IMPS2015
Beijing Normal U. China
July 13 – 16, 2015

Congratulations on the 80th anniversary of the Psychometric Society!

Developments in multivariate missing data analysis

Yutaka Kano, Osaka University, Japan

Outline of the talk

---Two topics being presented---

- *Part I. Bias of the observed-likelihood (FIML) MLE for NMAR missingness*
  - Effects of auxiliary variables in the analysis of missing data
  - Approximate population bias (APB) of the MLE in the analysis of nonignorable missing data
    - Joint work with Yoshiharu Takagi
- *Part II. A huge amount of missing values*
  - Full information maximum likelihood estimation in factor analysis with a large number of missing values
    - *To appear in Journal of Statistical Computation and Simulation*, 2015
    - Kei Hirose, Yutaka Kano (Osaka U, Japan)
    - Sunyong Kim, Miyuki Imada, Manabu Yoshida and Masato Matsuo (NTT, Japan)

Missing data analysis is very important!!

- Publication Manual of the APA (2010, 6th edition)
  - Similarly, missing data can have a detrimental effect on the legitimacy of the inferences drawn by statistical tests. For this reason, it is critical that the frequency or percentages of missing data be reported along with any empirical evidence and/or theoretical arguments for the causes of data that are missing. For example, data might be described as MCAR; MAR; or NMAR. It is also important to describe the methods for addressing missing data, if any were used (e.g., multiple imputation).

- Publication Manual of the Japanese Psychological Association (revised in 2015)
  - データの欠測は分析の結果に大きな影響をしばしば与える。欠測を伴うデータを分析する場合には，欠測の頻度や件数を示すとともに，欠測の発生について経験的あるいは理論的な説明を記述する。分析において採用した欠測モデルの性質(MCAR, MAR, NMAR の区分)や，欠測に対応するために採用した方法(多重埋め合わせなど)について記述することが望ましい。

Part I. Bias of the observed-likelihood (FIML) MLE for NMAR missingness

Yutaka kano and Yoshiharu Takagi

Basic theory of missing data analysis

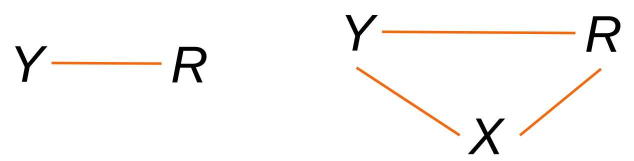
- observable variables: $\boldsymbol{Y} = (Y_{ij}) = [\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}]$
- response (or missing) indicator: $\boldsymbol{R} = (R_{ij})$

$$\begin{cases} R_{ij} = 0 \implies Y_{ij} \text{ is missing} \\ R_{ij} = 1 \implies Y_{ij} \text{ is observed} \end{cases}$$

- Missing-data mechanism (MDM): $P(\boldsymbol{R}|\boldsymbol{Y})$
- MAR: $P(\boldsymbol{R}|\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}) = P(\boldsymbol{R}|\boldsymbol{Y}_{obs})$
- FIML(Observed Likelihood): $L(\theta|\boldsymbol{Y}_{obs}) = f(\boldsymbol{Y}_{obs}|\theta)$
- MLE: $\hat{\theta} := \underset{\theta}{\text{argmax}}\ L(\theta|\boldsymbol{Y}_{obs})$
- If MDM is MAR, then MLE is consitent;
  if not, the MLE is biased.
- Any useful way of alleviating the bias is wanted.

Auxiliary-Variable method

- The auxiliary-variable (AV) method aims at reducing the bias of the estimators based on ML or MI by including external variables into a statistical model which are not of direct interest in the statistical analysis.
- An example:
  $Y \sim f(y|\theta)$, with θ a parameter of interest



Model without AV *X*　　　Model with AV *X*

Literature on the AV method in missing data analysis

- Original literature
  - Ibrahim, Lipsitz and Horton (2001) Appl. Statist.
  - Rubin (1996), Meng (1994)
- Psychometrics
  - Hoo (2009), Graham (2003, 2009)
    Collins, Shafer and Kam (2001)
- Clinical or Medical statistics
  - O'Neill and Temple (2012)
    Hardt, Herke and Leonhart (2012)
    Wang and Hall (2010), Daniels and Hogan (2008, sec.5.4)
- Sociology
  - Mustillo (2012).

Mechanism of reducing bias, How to choose AVs

- Collins, Schafer and Kam (2001, Psych. Meth.)
  - While not part of the substantive model, they can improve the performance of FIML because:
    - Making the MAR assumption more reasonable
    - Acting as proxies for x, even if MAR is violated
    - Increase efficiency by reducing uncertainty due to missingness
- Enders (2006, AERA extended Course)
  - Incorporating AVs can make MAR more plausible
  - A useful AV is either a potential cause or correlate of missingness, or a correlate of the variable that is missing
- Mustillo (2012, SMR, p.342)
  - Key factors
    - the magnitude of the correlation
    - the proportion of missingness
    - missingness pattern/type of missingness
    - the nature of auxiliary variables

Controversy

- Including auxiliary variables is useful with reducing the bias for NMAR?
  - The simulation showed that the inclusive strategy is to be greatly preferred. (Collins et al 2001, p.330) ;
  - The inclusion of auxiliary variables may not be necessary in many analytic situations. (Mustillo 2012, SMR, p.335)
    - Unrealistic too large correlations will be needed.
  - Auxiliary variables exhibit the surprising property of increasing bias in missing data problems.(Thoemmes and Rose 2014, MBR, p.443)
- No mathematical derivation has been given to compute the bias of estimators under a possibly misspecified missing-data mechanism, as long as I know.
  - i.e., MLE with AVs versus MLE without AVs under NMAR missingness

Purpose of this talk

- Difficulties arise when evaluating the bias, because
  - there are a wide variety of specifications of a missing-data mechanism (MDM);
  - difficult is computation of expectation of the terms involving MDM.
- Here we first take a simple model to compute the bias, where
  - MDM is almost arbitrary
  - A shared-parameter model for MDM is employed
    - e.g. a probit model rather than logit

Purpose of this talk

- Then applying the derived formulae to the simple model, we attempt to answer the questions bellow:
  - When inclusion of AVs reduces the bias for NMAR missingness, and how much?
  - What does "MAR becomes more plausible" mean with math words?
  - What does "proxy variable" and "reduction of uncertainty of missingness mean?
- Finally, a general theory to evaluate the bias of the MLE(FIML) will be developed under NMAR missingness.

A simple example and notation

X	Y	R
X_1	Y_1	1
$\vdots$	$\vdots$	$\vdots$
X_m	Y_m	1
X_{m+1}	missing	0
$\vdots$	$\vdots$	$\vdots$
X_n	missing	0

statistics of complete cases

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \quad \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

$$s_{xx} = \frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2$$

$$s_{yy} = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$$

$$s_{xy} = \frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})(Y_i - \bar{Y})$$

Full Information ML with X (observed likelihood)

A random sample on $\begin{bmatrix} Y \\ X \end{bmatrix}$:

$$\begin{bmatrix} Y_1 \\ X_1 \end{bmatrix}, \ldots, \begin{bmatrix} Y_m \\ X_m \end{bmatrix}, X_{m+1}, \ldots, X_n$$

FIML(Observed Likelihood):

$$L(\varphi|Y,X) = \prod_{i=1}^{m} N_2\left(\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \middle| \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} \sigma_{yy} & \sigma_{yx} \\ \sigma_{xy} & \sigma_{xx} \end{bmatrix}\right) \times \prod_{i=m+1}^{n} N(X_i|\mu_x, \sigma_{xx})$$

where $\varphi = (\mu_x, \mu_y, \sigma_{xx}, \sigma_{yy}, \sigma_{xy})$

MLE based on FIML

MLE for μ_y

$$\hat{\mu}_y = \bar{Y} + \frac{s_{xy}}{s_{xx}}(\hat{\mu}_x - \bar{X}),$$

$$\text{where} \quad \hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

e.g., Anderson (1957). JASA, 52, 200-203.

Full Information ML without X (observed likelihood)

A random sample on Y:

$$Y_1, \ldots, Y_m, Y_{mis}, \ldots, Y_{mis}$$

Observed Likelihood:

$$L(\varphi|Y) = \prod_{i=1}^{m} N(Y_i|\mu_y, \sigma_{yy})$$

MLE for μ_y:

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

Summary of the set-up

- $\mu_y := E[Y]$ is an interesting parameter.
- Data on Y involve NMAR missing values.
- We also have data on X which would not be used if data on Y were complete.
- X is said to be an auxiliary variable(AV).
- Should one include the data on X to estimate μ_y?
- FIML(OL) under normality

X	Y	R
X_1	Y_1	1
⋮	⋮	⋮
X_m	Y_m	1
X_{m+1}	missing	0
⋮	⋮	⋮
X_n	missing	0

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

versus

$$\hat{\mu}_y = \bar{Y} + \frac{s_{xy}}{s_{xx}}(\hat{\mu}_x - \bar{X})$$

Latent variable formulation for missing-data mechanism, without X

A shared-parameter model:

$Y \sim f(y|\theta)$, θ is a parameter of interest

$Y \perp\!\!\!\perp R|Z$ (conditional independence)

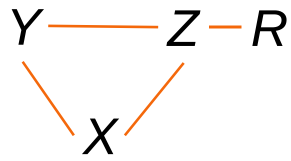
$R|Z = z \sim P(R = 1|z; \tau) = h(z|\tau)$

When $h(z|\tau) = \begin{cases} 1, & z \le \tau \quad (Y\text{: obs}) \\ 0, & z > \tau \quad (Y\text{: mis}) \end{cases}$,

it represents a probit model under normality, i.e.,

$$P(R = 1|Y = y) = \Phi(\alpha + \beta y)$$

Latent variable formulation for missing-data mechanism, with X



A shared-parameter model:

$(Y, X) \sim g(y, x|\theta, \theta_a)$, θ_a is a parameter of X

$(Y, X) \perp\!\!\!\perp R|Z$ (conditional independence)

$R|Z = z \sim P(R = 1|z; \tau) = h(z|\tau)$

Notice that

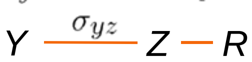
$$f(y|\theta) = \int g(y, x|\theta, \theta_a)dx$$

Bias of the MLEs under normality

- MLE without an AV

$$E[\bar{Y}] = E[Y|R = 1] = \mu_y + \frac{\sigma_{yz}}{\sigma_{zz}} E[Z - \mu_z|R = 1]$$

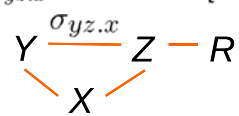
  - Bias increases linearly with σ_{yz} and with $E[Z - \mu_z|R = 1]$.



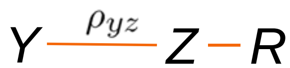
- MLE with an AV

$$E[\hat{\mu}_y] = \mu_y + \frac{\sigma_{yz.x} E[Z - \mu_z|R = 1]}{(1 - \rho_{xz}^2)\sigma_{zz} + \rho_{xz}^2 V(Z|R = 1)} + o(1)$$

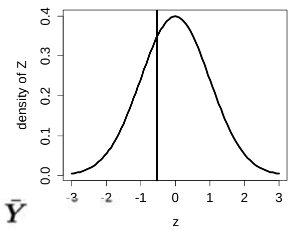
  - Bias increases linearly with $\sigma_{yz.x}$ and with $E[Z - \mu_z|R = 1]$.



Size of the bias standardized



$$\frac{E[\bar{Y}] - \mu_y}{\sigma_{yy}^{1/2}} = \rho_{yz} \times \frac{E[Z|R = 1] - \mu_z}{\sigma_{zz}^{1/2}}$$



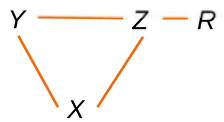
Size of the Standardized Bias of $\bar{Y}$

ρ_{yz}	Missing Rate (Bias for μ_z)					
	0% (0.00)	10% (0.19)	20% (0.35)	30% (0.50)	40% (0.64)	50% (0.80)
0.0	0.00	0.00	0.00	0.00	0.00	0.00
0.2	0.00	0.04	0.07	0.10	0.13	0.16
0.4	0.00	0.08	0.14	0.20	0.26	0.32
0.6	0.00	0.12	0.21	0.30	0.39	0.48

Effect of inclusion of an AV

$$\left|\frac{E[\hat{\mu}_y]-\mu_y}{E[\bar{Y}]-\mu_y}\right| = \left|\frac{\rho_{yz}-\rho_{yx}\rho_{xz}}{\rho_{yz}}\right| \frac{1}{(1-\rho_{xz}^2)+\rho_{xz}^2 B}$$

with $\quad B = \dfrac{V(Z|R=1)}{V(Z)}$



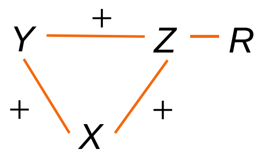
- AV does not necessarily reduce the bias of the MLE
  - When $|\rho_{yz}| > |\rho_{yz}-\rho_{yx}\rho_{xz}|$, AV could reduce the bias
  - When $|\rho_{yz}| < |\rho_{yz}-\rho_{yx}\rho_{xz}|$, AV could increase the bias

Effect of inclusion of an AV

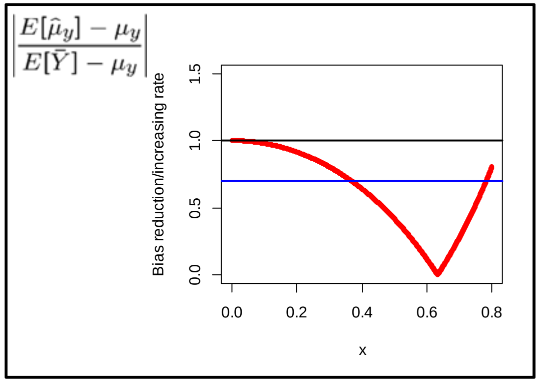
Bias reduction/increasing rate, when

$B=0.5,\ \rho_{yz}=0.4,\ x=\rho_{yx}=\rho_{zx}.$

One can get 30% bias reduction when $0.37 < x < 0.78$.



$$\begin{matrix} X & Y & Z \end{matrix}$$
$$\begin{bmatrix}1 & x & x\\ & 1 & 0.4\\ & & 1\end{bmatrix}$$



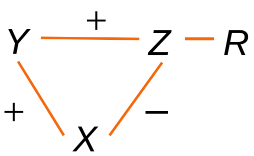
Effect of inclusion of an AV

Bias reduction/increasing rate, when

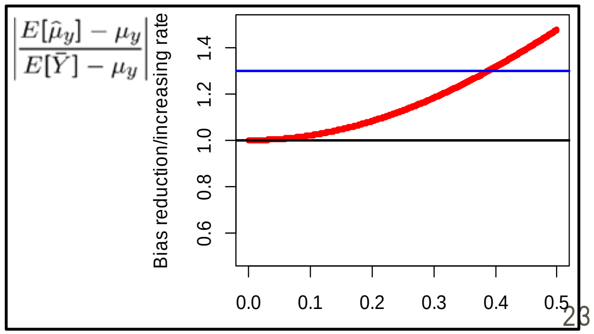
$B=0.5,\ \rho_{yz}=0.4,\ x=\rho_{yx}=-\rho_{zx}.$

The bias monotonically increases as x grows.

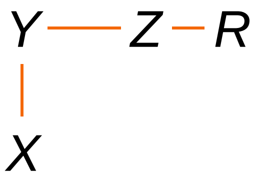
The bias enlarges by 30% when $x > 0.39$.



$$\begin{matrix} X & Y & Z \end{matrix}$$
$$\begin{bmatrix}1 & x & -x\\ & 1 & 0.5\\ & & 1\end{bmatrix}$$



In case of $X \perp\!\!\!\perp Z | Y$



Key relations: $\rho_{xz} = \rho_{xy}\rho_{yz}$

$$\left|\frac{E[\hat{\mu}_y]-\mu_y}{E[\bar{Y}]-\mu_y}\right| = \frac{1-\rho_{xy}^2}{(1-\rho_{xy}^2\rho_{yz}^2)+\rho_{xy}^2\rho_{yz}^2 B} \le 1$$

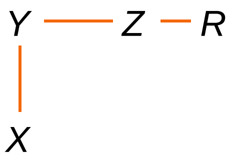
- AV always reduces the bias.
- AV can be called a proxy variable.
- Larger is Cor(Y,X), more bias reduction is obtained.

Effect of inclusion of an AV

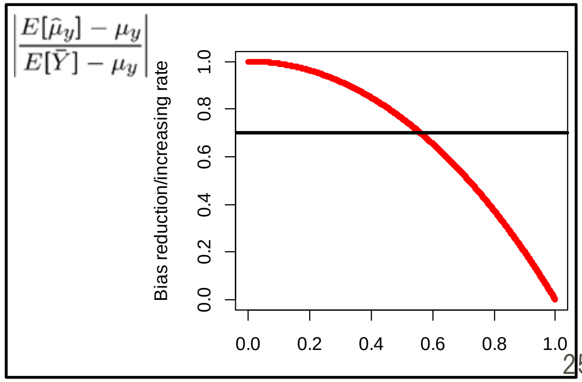
Bias reduction/increasing rate, when

$B = 0.5,\ \rho_{yz} = 0.4,\ x = \rho_{yx}.$

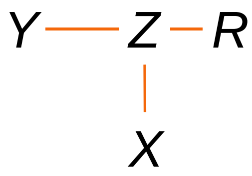
One can get 30% bias reduction when $x > 0.56$



$$\begin{array}{ccc} X & Y & Z \end{array}$$
$$\begin{bmatrix} 1 & x & 0.4x \\ & 1 & 0.4 \\ & & 1 \end{bmatrix}$$



In case of $X \perp\!\!\!\perp Y | Z$



Key relations: $\rho_{xy} = \rho_{xz}\rho_{yz}$

$$\left|\frac{E[\hat{\mu}_y] - \mu_y}{E[\bar{Y}] - \mu_y}\right| = \frac{1 - \rho_{xz}^2}{(1 - \rho_{xz}^2) + \rho_{xz}^2 B} \le 1$$

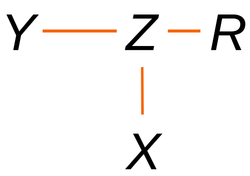
- AV always reduces the bias.
- AV can reduce variation of Z, so that connection between Y and R gets weak.
- Larger is Cor(X,Z), more bias reduction is obtained.

In case of $X \perp\!\!\!\perp Y | Z$

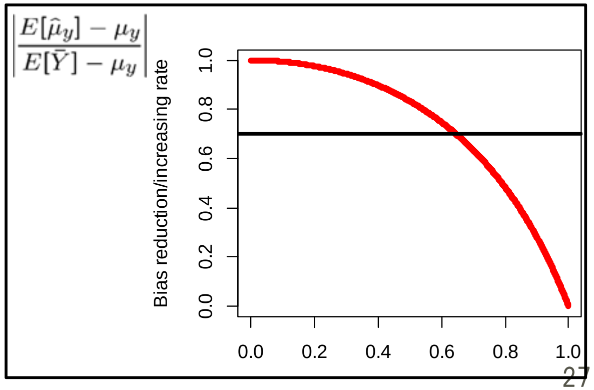
Bias reduction/increasing rate, when

$B = 0.5,\ \rho_{yz} = 0.4,\ x = \rho_{xz}.$

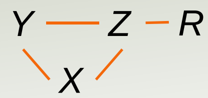
One can get 30% bias reduction when $x > 0.64$



$$\begin{array}{ccc} X & Y & Z \end{array}$$
$$\begin{bmatrix} 1 & 0.4x & x \\ & 1 & 0.4 \\ & & 1 \end{bmatrix}$$



Summary



- We have studied the effects of inclusion of an AV on the reduction of the bias of the MLE(FIML) in a simple setting.
- The bias can be expressed in the closed form:

$$E[\bar{Y}|R=1] - \mu_y = \frac{\sigma_{yz}}{\sigma_{zz}} E[Z - \mu_z | R = 1]$$

$$E[\hat{\mu}_y | R = 1] - \mu_y = \frac{\sigma_{yz.x} E[Z - \mu_z | R = 1]}{(1 - \rho_{xz}^2)\sigma_{zz} + \rho_{xz}^2 V(Z|R=1)} + o(1)$$

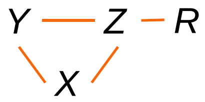
- The latent variable formulation of the missing-data mechanism enables us to successfully derive the simple but useful formulas.
- No particular missing-data mechanism is assumed.

Summary

- The four cases should be distinguished:
- Case I: $|\rho_{yz}| > |\rho_{yz} - \rho_{yx}\rho_{xz}|$

  $+ \quad + \; +$
  - AV makes MAR more plausible and reduces the bias.
  - In particular, if $\rho_{yz} - \rho_{yx}\rho_{xz} = 0$, the MAR holds for the model with AV.
- Case II: $|\rho_{yz}| < |\rho_{yz} - \rho_{yx}\rho_{xz}|$

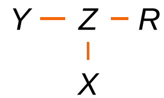
  $+ \quad + \; -$
  - AV enlarges the bias.



Summary

- Case III: $X \perp\!\!\!\perp Z | Y$

  - AV may be called a proxy variable for Y, and always reduces the bias.
  - Large correlation between X and Y is needed.
- Case IV: $X \perp\!\!\!\perp Y | Z$



  - AV reduces the variation of Z, and always reduces the bias.
  - Large correlation between X and Z is needed.

General framework to study the bias

Introduction of Approximate Population Bias

Purpose of this section

- The bias evaluation in the preceding section deals only with estimators expressed in a closed form:

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i, \qquad \hat{\mu}_y = \bar{Y} + \frac{s_{xy}}{s_{xx}}(\hat{\mu}_x - \bar{X})$$

- Here we discuss a method of deriving the bias of estimators implicitly defined, that is, those defined by optimizing functions or estimating equations

$$\hat{\theta} := \underset{\theta\in\Theta}{\text{argmax}}\; L(\theta|X), \qquad \frac{\partial}{\partial\theta}\log L(\theta|X)\Big|_{\theta=\hat{\theta}} = 0$$

Notation

Observable r.v.: $\boldsymbol{Y} = [Y_1, \ldots, Y_p]^T \sim f(\boldsymbol{y}|\boldsymbol{\theta}_0)$

Missing (response) indicator: $\boldsymbol{R} = [R_1, \ldots, R_p]^T$

$R_i = 1$ (or 0) $\Longleftrightarrow$ Y_i is observed (missing)

Missing-data pattern: $\boldsymbol{R} = \boldsymbol{r}^{(1)}, \ldots, \boldsymbol{r}^{(L)}, \quad (1 \le L \le 2^p)$

Joint distribution:

$$(\boldsymbol{Y}, \boldsymbol{R}) \sim g(\boldsymbol{y}, \boldsymbol{r}^{(\ell)}|\boldsymbol{\tau}_0, \boldsymbol{\theta}_0) = P\left(\boldsymbol{R} = \boldsymbol{r}^{(\ell)}|\boldsymbol{y}; \boldsymbol{\tau}_0, \boldsymbol{\theta}_0\right) f(\boldsymbol{y}|\boldsymbol{\theta}_0)$$

Observed components ($\boldsymbol{Y}_{obs}$):

$$D_{\boldsymbol{R}_i}\boldsymbol{Y}_i = D_{\boldsymbol{r}^{(\ell_i)}}\boldsymbol{Y}_i = \boldsymbol{Y}_i^{(\ell_i)} \quad \left(\boldsymbol{Y}_i = \left\{\boldsymbol{Y}_i^{(\ell_i)}, \boldsymbol{Y}_i^{(-\ell_i)}\right\}\right)$$

Notation

$D_{\boldsymbol{R}}$: Selection Matrix

(selecting observed variables)

$$\boldsymbol{Y}^{(\ell)} := D_{\boldsymbol{r}^{(\ell)}}\boldsymbol{Y} \ \left(= \boldsymbol{Y}_{obs}\right); \quad \boldsymbol{Y}^{(-\ell)} = \boldsymbol{Y}_{mis}$$

Example

$\boldsymbol{R} = [1, 0, 1, 0, \ldots, 0]'$

$$D_{\boldsymbol{R}}\boldsymbol{Y} = D_{[1,0,1,0,\ldots,0]'}\boldsymbol{Y}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix}$$

Likelihood Inference and MAR

Sample:

$(D_{\boldsymbol{R}_1}\boldsymbol{Y}_1, \boldsymbol{R}_1), \ldots, (D_{\boldsymbol{R}_n}\boldsymbol{Y}_n, \boldsymbol{R}_n)$

$(\boldsymbol{Y}_1^{(\ell_1)}, \boldsymbol{R}_1), \ldots, (\boldsymbol{Y}_n^{(\ell_n)}, \boldsymbol{R}_n)$

Direct Likelihood (Observed Likelihood):

$$DL_n(\boldsymbol{\theta}) = f(\boldsymbol{Y}_{obs}) = \prod_{i=1}^{n} f_{\boldsymbol{r}^{(\ell_i)}}(\boldsymbol{Y}_i^{(\ell_i)}|\boldsymbol{\theta})$$

MAR:

$$P\left(\boldsymbol{R} = \boldsymbol{r}^{(\ell)}\middle|\boldsymbol{Y}^{(\ell)}, \boldsymbol{Y}^{(-\ell)}\right) = P\left(\boldsymbol{R} = \boldsymbol{r}^{(\ell)}\middle|\boldsymbol{Y}^{(\ell)}\right)$$

$(\ell = 1, \ldots, L)$

Direct MLE and its limit

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}}{\text{argmax}}\ f(\boldsymbol{Y}_{obs}) = \underset{\boldsymbol{\theta}}{\text{argmax}} \prod_{i=1}^{n} f_{\boldsymbol{r}^{(\ell_i)}}(D_{\boldsymbol{r}^{(\ell_i)}}\boldsymbol{Y}_i|\boldsymbol{\theta})$$

$$\tilde{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}}{\text{argmax}}\ \boxed{E[\log f_{\boldsymbol{R}}(D_{\boldsymbol{R}}\boldsymbol{Y}|\boldsymbol{\theta})|\boldsymbol{\tau}_0, \boldsymbol{\theta}_0]} = \rho(\boldsymbol{\theta})$$

$\hat{\boldsymbol{\theta}} \xrightarrow{P} \tilde{\boldsymbol{\theta}}$

$$\rho(\boldsymbol{\theta}) = \sum_{\ell=1}^{L} \int \log f_{\boldsymbol{r}^{(\ell)}}(\boldsymbol{y}^{(\ell)}|\boldsymbol{\theta}) \times P(\boldsymbol{r}^{(\ell)}|\boldsymbol{y}^{(\ell)}; \boldsymbol{\tau}_0, \boldsymbol{\theta}_0) f_{\boldsymbol{r}^{(\ell)}}(\boldsymbol{y}^{(\ell)}|\boldsymbol{\theta}_0) d\boldsymbol{y}^{(\ell)}$$

where $\boldsymbol{y}^{(\ell)} = D_{\boldsymbol{r}^{(\ell)}}\boldsymbol{y}$

Direct MLE and its limit

For Not MAR missingness,

$$\left.\frac{\partial \rho(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \neq 0$$

Under some regularity conditions, we have

$$\left.\frac{\partial \rho(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = E\left[\left.\frac{\partial \log f_{\boldsymbol{R}}(D_{\boldsymbol{R}}\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|\boldsymbol{\tau}_0, \boldsymbol{\theta}_0\right]\Bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = 0,$$

and

$$\tilde{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$$

Approximate Population Bias (APB)

Taylor approximation:

$$0 = \frac{\partial \rho(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \approx \frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

Approximate Population Bias (APB):

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \approx -\left(\frac{\partial^2 \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right)^{-1} \frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

$$= E\left[\left.\frac{-\partial^2 \log f_{\boldsymbol{R}}(D_{\boldsymbol{R}}\boldsymbol{Y}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right|\boldsymbol{\tau}_0, \boldsymbol{\theta}_0\right]^{-1} E\left[\left.\frac{\partial \log f_{\boldsymbol{R}}(D_{\boldsymbol{R}}\boldsymbol{Y}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right|\boldsymbol{\tau}_0, \boldsymbol{\theta}_0\right]$$

$$= I(\boldsymbol{\theta}_0)^{-1} \frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

$$=: \text{bias}(\tilde{\boldsymbol{\theta}})$$

Approximate Population Bias (APB)

$$\text{APB}^2 := ||\text{bias}(\tilde{\boldsymbol{\theta}})||^2 = \left\| I(\boldsymbol{\theta}_0)^{-1} \frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\|^2$$

$$= \left(\frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}^T I(\boldsymbol{\theta}_0)^{-1}\right) I(\boldsymbol{\theta}_0) \left(I(\boldsymbol{\theta}_0)^{-1} \frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)$$

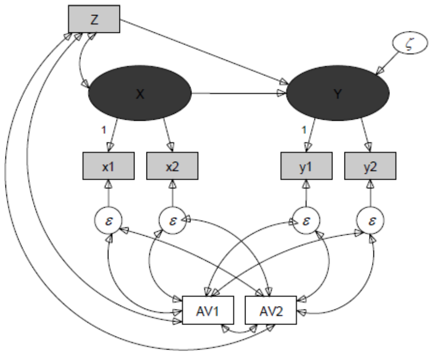
$$= \left(\frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^T I(\boldsymbol{\theta}_0)^{-1} \frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}},$$

provided that $I(\boldsymbol{\theta}_0) > 0$.

- Basically the APB can be calculated for any estimates once a MDM and population values are given.
- It is yet not certain to prove general properties with the use of APB.
- We hope to show in the future that
  - Saturated-correlated model reduces the bias.
  - Extra dependent variable model reduces the bias.

Saturated Correlates Model

- Graham and Coffman (2012) and Graham (2003)



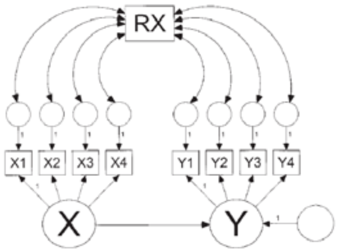
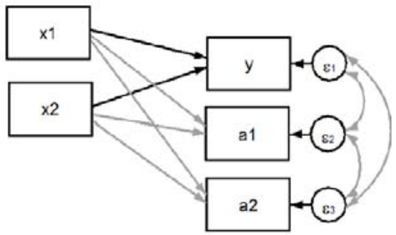
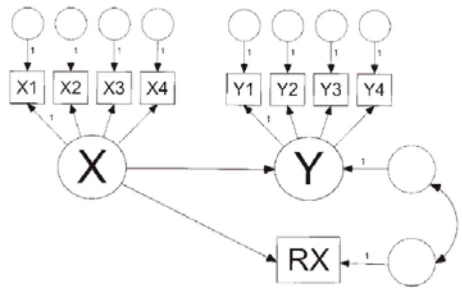


FIGURE 6 Model 3: "saturated correlates" model (latent variable version).

Extra DV model

- Graham and Coffman (2012) and Graham (2003)

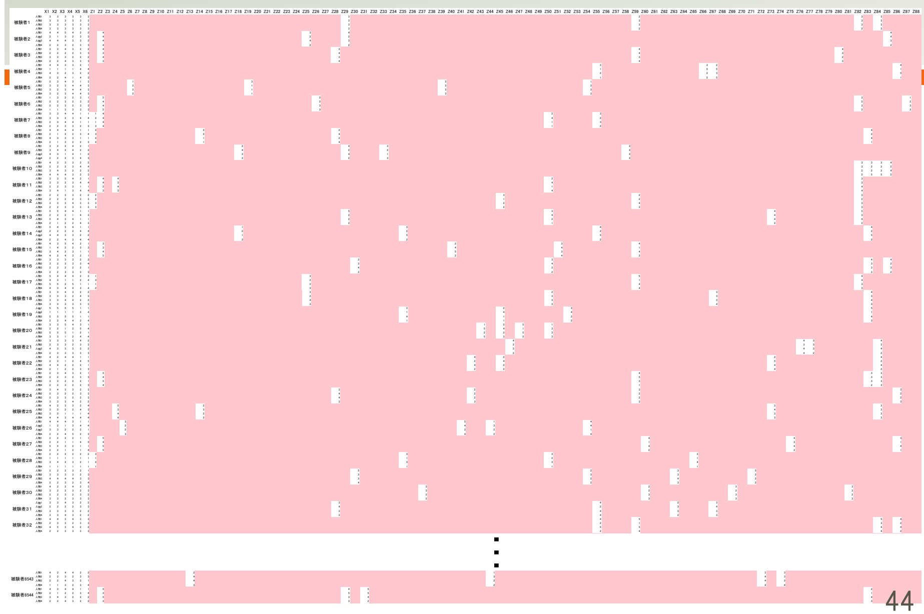




Part II.
A huge amount of missing values

Osaka University and NTT Network Innovation Laboratories

An example of a huge amount of missing values



Uninteresting items and/or too many items

- One should avoid responders to be forced to respond
  - too many questionnaire items
  - questionnaire Items that do not interest responders;
- That will cause the troubles:
  - Nonresponses (missing values)
  - One particular choice in many items, such as the neutral response
  - Random choice in many items
  - …

Real data example: Exploring the structure of first expressions of people

- A variety of questionnaire items can be considered to assess first impressions of people
- Suitable items may depend on responders
- One strategy to handle the situation is to prepare many items in a questionnaire, so that a responder can select several items that interest the responder
- Then, missing values take place for the items unselected
- We prepared 94 items regarding first expressions, among which 6 items are common to all responders, and the responders are requested to choose 4 items among the other 88 items
  - Observed variables are 10 in number
  - Missing variables are 84 in number

Description of the questionnaire data

- After selecting the 4 items, responders see virtual people stimuli such as photos, descriptions and twitter messages
- Assess the people by the 10 items with 5-point scale
- Common items (6 items)
  - pleasant-unpleasant, friendly-unfriendly, careful-hasty
  - sensible-insensible, active-passive, confident-unconfident
- Selective items (88 items)
  - Laid-Back-Rash, Frank-Formal, Incompetent-Competent
  - Mean-Nice, Disgusting-Delightful, Acid-Round,
  - Bad Feeling-Good Feeling, Serious-Frivolous
  - Simple-Complex, Neat-Untidy, ……….

People stimuli (in Japanese)

Survey data collection and results

- Web-based survey was conducted
  - December 28, 2011 to January 10, 2012
  - N=2362
- The following three factors were expected and actually were identified for the 94 items:
  - F1: Personality (30 items)
  - F2: Intelligence (33 items)
  - F3: Activeness (31 items)

F1: Personality (30 items)

SD Adjective Pair	SD項目	F1	F2	F3
91.Friend – Enemy	91.仲間である–敵である	0.88	-0.17	0.11
88.Feel at Ease – Frustratir	88.安らげる–いらいらする	0.80	0.03	0.01
84.Similar to Myself – Diffe	84.自分に似ている–自分	0.80	-0.20	0.00
85.Agree with Each Other –	85.話が合う–話が合わな	0.80	-0.09	0.01
93.Friendly – Unfriendly	93.親しみやすい–親しみ	0.79	-0.13	0.21
68.Soft – Hard	68.やわらかい–かたい	0.77	-0.05	0.02
67.Patient – Impatient	67.気長な–短気な	0.76	-0.30	-0.01
87.Empathetic – Lack Empa	87.共感できる–共感でき	0.75	0.10	-0.02
61.Kind – Unkind	61.親切な–不親切な	0.73	0.09	-0.03
92.Pleasant – Unpleasant	92.感じのよい–感じの悪い	0.73	0.12	0.12
13.Wan – Robust	13.弱々しい–たくましい	0.72	-0.49	-0.16
50.Modest – Immodest	50.控えめな–でしゃばりな	0.68	0.16	-0.38
86.Same Ways of Thinking	86.考え方が合う–考え方	0.68	0.12	-0.12
63.Warm – Cold	63.あたたかい–つめたい	0.64	0.15	-0.24
33.Bad Feeling – Good Fee	33.気持ち悪い–気持ち良	-0.64	-0.12	0.08

F2: Intelligence (33 items)

SD Adjective Pair	SD項目	F1	F2	F3
40.Simple – Complex	40.単純な–複雑な	-0.06	0.86	-0.28
12.Weak – Strong	12.弱い–強い	0.08	-0.85	0.32
29.Serious – Frivolous	29.真面目な–不真面目な	0.20	0.80	-0.07
73.Forgetful – Long-Memo	73.忘れっぽい–物覚えの	0.45	-0.79	-0.23
26.Neat – Untidy	26.きちんとしている–だら	0.00	0.77	-0.01
43.Disorganized – Organize	43.いい加減な–几帳面な	0.21	-0.74	0.13
42.New – Old	42.新しい–古い	0.47	-0.72	0.38
59.Intellectual – Sensuous	59.理知的な–感覚的な	-0.32	0.64	0.17
72.Logical – Emotional	72.論理的な–感情的な	-0.08	0.64	0.14
11.Short – Tall	11.背が低い–背が高い	-0.09	-0.64	-0.21
27.Elegant – Ungracious	27.上品な–下品な	0.19	0.62	-0.04
95.Sensible – Insensible	95.分別のある–分別のな	0.29	0.62	-0.04
36.Careful – Careless	36.注意深い–不注意な	0.10	0.62	-0.10
32.Responsible – Irresponsi	32.責任感の強い–無責任	0.13	0.59	0.12
94.Careful – Hasty	94.慎重な–軽率な	0.28	0.58	-0.14

F3: Activeness (31 items)

SD Adjective Pair	SD項目	F1	F2	F3
17.Strong – Weak	17.強気な–弱気な	-0.18	0.07	0.95
52.Exhibitionist – Quiet	52.目立ちたがり–大人しい	-0.15	-0.08	0.85
39.Skeptical – Credulous	39.懐疑的な–信じやすい	0.02	-0.69	0.84
47.Extrovert – Introvert	47.外向的な–内向的な	-0.12	-0.02	0.81
96.Active – Passive	96.積極的な–消極的な	-0.22	0.11	0.80
46.Loud – Quiet	46.にぎやかな–静かな	0.32	0.08	0.78
20.Bold – Timid	20.大胆な–小心な	0.15	-0.58	0.77
14.Healthy – Sickly	14.元気な–病弱な	0.02	0.02	0.75
97.Confident – Unconfident	97.自信のある–自信のな	-0.26	0.22	0.74
81.Superior – Inferior	81.優れている–劣ってい	-0.11	0.19	0.69
41.Clear – Vague	41.はっきりした–ぼんやりし	-0.16	0.29	0.68
55.Cheerful – Dismal	55.陽気な–陰気な	0.40	-0.17	0.66
53.Bright – Dark	53.明るい–暗い	0.29	-0.36	0.66
03.Sober – Flashy	03.地味な–派手な	0.12	0.60	-0.63
21.Masculine – Feminine	21.男性的な–女性的な	-0.20	0.51	0.62

Factor analysis by FIML with missing values

Exploratory factor analysis model

$$\underset{p\times 1}{\boldsymbol{Y}} = \underset{p\times 1}{\boldsymbol{\mu}} + \underset{p\times m}{\Lambda}\underset{m\times 1}{\boldsymbol{f}} + \underset{p\times 1}{\boldsymbol{u}}$$

$$\boldsymbol{f} \sim N_m(\boldsymbol{0}, \Phi), \quad \boldsymbol{u} \sim N_p(\boldsymbol{0}, \Psi)$$

$$\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \Lambda\Phi\Lambda' + \Psi)$$

FIML(or observed likelihood) under normality

$$\boldsymbol{Y}_k = [\underset{p^{(\ell_k)}\times 1}{\overset{\text{obs}}{\boldsymbol{Y}^{(\ell_k)}}}, \underset{p^{(-\ell_k)}\times 1}{\overset{\text{mis}}{\boldsymbol{Y}^{(-\ell_k)}}}], \quad k = 1, \ldots, n$$

$$E[\boldsymbol{Y}^{(\ell_k)}] = \boldsymbol{\mu}^{(\ell_k)}, \quad \text{Var}[\boldsymbol{X}^{(\ell_k)}] = \Lambda^{(\ell_k)}\Phi\Lambda^{(\ell_k)\prime} + \Psi^{(\ell_k)} = \Sigma^{(\ell_k)}$$

$$L(\boldsymbol{\mu}, \Lambda, \Psi | Y) = \prod_{k=1}^{n} N_{p^{(\ell_k)}}(\boldsymbol{Y}^{(\ell_k)} | \boldsymbol{\mu}^{(\ell_k)}, \Sigma^{(\ell_k)})$$

FIML can be applied for MAR missingness

A model for observable variables

$$Y_i = R_i\left(\mu_i + \sum_{j=1}^{m}\lambda_{ij}F_j + u_i\right) + (1 - R_i)Z_i, \quad i = 1, \ldots, p$$

- $R_i = 1$
  - A responder IS interested in item X_i and can assess with it appropriately
  - (S)he selects Y_i, so that Y_i is observed
- $R_i = 0$
  - A responder IS NOT interested in the item Y_i and his or her response will not follow according to the FA model
  - Z denotes a rv representing such a response
  - (S)he does not select Y_i, so that Y_i is missing

$$Y_i = R_i\left(\mu_i + \sum_{j=1}^{m}\lambda_{ij}F_j + u_i\right) + (1 - R_i)Z_i, \quad i = 1, \ldots, p$$

The marginal distribution for $\boldsymbol{Y}$:

$$\boldsymbol{Y} \sim P(\boldsymbol{R} = \boldsymbol{1}_p)N_p(\boldsymbol{\mu}, \Sigma) + \cdots + P(\boldsymbol{R} = \boldsymbol{0})F_z$$

NMAR missingness:

$$f(\boldsymbol{R}|\boldsymbol{Y}) \neq f(\boldsymbol{R}|\boldsymbol{Y}_{obs})$$

For the first impression data,

$$P(\boldsymbol{R} = \boldsymbol{1}_p) = 0$$

$$Y_i = R_i\left(\mu_i + \sum_{j=1}^{m}\lambda_{ij}F_j + u_i\right) + (1 - R_i)Z_i, \quad i = 1, \ldots, p$$

Assumptions:

$\boldsymbol{R} = (R_1, \ldots, R_p) \perp\!\!\!\perp (\boldsymbol{f}, \boldsymbol{u})$

$P(R_1, \ldots, R_p)$ is unrelated to $(\mu_i, \lambda_{ij}, \psi_{ii})$'s

Likelihood:

$$L = f(\boldsymbol{Y}_{obs}, \boldsymbol{R}) = f(\boldsymbol{Y}_{obs}|\boldsymbol{R})P(\boldsymbol{R}) \propto f(\boldsymbol{Y}_{obs}|\boldsymbol{R})$$

$$= \prod_{k=1}^{n} N_{p^{(\ell_k)}}(\boldsymbol{Y}^{(\ell_k)} | \boldsymbol{\mu}^{(\ell_k)}, \Sigma^{(\ell_k)})$$

$=$ FIML(Observed Likelihood)

Technical aspects

- Simulation studies showed the following:
  - The EM algorithm with only common factors as missing works well
  - Sample sizes should be more than a few thousands to obtain stable estimates for p=90, m=3, missing rate=90%
  - Asymptotic standard errors appear to be accurate when sample sizes are as large as 10,000 for p=90, m=3, missing rate=90%
  - Introduction of some common items can allow us to estimate more stably than analysis of all selective items

Summary

- Usually it would be difficult to extract useful information from data when the missing rate is 90%
- The analysis of the First Impression Data with FIML (OL) would be successful, because
  - responders select questionnaire items that they can evaluate, so that the assumption of a factor analysis model would be reasonable for the items selected;
  - introduce common items to all responders;
  - the sample size is enough large;
  - the EM algorithm with only common factors as missing works well

Thank you for your attention