

Statistical Inference with Monotone 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 \div Expected weight
- \vdots $\quad \quad \quad \vdots$
- m White blood cell count

We obtain data:

Patient	1	2	3	...	N
	$\begin{pmatrix} v_{1,1} \\ v_{1,2} \\ \vdots \\ v_{1,m} \end{pmatrix}$	$\begin{pmatrix} v_{2,1} \\ v_{2,2} \\ \vdots \\ v_{2,m} \end{pmatrix}$	$\begin{pmatrix} v_{3,1} \\ v_{3,2} \\ \vdots \\ v_{3,m} \end{pmatrix}$...	$\begin{pmatrix} v_{N,1} \\ v_{N,2} \\ \vdots \\ v_{N,m} \end{pmatrix}$

Vector notation: V_1, V_2, \dots, V_N

V_1 : The measurements on patient 1, stacked into 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:

μ : The population mean vector

Σ : The population covariance matrix

For a given data set, μ and Σ are unknown

We wish to perform inference about μ and Σ

Construct confidence regions for, and test hypotheses about, μ and Σ

Anderson (2003). *An Introduction to Multivariate Statistical Analysis*

Eaton (1984). *Multivariate Statistics: A Vector-Space Approach*

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, V_2, \dots, V_N : Measurements on N randomly chosen patients

Estimate μ and Σ using Fisher's maximum likelihood principle

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

Maximum likelihood estimator: The value of (μ, Σ) that maximizes L

$\hat{\mu} = \frac{1}{N} \sum_{j=1}^N V_j$: The sample mean and MLE of μ

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

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

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

LLN: As $N \rightarrow \infty$, $\frac{1}{N} \Sigma \rightarrow \mathbf{0}$ and hence $\hat{\mu} \rightarrow \mu$, a.s.

$N\hat{\Sigma}$ has a Wishart distribution, a generalization of the χ^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

$$\begin{pmatrix} v_{1,1} \\ v_{1,2} \\ v_{1,3} \\ \vdots \\ v_{1,m} \end{pmatrix} \begin{pmatrix} * \\ v_{2,2} \\ v_{2,3} \\ \vdots \\ v_{2,m} \end{pmatrix} \begin{pmatrix} * \\ * \\ v_{3,2} \\ \vdots \\ v_{3,m} \end{pmatrix} \dots \begin{pmatrix} * \\ * \\ * \\ \vdots \\ v_{N,m} \end{pmatrix}$$

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)

Oxygen RunTime RunPulse

44.609	11.37	178		39.407	12.63	174
45.313	10.07	185		46.080	11.17	156
54.297	8.65	156		45.441	9.63	164
51.855	10.33	166		54.625	8.92	146
49.156	8.95	180		39.442	13.08	174
40.836	10.95	168		60.055	8.63	170
44.811	11.63	176		37.388	14.03	186
45.681	11.95	176		44.754	11.12	176
39.203	12.88	168		46.672	10.00	*
45.790	10.47	186		46.774	10.25	*
50.545	9.93	148		45.118	11.08	*
48.673	9.40	186		49.874	9.22	*
47.920	11.50	170		49.091	10.85	*
47.467	10.50	170		59.571	*	*
50.388	10.08	168		50.541	*	*
				47.273	*	*

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

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

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

Goal: Maximum likelihood inference for μ and Σ , 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

⋮

We have n observations on $\begin{pmatrix} X \\ Y \end{pmatrix}$ and $N - n$ additional observations on Y

Difficulty: The likelihood function is more complicated

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

Partition μ and Σ 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

$$\mu_{1\cdot 2} = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mu_2 - \mu_2), \quad \Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$Y \sim N_q(\mu_2, \Sigma_{22}), \quad X|Y \sim N_p(\mu_{1\cdot 2}, \Sigma_{11\cdot 2})$$

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

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

Sample means:

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

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

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 MLE's of μ and Σ

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 μ_1

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

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

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

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

$$\hat{\Sigma}_{22} = \frac{1}{N} A_{22,N}$$

Seventy year-old unsolved problems

Explicit confidence levels for elliptical confidence regions for μ

In testing hypotheses on μ or Σ , 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 $N \rightarrow \infty$

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

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

The exact distribution of $\hat{\mu}$

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

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

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

$$V_1 \sim N_{p+q}(\mathbf{0}, \Omega), \quad V_2 \sim N_p(\mathbf{0}, \Sigma_{11 \cdot 2}), \quad Q_1 \sim \chi_{n-q}^2, \quad Q_2 \sim \chi_q^2;$$

$$\Omega = \frac{1}{N}\Sigma + \left(\frac{1}{n} - \frac{1}{N}\right) \begin{pmatrix} \Sigma_{11 \cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Corollary: $\hat{\mu}$ is an unbiased estimator of μ . Also, $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent iff $\Sigma_{12} = \mathbf{0}$.

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 $\hat{\mu}$:

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

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

$$\sqrt{N}(\hat{\mu} - \mu) \xrightarrow{\mathcal{L}} N_{p+q} \left(\mathbf{0}, \Sigma + (\delta - 1) \begin{pmatrix} \Sigma_{11.2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right)$$

The analog of Hotelling's T^2 -statistic

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

where

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

An obvious ellipsoidal confidence region for μ 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

Reminder:

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

Define

$$\hat{C} \equiv \begin{pmatrix} I_p & -\bar{\tau} A_{12} A_{22,n}^{-1} & \bar{\tau} A_{12} A_{22,n}^{-1} \\ \mathbf{0} & \tau I_q & \bar{\tau} I_q \end{pmatrix}$$

Then

$$\hat{\mu} = \hat{C} \begin{pmatrix} \bar{X} \\ \bar{Y}_1 \\ \bar{Y}_2 \end{pmatrix}$$

Let

$$\widehat{\Psi} = \widehat{C} \left(\frac{1}{n} I_{p+q} \oplus \frac{1}{N-n} I_q \right) \widehat{C}'$$

$\widehat{\theta}$: The largest eigenvalue of $\widehat{\text{Cov}}(\widehat{\mu})^{-1/2} \widehat{\Psi} \widehat{\text{Cov}}(\widehat{\mu})^{-1/2}$

Theorem: The confidence level of the region

$$\left\{ \nu \in \mathbb{R}^{p+q} : \frac{(\widehat{\mu} - \nu)' \widehat{\text{Cov}}(\widehat{\mu})^{-1} (\widehat{\mu} - \nu)}{\widehat{\theta}} \leq \chi_{p+q}^2(\alpha) \right\}$$

is at least $1 - \alpha$

Simultaneous confidence intervals

$$\left\{ \left(\hat{\mu}_j \pm t_j \widehat{\text{SE}}(\hat{\mu}_j) \right) : 1 \leq j \leq p + q \right\}$$

for the components of $\hat{\mu}$, we need lower bounds on

$$P \left(\left| \frac{\hat{\mu}_j - \mu_j}{\widehat{\text{SE}}(\hat{\mu}_j)} \right| \leq t_j; 1 \leq j \leq p + q \right)$$

We have lower bounds of the form:

$$\begin{aligned} c_1 \times \prod_{j=1}^p t_j \left(1 + \frac{2}{n} t_j^2 \right)^{-(n-q)/2} \times \prod_{j=p+1}^{p+q} t_j \left(1 + \frac{2}{N} t_j^2 \right)^{-N/2} \\ \times \exp \left[-c_2 \left(\sum_{j=1}^p \left(1 + \frac{2}{n} t_j^2 \right) \right) \left(\sum_{j=p+1}^{p+q} \left(1 + \frac{2}{N} t_j^2 \right) \right) \right] \end{aligned}$$

A decomposition of $\widehat{\Sigma}$

Notation: $A_{11 \cdot 2, n} := A_{11} - A_{12}A_{22, n}^{-1}A_{21}$

$$n\widehat{\Sigma} = \tau \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22, n} \end{pmatrix} + \bar{\tau} \begin{pmatrix} A_{11 \cdot 2, n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ + \tau \begin{pmatrix} A_{12}A_{22, n}^{-1} & \mathbf{0} \\ \mathbf{0} & I_q \end{pmatrix} \begin{pmatrix} B & B \\ B & B \end{pmatrix} \begin{pmatrix} A_{22, n}^{-1}A_{21} & \mathbf{0} \\ \mathbf{0} & I_q \end{pmatrix}$$

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, $N\widehat{\Sigma}_{22} \sim W_q(N-1, \Sigma_{22})$

$$\begin{aligned}
A_{22,N} &= \sum_{j=1}^n (Y_j - \bar{Y}_1 + \bar{Y}_1 - \bar{Y})(Y_j - \bar{Y}_1 + \bar{Y}_1 - \bar{Y})' \\
&\quad + \sum_{j=n+1}^N (Y_j - \bar{Y}_2 + \bar{Y}_2 - \bar{Y})(Y_j - \bar{Y}_2 + \bar{Y}_2 - \bar{Y})'
\end{aligned}$$

$$A_{22,N} = A_{22,n} + B$$

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

Verify that the terms in the decomposition of $\hat{\Sigma}$ are independent

The marginal distribution of $\widehat{\Sigma}_{11}$ is non-trivial

If $\Sigma_{12} = \mathbf{0}$ then $A_{22,n}$, B , $A_{11 \cdot 2,n}$, $A_{12}A_{22,n}^{-1}A_{21}$, \bar{X} , \bar{Y}_1 , and \bar{Y}_2 are independent

Matrix F -distribution: $F_{a,b}^{(q)} = W_2^{-1/2}W_1W_2^{-1/2}$

where $W_1 \sim W_q(a, \Sigma_{22})$ and $W_2 \sim W_q(b, \Sigma_{22})$

Theorem: Suppose that $\Sigma_{12} = \mathbf{0}$. Then

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

where W_1 , W_2 , and F are independent, and

$$W_1 \sim W_p(n - q - 1, I_p), \quad W_2 \sim W_p(q, I_p), \quad \text{and}$$

$$F \sim F_{N-n, n-q+p-1}^{(p)}$$

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

$$+ \Sigma_{11}^{-1/2} A_{12} A_{22, n}^{-1} (A_{22, n} + B) A_{22, n}^{-1} A_{21} \Sigma_{11}^{-1/2}$$

Theorem: With no assumptions on Σ_{12} ,

$$\widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1} \stackrel{\mathcal{L}}{=} \Sigma_{12}\Sigma_{22}^{-1} + \Sigma_{11.2}^{1/2}W^{-1/2}K\Sigma_{22}^{-1/2}$$

where W and K are independent, and

$$W \sim W_p(n - q + p - 1, I_p), \quad K \sim N_{pq}(\mathbf{0}, I_p \otimes I_q)$$

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

The general distribution of $\widehat{\Sigma}$ requires the hypergeometric functions of matrix argument

Saddlepoint approximations

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

$$|\widehat{\Sigma}| = |\widehat{\Sigma}_{11 \cdot 2}| \cdot |\widehat{\Sigma}_{22}|$$

$|\widehat{\Sigma}_{11 \cdot 2}|$ and $|\widehat{\Sigma}_{22}|$ are independent; each is a product of independent χ^2 variables

Hao and Krishnamoorthy (2001):

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

It now is plausible that tests of hypothesis on Σ are unbiased

Testing $\Sigma = \Sigma_0$

Data: Two-step, monotone incomplete sample

Σ_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

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

Is the LRT unbiased? If C is a critical region of size α , is

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

E. J. G. Pitman: With complete data, λ_1 is not unbiased

λ_1 becomes unbiased if sample sizes are replaced by degrees of freedom

With two-step monotone data, perhaps a similarly modified statistic, λ_2 , is unbiased?

Answer: Still unknown.

Theorem: If $|\Sigma_{11}| < 1$ then λ_2 is unbiased

With monotone incomplete data, further modification is needed

Theorem: The modified LRT,

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

is unbiased. Also, λ_1 is not unbiased

For diagonal $\Sigma = \text{diag}(\sigma_{jj})$, the power function of λ_3 increases monotonically as any $|\sigma_{jj} - 1|$ increases, $j = 1, \dots, p + q$.

With monotone two-step data, test

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

where μ_0 and Σ_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, λ_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 $N_{p+q}(\mu, \Sigma)$

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

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

Various results available on shrinkage estimation of Σ with incomplete data, but no such results available for μ

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 ν , calculated from the data

Testing for multivariate normality

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

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

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

$$\begin{aligned} \hat{\beta} &= \sum_{j=1}^n \left[\left(\begin{pmatrix} X_j \\ Y_j \end{pmatrix} - \hat{\mu} \right)' \hat{\Sigma}^{-1} \left(\begin{pmatrix} X_j \\ Y_j \end{pmatrix} - \hat{\mu} \right) \right]^2 \\ &\quad + \sum_{j=n+1}^N \left[(Y_j - \hat{\mu}_2)' \hat{\Sigma}_{22}^{-1} (Y_j - \hat{\mu}_2) \right]^2 \end{aligned}$$

An alternative to $\hat{\beta}$

Impute each missing X_j using linear regression:

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

Construct

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

Remarkably, $\hat{\beta} \equiv \hat{\beta}_*$

Theorem (Yamada, Romer, D.R.): With certain regularity conditions, constants c_1, c_2 ,

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

References

W.-Y. Chang and D.R. (2007). [Finite-sample inference with monotone incomplete multivariate normal data, I](#). Preprint, Penn State University.

W.-Y. Chang and D.R. (2008). [Finite-sample inference with monotone incomplete multivariate normal data, II](#). Preprint, Penn State University.

D.R. and T. Yamada (2008). [The Stein phenomenon for monotone incomplete multivariate normal data](#). Preprint, Penn State University.

T. Yamada, M. Romer, and D.R. (2008). [Measures of multivariate kurtosis with monotone incomplete data](#). In preparation.