

NMAR の下での尤度法

狩野 裕*

Likelihood Method in NMAR Missingness

Yutaka Kano*

欠測値データ解析において欠測メカニズムが MAR(missing at random) でないとき統計的推測は難しい。典型的な方法は欠測メカニズムを組み込んだ尤度による最尤法を適用することであるが、この方法には理論的にも数値的にも困難が生じることが多い。そこで、本論文では欠測メカニズムを明示的に用いずに推測する方法を議論する。まず最初に、潜在変数モデルにおいて、欠測メカニズムが潜在変数に依存するという MAR でない欠測において、欠測メカニズムを用いない修正直接最尤法を提案する。同方法は全てのパラメータの一致推定量を構成する。

つぎに MAR 条件の緩和を試みる。MAR は直接最尤推定量が一致性を有するための十分条件に過ぎない。数学的により弱い条件を提出しその性質を調べる。

It is difficult to draw statistical inference for data with missing values when the missing-data mechanism is not MAR (missing at random). Typically one can apply the method of likelihood with a missing-data mechanism to analyze such data. The likelihood method, however, encounters difficulties with identifying an appropriate model for the missing-data mechanism and with model identification or numerical optimization. Thus, we discuss the estimation problem without explicit use of missing-data mechanisms for not MAR missingness. First, a modified direct likelihood method is proposed to estimate a linear latent variate model with not MAR missingness. The method is shown to provide consistent estimators for all model parameters.

The MAR is known to be merely a sufficient condition for the direct likelihood MLE to be consistent. Weakening the MAR condition is explored, where the consistency still holds. Several properties of the weakened MAR condition are discussed.

キーワード: 欠測メカニズム, 情報量不等式, 一致推定量, 因果推論の基本的問題, 条件付き独立, 強い意味での無視可能

1. はじめに

不完全データとは、当初想定した形で得られないデータのことをいい、いくつかのパターンがある。典型的には、欠測値（欠損値）データ (missing data), そして、打ち切りや連

* 大阪大学大学院基礎工学研究科数理科学領域: 〒 560-8531 豊中市待兼山町 1-3 (E-mail: kano@sigmath.es.osaka-u.ac.jp).

続データのグループ化など精度の悪いデータ (coarse data) がある (e.g., Schafer (1997), 岩崎 (2002)).

欠測値データ解析においては観測変数が与えられた下での欠測指標の条件付き確率—欠測メカニズムという—が重要な役割を果たす。しかし、まず第一に、欠測メカニズムを正しくモデリングすることが困難であることが多く、もしモデルを誤特定したならば推定量に無視できないバイアスが生じ得るという問題がある (e.g., Tang *et al.* (2003, p. 750)). 第二に、たとえ欠測メカニズムがモデル化できたとしても推定に困難が生じる。すなわち、欠測メカニズムを含めたモデルが識別可能でないことがある (e.g., Ibrahim *et al.* (2001)). 識別できたとしても識別が弱いことがある (e.g., 森川・狩野 (2013)). こういった状況の下では、もし可能であるならば、欠測メカニズムを用いることなく興味あるパラメータを推測することが望ましい。そのための数学的条件が MAR (missing at random) である (Rubin (1976), Little and Rubin (2002))¹⁾。しかし、MAR は厳しい条件で、常に適用できるとは限らない。なお、MAR が成立しないとき、欠測メカニズムは NMAR (Not MAR) といわれる。

このような問題に対していくつかのアプローチがある。一つ目は、欠測メカニズムを用いず、尤度を修正したり、完全データ解析 (complete case analysis, CCA) を用いたりして、NMAR の欠測データの下で興味あるパラメータの一致推定量を構成するものである。たとえば Chen (2001) や Tang *et al.* (2003) がある。Allison (2002) は特に回帰分析において CCA を推奨している。確かに、尤度法は不適だが CCA が一致推定量を構成する場合がある。しかし、完全データ (complete case) が多くない場合、CCA の推定効率が大幅にダウンすることを注意しておかなければならない。本論文では第 3 節で、潜在変数モデルにおいて、NMAR である欠測メカニズムの下で一致推定が可能となる、欠測メカニズムを用いない修正尤度法を提案する。

二つ目のアプローチは MAR 条件の緩和であろう。MAR は一致推定のための十分条件に過ぎず、数学的に MAR 条件を緩和することが可能である。本論文第 4 節では具体的な条件を提示しその性質を調べる。

その他のアプローチとして、NMAR の下での欠測メカニズムを用いない MLE に生じるバイアスを理論的に評価し、バイアスを減少させるための手立てを講じる方法がある。バイアス減少のための手段として補助変数法 (e.g., Ibrahim *et al.* (2001), Graham (2003)) がある。また、実際的な方法として感度分析がある (e.g., Daniels and Hogan (2008)). これらのアプローチは別の機会に議論したい。

¹⁾ 粗いデータの場合は CAR (coarsened at random) といわれる (Heitjan and Rubin (1991), Tsiatis (2006)).

2. 記号・定義と最尤推定

本節では本論文を通じて用いる定義と記号を示す. p 次元の観測変数の確率ベクトルを $\mathbf{Y} = [Y_1, \dots, Y_p]'$ とし, その値域を $\mathcal{Y} (\subset \mathbb{R}^p)$ とする. $\mathbf{Y}$ の観測 (欠測) 指標ベクトルを $\mathbf{R} = [R_1, \dots, R_p]'$ とする. すなわち, Y_i が欠測 (観測) のとき $R_i = 0$ (1) である. $\mathbf{Y}$ の欠測パターンの総数を L ($1 \leq L \leq 2^p$) とし, $\mathbf{R}$ の取り得る値を $\mathbf{r}^{(\ell)}$ ($1 \leq \ell \leq L$) とする. たとえば, $\mathbf{r}^{(\ell)} = [1, \dots, 1]'$ は完全データ (complete case) を表す.

$D_{\mathbf{r}^{(\ell)}}$ を ℓ 番目の欠測パターンにおいて観測される変数を選択する行列とする. すなわち, $D_{\mathbf{r}^{(\ell)}} \mathbf{Y}$ は ℓ 番目の欠測パターンにおける観測変数ベクトルである. $D_{\mathbf{r}^{(\ell)}} \mathbf{Y}$ を $\mathbf{Y}^{(\ell)}$ と書くことがある. また, ℓ 番目の欠測パターンにおける欠測変数ベクトルを $\mathbf{Y}^{(-\ell)}$ と書く. 集合の意味で $\mathbf{Y} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}) = (\mathbf{Y}^{(\ell)}, \mathbf{Y}^{(-\ell)})$ ($\ell = 1, \dots, L$) である²⁾. なお, 確率ベクトルの実現値についても同様に $\mathbf{y}^{(\ell)}, \mathbf{y}^{(-\ell)}$ なる記号を用いる.

$(\mathbf{Y}, \mathbf{R})$ の同時分布を

$$P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{y}; \boldsymbol{\varphi}) f(\mathbf{y} | \boldsymbol{\theta}), \quad \ell = 1, \dots, L, \mathbf{y} \in \mathcal{Y} \quad (2.1)$$

$$\boldsymbol{\theta} \in \Theta (\subset \mathbb{R}^q), \quad \boldsymbol{\varphi} = [\boldsymbol{\theta}', \boldsymbol{\tau}']' \in \Xi (\subset \mathbb{R}^{q+r})$$

と書く. $\mathbf{Y}$ の値域 $\mathcal{Y}$ は $\mathbb{R}^p$ において縮退せず, 確率密度 $f(\mathbf{y} | \boldsymbol{\theta})$ は $\mathcal{Y} \times \Theta$ において正値をとり, $\mathcal{Y}$ は $\boldsymbol{\theta}$ に依存しないものとする. 本論文では $f(\mathbf{y} | \boldsymbol{\theta})$ はルベグ測度に対する微分とするが, 以下の議論は離散分布等でも成立する.

$\mathbf{Y}$ の部分である $\mathbf{Y}^{(\ell)}$ の分布を

$$f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)} | \boldsymbol{\theta}) = \int f(\mathbf{y} | \boldsymbol{\theta}) d\mathbf{y}^{(-\ell)}, \quad \ell = 1, \dots, L$$

と書く. なお, $\mathbf{Y}^{(\ell)}$ が完全データのときは積分を施さないものとする.

条件付き確率

$$P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y} = \mathbf{y}; \boldsymbol{\varphi}), \quad \ell = 1, \dots, L \quad (2.2)$$

$$P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y}^{(\ell)} = \mathbf{y}^{(\ell)}; \boldsymbol{\varphi}), \quad \ell = 1, \dots, L \quad (2.3)$$

を考える. (2.2) を欠測メカニズム (missing-data mechanism) という. 曖昧さが残るが, ときに (2.3) も欠測メカニズムとよばれることがある.

欠測メカニズムが MAR (missing at random) であるとは,

$$P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y} = \mathbf{y}; \boldsymbol{\varphi}) = P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y}^{(\ell)} = \mathbf{y}^{(\ell)}; \boldsymbol{\varphi}),$$

$$\ell = 1, \dots, L, \mathbf{y} \in \mathcal{Y}, \boldsymbol{\varphi} \in \Xi \quad (2.4)$$

²⁾ 観測値を表す $\mathbf{y}_{\text{obs}}$ や欠測値を表す $\mathbf{y}_{\text{mis}}$ はしばしば用いられ直感的に分かりやすい. しかし, 数学的に曖昧さがあ本論文の議論に不便であるから用いない (e.g., Seaman *et al.* (2013)).

が成立するときに言う (e.g., Rubin (1976), Little and Rubin (2002), 岩崎 (2002))³⁾. MAR が成立しないとき, NMAR (not MAR), または MNAR (missing not at random) という. MAR かつ (2.2) の条件付き確率が θ に依存しないとき⁴⁾, 欠測メカニズムは無視可能 (ignorable) といわれる (e.g., Little and Rubin (2002, p. 119), Lu and Copas (2004)).

MAR は, $\mathbf{Y}^{(\ell)}$ を与えた下で $\mathbf{R}$ と $\mathbf{Y}^{(-\ell)}$ とが条件付き独立であることを一般に意味しないことに注意する. というのは, $\mathbf{r}^{(\ell)}$ と $\mathbf{Y}^{(\ell)}$ が連動しているからである. いくつかの文献において混乱が見られる (e.g., Seaman *et al.* (2013, p. 260)).

なお, $L = 2$ のときは例外的に条件付き独立が成立する. 実際, MAR 条件

$$P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y} = \mathbf{y}; \boldsymbol{\varphi}) = P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y}^{(\ell)} = \mathbf{y}^{(\ell)}; \boldsymbol{\varphi}), \quad \mathbf{y} \in \mathcal{Y}, \boldsymbol{\varphi} \in \Xi$$

が成り立てば, 両辺を 1 から減じることで,

$$P(\mathbf{R} = \mathbf{r}^{(\ell')} | \mathbf{Y} = \mathbf{y}; \boldsymbol{\varphi}) = P(\mathbf{R} = \mathbf{r}^{(\ell')} | \mathbf{Y}^{(\ell)} = \mathbf{y}^{(\ell)}; \boldsymbol{\varphi}), \quad \mathbf{y} \in \mathcal{Y}, \boldsymbol{\varphi} \in \Xi$$

を確認できる. ここで $(\ell, \ell') = (1, 2)$ or $(2, 1)$ である. それゆえ

$$\mathbf{R} \perp\!\!\!\perp \mathbf{Y}^{(-\ell)} | \mathbf{Y}^{(\ell)}, \quad \ell = 1, 2 \quad (2.5)$$

を得る⁵⁾.

$\mathbf{Y}$ と $\mathbf{R}$ をつなぐ潜在変数ベクトル $\mathbf{Z}$ を導入し, $\mathbf{Z}$ を与えた下で $\mathbf{Y}$ と $\mathbf{R}$ とが条件付き独立であることを仮定する. $\mathbf{Z}$ を与えた下での $(\mathbf{Y}, \mathbf{R})$ の分布を $P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{z}) f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} | \mathbf{z})$ とし, $\mathbf{Z}$ の周辺分布を $f_{\mathbf{Z}}(\mathbf{z})$ とする. なお, ここではパラメータを省略している. このとき, $(\mathbf{Y}, \mathbf{R})$ の分布は,

$$\int P(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{z}) f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} | \mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}, \quad \ell = 1, \dots, L, \mathbf{y} \in \mathcal{Y} \quad (2.6)$$

と表すことができる. このモデルを shared-parameter model と言い, 一般に, その欠測メカニズムは MAR でないことが知られている (e.g., Albert and Follman (2009), Kano and Takai (2011)). Creemers *et al.* (2011) は shared-parameter model の拡張モデルを定義している. shared-parameter model による欠測メカニズムは潜在変数モデルと整合性が高い. 本論文では第 3 節で潜在変数モデルを取り上げる.

母集団分布 (2.1) からのサイズ n の独立同一標本を $(\mathbf{Y}_1, \mathbf{R}_1), \dots, (\mathbf{Y}_n, \mathbf{R}_n)$ と書く. 欠測を含む標本は $(D_{\mathbf{R}_1} \mathbf{Y}_1, \mathbf{R}_1), \dots, (D_{\mathbf{R}_n} \mathbf{Y}_n, \mathbf{R}_n)$ と書くことができる.

³⁾ MLE が真値に確率収束するための条件としては, $\boldsymbol{\varphi}$ が真値一点においてこの等式が成立することで十分である. しかし, 真値は未知であることから, 統計学の慣例を鑑み $\boldsymbol{\varphi}$ はパラメータ空間 Ξ の任意の点とした.

⁴⁾ parameter separation または parameter distinctness が成立するという. 本論文では無視可能の概念は重要でなく, したがって, parameter separation は仮定しない.

⁵⁾ $\perp\!\!\!\perp$ は Dawid の記号と言われ, (条件付き) 独立の関係を表す (e.g., Lauritzen (1996, p. 28), Dawid (1979)).

二つの条件付き確率を

$$\begin{aligned} m(\mathbf{r}^{(\ell)}|\mathbf{y};\varphi) &= P\left(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y} = \mathbf{y}; \varphi\right), \quad \ell = 1, \dots, L \\ m_{\mathbf{r}^{(\ell)}}(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)};\varphi) &= P\left(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y}^{(\ell)} = \mathbf{y}^{(\ell)}; \varphi\right) \\ &= \int m(\mathbf{r}^{(\ell)}|\mathbf{y};\varphi) \frac{f(\mathbf{y}|\varphi)}{f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\varphi)} d\mathbf{y}^{(-\ell)}, \quad \ell = 1, \dots, L \end{aligned}$$

と書く.

これらの記号の下で, 下記 2 種類の尤度を定義する.

$$FL_n(\varphi) = \prod_{i=1}^n m_{\mathbf{R}_i}(\mathbf{R}_i | D_{\mathbf{R}_i} \mathbf{Y}_i; \varphi) f_{\mathbf{R}_i}(D_{\mathbf{R}_i} \mathbf{Y}_i | \theta) \quad (2.7)$$

$$DL_n(\theta) = \prod_{i=1}^n f_{\mathbf{R}_i}(D_{\mathbf{R}_i} \mathbf{Y}_i | \theta) \quad (2.8)$$

ここで, $D_{\mathbf{R}_i} \mathbf{Y}_i$ が空のときは, $f_{\mathbf{R}_i}(D_{\mathbf{R}_i} \mathbf{Y}_i | \theta) = 1$ と約束する. (2.7) は欠測メカニズムを含めた尤度であり, Full Likelihood とよぶことにする. (2.8) は欠測メカニズムを含めない尤度で Direct Likelihood—直接尤度—という. なお, 後者は Observed-Data Likelihood, Full-Information Likelihood とよばれることがある (e.g., Arbuckle (1996), Schafer (1997)). 欠測メカニズムを正しく特定できれば, 識別性などの正則条件の下で, $FL_n(\varphi)$ を最大にする解 $\hat{\varphi}$ は真値 φ の一致推定量となる. また, 条件 MAR の下では, 識別性などの正則条件の下で, $DL_n(\theta)$ を最大にする解 $\hat{\theta}$ は真値 θ の一致推定量となる (Rubin (1976), Little and Rubin (2002), 岩崎 (2002), Takai and Kano (in press)). なお, 本論文では, (2.7) または (2.8) を最大にする解として定められる推定量を MLE とよぶことにする.

3. 潜在変数モデルの欠測値問題

潜在変数 (latent variable) を扱うモデルは広範で有用である. 潜在変数モデルにはいくつかのルーツがある. 一つは 20 世紀初頭に開発された因子分析モデル (Spearman (1904)) とそれを発展させた構造方程式モデル (e.g., Bollen (1989), Hoyle (2012)) である. もう一つは実験計画法や分散分析において登場するランダム効果 (ブロック効果, e.g., Searle (1971)) であり, それを発展させた (線形) 混合モデルである (e.g., Laird and Ware (1982), McCulloch *et al.* (2008)). 近年注目されている離散型潜在変数は有用な有限混合モデルを提供する (Muthén (2002)). あまり意識はしないが誤差変数も潜在変数である. 潜在変数モデルにおける欠測値問題は多くの文献で議論されている (e.g., Muthén *et al.* (1987), Arminger and Sobel (1990), Jamshidian (2004), Yuan and Lu (2008), Savalei and Bentler (2009)).

本節では Kano and Takai (2011) に従い, 下記で定義される線形潜在変数モデル (Linear latent variate model, LLVM) の欠測値問題を考える. $\mathbf{Y}$ を p 次元の観測変数ベクトル, $\mathbf{Z}_g$

を k_g 次元の潜在変数ベクトル ($g = 1, \dots, G$) で互いに無相関, e を p 次元の誤差ベクトルとし, 下記のモデルを考える.

$$\begin{aligned}
 \mathbf{Y} &= \boldsymbol{\mu} + \sum_{g=1}^G \Lambda_g(\boldsymbol{\lambda}) \mathbf{Z}_g + \mathbf{e} \\
 E(\mathbf{Z}_g) &= \mathbf{0}, \quad E(\mathbf{e}) = \mathbf{0}, \quad V(\mathbf{Z}_g) = \Phi_g, \quad V(\mathbf{e}) = \Psi = \text{diag}(\psi_{11}, \dots, \psi_{pp}), \\
 \text{Cov}(\mathbf{Z}_g, \mathbf{e}) &= \mathbf{0}, \quad \text{Cov}(\mathbf{Z}_g, \mathbf{Z}_h) = \mathbf{0}, \quad g \neq h, \\
 E[(\mathbf{Z}_g' \mathbf{Z}_g)^2] &< \infty, \quad E[(\mathbf{e}' \mathbf{e})^2] < \infty
 \end{aligned} \tag{3.1}$$

このとき

$$\begin{aligned}
 E(\mathbf{Y}) &= \boldsymbol{\mu} \\
 V(\mathbf{Y}) &= \sum_{g=1}^G \Lambda_g(\boldsymbol{\lambda}) \Phi_g \Lambda_g(\boldsymbol{\lambda})' + \Psi \quad (= \Sigma(\boldsymbol{\theta}), \text{ say}), \\
 \boldsymbol{\theta} &= [\boldsymbol{\lambda}', \psi_{11}, \dots, \psi_{pp}, v(\Phi_1)', \dots, v(\Phi_G)']'
 \end{aligned} \tag{3.2}$$

を得る. ここで, $\boldsymbol{\lambda}$ は s 次元のベクトルパラメータ, $\Lambda(\boldsymbol{\lambda})$ は $p \times k_g$ の行列値関数で連続, そして, $v(\cdot)$ は vec-operator (Magnus and Neudecker (1999)) である. また, Φ_g は構造をもたないことを仮定しておく. LLVM は有名な LISREL モデル (Jöreskog and Sörbom (1993)) や EQS モデル (Bentler (2006)) などのほとんどを実現できる広範なモデルである (Browne and Shapiro (1988), Anderson (1987, 1989)). なお, $G = 1$ のとき LLVM は検証的因子分析モデル (confirmatory factor analysis model; CFA) を表す.

さて, 観測変数ベクトル $\mathbf{Y}$ に欠測が生じるとしよう. 第2節での定義や記号を引き継ぐ. 正規性の仮定の下で観測変数ベクトル $D_{\mathbf{r}^{(\ell)}} \mathbf{Y}$ の確率分布は

$$N_{p^{(\ell)}} \left(D_{\mathbf{r}^{(\ell)}} \mathbf{y} \mid D_{\mathbf{r}^{(\ell)}} \boldsymbol{\mu}, D_{\mathbf{r}^{(\ell)}} \Sigma(\boldsymbol{\theta}) D_{\mathbf{r}^{(\ell)}}' \right)$$

である. ここで, $p^{(\ell)}$ は欠測パターン $\mathbf{r}^{(\ell)}$ において観測される変数の数である. サイズ n の標本 $(D_{\mathbf{R}_1} \mathbf{Y}_1, \mathbf{R}_1), \dots, (D_{\mathbf{R}_n} \mathbf{Y}_n, \mathbf{R}_n)$ が与えられているとし, $I_\ell = \{1 \leq i \leq n \mid \mathbf{R}_i = \mathbf{r}^{(\ell)}\}$ を定義し, n_ℓ を I_ℓ の要素の数とする. このとき, $\bigcup_{\ell=1}^L I_\ell = \{1, \dots, n\}$, そして, $\sum_{\ell=1}^L n_\ell = n$

である。この記号の下で、Direct Likelihood は

$$\begin{aligned}
 DL(\boldsymbol{\theta}, \boldsymbol{\mu}) &= \prod_{i=1}^n N\left(D_{\mathbf{R}_i} \mathbf{Y}_i \mid D_{\mathbf{R}_i} \boldsymbol{\mu}, D_{\mathbf{R}_i} \Sigma(\boldsymbol{\theta}) D_{\mathbf{R}_i}'\right) \\
 &= \prod_{i=1}^n \sum_{\ell=1}^L 1_{\{\mathbf{R}_i = \mathbf{r}^{(\ell)}\}} N_{p^{(\ell)}}\left(D_{\mathbf{r}^{(\ell)}} \mathbf{Y}_i \mid D_{\mathbf{r}^{(\ell)}} \boldsymbol{\mu}, D_{\mathbf{r}^{(\ell)}} \Sigma(\boldsymbol{\theta}) D_{\mathbf{r}^{(\ell)}}'\right) \\
 &= \prod_{\ell=1}^L \prod_{i \in I_\ell} N_{p^{(\ell)}}\left(D_{\mathbf{r}^{(\ell)}} \mathbf{Y}_i \mid D_{\mathbf{r}^{(\ell)}} \boldsymbol{\mu}, D_{\mathbf{r}^{(\ell)}} \Sigma(\boldsymbol{\theta}) D_{\mathbf{r}^{(\ell)}}'\right) \quad (3.3) \\
 \boldsymbol{\theta} &= [\boldsymbol{\lambda}', \psi_{11}, \dots, \psi_{pp}, v(\Phi_1)', \dots, v(\Phi_G)']'
 \end{aligned}$$

となる。(3.3) から、Direct Likelihood は L 個の欠測パターンを L 個の母集団と考えた多母集団の同時分析と同等であることがわかる (Muthén *et al.* (1987), Satorra (2002)). なお、すべてのパラメータ $\boldsymbol{\theta}$ と $\boldsymbol{\mu}$ を母集団間で等置していることに注意する。

次の仮定をおく。

A1. $\mathbf{Z}_1, \dots, \mathbf{Z}_G$, $\mathbf{e}$ は互いに独立

A2. $(\mathbf{Z}_1, \dots, \mathbf{Z}_G)$ を与えた下で $\mathbf{R}$ と $\mathbf{Y}$ は条件付き独立 (shared-parameter model)

A3. $P(\mathbf{R} = \mathbf{r} \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_G = \mathbf{z}_G) = \prod_{g=1}^G h_g(\mathbf{r} \mid \mathbf{z}_g, \boldsymbol{\tau}, \boldsymbol{\theta})$

前節で述べたように、shared-parameter model から導かれる欠測メカニズムは NMAR である。欠測メカニズムが潜在変数に依存するモデルは多くの状況で妥当であると思われる。たとえば、変量内誤差モデル (e.g., Fuller (1987)) では真の変数は必ず誤差を伴ってしか観測されない状況を扱う。すなわち、「観測変数 = 真の変数 + 誤差」である。欠測メカニズムは誤差を含んだ観測変数に規定され则认为よりも誤差フリーの真の変数 (潜在変数) に依存する则认为の方が自然であろう。なお、潜在変数モデルとしてポピュラーな CFA ($G = 1$) の場合、仮定 A3 は不要である。Kano and Takai (2011, Section 6) は仮定 A3 について具体的なモデルを用いて議論している。

上記の仮定の下で、次の結果を得る。

$\mathbf{R}$ を与えた下で $\mathbf{Z}_1, \dots, \mathbf{Z}_G$, $\mathbf{e}$ は互いに条件付き独立

$$\begin{aligned}
 E(\mathbf{Y} \mid \mathbf{R} = \mathbf{r}^{(\ell)}) &= \boldsymbol{\mu} + \sum_{g=1}^G \Lambda_g(\boldsymbol{\lambda}) \boldsymbol{\nu}_g^{(\ell)} \left(= \boldsymbol{\mu}^{(\ell)}, \text{ say} \right) \\
 V(\mathbf{Y} \mid \mathbf{R} = \mathbf{r}^{(\ell)}) &= \sum_{g=1}^G \Lambda_g(\boldsymbol{\lambda}) \Phi_g^{(\ell)} \Lambda_g(\boldsymbol{\lambda})' + \Psi \left(= \Sigma(\boldsymbol{\theta}^{(\ell)}), \text{ say} \right), \quad (3.4) \\
 \boldsymbol{\theta}^{(\ell)} &= [\boldsymbol{\lambda}', \psi_{11}, \dots, \psi_{pp}, v(\Phi_1^{(\ell)})', \dots, v(\Phi_G^{(\ell)})']', \quad \ell = 1, \dots, L
 \end{aligned}$$

ここで, $\nu_g^{(\ell)} = E[\mathbf{Z}_g | \mathbf{R} = \mathbf{r}^{(\ell)}]$ と $\Phi_g^{(\ell)} = V[\mathbf{Z}_g | \mathbf{R} = \mathbf{r}^{(\ell)}]$ であり, これらは (3.2) のモデルのパラメータである θ と欠測メカニズムを規定するパラメータと関係する複雑な構造をもっている. ここでは, その構造を無視して飽和モデルとして扱うこととする. 以上のことから, 潜在変数間の独立性を仮定した LLVM(3.1) は, $\mathbf{R} = \mathbf{r}^{(\ell)}$ の下でも潜在変数間の独立性が保たれた LLVM になる. すなわち, LLVM は $\mathbf{R}$ による条件付けという操作に関して閉じていることが理解されよう.

以上により, (擬似) 尤度を下記のように定義すれば, パラメータ $(\lambda, \psi_{11}, \dots, \psi_{pp})$ の一致推定が可能であることが分かる.

$$QDL(\theta, M) = \prod_{\ell=1}^L \prod_{i \in I_\ell} N_{p^{(\ell)}} \left(D_{\mathbf{r}^{(\ell)}} \mathbf{Y}_i \mid D_{\mathbf{r}^{(\ell)}} \boldsymbol{\mu}^{(\ell)}, D_{\mathbf{r}^{(\ell)}} \Sigma(\boldsymbol{\theta}^{(\ell)}) D_{\mathbf{r}^{(\ell)}}' \right) \quad (3.5)$$

$$M = \left[\boldsymbol{\mu}^{(1)'} , \dots, \boldsymbol{\mu}^{(L)'} \right]',$$

$$\Theta = \left[\lambda', \psi_{11}, \dots, \psi_{pp}, v(\Phi_1^{(1)})', \dots, v(\Phi_G^{(1)})', \dots, v(\Phi_1^{(L)})', \dots, v(\Phi_G^{(L)})' \right]'$$

MAR の際に適用できる Direct Likelihood (3.3) と新しい擬似尤度 (3.5) を比較すると, (3.3) は (3.5) において

$$\boldsymbol{\mu}^{(1)} = \dots = \boldsymbol{\mu}^{(L)} \quad (\text{or } \nu_g^{(\ell)} \equiv 0)$$

$$\Phi_g^{(1)} = \dots = \Phi_g^{(L)}, \quad g = 1, \dots, G$$

なる制約を付加することによって得られることが分かる. すなわち, (3.3) は欠測パターンによって定義された母集団間において すべて のパラメータを等置するのに対して, (3.5) は 一部 のパラメータ $(\lambda, \psi_{11}, \dots, \psi_{pp})$ のみを等置する. いずれの場合も, SEM の標準的なプログラムで分析可能である (狩野・三浦 (2002)).

各欠測パターンにおける潜在変数の平均 $\nu_g^{(\ell)} = E[\mathbf{Z}_g | \mathbf{R} = \mathbf{r}^{(\ell)}]$ は

$$\boldsymbol{\mu}^{(\ell)} = \boldsymbol{\mu} + \sum_{g=1}^G \Lambda_g(\lambda) \nu_g^{(\ell)}, \quad \ell = 1, \dots, L$$

を利用することによって推定することができる. このとき, 潜在変数の分散行列 $V(\mathbf{Z}_g) = \Phi_g$ は下記の公式によって推定できる.

$$\begin{aligned} \Phi_g &= V(\mathbf{Z}_g) = E[V(\mathbf{Z}_g | \mathbf{R})] + V[E(\mathbf{Z}_g | \mathbf{R})] \\ &= \sum_{\ell=1}^L P(\mathbf{R} = \mathbf{r}^{(\ell)}) \Phi_g^{(\ell)} + \sum_{\ell=1}^L P(\mathbf{R} = \mathbf{r}^{(\ell)}) (\nu_g^{(\ell)} - \bar{\nu}_g)(\nu_g^{(\ell)} - \bar{\nu}_g)', \end{aligned}$$

ここで $\bar{\nu}_g = \sum_{\ell=1}^L P(\mathbf{R} = \mathbf{r}^{(\ell)}) \nu_g^{(\ell)}$ である.

(3.4) から, リストワイズ削除した完全データによる解析 (CCA) も可能であることが分かる (Kano and Takai (2009)). ただし, 潜在変数の分散行列 Φ_g の推定にはバイアスが生じ, また, 一般論として CCA は精度が落ちる.

潜在変数の平均は識別性の問題を考慮して推定しなければならない. 推定量の漸近分布など含めて本分析方法の詳細については Kano and Takai (2011) を参照されたい.

4. MAR 条件の緩和

狩野 (2011) と Kano (2013) は MAR より弱い下記の条件を提案した.

$$\sum_{\ell=1}^L P\left(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y}^{(\ell)} = \mathbf{y}^{(\ell)}; \boldsymbol{\varphi}\right) \left(= \sum_{\ell=1}^L m_{\mathbf{r}^{(\ell)}}(\mathbf{r}^{(\ell)} | \mathbf{y}^{(\ell)}; \boldsymbol{\varphi})\right) = 1, \quad \mathbf{y} \in \mathcal{Y}, \boldsymbol{\varphi} \in \Xi \quad (4.1)$$

この条件は, 左辺が $\mathbf{y}$ に依存しない定数であることと同等である. 実際, 期待値を取ると

$$\int \sum_{\ell=1}^L P\left(\mathbf{R} = \mathbf{r}^{(\ell)} | \mathbf{Y}^{(\ell)} = \mathbf{y}^{(\ell)}; \boldsymbol{\varphi}\right) f(\mathbf{y} | \boldsymbol{\varphi}) d\mathbf{y} = \sum_{\ell=1}^L P\left(\mathbf{R} = \mathbf{r}^{(\ell)} | \boldsymbol{\varphi}\right) = 1 \quad (4.2)$$

を得て, 定数 = 1 であることが示される.

(2.4) の MAR 条件を (4.1) の最左辺へ代入すれば, すべての $\mathbf{Y}^{(\ell)}$ が $\mathbf{Y}$ に置き換わり (4.1) が成立することが容易に確かめられる. MAR の条件は L 個の式から構成されているのに対して, (4.1) は一つの式であることから, (4.1) は MAR と比して, 少なくとも数学的には, より弱い条件であることが理解される.

条件 (4.1) の下で Direct Likelihood (2.8) が一致推定量を生み出すことは, 第 5 節で示される.

本節も, 基本的に第 2 節の定義と記号を引き継ぐが, 特に, $\mathcal{Y} \times \Theta$ 上で $f(\mathbf{y} | \boldsymbol{\theta}) > 0$ が仮定されていることに注意する.

命題 4.1 $p = 2, L = 4, \mathbf{Y} = [Y_1, Y_2]', \mathbf{R} = [R_1, R_2]'$ とし, 欠測メカニズムを下記で与える.

$$P(\mathbf{R} = (i, j) | y_1, y_2) = g_{ij}(y_1) + h_{ij}(y_2), \quad (4.3)$$

$$0 \leq g_{ij}(y_1) \leq 1 \text{ and } 0 \leq h_{ij}(y_2) \leq 1, \quad i, j = 0, 1, [y_1, y_2]' \in \mathcal{Y}$$

条件 (4.1) は

$$P(\mathbf{R} = (1, 1) | y_1, y_2) + P(\mathbf{R} = (1, 0) | y_1) + P(\mathbf{R} = (0, 1) | y_2) + P(\mathbf{R} = (0, 0)) = 1 \quad (4.4)$$

となり, これをみたすように $g_{ij}(y_1)$ と $h_{ij}(y_2)$ をとることができる. なお, 4 つの関数

$$g_{01}(y_1), h_{10}(y_2), g_{00}(y_1), h_{00}(y_2)$$

がすべて定数関数であるとき、また、そのときに限り、欠測メカニズムは MAR となる。

証明 4.1 (4.3) が MAR 条件をみたすための条件は明らかである。

定数 $0 < c_1, c_2 < 1/4$ を定め、

$$0 \leq g_{01}(y_1) \leq c_1, \quad 0 \leq h_{10}(y_2) \leq c_2, \quad \forall [y_1, y_2]' \in \mathcal{Y} \quad (4.5)$$

をみたす関数 $g_{01}(y_2)$, $h_{10}(y_1)$ をとる。下記左辺の関数を右辺によって定義する⁶⁾。

$$\begin{aligned} g_{11}(y_1) + g_{10}(y_1) &= c_2 - E[h_{10}(Y_2)|Y_1 = y_1] \\ h_{11}(y_2) + h_{01}(y_2) &= c_1 - E[g_{01}(Y_1)|Y_2 = y_2] \end{aligned} \quad (4.6)$$

これらはそれぞれ区間 $[0, c_2]$ と $[0, c_1]$ に収まっているから、

$$\sum_{i,j} g_{ij}(y_1) = \left(c_2 - E[h_{10}(Y_2)|Y_1 = y_1] \right) + g_{01}(y_1) + g_{00}(y_1) = \frac{1}{2} \quad (4.7)$$

$$\sum_{i,j} h_{ij}(y_2) = \left(c_1 - E[g_{01}(Y_1)|Y_2 = y_2] \right) + h_{10}(y_2) + h_{00}(y_2) = \frac{1}{2} \quad (4.8)$$

が成立するように、 $g_{00}(y_1)$ と $h_{00}(y_2)$ を定めることができる。以上により、 $\sum_{i,j} (g_{ij}(y_1) + h_{ij}(y_2)) = 1$ となることが分かった。条件 (4.4) をみたすことを確認する。実際

$$\begin{aligned} & \left(g_{11}(y_1) + h_{11}(y_2) \right) + E \left[g_{10}(Y_1) + h_{10}(Y_2) | Y_1 = y_1 \right] \\ & + E \left[g_{01}(Y_1) + h_{01}(Y_2) | Y_2 = y_2 \right] + E \left[g_{00}(Y_1) + h_{00}(Y_2) \right] \\ & = g_{11}(y_1) + g_{10}(y_1) + E \left[h_{10}(Y_2) | Y_1 = y_1 \right] \\ & + h_{11}(y_2) + h_{01}(y_2) + E \left[g_{01}(Y_1) | Y_2 = y_2 \right] \\ & + E \left[g_{00}(Y_1) + h_{00}(Y_2) \right] \\ & = c_1 + c_2 + E \left[g_{00}(Y_1) + h_{00}(Y_2) \right] = 1 \end{aligned}$$

が成り立つ。最後の等式は、(4.7) と (4.8) の両辺の期待値を取り加算することにより容易に得られる。□

証明から分かるように、条件 (4.4) をみたす関数 $g_{ij}(y_1)$ と $h_{ij}(y_2)$ にはかなりの自由度がある。実際、(4.5) と (4.6) は、値域の制限を考えなければ、(少なくとも) $g_{01}(y_1)$, $h_{10}(y_2)$, $g_{11}(y_1)$ (or $g_{10}(y_1)$), $h_{11}(y_2)$ (or $h_{01}(y_2)$) が自由に決められることを示す。

次の命題は、緩和された MAR が元の MAR と同値になることがあることを示す。

⁶⁾ 左辺の個々の関数の選択には自由度がある。和のみ定めればよい。

命題 4.2 $p \geq 2$, $L = 2$, $\mathbf{r}^{(1)} = [1, \dots, 1]'$ とする. すなわち, $\mathbf{Y}^{(1)} = \mathbf{Y}$. このとき, MAR と条件 (4.1) は同値である.

証明 4.2 まず

$$P(\mathbf{R} = \mathbf{r}^{(1)} | \mathbf{Y}^{(1)} = \mathbf{y}^{(1)}) = P(\mathbf{R} = \mathbf{r}^{(1)} | \mathbf{Y} = \mathbf{y}) = 1 - P(\mathbf{R} = \mathbf{r}^{(2)} | \mathbf{Y} = \mathbf{y})$$

である. この式を条件 (4.1) へ代入すると

$$P(\mathbf{R} = \mathbf{r}^{(2)} | \mathbf{Y} = \mathbf{y}) = P(\mathbf{R} = \mathbf{r}^{(2)} | \mathbf{Y}^{(2)} = \mathbf{y}^{(2)}), \quad \mathbf{y} \in \mathcal{Y}$$

を得る. □

$\mathbf{Y}^{(1)}$ が完全データのとき, $\ell = 1$ に対する MAR 条件は自明であるから MAR の条件式の個数は $L - 1 = 1$ となる. したがって, 命題 4.2 の状況では条件式の個数が等しいことから, 同値性は予想される結果とも言える. しかし, 命題 4.3 に示すように, $L = 2$ のときでも一般にはこの同値性は成立しない.

欠測データ解析の重要な応用に因果推論 (causal inference) がある (Rubin (1974, 1978)). ある母集団から無作為抽出された標本を対照群 ($X = 0$) に割り付けたときの特性値を Y_1 , 処置群 ($X = 1$) に割り付けたときの特性値を Y_2 とする. 割り付けは無作為とは限らず, また, ひとつの標本にはどちらか一方のみが割り付け可能であるとする. この設定は, たとえば, 投薬の効果をみるとき, 一人の被験者に投薬 (処置群) と非投薬 (またはプラセボ投薬, 対照群) の両者を割り付けることが一般には不可能である事実を反映している. このとき, Y_1 または Y_2 のどちらかが必ず欠測するという状況で群間比較を行わなければならない. これは因果推論の基本的問題 (fundamental problem of causal inference) と言われている (e.g., Holland (1986)). なお, ここで群を表す変数 X は観測 (欠測) 指標変数 $\mathbf{R}$ と同等であることに注意する.

通常, この設定の下で因果推論を可能とするため, 強い意味で無視可能 (strongly ignorable, SI) とよばれる仮定をおく. 共変量を表す確率ベクトルを $\mathbf{Y}_3$ とし, $\mathbf{Y}_3$ が与えられた下で, X と (Y_1, Y_2) とが条件付き独立, すなわち $X \perp\!\!\!\perp (Y_1, Y_2) | \mathbf{Y}_3$, であるとき, 欠測メカニズムは強い意味で無視可能という (e.g., Rosenbaum and Rubin (1983), 狩野 (2003), 宮川 (2004), 星野 (2009))⁷⁾. なお, (SI) は MAR の十分条件である. 次の命題において $\mathbf{Y}_3$ は空でもよい.

⁷⁾ より正確には, 条件間の比較を可能とするため $0 < P(X = 1 | \mathbf{Y}_3 = \mathbf{y}_3) < 1$ なる条件が置かれている. 上記文献では本論文でのように共変量 $\mathbf{Y}_3$ は欠測を含まないことを仮定しているが, Rubin (1978) では $\mathbf{Y}_3$ に欠測を許すより一般的な状況を扱っている.

命題 4.3 $p \geq 2$ とし, p 次元観測変数ベクトルを $\mathbf{Y} = [Y_1, Y_2, \mathbf{Y}_3']'$ とする. 欠測パターン数を $L = 2$ とし, $\mathbf{r}^{(1)} = [1, 0, 1, \dots, 1]'$, $\mathbf{r}^{(2)} = [0, 1, 1, \dots, 1]'$ とする. このとき, 強い意味で無視可能 (SI), MAR, 条件 (4.1) は下記のように表すことができる.

$$(SI): \mathbf{R} \perp\!\!\!\perp (Y_1, Y_2) \mid \mathbf{Y}_3$$

$$MAR: \begin{cases} \mathbf{R} \perp\!\!\!\perp Y_1 \mid (Y_2, \mathbf{Y}_3) \\ \mathbf{R} \perp\!\!\!\perp Y_2 \mid (Y_1, \mathbf{Y}_3) \end{cases} \quad (4.9)$$

$$\text{条件 (4.1): } \begin{cases} \mathbf{R} \perp\!\!\!\perp Y_1 \mid \mathbf{Y}_3 \\ \mathbf{R} \perp\!\!\!\perp Y_2 \mid \mathbf{Y}_3 \end{cases} \quad (4.10)$$

さらに, (SI) と MAR は同値であり, 一般に条件 (4.1) は (SI) や MAR より弱い.

証明 4.3 MAR の条件は

$$P(\mathbf{R} = \mathbf{r}^{(1)} \mid Y_1 = y_1, Y_2 = y_2, \mathbf{Y}_3 = \mathbf{y}_3) = P(\mathbf{R} = \mathbf{r}^{(1)} \mid Y_1 = y_1, \mathbf{Y}_3 = \mathbf{y}_3)$$

$$P(\mathbf{R} = \mathbf{r}^{(2)} \mid Y_1 = y_1, Y_2 = y_2, \mathbf{Y}_3 = \mathbf{y}_3) = P(\mathbf{R} = \mathbf{r}^{(2)} \mid Y_2 = y_2, \mathbf{Y}_3 = \mathbf{y}_3)$$

であり, $L = 2$ のとき MAR が条件付き独立 (4.9) を意味することは, (2.5) において既に指摘している. また, 二つの条件付き独立 (4.9) と (SI) が同値であることは block independence lemma とよばれ, よく知られた命題である⁸⁾. たとえば Whittaker (1990, Proposition 2.2.5) や宮川 (1997, 定理 2.3) をみよ.

つぎに (4.10) を示す. 条件 (4.1) は

$$P(\mathbf{R} = \mathbf{r}^{(1)} \mid Y_1 = y_1, \mathbf{Y}_3 = \mathbf{y}_3) + P(\mathbf{R} = \mathbf{r}^{(2)} \mid Y_2 = y_2, \mathbf{Y}_3 = \mathbf{y}_3) = 1, \quad [y_1, y_2, \mathbf{y}_3']' \in \mathcal{Y}$$

と書ける. 左辺第 1 項は y_1 と $\mathbf{y}_3$ の関数, 左辺第 2 項は y_2 と $\mathbf{y}_3$ の関数であるが, それらの和が定数であるため, それぞれの項は y_1 と y_2 には依存しないことになる. よって

$$P(\mathbf{R} = \mathbf{r}^{(1)} \mid Y_1 = y_1, \mathbf{Y}_3 = \mathbf{y}_3) = P(\mathbf{R} = \mathbf{r}^{(1)} \mid \mathbf{Y}_3 = \mathbf{y}_3)$$

が成立する. $L = 2$ であるから

$$P(\mathbf{R} = \mathbf{r}^{(2)} \mid Y_1 = y_1, \mathbf{Y}_3 = \mathbf{y}_3) = P(\mathbf{R} = \mathbf{r}^{(2)} \mid \mathbf{Y}_3 = \mathbf{y}_3)$$

が得られ, $\mathbf{R}$ と Y_1 が条件付き独立であることが示される.

数学的には, 条件 (4.1) から (SI), すなわち, 周辺分布の独立性から同時分布の独立性が導けないことは明らかである. たとえば, Whittaker (1990, p. 31) に反例がある. $\square$

⁸⁾ 条件付き独立が有するいくつかの関係式は graphoid という代数的性質をみたす. その中で, 本命題の証明で用いた性質は intersection とよばれているものである. 詳細は Studený (2005) を参照されたい.

MAR から強い意味の無視可能性 (SI) が導けることは注目に値する⁹⁾.

予想通り, 数学的には条件 (4.1) は MAR より弱いことが分かった. しかし, とくに $L = 2$ のときは, 応用上それらの違いを認識できない可能性はある. Rubin (1978) は因果推論において $T(\geq 2)$ 個の群の比較を論じている. 命題 4.3 では $T = 2$ のケースを扱ったわけであるが, $T \geq 3$ のときは, MAR と条件 (4.1) は, 命題 4.3 のような確率変数の条件付き独立性によって記述することはできない. さらに, (SI) と MAR は同値ではなく, それゆえ 2 つの概念を区別する意味がある.

さて, 欠測値データ解析の理論研究においても 条件 (4.1) が有用であることを紹介する. Molenberghs *et al.* (2008) は, 欠測メカニズムが NMAR であるとき, まったく同一の適合を導く $(\mathbf{Y}, \mathbf{R})$ の同時分布で MAR をみたすものを構成した. 彼らは, それを MAR counterpart とよんだ. 以下では, (2.1)–(2.3) の記号を踏襲しパラメータは省略する. また, 複雑な記号を避けるため, $f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)})$ と $m_{\mathbf{r}^{(\ell)}}(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)})$ における添え字 $\mathbf{r}^{(\ell)}$ を省略する.

$(\mathbf{Y}, \mathbf{R})$ の同時分布を

$$m(\mathbf{r}^{(\ell)}|\mathbf{y})f(\mathbf{y}), \quad \ell = 1, \dots, L, \quad \mathbf{y} \in \mathcal{Y} \quad (4.11)$$

と書き, この分布から導かれる欠測メカニズムは NMAR であるとする. Molenberghs *et al.* (2008) は同時分布 (4.11) の MAR counterpart を

$$m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)})f(\mathbf{y}), \quad \ell = 1, \dots, L, \quad \mathbf{y} \in \mathcal{Y} \quad (4.12)$$

によって定義した¹⁰⁾. ここで

$$m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}) = \int m(\mathbf{r}^{(\ell)}|\mathbf{y}) \frac{f(\mathbf{y})}{f(\mathbf{y}^{(\ell)})} d\mathbf{y}^{(-\ell)}$$

である. (4.12) が確率分布であること, そして, (4.11) と (4.12) にもとづく尤度 (Full Likelihood) が一致することを容易に確認できる.

坂口・狩野 (2013) は次を示した.

命題 4.4 同時分布 (4.12) から導かれる欠測メカニズムが MAR であるための必要十分条件は (4.1) がみたされることである. すなわち

$$\sum_{\ell=1}^L m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}) = 1, \quad \mathbf{y} \in \mathcal{Y} \quad (4.13)$$

である.

⁹⁾ 星野 (2009, p. 44) は, 社会科学分野の応用においては MAR と (SI) は実質的に等値と述べている.

¹⁰⁾ Molenberghs *et al.* (2008) の記号を用いると ((9) 式と (10) 式), MAR counterpart は下記のように書ける.

$$f(\mathbf{y}^{(\ell)}|\mathbf{r}^{(\ell)})f(\mathbf{r}^{(\ell)})f(\mathbf{y}^{(-\ell)}|\mathbf{y}^{(\ell)})$$

これが (4.12) と同等であることはすぐに分かる.

注意 4.4 Molenberghs *et al.* (2008) は条件 (4.13) を明記していない. この命題から, どんな同時分布 $m(\mathbf{r}^{(\ell)}|\mathbf{y})f(\mathbf{y})$ に対しても (4.12) が MAR counterpart になるわけではないことが理解される.

証明 4.4 (Y, R) の同時分布が (4.12) で与えられるとき, Y の周辺分布 $g(\mathbf{y})$ は

$$g(\mathbf{y}) = f(\mathbf{y}) \sum_{\ell=1}^L m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}) \quad (4.14)$$

である. (4.12) と (4.14) より

$$P(R = \mathbf{r}^{(\ell')} | Y = \mathbf{y}) = \frac{m(\mathbf{r}^{(\ell')}|\mathbf{y}^{(\ell')})f(\mathbf{y})}{g(\mathbf{y})} = \frac{m(\mathbf{r}^{(\ell')}|\mathbf{y}^{(\ell')})}{\sum_{\ell=1}^L m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)})}, \quad \ell' = 1, \dots, L$$

一方

$$\begin{aligned} P(R = \mathbf{r}^{(\ell')} | Y^{(\ell')} = \mathbf{y}^{(\ell')}) &= \frac{m(\mathbf{r}^{(\ell')}|\mathbf{y}^{(\ell')})f(\mathbf{y}^{(\ell')})}{\int g(\mathbf{y})d\mathbf{y}^{(-\ell')}} \\ &= \frac{m(\mathbf{r}^{(\ell')}|\mathbf{y}^{(\ell')})f(\mathbf{y}^{(\ell')})}{\int f(\mathbf{y}) \sum_{\ell=1}^L m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)})d\mathbf{y}^{(-\ell')}}, \quad \ell' = 1, \dots, L \end{aligned}$$

である. したがって, 欠測メカニズムが MAR であるための条件は, これらを等置して

$$\sum_{\ell=1}^L m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}) = \int \frac{f(\mathbf{y})}{f(\mathbf{y}^{(\ell')})} \left(\sum_{\ell=1}^L m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}) \right) d\mathbf{y}^{(-\ell')}, \quad \ell' = 1, \dots, L, \mathbf{y} \in \mathcal{Y} \quad (4.15)$$

となる. したがって, (4.13) が MAR の十分条件になることは明らかである.

必要性を示す. 全ての欠測パターンにおいて共通に観測される変数を $\mathbf{y}_{(O)}$ とする. ベクトル $\mathbf{y}$ のすべての要素からなる集合を $\{\mathbf{y}\}$ と表すことにすると, $\{\mathbf{y}_{(O)}\} = \bigcap_{\ell=1}^L \{\mathbf{y}^{(\ell)}\}$ である. 任意の ℓ' に対して (4.15) の右辺は $\mathbf{y}^{(\ell')}$ の関数であるから,

$$\sum_{\ell=1}^L m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}) = h(\mathbf{y}_{(O)}) \quad (4.16)$$

と書くことができる. $\{\mathbf{y}_{(O)}\}$ が空のときは $h(\mathbf{y}_{(O)})$ は定数となるから, 第4節冒頭で述べたように, 証明は完結する. $\{\mathbf{y}_{(O)}\} \neq \emptyset$ とする. $\{\mathbf{y}_{(O)}\}$ の補集合, すなわち, ある欠測パターンにおいて欠測する変数の集合を $\{\mathbf{y}_{(-O)}\}$ と書く. $\{\mathbf{y}_{(-O)}\}$ は空でないとしてよいだろう. さらに $\{\mathbf{y}_{(-O)}^{(\ell)}\} = \{\mathbf{y}^{(\ell)}\} \setminus \{\mathbf{y}_{(O)}\}$ を定義すると

$$\{\mathbf{y}^{(\ell)}\} = \{\mathbf{y}_{(O)}\} \cup \{\mathbf{y}_{(-O)}^{(\ell)}\}, \quad \{\mathbf{y}_{(-O)}\} = \{\mathbf{y}^{(-\ell)}\} \cup \{\mathbf{y}_{(-O)}^{(\ell)}\}, \quad \ell = 1, \dots, L$$

を得る (図1を参照). (4.16) より

$$\sum_{\ell=1}^L \int m(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)})f(\mathbf{y})d\mathbf{y}_{(-O)} = \int h(\mathbf{y}_{(O)})f(\mathbf{y})d\mathbf{y}_{(-O)} = h(\mathbf{y}_{(O)})f(\mathbf{y}_{(O)}) \quad (4.17)$$

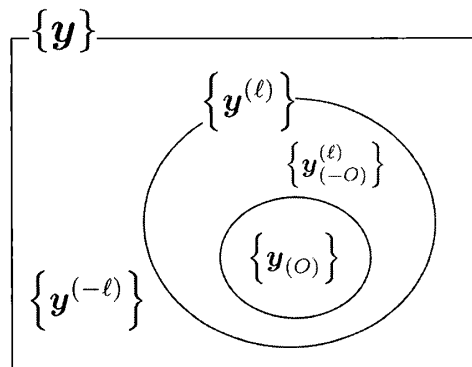


図1 変数の集合の包含関係.

となる. (4.17) の最左辺において

$$\begin{aligned}
 & \int m\left(\mathbf{r}^{(\ell)}|\mathbf{y}_{(O)}, \mathbf{y}_{(-O)}^{(\ell)}\right) f\left(\mathbf{y}_{(O)}, \mathbf{y}_{(-O)}^{(\ell)}, \mathbf{y}^{(-\ell)}\right) d\mathbf{y}^{(-\ell)} d\mathbf{y}_{(-O)}^{(\ell)} \\
 &= \int m\left(\mathbf{r}^{(\ell)}|\mathbf{y}_{(O)}, \mathbf{y}_{(-O)}^{(\ell)}\right) f\left(\mathbf{y}_{(O)}, \mathbf{y}_{(-O)}^{(\ell)}\right) d\mathbf{y}_{(-O)}^{(\ell)} \\
 &= m\left(\mathbf{r}^{(\ell)}|\mathbf{y}_{(O)}\right) f\left(\mathbf{y}_{(O)}\right)
 \end{aligned}$$

となるから, (4.17) の最左辺は $f(\mathbf{y}_{(O)})$ となり, それゆえ, $h(\mathbf{y}_{(O)}) = 1$ が帰結される. $\square$

坂口・狩野 (2013) では同時分布 (4.11) と (4.12) そして MAR と条件 (4.1) の関係を詳細に検討している.

5. 条件 (4.1) の十分性の証明

本節では, 欠測メカニズムを用いない尤度 (2.8) を最大にする解として定められる MLE $\hat{\boldsymbol{\theta}}$ が, 条件 (4.1) の下で真値 $\boldsymbol{\theta}$ の一致推定量になることを示す. 前節までの仮定と記号を引継ぐ. 特に, $\mathcal{Y} \times \Theta$ 上で $f(\mathbf{y}|\boldsymbol{\theta}) > 0$ であり, $\mathcal{Y}$ は $\boldsymbol{\theta}$ に依存しないことが仮定されていることに注意する. 未知パラメータの真値を $\boldsymbol{\varphi}_0 = [\boldsymbol{\theta}'_0, \boldsymbol{\tau}'_0]'$ ($\in \Xi$) とする. 真値はパラメータ空間 Ξ の任意の点とするが固定されているものとする.

まず, 任意の $\boldsymbol{\theta}$ ($\in \Theta$) に対して, 大数の法則より

$$\begin{aligned}
 \frac{1}{n} \log DL_n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \log f_{\mathbf{R}_i}(D_{\mathbf{R}_i} \mathbf{Y}_i | \boldsymbol{\theta}) \\
 &\xrightarrow{P} E[\log f_{\mathbf{R}}(D_{\mathbf{R}} \mathbf{Y} | \boldsymbol{\theta}) | \boldsymbol{\varphi}_0] \quad (n \rightarrow \infty)
 \end{aligned} \tag{5.1}$$

を得る. 不等式 $\log x \leq x - 1$ ($x > 0$) を用いると, 任意の $\boldsymbol{\theta} (\in \Theta)$ に対して

$$\begin{aligned}
 & E[\log f_{\mathbf{R}}(D_{\mathbf{R}}\mathbf{Y}|\boldsymbol{\theta})|\boldsymbol{\varphi}_0] - E[\log f_{\mathbf{R}}(D_{\mathbf{R}}\mathbf{Y}|\boldsymbol{\theta}_0)|\boldsymbol{\varphi}_0] \\
 &= \sum_{\ell=1}^L \int \log \frac{f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta})}{f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta}_0)} m_{\mathbf{r}^{(\ell)}}(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}; \boldsymbol{\varphi}_0) f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta}_0) d\mathbf{y}^{(\ell)} \\
 &\leq \sum_{\ell=1}^L \int \left(\frac{f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta})}{f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta}_0)} - 1 \right) m_{\mathbf{r}^{(\ell)}}(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}; \boldsymbol{\varphi}_0) f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta}_0) d\mathbf{y}^{(\ell)} \\
 &= \sum_{\ell=1}^L \int \left(f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta}) - f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta}_0) \right) m_{\mathbf{r}^{(\ell)}}(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}; \boldsymbol{\varphi}_0) d\mathbf{y}^{(\ell)} \\
 &= \int \left(f(\mathbf{y}|\boldsymbol{\theta}) - f(\mathbf{y}|\boldsymbol{\theta}_0) \right) \sum_{\ell=1}^L m_{\mathbf{r}^{(\ell)}}(\mathbf{r}^{(\ell)}|\mathbf{y}^{(\ell)}; \boldsymbol{\varphi}_0) d\mathbf{y} \tag{5.2}
 \end{aligned}$$

を得る. $\{\mathbf{y}^{(\ell)}\}$ が空のときは, $f_{\mathbf{r}^{(\ell)}}(\mathbf{y}^{(\ell)}|\boldsymbol{\theta}) = 1$ であり, 積分 $d\mathbf{y}^{(\ell)}$ を施さないことに注意する. 条件 (4.1) の下で (5.2) は 0 となる. よって

$$E[\log f_{\mathbf{R}}(D_{\mathbf{R}}\mathbf{Y}|\boldsymbol{\theta})|\boldsymbol{\varphi}_0] \leq E[\log f_{\mathbf{R}}(D_{\mathbf{R}}\mathbf{Y}|\boldsymbol{\theta}_0)|\boldsymbol{\varphi}_0], \quad \forall \boldsymbol{\theta} \in \Theta \tag{5.3}$$

を得る.

一貫性の証明には, パラメータの識別性, (5.1) における収束の一様性, パラメータ空間のコンパクト性など様々な数学的正則条件が必要である. その詳細はいくつかの成書等を参照されたい (e.g., Andersen and Gill (1982), Hampel *et al.* (1986), van der Vaart (2000), 稲垣 (2003), 吉田 (2006)). 本節では, 一貫性の証明において最も重要と考えられる一種の情報量不等式の証明を与えた.

当然のことながら, MAR の下でも (5.3) が成立する. ここでは parameter separation の仮定を用いていないことに注意する.

上記証明から, (4.1) は依然として十分条件であることが分かる. Kano (2013) はさらに弱い十分条件を導出しているがここでは省略する.

6. 議論

欠測値データ解析の理論研究は Rubin (1976) に始まり 40 年の歴史を刻む. 近年でも盛んに研究されている (e.g., Kim and Shao (2013)). Little (2011) はその歴史を次の 4 つに分けている. (i) Pre-EM Algorithm (Pre-1970s), (ii) The Maximum Likelihood Era (1970s–Mid 1980s), (iii) Bayes and Multiple Imputation (Mid 1980s–Present), (iv) Robustness Concerns (1990s–Present). これらの発展における基本的な仮定は MAR ないしは無視可能性である. これだけ重要な仮定でありながら, MAR 条件を緩和するための試みが報告されることはなかったように思われる. それは, Seaman *et al.* (2013) が指摘するように,

MAR の定義に曖昧さがあったことが影響しているのかもしれない。第 4 節で示したように、欠測メカニズムを用いない尤度にもとづく MLE が一貫性をもつための条件としては MAR は強すぎる条件で、少なくとも、数学的には緩和可能である。本論文では、緩和された MAR 条件 (4.1) をみたすが MAR は満足しないモデルをいくつか指摘したが、今後、さらに研究を進める必要がある。

なお、本論文で指摘した問題は尤度法に限らず、多重代入法 (Rubin (1987), Schafer (1999), Hojo *et al.* (in press)) であっても同様であろう。すなわち、MAR を仮定した多重代入法を行うときは MAR の仮定が問題になり、欠測メカニズムを考慮した尤度を用いる場合は識別性が問題になる。識別性が弱いと、Bayes 法を適用しても事前分布の選択に推定結果が大きく依存することがある。

Little (2011) が (iv) に計上したロバストネスは重要である。鍵の成果は、doubly robust estimation とよばれ、目的の確率モデルと欠測メカニズム (傾向スコア) のどちらかが正しければ一致推定量が構成できるという魅力的なものであり、多くの研究がある。この結果における重要な仮定が (SI) したがって MAR である。inverse-probability-weight 法は ML に代る推定法である (e.g., Rubin (1985), Hirano *et al.* (2003)). ここでも (SI) したがって MAR が基本的な仮定と解される。いずれの研究においても MAR の緩和は有用な結果となるであろう。

謝辞

本論文の執筆にあたっては、多くの研究者との議論が刺激になった。特に、Dr. Roderick Little (University of Michigan), Dr. Mortaza Jamshidian (California State University), 高井啓二博士 (関西大学), 森川耕輔氏 (大阪大学), 坂口弘樹氏 (大阪大学) に感謝を申し上げる。査読者の適切な指摘と提言に感謝する。

参 考 文 献

- Albert, P. S. and Follman, D. A. (2009). Shared-parameter models, in *Longitudinal Data Analysis* (G. Fitzmaurice *et al.*, eds.), pp. 433–452, Chapman & Hall/CRC Press, Boca Raton, London.
- Allison, P. D. (2002). *Missing Data*, Sage Publications, Thousand Oaks, London, New Delhi.
- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study, *Ann. Stat.*, **10**, 1100–1120.
- Anderson, T. W. (1987). Multivariate linear relations, in *Proceedings of the Second International Conference in Statistics* (T. Pukkila and S. Puntanen, eds.), pp. 9–36, University of Tampere, Tampere, Finland.
- Anderson, T. W. (1989). Linear latent variable models and covariance structures, *J. Econom.*, **41**, 91–119.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data, in *Advanced Structural Equation Modeling: Issues and Techniques* (G. A. Marcoulides and R. E. Schumacker, eds.), pp. 243–277, Lawrence Erlbaum Associates, Mahwah, NJ.
- Arminger, G. and Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data, *J. Am. Stat. Assoc.*, **85**, 195–203.
- Bentler, P. M. (2006). *EQS Structural Equations Program Manual*, Multivariate Software, Inc., Encino, CA.

- Bollen, K. A. (1989). *Structural Equations with Latent Variables*, John Wiley & Sons, New York.
- Browne, M. W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variable models, *Br. J. Math. Stat. Psychol.*, **41**, 193–208.
- Chen, K. (2001). Parametric models for response-biased sampling, *J. R. Stat. Soc. B*, **63**, 775–789.
- Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and Kenward, M. G. (2011). Generalized shared-parameter models and missingness at random, *Stat. Model.*, **11**, 279–310.
- Daniels, M. J. and Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Chapman and Hall, Boca Raton, FL.
- Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion), *J. R. Stat. Soc. B*, **41**, 1–31.
- Fuller, W. A. (1987). *Measurement Error Models*, John Wiley & Sons, New York.
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models, *Struct. Equ. Model.*, **10**, 80–100.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, New York.
- Hirano, K., Imbens, G. W. and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica*, **71**, 1161–1189.
- Heitjan, D. F. and Rubin, D. (1991). Ignorability and coarse data, *Ann. Stat.*, **19**, 2244–2253.
- Hojo, S., Yamamoto, M. and Kano, Y. (in press). Effect of violation of the normal assumption on MI and ML estimators in the analysis of incomplete data, *Commun. Stat. Theo. Meth.*
- Holland, P. W. (1986). Statistics and causal inference (with discussion), *J. Am. Stat. Assoc.*, **81**, 945–970.
- 星野崇宏 (2009). 『調査観察データの統計科学—因果推論・選択バイアス・データ融合』岩波書店.
- Hoyle, R. H. (2012). *Handbook of Structural Equation Modeling*, The Guilford Press, New York.
- Ibrahim, J. G., Lipsitz, S. R. and Horton, N. (2001). Using auxiliary data for parameter estimation with non-ignorably missing outcomes, *Appl. Stat.*, **50**, 361–373.
- 稲垣宣生 (2003). 『数理統計学 (改訂版)』裳華房.
- 岩崎学 (2002). 『不完全データの統計解析』エコノミスト社.
- Jamshidian, M. (2004). Strategies for analysis of missing data, in *Handbook of Data Analysis* (M. Hardy and A. Bryman, eds.), pp. 113–130, Sage Publication, London, California, New Delhi.
- Jöreskog, K. G. and Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with SIMPLIS Command Language*, Scientific Software International, Chicago.
- 狩野裕 (2003). 「構造方程式モデリング, 因果推論, そして非正規性」『多変量解析の展開—隠れた構造と因果を推理する』甘利俊一, 竹内啓他編, pp. 65–129, 岩波書店.
- 狩野裕 (2011). What is MAR?, 2011 年度統計関連学会連合大会, 九州大学.
- Kano, Y. (2013). Weakening the MAR condition (submitted).
- 狩野裕, 三浦麻子 (2002). 『グラフィカル多変量解析 (増補版)』現代数学社.
- Kano, Y. and Takai, K. (2009). Should complete-case analysis always be avoided?—SEM for incomplete data with nonignorable missing, A paper presented at the IMPS2009, Cambridge.
- Kano, Y. and Takai, K. (2011). Analysis of NMAR missing data without specifying missing-data mechanisms in a linear latent variate model, *J. Multivar. Anal.*, **102**, 1241–1255.
- Kim, J.-K. and Shao, J. (2013). *Statistical Methods for Handling Incomplete Data*, Chapman & Hall/CRC, Boca Raton, London.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics*, **38**, 963–974.
- Lauritzen, S. L. (1996). *Graphical Models*, Clarendon Press, Oxford, UK.
- Little, R. J. A. (2011). Calibrated Bayes, for statistics in general, and missing data in particular, *Stat. Sci.*, **26**, 162–174.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd Ed.), John Wiley & Sons, New Jersey.

- Lu, G. and Copas, J. B. (2004). Missing at random, likelihood ignorability and model completeness, *Ann. Stat.*, **32**, 754–765.
- Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons Ltd., Baffins Lane, Chichester.
- McCulloch, C. E., Searle, S. R. and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models* (2nd Ed.), John Wiley & Sons, New Jersey (土居正明他 翻訳 (2011). 『線形モデルとその拡張』 CAC).
- 宮川雅巳 (1997). 『グラフィカルモデリング』 朝倉書店.
- 宮川雅巳 (2004). 『統計的因果推論』 朝倉書店.
- Molenberghs, G., Beunckens, C., Sotito, C. and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit, *J. R. Stat. Soc. B*, **70**, 371–388.
- 森川耕輔, 狩野裕 (2013). Identification problem for the analysis of binary repeated measure data with non-ignorable missing, 第 7 回日本統計学会春季集会, 学習院大学.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling, *Behaviormetrika*, **29**, 81–117.
- Muthén, B. O., Kaplan, D. and Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random, *Psychometrika*, **52**, 431–462.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70**, 41–55.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, *J. Edu. Psychol.*, **66**, 688–701.
- Rubin, D. B. (1976). Inference and missing data (with discussion), *Biometrika*, **63**, 581–592.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization, *Ann. Stat.*, **6**, 34–58.
- Rubin, D. B. (1985). The use of propensity scores in applied Bayesian inference, in *Bayesian Statistics 2*, (J. Bernardo, M. DeGroot, D. Lindley and A. Smith, eds.), pp. 463–472, North-Holland/Elsevier, Amsterdam, New York.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.
- 坂口弘樹, 狩野裕 (2013). On MAR counterpart, An unpublished manuscript.
- Satorra, A. (2002). Asymptotic robustness in multiple group linear-latent variable models, *Econ. Theory*, **18**, 297–312.
- Savalei, V. and Bentler, P. M. (2009). A two-stage approach to missing data: Theory and applications to auxiliary variables, *Struct. Equ. Model.*, **16**, 477–497.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*, Chapman & Hall/CRC, Boca Raton, New York.
- Schafer, J. L. (1999). Multiple imputation: A primer, *Stat. Meth. Medical Res.*, **8**, 3–15.
- Seaman, S., Galati, J., Jackson, D. and Carlin, J. (2013). What is meant by “missing at random”?, *Stat. Sci.*, **28**, 257–268.
- Searle, S. R. (1971). *Linear Models*, John Wiley, New York.
- Spearman, C. (1904). “General intelligence,” objectively determined and measured, *Am. J. Psychol.*, **15**, 201–292.
- Studený, M. (2005). *On Probabilistic Conditional Independence Structures*, Springer-Verlag, London.
- Takai, K. and Kano, Y. (in press). Asymptotic inference and missing data, *Commun. Stat. Theo. Meth.*
- Tang, G., Little, R. J. A. and Raghunathan, T. E. (2003). Analysis of multivariate missing data with nonignorable nonresponse, *Biometrika*, **90**, 747–764.
- Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*, Springer, New York.
- van der Vaart, A. W. (2000). *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, John Wiley & Sons, New York.
- 吉田朋広 (2006). 『数理統計学』 朝倉書店.
- Yuan, K.-H. and Lu, L. (2008). SEM with missing data and unknown population distributions using two-stage ML: Theory and its application, *Multi. Behav. Res.*, **43**, 621–652.