case of complete data, it is well-known that the LRT is not unbiased

E. J. G. Pitman: $\lambda_1$ becomes unbiased if the sample sizes are replaced by the corresponding degrees of freedom

With monotone incomplete data, perhaps a similarly modified statistic, $\lambda_2$, is unbiased?
Answer: Not always; $\lambda_2$ is unbiased if $|\Sigma_{11}| < 1$

With monotone incomplete data, a further modification is necessary: The modified LRT

$$\lambda_3 \propto |A_{22,N}|^{(N-1)/2} \exp \left( -\frac{1}{2} \text{tr} \ A_{22,N} \right)$$
$$\times |A_{11\cdot2,n}|^{(n-q-1)/2} \exp \left( -\frac{1}{2} \text{tr} \ A_{11\cdot2,n} \right)$$
$$\times |A_{12} A_{22,n}^{-1} A_{21}|^{q/2} \exp \left( -\frac{1}{2} \text{tr} \ A_{12} A_{22,n}^{-1} A_{21} \right),$$

is unbiased. Also, $\lambda_1$ is never unbiased for 2-step monotone data.

For diagonal $\Sigma = \text{diag}(\sigma_{jj})$, the power function of $\lambda_3$ increases monotonically as any $|\sigma_{jj} - 1|$ increases, $j = 1, \ldots, p + q$. 
\( \mu_0 \) and \( \Sigma_0 \) are completely specified

With monotone 2-step data, test

\[ H_0 : (\mu, \Sigma) = (\mu_0, \Sigma_0) \text{ vs. } H_a : (\mu, \Sigma) \neq (\mu_0, \Sigma_0) \]

The LRT is

\[ \lambda_4 = \lambda_1 \exp \left( -\frac{1}{2} (n\bar{X}'\bar{X} + N\bar{Y}'\bar{Y}) \right) \]

Remarkably, \( \lambda_4 \) is unbiased
The sphericity test, $H_0 : \Sigma \propto I_{p+q}$

The LRT statistic is known but whether or not it is unbiased remains an open question