

確率統計基礎講義B

欠損データの統計解析

若木宏文

2012年度後期

参考書

R. R. A. Little and D. B. Rubin (2002), *Statistical analysis with missing data*, Wiley

目次

1. はじめに
2. 分散分析における欠測値の扱い
3. 完全データ分析と部分完全データ分析
4. 欠測値の補完
- 5-6. 補完データによる推定量の変動
- 7-9. 尤度に基づく推測
- 10-12. 単調欠測データの推測
- 13-15. EM アルゴリズムとその拡張

1 はじめに

1.1 欠損データとは

データ行列

行：人や物などの観測対象を表す

列：変数を表す

(i, j) 成分： i 番目の対象の j 番目の変数の値

量的データ (収入, 年齢), 質的データ (血液型:無順序, 教育レベル:順序付)

欠測値 (missing value) : 観測されなかった値

欠損データ (missing data) : 欠測値を含むデータ

完全データ (complete data) : 欠測値を含まないデータ

欠測値の例

世帯調査で収入を記入しなかった, 測定機械の不良で観測できなかった

⇒ 欠測値として扱う.

嗜好調査で, 選択肢に好みのものがなかった

⇒ 欠測値とせず, 「該当なし」の項目を追加すべき.

場合によっては, 「分からない」, 「答えたくない」など複数追加.

1.2 欠損データのパターン

データ行列を $Y = (y_{ij})$ とし, 欠測の有無を表すため 0-1 行列 $M = (m_{ij})$ を用意する.

$$m_{ij} = \begin{cases} 1 & y_{ij} \text{ が観測された} \\ 0 & y_{ij} \text{ が観測されなかった} \end{cases}$$

(a) 1 変数のみに欠測

$$M = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

(b) 複数の変数に同時に欠測

$$M = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}$$

(c) 欠測が単調に増加

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(d) ばらばら

$$M = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

(e) 同時に観測されない変数の組がある

$$M = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

(f) 因子分析

$$M = \begin{pmatrix} Y_1 & \cdots & Y_p & X_1 & X_2 \\ 1 & \cdots & 1 & 0 & 0 \\ 1 & \cdots & 1 & 0 & 0 \\ 1 & \cdots & 1 & 0 & 0 \end{pmatrix}$$

Y_1, \dots, Y_p : 観測される変数 , X_1, X_2 潜在変数

因子分析モデル

$$\begin{cases} Y_{1i} = a_{11}X_{1i} + a_{12}X_{2i} + \varepsilon_{1i} \\ Y_{2i} = a_{21}X_{1i} + a_{22}X_{2i} + \varepsilon_{2i} \\ \vdots \\ Y_{pi} = a_{p1}X_{1i} + a_{p2}X_{2i} + \varepsilon_{pi} \end{cases} \quad (i = 1, \dots, n)$$

例 1.1 (農業試験) Y_1, \dots, Y_K : 肥料, 気温, 日照時間等の説明変数
 Y_{K+1}, \dots, Y_{K+L} : 収穫された果実の糖度, 重量などの目的変数
⇒ パターン (a), あるいは (b) となる.

例 1.2 (追跡調査) 患者の集団に対して, 毎年同じ内容の検査を行う場合, 転居, 治療方法の変更, 死亡等によって研究期間が終わる前に調査対象から脱落する.
⇒ パターン (c) となる.

仮定 1.1 欠測値は, 意味のある真値が観測されなかったもの

例 1.3 (心臓まひのリスクに関する研究)

(Woolson and Clark (1984) より)

5つの変数：性別, 年齢 (5 カテゴリー), 肥満の有無 (3 時点で測定)

欠測パターン

Pattern	変数					
	Age	Gender	Weight 1	Weight 2	Weight 3	子供の数
A	0	0	0	0	0	1770
B	0	0	0	0	1	631
C	0	0	0	1	0	184
D	0	0	1	0	0	645
E	0	0	0	1	1	756
F	0	0	1	0	1	370
G	0	0	1	1	0	500

Woolson and Clark の分析

応答変数： O(肥満), N(肥満でない), M(欠測値) が 3 時点あるので
NNN, NNO, NON, ..., MMO の $3^3 - 1 = 26$ 通り

説明変数： 年齢 (3-5, 5-7, 7-9, 9-11, 11-13, 13-15) と性別 (男女) の組み合わせ
で 15 通り

15 × 26 の分割表データとして分析

*. 仮定 1.1 が妥当なので, 欠測をカテゴリーに加えるより O または N が
観測されなかったとして, 推定すべき.

例 1.4 (生存と生活の質に対する処置の因果効果)

薬と偽薬のどちらかをランダムに患者に割り当てる

$$T = \begin{cases} 0 & \text{偽薬 (placebo, control)} \\ 1 & \text{薬 (treatment)} \end{cases}$$

D_i : 患者 i の 1 年後の生死を表す変数

$$D_i(t) = \begin{cases} 0 & T = t \text{ の患者が生存} \\ 1 & T = t \text{ の患者が死亡} \end{cases} \quad (t = 0, 1)$$

因果効果

$$D_i(1) - D_i(0)$$

- *. $D_i(0)$ または $D_i(1)$ のどちらか一方しか観測されない
⇒ パターン (e) の欠損データであり, 因果効果も観測されないが,
ランダムに処置を割り当てたので, 因果効果の期待値は推定できる

さらに、1年後の生活の質を表す変数 $Y > 0$ を観測
死亡した患者に対しては Y の値は定義されないので、仮定 1.1 は成り立た
ない。

個々の患者の処置に対する生存と死亡の可能性

1. LL — $D_i(0) = D_i(1) = 0$ 両方生存
2. DD — $D_i(0) = D_i(1) = 1$ 両方死亡
3. LD — $D_i(0) = 0, D_i(1) = 1$ 偽薬では生存, 真薬で死亡
4. DL — $D_i(0) = 1, D_i(1) = 0$ 偽薬では死亡, 真薬で生存

LL では $Y_i(0), Y_i(1)$ の 2 変量の分布を考えることができる。

ただし、パターン (e) の欠損データとなり、その関連 (相関係数など) は
推定不可能。

DD では、生活の質は定義されない

LD または DL では、生存に対応した Y_i の 1 変量分布が考えられ、
観測される。

例 1.5 (国民投票の事前アンケート)

何かの政策についての是非を問う国民投票が行われる。

事前アンケートで、選択肢は「賛成」と「反対」

調査の結果は、「賛成」、「反対」と「無回答」となる。

無回答の内訳：(a)

- (a) 投票に行く心算がない(興味がないか、「賛成」、「反対」のどちらも選べない)
- (b) 「賛成」、「反対」を表明したくない
 - (a) に対しては仮定 1.1 は妥当ではない。
 - (b) に対しては仮定 1.1 は妥当で、欠測値として推定の対象となる。
 - 投票するかしないかも合わせて調査することで、「賛成」と「反対」の比率を推定するときは、「投票しない」人のデータを除外した上で欠損データとして分析できる。
 - 「投票する」、「投票しない」の選択肢でも「無回答」はあり得るが、比率の推定も欠損データとして分析できる。

1.3 欠損データのしくみ

仮定 1.1 が成り立つとする.

$Y = (y_{ij})$: 完全データ行列

$M = (m_{ij})$: 欠測の有無を表す 確率行列

$f(M|Y, \phi)$: Y が与えられたときの M の条件付き確率 (密度) 関数
ただし, ϕ は未知母数を表す.

MCAR (Missing completely at random)

$$f(M|Y, \phi) = f(M|\phi) \quad (\forall Y, \phi)$$

MAR (Missing at random)

$$f(M|Y_{obs}, Y_{mis}, \phi) = f(M|Y_{obs}, \phi) \quad (\forall Y_{mis}, \phi)$$

Y_{obs} : 観測されたデータ , Y_{mis} : 欠測値

NMAR (Not missing at random)

1 変量データの場合

$Y = (y_1, \dots, y_n)^T$: 観測ベクトル

$M = (M_1, \dots, M_n)^T$: missing pattern

$(y_1, M_1), \dots, (y_n, M_n)$ は *i.i.d.*(独立同一分布) とする.

$$f(M, Y | \theta, \phi) = \prod_{i=1}^n f(M_i, y_i | \theta, \phi) = \prod_{i=1}^n f(y_i | \theta) f(M_i | y_i, \phi)$$

$f(M_i | y_i, \phi)$ は, ベルヌーイ分布 (2 値) の確率関数

*. この設定では, MCAR と MAR は同じしくみを表す.

MCAR の場合 : $f(M_i = 1 | y_i, \phi) = \phi$

$$\frac{\sum_{i=1}^n (1 - M_i) y_i}{\sum_{i=1}^n (1 - M_i)} : E(Y_i) \text{ の不偏推定量}$$

NMAR では, 偏りのある推定量となる

例 1.6 (人工データの茎葉図 (幹葉図) と標本平均)

$N(0, 1)$ からの 100 個のデータ

(a) $f(M_i = 1|y_i) = 0$ (欠測値なし)

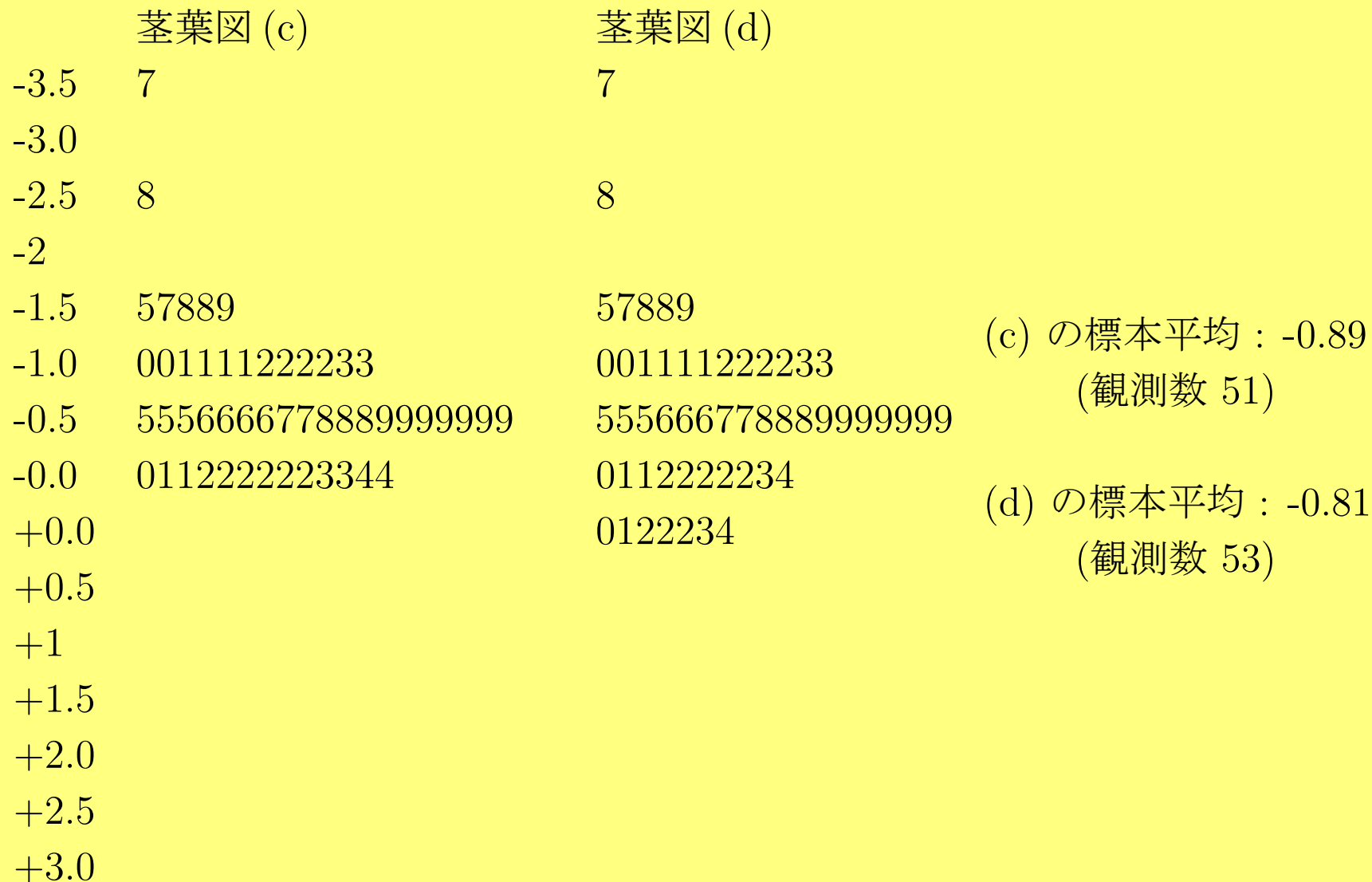
(b) $P(M_i = 1|y_i, \phi) = 0.5 \quad (\forall y_i)$ (MCAR)

(c) $P(M_i = 1|y_i, \phi) = \begin{cases} 1 & y_i > 0 \\ 0 & y_i \leq 0 \end{cases}$ (NMAR, censored data)

(d) $P(M_i = 1|y_i, \phi) = \Phi(2.05y_i)$ (NMAR, stochastic censored data)

ただし, Φ は標準正規分布関数

	茎葉図 (a)	茎葉図 (b)	
-3.5	7		
-3.0			
-2.5	8		
-2			
-1.5	57889	578	
-1.0	001111222233	1112233	(a) の標本平均 : -0.03
-0.5	5556666778889999999	5667888999999	
-0.0	0112222223344	011234	(b) の標本平均 : -0.11
+0.0	0011222222233344444	0122222234	(観測数 52)
+0.5	56777778899	677789	
+1	0011113444	11144	
+1.5	56778	6	
+2.0	023	02	
+2.5			
+3.0	3		



- 不完全データ (c) であっても, 母集団分布が対称で, 母平均のところで打ち切られたことが分かっているなら, 偏りの修正が可能.
- 母集団分布が正規分布であることが分かっているなら, どこで打ち切ったか分からなくても修正可能.
- 母集団分布が対称かどうかも分からなければ, 標本平均が偏るとも限らない.

例 1.7

$$\mathbf{Y}_i = (Y_{1i}, \dots, Y_{Ki})$$

$Y_{i1}, \dots, Y_{i,K-1}$: 欠損なし

Y_{iK} : 欠損あり ... パターン (a)

$$M = (M_1, \dots, M_n)$$

仮定 : $(\mathbf{Y}_1, M_1), \dots, (\mathbf{Y}_n, M_n)$ は互いに独立

$$\text{MCAR} : P(M_i = 1 | y_{i1}, \dots, y_{iK} : \phi) = \phi$$

計画的な欠損

$\mathbf{Y}_1, \dots, \mathbf{Y}_{K-1}$: 観測コスト小, Y_K : 観測するコスト大

⇒ 大標本を抽出して $\mathbf{Y}_1, \dots, \mathbf{Y}_{K-1}$ を観測し,

その一部を (無作為に) 選び, Y_K を観測

$$\text{MAR} : P(M_i = 1 | y_{i1}, \dots, y_{iK} : \phi) = P(M_i = 1 | y_{i1}, \dots, y_{i,K-1} : \phi)$$

MAR の確認

Y_K の欠損が計画されたものでない場合

Y_1, \dots, Y_{K-1} による Y_K と M の予測を行い, Y_K の予測値によって M の予測値が 1 となる確率が変化するかどうかを見る.

例 1.8 (単調欠測パターン)

Y_1, \dots, Y_K を K 時点での測定とする.

第 j 時点で欠測が起これば $j+1$ 時点以降も欠測が起こる場合

欠測の有無 (0,1) を表す代わりに, $M = j$ によって j 時点以降欠測が起こったとする. $M = K+1$ は完全データを表す.

$$\text{MCAR} : P(M_i = j | y_{i1}, \dots, y_{iK} : \phi) = \phi$$

$$\text{MAR} : P(M_i = j | y_{i1}, \dots, y_{iK} : \phi) = P(M_i = j | y_{i1}, \dots, y_{i,K-1} : \phi)$$

例 1.9 (2 変量データの欠損)

$$\mathbf{Y}_i = (Y_{i1}, Y_{i2}), \quad M_i = (M_{i1}, M_{i2})$$

仮定 : $(\mathbf{Y}_1, M_1), \dots, (\mathbf{Y}_n, M_n)$ は互いに独立

$$P(M_{i1} = r, M_{i2} = s | y_{i1}, y_{i2} : \phi) = g_{rs}(y_{i1}, y_{i2} : \phi) \quad r, s = 1, 2$$

MAR が成り立つとすると

$$g_{11}(y_{i1}, y_{i2} : \phi) = g_{11}(\phi)$$

$$g_{10}(y_{i1}, y_{i2} : \phi) = g_{10}(y_{i2} : \phi)$$

$$g_{01}(y_{i1}, y_{i2} : \phi) = g_{01}(y_{i1} : \phi)$$

$$g_{00}(y_{i1}, y_{i2} : \phi) = 1 - g_{11}(\phi) - g_{10}(y_{i2} : \phi) - g_{01}(y_{i1} : \phi)$$

g_{10}, g_{01} に関する仮定は不自然

より自然な仮定 (NMAR) は

Y_1 が欠損することと, Y_2 が欠損することが独立 :

$$g_{11}(y_{i1}, y_{i2} : \phi) = g_{1+}(y_{i1} : \phi)g_{+1}(y_{i2} : \phi)$$

$$g_{10}(y_{i1}, y_{i2} : \phi) = g_{1+}(y_{i1} : \phi)\{1 - g_{+1}(y_{i2} : \phi)\}$$

$$g_{01}(y_{i1}, y_{i2} : \phi) = \{1 - g_{1+}(y_{i1} : \phi)\}g_{+1}(y_{i2} : \phi)$$

$$g_{00}(y_{i1}, y_{i2} : \phi) = \{1 - g_{1+}(y_{i1} : \phi)\}\{1 - g_{+1}(y_{i2} : \phi)\}$$

1.4 欠損データ解析の分類

1. (完全データのみを用いる, procedures based on complete data) 欠測値を含む観測対象のデータはすべて除外

2. (加重法, weighting procedures)

母集団からの対象の抽出が無作為ではなく, 対象毎に抽出される確率が異なる場合

母平均の推定量として

$$\frac{\sum_{i=1}^n \pi_i^{-1} y_i}{\sum_{i=1}^n \pi_i^{-1}}, \quad \pi_i \text{ は対象 } i \text{ が選択される既知の確率}$$

が用いられる.

欠損が起こる場合, 欠損確率も抽出確率の一部と考える

$$\frac{\sum_{i=1}^n (\pi_i \hat{p}_i)^{-1} y_i}{\sum_{i=1}^n (\pi_i \hat{p}_i)^{-1}}, \quad \hat{p}_i \text{ は対象 } i \text{ が観測される確率}$$

3. (補完法, imputation-based procedures)

欠測値を適当な値で置き換えて完全データの分析方法を用いるもの.

平均値による補完 (mean imputation) : 対応する変数の観測値の平均値を代入

回帰による補完 (regression imputation) : 観測された他の変数による予測値を代入

4. (モデルに基づく方法, model-based procedures)

母数モデルを仮定して, 最尤法あるいはベイズ法を基に分析する.

2 分散分析における欠測値の扱い

2.1 分散分析とは

2元配置分散分析

例 2.1 2種類の肥料 A, B を濃度を変えて組み合わせて使用したときの作物の収穫量に対する影響を調べたい.

A\B	1	2	3
1	N_{11}	N_{12}	N_{13}
2	N_{21}	N_{22}	N_{23}
3	N_{31}	N_{32}	N_{33}

N_{ij} : A の濃度 i , B の濃度 j のときの実験回数

モデル

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

$$\varepsilon_{ijk} \ (i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, N_{ij}) \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$$

釣り合い型

$$N_{ij}N_{++} = N_{i+}N_{+j} \ (i = 1, \dots, a; j = 1, \dots, b),$$

$$\text{where } N_{i+} = \sum_{j=1}^b N_{ij}, \ N_{+j} = \sum_{i=1}^a N_{ij}, \ N_{++} = \sum_{i=1}^a \sum_{j=1}^b N_{ij}$$

識別性

パラメータの組み合わせ $(\mu, \alpha_i, \beta_j, \gamma_{ij})$ と $(\mu + c, \alpha_i - c, \beta_j)$ は同じ分布となり, データから区別できない (識別性の欠如)

$$\Rightarrow \sum_{i=1}^a N_{i+} \alpha_i = \sum_{j=1}^b N_{+j} \beta_j = 0, \quad \sum_{i=1}^a N_{i+} \gamma_{ij} = 0 \ (\forall j), \quad \sum_{j=1}^b N_{+j} \gamma_{ij} = 0 \ (\forall i),$$

推定量

$$\bar{y}_{ij} = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} y_{ijk}, \quad \bar{y}_{i*} = \frac{1}{N_{i+}} \sum_{j=1}^b \sum_{k=1}^{N_{ij}} y_{ijk},$$
$$\bar{y}_{*j} = \frac{1}{N_{+j}} \sum_{i=1}^a \sum_{k=1}^{N_{ij}} y_{ijk}, \quad \bar{y}_{**} = \frac{1}{N_{++}} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} y_{ijk}$$

$$\Rightarrow \hat{\mu} = \bar{y}_{**}, \quad \hat{\alpha}_i = \bar{y}_{i*} - \bar{y}_{**}, \quad \hat{\beta}_j = \bar{y}_{*j} - \bar{y}_{**},$$
$$\hat{\gamma}_{ij} = \bar{y}_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y}_{**}$$

変動の分解

$$S_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (y_{ijk} - \bar{y}_{**})^2 = S_W + S_A + S_B + S_{AB},$$

$$S_W = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (y_{ijk} - \bar{y}_{ij})^2, \quad S_A = \sum_{i=1}^a N_{i+} (\bar{y}_{i*} - \bar{y}_{**})^2,$$

$$S_B = \sum_{j=1}^b N_{+j} (\bar{y}_{*j} - \bar{y}_{**})^2, \quad S_{AB} = \sum_{i=1}^a \sum_{j=1}^b N_{ij} (\bar{y}_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y}_{**})^2$$

S_W, S_A, S_B, S_{AB} は独立, $S_W/\sigma^2 \sim \chi_{N_{+++}-ab}^2$,

$H_A : \alpha_i = 0 \ (i = 1, \dots, a) \Rightarrow S_A/\sigma^2 \sim \chi_{a-1}^2$,

$H_B : \beta_j = 0 \ (j = 1, \dots, b) \Rightarrow S_B/\sigma^2 \sim \chi_{b-1}^2$,

$H_{AB} : \gamma_{ij} = 0 \ (i = 1, \dots, a; j = 1, \dots, b) \Rightarrow S_{AB}/\sigma^2 \sim \chi_{ab-a-b+1}^2$

分散分析

$$V_W = \frac{1}{N_{++++} - ab} S_W : \text{級内分散},$$

$$V_a = \frac{1}{a - 1} S_a : \text{主効果 A に関する分散},$$

$$V_b = \frac{1}{b - 1} S_b : \text{主効果 B に関する分散},$$

$$V_{ab} = \frac{1}{ab - a - b + 1} S_{ab} : \text{交互作用に関する分散}$$

H_A の検定

$$\frac{V_a}{V_W} > F_{a-1, N_{++++}-ab}(\alpha) \Rightarrow \text{優位水準 } \alpha \text{ 棄却}$$

H_B の検定

$$\frac{V_b}{V_W} > F_{b-1, N_{+++}-ab}(\alpha) \Rightarrow \text{優位水準 } \alpha \text{ 棄却}$$

H_{AB} の検定

$$\frac{V_{ab}}{V_W} > F_{ab-a-b+1, N_{+++}-ab}(\alpha) \Rightarrow \text{優位水準 } \alpha \text{ 棄却}$$

注. $N_{ij} = 1$ ($\forall i, j$) のときは, $S_W \equiv 0$ となるので, V_W の代わりに V_{AB} を用いる.

交互作用の検定はできない.

注. 2元配置分散分析モデルは線形回帰モデルなので、釣り合い型でなくても、 $\mu, \alpha_i, \beta_j, \gamma_{ij}$ 等の推定や H_A, H_B, H_{AB} 等の仮説検定は可能であるが、推定量や検定統計量の形が複雑になり、また、 H_A の解釈も異なる。

注. 欠測は通常、応答変数 (y_{ijk}) に起こるが、完全データのみで解析しようとするとは釣り合い型ではなくなってしまう。

⇒ 欠測値を、完全データから推定したパラメータによる予測値で置き換えて、釣り合い型の手法を用いる。

最小 2 乗法 2 元配置分散分析で, $a = b = 2$, $N_{ij} = 2$ とする. 交互作用はないものとする ($\gamma_{ij} = 0$). モデルの識別性のために, $\alpha_1 = 0, \beta_1 = 0$ とするとモデルは

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\begin{aligned}\mathbf{y} &= (y_{111}, y_{112}, y_{121}, y_{122}, y_{211}, y_{212}, y_{221}, y_{222})^T \\ &= (y_1, y_2, \dots, y_n)^T \quad (n = 8)\end{aligned}$$

$$\boldsymbol{\varepsilon} = (\varepsilon_{111}, \varepsilon_{112}, \varepsilon_{121}, \varepsilon_{122}, \varepsilon_{211}, \varepsilon_{212}, \varepsilon_{221}, \varepsilon_{222})^T,$$

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}^T = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n),$$

$$\boldsymbol{\beta} = (\mu, \alpha_2, \beta_2)^T.$$

推定量

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y},$$

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n - p}, \quad (p = 3),$$

$$\hat{y}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}},$$

$$V := \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (X^T X)^{-1}$$

線形仮説の検定

$$H_0 : C^T \boldsymbol{\beta} = \mathbf{0}, \quad (C : p \times w)$$

の検定は, $F > F_{w, n-p}(\alpha) \Rightarrow$ 棄却. ただし

$$F = \frac{S/w}{\hat{\sigma}^2}, \quad S = (C^T \hat{\boldsymbol{\beta}})^T \{C^T (X^T X)^{-1} C\}^{-1} (C^T \hat{\boldsymbol{\beta}})$$

2.2 欠測値の予測

仮定

- 欠損構造：MAR を仮定
- 欠損の有無 (M) の分布は, 回帰モデルの未知母数に依存しない.

欠測値： $y_1, \dots, y_m,$

観測値： y_{m+1}, \dots, y_{m+r} ($r + m = n$)

$\hat{\beta}_*, \hat{\sigma}_*^2, V_*, S_*$: 完全データ $(y_1, \mathbf{x}_1), \dots, (y_r, \mathbf{x}_r)$ のみで計算された
 $\hat{\beta}, \hat{\sigma}^2, V, S$ の値

Yates (1933) の方法

$$SS(\boldsymbol{\beta}, y_1, \dots, y_m) = \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2$$

を最小とする y_{r+1}, \dots, y_n と $\boldsymbol{\beta}$ を求める.

$$\begin{aligned} \min SS(\boldsymbol{\beta}, y_1, \dots, y_m) &= SS(\hat{\boldsymbol{\beta}}_*, \hat{y}_1, \dots, \hat{y}_m), \\ \hat{y}_i &= \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_* \quad (i = 1, \dots, m) \end{aligned}$$

欠測値 y_i を予測値 \hat{y}_i で置き換えて最小 2 乗法を適用すると,

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_*$$

となるが, そのときの $\hat{\sigma}^2$ は不偏推定量とならず,

$$\hat{\sigma}_*^2 = \hat{\sigma}^2 \frac{n-p}{r-p}$$

反復法による欠測値の予測 (Healy and Westmacott (1956))

初期値 $\tilde{y}_i^{(0)}$ ($i = 1, \dots, m$) を適当に定め,

$k = 0, 1, \dots$, として, 収束するまで (1), (2) を反復する方法

(1) 欠測値 y_i に $\tilde{y}_i^{(k)}$ を代入し, 標本数 n の完全データと違って, 推定量 ($= \boldsymbol{\beta}^{(k)}$) を計算する.

$$(2) \tilde{y}_i^{(k+1)} = \mathbf{x}_i^T \boldsymbol{\beta}^{(k)}$$

欠損パターンが複雑でも適用可能 (EM アルゴリズム)

Bartlett の ANCOVA 法

$$Z = \begin{pmatrix} I_m \\ O \end{pmatrix} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^T : \text{missing-value covariate}$$

Z を説明変数に加えたモデル

$$\mathbf{y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \mathbf{e}$$

欠測値の初期値 \tilde{y}_i ($i = 1, \dots, m$) を適当に定め

$$SS(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^m (\tilde{y}_i - \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma})^2 + \sum_{i=m+1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma})^2$$

を最小とする $\boldsymbol{\beta}, \boldsymbol{\gamma}$ を求めると,

$$\min SS(\boldsymbol{\beta}, \boldsymbol{\gamma}) = SS(\hat{\boldsymbol{\beta}}_*, \hat{\boldsymbol{\gamma}}), \quad \hat{\gamma}_i = \tilde{y}_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_*, \quad i = 1, \dots, m$$

$$\Rightarrow \hat{y}_i = \tilde{y}_i - \hat{\gamma}_i$$

何の役に立つの？

$$\hat{\sigma}_*^2 = \frac{1}{n - m - p} \text{SS}(\hat{\beta}_*, \hat{\gamma})$$

ハット行列 $\{(X, Z)^T(X, Z)\}^{-1}$ の左上の $p \times p$ 行列を U とすると

$$V_* = \hat{\sigma}_*^2 U$$

$$\hat{\boldsymbol{\gamma}} = B^{-1}\boldsymbol{\rho}$$

ただし, $B = (B_{jk})$ は, \mathbf{z}_i

$$\mathbf{z}_i = X\boldsymbol{\delta} + \mathbf{e}$$

による予測値を $\hat{\mathbf{z}}_i = (\hat{z}_{1i}, \dots, \hat{z}_{ni})$ とすると

$$b_{jk} = \sum_{i=1}^n (z_{ij} - \hat{z}_{ij})(z_{ik} - \hat{z}_{ik})$$

$$\boldsymbol{\rho} = (\tilde{y}_1 - \mathbf{x}_1^T \hat{\boldsymbol{\beta}}, \dots, \tilde{y}_m - \mathbf{x}_m^T \hat{\boldsymbol{\beta}})^T,$$

$\hat{\boldsymbol{\beta}}$ は, 欠測値を \tilde{y}_i で置き換えた完全データから得られる最小2乗推定値

例 2.2 酸化カリウムの散布量の綿花強度に関する影響に関する実験
(乱塊法 (randomized block design))

散布量	ブロック			Total
	1	2	3	
36	u_1	8.00	7.93	15.93
54	8.14	8.15	7.87	24.16
72	7.76	u_2	7.74	15.50
108	7.17	7.75	7.80	22.54
144	7.46	7.68	7.21	22.35
Total	30.53	31.40	38.55	100.48

u_1, u_2 (欠測値) に初期値 $\bar{y} = 7.7292$ (総平均) を代入して, u_1, u_2 の予測残差を計算すると

$$\boldsymbol{\rho} = -(0.0798, 0.1105)^T$$

$u_1 = 1$ を代入し, 他のすべてを 0 とした \mathbf{z}_1 の予測残差, $u_2 = 1$ を代入し, 他のすべてを 0 とした \mathbf{z}_2 の予測残差を計算すると

$$B = \begin{pmatrix} 0.5333 & 0.00667 \\ 0.00667 & 0.5333 \end{pmatrix}, \quad B^{-1} = \begin{pmatrix} 1.9408 & -0.2381 \\ -0.2381 & 1.9408 \end{pmatrix}$$

となり, 欠測値の予測値は

$$(\bar{y}, \bar{y})^T - B^{-1}\boldsymbol{\rho} = (7.8549, 7.9206)^T$$

釣り合い型なので, 行平均, 列平均の計算だけで B や $\boldsymbol{\rho}$ の計算ができる

3 完全データ分析と部分完全データ分析

3.1 完全データのみによる分析

利点：既存の分析法がそのまま使える。

欠点：精度が落ちる（利用できる情報を生かしていない）

MCAR でないとき、完全データは全データからのランダム標本と見なせないので偏りが生じる。

情報の損失

θ ：パラメータ

$\hat{\theta}_{CC}$ ：完全データによる推定量

$\hat{\theta}_{NM}$ ：欠損が起こらなかった場合の推定量

$\hat{\theta}_{EFF}$ ：利用できる全データを用いた有効推定量（最尤法など）

$$\text{Var}(\hat{\theta}_{CC}) = \text{Var}(\hat{\theta}_{NM})(1 + \Delta_{CC}^*),$$

$$\text{Var}(\hat{\theta}_{CC}) = \text{Var}(\hat{\theta}_{EFF})(1 + \Delta_{CC}),$$

例 3.1 単調欠測データの効率 (2 変量正規分布)

$$(Y_{i1}, Y_{i2})^T \quad (i = 1, 2, \dots, n) \quad \overset{i.i.d.}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

n 個中, r 個が完全データ, $n - r$ 個は, Y_2 のみ欠測値

MCAR を仮定

$E(Y_1)$ を標本平均で推定するとき

$$\Delta_{\text{CC}}^* = \Delta_{\text{CC}} = \frac{n - r}{r}$$

$E(Y_2)$ を標本平均で推定するとき

$$\Delta_{\text{CC}}^* = \frac{n - r}{r}, \quad \Delta_{\text{CC}} \approx \frac{(n - r)\rho^2}{n(1 - \rho^2) + r\rho^2}, \quad (0 \text{ と } \Delta_{\text{CC}}^* \text{ の間の値をとる})$$

$\rho = \text{Corr}(Y_1, Y_2)$ (相関係数)

(導出は後日)

※ MCAR の妥当性 : 完全データの Y_1 の分布と 不完全データの Y_1 の分布を比較. (MAR の妥当性はわからない)

例 3.2 母平均の推定におけるバイアス

$E(Y)$ の推定

$$\mu_{CC} = E(Y|M = 1), \quad \mu_{IC} = E(Y|M = 0), \quad \pi_{CC} = P(M = 1)$$

$$\Rightarrow \mu = E(Y) = \pi_{CC}\mu_{CC} + (1 - \pi_{CC})\mu_{IC}$$

なので、完全データのみでの標本平均のバイアスは

$$\mu_{CC} - \mu = (1 - \pi_{CC})(\mu_{CC} - \mu_{IC})$$

MCAR ならば $\mu_{CC} = \mu_{IC}$ より バイアスは 0.

∴

$$(Y_1, M_1), \dots, (Y_n, M_n) : \overset{i.i.d.}{\sim} (Y, M), \quad \bar{Y}_{CC} = \frac{\sum_{i=1}^n Y_i 1_{M_i=1}}{\sum_{i=1}^n 1_{M_i=1}},$$

$$E(\bar{Y}_{CC} | M_1, \dots, M_n) = \mu_{CC}$$

例 3.3 回帰係数の推定におけるバイアス

Y : 目的変数, X_1, \dots, X_p : 説明変数

Y にも, X_1, \dots, X_p にも欠損が起こりうる.

回帰モデルが正しいものと仮定する.

(X_1, \dots, X_p, Y) が完全データとなる確率が Y に依存しなければ完全データのみによる回帰係数の推定量にバイアスは生じない. \Leftarrow NMAR でも良い.
(Glynn and Laird, 1986)

例 3.4 オッズ比の推定におけるバイアス

Y_1, Y_2 2 値変数 (0 or 1), $\frac{P(Y_1 = 1)P(Y_1 = 0)}{P(Y_2 = 1)/P(Y_2 = 0)}$ の推定

$$\log P(M_1 = 1, M_2 = 1 | Y_1, Y_2) = g_1(Y_1) + g_2(Y_2)$$

ならば, 完全データのみで推定してもバイアスは生じない \Rightarrow Case-control Study の妥当性 (Kleinbaum, Morgenstern and Kupper, 1981)

3.2 加重法による完全データの分析

加重法

例 3.5 層別抽出法 (stratified random sampling)

有限母集団 $Y = (y_1, \dots, y_N)$

抽出の有無 $I = (I_1, \dots, I_N)$: 確率変数

$$I_i = \begin{cases} 1 & y_i \text{ を抽出する} \\ 0 & y_i \text{ を抽出しない} \end{cases}$$

層の指標 $Z = (z_1, \dots, z_N)$

$$y_i \text{ が第 } j \text{ 層に所属} \Rightarrow z_i = j \quad (j = 1, \dots, J), \quad N_j = \sum_{i=1}^N 1_{z_i=j}$$

$$f(I|Y, Z) = \begin{cases} \prod_{j=1}^J \binom{N_j}{n_j}^{-1} & \text{if } \sum_{i:z_i=j} I_i = n_j \text{ for all } j \\ 0 & \text{otherwise} \end{cases}$$

母平均の推定 $T = \bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$

$$t = \bar{y}_{st} = \frac{1}{N} \sum_{j=1}^J N_j \bar{y}_j, \quad \bar{y}_j = \frac{1}{n_j} \sum_{i:z_i=j} I_i y_i$$

\bar{y}_{st} の分散の不偏推定

$$\text{Var}(\bar{y}_{st}) = \frac{1}{N^2} \sum_{j=1}^J N_j^2 \left(\frac{1}{n_j} - \frac{1}{N_j} \right) S_{yj}^2, \quad S_{yj}^2 : \text{第 } j \text{ 層の分散},$$

$$v_{st} = \widehat{\text{Var}}(\bar{y}_{st}) = \frac{1}{N^2} \sum_{j=1}^J N_j^2 \left(\frac{1}{n_j} - \frac{1}{N_j} \right) s_{yj}^2, \quad s_{yj}^2 : \text{第 } j \text{ 層の標本(不偏)分散},$$

信頼区間

$$\bar{y}_{st} \pm 1.96 \sqrt{v_{st}}$$

欠損データへの応用

$$\bar{y}_{\text{st}} = \frac{1}{N} \sum_{j=1}^J \frac{N_j}{n_j} \sum_{i:z_i=j} I_i y_i = \frac{1}{N} \sum_{j=1}^J \sum_{i:z_i=j} \pi_i^{-1} I_i y_i = \frac{1}{N} \sum_{i=1}^N \pi_i^{-1} I_i y_i$$

$$\pi_i = \text{P}(I_i = 1) = \frac{n_j}{N_j} \quad (\text{if } z_i = j)$$

$$N = \sum_{j=1}^J n_j \frac{N_j}{n_j} = \sum_{j=1}^J \sum_{i:z_i=j} I_i \frac{N_j}{n_j} = \sum_{i=1}^N I_i \pi_i^{-1}$$

大きさ $n = \sum_j^J n_j$ の標本を Y_1, \dots, Y_n とすると

$$\bar{y}_{\text{st}} = \frac{1}{n} \sum_{i=1}^n w_i Y_i, \quad w_i = \frac{n \pi_i^{-1}}{\sum_{k=1}^n \pi_k^{-1}}$$

y_i が選択され、観測される (欠測値とならない) 確率を

$$P(\text{選択かつ観測される}) = P(\text{選択}) \times P(\text{観測される} \mid \text{選択}) = \pi_i \times \phi_i$$

と表すと、欠損データによる母平均の推定量は

$$\bar{y}_w = \frac{1}{r} \sum_{i=1}^r w_i Y_i,$$

$$r : \text{観測されたデータ数}, \quad w_i = \frac{r(\pi_i \phi_i)^{-1}}{\sum_{k=1}^r (\pi_k \phi_k)^{-1}}$$

ϕ_i は未知であるが、各層で一定とする。

$$\hat{w}_i = \frac{r(\pi_i \hat{\phi}_i)^{-1}}{\sum_{k=1}^r (\pi_k \hat{\phi}_k)^{-1}}, \quad \hat{\phi}_i = \frac{r_j}{n_j}, \quad r_j : \text{第 } j \text{ 層から観測されたデータ数}$$

単純無作為抽出であるが, 個体ごとに欠測値となる確率が異なる場合

仮定 3.1 母集団を J 個のクラスに分けたとき, クラス内で個体が欠測値となる確率が等しい.

仮定 3.1 のクラスを加重クラスと呼ぶ.

C : 加重クラスを表す変数

n_j : 大きさ n の標本中, $C = j$ であったものの個数

r_j : n_j 個の標本中, 観測されたもの ($n_j - r_j$ 個は欠測値)

$$r = \sum_{j=1}^J r_j$$

各個体が標本に含まれる確率 π_i は全て等しいので

$$\hat{w}_i = \frac{r\hat{\phi}_i^{-1}}{\sum_{k=1}^r \hat{\phi}_k^{-1}} = \frac{r\hat{\phi}_i^{-1}}{n}, \quad \hat{\phi}_i = \frac{r_j}{n_j} \text{ if } C_i = j$$

$$\bar{y}_{wc} = \frac{1}{r} \sum_{i=1}^r \hat{w}_i Y_{ic} = \frac{1}{n} \sum_{j=1}^J \sum_{i:C_i=j} \frac{n_j}{r_j} Y_{iR} = \frac{1}{n} \sum_{j=1}^J n_j \bar{Y}_{jR}$$

ただし,

Y_{1R}, \dots, Y_{rR} : 観測された値

\bar{Y}_{jR} : 加重クラス j から観測された変数の標本平均.

$$\text{Var}(\bar{Y}_{wc}) = \sum_{j=1}^J \left(\frac{n_j}{n}\right)^2 f_j S_j^2,$$

S_j^2 : 加重クラス j 内での Y の母分散,

N_j : 加重クラス j 内の個体数, $f_j = r_j^{-1} - N_j^{-1}$

(Oh and Scheuren, 1983)

$$\widehat{\text{mse}}(\bar{Y}_{\text{wc}}) = \sum_{j=1}^J \left(\frac{n_j}{n}\right)^2 \left(1 - \frac{r_j n}{n_n N}\right) \frac{s_{jR}^2}{r_j} + \frac{N - n}{(N - 1)n^2} \sum_{j=1}^J n_j (\bar{Y}_{jR} - \bar{Y}_{\text{wc}})^2,$$

s_{jR}^2 : 加重クラス j 内での Y の標本分散

(Oh and Scheuren, 1983)

傾向スコア (Propensity Score)

X : 欠測が起こらない変数

MAR の仮定 $P(M|X, Y) = P(M|X)$

X の取りうる値が有限個, $c_1, \dots, c_J \Rightarrow$

$$C = j \quad \text{if } X = c_j$$

問題点 : J が大きいとき n_j が小さくなるので $\hat{\phi}_i$ が不安定.

傾向スコア

$$p(x) = P(M = 0|X = x)$$

$$\begin{aligned} P(M = 0|Y, p(X)) &= E[P(M = 0)|Y, X] | Y, p(X)] \\ &= E[P(M = 0)|X] | Y, p(X)] \\ &= E[P(X)|Y, p(X)] = p(X) = P(M = 0|p(X)) \end{aligned}$$

加重クラス

$J = 5 \sim 6$, $0 = p_0 < p_1 < \dots < p_J = 1$ として

$$C = j \text{ if } p_{j-1} < (\leq) p(X) \leq p_j$$

$p(x)$ の推定

M を目的変数, X を説明変数として ロジスティック回帰あるいはプロビット回帰により推定.

例 3.6 一般化推定方程式 (GEE) に対する加重法

$\mathbf{y}_i = (y_{i1}, \dots, y_{iK})^T$: 個体 i の観測値 ($i = 1, \dots, n$)

$i = 1, \dots, r \Rightarrow$ 完全データ

$i = r + 1, \dots, n$ 一部, あるいは全てが欠測値

$$m_i = \begin{cases} 1 & \mathbf{y}_i \text{が不完全データ} \\ 0 & \mathbf{y}_i \text{が完全データ} \end{cases}$$

GEE ($r = n$ の場合)

$$\sum_{i=1}^n D_i(\mathbf{x}_i, \boldsymbol{\beta}) [\mathbf{y}_i - \mathbf{g}(\mathbf{x}_i, \boldsymbol{\beta})] = \mathbf{0},$$

$\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$: 説明変数, 完全データ ($i = 1, \dots, n$),

$\boldsymbol{\beta} : d \times 1$ 未知回帰係数ベクトル, $D_i(\mathbf{x}_i, \boldsymbol{\beta}) : d \times K$ 行列

$\mathbf{g}(\mathbf{x}_i, \boldsymbol{\beta}) : K \times 1$ 回帰関数

Weighted GEE ($r < n$) (Robins, Rotnitzky and Zhao, 1995)

$\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^T$: 補助変数, 完全データ ($i = 1, \dots, n$)

\mathbf{y}_i が完全データとなるかどうかには影響を持つ可能性
があるが, 回帰モデルには組み込まれない変数

$$P(m_i = 1 | \mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i) = P(m_i = 1 | \mathbf{x}_i)$$

$$\Rightarrow \sum_{i=1}^r D_i(\mathbf{x}_i, \boldsymbol{\beta}) [\mathbf{y}_i - \mathbf{g}(\mathbf{x}_i, \boldsymbol{\beta})] = \mathbf{0} \text{ は一貫性を持つ}$$

$$P(m_i = 1 | \mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i) = P(m_i = 1 | \mathbf{x}_i, \mathbf{z}_i)$$

$$\Rightarrow \sum_{i=1}^n w_i(\hat{\alpha}) D_i(\mathbf{x}_i, \boldsymbol{\beta}) [\mathbf{y}_i - \mathbf{g}(\mathbf{x}_i, \boldsymbol{\beta})] = \mathbf{0},$$

$$w_i(\hat{\alpha}) = \frac{1}{p(\mathbf{x}_i, \mathbf{z}_i, \hat{\alpha})} : \text{完全データとなる確率の推定量}$$

$\hat{\alpha}$: m_i の $\mathbf{x}_i, \mathbf{z}_i$ によるロジスティック回帰係数

加重法による分散の増加

$$Y_1, \dots, Y_r : i.i.d., \text{Var}(Y_i) = \sigma^2$$

$$\text{Var}\left(\frac{1}{r} \sum_{i=1}^r w_i Y_i\right) = \frac{\sigma^2}{r^2} \sum_{i=1}^r w_i^2 = \frac{\sigma^2}{r^2} \{1 + cv^2(W)\},$$

$cv^2(W)$: 加重値の変動係数

Post-Stratification and Raking to Known Margins

加重クラス推定量

$$\bar{y}_{wc} = \sum_{j=1}^J \frac{n_j}{n} \bar{Y}_{jR}$$

は、母集団での加重クラスの比率 N_j/N を、標本比率 n_j/n で推定したもの

post-stratified mean (N_j/N が既知のとき)

$$\bar{y}_{ps} = \frac{1}{N} \sum_{j=1}^J N_j \bar{Y}_{jR},$$

$$\text{Var}(\bar{y}_{ps}) = \frac{1}{N^2} \sum_{j=1}^J N_j^2 \left(1 - \frac{r_j}{N_j}\right) \frac{S_{jR}^2}{r_j}$$

例 3.7 Raing Ratio Estimation

加重クラス $(j, l) : X_1 = j, X_2 = l \quad (i = 1, \dots, J; l = 1, \dots, L)$

$$\bar{y}_{\text{ps}} = \frac{1}{N} \sum_{j=1}^J \sum_{l=1}^L N_{jl} \bar{Y}_{jlR}, \quad \bar{y}_{\text{wc}} = \frac{1}{n} \sum_{j=1}^J \sum_{l=1}^L n_{jl} \bar{Y}_{jlR},$$

$N_{j+} = \sum_{l=1}^L N_{jl}, N_{+l} = \sum_{j=1}^J N_{jl}$ は既知だが, N_{jl} は未知の場合

$$\left\{ \begin{array}{l} N_{jl}^* = a_j b_l n_{jl}, \quad j = 1, \dots, J; l = 1, \dots, L, \\ N_{j+}^* = \sum_{l=1}^L N_{jl}^* = N_{j+}, \quad j = 1, \dots, J, \\ N_{+l}^* = \sum_{j=1}^J N_{jl}^* = N_{+l}, \quad l = 1, \dots, L, \end{array} \right.$$

$$\Rightarrow \bar{y}_{\text{rake}} = \frac{1}{N} \sum_{j=1}^J \sum_{l=1}^L N_{jl}^* \bar{Y}_{jlR},$$

3.3 部分完全データ分析 (Avalable case analysis)

例.

$$\begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & * \\ y_{41} & y_{42} & * \\ y_{51} & * & y_{53} \end{pmatrix}$$

Cov(Y_1, Y_2) の推定

完全データ分析 : $\mathbf{y}_1, \mathbf{y}_2$ のみ利用

部分完全データ分析 : $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4$ を利用

Y_3 の欠損が Y_1, Y_2 に依存しているとき,

(y_{11}, y_{12}) と布と (y_{31}, y_{32}) は同一ではない. (MCAR の仮定が必要)

相関係数の推定

I_{jk} : Y_j と Y_k が共に観測されている個体番号の集合 ($j \neq k$)

I_j : Y_j が観測されている個体番号の集合

$$s_{jk}^{(jk)} = \sum_{i \in I_{jk}} (y_{ij} - \bar{y}_j^{(jk)})(y_{ik} - \bar{y}_k^{(jk)}) / (n^{(jk)} - 1),$$
$$r_{jk}^* = \frac{s_{jk}^{(jk)}}{\sqrt{s_{jj}^{(j)} s_{kk}^{(k)}}}, \quad r_{jk} = \frac{s_{jk}^{(jk)}}{\sqrt{s_{jj}^{(jk)} s_{kk}^{(jk)}}}$$

. $r_{jk}^ \in [-1, 1]$ とは限らない!

(その他の分散の推定)

$$s_{jk}^* = r_{jk}^{(jk)} \sqrt{s_{jj}^{(j)} s_{kk}^{(k)}}, \quad \tilde{s}_{jk}^{(jk)} = \sum_{i \in I_{jk}} (y_{ij} - \bar{y}_j^{(j)})(y_{ik} - \bar{y}_k^{(k)}) / (n^{(jk)} - 1)$$

$r_{jk}^{(jk)}$ も問題が生じる

Y_1	1	2	3	4	1	2	3	4	?	?	?	?
Y_2	1	2	3	4	?	?	?	?	1	2	3	4
Y_3	?	?	?	?	1	2	3	4	4	3	2	1

$$r_{12}^{(12)} = r_{13}^{(13)} = 1, r_{23}^{(2,3)} = -1$$

$$\text{Cov}(Y_1, Y_2) = \text{Cov}(Y_1, Y_3) = 1 \Rightarrow \text{Cov}(Y_2, Y_3) = 1$$

4 欠測値の補完

欠測値の補完 (imputaion) : 観測値を用いて欠測値を予測し, その値を欠測値の代用とする方法

観測値を用いて予測分布を構成し, 予測分布の平均値, あるいは, 予測分布に従う乱数により補完する.

予測分布の構成方法 : Explicit modeling, Implicit modeling

Explicit modeling : 正規分布など, 統計モデルを仮定して予測分布を構成する方法

Implicit modeling : 補完法のアルゴリズムを与える. 予測分布の統計モデルを仮定したことになる.

Explicit imputation の代表的な方法

Mean imputation

観測データの標本平均で補完する方法

Regression imputation

補完しようとしている変数を目的変数, 他の変数を説明変数とする回帰により予測する方法

Stochastic-regression imputation

回帰による予測値に乱数を用いた残差を加えて値で補完する方法.
正規性を仮定した線形回帰の場合
 $N(0, \hat{\sigma})$ に従う乱数を用いる. $\hat{\sigma}$ は残差平方和による分散の推定値

Implicit imputation

Hot deck imputaion

「良く似た個体」から抽出した値で補完する方法

Substitution

欠損が起こった個体と同様の個体を標本外から選び、観測値を得る.

Cold deck imputaion

過去の同様な調査で得ていた値で置き換えるなど, 外部から得た定数値で補完する方法

Composit methods

上記の組み合わせ. 回帰による予測値+残差からの無作為抽出で補完するなど.

4.1 予測分布の平均による補完

平均値による補完

y_{ij} : 欠測値 $\Leftarrow \bar{y}_j^{(j)}$

*補完後に、完全データと思って分散を推定すると実際には

$$s_{jj}^{(j)} \frac{n^{(j)} - 1}{n - 1}, \quad n^{(j)} : \text{観測された } Y_j \text{ の個体数}$$

MCAR の仮定の下で $s_{ij}^{(j)}$ は分散の一致推定量なので、補完後に完全データとして分散を推定すると過小評価 $\Rightarrow \frac{n-1}{n^{(j)}-1}$ 倍して調整

*補完後に、完全データと思って Y_j と Y_k の共分散を推定すると

$$\tilde{s}_{jk}^{(jk)} \frac{n^{(jk)} - 1}{n - 1}, \quad n^{(jk)} : Y_j, Y_k \text{ が共に観測された個体数}$$

MCAR の仮定の下で、共分散も過小評価 $\Rightarrow \frac{n-1}{n^{(jk)}-1}$ 倍して調整

条件付平均による補完

観測値が与えられた条件付平均で補完する方法

例 4.1 補正クラスを用いた補完

クラス内で欠損確率が同じとなるように, J 個のクラスに分割できたとする.

n_j : 第 j クラスの個体数

r_j : 第 j クラスからの観測値の個数

\bar{y}_{jR} : 第 j クラスからの観測値の標本平均 $= \frac{1}{r_j} \sum_{i=1}^{r_j} y_{ij}$

\Rightarrow クラスごとに $n_j - r_j$ 個の欠測値を \bar{y}_{jR} で補完

補完後の標本平均は

$$\frac{1}{n} \sum_{j=1}^J \left(\sum_{i=1}^{r_j} y_{ij} + \sum_{i=r_j+1}^{n_j} \bar{y}_{jR} \right) = \frac{1}{n} \sum_{j=1}^J n_j \bar{y}_{jR} = \bar{y}_{wc}$$

加重クラスを用いた平均の推定法と同じ結果を得る

例 4.2 回帰による補完

Y_1, \dots, Y_{K-1}, Y_K の内, Y_K のみに欠損が起こるとする.
完全データから回帰係数を推定し欠測値 y_{iK} を

$$\hat{y}_{iK} = \tilde{\beta}_{K0 \cdot 12 \dots K-1} + \sum_{j=1}^{K-1} \tilde{\beta}_{Kj \cdot 12 \dots K-1} y_{ij}$$

で補完

*. Y_1, \dots, Y_{K-1} がカテゴリー変数の場合, 例 4.1 と同じ形となる.

例 4.3 Buck's method (Buck, 1960)

完全データから, 平均ベクトル $\boldsymbol{\mu}$ と共分散行列 Σ を推定
欠損パターンごとに, 欠測値を線形回帰で予測

ex. Y_1, Y_3 が欠測値, Y_2, Y_4, Y_5 が観測されていれば
 $(Y_1, Y_3)^T$ を Y_2, Y_4, Y_5 で予測

*. 同時分布が正規分布であるとき有効

*. 補完後の分散や共分散は過小評価となるが, 平均による補完よりは過小の程度が小さい.

4.2 予測分布からの無作為抽出

予測分布の平均値による補完：

平均予測誤差を最小とする \Leftrightarrow 補完後の分散・共分散は過小評価となる

\Rightarrow 予測分布からの無作為標本により補完

例 4.4 確率的回帰補完

Y_1, \dots, Y_{K-1}, Y_K の内, Y_K のみに欠損が起こるとする.
完全データから回帰係数を推定し欠測値 y_{iK} を

$$\hat{y}_{iK} = \tilde{\beta}_{K0 \cdot 12 \dots K-1} + \sum_{j=1}^{K-1} \tilde{\beta}_{Kj \cdot 12 \dots K-1} y_{ij} + z_{iK},$$

$$z_{iK} \sim N(0, \tilde{\sigma}_{KK \cdot 12 \dots K-1}),$$

$\tilde{\sigma}_{KK \cdot 12 \dots K-1}$: 完全データによる $\text{Cov}(Y_K)$ の推定量

例 4.5 2変量単調欠測データでの比較

$$(Y_1, Y_2) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

Y_1 : 欠損なし

$$\text{MCAR を仮定 : } P(Y_2 \text{ が欠測}) = \lambda = \frac{n - r}{n}$$

比較手法

1. Umean : Y_2 の欠測値を 観測された Y_2 の標本平均 \bar{y}_{2R} で補完
2. Udraw : $\bar{y}_{2R} + Z_2$ で補完
 $Z_2 \sim N(0, \tilde{\sigma}_{22})$, σ_{22} : 完全データによる (Y_2) の推定値
3. Cmean : Conditional mean, 例 4.2 の 回帰による予測値で補完
4. Cdraw : Conditonal draw, 例 4.4 による補完

補完後の Y_2 に関するパラメータの推定量の漸近バイアス

Method	パラメータ			
	μ_2	σ_{22}	$\beta_{21.1}$	$\beta_{11.2}$
Umean	0^*	$-\lambda\sigma_{22}$	$-\lambda\beta_{21.1}$	0^*
Udraw	0	0	$-\lambda\beta_{21.1}$	$-\lambda\beta_{11.2}$
Cmean	0	$-\lambda(1 - \rho^2)\sigma_{22}$	0^*	$\frac{\lambda(1-\rho^2)}{1-\lambda(1-\rho^2)}\beta_{12.2}$
Cdraw	0	0	0	0

0^* : 完全データによる推定値と同じ

$$E(Y_2|Y_1) = \beta_{20.1} + \beta_{21.1}Y_1$$

$$E(Y_1|Y_2) = \beta_{10.2} + \beta_{11.2}Y_2$$

*. 第4列, Y_1 を目的変数, Y_2 を説明変数とするとき, 説明変数に欠損が起こるときの回帰式による補完法は, 確率化が必要

4.3 統計モデルを指定しない抽出法

設定：有限母集団の有限母集団からの標本選択と部分測定

N ：母集団の個体数

n ：標本数

r ：標本の内、実際に測定が行われる個体数

y_1, \dots, y_r ：観測値

*. n, r は定数とする.

Hot deck 推定量

$$\bar{y}_{HD} = \frac{1}{n} \{ r \bar{y}_R + (n - r) \bar{y}_{NR}^* \}, \quad \bar{y}_{NR}^* = \sum_{i=1}^r \frac{H_i y_i}{n - r},$$

$$H_i : y_i \text{ が代入される回数}, \quad \sum_{i=1}^r H_i = n - r$$

*. \bar{y}_{HD} の性質は (H_1, \dots, H_r) の分布に依存する

$$E(\bar{y}_{HD}) = E[E(\bar{y}_{HD}|Y_{obs})],$$

$$\text{Var}(\bar{y}_{HD}) = \text{Var}[E(\bar{y}_{HD}|Y_{obs})] + E[\text{Var}(\bar{y}_{HD}|Y_{obs})]$$

例 4.6 単純無作為抽出による Hot deck 補完法

$(H_1, \dots, H_r) \sim \text{Mt}_r(n - r, (1/r, \dots, 1/r))$ 多項分布

$$E(H_i|Y_{obs}) = \frac{n - r}{r}, \quad \text{Var}(H_i|Y_{obs}) = (n - r) \left(1 - \frac{1}{r}\right) \frac{1}{r},$$

$$\text{Cov}(H_i, H_j|Y_{obs}) = -\frac{n - r}{r^2}$$

$$\Rightarrow E(\bar{y}_{HD}|Y_{obs}) = \bar{y}_R,$$

$$\text{Var}(\bar{y}_{HD}|Y_{obs}) = \left(1 - \frac{1}{r}\right) \left(1 - \frac{r}{n}\right) \frac{s_{yR}^2}{n}$$

MCAR (等確率で r 個が観測される) の場合

$$E(\bar{y}_{HD}) = \bar{y} : \text{母平均},$$

$$\text{Var}(\bar{y}_{HD}) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 + \left(1 - \frac{1}{r}\right) \left(1 - \frac{r}{n}\right) \frac{S_y^2}{n}, \quad S_y^2 : \text{母分散}$$

例 4.7 補正クラスごとの Hot deck 補完法

仮定:母集団を J 個のクラスに分割.

各クラス内では標本の観測 (欠測) 確率が等しい

各クラスごとに, 欠測値を Hot deck 法で補完

例 4.8 最近隣 Hot deck 補完法

共変量 $\mathbf{x}_i = (x_{i1}, \dots, x_{iK})^T$ による個体間の適当な距離 $d(i, j)$ を定義
欠測値 y_i に代入する値を $\{y_j; d(i, j) \leq \delta\}$ から抽出する.

\mathbf{x} によって補正クラスを定義する場合

$$d(i, j) = \begin{cases} 0 & : i, j \text{ が同じ補正クラス内} \\ 1 & : i, j \text{ が異なる補正クラス} \end{cases}, \quad \delta = 0$$

その他の (擬) 距離

Maximum deviation: $d(i, j) = \max_k |x_{ik} - x_{jk}|$

Mahalanobis: $d(i, j) = (\mathbf{x}_i - \mathbf{x}_j)^T S_{xx}^{-1} (\mathbf{x}_i - \mathbf{x}_j)$

Predictive Mean: $d(i, j) = \{\hat{y}(\mathbf{x}_i) - \hat{y}(\mathbf{x}_j)\}^2$

例 4.9 脱落データの補完

$\mathbf{y}_i = (y_{i1}, \dots, y_{it}, \dots, y_{iK})$: 経時データ

M_i : 欠損の有無を表す変数

$M_i = 0 \Rightarrow$ 完全データ

$M_i = k \Rightarrow y_{ik}, \dots, y_{iK}$ が欠損 ($k = 1, \dots, K$)

最後の観測値による補完

$$M_i = k \Rightarrow \hat{y}_{it} = y_{i,k-1} \quad (t = k \dots, K)$$

(測定時刻による影響がないことが前提)

個体差と時間効果を取り入れた補完

$$\tilde{y}_{it} = \bar{y}_{\text{obs},i} - \bar{y}_{\text{obs},+}^{(\text{cc})} + \bar{y}_{+t}^{(\text{cc})},$$

$$\tilde{y}_{\text{obs},i} = \frac{1}{k-1} \sum_{t=1}^{k-1} y_{it} : \text{個体 } i \text{ の } k-1 \text{ 時点までの平均}$$

$$\bar{y}_{\text{obs},+}^{(\text{cc})} = \frac{1}{r} \sum_{l=1}^r \tilde{y}_{\text{obs},l} : \text{欠損のない個体の } k-1 \text{ 時点までの平均の個体間平均}$$

$$\bar{y}_{+t}^{(\text{cc})} = \frac{1}{r} \sum_{l=1}^r y_{lt} \text{ 欠損のない個体の } t \text{ 時点の観測値の個体間平均}$$

これに、無作為抽出した残差 $y_{lt} - \tilde{y}_{lt}$ を加えて

$$\hat{y}_{it} = \tilde{y}_{it} + (y_{lt} - \tilde{y}_{lt}) = y_{lt} + (\bar{y}_{\text{obs},i} - \bar{y}_{\text{obs},l}),$$

$$\text{ただし } \tilde{y}_{lt} = \bar{y}_{\text{obs},l} - \bar{y}_{\text{obs},+}^{(\text{cc})} + \bar{y}_{+t}^{(\text{cc})}$$

⇒ hot-deck による抽出値を個体効果で補正したもの

5 補完データによる推定量の変動

5.1 Ultimate cluster(UC, 最終集落単位)による分散の推定

補題

$\hat{\theta}_1, \dots, \hat{\theta}_k : \mu$ の不偏推定量,

$\text{Cov}(\hat{\theta}_i, \hat{\theta}_j) = 0 \ (i \neq j)$

$$\hat{\theta} = \frac{1}{k} \sum_{j=1}^k \hat{\theta}_j, \quad \hat{v}(\hat{\theta}) = \frac{1}{k(k-1)} \sum_{j=1}^k (\hat{\theta}_j - \hat{\theta})^2$$
$$\Rightarrow \quad \text{E}(\hat{\theta}) = \mu, \quad \text{E}[\hat{v}(\hat{\theta})] = \text{Var}(\hat{\theta})$$

例 5.1 有限母集団： K 個の UC から成る.

t_j : 第 j UC の変数 Y の総和

$T = \sum_{j=1}^K t_j$ を推定したい.

標本抽出法：復元抽出により, k 個の UC を抽出

第 j UC が選ばれたとき, t_j の不偏推定量 \hat{t}_j を構成

Horvitz-Tohmpson 推定量

$$\hat{t}_{HT} = \sum_{i=1}^k \frac{\hat{t}_{j_i}}{\pi_{j_i}}, \quad \pi_j = \frac{k}{K} : \text{第 } j \text{ UC が選択される確率}$$

$$\hat{\theta}_i = k\hat{t}_{j_i}/\pi_{j_i} \Rightarrow E(\hat{\theta}_i) = E[E(k\hat{t}_{j_i}/\pi_{j_i}|j_i)] = \sum_{j=1}^K \frac{1}{K} (kE(\hat{t}_j)/\pi_j) = \sum_{j=1}^K t_j$$

復元抽出なので, $\hat{\theta}_1, \dots, \hat{\theta}_k$ は独立

補題より $\text{Var}(\hat{t}_{HT})$ の不偏推定量

$$\hat{v}(\hat{t}_{HT}) = \sum_{i=1}^k \frac{(k\hat{t}_{j_i}/\pi_{j_i} - \hat{t}_{HT})^2}{k(k-1)}$$

を得る

選ばれた UC からの標本に欠損がある場合にも,

条件 1 加重法や補完法により t_j の不偏推定量 \hat{t}_j が得られ,

条件 2 加重や補完が各 UC 内で独立に行われるならば

\hat{t}_{HT} は, T の不偏推定量となり, その分散の不偏推定量として $\hat{v}(\hat{t}_{HT})$ を用いることができる.

例 5.2 層別標本抽出の場合

有限母集団が H 個の層から成り, 各層 (h) は K_h 個の UC から成るとする.
層 h から復元抽出で k_h 個の UC を選び, \hat{t}_{hj} を構成

$$\hat{t} = \sum_{h=1}^H \sum_{i=1}^{k_h} \frac{\hat{t}_{hji}}{\pi_{hji}} = \sum_{h=1}^H \hat{t}_h,$$

$$\hat{v}(\hat{t}) = \sum_{h=1}^H \sum_{i=1}^{k_h} \frac{(k_h \hat{t}_{hji} / \pi_{hji} - \hat{t}_h)^2}{k_h(k_h - 1)}$$

5.2 リサンプリングによる分散の推定

ブートストラップ法

例 5.3 完全データに対するブートストラップ

$S = \{i; i = 1, \dots, n\}$: 無作為抽出された標本

$\hat{\theta}$: 標本 S の観測値による母数 θ の一致推定量

$S^{(b)}$: S からの復元抽出による大きさ n の標本 ($b = 1, \dots, B$)

$\hat{\theta}^{(b)}$: S の代わりに $S^{(b)}$ を用いて計算した $\hat{\theta}$

$\hat{\theta}$ のバイアス補正

$$\hat{\theta}_{boot} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{(b)}$$

$\hat{\theta}$ あるいは $\hat{\theta}_{boot}$ の分散の推定量

$$\hat{V}_{boot} = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{(b)} - \hat{\theta}_{boot})^2$$

正規近似による信頼区間

$$I_{norm}(\theta) = \hat{\theta} \pm z_{1-\alpha/2} \sqrt{\hat{V}_{boot}}$$

ブートストラップ分布による信頼区間

$$I_{emp}(\theta) = (\hat{\theta}^{(b,l)}, \hat{\theta}^{(b,u)}),$$

$\hat{\theta}^{(b,l)}$: $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(b)}$ の $\alpha/2$ 分位点 ,

$\hat{\theta}^{(b,u)}$: $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(b)}$ の $1 - \alpha/2$ 分位点 ,

例 5.4 補完された欠損データに対するブートストラップ法

$S = \{i; i = 1, \dots, n\}$: 無作為抽出された標本 (一部欠損あり)

Imp : 補完法, $\hat{S} = \text{Imp}(S)$: 補完されたデータ

$\hat{\theta} = \hat{\theta}(\hat{S})$: \hat{S} を用いた θ の一致推定量

(a) $S^{(b)}$: S からのリサンプリング

(b) $\hat{S}^{(b)} = \text{Imp}(S^{(b)})$

(c) $\hat{\theta}^{(b)} = \hat{\theta}(\hat{S}^{(b)})$

例 5.3 と同様に, $\hat{\theta}_{boot}, \hat{V}_{boot}$ 等を計算

注 1. $\hat{S}^{(b)}$ を \hat{S} からリサンプリングしてはいけない

注 2. $\hat{\theta}(\hat{S})$ の一致性は, Imp にも依存

ジャックナイフ法

例 5.5 完全データに対するジャックナイフ法

$S = \{i; i = 1, \dots, n\}$: 無作為抽出された標本

$S^{(\setminus j)} = S \setminus \{j\}$ ($j = 1, \dots, n$)

$\hat{\theta}$: 標本 S の観測値による母数 θ の一致推定量

$\hat{\theta}^{(\setminus j)}$: 標本 $S^{(\setminus j)}$ の観測値による母数 θ の一致推定量

$$\tilde{\theta}_j = n\hat{\theta} - (n-1)\hat{\theta}^{(\setminus j)} \text{ Jackknife pseudovalue,}$$

$\hat{\theta}$ のバイアス補正

$$\hat{\theta}_{jack} = \frac{1}{n} \sum_{j=1}^n \tilde{\theta}_j,$$

$\hat{\theta}$ あるいは $\hat{\theta}_{jack}$ の分散の推定量

$$\hat{V}_{jack} = \frac{1}{n(n-1)} \sum_{j=1}^n (\tilde{\theta}_j - \hat{\theta}_{jack})^2$$

正規近似による信頼区間

$$I_{\text{norm}}(\theta) = \hat{\theta} \pm z_{1-\alpha/2} \sqrt{\hat{V}_{\text{jack}}}$$

例 5.6 補完された欠損データに対するジャックナイフ法

$S = \{i; i = 1, \dots, n\}$: 無作為抽出された標本 (一部欠損あり)

Imp : 補完法, $\hat{S} = \text{Imp}(S)$: 補完されたデータ

$\hat{\theta} = \hat{\theta}(\hat{S})$: \hat{S} を用いた θ の一致推定量

(a) $S^{(\setminus j)}$: 個体 j を削除

(b) $\hat{S}^{(\setminus j)} = \text{Imp}(S^{(\setminus j)})$

(c) $\hat{\theta}^{(\setminus j)} = \hat{\theta}(\hat{S}^{(\setminus j)})$

例 5.5 と同様に, $\tilde{\theta}_j, \hat{\theta}_{\text{jack}}, \hat{V}_{\text{jack}}$ 等を計算

例 5.7 層別標本抽出に対するジャックナイフ法

表記を簡単にするため、個体番号を付け替えて j_1, \dots, j_{k_h} を $1, 2, \dots, k_h$ と表す

$$\hat{t} = \hat{t}(S) = \sum_{h=1}^H \sum_{i=1}^{k_h} \frac{\hat{t}_{hi}(S)}{\pi_{hi}} = \sum_{h=1}^H \hat{t}_h(S),$$

UC 内の個体の観測値は相関ある場合が多い

⇒ 個体を抜き取るのではなく、UC 単位で抜き取る

$$\tilde{t}^{(\setminus hj)} = \sum_{h' \neq h}^H \hat{t}_{h'} + \hat{t}_h^{(\setminus hj)},$$
$$\hat{t}_{h'} = \sum_{j=1}^{k_{h'}} \frac{\hat{t}_{h'j'}}{\pi_{h'j'}}, \quad \hat{t}_h^{(\setminus hj)} = \sum_{j' \neq j}^{k_h} \left(\frac{k_h}{k_h - 1} \right) \frac{\hat{t}_{hj'}}{\pi_{hj'}}$$

$\theta = \theta(T)$ の推定量 $\hat{\theta} = \theta(\hat{t})$ の分散のジャックナイフ推定量

$$\hat{V}_{jack} = \sum_{h=1}^H \frac{k_h - 1}{k_h} \sum_{j=1}^{k_h} (\hat{\theta}^{(\setminus hj)} - \hat{\theta})^2, \quad \hat{\theta}^{(\setminus hj)} = \theta(\hat{t}^{(\setminus hj)})$$

欠損データの場合

$\hat{S} = \text{Imp}(S)$: 欠測値データを補完

$$\hat{t} = \hat{t}(\hat{S}) = \sum_{h=1}^H \sum_{i=1}^{k_h} \frac{\hat{t}_{hi}(\hat{S})}{\pi_{hi}} = \sum_{h=1}^H \hat{t}_h(\hat{S}),$$

$h = 1, \dots, H; i = 1, \dots, k_h$ に対して

(a) $S^{(\setminus hj)}$: 第 h 層の第 j UC を除外した標本

(b) $\hat{S}^{(\setminus hj)} = \text{Imp}(S^{(\setminus hj)})$

(c) $\tilde{t}^{(\setminus hj)} = \sum_{h' \neq h}^H \hat{t}_{h'}(\hat{S}^{(\setminus hj)}) + \hat{t}_h^{(\setminus hj)}(\hat{S}^{(\setminus hj)}),$

$$\hat{t}_{h'}(\hat{S}^{(\setminus hj)}) = \sum_{j=1}^{k_{h'}} \frac{\hat{t}_{h'j'}(\hat{S}^{(\setminus hj)})}{\pi_{h'j'}}, \quad \hat{t}_h^{(\setminus hj)}(\hat{S}^{(\setminus hj)}) = \sum_{j' \neq j}^{k_h} \left(\frac{k_h}{k_h - 1} \right) \frac{\hat{t}_{hj'}(\hat{S}^{(\setminus hj)})}{\pi_{hj'}}$$

(d) $\hat{V}_{jack} = \sum_{h=1}^H \frac{k_h - 1}{k_h} \sum_{j=1}^{k_h} (\hat{\theta}^{(\setminus hj)} - \hat{\theta})^2, \quad \hat{\theta}^{(\setminus hj)} = \theta(\hat{t}^{(\setminus hj)})$

5.3 多重補完

Y_{mis} : 欠測値の全体 ,

$\hat{Y}_{mis}^{(1)}, \dots, \hat{Y}_{mis}^{(D)}$: 補完候補

$$Y^{(d)} = (Y_{res}, \hat{Y}_{mis}^{(d)}), \quad (d = 1, \dots, D)$$

補完候補が無作為抽出を利用した場合

⇒ 分析結果に欠測値の変動を取り込むことができる

補完候補が異なるモデルに基づく場合

⇒ 分析結果へのモデルの違いによる影響を見ることができる

$\hat{\theta}_d : Y^{(d)}$ を用いた推定値

$W_d : \hat{\theta}_d$ の分散の推定値 $d = 1, \dots, D$

$$\bar{\theta}_D = \frac{1}{D} \sum_{d=1}^D \hat{\theta}_d,$$

$$\bar{W}_D = \frac{1}{D} \sum_{d=1}^D W_d, \quad B_D = \frac{1}{D-1} \sum_{d=1}^D (\hat{\theta}_d - \bar{\theta}_D)^2,$$

$$T_D = \bar{W}_D + \frac{D+1}{D} B_D$$

\Rightarrow

$$(\theta - \bar{\theta}_D) T_D^{-1/2} \sim t_\nu, \quad \nu = (D-1) \left(1 + \frac{1}{D+1} \frac{\bar{W}_D}{B_D}\right)^2$$

例 5.8 層別標本抽出に対する多重補完

設定：有限母集団が H 個の層から成る.

N_h : 第 h 層の個体数, $N = \sum_{h=1}^H N_h$,

n_h : 第 h 層から抽出される標本数, $n = \sum_{h=1}^H n_h$,

\bar{Y} : 母平均

欠損がない場合

$$\bar{y}_{ST} = \sum_{h=1}^H P_h \bar{y}_h : \text{母平均の推定量}$$

$$\bar{y}_h : \text{第 } h \text{ 層からの標本平均}, \quad P_h = \frac{N_h}{N},$$

$$\widehat{\text{Var}}(\bar{y}_{st}) = \sum_{h=1}^H P_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h}, \quad s_h^2 : \text{第 } h \text{ 層からの標本分散}$$

欠損がある場合

r_h 第 h 層からの観測数, ($n_h - r_h$ が欠測),

$$\hat{Y}_{MI} = \frac{1}{D} \sum_{d=1}^D \left(\sum_{h=1}^H P_h \bar{y}_{h(d)} \right),$$

$$T_D = \frac{1}{D} \sum_{d=1}^D \sum_{h=1}^H P_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{s_{h(d)}^2}{n_h} \\ + \frac{D+1}{D} \frac{1}{D-1} \left(\sum_{h=1}^H P_h \bar{Y}_{h(d)} - \hat{Y}_{MI} \right)^2$$

層ごとに MCAR である場合
最適な母平均の推定量は

$$\hat{Y}_{ST} = \sum_{h=1}^H P_h \bar{y}_{hR} : \text{母平均の推定量}$$

\bar{y}_h : 第 h 層からの観測値の標本平均, $P_h = \frac{N_h}{N}$,

$$\widehat{\text{Var}}(\hat{Y}_{st}) = \sum_{h=1}^H P_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{s_{hR}^2}{n_h},$$

s_{hR}^2 : 第 h 層からの観測値の標本分散

層ごとの Hot deck 法では、適当な条件下で

$$\hat{Y}_{MI} \rightarrow \hat{Y}_{ST} \quad D \rightarrow \infty$$

ただし, T_D は $D \rightarrow \infty$ でも $\widehat{\text{Var}}(\hat{Y}_{st})$ より小さな値となる.

(Herog and Rubin, 1983)

6 尤度に基づく推測

6.1 完全データに対する最尤法

Y : 観測変量 (確率変数, 確率ベクトル, 確率行列)

$f(Y|\theta)$: Y の確率密度関数, $\theta \in \Omega_\theta$

Ω_θ : 母数空間. 特に制限しない限り, $f(Y|\theta)$ が確率密度関数となるような θ の全体とする.

定義 6.1. $L(\theta|Y) = f(Y|\theta)$ (あるいは, $f(Y|\theta)$ の定数倍) を尤度関数と呼ぶ.

定義 6.2. $\ell(\theta|Y) = \log L(\theta|Y)$: 対数尤度関数.

例 6.1 正規分布

$$Y = (y_1, \dots, y_n)^T, \theta = \mu, \sigma^2$$

$$f(Y|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2}\right),$$

$$\ell(\mu, \sigma^2|Y) = \log f(Y|\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2}$$

(定数項を無視)

例 6.2 指数分布 $\theta > 0$

$$f(Y|\theta) = \theta^{-n} \exp\left(-\sum_{i=1}^n \frac{y_i}{\theta}\right),$$

$$\ell(\theta|Y) = -n \log \theta - \sum_{i=1}^n \frac{y_i}{\theta}.$$

例 6.3 多項分布

$$Y = (y_1, \dots, y_n)^T,$$

y_i の取り得る値 : $c = 1, 2, \dots, C$

$$n_c = \#\{i; y_i = c\}, \quad (c = 1, \dots, C)$$

$$\theta = (\pi_1, \dots, \pi_{C-1}), \quad (\pi_C = 1 - \sum_{c=1}^{C-1} \pi_c)$$

$$f(Y|\theta) = \frac{n!}{n_1! \cdots n_C!} \prod_{c=1}^C \pi_c^{n_c},$$

$$\ell(\theta|Y) = \sum_{c=1}^C n_c \log \pi_c$$

例 6.4 多変量正規分布 $Y = (y_{ij}), i = 1, \dots, n; j = 1, \dots, K$
 $\theta = (\boldsymbol{\mu}, \Sigma), \boldsymbol{\mu} = (\mu_1 \dots, \mu_K), \Sigma = (\sigma_{ij})$

$$f(Y|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-nK/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\mu})^T\right),$$

$$\ell(\boldsymbol{\mu}, \Sigma|Y) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\mu})^T$$

定義 6.3. 最尤推定量

$$\hat{\theta} = \operatorname{argmax} \ell(\theta|Y)$$

$\ell(\theta|Y)$ が有界, 微分可能, Ω_θ が開集合ならば

$$D_\ell(\theta)|_{\theta=\hat{\theta}} = 0, \quad D_\ell(\theta) = \frac{\partial \ell(\theta|Y)}{\partial \theta}$$

$D_\ell(\theta)$: スコア関数

$D_\ell(\theta) = 0$: 尤度方程式

カルバックライブラー擬距離 $Y \sim g(Y)$

$$\text{KL}(g, f) = \mathbb{E} \left[\log \frac{g(Y)}{f(Y)} \right]$$

を f の g からのカルバックライブラー擬距離と呼ぶ.

$$\text{KL}(g, f) \geq 0,$$

$$\text{KL}(g, f) = 0 \Leftrightarrow \mathbb{P}(f(Y) = g(Y)) = 1$$

$$Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} f(Y|\theta_0)$$

$$\frac{1}{n} \sum_{i=1}^n \log \frac{f(Y_i|\theta)}{f(Y_i|\theta_0)} \rightarrow -\text{KL}(f(*|\theta_0), f(*|\theta)) \quad n \rightarrow \infty \text{ (大数の法則)},$$

$$\text{argmax}_{\theta} \ell(\theta|Y_1, \dots, Y_n) = \text{argmax}_{\theta} \frac{1}{n} \sum_{i=1}^n \log \frac{f(Y_i|\theta)}{f(Y_i|\theta_0)},$$

$$\text{argmax}_{\theta} \{-\text{KL}(f(*|\theta_0), f(*|\theta))\} = \theta_0$$

適当な条件の下で

$$\hat{\theta} \xrightarrow{as} \theta_0, \quad n \rightarrow \infty$$

例 6.5 指数分布 (例 6.2)

$$D(\theta|Y) = -\frac{n}{\theta} + \sum_{i=1}^n \frac{y_i}{\theta^2} = 0, \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i$$

例 6.6 多項分布 (例 6.3)

$$\frac{\partial \ell(\theta|Y)}{\partial \pi_c} = \frac{n_c}{\pi_c} - \frac{n_C}{\pi_C} = 0 \quad (c = 1, \dots, C-1),$$
$$\hat{\pi}_c \propto n_c \quad \Rightarrow \quad \hat{\pi}_c = \frac{n_c}{n}$$

例 6.7 正規分布 (例 6.1)

$$\ell(\mu, \sigma^2|Y) = -\frac{n}{2} \log \sigma^2 - \frac{n(\bar{y} - \mu)^2}{2} - \frac{(n-1)s^2}{2\sigma^2},$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

$$\hat{\mu} = \bar{y}, \quad \hat{\sigma}^2 = \frac{n-1}{n} s^2$$

例 6.8 多変量正規分布 (例 6.4)

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n} S, \quad S = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^{\text{T}}$$

性質 6.1 $g(\theta)$ の最尤推定量は $g(\hat{\theta})$

例 6.9 単回帰と条件付き平均・分散

$$(y_{i1}, y_{i2}), \quad i = 1, \dots, n \stackrel{i.i.d.}{\sim} N((\mu_1, \mu_2), \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

$$\hat{\mu}_j = \bar{y}_j \quad (j = 1, 2),$$

$$\hat{\sigma}_{jk} = \frac{s_{jk}}{n} = \frac{1}{n} \sum_{i=1}^n (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k)$$

$$\begin{aligned}
E[y_{i2}|y_{i1}] &= \mu_2 + \beta_{21.1}(y_{i1} - \mu_1), & \text{Var}[y_{i2}|y_{i1}] &= \sigma_{22.1}, \\
\beta_{21.1} &= \sigma_{12}/\sigma_{11}, & \sigma_{22.1} &= \sigma_{22} - \sigma_{12}^2/\sigma_{11}, \\
\hat{\beta}_{21.1} &= s_{12}/s_{11}, & \hat{\sigma}_{22.1} &= \frac{1}{n}(s_{22} - s_{12}^2/s_{11})
\end{aligned}$$

単回帰モデルと最小 2 乗法

$$y_{i2} = \beta_{21.0} + \beta_{21.1}y_{i1} + \varepsilon_i, \quad \varepsilon_i, i = 1, \dots, n \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$$

$$\beta_{21.1}^{(ls)} = s_{12}/s_{11}, \quad \beta_{21.0}^{(ls)} = \bar{y}_2 - \beta_{21.1}^{(ls)}\bar{y}_1$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\text{RSS} \quad (\Leftrightarrow \quad \hat{\sigma}_{22.1} = \frac{1}{n}\text{RSS})$$

$$\text{RSS} = \sum_{i=1}^n \{y_{i1} - \bar{y}_2 - \beta_{21.1}(y_{i1} - \bar{y}_1)\}^2 = s_{22} - s_{12}^2/s_{11}$$

例 6.10 重回帰と条件付き平均・分散

$$(y_i, x_{i1}, \dots, x_{ip}), i = 1, \dots, n \stackrel{i.i.d.}{\sim} N \left[(\mu_y, \mu_{x1}, \dots, \mu_{xp}), \begin{pmatrix} \sigma_{yy} & \boldsymbol{\sigma}_{yx} \\ \boldsymbol{\sigma}_{xy}^T & \Sigma_{xx} \end{pmatrix} \right]$$

$$\hat{\sigma}_{yy} = \frac{s_{yy}}{n} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2,$$

$$\hat{\boldsymbol{\sigma}}_{yx} = \frac{\mathbf{s}_{yx}}{n} = \frac{1}{n} \left(\sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1), \dots, \sum_{i=1}^n (y_i - \bar{y})(x_{ip} - \bar{x}_p) \right),$$

$$\hat{\Sigma}_{xx} = \frac{1}{n} S_{xx} = \frac{1}{n} \left(\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \right)_{j,k=1,\dots,p}$$

$$\mu_{y|x_1, \dots, x_p} \equiv E[y_{i1} | x_{i1}, \dots, x_{ip}], \quad \sigma_{y|x_1, \dots, x_p} \equiv \text{Var}[y_{i1} | x_{i1}, \dots, x_{ip}]$$

$$\Rightarrow \hat{\mu}_{y|x_1, \dots, x_p} = \hat{y}_i, \quad \hat{\sigma}[y_{i1} | x_{i1}, \dots, x_{ip}] = \frac{1}{n} \left\{ s_{yy} - \mathbf{s}_{yx} S_{xx}^{-1} \mathbf{s}_{xy}^T \right\}$$

$$\hat{y}_i = \bar{y} + (x_{i1} - \bar{x}_1, \dots, x_{ip} - \bar{x}_p) \hat{\boldsymbol{\beta}}, \quad \hat{\boldsymbol{\beta}} = S_{xx}^{-1} \mathbf{s}_{xy}$$

重回帰モデルと最小2乗法

$$y_i \sim N(\mu_i, \sigma^2), \quad i = 1, \dots, n \text{ 独立} \quad \mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

$$Y = (y_1, \dots, y_n), \quad \theta = (\beta_0, \dots, \beta_p, \sigma^2)$$

$$\Rightarrow \ell(\theta|Y) = -\frac{n}{2} \log \sigma^2 - \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} + \dots + \beta_p x_{ip})^2 / (2\sigma^2)$$

$$\hat{\beta}_0 = \bar{y} - (\bar{x}_1, \dots, \bar{x}_p) \hat{\beta}, \quad \hat{\beta} = S_{xx}^{-1} \mathbf{s}_{xy},$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \text{RSS}, \quad \text{RSS} = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip})^2 \\ &= \left\{ s_{yy} - \mathbf{s}_{yx} S_{xx}^{-1} \mathbf{s}_{xy}^T \right\} \end{aligned}$$

$$\sigma^2 \text{ の不偏推定量は } \frac{1}{n-p-1} \text{RSS}$$

一般化最小2乘法

$$y_i \sim N(\mu_i, w_i^{-1}\sigma^2), \quad i = 1, \dots, n \text{ 独立}$$

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad w_1, \dots, w_n : \text{ 既知}$$

$$Y = (y_1 \dots, y_n), \quad \theta = (\beta_0, \dots, \beta_p, \sigma^2)$$

$$\Rightarrow \ell(\theta|Y) = -\frac{n}{2} \log \sigma^2 - \sum_{i=1}^n w_i (y_i - \beta_0 - \beta_1 x_{i1} + \dots + \beta_p x_{ip})^2 / (2\sigma^2)$$

$$\hat{\beta}_0 = \bar{y}^{(w)} - \sum_{j=1}^p \hat{\beta}_j \bar{x}_j^{(w)}, \quad \bar{y}^{(w)} = \sum_{i=1}^n w_i y_i / \left(\sum_{i=1}^n w_i \right),$$

$$\bar{x}_j^{(w)} = \sum_{i=1}^n w_i x_{ij} / \left(\sum_{i=1}^n w_i \right), \quad j = 1, \dots, p$$

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^\top = (X_*^\top W X_*)^{-1} (X_*^\top W Y_*), \quad W = \text{diag}(w_1, \dots, w_n),$$

$$X_* = (x_{ij} - \bar{x}_j^{(w)}), \quad Y_* = (y_1 - \bar{y}^{(w)}, \dots, y_n - \bar{y}^{(w)})^\top,$$

$$\hat{\sigma}^2 = (Y_* - X_* \hat{\boldsymbol{\beta}})^\top W (Y_* - X_* \hat{\boldsymbol{\beta}}) / n$$

例 6.11 一般化線形モデル $(y_i, x_{i1}, \dots, x_{ip}), i = 1, \dots, n$: 独立

$$f(y_i | \mathbf{x}_i, \boldsymbol{\beta}, \phi) = \exp \left[\frac{1}{\phi} \{ y_i \delta(\mathbf{x}_i, \boldsymbol{\beta}) - b(\delta(\mathbf{x}_i, \boldsymbol{\beta})) \} + c(y_i, \phi) \right],$$

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T, \quad \boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T,$$

$\delta(\cdot, \cdot), b(\cdot), c(\cdot)$: 既知関数, $\phi > 0$: 尺度母数 (scale parameter)

$$\mu_i = \mathbb{E}[y_i | \mathbf{x}_i, \boldsymbol{\beta}, \phi] = g^{-1} \left(\beta_0 + \sum_{j=1}^p \beta_j x_{ij} \right), \quad g : \text{リンク関数}$$

$$\delta_i = \delta(\mathbf{x}_i, \boldsymbol{\beta})$$

$$\Rightarrow \mu_i = b'(\delta_i), \quad \sigma_i^2 = \text{Var}(y_i | \delta_i, \phi) = \phi b''(\delta_i)$$

$$g_c : \text{標準リンク (canonical link)} \quad \Leftrightarrow \quad g_c(\mu_i) = \delta(\mathbf{x}_i, \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$$

$$(\Rightarrow g_c^{-1}(\delta_i) = b'(\delta_i))$$

Normal linear regression : y_i : 正規分布

$$g_c(\mu_i) = \mu_i, b(\delta) = \delta^2/2, \phi = \sigma^2$$

Poisson regression : y_i : ポアソン分布

$$g_c(\mu_i) = \log \mu_i, b(\delta) = \exp(\delta), \phi = 1$$

Logistic regression : $y_i \in \{0, 1\}$: 2項分布

$$g_c(\mu_i) = \text{logit}(\mu_i) = \log \frac{\mu_i}{1 - \mu_i}, b(\delta) = \log\{1 + \exp(\delta)\}, \phi = 1$$

対数尤度関数

$$\ell(\theta|Y) = \sum_{i=1}^n \left[\frac{1}{\phi} \{y_i \delta(\mathbf{x}_i, \boldsymbol{\beta}) - b(\delta(\mathbf{x}_i, \boldsymbol{\beta}))\} + c(y_i, \phi) \right]$$

ベイズ推定 母数 θ も確率変数として扱う

$\theta \sim p(\theta)$: 事前分布 (確率密度関数)

$f(Y|\theta)$: θ が与えられた条件付き密度関数

$$p(\theta|Y) = \frac{p(\theta)f(Y|\theta)}{p(Y)} : \text{事後分布}$$

$$p(Y) = \int p(\theta)f(Y|\theta)d\theta$$

推定問題

$\text{loss}(\hat{\theta}, \theta)$: 損失関数

リスク

$$\begin{aligned} E[\text{loss}(\hat{\theta}, \theta)] &= \iint \text{loss}(\hat{\theta}, \theta)p(Y|\theta)p(\theta)dYd\theta \\ &= \int \left\{ \int \text{loss}(\hat{\theta}, \theta)p(\theta|Y)d\theta \right\} p(Y)dY \end{aligned}$$

最小リスク推定量

$$\hat{\theta}_B(Y) = \operatorname{argmin}_{\hat{\theta}} \int \operatorname{loss}(\hat{\theta}, \theta) p(\theta|Y) d\theta$$

はリスクを最小とする.

$$\operatorname{loss}(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 \quad \Rightarrow \quad \hat{\theta}_B = E[\theta|Y] : \text{事後平均}$$

$$\operatorname{loss}(\hat{\theta}, \theta) = |\hat{\theta} - \theta| \quad \Rightarrow \quad \hat{\theta}_B = F_{\theta|Y}^{-1}\left(\frac{1}{2}\right) : \text{事後分布の中央値}$$

$$\operatorname{loss}(\hat{\theta}, \theta) = \begin{cases} 1 & |\hat{\theta} - \theta| > \delta/2 \\ 0 & |\hat{\theta} - \theta| \leq \delta/2 \end{cases}$$

$$\Rightarrow \quad \hat{\theta}_B^{(\delta)} \rightarrow \operatorname{argmax}_{\theta} p(\theta|Y) \quad (\delta \rightarrow 0) : \text{事後分布の最頻値 (モード)}$$

*. $p(\theta) = \text{定数}$ のとき, 事後分布の最頻値は最尤推定量と一致

大標本近似

近似 6.1.

$$(\theta - \hat{\theta}) \approx N(0, C), \quad C = \text{Cov}[(\theta - \hat{\theta})]$$

Bayesian の解釈

θ : 確率変数, $\hat{\theta}$: 事後分布の最頻値

「 θ の事後分布は, 平均 $\hat{\theta}$, 共分散行列 C の正規分布で近似できる。」

$$\ell(\theta|Y) = \ell(\hat{\theta}|Y) + (\theta - \hat{\theta})^T D_\ell(\hat{\theta}|Y) - \frac{1}{2}(\theta - \hat{\theta})^T I(\hat{\theta}|Y)(\theta - \hat{\theta}) + r(\theta|Y),$$

$$D_\ell(\theta|Y) : \text{スコア関数}, \quad I(\theta|Y) = -\frac{\partial^2 \ell(\theta|Y)}{\partial \theta \partial \theta^T} : \text{observed information}$$

$D_\ell(\hat{\theta}|Y) = 0$ なので, $r(\theta|Y)$ が無視でき, $p(\theta)$ が十分フラットならば

$$f(\theta|Y) \propto \exp \left[-\frac{1}{2}(\theta - \hat{\theta})^T I(\hat{\theta}|Y)(\theta - \hat{\theta}) \right], \quad C = I(\hat{\theta}|Y)^{-1}$$

性質 6.2. $\text{Cov}(\theta - \hat{\theta}) \approx C$ ならば

$$\text{Cov}[g(\theta) - g(\hat{\theta})] \approx D_g(\hat{\theta})CD_g(\hat{\theta})^T, \quad D_g(\theta) = \frac{\partial g(\theta)}{\partial \theta}$$

近似 6.2.

$$g(\theta) - g(\hat{\theta}) \approx N[0, D_g(\hat{\theta})CD_g(\hat{\theta})^T]$$

frequentist の解釈

$$0 = D_\ell(\hat{\theta}|Y) = D_\ell(\theta|Y) - I(\theta|Y)(\hat{\theta} - \theta) + r(\hat{\theta}|Y),$$

$$D_\ell(\theta|Y) \approx I(\theta|Y)(\hat{\theta} - \theta),$$

$$D_\ell(\theta|Y) \approx N(0, J(\theta)),$$

$$J(\theta) = \mathbb{E}[I(\theta|Y)|\theta] = \int I(\theta|y)f(y|\theta)dy : \text{expected information matrix},$$

$$I(\hat{\theta}) \approx J(\hat{\theta}) \approx J(\theta)$$

*. 近似 6.1, 6.2 は, Y の真の密度関数が $f(Y|\theta_0)$ ($\exists\theta_0$) である場合に成り立つ.

$Y \sim f^*(Y) \neq f(Y|\theta)$ ($\forall\theta$) の場合

近似 6.3.

$$(\hat{\theta}|f^*) \approx N(\theta^*, C^*),$$

$$C^* = J^{-1}(\theta)K(\theta)J^{-1}(\theta), \quad K(\theta) = E[D_\ell(\theta)D_\ell(\theta)^T]$$

C^* の一致推定量

$$\hat{C}^* = I^{-1}(\hat{\theta})\hat{K}(\hat{\theta})I^{-1}(\hat{\theta}), \quad \hat{K}(\hat{\theta}) = D_\ell(\hat{\theta})D_\ell(\hat{\theta})^T$$

例 6.12 指数分布 (例 6.2)

$$I(\theta|Y) = -\frac{n}{\theta^2} + 2 \sum \frac{y_i}{\theta^3}, \quad J(\theta) = \mathbb{E}[I(\theta|Y)] = \frac{n}{\theta^2},$$

$$I(\hat{\theta}) = J(\hat{\theta}) = \frac{n}{\bar{y}^2} (\hat{\theta} = \bar{y})$$

$$\Rightarrow \theta - \hat{\theta} \approx N(0, \bar{y}^2/n)$$

例 6.13 正規分布 (例 6.1)

$$I(\hat{\mu}, \log \hat{\sigma}^2|Y) = J(\hat{\mu}, \log \hat{\sigma}^2|Y) = \begin{pmatrix} n/\hat{\sigma}^2 & 0 \\ 0 & n/2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \mu - \hat{\mu} \\ \log \sigma^2 - \log \hat{\sigma}^2 \end{pmatrix} \approx N \left[\mathbf{0}, \begin{pmatrix} n/\hat{\sigma}^2 & 0 \\ 0 & n/2 \end{pmatrix} \right]$$

仮説検定

帰無仮説 $H_0 : \theta = \theta_0, (\dim \theta = d)$

Wald 検定

$$p_C = P(\chi_d^2 > W(\theta_0, \hat{\theta}) | \theta = \theta_0) \quad (p\text{-値})$$

$$W(\theta_0, \hat{\theta}) = (\theta_0 - \hat{\theta})^T C^{-1} (\theta_0 - \hat{\theta}) : \text{Wald statistic}$$

$$p_C < \alpha \quad \Rightarrow \quad H_0 \text{ を棄却} \quad (\alpha : \text{有意水準})$$

尤度比検定

$$p_L = P(\chi_d^2 > \text{LR}(\theta_0, \hat{\theta}) | \theta = \theta_0) \quad (p\text{-値})$$

$$\text{LR}(\theta_0, \hat{\theta}) = 2[l(\hat{\theta}|Y) - l(\theta_0|Y)]$$

$$p_L < \alpha \quad \Rightarrow \quad H_0 \text{ を棄却} \quad (\alpha : \text{有意水準})$$

帰無仮説 $H_0 : \theta_{(1)} = \theta_{(1),p}$, $\theta = (\theta_{(1)}, \theta_{(2)})$, $\dim\theta_{(1)} = q < d$

$$p_C(\theta_{(1),0}) = P\{\chi_q^2 > (\theta_{(1),0} - \hat{\theta}_{(1)})^T C_{(11)}^{-1} (\theta_{(1),0} - \hat{\theta}_{(1)})\},$$

$$C_{11} = \text{Cov}(\hat{\theta}_{(1)})$$

$$p_L(\theta_{(1),0}) = P\{\chi_q^2 > \text{LR}(\hat{\theta}, \tilde{\theta})\},$$

$$\text{LR}(\hat{\theta}, \tilde{\theta}) = 2\{\ell(\hat{\theta}|Y) - \ell(\tilde{\theta}|Y)\},$$

$$\tilde{\theta} = (\theta_{(1),0}, \hat{\theta}_{(2)}), \quad \hat{\theta}_{(2)} = \underset{\theta_{(2)}}{\text{argmax}} \ell((\theta_{(1)}, \theta_{(2)})|Y)$$

$p_C < \alpha$ or $p_L < \alpha \Rightarrow$ 棄却

例 6.14 正規分布 (例 6.1)

$$\theta = (\mu, \sigma^2), \theta_{(1)} = \mu, \theta_{(2)} = \sigma^2$$

$$H_0 : \mu = \mu_0$$

$$\begin{aligned} \text{LR} &= 2 \left\{ \left(-\frac{n}{2} \log \frac{(n-1)s^2}{n} - \frac{n}{2} \right) - \left(-\frac{n}{2} \log s_0^2 - \frac{n}{2} \right) \right\} \\ &= n \log \frac{ns_0^2}{(n-1)s^2}, \end{aligned}$$

$$s_0^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_0)^2 = \frac{(n-1)s^2}{n} + (\bar{y} - \mu_0)^2$$

$$\text{LR} = n \log \left(1 + \frac{t^2}{n} \right) \approx t^2, \quad t^2 = \frac{n^2(\bar{y} - \mu_0)^2}{(n-1)s^2}$$

事後分布に基づく Bayes inference

θ の点推定

事後分布の平均, 中央値, 最頻値

θ の $1 - \alpha$ 信頼区間

$$[\theta_{\alpha/2}(Y), \theta_{1-\alpha/2}(Y)],$$

$\theta_{\alpha}(Y)$: 事後分布 $p(\theta|Y)$ の 100α パーセント点

$H_0 : \theta = \theta_0$ の p -値

$$\int_R p(\theta|Y) d\theta, \quad R = \{\theta \mid p(\theta|Y) < p(\theta_0|Y)\}$$

共役事前分布

事前分布と事後分布が同じ分布族となるような事前分布

Jeffrey's prior

$$p(\theta) \propto \sqrt{|J(\theta)|}$$

*. パラメータ変換で不変

$$\varphi = g(\theta), p(\theta) \propto \sqrt{|J(\theta)|} \Rightarrow p(\varphi) \propto \sqrt{|J(\varphi)|}$$

*. 有限測度にならない場合がある (improper prior)

例 6.15 共役事前分布による正規分布に関する推測 (例 6.1)

$$\theta = (\mu, \sigma^2)$$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu|\sigma^2),$$

$$\sigma^2 \sim \frac{1}{\chi^2(\nu_0, \sigma_0^2)}, \quad (\mu|\sigma^2) \sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)$$

$$\left(\nu_0 : \text{自由度 } \sigma_0^2 : \text{尺度母数}, \quad \frac{\sigma_0^2}{\sigma^2} \sim \chi_{\nu_0}^2\right)$$

$$\Rightarrow p(\mu, \sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{1}{2\sigma^2}(\nu_0\sigma_0^2 + \kappa_0(\mu_0 - \mu)^2)\right\}$$

事後分布

$$(\sigma^2|Y) \sim \frac{1}{\chi^2(\nu_n, \sigma_n^2)},$$

$$\nu_n = \nu_0 + n, \quad \nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)^2$$

$$(\mu|\sigma^2, Y) \sim N(\mu_n, \sigma^2/\kappa_n),$$

$$\kappa_n = \kappa_0 + n, \quad \mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}$$

$$(\mu|Y) \sim t(\mu_n, \sigma_n^2/\kappa_n, \nu_n)$$

$$\left(\mu_n: \text{平均}, \sigma_n^2/\kappa_n: \text{尺度母数}, \nu_n: \text{自由度}, \frac{(\mu|Y) - \mu_n}{\sigma_n^2/\kappa_n} \sim t_{\nu_n} \right)$$

(μ, σ^2) の Jeffrey's prior

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$$

(共役事前分布において $\kappa_0 = 0, \nu_0 = -1, \sigma_0^2 = 0$ としたもの)

事後分布

$$\begin{aligned}(\sigma^2|Y) &\sim \frac{1}{\chi^2(n-1, s^2)}, \\(\mu|\sigma^2, Y) &\sim N(\bar{y}, \sigma^2/n), \\(\mu|Y) &\sim t(\bar{y}, s^2/n, n-1)\end{aligned}$$

事後分布に基づく μ の信頼区間

$$\mu = \bar{y} \pm t_{n-1}(\alpha/2) \sqrt{\frac{s^2}{n}}$$

例 6.16 重回帰モデル (例 6.10)

$$y_i \sim N(\mu_i, \sigma^2), \quad i = 1, \dots, n \text{ 独立} \quad \mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip},$$

Jeffrey's prior

$$p(\beta_0, \beta_1, \dots, \beta_p, \sigma^2) \propto \frac{1}{\sigma^2}$$

事後分布

$$(\sigma^2 | Y) \sim \frac{1}{\chi^2(n - p - 1, s^2)},$$

$$(\boldsymbol{\beta} | \sigma^2, Y) \sim N_{p+1}(\hat{\boldsymbol{\beta}}, (X^T X)^{-1} \sigma^2),$$

$$(\boldsymbol{\beta} | Y) \sim t_{p+1}(\hat{\boldsymbol{\beta}}, (X^T X)^{-1} s^2, n - p - 1)$$

$$\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

p 次元 t -分布

$$\mathbf{X} \sim t_p(\boldsymbol{\mu}, \Sigma, \nu)$$

$$p(\mathbf{x}) \propto |\Sigma|^{-1/2} \left\{ 1 + \frac{1}{\nu} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}^{-(\nu+p)/2}$$

例 6.17 多項分布 (例 6.3)

共役事前分布 (Dirichlet 分布)

$$p(\pi_1, \dots, \pi_C) \propto \prod_{c=1}^C \pi_c^{\alpha_c - 1}, \quad \pi_c > 0, \quad \sum_{c=1}^C \pi_c = 1$$

事後分布

$$p(\pi_1, \dots, \pi_C | Y) \propto \prod_{c=1}^C \pi_c^{n_c + \alpha_c - 1}, \quad \pi_c > 0, \quad \sum_{c=1}^C \pi_c = 1,$$

$$\mathbf{E}[\pi_c | Y] = \frac{n_c + \alpha_c}{n_+ + \alpha_+}, \quad n_+ = \sum_{c=1}^C n_c = n, \quad \alpha_+ = \sum_{c=1}^C \alpha_c$$

*. $\alpha_c = 1 \Rightarrow$ 一様分布. $\alpha_c = 0.5 \Rightarrow$ Jeffrey's prior

例 6.18 多変量正規分布 $N_K(\boldsymbol{\mu}, \Sigma)$ (例 6.4)

Jeffrey's prior

$$p(\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-(K+1)/2}$$

事後分布

$$(\Sigma|Y) \sim W_K^{-1}(S, n-1),$$

$$(\boldsymbol{\mu}|\Sigma, Y) \sim N_K(\bar{\boldsymbol{y}}, \Sigma/n)$$

シミュレーションによる事後分布の特性値の導出 $\theta = (\theta_1, \theta_2)$,
 θ_1 の事後分布

$$p(\theta_1|Y) = \int p(\theta)L(\theta|Y)d\theta_2 / \int p(\theta)L(\theta|Y)d\theta$$

θ_2 に関する積分が困難な場合

$$\theta^{(d)} = (\theta_1^{(d)}, \theta_2^{(d)}), \quad d = 1, 2, \dots, D \stackrel{i.i.d.}{\sim} p(\theta|Y)$$

事後平均 : $\sum_{d=1}^D \theta_1^{(d)} / D$

信頼区間 : $(\hat{\theta}_{1,2.5}, \hat{\theta}_{1,97.5})$

$\hat{\theta}_{1,2.5}, \hat{\theta}_{1,97.5}$: $\{\theta^{(d)}; d = 1, \dots, D\}$ の標本 100 分位点

性質 6.1B. $\lambda = g(\theta)$: θ の関数

λ の事後分布からの標本は $\lambda^{(d)} = g(\theta^{(d)})$ によって得られる

例 6.19 重回帰モデル (例 6.12)

$$(\sigma^2|Y) \sim \frac{1}{\chi^2(n-p-1, s^2)}$$

$$\Rightarrow \sigma^{(d)2} = s^2/v, \quad v \sim \chi_{n-p-1}^2$$

$$(\boldsymbol{\beta}|\sigma^2, Y) \sim N_{p+1}(\hat{\boldsymbol{\beta}}, (X^T X)^{-1}\sigma^2)$$

$$\Rightarrow \boldsymbol{\beta}^{(d)} = \hat{\boldsymbol{\beta}} + A^T \boldsymbol{z}\sigma^{(d)}, \quad \boldsymbol{z} \sim N_p(\mathbf{0}, I_p), \quad A^T A = (X^T X)^{-1}$$

$$\lambda = \beta_1/\beta_2$$

$$\Rightarrow \lambda^{(d)} = \beta_1^{(d)}/\beta_2^{(d)}$$

例 6.20 多項分布 (例 6.17)

$$p(\pi_1, \dots, \pi_C | Y) \propto \prod_{c=1}^C \pi_c^{n_c + \alpha_c - 1}$$
$$\Rightarrow \pi_c^{(d)} = v_c / \sum_{j=1}^C v_j, \quad v_j \sim \chi_{2(n_c + \alpha_c)}^2, \quad j = 1, \dots, C, \quad \text{独立}$$

例 6.21 多變量正規分布 (例 6.18)

$$(\Sigma | Y) \sim W_K^{-1}(S, n - 1)$$
$$\Rightarrow \Sigma^{(d)} = (B^T B)^{-1} A, \quad A^T A = S^{-1},$$
$$B = (b_{jk}) : \text{上三角行列, } b_{jj}^2 \sim \chi_{n-j}^2, \quad b_{jk} \sim N(0, 1) \quad (j < k)$$

6.2 尤度に基づく推測 (不完全データ)

$Y = (Y_{\text{obs}}, Y_{\text{mis}})$ Y_{obs} : 観測値, Y_{mis} : 欠測値

$$f(Y_{\text{obs}}|\theta) = \int f(Y_{\text{obs}}, Y_{\text{mis}}|\theta) dY_{\text{mis}}$$

欠測機構を無視した尤度

$$L_{\text{ign}}(\theta|Y_{\text{obs}}) \propto f(Y_{\text{obs}}|\theta), \theta \in \Omega_{\theta}.$$

最尤推定量

$$\hat{\theta}_{\text{ign}} = \underset{\theta}{\operatorname{argmax}} L_{\text{ign}}(\theta|Y_{\text{obs}})$$

事後分布

$$p(\theta|Y_{\text{obs}}) \propto p(\theta) \times L_{\text{ign}}(\theta|Y_{\text{obs}})$$

M : 欠測を表す変数

$$M_{ij} = \begin{cases} 1 & y_{ij} : \text{欠測値} \\ 0 & y_{ij} : \text{観測値} \end{cases}$$

$$f(Y, M|\theta, \psi) = f(Y|\theta)f(M|Y, \psi), \quad (\theta, \psi) \in \Omega_{\theta, \psi},$$

$$f(Y_{\text{obs}}, M|\theta, \psi) = \int f(Y_{\text{obs}}, Y_{\text{mis}}|\theta)f(M|Y_{\text{obs}}, Y_{\text{mis}}, \psi)dY_{\text{mis}}$$

欠測機構を取り入れた尤度

$$L_{\text{full}}(\theta, \psi|Y_{\text{obs}}, M) \propto f(Y_{\text{obs}}, M|\theta, \psi), \quad (\theta, \psi) \in \Omega_{\theta, \psi}$$

MAR(missing at random)

$$f(M|Y_{\text{obs}}, Y_{\text{mis}}, \psi) = f(M|Y_{\text{obs}}, \psi) \text{ for all } Y_{\text{mis}}$$

$$\begin{aligned} \Rightarrow f(Y_{\text{obs}}, M|\theta, \psi) &= f(M|Y_{\text{obs}}, \psi) \int f(Y_{\text{obs}}, Y_{\text{mis}}|\theta) dY_{\text{mis}} \\ &= f(M|Y_{\text{obs}}, \psi) f(Y_{\text{obs}}|\theta) \end{aligned}$$

$\Omega_{\theta, \psi} = \Omega_{\theta} \times \Omega_{\psi}$ ならば

$$\operatorname{argmax}_{\theta} L_{\text{full}}(\theta, \psi|Y_{\text{obs}}, M) = \operatorname{argmax}_{\theta} L_{\text{ign}}(\theta|Y_{\text{obs}}) \quad \text{for all } \psi$$

定義 6.4. 最尤法において欠測機構が無視可能 (ignorable) とは

(a) MAR : missing at random

(b) Distinctness : $\Omega_{\theta, \psi} = \Omega_{\theta} \times \Omega_{\psi}$

事後分布

$$p(\theta, \psi | Y_{\text{obs}}, M) \propto p(\theta, \psi) \times L_{\text{full}}(\theta, \psi | Y_{\text{obs}}, M)$$

MAR かつ 事前独立 ($p(\theta, \psi) = p(\theta)p(\psi)$) \Rightarrow 事後独立

$$\begin{aligned} p(\theta, \psi | Y_{\text{obs}}, M) &\propto \{p(\theta)L(\theta | Y_{\text{obs}})\} \{p(\psi)L(\psi | Y_{\text{obs}}, M)\} \\ &\propto p(\theta | Y_{\text{obs}})p(\psi | Y_{\text{obs}}, M) \end{aligned}$$

定義 6.5. ベイズ法において欠測機構が無視可能 (ignorable) とは

(a) MAR : missing at random

(b) 事前独立 : $p(\theta, \psi) = p(\theta)p(\psi)$

例 6.22 指数分布 (例 6.2)

$$Y = (Y_1, \dots, Y_n)^T$$

$$f(Y|\theta) = \theta^{-n} \exp\left(-\sum_{i=1}^n \frac{y_i}{\theta}\right)$$

$$Y_{\text{obs}} = (Y_1, \dots, Y_r)^T, \quad Y_{\text{mis}} = (Y_{r+1}, \dots, Y_n)^T$$

$$f(Y_{\text{obs}}|\theta) = \theta^{-r} \exp\left(-\sum_{i=1}^r \frac{y_i}{\theta}\right)$$

$$M = (M_1, \dots, M_n)^T, \text{ MCAR}$$

$$f(M|Y, \psi) = \psi^r (1 - \psi)^{n-r}$$

$$\Rightarrow f(Y_{\text{obs}}, M|\theta) = \psi^r (1 - \psi)^{n-r} \theta^{-r} \exp\left(-\sum_{i=1}^r \frac{y_i}{\theta}\right)$$

$$\operatorname{argmax}_{\theta} L_{\text{full}}(\theta, \psi|Y_{\text{obs}}, M) = \frac{1}{r} \sum_{i=1}^r y_i$$

Censored data

$$f(M|Y, \psi) = \prod_{i=1}^n f(M_i|y_i, \psi),$$

$$f(M_i|y_i, \psi) = \begin{cases} 1, & M_i = 1 \text{ and } y_i \geq c, \text{ or } M_i = 0 \text{ and } y_i < c \\ 0, & \text{その他} \end{cases}$$

\Rightarrow

$$\begin{aligned} f(Y_{\text{obs}}, M|\theta) &= \prod_{i=1}^r f(y_i, M_i|\theta) \prod_{i=r+1}^n f(M_i|\theta) \\ &= \prod_{i=1}^r f(y_i|\theta) P(y_i < c|y_i) \prod_{i=r+1}^n P(y_i \geq c|y_i) \\ &= \theta^{-r} \exp\left(-\sum_{i=1}^r \frac{y_i}{\theta}\right) \exp\left(-\frac{(n-r)c}{\theta}\right) \end{aligned}$$

最尤推定値

$$\hat{\theta} = \frac{1}{r} \left\{ \sum_{i=1}^r y_i + (n-r)c \right\}$$

例 6.23 2変量正規分布, 単調欠損データ

$$Y = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_r \\ y_{r+1,1} & ? \\ \vdots & \vdots \\ y_{n,1} & ? \end{pmatrix} \quad \begin{aligned} \ell_{\text{ign}}(\boldsymbol{\mu}, \Sigma | Y_{\text{obs}}) &= \log L_{\text{ign}}(\boldsymbol{\mu}, \Sigma | Y_{\text{obs}}) \\ &= -\frac{1}{2}r \log |\Sigma| - \frac{1}{2} \sum_{i=1}^r (\mathbf{y}_i - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\mu})^T \\ &\quad - \frac{1}{2}(n-r) \log \sigma_{11} - \frac{1}{2} \sum_{i=r+1}^n \frac{(y_{i1} - \mu_1)^2}{\sigma_{11}} \end{aligned}$$

$\theta = (\boldsymbol{\mu}, \Sigma)$, とし, 定義 6.4 を満たすなら $\ell_{\text{ign}}(\boldsymbol{\mu}, \Sigma | Y_{\text{obs}})$ を最大にすることで最尤推定量が得られる. Bayes 法では, MAR に加えて, $p(\boldsymbol{\mu}, \Sigma, \psi) = p(\boldsymbol{\mu}, \Sigma)p(\psi)$ であれば事後密度は $L_{\text{ign}}(\boldsymbol{\mu}, \Sigma | Y_{\text{obs}})$ に比例する.

例 6.24 一元配置分散分析, 欠損が群平均に依存する場合

$$X = \{x_{ij}; i = 1, \dots, I, j = 1, \dots, n_i\}$$

ランダム効果一元配置分散分析モデル

$$(x_{ij}|\theta) \sim_{\text{ind}} N(\mu_i, \sigma^2),$$

$$(\mu_i|\theta) \sim_{\text{ind}} N(\mu, \tau^2),$$

$\theta = (\mu, \sigma^2, \tau^2)$: 未知母数, μ_i : 観測されない群平均

$$P(m_{ij} = 1|X, \mu_i, \psi) \equiv \pi(\mu_i; \psi) = \frac{\exp(\psi_0 + \psi_1 \mu_i)}{1 + \exp(\psi_0 + \psi_1 \mu_i)}$$

$Y = (X, \{\mu_i\})$: 全変数, $Y_{\text{obs}} = X_{\text{obs}}$ and $Y_{\text{mis}} = (X_{\text{mis}}, \{\mu_i\})$

欠損機構は, μ_i (欠測値の一部) に依存 \Rightarrow “無視” できない

$$L_{\text{full}}(\theta, \psi | X_{\text{obs}}, M) = \sigma^{-r} \tau^{-1} \prod_{i=1}^I \int \pi(\mu_i; \psi)^{r_i} [1 - \pi(\mu_i; \psi)]^{n-r_i} \exp\left\{-\frac{r_i}{2\sigma^2}(\bar{x}_{\text{obs},i} - \mu_i)^2 - \frac{1}{2\tau^2}(\mu_i - \mu)^2\right\} d\mu_i$$

r_i : 第 i 群からの観測値の個数, $r = \sum_{i=1}^I r_i$,

$\bar{x}_{\text{obs},i}$: 第 i 群の観測値の標本平均

固定効果分散分析モデル

$$x_{ij} | \mu_i, \sigma^2 \sim_{\text{ind}} N(\mu_i, \sigma^2),$$

$\theta^* = (\{\mu_i\}, \sigma^2)$: 未知パラメータ

$$Y = X, \quad Y_{\text{obs}} = X_{\text{obs}}, Y_{\text{mis}} = X_{\text{mis}}$$

$$\begin{aligned}
L_{\text{full}}(\theta^*, \psi | X_{\text{obs}}, M) &= \sigma^{-r} \prod_{i=1}^I \exp \left\{ -\frac{r_i}{2\sigma^2} (\bar{x}_{\text{obs},i} - \mu_i)^2 \pi(\mu_i; \psi)^{r_i} [1 - \pi(\mu_i; \psi)]^{n_i - r_i} \right\}, \\
L_{\text{ign}}(\theta^* | X_{\text{obs}}, M) &= \sigma^{-r} \prod_{i=1}^I \exp \left\{ -\frac{r_i}{2\sigma^2} (\bar{x}_{\text{obs},i} - \mu_i)^2 \right\}
\end{aligned}$$

- * 固定効果モデルでは、MAR であるが定義 6.4 (b) を満たしていない。
- * MAR なので、 L_{ign} による最尤推定量を用いることができるが、各群からの欠測値の個数も、 μ_i に関する情報を持っているので、有効推定量にはならない。

6.3 パラメータと欠測値に関する最大化

手法

欠損機構を無視できる場合

$$L_{\text{mispar}}(\theta, Y_{\text{mis}}|Y_{\text{obs}}) = L(\theta|Y_{\text{obs}}, Y_{\text{mis}}) = f(Y_{\text{obs}}, Y_{\text{mis}}|\theta),$$

Y_{mis} もパラメータと見なして $L_{\text{mispar}}(\theta, Y_{\text{mis}}|Y_{\text{obs}})$ を θ, Y_{mis} の両方に関して最大化する.

欠損機構が無視できない場合

$$\begin{aligned} L_{\text{mispar}}(\theta, \psi, Y_{\text{mis}}|Y_{\text{obs}}, M) &= L(\theta, \psi|Y_{\text{obs}}, Y_{\text{mis}}, M) \\ &= f(Y_{\text{obs}}, Y_{\text{mis}}|\theta)f(M|Y_{\text{obs}}, Y_{\text{mis}}, \psi) \end{aligned}$$

*. 妥当な推定方法を与えるとは限らない.

例 6.25 1次元正規分布

$Y = (Y_{\text{obs}}, Y_{\text{mis}})$, MAR, distinctness を仮定

$Y_{\text{obs}} = (y_1, \dots, y_r)$: 観測値, $Y_{\text{mis}} = (y_{r+1}, \dots, y_n)$: 欠測値

$Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2), \theta = (\mu, \sigma^2)$

$$f(Y|\theta) = \prod_{i=1}^r f(y_i|\theta) \times \prod_{i=r+1}^n f(y_i|\theta)$$

$$\Rightarrow L_{\text{ign}}(\theta|Y_{\text{obs}}) = \prod_{i=1}^r f(y_i|\theta) : \text{観測値の尤度}$$

$L_{\text{ign}}(\theta|Y_{\text{obs}})$ を θ に関して最大とすると

$$\hat{\mu} = \frac{1}{r} \sum_{i=1}^r y_i, \quad \hat{\sigma}^2 = \frac{1}{r} \sum_{i=1}^r (y_i - \hat{\mu})^2$$

$$L_{\text{pisparr}}(\theta, Y_{\text{mis}}|Y_{\text{obs}}) = f(Y_{\text{obs}}|\theta)f(Y_{\text{mis}}|\theta)$$

を Y_{mis} に関して最大化すると

$$Y_{\text{mis}} = (\tilde{y}_{r+1}, \dots, \tilde{y}_n), \quad \tilde{y}_i = \mu, \quad (i = r + 1, \dots, n)$$

が得られる. これを L_{mispar} に代入して μ, σ^2 について最大化すると

$$\tilde{\mu} = \hat{\mu}, \quad \tilde{\sigma}^2 = \frac{r}{n}\hat{\sigma}^2$$

となる.

*. $n \rightarrow \infty$ のとき, $r \rightarrow \infty$ となるなら, $\hat{\sigma}^2 \rightarrow \sigma^2$.

*. $r \rightarrow \infty$ であっても, $r/n \rightarrow 1$ でない場合には, $\tilde{\sigma}^2$ は σ^2 に収束しない.

例 6.26 回帰分析

例 6.25 に加えて, 共変量 $X = (x_1, \dots, x_n)$ が観測されるとする.

$$Y_i|_{X=x_i} \sim N(\beta_0 + x_i\beta_1, \sigma^2) \quad i = 1, \dots, n$$

$\theta = (\beta_0, \beta_1, \sigma^2)$ の最尤推定値は $L_{\text{ign}}(\theta|Y_{\text{obs}})$ の最大化によって得られる.

*. L_{mispar} を最大とする β_0, β_1 は最尤推定値と一致するが, σ^2 は, 観測値の比率 r/n を掛けたものとなる.

例 6.27 指数分布の打ち切りデータ (例 6.22)

$$f(Y|\theta) = \theta^{-n} \exp\left(-\sum_{i=1}^n \frac{y_i}{\theta}\right)$$

Censored data

$$f(M|Y, \psi) = \prod_{i=1}^n f(M_i|y_i, \psi),$$

$$f(M_i|y_i, \psi) = \begin{cases} 1, & M_i = 1 \text{ and } y_i \geq c, \text{ or } M_i = 0 \text{ and } y_i < c \\ 0, & \text{その他} \end{cases}$$

$$\Rightarrow \hat{\theta} = \frac{1}{r} \left\{ \sum_{i=1}^r y_i + (n - r)c \right\}$$

$$\begin{aligned}
L_{\text{mispar}}(\theta, Y_{\text{mis}} | Y_{\text{obs}}) &= f(Y_{\text{obs}} | \theta) f(Y_{\text{mis}} | \theta) \\
&= \theta^{-n} \exp\left(-\sum_{i=1}^r \frac{y_i}{\theta}\right) \exp\left(-\sum_{i=r+1}^n \frac{y_i}{\theta}\right)
\end{aligned}$$

y_{r+1}, \dots, y_n について最大化すると $\tilde{y}_i = c$ ($i > r$) となり, $L_{\text{mispar}}(\theta, \tilde{Y}_{\text{mis}} | Y_{\text{obs}})$ を θ について最大化すると

$$\tilde{\theta} = \frac{r}{n} \hat{\theta}$$

が得られる.

例 6.28 一般化線形混合モデル

$$f(y_i|x_i, u_i, \beta, \phi) = \exp\left\{\frac{y_i\delta(x_i, u_i, \beta) - b(\delta(x_i, u_i, \beta))}{\phi} + c(y_i, \phi)\right\},$$

u_i : ランダム効果 $\sim f(u_i|x_i, \alpha)$ (α : パラメータ)

$$\mu_i = E(y_i|x_i, u_i, \beta, \phi) = g^{-1}\left(\beta_0 + \sum_{j=1}^p \beta_j x_{ij} + u_i\right),$$

対数尤度関数

$$\ell(\beta, \phi, \alpha|Y) = \sum_{i=1}^n \log\left\{f(y_i|x_i, u_i, \beta, \phi)f(u_i|x_i, \alpha)du_i\right\}$$

Lee and Nelder (1996)

$$Y_{\text{obs}} = (y_1, \dots, y_n), \quad Y_{\text{mis}} = (u_1, \dots, u_n)$$

$$\begin{aligned} \log L_{\text{mispar}}(\beta, \phi, \alpha, Y_{\text{mis}} | Y_{\text{obs}}) &= h(\beta, \phi, \alpha, u | Y) \\ &= \sum_{i=1}^n \log f(y_i | x_i, u_i, \beta, \phi) + \log f(u_i | x_i, \alpha) \end{aligned}$$

*. $h(\beta, \phi, \alpha, u | Y)$ を最大とする推定値は一致推定量となるとは限らない.

Coarsened data (粗雑化データ : 欠損データの一般化)

$\mathcal{Y} : Y$ の値域 , G : 粗雑化変数

$\tilde{Y}(Y, G) : \text{coarsened data}$, $\tilde{Y} : (Y, G) \mapsto \mathcal{A} \in \mathcal{Y}$

M : 欠測の有無を表す変数

$$\Rightarrow \tilde{Y}(Y, M) = \begin{cases} \{Y\} & \text{if } M = 0 \\ \mathcal{Y} & \text{if } M = 1 \end{cases}$$

例 6.29 Stochastic Censoring Time

Y : ある事象が起こるまでの時間

G : 打ち切りがおこるまでの時間

$(y_1, g_1), \dots, (y_n, g_n) : \text{i.i.d.}$

$$y_{\text{obs},i} = y_{\text{obs},i}(y_i, g_i) = \begin{cases} \{y_i\} & \text{if } y_i \leq g_i \\ (g_i, \infty) & \text{if } y_i > g_i \end{cases}$$

$$Y = (Y_{\text{obs}}, Y_{\text{mis}}), \quad G = (G_{\text{obs}}, G_{\text{mis}})$$

打ち切りが起こるまでに事象が起こる $\Rightarrow G$ は観測されない

事象が起こる前に打ち切られる $\Rightarrow Y$ は観測されない

$f(Y|\theta)$: Y の分布, θ : 未知母数

$f(g|y; \phi)$: $Y = y$ のときの G の条件付分布, ϕ : 未知母数

$$L_{\text{full}}(\theta, \phi | Y_{\text{obs}}, G_{\text{obs}}) = \iint f(G|Y; \phi) f(Y|\theta) dY_{\text{mis}} dG_{\text{mis}},$$

$$L_{\text{ign}}(\theta, \phi | Y_{\text{obs}}) = \int f(Y|\theta) dY_{\text{mis}}$$

定義 6.6

$$f(g|y_{\text{obs}}, y_{\text{mis}}, \phi) = f(g|y_{\text{obs}}, \theta) \text{ for all } y_{\text{mis}}$$

が成り立つとき, coarsened at random (CAR) と言う.

定義 6.7 L_{ign} による推測と L_{full} による推測が一致するとき, coarsening mechanism は無視可能という.

*. 最尤法において coarsening mechanism が無視可能であるための十分条件は CAR であって, かつ, θ と ϕ が distinct. ベイズ法において無視可能であるための十分条件は CAR であって, かつ, θ と ϕ が事前独立であること.

例 6.30 (例 6.29 の続き) $i \neq j$ のとき (y_i, g_i) と (y_j, g_j) は独立であるとする

$$L_{\text{full}}(\theta, \phi | Y_{\text{obs}}, G_{\text{obs}}) = \prod_{i: g_i \geq y_i} \int_{g \geq y_i} f(y_i | x_i, \theta) f(g | x_i, y_i, \theta) dg$$

$$\times \prod_{i: g_i < y_i} \int_{g < y_i} f(g_i | x_i, y, \theta) f(y_i | x_i, \theta) dy$$

x_i : 共変量

$$L_{\text{ign}}(\theta | Y_{\text{obs}}) = \prod_{i: g_i \geq y_i} f(y_i | x_i, \theta) \prod_{i: g_i < y_i} \int_{g < y_i} f(y_i | x_i, \theta) dy$$

Data が CAR :

$$f(g_i | y_i, x_i, \psi) = f(g_i | x_i, \psi)$$

7 単調欠測データの推測

7.1 対数尤度の分解

欠損機構が無視できる場合: $\ell(\theta|Y_{\text{obs}}) = \ell_{\text{ign}}(\theta|Y_{\text{obs}})$

パラメータ変換 $\phi = \phi(\theta) = (\phi_1, \dots, \phi_J)$

以下を満たすものがあるとする:

$$\ell(\phi|Y_{\text{obs}}) = \ell_1(\phi_1|Y_{\text{obs}}) + \dots + \ell_J(\phi_J|Y_{\text{obs}})$$

1. ϕ_1, \dots, ϕ_J は distinct (母数空間が直積) で, Bayes 法においては, 事前独立
2. 各 $\ell_j(\phi_j|Y_{\text{obs}})$ は完全データの対数尤度に対応

$$\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_J), \quad \hat{\phi}_j = \operatorname{argmax}_{\phi_j} \ell_j(\phi_j | Y_{\text{obs}}) \quad (j = 1, \dots, J)$$

$$I(\phi | Y_{\text{obs}}) = \begin{bmatrix} I(\phi_1 | Y_{\text{obs}}) & & & O \\ & I(\phi_2 | Y_{\text{obs}}) & & \\ & & \ddots & \\ O & & & I(\phi_J | Y_{\text{obs}}) \end{bmatrix}$$

$$\operatorname{Cov}(\phi - \hat{\phi} | Y_{\text{obs}}) \approx \begin{bmatrix} I^{-1}(\hat{\phi}_1 | Y_{\text{obs}}) & & & O \\ & I^{-1}(\hat{\phi}_2 | Y_{\text{obs}}) & & \\ & & \ddots & \\ O & & & I^{-1}(\hat{\phi}_J | Y_{\text{obs}}) \end{bmatrix}$$

$$\text{Cov}(\theta - \hat{\theta} | Y_{\text{obs}}) \approx D(\hat{\theta}) \text{Cov}(\phi - \hat{\phi} | Y_{\text{obs}}) D(\hat{\theta})^T,$$

$$D(\hat{\theta}) = \{d_{jk}(\hat{\theta})\}, \quad d_{jk}(\theta) = \frac{\partial \theta_j}{\partial \phi_k}$$

7.2 2変量正規分布, 単調欠測データ

最尤推定量

例 7.1 2変量正規分布, 単調欠測データ (例 6.23)

$$(y_{i1}, y_{i2}) \quad i = 1, \dots, n \quad \overset{i.i.d.}{\sim} N_2(\boldsymbol{\mu}, \Sigma),$$

$$Y_{\text{obs}} = [\{(y_{i1}, y_{i2}); i = 1, \dots, r\}, \{y_{i1}; i = r + 1, \dots, n\}]$$

$$Y_{\text{mis}} = \{y_{i2}; i = r + 1, \dots, n\}$$

$$f(y_{i1}, y_{i2} | \boldsymbol{\mu}, \Sigma) = f(y_1 | \mu_1, \sigma_{11}) f(y_{i2} | y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}),$$

$$y_{i1} \sim N(\mu_1, \sigma_{11}), \quad y_{i2} | y_{i1} \sim N(\beta_{20.1} + \beta_{21.1} y_{i1}, \sigma_{22.1})$$

$$\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^T \quad \Leftrightarrow \quad \boldsymbol{\phi} = (\mu_1, \sigma_{11}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1})^T$$

$$\begin{aligned} \mu_2 &= \beta_{20.1} + \beta_{21.1} \mu_1, & \beta_{21.1} &= \sigma_{12} / \sigma_{11}, \\ \sigma_{12} &= \beta_{21.1} \sigma_{11}, & \beta_{20.1} &= \mu_2 - \beta_{21.1} \mu_1 \\ \sigma_{22} &= \sigma_{22.1} + \beta_{21.1}^2 \sigma_{11} & \sigma_{22.1} &= \sigma_{11} - \sigma_{12}^2 / \sigma_{11} \end{aligned}$$

$$\begin{aligned} f(Y_{\text{obs}} | \boldsymbol{\theta}) &= \prod_{i=1}^r f(y_{i1}, y_{i2} | \boldsymbol{\theta}) \prod_{i=r+1}^n f(y_{i1} | \boldsymbol{\theta}) \\ &= \prod_{i=1}^n f(y_{i1} | \mu_1, \sigma_{11}) \prod_{i=1}^r f(y_{i2} | y_{i1}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}) \\ &= \ell_1(\mu_1, \sigma_{11} | Y_{\text{obs}}) \ell_2(\beta_{20.1}, \beta_{21.1}, \sigma_{22.1} | Y_{\text{obs}}) \end{aligned}$$

l_1 の最大化

$$\Rightarrow \hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n y_{i1}, \quad \hat{\sigma}_{11} = \frac{1}{n} \sum_{i=1}^n (y_{i1} - \hat{\mu}_1)^2$$

l_2 の最大化 (r 個の完全データに基づく回帰)

$$\Rightarrow \hat{\beta}_{21 \cdot 1} = s_{12}/s_{11}, \quad \hat{\beta}_{20 \cdot 1} = \bar{y}_2 - \hat{\beta}_{21 \cdot 1} \bar{y}_1, \quad \hat{\sigma}_{22 \cdot 1} = s_{22 \cdot 1},$$

$$\bar{y}_j = \frac{1}{r} \sum_{i=1}^r y_{ij},$$

$$s_{jk} = \frac{1}{r} \sum_{i=1}^r (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k), \quad s_{22 \cdot 1} = s_{22} - s_{12}^2/s_{11}$$

$$\hat{\phi} \Rightarrow \hat{\theta}$$

$$\hat{\mu}_2 = \hat{\beta}_{20.1} + \hat{\beta}_{21.1}\hat{\mu}_1 = \bar{y}_2 + \hat{\beta}_{21.1}(\hat{\mu}_1 - \bar{y}_1)$$

$$\hat{\sigma}_{12} = \hat{\beta}_{21.1}\hat{\sigma}_{11},$$

$$\hat{\sigma}_{22} = \hat{\sigma}_{22.1} + \hat{\beta}_{21.1}^2\hat{\sigma}_{11} = s_{22} + \hat{\beta}_{21.1}^2(\hat{\sigma}_{11} - s_{11})$$

相関係数

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = \beta_{21.1}\sqrt{\frac{\sigma_{11}}{\sigma_{22.1} + \beta_{21.1}^2\sigma_{11}}}$$

$$\hat{\rho} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}\sqrt{\frac{\hat{\sigma}_{11}}{s_{11}}}\sqrt{\frac{s_{22}}{\hat{\sigma}_{22}}}$$

imputation

$$\hat{\mu}_2 = \frac{1}{n}\left\{\sum_{i=1}^r y_{i2} + \sum_{i=r+1}^n \hat{y}_{i2}\right\}, \quad \hat{y}_{i2} = \bar{y}_2 + \hat{\beta}_{21.1}(y_{i1} - \bar{y}_1)$$

大標本に基づく共分散行列の推定

$$\begin{aligned} \ell(\phi|Y_{\text{obs}}) &= -\frac{1}{2\sigma_{22.1}} \sum_{i=1}^r (y_{i2} - \beta_{20.1} - \beta_{21.1}y_{i1})^2 - \frac{r}{2} \log \sigma_{22.1} \\ &\quad - \frac{1}{2\sigma_{11}} \sum_{i=1}^n (y_{i1} - \mu_1)^2 - \frac{r}{2} \log \sigma_{11} \end{aligned}$$

$$I(\hat{\phi}|Y_{\text{obs}}) = \begin{bmatrix} I(\hat{\mu}_1, \hat{\sigma}_{11}|Y_{\text{obs}}) & O \\ O & I(\hat{\beta}_{20.1}, \hat{\beta}_{20.1}, \hat{\sigma}_{22.1}|Y_{\text{obs}}) \end{bmatrix}$$

$$I(\hat{\mu}_1, \hat{\sigma}_{11}|Y_{\text{obs}}) = \begin{bmatrix} n/\hat{\sigma}_{11} & 0 \\ 0 & n/(2\hat{\sigma}_{11}^2) \end{bmatrix},$$

$$I(\hat{\beta}_{20.1}, \hat{\beta}_{20.1}, \hat{\sigma}_{22.1}|Y_{\text{obs}}) = \begin{bmatrix} r/\hat{\sigma}_{22.1} & r\bar{y}_1/\hat{\sigma}_{22.1} & 0 \\ r\bar{y}_1/\hat{\sigma}_{22.1} & \sum_{i=1}^r y_{i1}^2/\hat{\sigma}_{22.1} & 0 \\ 0 & 0 & r/(2\hat{\sigma}_{22.1}^2) \end{bmatrix}$$

$$\text{Cov}(\phi - \hat{\phi}) \approx \begin{bmatrix} I^{-1}(\hat{\mu}_1, \hat{\sigma}_{11}|Y_{\text{obs}}) & O \\ O & I^{-1}(\hat{\beta}_{20.1}, \hat{\beta}_{20.1}, \hat{\sigma}_{22.1}|Y_{\text{obs}}) \end{bmatrix}$$

$$I^{-1}(\hat{\mu}_1, \hat{\sigma}_{11}|Y_{\text{obs}}) = \begin{bmatrix} \hat{\sigma}_{11}/n & 0 \\ 0 & 2\hat{\sigma}_{11}^2/n \end{bmatrix}$$

$$I^{-1}(\hat{\beta}_{20.1}, \hat{\beta}_{20.1}, \hat{\sigma}_{22.1}|Y_{\text{obs}})$$

$$= \begin{bmatrix} \hat{\sigma}_{22.1}(1 + \bar{y}_1^2/s_{11})/r & -\bar{y}_1\hat{\sigma}_{22.1}/(rs_{11}) & 0 \\ -\bar{y}_1\hat{\sigma}_{22.1}/(rs_{11}) & \hat{\sigma}_{22.1}/(rs_{11}) & 0 \\ 0 & 0 & 2\hat{\sigma}_{22.1}^2/r \end{bmatrix}$$

$$\text{Cov}(\theta - \hat{\theta} | Y_{\text{obs}}) \approx D(\hat{\theta}) \text{Cov}(\phi - \hat{\phi} | Y_{\text{obs}}) D(\hat{\theta})^T,$$

$$\begin{aligned} D(\mu_2) &= \left(\frac{\partial \mu_2}{\partial \mu_1}, \frac{\partial \mu_2}{\partial \sigma_{11}}, \frac{\partial \mu_2}{\partial \beta_{20.1}}, \frac{\partial \mu_2}{\partial \beta_{21.1}}, \frac{\partial \mu_2}{\partial \sigma_{22.1}} \right) \\ &= (\beta_{21.1}, 0, 1, \mu_1, 0) \end{aligned}$$

$$\begin{aligned} \text{Var}(\mu_2 - \hat{\mu}_2) &= D(\mu_2) \text{Cov}(\phi - \hat{\phi}) D(\mu_2)^T \\ &\approx \hat{\sigma}_{22.1} \left[\frac{1}{r} + \frac{\hat{\rho}^2}{n(1 - \hat{\rho}^2)} + \frac{(\bar{y}_1 - \hat{\mu}_1)^2}{r s_{11}} \right] = \frac{\hat{\sigma}_{22}}{r} \left(1 - \hat{\rho}^2 \frac{n - r}{n} \right) \end{aligned}$$

信賴区間

$$\mu_2 = \hat{\mu}_2 \pm 1.96 \sqrt{\widehat{\text{Var}}(\mu_2 - \hat{\mu}_2)},$$

$$\log \sigma_{22} = \log \hat{\sigma}_{22} \pm 1.96 \sqrt{\widehat{\text{Var}}(\log \sigma_{22} - \log \hat{\sigma}_{22})}$$

7.3 小標本での推測

事前分布

$$f(\mu_1, \sigma_{11}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1}) \propto \sigma_{11}^{-a} \sigma_{22.1}^{-c}$$

($a = c = 1$ のときは Jeffreys' prior)

事後分布

$$(1) \quad n\hat{\sigma}_{11}/\sigma_{11} \sim \chi_{n+2a-3}^2$$

$$(2) \quad \mu_1|\sigma_{11} \sim N(\hat{\mu}_1, \sigma_{11}/n)$$

$$(3) \quad r\hat{\sigma}_{22.1}/\sigma_{22.1} \sim \chi_{r+2c-4}^2$$

$$(4) \quad \beta_{21.1}|\sigma_{22.1} \sim N(\hat{\beta}_{22.1}, \sigma_{22.1}/(r s_{11}))$$

$$(5) \quad \beta_{20.1}|(\beta_{21.1}, \sigma_{22.1}) \sim N(\bar{y}_2 - \beta_{22.1}\bar{y}_1, \sigma_{22.1}/r)$$

(6) (σ_{11}, μ_1) と $(\sigma_{22.1}, \beta_{21.1}, \beta_{20.1})$ は事後独立

シミュレーションによる推測 $d = 1, 2, \dots, D$ に対して

$$1. x_1^{(d)} \sim \chi_{n+2a-3}^2, \quad x_2^{(d)} \sim \chi_{r+2c-4}^2, \quad z_1^{(d)}, z_2^{(d)}, z_3^{(d)} \stackrel{i.i.d.}{\sim} N(0, 1)$$

$$2. \phi^{(d)} = (\mu_1^{(d)}, \sigma_{11}^{(d)}, \beta_{20 \cdot 1}^{(d)}, \beta_{21 \cdot 1}^{(d)}, \sigma_{22 \cdot 1}^{(d)})$$

$$\sigma_{11}^{(d)} = \frac{n \hat{\sigma}_{11}}{x_1^{(d)}} \quad \mu_1^{(d)} = \hat{\mu}_1 + z_1^{(d)} \sqrt{\sigma_{11}/n}$$

$$\sigma_{22 \cdot 1}^{(d)} = \frac{r \hat{\sigma}_{22 \cdot 1}}{x_2^{(d)}}$$

$$\beta_{21 \cdot 1} = \hat{\beta}_{22 \cdot 1} + z_2^{(d)} \sqrt{\sigma_{22 \cdot 1}/(r s_{11})}$$

$$\beta_{20 \cdot 1} = \bar{y}_2 - \beta_{22 \cdot 1} \bar{y}_1 + z_3^{(d)} \sqrt{\sigma_{22 \cdot 1}/r}$$

$$3. g^{(d)} = g(\phi^{(d)})$$

$$\text{例えば, } \mu_2^{(d)} = \beta_{20 \cdot 1}^{(d)} + \beta_{21 \cdot 1}^{(d)} \mu_1^{(d)}$$

$\{\mu_2^{(d)}; d = 1, \dots, D\}$ の経験分布に基づいて μ_2 の信頼区間等を構築

7.4 3変量以上の単調欠測データ

1変数のみに欠測がある場合

$$M = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

例 7.2 $K + 1$ 変量正規分布

$$(\mathbf{y}_{i1}^T, y_{i2})^T \quad i = 1, \dots, n \stackrel{i.i.d.}{\sim} N_{K+1}(\boldsymbol{\mu}, \Sigma),$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{12}^T & \sigma_{22} \end{pmatrix}$$

$$Y_{\text{obs}} = [\{(\mathbf{y}_{i1}^T, y_{i2}); i = 1, \dots, r\}, \{y_{i1}; i = r + 1, \dots, n\}]$$

$$Y_{\text{mis}} = \{y_{i2}; i = r + 1, \dots, n\}$$

最尤推定量

$$\hat{\mu}_2 = \bar{y}_2 + (\hat{\mu}_1 - \bar{y}_1)^T \hat{\beta}_{21 \cdot 1}$$

$$\hat{\sigma}_{22} = s_{22} + \hat{\beta}_{21 \cdot 1}^T (\hat{\Sigma}_{11} - S_{11}) \hat{\beta}_{21 \cdot 1}$$

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_{i1}, \quad \hat{\Sigma}_{11} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_{i1} - \hat{\mu}_1)(\mathbf{y}_{i1} - \hat{\mu}_1)^T, \quad \hat{\beta}_{21 \cdot 1} = S_{11}^{-1} \mathbf{s}_{12},$$

$$\bar{\mathbf{y}}_1 = \frac{1}{r} \sum_{i=1}^r \mathbf{y}_{i1}, \quad \bar{y}_2 = \frac{1}{r} \sum_{i=1}^r y_{i2},$$

$$\mathbf{s}_{12} = \frac{1}{r} \sum_{i=1}^r (\mathbf{y}_{ik} - \bar{\mathbf{y}}_1)(y_{i2} - \bar{y}_2), \quad S_{11} = \frac{1}{r} \sum_{i=1}^r (\mathbf{y}_{ik} - \bar{\mathbf{y}}_1)(\mathbf{y}_{ik} - \bar{\mathbf{y}}_1)^T$$

*. 同時分布が正規分布でなくても、次の条件の下で、 $\hat{\mu}_2$ は $E(y_{i2})$ の最尤推定量となる。

1. $y_{i2} | \mathbf{y}_{i1} \sim N(\beta_{20 \cdot 1} + \mathbf{y}_{i1}^T \boldsymbol{\beta}_{21 \cdot 1}, \sigma_{22 \cdot 1})$

2. \mathbf{y}_{i1} の分布は次を満たしていれば正規分布でなくてもよい

- (i) $\hat{\boldsymbol{\mu}}_1$ が $E(\mathbf{y}_{i1})$ の最尤推定量

- (ii) $\mu_2, \beta_{20 \cdot 1}, \boldsymbol{\beta}_{21 \cdot 1}, \sigma_{22 \cdot 1}$ は、 \mathbf{y}_{i1} の分布のパラメータと “distinct”

ダミー変数

\mathbf{y}_{i1} : 個体 i が $K + 1$ 個のグループのどれに所属するかを表す

個体 i が第 1 群 $\Rightarrow \mathbf{y}_{i1} = (1, 0, \dots, 0)^T$,

個体 i が第 2 群 $\Rightarrow \mathbf{y}_{i1} = (0, 1, \dots, 0)^T$,

⋮

$$E[y_{i1k}] = P(\text{第 } k \text{ 群に所属})$$

個体 i が第 K 群 $\Rightarrow \mathbf{y}_{i1} = (0, 0, \dots, 0, 1)^T$,

個体 i が第 $K + 1$ 群 $\Rightarrow \mathbf{y}_{i1} = (, 0, \dots, 0)^T$

尤度関数の分解 (一般の単調欠損構造の場合)

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{aligned} & \prod_{i=1}^n f(y_{i1}, \dots, y_{iJ} | \phi) \\ &= \prod_{i=1}^n f(y_{i1} | \phi_1) \prod_{i=1}^{r_2} f(y_{i2} | y_{i1}, \phi_2) \cdots \prod_{i=1}^{r_J} f(y_{iJ} | y_{i1}, \dots, y_{i,J-1}, \phi_J) \end{aligned}$$

例 7.3 多変量正規分布

$f(y_{i1}|\phi_1), f(y_{i2}|y_{i1}, \phi_2), \dots, f(y_{iJ}|y_{i1}, \dots, y_{i,J-1}, \phi_J)$ はいずれも正規分布

1. ϕ_1 : y_1 の平均ベクトル, 共分散行列

y_{i1} ($i = 1, \dots, n$) により推定

2. ϕ_2 : y_2 を目的変数, y_1 を説明変数とした線形回帰による
回帰係数ベクトル (行列) と誤差項の分散 (共分散行列)

(y_{i1}, y_{i2}) ($i = 1, \dots, r_1$) により推定

3. ϕ_3 : y_3 を目的変数, y_1, y_2 を説明変数とした線形回帰

(y_{i1}, y_{i2}, y_{i3}) ($i = 1, \dots, r_3$) により推定

⋮

Sweep Operator による 最尤推定量の計算

$G = (g_{ij}) : p \times p$ 対称行列

$1 \leq k \leq p$

$$H = (h_{ij}) = \text{SWP}[k]G$$

$$\begin{cases} h_{kk} = -1/g_{kk}, \\ h_{jk} = h_{kj} = g_{jk}/g_{kk} \quad j \neq k, \\ h_{jl} = g_{jl} - g_{jk}g_{lk}/g_{kk} \quad j \neq k, l \neq k \end{cases}$$

例. $p = 3$

$$\text{SWP}[1]G = \begin{bmatrix} -1/g_{11} & g_{12}/g_{11} & g_{13}/g_{11} \\ g_{12}/g_{11} & g_{22} - g_{12}^2/g_{11} & g_{23} - g_{13}g_{12}/g_{11} \\ g_{13}/g_{11} & g_{23} - g_{13}g_{12}/g_{11} & g_{33} - g_{13}^2/g_{11} \end{bmatrix}$$

$\text{SWP}[i]\text{SWP}[j] \cdots \text{SWP}[k]G = \text{SWP}[i, j, \dots, k]$ と書く

性質

$$\text{SWP}[i, j] = \text{SWP}[j, i]$$

$$G = \begin{bmatrix} 1 & \bar{y}_1 & \cdots & \bar{y}_j & \cdots & \bar{y}_K \\ \bar{y}_1 & n^{-1} \sum_i y_{1i}^2 & & & \cdots & n^{-1} \sum_i y_{Ki} y_{1i} \\ \vdots & \vdots & \ddots & & & \\ \bar{y}_k & & & n^{-1} \sum_i y_{ji} y_{ki} & & \\ \vdots & & & & \ddots & \vdots \\ \bar{y}_K & n^{-1} \sum_i y_{1i} y_{Ki} & & & \cdots & n^{-1} \sum_i y_{Ki}^2 \end{bmatrix},$$

$$\bar{y}_j = \sum_{i=1}^n y_{ji}$$

*. 行, 列を 第 0 行~第 K 行, 第 0 列~第 K 列と数える

$$\text{SWP}[0]G = \begin{bmatrix} -1 & \bar{y}_1 & \cdots & \bar{y}_j & \cdots & \bar{y}_K \\ \bar{y}_1 & s_{11} & & & \cdots & s_{K1} \\ \vdots & & \ddots & & & \vdots \\ \bar{y}_k & & & s_{jk} & & \\ \vdots & & & & \ddots & \\ \bar{y}_K & s_{1K} & & & \cdots & s_{KK} \end{bmatrix}$$

$$s_{jk} = \frac{1}{n} \sum_{i=1}^n (y_{ji} - \bar{y}_j)(y_{ki} - \bar{y}_k)$$

SWP[0, 1]G

$$= \begin{bmatrix} -(1 + \bar{y}_1^2/s_{11}) & \bar{y}_1/s_{11} & \bar{y}_2 - (s_{12}/s_{11})\bar{y}_1 & \cdots & \bar{y}_K - (s_{1K}/s_{11})\bar{y}_1 \\ & -1/s_{11} & s_{12}/s_{11} & \cdots & s_{1K}/s_{11} \\ & \vdots & s_{22} - s_{12}^2/s_{11} & \cdots & s_{2K} - s_{1K}s_{12}/s_{11} \\ & & & & \vdots \\ \bar{y}_K - (s_{1K}/s_{11})\bar{y}_1 & & \cdots & & s_{KK} - s_{1K}^2/s_{11} \end{bmatrix}$$
$$= \begin{bmatrix} -A & B \\ B^T & C \end{bmatrix}, \quad A : 2 \times 2, \quad B : 2 \times (K - 1), \quad C : (K - 1) \times (K - 1)$$

$\Rightarrow Y_2, \dots, Y_K$ の Y_1 の上への回帰

B : 回帰係数行列, C 残差の共分散行列

$$\text{SWP}[0, 1, \dots, q]G = \begin{bmatrix} -D & E \\ E^T & F \end{bmatrix},$$

$$D : (q + 1) \times (q + 1), \quad E : (q + 1) \times (K - q), \quad F : (K - q) \times (K - q)$$

$\Rightarrow Y_{q+1}, \dots, Y_K$ の Y_1, \dots, Y_q の上への回帰

E : 回帰係数行列, F 残差の共分散行列

reverse sweep operator

$$H = (h_{ij}) = \text{RSW}[k]G$$

$$\begin{cases} h_{kk} = -1/g_{kk}, \\ h_{jk} = h_{kj} = -g_{jk}/g_{kk} \quad j \neq k, \\ h_{jl} = g_{jk}g_{lk}/g_{kk} \quad j \neq k, l \neq k \end{cases}$$

$$\Rightarrow (\text{RSW}[k])(\text{SWP}[k])G = (\text{SWP}[k])(\text{RSW}[k])G = G$$

例 7.4 2 変量正規分布, 単調欠測データ (例 7.1)

$$\theta = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^T \quad \leftrightarrow \quad \phi = (\mu_1, \sigma_{11}, \beta_{20.1}, \beta_{21.1}, \sigma_{22.1})^T$$

$$\mu_2 = \beta_{20.1} + \beta_{21.1}\mu_1,$$

$$\beta_{21.1} = \sigma_{12}/\sigma_{11},$$

$$\sigma_{12} = \beta_{21.1}\sigma_{11},$$

$$\beta_{20.1} = \mu_2 - \beta_{21.1}\mu_1$$

$$\sigma_{22} = \sigma_{22.1} + \beta_{21.1}^2\sigma_{11}$$

$$\sigma_{22.1} = \sigma_{11} - \sigma_{12}^2/\sigma_{11}$$

$$\theta^* = \begin{bmatrix} -1 & \mu_1 & \mu_2 \\ \mu_1 & \sigma_{11} & \sigma_{12} \\ \mu_2 & \sigma_{12} & \sigma_{22} \end{bmatrix}$$

$$\text{SWP}[1]\theta^* = \begin{bmatrix} -(1 + \mu_1^2/\sigma_{11}) & \mu_1/\sigma_{11} & \mu_2 - \mu_1\sigma_{12}/\sigma_{11} \\ \mu_1/\sigma_{11} & -\sigma_{11}^{-1} & \sigma_{12}/\sigma_{11} \\ \mu_2 - \mu_1\sigma_{12}/\sigma_{11} & \sigma_{12}/\sigma_{11} & \sigma_{22} - \sigma_{12}^2/\sigma_{11} \end{bmatrix}$$

$$\phi^* = \text{SWP}[1]\theta^* = \begin{bmatrix} \text{SWP}[1] \begin{bmatrix} -1 & \mu_1 \\ \mu_1 & \sigma_{11} \end{bmatrix} & \begin{bmatrix} \beta_{20.1} \\ \beta_{21.1} \\ \sigma_{22.1} \end{bmatrix} \end{bmatrix}$$

$$\hat{\phi}^* = \text{SWP}[1]\hat{\theta}^* = \begin{bmatrix} \text{SWP}[1] \begin{bmatrix} -1 & \hat{\mu}_1 \\ \hat{\mu}_1 & \hat{\sigma}_{11} \end{bmatrix} & \begin{bmatrix} \hat{\beta}_{20.1} \\ \hat{\beta}_{21.1} \\ \hat{\sigma}_{22.1} \end{bmatrix} \end{bmatrix}$$

$$\hat{\theta}^* = \text{RSW}[1] \begin{bmatrix} \text{SWP}[1] \begin{bmatrix} -1 & \hat{\mu}_1 \\ \hat{\mu}_1 & \hat{\sigma}_{11} \end{bmatrix} & \begin{bmatrix} \hat{\beta}_{20.1} \\ \hat{\beta}_{21.1} \\ \hat{\sigma}_{22.1} \end{bmatrix} \\ \hat{\beta}_{20.1} & \hat{\beta}_{21.1} & \hat{\sigma}_{22.1} \end{bmatrix}$$

例 7.5 多変量正規分布, 単調欠測データ (例 7.3)

3 階層の場合

$$Y = \begin{pmatrix}
 y_{11} & \cdots & y_{1q_1} & y_{1,q_1+1} & \cdots & y_{1,q_2} & y_{1,q_2+1} & \cdots & y_{1p} \\
 \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
 y_{r_1 1} & \cdots & y_{r_1 q_1} & y_{r_1, q_1+1} & \cdots & y_{r_1 q_2} & y_{r_1, q_2+1} & \cdots & y_{r_1 p} \\
 y_{r_1+1, 1} & \cdots & y_{r_1+1, q_1} & y_{r_1+1, q_1+1} & \cdots & y_{r_1+1, q_2} & * & \cdots & * \\
 \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
 y_{r_2 1} & \cdots & y_{r_2 q_1} & y_{r_2, q_1+1} & \cdots & y_{r_2, q_2} & * & \cdots & * \\
 y_{r_2+1, 1} & \cdots & y_{r_2+1, q_1} & * & \cdots & * & * & \cdots & * \\
 \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
 y_{n, 1} & \cdots & y_{n, q_1} & * & \cdots & * & * & \cdots & *
 \end{pmatrix}$$

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} Y = [Y_1 \ Y_2 \ Y_3] = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & * \\ Y_{31} & * & * \end{bmatrix}$$

STEP 1. ブロック 1 のデータから, $\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_{11}$ を推定

$$B_1 = \frac{1}{n} \begin{bmatrix} Y_{11}^T & Y_{21}^T & Y_{31}^T \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \end{bmatrix}$$

$$\text{SWP}[0]B_1 = \begin{bmatrix} -1 & \hat{\boldsymbol{\mu}}_1^T \\ \hat{\boldsymbol{\mu}}_1 & \hat{\boldsymbol{\Sigma}}_{11} \end{bmatrix}$$

STEP 2. ブロック 2 のデータにより, Y_2 を Y_1 上に回帰して,
 $\hat{\boldsymbol{\beta}}_{20\cdot1}, \hat{\boldsymbol{\beta}}_{21\cdot1}, \hat{\boldsymbol{\Sigma}}_{22\cdot1}$ を推定

$$B_2 = \frac{1}{n} \begin{bmatrix} Y_{11}^T & Y_{21}^T \\ Y_{12}^T & Y_{22}^T \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$$

$$\text{SWP}[0, 1, \dots, q_1]B_2 = \begin{bmatrix} * & \hat{\boldsymbol{\beta}}_{20\cdot1}^T \\ * & \hat{\boldsymbol{\beta}}_{21\cdot1}^T \\ \hat{\boldsymbol{\beta}}_{20\cdot1} & \hat{\boldsymbol{\beta}}_{21\cdot1} & \hat{\boldsymbol{\Sigma}}_{22\cdot1} \end{bmatrix}$$

STEP 3. ブロック 3 のデータ により, Y_3 を Y_1, Y_2 上に回帰して,
 $\hat{\beta}_{30.12}, \hat{\beta}_{31.12}, \hat{\beta}_{32.12}, \hat{\Sigma}_{33.12}$ を推定

$$B_3 = \frac{1}{n} \begin{bmatrix} Y_{11}^T \\ Y_{12}^T \\ Y_{13}^T \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \end{bmatrix}$$

$$\text{SWP}[0, 1, \dots, q_2] B_3 = \begin{bmatrix} & & & \hat{\beta}_{30.12}^T \\ & * & & \hat{\beta}_{31.12}^T \\ & & & \hat{\beta}_{32.12}^T \\ \hat{\beta}_{30.12} & \hat{\beta}_{31.12} & \hat{\beta}_{32.12} & \hat{\Sigma}_{33.12} \end{bmatrix}$$

STEP 4.

$$A_1 = \text{SWP}[1, \dots, q_1] \begin{bmatrix} -1 & \hat{\mu}_1^T \\ \hat{\mu}_1 & \hat{\Sigma}_{11} \end{bmatrix}$$

STEP 5.

$$A_2 = \text{SWP}[q_1 + 1, \dots, q_2] \begin{bmatrix} & & \hat{\beta}_{20 \cdot 1}^T \\ & A_1 & \hat{\beta}_{21 \cdot 1}^T \\ \hat{\beta}_{20 \cdot 1} & \hat{\beta}_{21 \cdot 1} & \hat{\Sigma}_{22 \cdot 1} \end{bmatrix}$$

STEP 6.

$$\begin{bmatrix} -1 & \hat{\mu}^T \\ \hat{\mu} & \hat{\Sigma} \end{bmatrix} = \text{RSW}[1, \dots, q_2] \begin{bmatrix} & & & \hat{\beta}_{30 \cdot 12}^T \\ & & & \hat{\beta}_{31 \cdot 12}^T \\ & & A_2 & \hat{\beta}_{32 \cdot 12}^T \\ \hat{\beta}_{30 \cdot 12} & \hat{\beta}_{31 \cdot 12} & \hat{\beta}_{32 \cdot 12} & \hat{\Sigma}_{33 \cdot 12} \end{bmatrix}$$

STEP4~STEP6 をまとめると

$$\begin{bmatrix} -1 & \hat{\boldsymbol{\mu}}^T \\ \hat{\boldsymbol{\mu}} & \hat{\boldsymbol{\Sigma}} \end{bmatrix} \\ = \text{RSW}[\sim 1, \sim 2] \left[\text{SWP}[\sim 2] \left[\text{SWP}[\sim 1] \begin{bmatrix} -1 & \hat{\boldsymbol{\mu}}_1^T \\ \hat{\boldsymbol{\mu}}_1 & \hat{\boldsymbol{\Sigma}}_{11} \end{bmatrix} \begin{matrix} \hat{\boldsymbol{\beta}}_{20 \cdot 1}^T \\ \hat{\boldsymbol{\beta}}_{21 \cdot 1}^T \\ \hat{\boldsymbol{\Sigma}}_{22 \cdot 1} \end{matrix} \right] \begin{matrix} \hat{\boldsymbol{\beta}}_{30 \cdot 12}^T \\ \hat{\boldsymbol{\beta}}_{31 \cdot 12}^T \\ \hat{\boldsymbol{\beta}}_{32 \cdot 12}^T \\ \hat{\boldsymbol{\Sigma}}_{33 \cdot 12} \end{matrix} \right]$$

多変量正規分布, 単調欠測データに対する Bayes 法

適当な事前分布を仮定して, ϕ の事後分布が得られたとする.

$d = 1, \dots, D$ に対して

$$\phi^{(d)} = (\boldsymbol{\mu}_1^{(d)}, \boldsymbol{\Sigma}_{11}^{(d)}, \boldsymbol{\beta}_{20 \cdot 1}^{(d)}, \boldsymbol{\beta}_{21 \cdot 1}^{(d)}, \boldsymbol{\Sigma}_{22 \cdot 1}^{(d)}, \boldsymbol{\beta}_{30 \cdot 12}^{(d)}, \boldsymbol{\beta}_{31 \cdot 12}^{(d)}, \boldsymbol{\beta}_{32 \cdot 12}^{(d)}, \boldsymbol{\Sigma}_{33 \cdot 12}^{(d)})$$

を生成し, sweep operator, reverse sweep operator により STEP4~STEP 6 により, $\theta^{(d)}$ を生成.

8 一般の欠損構造に対する最尤法

8.1 反復法

MAR を仮定, 欠損構造を無視する (distinct なら OK)

$$L(\theta|Y_{\text{obs}}) = \int f(Y_{\text{obs}}, Y_{\text{mis}}|\theta)dY_{\text{mis}}$$

尤度方程式

$$D_{\ell}(\theta|Y_{\text{obs}}) = \frac{\partial \ell(\theta|Y_{\text{obs}})}{\partial \theta} = 0, \quad \ell(\theta|Y_{\text{obs}}) = \log L(\theta|Y_{\text{obs}})$$

ニュートン・ラフソン法

$$\theta^{(t+1)} = \theta^{(t)} + I^{-1}(\theta^{(t)}|Y_{\text{obs}})D_{\ell}(\theta^{(t)}|Y_{\text{obs}}),$$

$$I(\theta|Y_{\text{obs}}) = -\frac{\partial^2 \ell(\theta|Y_{\text{obs}})}{\partial \theta \partial \theta}$$

スコア法

$$\theta^{(t+1)} = \theta^{(t)} + J^{-1}(\theta^{(t)})D_{\ell}(\theta^{(t)}|Y_{\text{obs}}),$$

$$J(\theta) = E[I(\theta|Y_{\text{obs}})|\theta] = -\int \frac{\partial^2 \ell(\theta|Y_{\text{obs}})}{\partial \theta \partial \theta} f(Y_{\text{obs}}|\theta) dY_{\text{obs}}$$

その他 (あまり良くない)

$$\theta^{(t+1)} = \theta^{(t)} + \lambda_t Q^{-1}(\theta^{(t)})D_{\ell}(\theta^{(t)}|Y_{\text{obs}}),$$

$$Q(\theta) = \sum_{i=1}^n \left(\frac{\partial \ell_i}{\partial \theta} \right) \left(\frac{\partial \ell_i}{\partial \theta} \right)^{\text{T}}, \quad \ell_i : \text{個体 } i \text{ の対数尤度}$$

λ_t : ステップ幅を調節

8.2 Expectation Maximization (EM) アルゴリズム

- $\ell(\theta|Y_{\text{obs}})$ の最大化を, $\ell(\theta|Y_{\text{obs}}, Y_{\text{mis}})$ を用いて行う方法
- Expectation ステップ (E ステップ) と Maximization ステップ (M ステップ) を交互に繰り返して行う.

利点

- コンピュータによる実現が容易
- 各ステップの統計的な解釈が容易
- 各ステップで Y_{obs} の対数尤度が減少しないことが保証される

欠点

- 欠測値の割合が多いと収束が遅い
- One Step estimator が漸近有効でない
(ニュートン・ラフソン法, スコア法では成り立つ)

8.3 EステップとMステップ

Eステップ

$\theta = \theta^{(t)}$ を用いて, “ Y_{mis} ” を Y_{obs} を与えた条件付期待値で「置き換える」

$$Q(\theta|\theta^{(t)}) = \int \ell(\theta|Y) f(Y_{\text{mis}}|Y_{\text{obs}}, \theta = \theta^{(t)}) dY_{\text{mis}}$$

Mステップ

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta|\theta^{(t)})$$

※ $\ell(\theta|Y_{\text{obs}}, Y_{\text{mis}})$ が Y_{mis} の線形関数ならば,

Eステップは Y_{mis} を $E[Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}]$ で置き換えることと同じ

例 8.1 1 変量正規分布

$$y_1, \dots, y_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2),$$

$$Y_{\text{obs}} = (y_1, \dots, y_r), \quad Y_{\text{mis}} = (y_{r+1}, \dots, y_n), \quad \theta = (\mu, \sigma^2)$$

$$\ell(\mu, \sigma^2 | Y) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{2\sigma^2}$$

$$\mathbb{E} \left[\sum_{i=1}^n y_i | Y_{\text{obs}}, \theta^{(t)} \right] = \sum_{i=1}^r y_i + (n-r)\mu^{(t)},$$

$$\mathbb{E} \left[\sum_{i=1}^n y_i^2 | Y_{\text{obs}}, \theta^{(t)} \right] = \sum_{i=1}^r y_i^2 + (n-r)[(\mu^{(t)})^2 + (\sigma^{(t)})^2]$$

$$\begin{aligned} Q(\theta | \theta^{(t)}) &= \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left[\sum_{i=1}^r y_i^2 + (n-r)\{(\mu^{(t)})^2 + (\sigma^{(t)})^2\} \right] \\ &\quad + \frac{\mu}{\sigma^2} \left[\sum_{i=1}^r y_i + (n-r)\mu^{(t)} \right] - \frac{n\mu^2}{2\sigma^2} \end{aligned}$$

$$\mu^{(t+1)} = \frac{1}{n} \left[\sum_{i=1}^r y_i + (n-r)\mu^{(t)} \right],$$

$$(\sigma^{(t+1)})^2 = \frac{1}{n} \left[\sum_{i=1}^r y_i^2 + (n-r)[(\mu^{(t)})^2 + (\sigma^{(t)})^2] \right] - (\mu^{(t+1)})^2$$

$(\mu^{(t)}, \sigma^{(t)}) \rightarrow (\hat{\mu}, \hat{\sigma})$ として、極限值を代入すると

$$\hat{\mu} = \frac{1}{r} \sum_{i=1}^r y_i,$$

$$\hat{\sigma}^2 = \frac{1}{r} \sum_{i=1}^r y_i^2 - \hat{\mu}^2$$

例 8.2 多項分布

$$Y = (y_1, y_2, y_3, y_4) \sim \text{Mt}_4\left[n, \left(\frac{1-\theta}{2}, \frac{\theta}{4}, \frac{\theta}{4}, \frac{1}{2}\right)\right],$$

$$Y_{\text{obs}} = (y_1, y_2, y_3 + y_4) = (38, 34, 125)$$

$$\ell(\theta|y_1, y_2, y_3, y_4) = y_1 \log \frac{1-\theta}{2} + y_2 \log \frac{\theta}{4} + y_3 \log \frac{\theta}{4} + y_4 \log \frac{1}{2}$$

$$\hat{\theta}_{\text{complete}} = \frac{y_2 + y_3}{y_1 + y_2 + y_3}$$

$$\text{E}[y_1|\theta, Y_{\text{obs}}] = 38, \quad \text{E}[y_2|\theta, Y_{\text{obs}}] = 34,$$

$$\text{E}[y_3|\theta, Y_{\text{obs}}] = 125 \cdot \frac{\theta/4}{\theta/4 + 1/2},$$

$$\text{E}[y_4|\theta, Y_{\text{obs}}] = 125 \cdot \frac{1/2}{\theta/4 + 1/2}$$

$$y_3^{(t)} = 125 \cdot \frac{\theta^{(t)}/4}{\theta^{(t)}/4 + 1/2},$$
$$\theta^{(t+1)} = \frac{y_2 + y_3^{(t)}}{y_1 + y_2 + y_3^{(t)}} = \frac{34 + y_3^{(t)}}{72 + y_3^{(t)}}$$

$\theta^{(t)} \rightarrow \hat{\theta}$ として, 代入すると $\hat{\theta}$ に関する 2 次方程式を得る.

$$\frac{\partial}{\partial \theta} \ell(\theta | Y_{\text{obs}}) = \frac{\partial}{\partial \theta} \left\{ y_1 \log \frac{1 - \theta}{2} + y_2 \log \frac{\theta}{4} + (y_3 + y_4) \log \frac{\theta + 2}{4} \right\}$$

と同じ解を持つ.

例 8.3 2 変量正規分布 $N_2(\boldsymbol{\mu}, \Sigma)$

$$Y_{\text{obs}} = \{(y_{i1}, y_{i2}); i = 1, \dots, r\} \\ \cup \{y_{i1}; i = r + 1, \dots, r + q\} \cup \{y_{i2}; i = r + q + 1, \dots, n\}$$

$\ell(\theta|Y)$: $s_1, s_2, s_{11}, s_{12}, s_{22}$ の一次式

$$s_1 = \sum_{i=1}^n y_{i1}, \quad s_2 = \sum_{i=1}^n y_{i2},$$

$$s_{11} = \sum_{i=1}^n y_{i1}^2, \quad s_{22} = \sum_{i=1}^n y_{i2}^2, \quad s_{12} = \sum_{i=1}^n y_{i1}y_{i2}$$

$$E[y_{i2}|y_{i1}, \boldsymbol{\mu}, \Sigma] = \beta_{20\cdot1} + \beta_{21\cdot1}y_{i1},$$

$$E[y_{i2}^2|y_{i1}, \boldsymbol{\mu}, \Sigma] = (\beta_{20\cdot1} + \beta_{21\cdot1}y_{i1})^2 + \sigma_{22\cdot1},$$

$$E[y_{i2}y_{i1}|y_{i1}, \boldsymbol{\mu}, \Sigma] = (\beta_{20\cdot1} + \beta_{21\cdot1}y_{i1})y_{i1},$$

E ステップ

$$s_1^{(t)} = \sum_{i=1}^{r+q} y_{i1} + \sum_{i=r+q+1}^n (\beta_{10.2}^{(t)} + \beta_{11.2}^{(t)} y_{i2}),$$

$$s_2^{(t)} = \sum_{i=1}^r y_{i2} + \sum_{i=r+q+1}^n y_{i2} + \sum_{i=r+1}^{r+q} (\beta_{20.1}^{(t)} + \beta_{21.1}^{(t)} y_{i1}),$$

$$s_{11}^{(t)} = \sum_{i=1}^{r+q} y_{i1}^2 + \sum_{i=r+q+1}^n \{(\beta_{10.2}^{(t)} + \beta_{11.2}^{(t)} y_{i2})^2 + \sigma_{11.2}^{(t)}\}$$

$$s_{12}^{(t)} = \sum_{i=1}^r y_{i1} y_{i2} + \sum_{i=r+1}^{r+q} (\beta_{20.1}^{(t)} + \beta_{21.2}^{(t)} y_{i1}) y_{i1} + \sum_{i=r+q}^n (\beta_{10.2}^{(t)} + \beta_{11.2}^{(t)} y_{i2}) y_{i2}$$

$$s_{22}^{(t)} = \sum_{i=1}^r y_{i2}^2 + \sum_{i=r+q+1}^n y_{i2}^2 + \sum_{i=r+1}^{r+q} \{(\beta_{20.1}^{(t)} + \beta_{21.1}^{(t)} y_{i1})^2 + \sigma_{22.1}^{(t)}\}$$

$$\hat{\mu}_1 = \frac{s_1}{n}, \quad \hat{\mu}_2 = \frac{s_2}{n},$$

$$\hat{\sigma}_1^2 = \frac{s_{11}}{n} - \hat{\mu}_1^2, \quad \hat{\sigma}_2^2 = \frac{s_{22}}{n} - \hat{\mu}_2^2, \quad \hat{\sigma}_{12} = \frac{s_{12}}{n} - \hat{\mu}_1 \hat{\mu}_2$$

M ステップ

$$\mu_1^{(t+1)} = \frac{s_1^{(t)}}{n}, \quad \mu_2^{(t+1)} = \frac{s_2^{(t)}}{n},$$

$$(\sigma_1^{(t+1)})^2 = \frac{s_{11}^{(t)}}{n} - (\mu_1^{(t+1)})^2, \quad (\sigma_2^{(t+1)})^2 = \frac{s_{22}^{(t)}}{n} - (\mu_2^{(t+1)})^2,$$

$$\sigma_{12}^{(t+1)} = \frac{s_{12}^{(t)}}{n} - \mu_1^{(t+1)} \mu_2^{(t+1)}$$

$$\Rightarrow \beta_{20.1}^{(t+1)}, \beta_{21.1}^{(t+1)}, \sigma_{22.1}^{(t+1)}, \dots$$

8.4 EM アルゴリズムの理論

EM アルゴリズムの収束

$$f(Y|\theta) = f(Y_{\text{obs}}, Y_{\text{mis}}|\theta) = f(Y_{\text{obs}}|\theta)f(Y_{\text{mis}}|Y_{\text{obs}}, \theta),$$

$$\ell(\theta|Y) = \ell(\theta|Y_{\text{obs}}, Y_{\text{mis}}) = \ell(\theta|Y_{\text{obs}}) + \log f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)$$

$$\ell(\theta|Y_{\text{obs}}) = \ell(\theta|Y) - \log f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)$$

$$= Q(\theta|\theta^{(t)}) - H(\theta|\theta^{(t)}),$$

$$Q(\theta|\theta^{(t)}) = \int \ell(\theta|Y)f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)})dY_{\text{mis}},$$

$$H(\theta|\theta^{(t)}) = \int \log f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)})dY_{\text{mis}}$$

Generalized EM (GEM) アルゴリズム

$\theta^{(t+1)} = M(\theta^{(t)})$ が次を満たすように定められているとする

$$Q(\theta^{(t+1)}|\theta^{(t)}) \geq Q(\theta^{(t)}|\theta^{(t)})$$

$$\begin{aligned} & \ell(\theta^{(t+1)}|Y_{\text{obs}}) - \ell(\theta^{(t)}|Y_{\text{obs}}) \\ &= \{Q(\theta^{(t+1)}|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})\} - \{H(\theta^{(t+1)}|\theta^{(t)}) - H(\theta^{(t)}|\theta^{(t)})\} \geq 0 \end{aligned}$$

∴

$$\begin{aligned} H(\theta^{(t)}|\theta^{(t)}) - H(\theta|\theta^{(t)}) &= \int \log \frac{f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)})}{f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)} f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}) dY_{\text{mis}} \\ &= \text{KL}[f(*|Y_{\text{obs}}, \theta^{(t)}), f(*|Y_{\text{obs}}, \theta)] \geq 0 \end{aligned}$$

定理 8.1. GEM アルゴリズムでは

$$\ell(\theta^{(t+1)}|Y_{\text{obs}}) \geq \ell(\theta^{(t)}|Y_{\text{obs}})$$

であり, 等号は

$$Q(\theta^{(t+1)}|\theta^{(t)}) = Q(\theta^{(t)}|\theta^{(t)})$$

のときにのみ成り立つ

系 1. $\forall \theta \ell(\theta^*|Y_{\text{obs}}) \geq \ell(\theta|Y_{\text{obs}})$ ならば

$$\ell(M(\theta^*)|Y_{\text{obs}}) = \ell(\theta^*|Y_{\text{obs}}),$$

$$Q(M(\theta^*)|\theta^*) = Q(\theta^*|\theta^*),$$

$$f(Y_{\text{mis}}|Y_{\text{obs}}, M(\theta^*)) = f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^*)$$

系 2. $\forall \theta \ell(\theta^*|Y_{\text{obs}}) \geq \ell(\theta|Y_{\text{obs}})$ ならば

$$M(\theta^*) = \theta^*$$

定理 8.2. EM アルゴリズムにおいて

(a) $D^{10}Q(\theta^{(t+1)}|\theta^{(t)}) = 0$

ただし

$$D^{10}Q(\theta^{(t+1)}|\theta^{(t)}) = \frac{\partial}{\partial \theta} Q(\theta|\theta^{(t)})|_{\theta=\theta^{(t+1)}}$$

(b) $\theta^{(t)} \rightarrow \theta^*$

(c) $f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)$ は θ の十分滑らかな関数

が成り立つならば

$$D_{\ell}(\theta^*|Y_{\text{obs}}) = \frac{\partial}{\partial \theta} \ell(\theta|Y_{\text{obs}})|_{\theta=\theta^*} = 0$$

証明

$$\begin{aligned} D_\ell(\theta^{(t+1)}|Y_{\text{obs}}) &= D^{01}Q(\theta^{(t+1)}|\theta^{(t)}) - D^{01}H(\theta^{(t+1)}|\theta^{(t)}) \\ &= -D^{01}H(\theta^{(t+1)}|\theta^{(t)}) \\ &= -\frac{\partial}{\partial\theta} \int \log f(Y_{\text{mis}}|Y_{\text{obs}}, \theta) f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}) dY_{\text{mis}} \Big|_{\theta=\theta^{(t+1)}} \end{aligned}$$

$t \rightarrow \infty$ とし, 微分と積分の交換をすると ((c))

$$D_\ell(\theta^*|Y_{\text{obs}}) = - \int \frac{\partial}{\partial\theta} f(Y_{\text{mis}}|Y_{\text{obs}}, \theta) dY_{\text{mis}} \Big|_{\theta=\theta^*} = 0$$

その他の収束定理

1. $\ell(\theta|Y_{\text{obs}})$ が有界ならば $\ell(\theta^{(t)}|Y_{\text{obs}})$ は収束する
2. $f(Y|\theta)$ が (曲) 指数型分布族で, $\ell(\theta|Y_{\text{obs}})$ が有界ならば $\ell(\theta^{(t)}|Y_{\text{obs}})$ は停留値に収束する
3. $f(Y|\theta)$ が正則な指数型分布族で, $\ell(\theta|Y_{\text{obs}})$ が有界ならば $\theta^{(t)}$ は停留点 θ^* に収束する

指数型分布に対する EM アルゴリズム

$$f(Y|\theta) = b(Y) \exp[s(Y)\theta - a(\theta)],$$

$s(Y) : 1 \times d$ ベクトル, $\theta : d \times 1$ ベクトル,

$a(\theta), b(Y) : \text{既知の関数}$

E ステップ

$$\ell(\theta|Y) = s(Y)\theta - a(\theta)$$

$$Q(\theta|\theta^{(t)}) = \mathbb{E}[\ell(\theta|Y)|Y_{\text{obs}}, \theta^{(t)}] = s^{(t+1)}\theta - a(\theta),$$

$$s^{(t+1)} = \mathbb{E}[s(Y)|Y_{\text{obs}}, \theta^{(t)}]$$

M ステップ

$$\frac{\partial}{\partial \theta} Q(\theta|\theta^{(t)}) = 0 \quad \Leftrightarrow \quad s^{(t+1)} = \frac{\partial a(\theta)}{\partial \theta^T} = \mathbb{E}[s(Y)|\theta]$$

例 8.4 t 分布に対する最尤推定量

$$X = (x_1, \dots, x_n), \quad x_1, \dots, x_n \stackrel{i.i.d.}{\sim} f(x|\theta), \quad \theta = (\mu, \sigma^2)$$

$$f(x|\theta) = \frac{\Gamma[\frac{\nu+1}{2}]}{\sqrt{\pi\nu\sigma^2}\Gamma[\frac{\nu}{2}]} \frac{1}{\{1 + \frac{(x-\mu)^2}{\nu\sigma^2}\}^{(\nu+1)/2}}$$

X を欠損データの観測部分とみなす：

$$Y = (Y_{\text{obs}}, Y_{\text{mis}}),$$

$$Y_{\text{obs}} = X, \quad Y_{\text{mis}} = W = (w_1, \dots, w_n)$$

$$(x_1, w_1), \dots, (x_n, w_n) : \text{独立}, \quad w_i \sim \frac{1}{\nu}\chi_\nu^2, \quad x_i|w_i \sim N\left(\mu, \frac{\sigma^2}{w_i}\right)$$

(x_i, w_i) の同時確率密度関数 :

$$\begin{aligned} f(x, w|\theta) &= \frac{\nu}{2\Gamma[\frac{\nu}{2}]} \exp(-\nu w/2) \left(\frac{\nu w}{2}\right)^{\nu/2-1} \frac{\sqrt{w}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{w}{2\sigma^2}(x-\mu)^2\right\} \\ &= \frac{\nu}{2\sqrt{\pi\nu\sigma^2}\Gamma[\frac{\nu}{2}]} \exp\left\{-\frac{\nu w}{2}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)\right\} \left(\frac{\nu w}{2}\right)^{(\nu+1)/2-1} \end{aligned}$$

$$f(x|\theta) = \int_0^\infty f(x, w|\theta) dw \quad : t \text{ 分布}$$

$$\ell(\theta|Y) = \sum_{i=1}^n \log f(x_i, w_i|\theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n w_i (x_i - \mu)^2 - \frac{n}{2} \log \sigma^2$$

$$= -\frac{\mu^2}{2\sigma^2} s_0 + \frac{\mu}{\sigma^2} s_1 - \frac{1}{2\sigma^2} s_2 - \frac{n}{2} \log \sigma^2,$$

$$s_0 = \sum_{i=1}^n w_i, \quad s_1 = \sum_{i=1}^n w_i x_i, \quad s_2 = \sum_{i=1}^n w_i x_i^2$$

E ステップ

$$\mathbb{E}[\ell(\theta|Y)|Y_{\text{obs}}, \theta^{(t)}] = -\frac{\mu^2}{2\sigma^2}s_0^{(t)} + \frac{\mu}{\sigma^2}s_1^{(t)} - \frac{1}{2\sigma^2}s_2^{(t)} - \frac{n}{2}\log \sigma^2,$$

$$s_0^{(t)} = \sum_{i=1}^n w_i^{(t)}, \quad s_1^{(t)} = \sum_{i=1}^n w_i^{(t)} x_i, \quad s_2^{(t)} = \sum_{i=1}^n w_i^{(t)} x_i^2,$$

$$w_i^{(t)} = \mathbb{E}[w_i|x_i, \theta^{(t)}] = \frac{\nu + 1}{\nu + (d_i^{(t)})^2}, \quad d_i^{(t)} = \frac{x_i - \mu^{(t)}}{\sigma^{(t)}}$$

∴

$$\begin{aligned} f(w|x, \theta) &= \frac{f(x, w|\theta)}{f(x|\theta)} \\ &= \frac{\nu}{2\Gamma[\frac{\nu+1}{2}]} \left\{1 + \frac{(x - \mu)^2}{\nu\sigma^2}\right\}^{(\nu+1)/2} \exp\left\{-\frac{\nu w}{2} \left(1 + \frac{(x - \mu)^2}{\nu\sigma^2}\right)\right\} \left(\frac{\nu w}{2}\right)^{(\nu+1)/2-1} \end{aligned}$$

$$w_i|x_i \sim \left\{\nu + \frac{(x_i - \mu)^2}{\sigma^2}\right\}^{-1} \chi_{\nu+1}^2$$

M ステップ

$$-\frac{\mu^2}{2\sigma^2}s_0^{(t)} + \frac{\mu}{\sigma^2}s_1^{(t)} - \frac{1}{2\sigma^2}s_2^{(t)} - \frac{n}{2}\log\sigma^2$$

の最大化

$$\mu^{(t+1)} = \frac{s_1^{(t)}}{s_0^{(t)}} = \frac{\sum_{i=1}^n w_i^{(t)} x_i}{\sum_{i=1}^n w_i^{(t)}},$$

$$(\sigma^{(t+1)})^2 = \frac{s_2^{(t)} - (s_1^{(t)})^2/s_0^{(t)}}{n} = \frac{1}{n} \sum_{i=1}^n w_i^{(t)} (x_i - \mu^{(t+1)})^2$$

- t 分布の最尤推定値は EM アルゴリズムによって計算できる。
- $\hat{\mu}$ は、外れ値の影響を小さくした加重平均となっている：

$$\hat{\mu} = \frac{\sum_{i=1}^n \hat{w}_i x_i}{\sum_{i=1}^n \hat{w}_i}, \quad \hat{w}_i = \frac{\nu + 1}{\nu + \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^2}, \quad (\hat{\sigma})^2 = \frac{1}{n} \sum_{i=1}^n \hat{w}_i (x_i - \hat{\mu})^2$$

EM アルゴリズムの収束レート

情報量の分解

$$-\ell(\theta|Y_{\text{obs}}) = -\ell(\theta|Y) + \log f(Y_{\text{mis}}, |Y_{\text{obs}}\theta)$$

$$I(\theta|Y_{\text{obs}}) = I(\theta|Y_{\text{obs}}, Y_{\text{mis}}) + \frac{\partial^2}{\partial\theta\partial\theta^T} \log f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)$$

$$= \mathbb{E}[I(\theta|Y_{\text{obs}}, Y_{\text{mis}})|Y_{\text{obs}}] + \mathbb{E}\left[\frac{\partial^2}{\partial\theta\partial\theta^T} \log f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)|Y_{\text{obs}}\right]$$

$$= -D^{20}Q(\theta|\theta) + D^{20}H(\theta|\theta)$$

$$Q(\theta|\theta^{(t)}) = \mathbb{E}[\ell(\theta|Y_{\text{obs}}, Y_{\text{mis}})|Y_{\text{obs}}, \theta^{(t)}],$$

$$H(\theta|\theta^{(t)}) = \mathbb{E}[\log f(Y_{\text{mis}}|Y_{\text{obs}}, \theta)|Y_{\text{obs}}, \theta^{(t)}]$$

$$i_{\text{obs}} = i_{\text{com}} - i_{\text{mis}},$$

$$i_{\text{obs}} = I(\theta|Y_{\text{obs}})|_{\theta=\theta^*}, \quad i_{\text{com}} = -D^{20}Q(\theta|\theta)|_{\theta=\theta^*},$$

$$i_{\text{mis}} = -D^{20}H(\theta|\theta)|_{\theta=\theta^*},$$

指数型分布の場合

$$f(Y|\theta) = b(Y) \exp[s(Y)\theta - a(\theta)],$$

$$i_{\text{com}} = \text{Var}[s(Y)|\theta], \quad i_{\text{mis}} = \text{Var}[s(Y)|Y_{\text{obs}}, \theta]$$

$$\Rightarrow \quad I(\theta|Y_{\text{obs}}) = \text{Var}[s(Y)|\theta] - \text{Var}[s(Y)|Y_{\text{obs}}, \theta]$$

収束レート

$$DM = i_{\text{mis}} i_{\text{com}}^{-1} = I - i_{\text{obs}} I_{\text{mis}}^{-1} : \text{欠損による情報量の損失率}$$

$$|\theta^{(t+1)} - \theta^*| = \lambda |\theta^{(t)} - \theta^*|, \quad \lambda : DM \text{ の最大固有値}$$

(Dempster, Laird and Rubin, 1977)

収束レートの推定 (Louis 1982)

$$\begin{aligned} -D^{20} H(\theta|\theta) &= \text{E}[D_{\ell}(\theta|Y_{\text{obs}}, Y_{\text{mis}}) D_{\ell}(\theta|Y_{\text{obs}}, Y_{\text{mis}})^{\text{T}} | Y_{\text{obs}}, \theta] \\ &\quad - D_{\ell}(\theta|Y_{\text{obs}}) D_{\ell}(\theta|Y_{\text{obs}})^{\text{T}} \end{aligned}$$

$$I(\hat{\theta}|Y_{\text{obs}}) = -D^{20} Q(\hat{\theta}|\hat{\theta}) - \text{E}[D_{\ell}(\theta|Y_{\text{obs}}, Y_{\text{mis}}) D_{\ell}(\theta|Y_{\text{obs}}, Y_{\text{mis}})^{\text{T}} | Y_{\text{obs}}, \theta] |_{\theta=\hat{\theta}}$$

例 8.5 多項分布 (例 8.2 の続き)

$$Y = (y_1, y_2, y_3, y_4) \sim \text{Mt}_4 \left[n, \left(\frac{1-\theta}{2}, \frac{\theta}{4}, \frac{\theta}{4}, \frac{1}{2} \right) \right],$$

$$\ell(\theta|Y) = y_1 \log(1-\theta) + (y_2 + y_3) \log \theta + \text{定数}$$

$$D_\ell(\theta|Y) = -\frac{y_1}{1-\theta} + \frac{y_2 + y_3}{\theta} \quad I(\theta|Y) = \frac{y_1}{(1-\theta)^2} + \frac{(y_2 + y_3)}{\theta^2}$$

$$E[I(\theta|Y)|Y_{\text{obs}}, \theta] = \frac{y_1}{(1-\theta)^2} + \frac{(y_2 + \hat{y}_3)}{\theta^2},$$

$$\hat{y}_3 = E[y_3|Y_{\text{obs}}, \theta] = (y_2 + y_3) \frac{\theta/4}{\theta/4 + 1/2}$$

$$E[D_\ell^2(\theta|Y)|Y_{\text{obs}}, \theta] = \frac{V}{\theta^2},$$

$$V = \text{Var}[y_3|Y_{\text{obs}}, \theta] = (y_3 + y_4) \frac{\theta/4}{\theta/4 + 1/2} \frac{1/2}{\theta/4 + 1/2}$$

$$Y_{\text{obs}} = (y_1, y_2, y_3 + y_4) = (38, 34, 125),$$

$$\hat{\theta} = 0.6268,$$

$$E[I(\theta|Y)|Y_{\text{obs}}, \theta]|_{\theta=\hat{\theta}} = 435.3, \quad E[D_\ell^2(\theta|Y)|Y_{\text{obs}}, \theta]|_{\theta=\hat{\theta}} = 57.8$$

$$I(\hat{\theta}|Y_{\text{obs}}) = 435.3 - 57.8 = 377.5$$

$$DM = \frac{57.8}{435.3} = 0.1328$$

t	$\theta^{(t)}$	$\theta^{(t)} - \hat{\theta}$	$(\theta^{(t+1)} - \hat{\theta}) / (t) - \hat{\theta}$
0	0.5000	0.1268	0.1465
1	0.6082	0.0186	0.1346
2	0.6243	0.002500	0.1330
3	0.6265	0.0003326	0.1328
4	0.6268	0.000005866	0.1328

8.5 EM アルゴリズムの拡張

ECM アルゴリズム

GEM アルゴリズム $\theta^{(t+1)} = M(\theta^{(t)})$:

$$Q(\theta^{(t+1)}|\theta^{(t)}) \geq Q(\theta^{(t)}|\theta^{(t)}) \quad \Rightarrow \quad \ell(\theta^{(t+1)}|Y_{\text{obs}}) \geq \ell(\theta^{(t)}|Y_{\text{obs}})$$

CM ステップ (conditional maximization step)

$g_s(\theta) = g_s(\theta^{(t)})$ を固定して, $Q(\theta|\theta^{(t)})$ を最大化

例. $\theta = (\theta_1, \dots, \theta_S)$ とし, $s = 1, \dots, S$ に対して $\theta_s = \theta_s^{(t)}$ を固定して,
残りの θ_j ($j \neq s$) について $Q(\theta|\theta^{(t)})$ を最大化,
あるいは, $\theta_j = \theta_j^{(t)}$ ($j \neq s$) を固定して, θ_s について $Q(\theta|\theta^{(t)})$ を最大化.

例 8.6 多変量正規線形回帰モデル

$$Y_i \sim N_K(X_i\beta, \Sigma), \quad i = 1, \dots, n \text{ 独立}, \quad \theta = (\beta, \Sigma)$$

$$\ell(\theta|Y) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n (Y_i - X_i\beta)^T \Sigma^{-1} (Y_i - X_i\beta)$$

スコア関数

$$\frac{\partial \ell(\theta|Y)}{\partial \beta} = \sum_{i=1}^n \left\{ X_i^T \Sigma^{-1} Y_i - X_i^T \Sigma^{-1} X_i \beta \right\}$$

$$\frac{\partial \ell(\theta|Y)}{\partial (\Sigma^{-1})} = \frac{n}{2} \Sigma - \frac{1}{2} \sum_{i=1}^n (Y_i - X_i\beta)(Y_i - X_i\beta)^T$$

完全データの場合

$\Sigma = \Sigma^{(t)}$ を固定して β に関して最大化すると

$$\beta^{(t+1)} = \left\{ \sum_{i=1}^n X_i^T (\Sigma^{(t)})^{-1} X_i \right\}^{-1} \left\{ \sum_{i=1}^n X_i^T (\Sigma^{(t)})^{-1} Y_i \right\}$$

$\beta = \beta^{(t+1)}$ を固定して Σ に関して最大化すると

$$\Sigma^{(t+1)} = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i \beta^{(t+1)})(Y_i - X_i \beta^{(t+1)})^T$$

CM1: $\ell(\beta^{(t+1)}, \Sigma^{(t)} | Y) \geq \ell(\beta^{(t)}, \Sigma^{(t)} | Y)$

CM2: $\ell(\beta^{(t+1)}, \Sigma^{(t+1)} | Y) \geq \ell(\beta^{(t+1)}, \Sigma^{(t)} | Y)$

欠損データの場合

E ステップ

$$\ell(\theta|Y) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n \{Y_i^T \Sigma^{-1} Y_i - 2\beta^T X_i^T \Sigma^{-1} Y_i + \beta^T X_i^T \Sigma^{-1} X_i \beta\}$$

$$Q(\theta|\theta^{(t)}) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n \left[\text{tr}(S_i^{(t)} \Sigma^{-1}) - 2\beta^T X_i^T \Sigma^{-1} Y_i^{(t)} + \beta^T X_i^T \Sigma^{-1} X_i \beta \right]$$

$$S_i^{(t)} = \text{E}[Y_i Y_i^T | Y_{\text{obs}}, \theta^{(t)}], \quad Y_i^{(t)} = \text{E}[Y_i | Y_{\text{obs}}, \theta^{(t)}]$$

$$Y_i \text{ が欠測} \Rightarrow S_i^{(t)} = \Sigma^{(t)} + X_i \beta^{(t)} (X_i \beta^{(t)})^T, \quad Y_i^{(t)} = X_i \beta^{(t)}$$

CM1 ステップ

$$\beta^{(t+1)} = \left\{ \sum_{i=1}^n X_i^T (\Sigma^{(t)})^{-1} X_i \right\}^{-1} \left\{ \sum_{i=1}^n X_i^T (\Sigma^{(t)})^{-1} Y_i^{(t)} \right\}$$

CM2 ステップ

$$\Sigma^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \left[S_i^{(t)} - Y_i^{(t)} (\beta^{(t+1)})^T X_i^T - (Y_i^{(t)} (\beta^{(t+1)})^T X_i^T)^T \right. \\ \left. + X_i \beta^{(t+1)} (\beta^{(t+1)})^T X_i^T \right]$$

$$Y_{\text{obs}} = (Y_1, \dots, Y_r), \quad Y_{\text{mis}} = (Y_{r+1}, \dots, Y_n)$$

$$\Rightarrow \quad \Sigma^{(t+1)} = \frac{1}{n} \sum_{i=1}^r (Y_i - X_i \beta^{(t+1)}) (Y_i - X_i \beta^{(t+1)})^T + \frac{n-r}{n} \Sigma^{(t)} \\ + \frac{1}{n} \sum_{i=r+1}^n X_i (\beta^{(t)} - \beta^{(t+1)}) (\beta^{(t)} - \beta^{(t+1)})^T X_i^T$$

例 8.7 自由度が未知の t 分布 (例 8.4 の続き)

$$X = (x_1, \dots, x_n), \quad x_1, \dots, x_n \stackrel{i.i.d.}{\sim} f(x|\theta), \quad \theta = (\mu, \sigma^2, \nu)$$

$$f(x|\theta) = \frac{\Gamma[\frac{\nu+1}{2}]}{\sqrt{\pi\nu\sigma^2}\Gamma[\frac{\nu}{2}]} \frac{1}{\{1 + \frac{(x-\mu)^2}{\nu\sigma^2}\}^{(\nu+1)/2}}$$

E ステップ

$$\mathbb{E}[\ell(\theta|Y)|Y_{\text{obs}}, \theta^{(t)}] = -\frac{\mu^2}{2\sigma^2} s_0^{(t)} + \frac{\mu}{\sigma^2} s_1^{(t)} - \frac{1}{2\sigma^2} s_2^{(t)} - \frac{n}{2} \log \sigma^2,$$

$$s_0^{(t)} = \sum_{i=1}^n w_i^{(t)}, \quad s_1^{(t)} = \sum_{i=1}^n w_i^{(t)} x_i, \quad s_2^{(t)} = \sum_{i=1}^n w_i^{(t)} x_i^2,$$

$$w_i^{(t)} = \mathbb{E}[w_i|x_i, \theta^{(t)}] = \frac{\nu^{(t)} + 1}{\nu^{(t)} + (d_i^{(t)})^2}, \quad d_i^{(t)} = \frac{x_i - \mu^{(t)}}{\sigma^{(t)}}$$

CM1 ステップ

$\nu = \nu^{(t)}$ を固定して, (μ, σ^2) について最大化 (例 8.4 と同じ)

CM2 ステップ

$(\mu, \sigma^2) = (\mu^{(t+1)}, (\sigma^{(t+1)})^2)$ を固定して $Q(\mu, \sigma^2, \nu | \mu^{(t+1)}, (\sigma^{(t+1)})^2, \nu^{(t)})$ を ν について最大化

$$\begin{aligned} \ell(\mu, \sigma^2, \nu | Y) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^n w_i \frac{(x_i - \mu)^2}{\sigma^2} \\ &\quad + \frac{n\nu}{2} \log \frac{\nu}{2} - n \log \Gamma\left[\frac{\nu}{2}\right] + \left(\frac{\nu}{2} - 1\right) \sum_{i=1}^n \log w_i - \frac{\nu}{2} \sum_{i=1}^n w_i \\ Q(\mu, \sigma^2, \nu | \mu^{(t)}, (\sigma^{(t)})^2, \nu^{(t)}) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^n w_i^{(t)} \frac{(x_i - \mu)^2}{\sigma^2} \\ &\quad + \frac{n\nu}{2} \log \frac{\nu}{2} - n \log \Gamma\left[\frac{\nu}{2}\right] + \left(\frac{\nu}{2} - 1\right) \sum_{i=1}^n z_i^{(t)} - \frac{\nu}{2} \sum_{i=1}^n w_i^{(t)}, \\ z_i^{(t)} &= E(\log w_i | x_i, \mu^{(t)}, (\sigma^{(t)})^2, \nu^{(t)}) \end{aligned}$$

ECME アルゴリズム (Expectation/Conditional Maximization Either)

例 8.8 自由度が未知の t 分布 (例 8.7 の続き)

$$X = (x_1, \dots, x_n), \quad x_1, \dots, x_n \stackrel{i.i.d.}{\sim} f(x|\theta), \quad \theta = (\mu, \sigma^2, \nu)$$

$$f(x|\mu, \sigma^2, \nu) = \frac{\Gamma[\frac{\nu+1}{2}]}{\sqrt{\pi\nu\sigma^2}\Gamma[\frac{\nu}{2}]} \frac{1}{\left\{1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right\}^{(\nu+1)/2}}$$

E ステップと CM1 ステップは ECM アルゴリズムと同じ

ECM2 ステップ

$$\begin{aligned} & Q(\mu, \sigma^2, \nu | \mu^{(t)}, (\sigma^{(t)})^2, \nu^{(t)}) \\ &= \int \ell(\mu^{(t+1)}, (\sigma^{(t+1)})^2, \nu | Y) f(Y_{\text{mis}} | Y_{\text{obs}}, \mu^{(t+1)}, (\sigma^{(t+1)})^2, \nu^{(t)}) dY_{\text{mis}} \end{aligned}$$

の代わりに次を最大化する

$$\sum_{i=1}^n \log f(x_i | \mu^{(t+1)}, (\sigma^{(t+1)})^2, \nu) = \int \ell(\mu^{(t+1)}, (\sigma^{(t+1)})^2, \nu | Y) f(Y_{\text{mis}} | \nu^{(t)}) dY_{\text{mis}}$$

PX-EM アルゴリズム (Parameter-Expanded EM アルゴリズム)

$f(Y|\theta)$: 元の確率密度関数

パラメータ α を追加してモデルを拡張する

$$\{f(Y|\theta)|\theta\} \subset \{f_{\text{ex}}(Y|\theta^*, \alpha)|\theta^*, \alpha\},$$

$$f_{\text{ex}}(Y|\theta^*, \alpha_0) = f(Y|\theta)$$

とする. ただし, 既知の関数 $\theta = R(\theta^*, \alpha)$ で

$$\alpha = \alpha_0 \quad \Rightarrow \quad R(\theta^*, \alpha_0) = \theta^*$$

を満たすとする. また,

$$f_{\text{ex}}(Y_{\text{obs}}|\theta^*, \alpha) = f_{\text{ex}}(Y_{\text{obs}}|\theta^*, \alpha') \text{ for all } \alpha, \alpha'$$

拡張モデルにおいて, $\phi = (\theta^*, \alpha)$ は識別可能とする.

PX-EM アルゴリズムは, $f_{\text{ex}}(Y|\theta^*, \alpha)$ に対して EM アルゴリズムを用いる

PX-E ステップ

$$Q(\phi|\phi^{(t)}) = \text{E}[\log f_{\text{ex}}(Y|\phi)|Y_{\text{obs}}, \phi^{(t)}]$$

PX-M ステップ

$$\phi^{(t+1)} = \underset{\phi}{\text{argmax}} Q(\phi|\phi^{(t)}), \quad \theta^{(t+1)} = R(\theta^{*(t+1)}, \alpha^{(t+1)})$$

例 8.9 自由度が既知の t 分布 (例 8.4 の続き)

$$(x_1, w_1), \dots, (x_n, w_n) : \text{独立}, \quad w_i \sim \frac{\alpha}{\nu} \chi_{\nu}^2, \quad x_i|w_i \sim N(\mu_*, \frac{\sigma_*^2}{w_i})$$

$\alpha = 1 \Rightarrow$ 例 8.4 と同じ,

(x_i, w_i) の同時確率密度関数 :

$$\theta^* = (\mu_*, \sigma_*), \quad \theta = R(\theta^*, \alpha) = (\mu_*, \sigma_* / \sqrt{\alpha})$$

$$\begin{aligned} f_{\text{ex}}(x, w | \theta^*, \alpha) &= \frac{\nu}{2\alpha \Gamma[\frac{\nu}{2}]} \exp\left(-\frac{\nu w}{2\alpha}\right) \left(\frac{\nu w}{2\alpha}\right)^{\nu/2-1} \frac{\sqrt{w}}{\sqrt{2\pi\sigma_*^2}} \exp\left\{-\frac{w}{2\sigma_*^2}(x - \mu_*)^2\right\} \\ &= \frac{\nu}{2\alpha \sqrt{\pi\nu\sigma_*^2} \Gamma[\frac{\nu}{2}]} \exp\left\{-\frac{\nu w}{2\alpha} \left(1 + \frac{\alpha(x - \mu_*)^2}{\nu\sigma_*^2}\right)\right\} \left(\frac{\nu w}{2\alpha}\right)^{(\nu+1)/2-1} \end{aligned}$$

$$f_{\text{ex}}(x | \theta^*, \alpha) = \int_0^\infty f_{\text{ex}}(x, w | \theta^*, \alpha) dw = f(x | \theta) \quad t \text{ 分布}$$

$$\ell_{\text{ex}}(\theta^*, \alpha | Y) = -\frac{n\nu}{2} \log \alpha - \frac{\nu}{2\alpha} \sum_{i=1}^n w_i \left(1 + \frac{\alpha(x_i - \mu_*)^2}{\nu\sigma_*^2}\right) - \frac{n}{2} \log \sigma_*^2$$

PX-E ステップ

$$w_i^{(t)} = \mathbb{E}[w_i | x_i, \phi^{(t)}] = \alpha^{(t)} \frac{\nu + 1}{\nu + d_i^{(t)} 2},$$

$$d_i^{(t)} = \sqrt{\alpha^{(t)}} \frac{x_i - \mu_*^{(t)}}{\sigma_*^{(t)}} = \frac{x_i - \mu^{(t)}}{\sigma^{(t)}} \quad (\mu^{(t)}, \sigma^{(t)}) = R(\theta^{*(t)}, \alpha^{(t)}),$$

$$\mathbb{E}[\ell_{\text{ex}}(\theta^*, \alpha | Y) | Y_{\text{obs}}, \phi^{(t)}]$$

$$= -\frac{n\nu}{2} \log \alpha - \frac{\nu}{2\alpha} \sum_{i=1}^n w_i^{(t)} \left(1 + \frac{\alpha(x_i - \mu_*)^2}{\nu\sigma_*^2} \right) - \frac{n}{2} \log \sigma_*^2$$

PX-M ステップ

$$\mu_*^{(t+1)} = \frac{\sum_{i=1}^n w_i^{(t)} x_i}{\sum_{i=1}^n w_i^{(t)}}, \quad \sigma_*^{(t+1)2} = \frac{1}{n} \sum_{i=1}^n w_i^{(t)} (x_i - \mu_*^{(t+1)})^2,$$

$$\alpha^{(t+1)} = \frac{1}{n} \sum_{i=1}^n w_i^{(t)}$$

$$(\mu^{(t+1)}, \sigma^{(t+1)}) = R(\theta^{*(t+1)}, \alpha^{(t+1)})$$

$$\Rightarrow \sigma^{(t+1)2} = \frac{\sum_{i=1}^n w_i^{(t)} (x_i - \mu_*^{(t+1)})^2}{\sum_{i=1}^n w_i^{(t)}}$$

$$\left(\text{EM アルゴリズムでは} \sigma^{(t+1)2} = \frac{\sum_{i=1}^n w_i^{(t)} (x_i - \mu_*^{(t+1)})^2}{n} \right)$$

収束レート

$$f_{\text{ex}}(Y_{\text{obs}}|\theta^*, \alpha) = f_{\text{ex}}(Y_{\text{obs}}|\theta^*, \alpha') \text{ for all } \alpha, \alpha'$$

より

$$I_{\text{obs}} = \begin{pmatrix} i_{\text{obs}} & 0 \\ 0 & 0 \end{pmatrix}, \quad I_{\text{com}} = \begin{pmatrix} i_{\text{obs}} & i_{\theta\alpha}^{\text{T}} \\ i_{\theta\alpha} & i_{\alpha\alpha} \end{pmatrix},$$

収束レートは, $I - I_{\text{obs}}I_{\text{com}}^{-1}$ の $\theta - \theta$ ブロック :

$$DM_{\text{ex}}(\theta) = I - i_{\text{obs}}(i_{\text{com}} - i_{\theta\alpha}^{\text{T}}i_{\alpha\alpha}^{-1}i_{\theta\alpha})^{-1}$$

$$DM(\theta) = I - i_{\text{obs}}i_{\text{com}}^{-1}$$

$$\Rightarrow DM_{\text{ex}}(\theta) < DM(\theta)$$

9 欠測値を含む分割表の分析

データ行列： $n \times V$, n —個体数, V —カテゴリー変数の個数

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1V} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{nV} \end{pmatrix}$$

$Y = (Y_1, \dots, Y_V)$: V 次元カテゴリー変数

$\{n_{jk\dots t}\}$: V 次元分割表

$n_{jk\dots t}$: $Y_1 = j, Y_2 = k, \dots, Y_V = t$ となる個体数

例. 性別と年齢で $2 \times 5 = 10$ グループに分割された子供が肥満かどうかを 3 回調査.

Y_j ($j = 1, 2, 3$): j 回目の調査で肥満であるかどうかを表す 2 値変数

Y_4 : 性別を表す 2 値変数

Y_5 : 年齢グループ (5-7, 7-9, 9-11, 11-13, 13-15)

Y_4, Y_5 は欠測値なし

$$M_j = \begin{cases} 1 & Y_j \text{が欠測値} \\ 0 & Y_j \text{が欠測値でない} \end{cases}$$

欠損パターン:

$$(M_1, M_2, M_3) = \begin{cases} (0, 0, 0) & \text{完全データ} \\ (1, 0, 0), (0, 1, 0), (0, 0, 1) & \text{欠測が 1 回} \\ (1, 1, 0), (1, 0, 1), (0, 1, 1) & \text{欠測が 2 回} \end{cases}$$

$Y_1, Y_2, Y_3, Y_4 = 0 \text{ or } 1, Y_5 = 1, 2, 3, 4 \text{ or } 5$ と表す.

欠損パターンが $(1, 0, 0)$ の場合

$n_{0jklmt} + n_{1jklmt}$ は観測されるが, n_{0jklmt}, n_{1jklmt} は観測されない
 $\Rightarrow 2 \times 2 \times 2 \times 5$ 分割表にまとめられる.

完全データ : 5次元の分割表 ($2 \times 2 \times 2 \times 2 \times 5$ 分割表)

欠測が1回 : 4次元の分割表 ($2 \times 2 \times 2 \times 5$ 分割表)

欠測が2回 : 3次元の分割表 ($2 \times 2 \times 5$ 分割表)

9.1 単調欠損データに対する尤度の分解

完全データの場合

$\{n_{jk\dots t}\} \sim \text{Mt}(n, \theta)$ (繰り返し数 n の多項分布)

$$\theta = \{\pi_{jk\dots t}\}, \quad \pi_{jk\dots t} = P(Y_1 = j, Y_2 = k, \dots, Y_V = t)$$

尤度関数

$$L(\theta|\{n_{jk\dots t}\}) = \prod_{j,k,\dots,t} \pi_{jk\dots t}^{n_{jk\dots t}}, \quad \sum_{j,k,\dots,t} \pi_{jk\dots t} = 1$$

最尤推定量

$$\hat{\pi}_{jk\dots t} = \frac{n_{jk\dots t}}{n}$$

分散の推定量

$$\widehat{\text{Var}}(\pi_{jk\dots t} - \hat{\pi}_{jk\dots t}) = \frac{1}{n} \hat{\pi}_{jk\dots t} (1 - \hat{\pi}_{jk\dots t})$$

ベイズ推測

事前分布 (ディリクレ分布)

$$p(\{\pi_{jk\dots t}\}) \propto \prod_{j,k,\dots,t} \pi_{jk\dots t}^{\alpha_{jk\dots t}-1}, \quad \pi_{jk\dots t} > 0, \quad \sum_{j,k,\dots,t} \pi_{jk\dots t} = 1$$

事後分布

$$p(\{\pi_{jk\dots t}\} | \{n_{jk\dots t}\}) \propto \prod_{j,k,\dots,t} \pi_{jk\dots t}^{\alpha_{jk\dots t} + n_{jk\dots t} - 1}, \quad \sum_{j,k,\dots,t} \pi_{jk\dots t} = 1$$

単調欠損データ

例 9.1 2次元分割表

$$Y_1 = j \in \{1, 2, \dots, J\}, \quad Y_2 = k \in \{1, 2, \dots, K\}$$

$\{(y_{i1}, y_{i2}); i = 1, 2, \dots, r\}$: 完全データ

$\{y_{i1}; i = r + 1, \dots, n\}$: Y_1 のみ観測

<u>完全データ</u>					<u>不完全データ</u>	
	Y_2				レベル	
レベル	1	...	K	計		
	1	:			Y_1	:
Y_1	:	...	r_{jk}	...	J	m_J
	J	:			計	$m = n - r$
計		r_{+k}		r		

尤度の分割

$$\theta = (\pi_{11}, \pi_{12}, \dots, \pi_{JK}), \quad \sum_{j=1}^J \sum_{k=1}^K \pi_{jk} = \pi_{++} = 1$$

$$\phi_1 = \{\pi_{j+}; j = 1, \dots, J\}, \quad \pi_{j+} = P(Y_1 = j) = \sum_{k=1}^K \pi_{jk}$$

$$\phi_2 = \{\pi_{k \cdot j}; j = 1, \dots, J, k = 1, \dots, K\}, \quad \pi_{k \cdot j} = P(Y_2 = k | Y_1 = j) = \frac{\pi_{jk}}{\pi_{j+}}$$

$$\theta \Leftrightarrow \phi = (\phi_1, \phi_2)$$

$$L(\phi | \{r_{jk}\}, \{m_j\}) = \left(\prod_{j=1}^J \pi_{j+}^{r_{j+} + m_j} \right) \times \left(\prod_{j=1}^J \prod_{k=1}^K \pi_{k \cdot j}^{r_{jk}} \right)$$

$$\hat{\pi}_{j+} = \frac{r_{j+} + m_j}{n}, \quad \hat{\pi}_{k \cdot j} = \frac{r_{jk}}{r_{j+}}$$

$$\Rightarrow \hat{\pi}_{jk} = \hat{\pi}_{j+} \hat{\pi}_{k \cdot j} = \frac{r_{jk} + (r_{jk}/r_{j+})m_j}{n}$$

ベイズ法

ディリクレ事前分布

$$p(\phi) \propto \left(\prod_{j=1}^J \pi_{j+}^{n_{j0}-1} \right) \times \left(\prod_{j=1}^J \prod_{k=1}^K \pi_{k \cdot j}^{r_{jk0}-1} \right)$$

事後分布

$$p(\phi|\text{data}) \propto \left(\prod_{j=1}^J \pi_{j+}^{n_{j0}+r_{j+}+m_j-1} \right) \times \left(\prod_{j=1}^J \prod_{k=1}^K \pi_{k \cdot j}^{r_{jk0}+r_{jk}-1} \right)$$

例 9.2 数値例

完全データ					不完全データ			
					レベル			
		Y_2						
レベル		1	2	3	計	Y_1	1	100
Y_1	1	20	30	40	90		2	90
	2	50	60	20	130			
	計	70	90	60	220	計	190	

$$\hat{\pi}_{1+} = \frac{190}{410}, \quad \hat{\pi}_{2+} = \frac{220}{410}$$

$$\hat{\pi}_{1.1} = \frac{20}{90}, \quad \hat{\pi}_{2.1} = \frac{30}{90}, \quad \hat{\pi}_{3.1} = \frac{40}{90}$$

$$\hat{\pi}_{1.2} = \frac{50}{130}, \quad \hat{\pi}_{2.2} = \frac{60}{130}, \quad \hat{\pi}_{3.2} = \frac{20}{130}$$

$$\hat{\pi}_{11} = 0.1030, \quad \hat{\pi}_{12} = 0.1545, \quad \hat{\pi}_{13} = 0.2060,$$

$$\hat{\pi}_{21} = 0.2064, \quad \hat{\pi}_{22} = 0.2377, \quad \hat{\pi}_{23} = 0.0826$$

完全データのみを用いた場合

$$\tilde{\pi}_{11} = 0.0909, \quad \tilde{\pi}_{12} = 0.1364, \quad \tilde{\pi}_{13} = 0.1818,$$

$$\tilde{\pi}_{21} = 0.2273, \quad \tilde{\pi}_{22} = 0.2727, \quad \tilde{\pi}_{23} = 0.0909$$

※. 観測データ全体を用いた最尤法

: MAR と分布のパラメータと欠損パラメータと分離が必要
完全データのみを用いた最尤法

: MCAR と分布のパラメータと欠損パラメータと分離が必要

MCAR であれば, $\hat{\theta} - \tilde{\theta} \approx 0$

⇒ 大標本近似を用いて検定を行うと p 値 < 0.01 となり
棄却 ($\chi_1^2 = 5.23$)

9.2 一般の欠損構造の場合

$$y_1, \dots, y_n \stackrel{i.i.d.}{\sim} \text{Mt}_C(1, \theta), \quad \theta = (\pi_1, \dots, \pi_C)$$

$$\{1, 2, \dots, C\} \supset S_1, S_2, \dots, S_K$$

y_1, \dots, y_r : 該当するカテゴリーが観測されている個体

y_{r+1}, \dots, y_n : S_1, \dots, S_K のどれかに該当することがわかっている

$$r_c = \#\{i; y_i = c\}, \quad m_k = \#\{i; y_i \in S_k\}$$

$$\delta(c \in S_k) = \begin{cases} 1 & c \in S_k \\ 0 & c \notin S_k \end{cases}$$

E ステップ

$n_c = \#\{i; y_i = c\}$: 未知の値

$$\theta^{(t)} = (\pi_1^{(t)}, \dots, \pi_C^{(t)})$$

$$n^{(t)} = \mathbb{E}[n_c | r_1, \dots, r_C, m_1, \dots, m_K, \theta^{(t)}]$$

$$= r_c + \sum_{k=1}^K m_k \psi_{c \cdot S_k}^{(t)}, \quad \psi_{c \cdot S_k}^{(t)} = \frac{\pi_c^{(t)} \delta(c \in S_k)}{\sum_{j=1}^K \pi_j^{(t)} \delta(j \in S_k)}$$

M ステップ

$$\pi^{(t+1)} = \frac{\pi^{(t)}}{n}$$

例. $C = 4$

観測データ $Y_{\text{obs}} = (r_1, r_2, r_3, r_4, m_1, m_2)$

$$S_1 = \{1, 2\}, S_2 = \{1, 3, 4\}$$

r_c : カテゴリー c ($c = 1, 2, 3, 4$) に属する個体数 (完全分類)

m_1 カテゴリー 1 または 2 に属する個体数 (部分的分類)

m_2 カテゴリー 1, 3, 4 のいずれかに属する個体数 (部分的分類)

観測されなかったデータ $Y_{\text{mis}} = (n_{11}, n_{12}, n_{21}, n_{23}, n_{24})$

n_{11} : m_1 個の個体中, カテゴリー 1 に属する個体数

$$n_{12} = m_1 - n_{11}$$

n_{21} : m_2 個の個体中, カテゴリー 1 に属する個体数

n_{23} : m_2 個の個体中, カテゴリー 3 に属する個体数

$$n_{24} = m_2 - n_{21} - n_{23}$$

対数尤度関数

$$\ell((\pi_1, \pi_2, \pi_3, \pi_4) | Y_{\text{obs}}, Y_{\text{mis}}) = n_1 \log \pi_1 + n_2 \log \pi_2 + n_3 \log \pi_3 + n_4 \log \pi_4$$

$$n_1 = r_1 + n_{11} + n_{21}, \quad n_2 = r_2 + n_{12},$$

$$n_3 = r_3 + n_{23}, \quad n_4 = r_4 + n_{24}$$

$$\theta^{(t)} = (\pi_1^{(t)}, \pi_2^{(t)}, \pi_3^{(t)}, \pi_4^{(t)})$$

$$n_1^{(t)} = \mathbb{E}[r_1 + n_{11} + n_{21} | r_1, r_2, r_3, r_4, m_1, m_2, \theta^{(t)}],$$

$$= r_1 + m_1 \frac{\pi_1^{(t)}}{\pi_1^{(t)} + \pi_2^{(t)}} + m_2 \frac{\pi_1^{(t)}}{\pi_1^{(t)} + \pi_3^{(t)} + \pi_4^{(t)}}$$

$$\psi_{c, S_1}^{(t)} = \frac{\pi_1^{(t)} \delta(1 \in S_1)}{\pi_1^{(t)} \delta(1 \in S_1) + \pi_2^{(t)} \delta(2 \in S_1) + \pi_3^{(t)} \delta(3 \in S_1) + \pi_4^{(t)} \delta(4 \in S_1)}$$

例 9.3 2×2 分割表

<u>完全データ</u>				<u>Y_1 のみ観測</u>			<u>Y_2 のみ観測</u>						
		Y_2		計			計	Y_2		計			
レベル		1	2		レベル	1		2	レベル		1	2	
Y_1	1	100	50	150	Y_1	1		30	Y_1	1			
	2	75	75	150		2		60		2			
	計	175	125	300		計		90		計	28	60	88

完全データのみで、初期値 $\theta^{(1)} = (\pi_{11}^{(1)}, \pi_{12}^{(1)}, \pi_{21}^{(1)}, \pi_{22}^{(1)})$ を定める

$$\pi_{11}^{(1)} = \frac{100}{300}, \quad \pi_{12}^{(1)} = \frac{50}{300}$$

$$\pi_{21}^{(1)} = \frac{75}{300}, \quad \pi_{22}^{(1)} = \frac{75}{300}$$

パラメータ(1)

レベル	Y ₂		計
	1	2	
Y ₁ 1	$\frac{100}{300}$	$\frac{50}{300}$	$\frac{150}{300}$
2	$\frac{75}{300}$	$\frac{75}{300}$	$\frac{150}{300}$
計	$\frac{175}{300}$	$\frac{125}{300}$	1

条件付期待値

レベル	Y ₂		計
	1	2	
Y ₁ 1	20	10	30
2	30	30	60
計			90

条件付期待値

レベル	Y ₂		計
	1	2	
Y ₁ 1	16	24	
2	12	36	
計	28	60	88

パラメータ(2)

レベル	Y ₂		計
	1	2	
Y ₁ 1	$\frac{100+20+16}{478}$	$\frac{50+10+24}{478}$	$\frac{220}{478}$
2	$\frac{75+30+12}{478}$	$\frac{75+30+36}{478}$	$\frac{258}{478}$
計	$\frac{253}{478}$	$\frac{225}{478}$	1

9.3 対数線形モデル

$\pi_{jk\dots t} = P(Y_1 = j, Y_2 = k, \dots, Y_V = t) : V$ 次元分割表のセル確率

独立構造

$$\pi_{jk\dots t} = \tau \tau_j^{(1)} \tau_k^{(2)} \cdots \tau_t^{(V)}$$

\Rightarrow

$$\log \pi_{jk\dots t} = \alpha + \alpha_j^{(1)} + \alpha_k^{(2)} + \cdots + \alpha_t^{(V)},$$

$$\alpha_j^{(1)} = \log \tau_j^{(1)}, \alpha_k^{(2)} = \log \tau_k^{(2)}, \dots$$

識別性のための制約条件

$$\sum_{j=1}^J \alpha_j^{(1)} = \cdots = \sum_{t=1}^T \alpha_t^{(V)} = 0$$

一般の対数線形モデル

($V = 3$ の場合)

$$\log \pi_{jkl} = \alpha + \alpha_j^{(1)} + \alpha_k^{(2)} + \alpha_l^{(3)} + \alpha_{jk}^{(12)} + \alpha_{jl}^{(13)} + \alpha_{kl}^{(23)} + \alpha_{jkl}^{(123)}$$
$$\sum_{j=1}^J \alpha_j^{(1)} = \dots = \sum_{j=1}^J \alpha_{jk}^{(12)} = \sum_{k=1}^K \alpha_{jk}^{(12)} = \dots = \sum_{j=1}^J \alpha_{jkl}^{(123)} = \dots = 0$$

$\alpha_j^{(1)}, \alpha_k^{(2)}, \alpha_l^{(3)}$ —主効果

$\alpha_{jk}^{(12)}, \alpha_{jl}^{(13)}, \alpha_{kl}^{(23)}$ —2次の関連

$\alpha_{jkl}^{(123)}$ —3次の関連

飽和モデル：主効果とすべての関連項を含むモデル

部分モデル：いくつかのパラメータを0としたモデル

階層構造モデル：例えば

$$\{12\} : \alpha_{jk}^{(12)} \neq 0 \Rightarrow \alpha_j^{(1)} \neq 0, \alpha_k^{(2)} \neq 0$$

$$\{12, 13\} : \alpha_j^{(1)}, \alpha_k^{(2)}, \alpha_l^{(3)}, \alpha_{jk}^{(12)}, \alpha_{jl}^{(13)} \text{ を含むモデル}$$

$\{123\}, \{12, 13, 23\}, \{12, 13\}, \{1, 23\}, \{12\}, \{1, 2, 3\}, \{1, 2\}, \{1\}, \{\emptyset\}$
と 1, 2, 3 を入れ替えたモデルがある。

※. $\{12, 13, 23\}$ では, 最尤推定値の導出に反復法が必要
他の階層構造モデルでは, 最尤推定値は陽に書ける

適合度検定

$$G^2 = 2 \sum_c n_c \log \frac{n_c}{\hat{n}_c} \quad \text{尤度比検定}$$

$$\chi^2 = \sum_c \frac{(n_c - \hat{n}_c)^2}{\hat{n}_c} \quad \text{カイ 2 乗検定}$$

$$\hat{n}_c = n \hat{\pi}_c$$

欠損データに対する EM アルゴリズム

E ステップ 一般の多項分布と同じ

M ステップ $\theta = (\alpha, \alpha_j^{(1)}, \dots)$ に関する最大化.

{12, 13, 23} では, 各 M ステップで反復法が必要となる
⇒ ECM アルゴリズム

適合度検定

$$G^2 = 2 \sum_c \tilde{n}_c \log \frac{\tilde{n}_c}{\hat{n}_c} \quad \text{尤度比検定}$$

$$\chi^2 = \sum_c \frac{(\tilde{n}_c - \hat{n}_c)^2}{\hat{n}_c} \quad \text{カイ 2 乗検定}$$

$\hat{n}_c = n\hat{\pi}_c$: 制限されたモデルでの期待度数

$\tilde{n}_c = n\tilde{\pi}_c$: 飽和モデルでの期待度数

例 9.4 3 次の関連性のない $2 \times 2 \times 2$ 分割表

$$\log \pi_{jkl} = \alpha + \alpha_j^{(1)} + \alpha_k^{(2)} + \alpha_l^{(3)} + \alpha_{jk}^{(12)} + \alpha_{jl}^{(13)} + \alpha_{kl}^{(23)} + \alpha_{jkl}^{(123)}$$

$$\log \frac{\pi_{j00}\pi_{j11}}{\pi_{j01}\pi_{j10}} = 4(\alpha_{00}^{(23)} + \alpha_{j00}^{(123)}) \quad (j = 0, 1)$$

$$\pi_{(j)kl} = \frac{\pi_{jkl}}{\pi_{+kl}} = P(Y_1 = j | (Y_2, Y_3) = (k, l)), \quad \pi_{+kl} = \pi_{0kl} + \pi_{1kl}$$

$$\log \frac{\pi_{j00}\pi_{j11}}{\pi_{j01}\pi_{j10}} = \log \frac{\pi_{(j)00}\pi_{(j)11}}{\pi_{(j)01}\pi_{(j)10}} + \log \frac{\pi_{+00}\pi_{+11}}{\pi_{+01}\pi_{+10}}$$

$$\alpha_{jkl}^{(123)} = 0 \quad \Leftrightarrow \quad \log \frac{\pi_{(0)00}\pi_{(0)11}}{\pi_{(0)01}\pi_{(0)10}} = \log \frac{\pi_{(1)00}\pi_{(1)11}}{\pi_{(1)01}\pi_{(1)10}}$$

$$y_{jkl} = \#\{i; y_{i1} = j, y_{i2} = k, y_{i3} = l\}, \quad (j, k, l = 0, 1)$$

$$\begin{aligned} \ell(\theta|Y) &= \sum_{j,k,l} y_{jkl} \log \pi_{jkl} \\ &= \sum_{j,k,l} y_{jkl} \log \pi_{(j)kl} + \sum_{k,l} y_{+kl} \log \pi_{+kl} \end{aligned}$$

$\pi_{(j)kl} = \pi_{(j)kl}^{(t)}$ を固定して, π_{jkl} について l を最大化すると

$$\pi_{jkl}^{(t+1)} = \pi_{(j)kl}^{(t)} \frac{y_{+kl}}{n} \quad \left(\pi_{+kl}^{(t+1)} = \frac{y_{+kl}}{n} \right)$$

$$\log \frac{\pi_{(0)00} \pi_{(0)11}}{\pi_{(0)01} \pi_{(0)10}} = \log \frac{\pi_{(1)00} \pi_{(1)11}}{\pi_{(1)01} \pi_{(1)10}} \quad \text{— (☆)}$$

※. $\pi_{(j)kl}^{(t)}$ が (☆) を満たす $\Rightarrow \pi_{(j)kl}^{(t+1)}$ も (☆) を満たす

ECM アルゴリズム

$$\text{初期値 } \pi_{jkl}^{(1)} = \frac{1}{8}$$

E ステップ : 多項分布の欠損データに対する E ステップと同じ

$$\Rightarrow y_{jkl}^{(t)}$$

CM ステップ

$$\text{CM1: } \pi_{jkl}^{(t+1/3)} = \pi_{jk(l)}^{(t)} \frac{y_{jk+}^{(t)}}{n}$$

$$\text{CM2: } \pi_{jkl}^{(t+2/3)} = \pi_{j(k)l}^{(t+1/3)} \frac{y_{j+l}^{(t)}}{n}, \quad \pi_{j(k)l}^{(t+1/3)} = \pi_{jkl}^{(t+1/3)} / \pi_{j+l}^{(t+1/3)}$$

$$\text{CM3: } \pi_{jkl}^{(t+3/3)} = \pi_{(j)kl}^{(t+2/3)} \frac{y_{+kl}^{(t)}}{n}, \quad \pi_{(j)kl}^{(t+2/3)} = \pi_{jkl}^{(t+2/3)} / \pi_{+kl}^{(t+2/3)}$$

まとめ

欠損データのしくみ

MCAR (Missing completely at random)

$$f(M|Y, \phi) = f(M|\phi) \quad (\forall Y, \phi)$$

MAR (Missing at random)

$$f(M|Y_{obs}, Y_{mis}, \phi) = f(M|Y_{obs}, \phi) \quad (\forall Y_{mis}, \phi)$$

Y_{obs} : 観測されたデータ , Y_{mis} : 欠測値

NMAR (Not missing at random)

欠損データの解析方法

完全データのみ用いる

NMAR の場合, バイアスが生じる

MCAR の場合, バイアスは生じないが情報の損失で, 推定精度が低下

加重法

欠損が起こらない共変量によって、個体の欠損確率を推定、バイアスを補正
傾向スコア

imputation

欠測値を予測値で置き換えて、完全データに対する手法を用いる。
推定値の標準誤差を過小評価 ⇒ 乱数を利用した imputation

統計モデルに基づく手法

最尤法, ベイズ法, 欠損構造が無視可能

単調欠測データの推測

パラメータ変換

$$\phi = \phi(\theta) = (\phi_1, \dots, \phi_J)$$

$$\ell(\phi|Y_{\text{obs}}) = \ell_1(\phi_1|Y_{\text{obs}}) + \dots + \ell_J(\phi_J|Y_{\text{obs}})$$

1. ϕ_1, \dots, ϕ_J は distinct (母数空間が直積) で, Bayes 法においては, 事前独立
2. 各 $\ell_j(\phi_j|Y_{\text{obs}})$ は完全データの対数尤度に対応

Sweep Operator

多変量正規モデル

EM アルゴリズム

E ステップ

$\theta = \theta^{(t)}$ を用いて, “ Y_{mis} ” を Y_{obs} を与えた条件付期待値で「置き換える」

$$Q(\theta|\theta^{(t)}) = \int \ell(\theta|Y) f(Y_{\text{mis}}|Y_{\text{obs}}, \theta = \theta^{(t)}) dY_{\text{mis}}$$

M ステップ

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta|\theta^{(t)})$$

EM アルゴリズムの拡張

ECM アルゴリズム, ECME アルゴリズム, PX-EM アルゴリズム

終わり