Finite-Sample Inference with Incomplete Multivariate Normal Data

This talk is based on joint work with my wonderful co-authors:

Wan-Ying Chang (Dept. of Fish & Wildlife, WA, USA)
Tomoya Yamada (Sapporo Gakuin University, Japan)
Megan Romer (Penn State University)
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 (low-density lipoprotein) cholesterol
3. Systolic blood pressure
4. Glucose intolerance
5. Insulin response to oral glucose
6. Actual weight ÷ Expected weight

\[ \vdots \]

$m$ White 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>(v_{1,1})</td>
<td>(v_{2,1})</td>
<td>(v_{3,1})</td>
<td>\cdots</td>
<td>(v_{N,1})</td>
</tr>
<tr>
<td></td>
<td>(v_{1,2})</td>
<td>(v_{2,2})</td>
<td>(v_{3,2})</td>
<td>\cdots</td>
<td>(v_{N,2})</td>
</tr>
<tr>
<td></td>
<td>\cdots</td>
<td>\cdots</td>
<td>\cdots</td>
<td>\cdots</td>
<td>\cdots</td>
</tr>
<tr>
<td></td>
<td>(v_{1,m})</td>
<td>(v_{2,m})</td>
<td>(v_{3,m})</td>
<td>\cdots</td>
<td>(v_{N,m})</td>
</tr>
</tbody>
</table>

Vector notation:  \(V_1, V_2, V_3, \ldots, V_N\)

\(V_1\): The measurements on patient 1, stacked in a column etc.
Classical multivariate analysis

Statistical analysis of $N m$-dimensional data vectors

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 probability density function of \( V \): For \( v \in \mathbb{R}^m \),

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

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

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

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

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

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

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

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

LLN: As \( N \to \infty \), \( \frac{1}{N} \Sigma \to 0 \) and hence \( \hat{\mu} \to \mu \), a.s.

\( N \hat{\Sigma} \) has a Wishart distribution, a generalization of the \( \chi^2 \)

\( \hat{\mu} \) and \( \hat{\Sigma} \) also are mutually independent.
Monotone incomplete data

Some patients were not measured completely

The resulting data set, with * denoting a missing observation

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

Monotone data: Each * is followed by *’s only

We may need to renumber patients to display the data in monotone form
Physical Fitness Data

A well-known data set from a SAS manual on missing 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>39.407</td>
<td>12.63</td>
<td>174</td>
</tr>
<tr>
<td>46.080</td>
<td>11.17</td>
<td>156</td>
</tr>
<tr>
<td>45.441</td>
<td>9.63</td>
<td>164</td>
</tr>
<tr>
<td>54.625</td>
<td>8.92</td>
<td>146</td>
</tr>
<tr>
<td>37.388</td>
<td>14.03</td>
<td>186</td>
</tr>
<tr>
<td>49.091</td>
<td>10.85</td>
<td>170</td>
</tr>
<tr>
<td>49.874</td>
<td>9.22</td>
<td>176</td>
</tr>
<tr>
<td>46.672</td>
<td>10.00</td>
<td>*</td>
</tr>
<tr>
<td>46.774</td>
<td>10.25</td>
<td>*</td>
</tr>
<tr>
<td>45.118</td>
<td>11.08</td>
<td>*</td>
</tr>
<tr>
<td>49.874</td>
<td>9.22</td>
<td>*</td>
</tr>
<tr>
<td>49.091</td>
<td>10.85</td>
<td>*</td>
</tr>
<tr>
<td>59.571</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>50.541</td>
<td>*</td>
<td>*</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 $q$

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

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

Goal: Maximum likelihood inference for $\mu$ and $\Sigma$, with analytical results as extensive and as explicit as in the classical setting
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

...
Monotone incomplete data:
\[
\left( \begin{array}{c} X_1 \\ Y_1 \end{array} \right), \ldots, \left( \begin{array}{c} X_n \\ Y_n \end{array} \right), \left( \begin{array}{c} * \\ Y_{n+1} \end{array} \right), \ldots, \left( \begin{array}{c} * \\ Y_N \end{array} \right)
\]

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:

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

Let

$$\mu_{1.2} = \mu_1 - \Sigma_{12} \Sigma_{22}^{-1} (Y - \mu_2), \quad \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

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

$\hat{\mu}$ and $\hat{\Sigma}$: Wilks, Anderson, Morrison, Olkin, Jinadasa, Tracy, ...

Anderson and Olkin (1985): An elegant derivation of $\hat{\Sigma}$
Sample means:

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

\[
\bar{Y}_1 = \frac{1}{n} \sum_{j=1}^{n} Y_j, \quad \bar{Y}_2 = \frac{1}{N - n} \sum_{j=n+1}^{N} Y_j
\]

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 MLE’s of $\mu$ and $\Sigma$

Notation: $\tau = n/N$, $\bar{\tau} = 1 - \tau$

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

$\hat{\mu}_1$ is called the *regression estimator of* $\mu_1$

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

$\hat{\Sigma}$ is more complicated:

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

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

$$
\hat{\Sigma}_{22} = \frac{1}{N} A_{22} N
$$
Seventy-five year-old unsolved problems

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

In testing hypotheses on $\mu$ or $\Sigma$, are the LRT statistics unbiased?

Calculate the higher moments of the components of $\hat{\mu}$

Determine the asymptotic behavior of $\hat{\mu}$ as $n$ or $N \to \infty$

The Stein phenomenon for $\hat{\mu}$?

The crucial obstacle: The exact distribution of $\hat{\mu}$
The exact distribution of $\hat{\mu}$

Chang and D.R.: For $n > p + q$,

$$
\hat{\mu} \sim \mu + V_1 + \left(\frac{1}{n} - \frac{1}{N}\right)^{1/2} \left(\frac{Q_2}{Q_1}\right)^{1/2} \begin{pmatrix} V_2 \\ 0 \end{pmatrix},
$$

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

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

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

Corollary: $\hat{\mu}$ is an unbiased estimator of $\mu$. Also, $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent iff $\Sigma_{12} = 0$. 
Computation of the higher moments of $\hat{\mu}$ now is straightforward.

Due to the term $1/Q_1$, even moments exist only up to order $n - q$.

The covariance matrix of $\hat{\mu}$:

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

Asymptotics for $\hat{\mu}$: If $n, N \to \infty$ with $N/n \to \delta \geq 1$ then

$$\sqrt{N}(\hat{\mu} - \mu) \xrightarrow{\mathcal{L}} N_{p+q} \left( 0, \Sigma + (\delta - 1) \begin{pmatrix} \Sigma_{11.2} & 0 \\ 0 & 0 \end{pmatrix} \right)$$
An analog of Hotelling’s $T^2$-statistic

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

where

$$\widehat{\text{Cov}}(\hat{\mu}) = \frac{1}{N} \hat{\Sigma} + \frac{(n - 2)\bar{\tau}}{n(n - q - 2)} \begin{pmatrix} \hat{\Sigma}_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

A natural ellipsoidal confidence region for $\mu$ is

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

Calculate the corresponding confidence level

Chang and D.R.: Lower bounds on the confidence coefficient; and asymptotics for $T^2$ under various assumptions, e.g.,

$n, N \to \infty \text{ with } n/N \to \delta, 0 < \delta \leq 1$
Some results of M. Romer

The probability distribution of $T^2$ does not depend on $\mu$ or $\Sigma$

A stochastic representation for the exact distribution of $T^2$

Improved lower bounds on the confidence coefficient of ellipsoidal confidence regions for $\mu$

A more algebraic derivation of the distribution of $\hat{\mu}$

Intricate algebraic calculations, repeated applications of the orthogonal invariance of various distributions
Inference for $\Sigma$

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

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

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
The distribution of $|\hat{\Sigma}|$ is much simpler:

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

$|\hat{\Sigma}_{11.2}|$ and $|\hat{\Sigma}_{22}|$ are independent; each is a product of independent $\chi^2$ variables

Hao and Krishnamoorthy (2001):

$$|\hat{\Sigma}| \overset{\mathcal{L}}{=} n^{-p} N^{-q} |\Sigma| \cdot \prod_{j=1}^{p} \chi_{n-q-j}^2 \cdot \prod_{j=1}^{q} \chi_{N-j}^2$$

It now is plausible that tests of hypothesis on $\Sigma$ are unbiased
Testing $\Sigma = \Sigma_0$

Data: Two-step, monotone incomplete sample

$\Sigma_0$: A given, positive definite matrix

Test $H_0: \Sigma = \Sigma_0$ vs. $H_a: \Sigma \neq \Sigma_0$ (WLOG, $\Sigma_0 = I_{p+q}$)

Hao and Krishnamoorthy (2001): The LRT statistic is

$$
\lambda_1 \propto |A_{22,N}|^{N/2} \exp \left( -\frac{1}{2} \text{tr} A_{22,N} \right) \\
\times |A_{11,2,n}|^{n/2} \exp \left( -\frac{1}{2} \text{tr} A_{11,2,n} \right) \\
\times \exp \left( -\frac{1}{2} \text{tr} A_{12}A_{22,n}^{-1}A_{21} \right).
$$

Is the LRT unbiased? If $C$ is a critical region of size $\alpha$, is

$$
P(\lambda_1 \in C|H_a) \geq P(\lambda_1 \in C|H_0)?$$
E. J. G. Pitman: With complete data, $\lambda_1$ is not unbiased

$\lambda_1$ becomes unbiased if sample sizes are replaced by degrees of freedom

With two-step monotone data, perhaps a similarly modified statistic, $\lambda_2$, is unbiased?

Answer: Still unknown.

Chang and D.R.: If $|\Sigma_{11}| < 1$ then $\lambda_2$ is unbiased

With monotone incomplete data, further modification is needed
Theorem: The modified LRT,

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

is unbiased. Also, $\lambda_1$ is not unbiased.

For diagonal $\Sigma = \text{diag}(\sigma_{jj})$, the power function of $\lambda_3$ increases monotonically as any $|\sigma_{jj} - 1|$ increases, $j = 1, \ldots, p + q$. 
With monotone two-step data, test

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

where \( \mu_0 \) and \( \Sigma_0 \) are given. The LRT statistic 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, \( H_0 : \Sigma \propto I_{p+q} \ vs. \ H_a : \not\propto I_{p+q} \)

The unbiasedness of the LRT statistic is an open problem
The Stein phenomenon for $\hat{\mu}$

$\hat{\mu}$: The mean of a complete sample from $N_m(\mu, I_m)$

Quadratic loss function: $L(\hat{\mu}, \mu) = \|\hat{\mu} - \mu\|^2$

Risk function: $R(\hat{\mu}) = E[L(\hat{\mu}, \mu)]$

C. Stein: $\hat{\mu}$ is inadmissible for $m \geq 3$

James-Stein estimator for shrinking $\hat{\mu}$ to $\nu \in \mathbb{R}^m$:

$$\hat{\mu}_c = \left(1 - \frac{c}{\|\hat{\mu} - \nu\|^2}\right) (\hat{\mu} - \nu) + \nu$$

Baranchik’s positive-part shrinkage estimator:

$$\hat{\mu}_c^+ = \left(1 - \frac{c}{\|\hat{\mu} - \nu\|^2}\right)_+ (\hat{\mu} - \nu) + \nu$$
We collect a monotone incomplete sample from $\mathcal{N}_{p+q}(\mu, \Sigma)$.

Does the Stein phenomenon hold for $\hat{\mu}$, the MLE of $\mu$?

The phenomenon seems almost universal: It holds for many loss functions, inference problems, and distributions.

Various results available on shrinkage estimation of $\Sigma$ with incomplete data, but no such results available for $\mu$.

The crucial impediment: The distribution of $\hat{\mu}$ was unknown.
Theorem (Yamada and D.R.): For $p \geq 2$, $n \geq q + 3$, and
$\Sigma = I_{p+q}$, both $\hat{\mu}$ and $\hat{\mu}_c$ are inadmissible:

$$R(\hat{\mu}) > R(\hat{\mu}_c) > R(\hat{\mu}_c^+)$$

for all $\nu \in \mathbb{R}^{p+q}$ and all $c \in (0, 2c^*)$, where

$$c^* = \frac{p - 2}{n} + \frac{q}{N}.$$  

Non-radial loss functions

Replace $\|\hat{\mu} - \nu\|^2$ by non-radial functions of $\hat{\mu} - \nu$

Shrinkage to a random vector $\nu$, calculated from the data
Testing for multivariate normality

Monotone incomplete data, i.i.d., unknown population:

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

A generalization of Mardia’s statistic for testing for kurtosis:

\[
\hat{\beta} = \sum_{j=1}^{n} \left[ \left( \begin{pmatrix}
X_j \\
Y_j
\end{pmatrix} - \hat{\mu} \right)' \hat{\Sigma}^{-1} \left( \begin{pmatrix}
X_j \\
Y_j
\end{pmatrix} - \hat{\mu} \right) \right]^2 + \sum_{j=n+1}^{N} \left[ (Y_j - \hat{\mu}_2)' \hat{\Sigma}_{22}^{-1} (Y_j - \hat{\mu}_2) \right]^2
\]
Alternatively, impute each missing \( X_j \) using linear regression:

\[
\hat{X}_j = \begin{cases} 
X_j, & 1 \leq j \leq n \\
\hat{\mu}_1 + \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} (Y_j - \hat{\mu}_2), & n + 1 \leq j \leq N
\end{cases}
\]

and construct

\[
\hat{\beta}_* = \sum_{j=1}^{N} \left[ \left( \begin{pmatrix} \hat{X}_j \\ Y_j \end{pmatrix} - \hat{\mu} \right)' \hat{\Sigma}^{-1} \left( \begin{pmatrix} \hat{X}_j \\ Y_j \end{pmatrix} - \hat{\mu} \right) \right]^2
\]

Yamada, Romer, D.R.: \( \hat{\beta} \equiv \hat{\beta}_* \). Also, with certain regularity conditions, and constants \( c_1, c_2 \),

\[
(\hat{\beta} - c_1)/c_2 \xrightarrow{\mathcal{L}} N(0, 1)
\]

as \( n, N \to \infty \).
Three-step monotone multivariate normal data

\[
\begin{pmatrix}
X_1 \\
Y_1 \\
Z_1
\end{pmatrix}, \ldots, \begin{pmatrix}
X_l \\
Y_l \\
Z_l
\end{pmatrix}, \begin{pmatrix}
* \\
Y_{l+1} \\
Z_{l+1}
\end{pmatrix}, \ldots, \begin{pmatrix}
* \\
Y_m \\
Z_m
\end{pmatrix}, \begin{pmatrix}
* \\
Y_{m+1} \\
Z_{m+1}
\end{pmatrix}, \ldots, \begin{pmatrix}
* \\
\end{pmatrix}
\]

Partition \( \mu \) and \( \Sigma \) in the usual way:

\[
\mu = \begin{pmatrix}
\mu_1 \\
\mu_2 \\
\mu_3
\end{pmatrix}, \quad \Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\
\Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\
\Sigma_{31} & \Sigma_{32} & \Sigma_{33}
\end{pmatrix}
\]

The MLEs for \( \mu \) and \( \Sigma \) can be derived recursively

Romer: \( \hat{\mu}_3 \) is independent of \( \{\hat{\mu}_1, \hat{\mu}_2\} \). Also, \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) are uncorrelated.
Non-monotone incomplete multivariate normal data

Eaton and Kariya (1983)

\[
\begin{pmatrix}
X_1 \\
Y_1
\end{pmatrix}, \ldots, \begin{pmatrix}
X_l \\
Y_l
\end{pmatrix}, \begin{pmatrix}
* \\
Y_{l+1}
\end{pmatrix}, \ldots, \begin{pmatrix}
* \\
Y_m
\end{pmatrix}, \begin{pmatrix}
X_{m+1} \\
*
\end{pmatrix}, \ldots, \begin{pmatrix}
X_n \\
*
\end{pmatrix}
\]

The ML equations $\mu$ and $\Sigma$ are rational, not polynomial, and will have multiple solutions.

If we reduce the rational equations to polynomial form, we can use Bezout's theorem to obtain bounds for the number of solutions.

We plan to apply BKK to obtain the exact number of solutions (at least in special cases).
References


Chang and Richards (2008). Finite-sample inference with monotone incomplete multivariate normal data, II.

