第二部 統計解析セッション

もしも、この機に統計担当者がNASレポートを振り返ったら･･･（後編）

武田薬品工業株式会社
黒田 晋吾

Contents

- 4章：Drawing Inference from Incomplete Data
  - 欠測のあるデータにおける推測
  - MARの仮定の下で用いられる解析手法
  - MARの仮定に関する留意点
- 5章：Principles and Methods of Sensitivity Analyses
  - 感度分析の考え方
  - Pattern Mixture ModelとSelection Model
  - Decision Making

Contents

- 4章：Drawing Inference from Incomplete Data
  - 欠測のあるデータにおける推測
  - MARの仮定の下で用いられる解析手法
  - MARの仮定に関する留意点
- 5章：Principles and Methods of Sensitivity Analyses
  - 感度分析の考え方
  - Pattern Mixture ModelとSelection Model
  - Decision Making

欠測のあるデータにおける推測

- 欠測のあるデータからの推測は、主観的な検証不可能な仮定に基づくものである
  - $p(y_{obs}, y_{mis}, m) = p(y_{obs}, m)p(y_{mis}|y_{obs}, m)$
- 欠測データの問題に対して、普遍的に用いることが可能な手法は存在しない
  - ICH E9ガイドラインでも同様の指摘
  - 欠測の発生を抑えることが大切
- アウトカムと欠測の有無の両方に影響を与える変数のデータを収集することが重要
  - MARの仮定の確からしさ

記号の定義

- $\boldsymbol{Y} = (Y_1, \cdots, Y_K)'$: 評価変数ベクトル
  $\boldsymbol{Y_{obs}}$: $\boldsymbol{Y}$の観測データ、$\boldsymbol{Y_{mis}}$: $\boldsymbol{Y}$の欠測データ
  $\boldsymbol{Y_t^-} = (Y_1, \cdots Y_{t-1})$
- $\boldsymbol{M} = (M_1, \cdots, M_K)'$: 欠測指示変数ベクトル
  Y_jが観測$\Leftrightarrow M_j = 0$、Y_jが欠測 $\Leftrightarrow M_j = 1$
  $\boldsymbol{M}$の要素はすべて観測される
- $\boldsymbol{X}$: 群や背景項目等のデザイン変数ベクトル
  主解析のモデルに含める変数
  すべて観測されると仮定
- $\boldsymbol{V}$: 補助変数ベクトル
  主解析のモデルに含めない変数、投与開始後のデータを含む
  $\boldsymbol{V_{obs}}$: $\boldsymbol{V}$の観測データ、$\boldsymbol{V_{mis}}$: $\boldsymbol{V}$の欠測データ
  （欠測パターンは$\boldsymbol{Y}$と$\boldsymbol{V}$で同じとする）
  $\boldsymbol{V_t^-} = (V_1, \cdots V_{t-1})$

MARの仮定の下で用いられる解析手法

手法	モデル	特徴	留意点
MMRM （尤度に基づく方法）	• 解析モデル $f(\boldsymbol{y}_{obs}\|\boldsymbol{x};\boldsymbol{\theta})$	• 欠測メカニズムのモデル化が不要	• 補助変数$\boldsymbol{v}$をモデルに含めることはできない • 推測の妥当性はFull Dataに対するモデル$f(\boldsymbol{y}\|\boldsymbol{x};\boldsymbol{\theta})$の妥当性に依存
Multiple Imputation （欠測データの補完による方法）	• 解析モデル $f(\boldsymbol{y}\|\boldsymbol{x};\boldsymbol{\theta})$ • 補完モデル $f(\boldsymbol{y}\|\boldsymbol{x},\boldsymbol{v};\boldsymbol{\psi})$	• 解析モデルには含めることができない補助変数を補完モデルで考慮可能	• 推測の妥当性はFull Dataに対するモデル$f(\boldsymbol{y}\|\boldsymbol{x};\boldsymbol{\theta})$の妥当性に依存 • 解析モデルと補完モデルでincompatibilityの問題が生じうる
IPW GEE （観測確率の重み付けによる方法）	• 平均構造のモデル $E[\boldsymbol{Y}\|\boldsymbol{X}]=g(\boldsymbol{X},\boldsymbol{\beta})$ • 観測確率モデル $P(M_t=0\|\boldsymbol{X},\boldsymbol{V}_t^-,\boldsymbol{Y}_t^-;\boldsymbol{\phi})$	• 解析時のモデルには含めることができない補助変数を観測確率モデルで考慮可能 • パラメトリックモデルを利用しない	• $\forall(t,\boldsymbol{X},\boldsymbol{V}_t^-,\boldsymbol{Y}_t^-)$, $P(M_t=0\|\boldsymbol{X},\boldsymbol{V}_t^-,\boldsymbol{Y}_t^-)>0$ • 欠測データの分布の台が観測データと同一である必要性 • 重みの値が大きくなることによって推定が不安定となり、偏りを生じたり推定量の分散が大きくなる可能性

MARの仮定に関する留意点

- 投与中止後のデータが収集されていない場合、MARの仮定の下でのいかなる解析も「すべての被験者が投与を継続した場合」の効果を推定することになる。一般に、これはintention-to-treat effectの推定として妥当ではない。
$[\boldsymbol{M}|\boldsymbol{X},\boldsymbol{V}_{\boldsymbol{obs}},\boldsymbol{V}_{\boldsymbol{mis}},\boldsymbol{Y}_{\boldsymbol{obs}},\boldsymbol{Y}_{\boldsymbol{mis}}]=[\boldsymbol{M}|\boldsymbol{X},\boldsymbol{V}_{\boldsymbol{obs}},\boldsymbol{Y}_{\boldsymbol{obs}}]$
$[\boldsymbol{Y}_{\boldsymbol{mis}},\boldsymbol{V}_{\boldsymbol{mis}}|\boldsymbol{X},\boldsymbol{V}_{\boldsymbol{obs}},\boldsymbol{Y}_{\boldsymbol{obs}},\boldsymbol{M}]=[\boldsymbol{Y}_{\boldsymbol{mis}},\boldsymbol{V}_{\boldsymbol{mis}}|\boldsymbol{X},\boldsymbol{V}_{\boldsymbol{obs}},\boldsymbol{Y}_{\boldsymbol{obs}}]$

- 投与中止後のデータも収集されている場合、MARの仮定の下でintention-to-treat effectを推定する際に、IPW法や欠測データの補完において投与中止後の情報が利用可能

MARの仮定に関する留意点

- Permutt (2015)
  - The mathematically sophisticated methods proposed in the NRC report were intended, we understand, to be used in combination with retrieved dropouts to estimate *de facto* estimands. … If applied straightforwardly to trials in which measurements are not taken after discontinuation of treatment, however, the methods would have little justification.

Contents

- 4章：Drawing Inference from Incomplete Data
  - 欠測のあるデータにおける推測
  - MARの仮定の下で用いられる解析手法
  - MARの仮定に関する留意点
- 5章：Principles and Methods of Sensitivity Analyses
  - 感度分析の考え方
  - Pattern Mixture ModelとSelection Model
  - Decision Making

感度分析

- $(Y, M|X, V)$の同時密度関数
  $f(y_{obs}, y_{mis}, m|x, v) = \underline{f(y_{obs}, m|x, v)}\,\underline{f(y_{mis}|y_{obs}, m, x, v)}$
  - 右辺第2項に対する仮定:
    欠測データに対する検証不可能な仮定（type(i)の仮定）
  - 右辺第1項に対する仮定:
    観測データに対する検証可能な仮定（type(ii)の仮定）
- 欠測のあるデータにおいて推測を行うためには観測データの情報を用いて欠測データに外挿するためのtype(i)の仮定が必要となる
  - つまり、(X, V)を与えた下でのMと(Y_{obs}, Y_{mis})との関係を特定する必要がある
  - MNARの仮定の下では、右辺第2項は観測データからだけでは推測できない

感度分析

- NASレポートで注目している感度分析
  - 欠測メカニズム、つまりtype(i)の仮定に対する感度分析
    - 結論がどれだけtype(i)の仮定に依存しているか？
    - type(i)の仮定に対する感度分析を行う際は、「仮定の理解の容易さ」と「仮定のもっともらしさ」が重要
  - Non-Future Dependenceを仮定

- Recommendation 15
  - Sensitivity analyses should be part of the primary reporting of findings from clinical trials. Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting.

PMMとSMの比較

	特徴	留意点
Pattern Mixture Model $f(y_{obs}, y_{mis}, m\|x)$ $= f(y_{mis}\|y_{obs}, m, x)$ $\times f(y_{obs}\|m, x) f(m\|x)$	• 観測データと欠測データは異なる分布に従うと考える • 欠測データがどのように補完されたかが明確 • 感度パラメータの解釈が容易 • 観測データの分布に対する確認が容易	• 欠測パターンごとの結果をまとめる必要が生じる場合、計算が困難となることがある • 補助変数を考慮することが困難であり、一般に追加の仮定が必要となる
Selection Modek $f(y_{obs}, y_{mis}, m\|x)$ $= f(y_{obs}, y_{mis}\|x)$ $\times f(m\|y_{obs}, y_{mis}, x)$	• 観測データと欠測データを区別せずに一つの分布に従うと考える • 補助変数を考慮するのが容易	• 欠測データがどのように補完されたかが明確でない • パラメトリックなSMはモデルの仮定に影響を受けやすい • 観測確率の逆数での重み付による解析は推定効率が悪い • 感度パラメータごとにtype(ii)の仮定を確認する必要性 • 感度パラメータの解釈が困難で妥当な範囲の設定が難しい

Decision Making

- 感度分析の結果を統合し、意思決定を行う方法にコンセンサスはない
  - どのような方法を用いた場合でも、主解析と各感度分析にどのような重みを置くかが重要
- 三つの意思決定の方法
  - 感度パラメータに対して妥当な値の範囲を設定し、その範囲での推定値の下限と上限を示す
  - MARの仮定の下での解析における結論が逆転する感度パラメータの値を表示
  - 各感度パラメータから算出される推定値を何らかの方法で平均化

Decision Making

- Permutt (2015)
  - In practice, so far as significance testing is concerned, an effect will confidently be considered significant only if it remains significant at the outer limit of a plausible range of deviations from missing-at-randomness.
  - Possibly the NRC panel wanted to emphasize estimation over significance testing. … Binary decisions are nevertheless needed in the regulatory setting. As a practical matter, regulators will need to act on the analyses at the unfavorable end of the plausible range, and applicants will need to anticipate and plan accordingly.

参考文献

- National Research Council (2010). *The prevention and treatment of missing data in clinical trials*. The National Academic Press.
- Permutt, T. (2015). A taxonomy of estimands for regulatory clinical trials with discontinuations. *Statistics in Medicine*, doi: 10.1002/sim.6841
- Permutt, T. (2015). Sensitivity analysis for missing data in regulatory submissions. Statistics in Medicine, doi: 10.1002/sim.6753

Back Up

欠測データに対処するための6つの原則

1. 欠測によって、解析において重要な真の値が得られなくなっているかを決定すること
   - 死亡後のデータを欠測データとして考慮することは適切？
2. causal estimandに対する適切な推測が可能な解析となっていること
3. 欠測が生じた理由が可能な限り記録されていること
4. 欠測メカニズムに対する主要な仮定が決定されていること
5. 主要な欠測メカニズムの仮定の下で妥当な解析が実施されていること
6. 感度分析により、治療効果に対する頑健性の評価がされていること

欠測メカニズム

- Missing Completely at Random（MCAR）

$$[M|X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M]$$

- Missing at Random（MAR）

$$[M|X, V_{obs}, V_{mis}, Y_{obs}, Y_{mis}] = [M|X, V_{obs}, Y_{obs}]$$
$$[Y_{mis}, V_{mis}|X, V_{obs}, Y_{obs}, M] = [Y_{mis}, V_{mis}|X, V_{obs}, Y_{obs}]$$

- Missing Not at Random（MNAR）

  上記以外

疑問

- MとY_{obs}はそもそも対応していた（Y_{obs}からMの値は決まる）はず？

欠測メカニズム

- Seaman et al. (2013)によるMARの定義（$\boldsymbol{X}$と$\boldsymbol{V}$は省略）
  - $o(\boldsymbol{Y},\boldsymbol{M})$:
    $\boldsymbol{M}$が0である時点の$\boldsymbol{Y}$の部分ベクトルを取り出す関数
  - $g_\phi(\boldsymbol{m}|\boldsymbol{y})$:
    $\boldsymbol{M}$の条件付き確率関数（ϕ:パラメータ）

  - データが（everywhere）MARであるとは、任意のϕについて以下が成立すること
    （$\boldsymbol{y},\boldsymbol{y}^*$はすべてが観測されているとは限らない）

$$g_\phi(\boldsymbol{m}|\boldsymbol{y}) = g_\phi(\boldsymbol{m}|\boldsymbol{y}^*),$$
$$\forall \boldsymbol{m},\boldsymbol{y},\boldsymbol{y}^* \text{ s.t. } o(\boldsymbol{y},\boldsymbol{m}) = o(\boldsymbol{y}^*,\boldsymbol{m})$$

尤度に基づく方法

- $f(\boldsymbol{y}, \boldsymbol{m}|\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\phi})$: $(\boldsymbol{Y}, \boldsymbol{M})$の同時密度関数
  $\boldsymbol{\theta}$: $\boldsymbol{Y}$の分布を規定するパラメータ
  $\boldsymbol{\phi}$: $\boldsymbol{M}$の分布を規定するパラメータ

- $f(\boldsymbol{y}_{obs}, \boldsymbol{m}|\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\phi})$: 観測データの同時密度関数
  $= \int f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}) f(\boldsymbol{m}|\boldsymbol{y}, \boldsymbol{v}; \boldsymbol{\phi}) d\boldsymbol{y}_{mis}$

- 観測データの対数尤度関数
  $\log L(\boldsymbol{\theta}, \boldsymbol{\phi})$
  $= \sum \log \int f\left(\boldsymbol{y}^i \middle| \boldsymbol{x}^i; \boldsymbol{\theta}\right) f\left(\boldsymbol{m}^i \middle| \boldsymbol{y}^i, \boldsymbol{x}^i; \boldsymbol{\phi}\right) d\boldsymbol{y}_{mis}^i$

尤度に基づく方法

- MARの下での観測データの対数尤度関数

$$
\begin{aligned}
&\log L(\boldsymbol{\theta}, \boldsymbol{\phi}) \\
&= \sum \log \int f\left(\boldsymbol{y}^i \middle| \boldsymbol{x}^i; \boldsymbol{\theta}\right) f\left(\boldsymbol{m}^i \middle| \boldsymbol{y}^i, \boldsymbol{x}^i; \boldsymbol{\phi}\right) d\boldsymbol{y}_{\boldsymbol{mis}}^i \\
&= \sum \log \int f\left(\boldsymbol{y}^i \middle| \boldsymbol{x}^i; \boldsymbol{\theta}\right) f\left(\boldsymbol{m}^i \middle| \boldsymbol{y}_{\boldsymbol{obs}}^i, \boldsymbol{x}^i; \boldsymbol{\phi}\right) d\boldsymbol{y}_{\boldsymbol{mis}}^i \\
&= \sum \log \int f\left(\boldsymbol{y}^i \middle| \boldsymbol{x}^i; \boldsymbol{\theta}\right) d\boldsymbol{y}_{\boldsymbol{mis}}^i f\left(\boldsymbol{m}^i \middle| \boldsymbol{y}_{\boldsymbol{obs}}^i, \boldsymbol{x}^i; \boldsymbol{\phi}\right) \\
&= \underline{\sum \log f\left(\boldsymbol{y}_{\boldsymbol{obs}}^i \middle| \boldsymbol{x}^i; \boldsymbol{\theta}\right)} + \sum \log f\left(\boldsymbol{m}^i \middle| \boldsymbol{y}_{\boldsymbol{obs}}^i, \boldsymbol{x}^i; \boldsymbol{\phi}\right)
\end{aligned}
$$

- $\boldsymbol{\theta}$と$\boldsymbol{\phi}$にdistinctness conditionを仮定すれば、尤度に基づく方法により観測データのみを利用して$\boldsymbol{\theta}$の推測が可能
  - 欠測メカニズムのモデル化が不要（無視可能）
  - ただし、$\boldsymbol{\phi}$に関する推測はできない

MMRM

- Mixed Model for Repeated Measures（MMRM）
  - 名称にはいろいろな意見があるよう･･･
- 前頁の$f(y_{obs}|x;\theta)$に線形混合モデルを仮定
  - 通常、変量効果はモデルに含めない
  - 時点をカテゴリ変数としてモデルに含める
  - 解析モデルに含めるXによってMARの仮定を満たす必要がある
    - 補助変数vをモデルに含めることはできない
    - MARの仮定の妥当性？

尤度に基づく方法の特徴と注意点

- 特徴
  - 欠測が無視可能(MARかつdistinctness conditionが成立)であれば、尤度に基づく方法は妥当な推測を与える

- 注意点
  - 推測の妥当性は、完全データに対するパラメトリックモデル$f(y|x;\boldsymbol{\theta})$の妥当性に依存する
    - 観測データから、モデル$f(y|x;\boldsymbol{\theta})$の妥当性を確認することはできないことに注意

Multiple Imputation

- Single Imputationにおける問題点
  - 補完法での仮定が適切でない場合、補完により生成された完全データに基づく推測は偏りが生じている可能性がある
  - 補完データは実際に得られたデータと見なされるため、精度が過大に評価される
（欠測データの不確実性が考慮されない）

- Recommendation 10
  - Single imputation methods like last observation carried forward and baseline observation carried forward should not be used as the primary approach to the treatment of missing data unless the assumptions that underlie them are scientifically justified.

Multiple Imputation

- MARの下でのMultiple Imputationの手順

$$[\boldsymbol{Y}_{\boldsymbol{mis}}, \boldsymbol{V}_{\boldsymbol{mis}} | \boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{obs}}, \boldsymbol{Y}_{\boldsymbol{obs}}, \boldsymbol{M}] = [\boldsymbol{Y}_{\boldsymbol{mis}}, \boldsymbol{V}_{\boldsymbol{mis}} | \boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{obs}}, \boldsymbol{Y}_{\boldsymbol{obs}}]$$

1. 解析モデル$f(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta})$を指定
2. 補完モデル$f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}; \boldsymbol{\psi})$を指定し、観測データにあてはめる
3. 2.から予測分布$f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}) = \int f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}; \boldsymbol{\psi}) f(\boldsymbol{\psi}) d\boldsymbol{\psi}$を構成
4. 予測分布を利用して欠測データの補完を複数回実施し、K個の完全データセットを作成
5. 1.のモデルをK個の完全データそれぞれにあてはめる
6. K個の解析データから算出されたK個の解析結果を統合

補完を複数回実施することで、欠測データの不確実性を考慮

Multiple Imputationの特徴と注意点

- 特徴
  - 欠測データの不確実性を反映可能
  - 解析時のモデルには含めることができない補助変数を補完モデルで考慮することが可能
    - MARの仮定がより確からしくなる

- 注意点
  - パラメトリックモデル$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v};\boldsymbol{\theta})$の妥当性
    - 補完値が実際にはとり得ない値となることも
  - 解析モデル$f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$と補完モデル$f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v};\boldsymbol{\psi})$でincompatibilityの問題が生じうる
    - $\int f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v};\boldsymbol{\psi})f(\boldsymbol{v}|\boldsymbol{x};\boldsymbol{\psi})d\boldsymbol{v} = f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$が成立しないこと

Proper ImputationとIncompatibility

- Proper Imputation
  - 経験的には以下の条件の下で"proper"なmultiple imputation(MIによる推測が妥当)になると考えられている(Rubin, 1987)
    1. 補完値がベイズ流予測分布(又はその近似)からの実現値として得られており、補完間のバラツキが適切に考慮されている
    2. 仮定する反応過程に関して、欠測データに対する適切な補完モデルが選択されていること
    3. 完全データに対する解析モデルとして適切なモデルが選択されていること
- Incompatibility(Meng, 1994)
  - Uncongenialityともいう
  - 上記の2.と3.に関連して、解析モデルから算出される$\boldsymbol{\theta}$の推定量$\hat{\boldsymbol{\theta}}$及び$\hat{\boldsymbol{\theta}}$の分散の推定量が、補完モデルにおける$\boldsymbol{\theta}$の事後平均及び事後分散と漸近的に等しくならない

$\hat{\boldsymbol{\theta}}_{imp}$の事後分布→$\boldsymbol{\theta}_j$をサンプリング→$\boldsymbol{Y}_{mis}$をサンプリング
→ $\hat{\boldsymbol{\theta}}_j$を算出→結果を統合して$\hat{\boldsymbol{\theta}}$を算出

Proper ImputationとIncompatibility

- 実際の解析では・・・（O’Kelly and Ratitch, 2014）
  - MARの仮定の下でMultiple Imputationを利用する際は、補完モデルは少なくとも解析モデル以上に複雑であることが必要
    - つまり、補完モデルに含めた因子を解析モデルでは除いてもよいが、解析モデルに含める因子を補完モデルから除いてはいけない
  - MNARの仮定の下での解析では、必ずしもこの要件を満たしていない
    - control-based imputationによる解析では、Incompatibilityのために分散が過大推定されることがある（Lu, 2014、Tang, 2015）

IPW GEE

- Inverse Probability Weighted Generalized Estimating Equations（Robins, et al. 1995）

$$\sum_i \left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)' V_i^{-1} \Delta_i (\boldsymbol{Y}_i - \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\beta})) = 0$$

  - $\boldsymbol{\beta}$: 回帰パラメータ
  - $E[\boldsymbol{Y}_i | \boldsymbol{X}_i] = \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\beta})$
  - V_i : 作業共分散行列
  - $\pi_{it} = \pi_{it}(\boldsymbol{X}_i, \boldsymbol{V}_{it}^-, \boldsymbol{Y}_{it}^-; \boldsymbol{\phi}) = P(M_{it} = 0 | \boldsymbol{X}_i, \boldsymbol{V}_{it}^-, \boldsymbol{Y}_{it}^-; \boldsymbol{\phi})$: 時点tでの観測確率（$\boldsymbol{V}_t^- = (V_1, \cdots, V_{t-1})$、$\boldsymbol{Y}_t^- = (Y_1, \cdots, Y_{t-1})$）
  - $$\Delta_i = \begin{pmatrix} \pi_{i1}^{-1}(1 - M_{i1}) & 0 & \cdots & 0 \\ 0 & \pi_{i2}^{-1}(1 - M_{i2}) & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \pi_{iK}^{-1}(1 - M_{iK}) \end{pmatrix}$$

<u>単調な欠測を仮定</u>

IPW GEEを用いた解析の流れ

- 以下の手順で実施
  1. 完全データの下で使用する回帰モデルを指定
  2. 観測確率モデルを選択
     $\lambda_t(\boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{t}}^{-}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}; \boldsymbol{\phi}) \quad = P(M_t = 0 | M_{t-1} = 0, \boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{t}}^{-}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}; \boldsymbol{\phi})$
  3. 観測確率モデルにおけるパラメータ$\boldsymbol{\phi}$を推定
  4. 時点tにおける観測確率を算出
     $\hat{\pi}_t = \hat{\pi}_t\left(\boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{t}}^{-}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}; \widehat{\boldsymbol{\phi}}\right) = \Pi_k^t \lambda_k(\boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{t}}^{-}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}; \widehat{\boldsymbol{\phi}})$
  5. $\Delta = \Delta(\hat{\pi}_1, \cdots, \hat{\pi}_K)$としてIPW GEEを解くことにより1.のモデルの回帰パラメータを推定
  6. ブートストラップ法によりSEを推定

$$\sum_i \left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)' V_i^{-1} \Delta_i (\boldsymbol{Y}_i - \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\beta})) = 0$$

$$\Delta_i = \begin{pmatrix} \pi_{i1}^{-1}(1 - M_{i1}) & 0 & \cdots & 0 \\ 0 & \pi_{i2}^{-1}(1 - M_{i2}) & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \pi_{iK}^{-1}(1 - M_{iK}) \end{pmatrix}$$

IPW GEEの特徴と注意点

- 特徴
  - 観測確率π_tに対するモデルが正しければ、大標本の下で偏りのない推定が可能
  - 解析時のモデルには含めることができない補助変数を観測確率モデルで考慮することが可能
    - MARの仮定がより確からしくなる
  - パラメトリックモデルを利用しない（GEEと同様）

IPW GEEの特徴と注意点

- 注意点
  - MARの仮定に加えて以下を仮定している
    - Y_tが観測不能となるような$(\boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{t}}^{-}, \boldsymbol{Y}_{\boldsymbol{t}}^{-})$が存在しないこと
      $$\forall(t, \boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{t}}^{-}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}), P(M_t = 0 | \boldsymbol{X}, \boldsymbol{V}_{\boldsymbol{t}}^{-}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}) > 0$$
    - 欠測データの分布の台（support）が観測データと同一であること
  - 観測確率モデルを正しく指定する必要がある
  - 有限標本の場合、重みの値が大きくなることによって推定が不安定となり、偏りを生じたり推定量の分散が大きくなる可能性がある

Augmented IPW

- IPW GEEでは
  - 欠測のある被験者の情報を十分に利用していない
    - 欠測のある被験者の「欠測したという情報」は $\pi_{it}(\boldsymbol{X_i}, \boldsymbol{V_{it}^{-}}, \boldsymbol{Y_{it}^{-}}; \boldsymbol{\phi})$の推定に利用される
    - 一方で、観測データを用いた予測によって得られる欠測データに関する情報は、解析モデルにおける回帰パラメータ$\boldsymbol{\beta}$の推定には利用されない

$$\sum_i \left(\frac{\partial \boldsymbol{g}(\boldsymbol{X_i}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)' V_i^{-1} \Delta_i (\boldsymbol{Y_i} - \boldsymbol{g}(\boldsymbol{X_i}, \boldsymbol{\beta})) = 0$$

Augmented IPW

- Augmented IPW法を用いることで、IPW GEE法における問題点を改善可能

$$\hat{\mu} = \frac{1}{n}\sum_{i} \frac{(1-M_{iT})Y_{iT}}{\pi_{iT}} + \frac{1}{n}\sum_{i} \left\{1 - \frac{1-M_{iT}}{\pi_{iT}}\right\} E\left(Y_T \middle| \boldsymbol{X}_i, \boldsymbol{V}_{\overline{iT}}, \boldsymbol{Y}_{\overline{iT}}\right)$$

- Augmentation termを追加することで、欠測のある被験者の欠測時点のデータに関する情報を解析モデルにおける回帰パラメータ$\boldsymbol{\beta}$の推定において利用し、推定効率を改善
- Doubly robustnessと呼ばれる性質をもつ
  - 観測確率モデル又はAugmentation termにおける予測モデルのいずれかが正しければ偏りのない推定が可能

Augmented IPW

- 他のAIPW

（Seaman and Copas, 2009）

$$\sum_i \left\{ \frac{1-M_{iK}}{\pi_{iK}} \left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)' V_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\beta})) + \sum_j^{K-1} \left(\frac{I(M_{ij}=0, M_{i,j+1}=1) - (1-\lambda_{i,j+1})(1-M_{ij})}{\pi_{i,j+1}} \right) \boldsymbol{H}_{ij}(\boldsymbol{\beta}) \right\} = 0$$

$$\boldsymbol{H}_j(\boldsymbol{\beta}) = \left(\frac{\partial \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)' V^{-1} \left(\boldsymbol{W}_j - \boldsymbol{g}(\boldsymbol{X}, \boldsymbol{\beta}) \right)$$

$$\boldsymbol{W}_j = \left(W_{j1}, \cdots, W_{jK} \right),\ W_{jt} = Y_t \quad (j < t),\ W_{jt} = E\left(Y_t \middle| \boldsymbol{X}, \boldsymbol{V}_{j+1}^{-}, \boldsymbol{Y}_{j+1}^{-} \right) \ (t \geq j)$$

IPW GEEに関するRecommendation

- Recommendation 13
  - Weighted generalized estimating equations methods should be more widely used in settings when missing at random can be well justified and a stable weight model can be determined, as a possibly useful alternative to parametric modeling.

MARの仮定について

- Recommendation 14
  - When substantial missing data are anticipated, auxiliary information should be collected that is believed to be associated with reasons for missing values and with the outcomes of interest. This could improve the primary analysis through use of a more appropriate missing at random model or help to carry out sensitivity analyses to assess the impact of missing data on estimates of treatment differences. In addition, investigators should seriously consider following up all or a random sample of trial dropouts, who have not withdrawn consent, to ask them to indicate why they dropped out of the study, and, if they are willing, to collect outcome measurements from them.

Non-future dependence

- 以下
  - 単調な欠測を仮定
  - 補助変数$\boldsymbol{V}$は省略
  - L：最終観測時点（つまり、$l > L$についてY_lはすべて欠測）
- Missing at Random（MAR）
  $$[L|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}},\boldsymbol{Y}_{\boldsymbol{mis}}] = [L|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}}]$$
  $$[\boldsymbol{Y}_{\boldsymbol{mis}}|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}},L] = [\boldsymbol{Y}_{\boldsymbol{mis}}|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}}]$$
  $$[Y_t|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{t}}^{-},L=l] = [Y_t|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{t}}^{-},L \geq t], \qquad t \geq 1, t > l$$
- Non-Future Dependence（NFD）（Kenward et al., 2003）
  $$[L|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}},\boldsymbol{Y}_{\boldsymbol{mis}}] = [L|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}},Y_{L+1}]$$
  $$[\boldsymbol{Y}_{\boldsymbol{mis}}|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}},Y_{L+1},L] = [\boldsymbol{Y}_{\boldsymbol{mis}}|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{obs}},Y_{L+1}]$$
  $$[Y_t|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{t}}^{-},L=l] = [Y_t|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{t}}^{-},L \geq t-1], \qquad t \geq 2,\ t > l+1$$

Non-future dependence

- Missing at Random（MAR）

$$[L|\boldsymbol{X},\boldsymbol{Y}_{obs},\boldsymbol{Y}_{mis}] = [L|\boldsymbol{X},\boldsymbol{Y}_{obs}]$$
$$[\boldsymbol{Y}_{mis}|\boldsymbol{X},\boldsymbol{Y}_{obs},L] = [\boldsymbol{Y}_{mis}|\boldsymbol{X},\boldsymbol{Y}_{obs}]$$
$$[Y_t|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{t}}^{-},L=l] = [Y_t|\boldsymbol{X},\boldsymbol{Y}_{\boldsymbol{t}}^{-},L \geq t], \qquad t \geq 1, t > l$$

脱落時点＼時点	l	$l+1$	$l+2$	$l+3$	$l+4$
$L=l$	○	×	×	×	×
$L=l+1$	○	○	×	×	×
$L=l+2$	○	○	○	×	×
$L=l+3$	○	○	○	○	×
$L=l+4$	○	○	○	○	○

○ : $\boldsymbol{Y}_{obs}$、× : $\boldsymbol{Y}_{mis}$

Non-future dependence

- Non-Future Dependence（NFD）
  $[L|\boldsymbol{X}, \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}] = [L|\boldsymbol{X}, \boldsymbol{Y}_{obs}, Y_{L+1}]$
  $[\boldsymbol{Y}_{mis}|\boldsymbol{X}, \boldsymbol{Y}_{obs}, Y_{L+1}, L] = [\boldsymbol{Y}_{mis}|\boldsymbol{X}, \boldsymbol{Y}_{obs}, Y_{L+1}]$
  $[Y_t|\boldsymbol{X}, \boldsymbol{Y}_t^-, L = l] = [Y_t|\boldsymbol{X}, \boldsymbol{Y}_t^-, L \geq t-1], \qquad t \geq 2, t > l+1$

脱落時点／時点	l	$l+1$	$l+2$	$l+3$	$l+4$
$L=l$	○	△	×	×	×
$L=l+1$	○	○	△	×	×
$L=l+2$	○	○	○	△	×
$L=l+3$	○	○	○	○	△
$L=l+4$	○	○	○	○	○

○ : $\boldsymbol{Y}_{obs}$、× : $\boldsymbol{Y}_{mis}$、△ : Y_{L+1}

- △部分は欠測データであることに注意

Pattern Mixture Model

- Pattern Mixture Model（PMM）

$$f(y_{obs}, y_{mis}, m|x) = f(y_{mis}|y_{obs}, m, x) f(y_{obs}|m, x) f(m|x)$$

  - 右辺第1項の形から、Y_{mis}の予測分布を用いて明示的に欠測データの補完を実施する解析がよく行われる
    - PMMとMIを組み合わせた解析

Pattern Mixture Model

- PMM＋NFDでの解析の例（連続量の場合、Ratitch, et al, 2013）
  - 以下に従って欠測データに対する補完を実施する
    - 観測データから、$E(Y_t|\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}, L \geq t) = \mu_t + \boldsymbol{\beta}_{\boldsymbol{t}}{}'\boldsymbol{Y}_{\boldsymbol{t}}^{-}$を推定
    - $L = t - 1$のパターンについて
      $$E(Y_t|\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}, L = t-1) = (\mu_t + \Delta_{\mu_t}) + \boldsymbol{\beta}_{\boldsymbol{t}}{}'\boldsymbol{Y}_{\boldsymbol{t}}^{-}$$
      により補完モデルを構築し欠測データを補完
    - 上記の補完データと観測データから、
      $$E(Y_t|\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}, L \geq t-1) = \tilde{\mu}_t + \tilde{\boldsymbol{\beta}}_{\boldsymbol{t}}{}'\boldsymbol{Y}_{\boldsymbol{t}}^{-}$$
      を推定
    - $L < t - 1$のパターンについて
      $$E(Y_t|\boldsymbol{x}, \boldsymbol{Y}_{\boldsymbol{t}}^{-}, L < t-1) = \tilde{\mu}_t + \tilde{\boldsymbol{\beta}}_{\boldsymbol{t}}{}'\boldsymbol{Y}_{\boldsymbol{t}}^{-}$$
      により補完モデルを構築し欠測データを補完
  - $\Delta_{\mu_t} = 0$はMARを意味する
  - Δ_{μ_t}を動かすことによって、感度分析を実施（Δ_{μ_t}は感度パラメータと呼ばれる）

	l	$l+1$	$l+2$	$l+3$	$l+4$
$L=l$	○	△	×	×	×
$L=l+1$	○	○	△	×	×
$L=l+2$	○	○	○	△	×
$L=l+3$	○	○	○	○	△
$L=l+4$	○	○	○	○	○

Selection Model

- Selection Model（SM）

$$f(\boldsymbol{y_{obs}}, \boldsymbol{y_{mis}}, \boldsymbol{m}|\boldsymbol{x}) = f(\boldsymbol{y_{obs}}, \boldsymbol{y_{mis}}|\boldsymbol{x}) f(\boldsymbol{m}|\boldsymbol{y_{obs}}, \boldsymbol{y_{mis}}, \boldsymbol{x})$$

- SM＋NFDでの解析の例①
  - Diggle and Kenward (1994)の方法
    - $\boldsymbol{Y}$: 多変量正規分布を仮定
    - $\boldsymbol{M}$と$(\boldsymbol{Y_{obs}}, \boldsymbol{Y_{mis}})$に以下の関係を仮定

      $$\mathrm{logit}\{P(L = t|L \geq t, \boldsymbol{Y_{obs}}, \boldsymbol{Y_{mis}})\}$$
      $$= \mathrm{logit}\{P(L = t|L \geq t, Y_{t+1}, Y_t)\} = \alpha + \alpha_1 Y_t + \alpha_2 Y_{t+1}$$

      - $\alpha_2 = 0$はMARを意味する
      - モデルの仮定は観測データからは確認できない
      - パラメトリックなSelection Modelはモデルの仮定に影響を受けやすい

Selection Model

- SM＋NFDでの解析の例②
  - NASレポートでの感度分析
    - $\boldsymbol{M}$と$(\boldsymbol{Y_{obs}}, \boldsymbol{Y_{mis}})$に以下の関係を仮定
      $\mathrm{logit}\{P(L = t|L \geq t, \boldsymbol{Y_{obs}}, \boldsymbol{Y_{mis}})\}$
      $= \mathrm{logit}\{P(L = t|L \geq t, \boldsymbol{Y_{t+1}^{-}}, Y_{t+1})\}$
      $= \alpha_{0t} + \boldsymbol{\alpha'_{1t}} \boldsymbol{Y_{t+1}^{-}} + \alpha Y_{t+1} = b_t(\boldsymbol{Y_{t+1}^{-}}; \boldsymbol{\gamma_t}) + \alpha Y_{t+1}$
      ただし、$b_t(\boldsymbol{Y_{t+1}^{-}}; \boldsymbol{\gamma_t}) = \alpha_{0t} + \boldsymbol{\alpha'_{1t}} \boldsymbol{Y_{t+1}^{-}}, \boldsymbol{\gamma_t} = (\alpha_{0t}, \boldsymbol{\alpha_{1t}})'$
    - これは以下のPattern Mixture Modelでの式に書き換え可能
      $f(y_{t+1}|L = t, Y_{t+1}^{-}) = f(y_{t+1}|L \geq t+1, Y_{t+1}^{-})$
      $\times \exp(\alpha y_{t+1}) \times C$（$C$は定数）
      - 観測データからαに関する情報は得られない
      - αが時点tでの中止例とt以降の継続例を関連づける役割

Selection Model

- SM＋NFDでの解析の例②
  - NASレポートでの感度分析
    - $E(Y_K) = E\left\{\frac{(1-M_K)Y_K}{\pi_K(\boldsymbol{Y}, b_1, \cdots, b_K; \alpha)}\right\}$
    - $\pi_K(\boldsymbol{Y}, b_1, \cdots, b_K; \alpha)$
      $= \Pi_{t=1}^{K-1}[1 - \mathrm{expit}\{b_t(\boldsymbol{Y}_{t+1}^{-}; \boldsymbol{\gamma}_t) + \alpha Y_{t+1}\}]$
    - $\hat{\mu}_{IPW} = (1/n)\sum_i \frac{(1-M_{iK})Y_{iK}}{\pi_K(\boldsymbol{Y}_i, \hat{b}_{i1}, \cdots, \hat{b}_{iK}; \alpha)}$
    - $\sum_i \sum_{t=1}^{K-1}(1 - M_{it})\frac{\partial b_t(\boldsymbol{Y}_{i,t+1}^{-}; \boldsymbol{\gamma}_t)}{\partial \boldsymbol{\gamma}_t}\left[\frac{1-M_{i,t+1}}{1-\mathrm{expit}\{b_t(\boldsymbol{Y}_{i,t+1}^{-}; \boldsymbol{\gamma}_t) + \alpha Y_{i,t+1}\}} - 1\right] = 0$

      を解くことによって$\hat{\boldsymbol{\gamma}}_t$を推定し、$\hat{b}_t = b_t(Y_{t+1}^{-}; \hat{\boldsymbol{\gamma}}_t)$を算出

ナニワ

データサイエンス研究会

－ End of File －