

Bayesian AB テスト入門 (モデルベースABテスト入門)

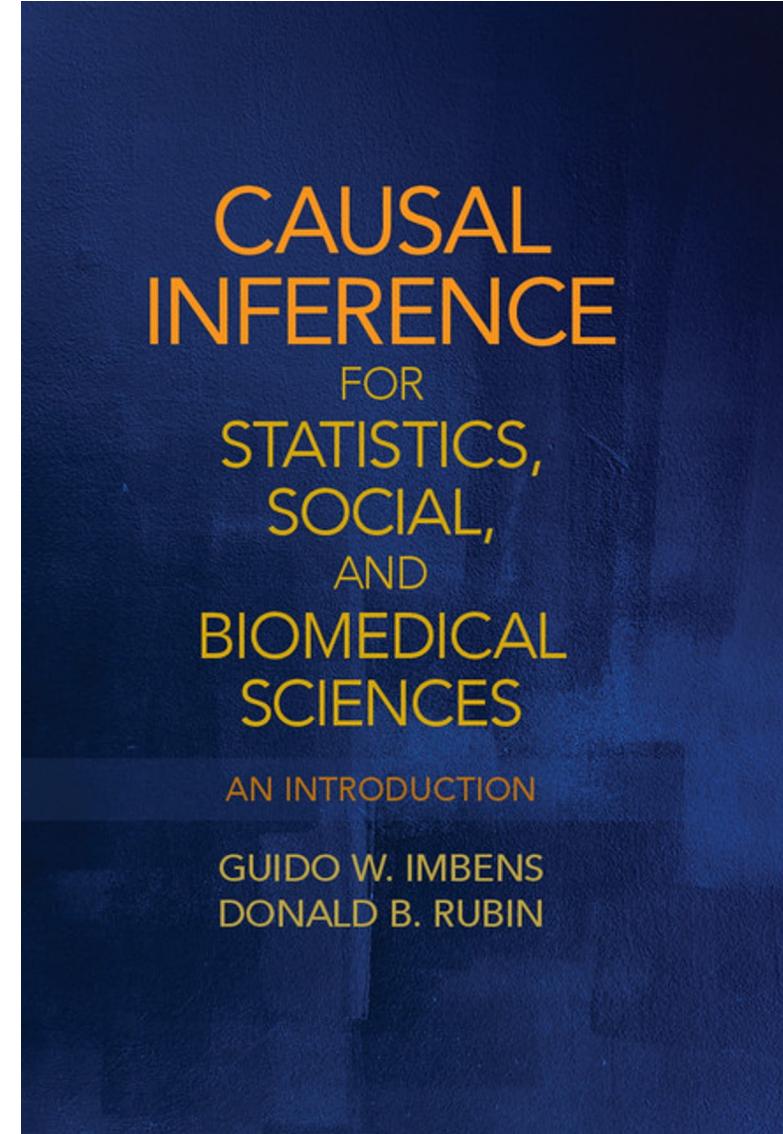
こっしー

参考文献

右の本の 8 章をがっつり転載します。

Rubin 大先生が著者の1人の本。

**Causal Inference For Statistics,
Social, And Biomedical Sciences**



対象者

- 因果推論の手法（の一つ）、潜在的結果変数フレームワークをぼんやり理解している人。
- 具体的には、因果効果とは $Y_i(1) - Y_i(0)$ で定義できることを何となく知っていたり、ATE という言葉に聞き馴染みがある人。
- ベイズを知っている人。事後予測分布と言われて思い当たる人。
- ベイズを知らなくても、聞き流せる人。

モデルベースな AB テストの簡易的な実装方法を知る

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4 ステップで ATE 推測
4. 具体例
5. シミュレーションベースで再計算
6. 潜在的結果変数に従属関係があった場合
7. 無限超母集団における ATE の推測
8. 実データで求めてみよう

1. 単純な解決策

- ある就職支援プログラムを行なった。
- $Y_i(0)$ はプログラム非参加後の収入
- $Y_i(1)$ はプログラム参加後の収入
- W_i はプログラムに参加したかを表す処置割り付け変数
- Y_i^{obs} は観測した収入
- 因果効果 ($Y_i(1) - Y_i(0)$) が ? のせいで分から = **因果推論の根本問題**

Table 8.2. First Six Observations from NSW Program Data

Unit	Potential Outcomes		Treatment W_i	Observed Outcome Y_i^{obs}
	$Y_i(0)$	$Y_i(1)$		
1	0	?	0	0
2	?	9.9	1	9.9
3	12.4	?	0	12.4
4	?	3.6	1	3.6
5	0	?	0	0
6	?	24.9	1	24.9

Note: Question marks represent missing potential outcomes.

1. 単純な解決策

- もし $\text{?} (= Y_i^{mis})$ が分かるとしたら、以下の式で簡単に ATE τ_{fs} は分かる。

$$\begin{aligned}\tau_{fs} &= \tilde{\tau}(Y^{obs}, Y^{mis}, W) \\ &= \frac{1}{6} \cdot \sum_{i=1}^N ((W_i \cdot Y_i^{obs} + (1 - W_i) \cdot Y_i^{mis}) \\ &\quad - ((1 - W_i) \cdot Y_i^{obs} + W_i \cdot Y_i^{mis})) \\ &= \frac{1}{6} \cdot \sum_{i=1}^N ((2 \cdot W_i - 1) \cdot (Y_i^{obs} - Y_i^{mis}))\end{aligned}$$

1. 単純な解決策

- なら Y_i^{mis} を埋めればいいでしょ?
- --> 各群の観測値の平均で Y_i^{mis} を埋めちゃおう ($= \hat{Y}_i^{mis}$)

$$\hat{\tau} = \tilde{\tau}(Y^{obs}, \hat{Y}^{mis}, W)$$

$$= \frac{1}{6} \cdot \sum_{i=1}^N ((2 \cdot W_i - 1) \cdot (Y_i^{obs} - \hat{Y}_i^{mis}))$$

- この $\hat{\tau}$ は、 $\bar{Y}_t^{obs} - \bar{Y}_c^{obs}$ に等しい。

Table 8.3. *The Average Treatment Effect Using Imputation of Average Observed Outcome Values within Treatment and Control Groups for the NSW Program Data*

Unit	Potential Outcomes		Treatment W_i	Observed Outcome Y_i^{obs}
	$Y_i(0)$	$Y_i(1)$		
1	0	(12.8)	0	0
2	(4.13)	9.9	1	9.9
3	12.4	(12.8)	0	12.4
4	(4.13)	3.6	1	3.6
5	0	(12.8)	0	0
6	(4.13)	24.9	1	24.9
Average	4.13	12.8		
Diff (ATE):		8.67		

1. 単純な解決策

問題点

- 一点のみの点推定しか分からず、不確実性を測れない
- Y_2^{mis} , Y_4^{mis} , Y_6^{mis} が全部 4.13 に一致するわけがない
 - Y_1^{mis} , Y_3^{mis} , $Y_5^{mis} = 12.8$ もあり得ない
- --> Y_i^{mis} の推定に変動がほしい
- --> Y_i^{mis} の分布が知りたい

1. 単純な解決策

解決策2

- Y_i^{obs} からランダムに選ぶ。
- ATE = 4.1, 10.7 --> バラついた!
- 右記 2通りのように、全ての組み合わせ $3^6 = 729$ 通りの ATE 全てを平均すると、 $\hat{\tau}^{dif} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}$ に等しくなる。
- 729 通りの ATE を用いれば、ATE の標準偏差が算出できる。

Table 8.4. The Average Treatment Effect Using Imputed Draws from the Empirical Distributions within Treatment and Control Groups for the First Six Units from the NSW Program Data

Unit	Potential Outcomes		Treatment W_i	Observed Outcome Y_i^{obs}
	$Y_i(0)$	$Y_i(1)$		
Panel A: First draw				
1	0	(3.6)	0	0
2	(12.4)	9.9	1	9.9
3	12.4	(9.9)	0	12.4
4	(12.4)	3.6	1	3.6
5	0	(9.9)	0	0
6	(0)	24.9	1	24.9
Average	6.2	10.3		
Diff (ATE):		4.1		
Panel B: Second draw				
1	0	(9.9)	0	0
2	(0)	9.9	1	9.9
3	12.4	(24.9)	0	12.4
4	(0)	3.6	1	3.6
5	0	(3.6)	0	0
6	(0)	24.9	1	24.9
Average	2.1	12.8		
Diff (ATE):		10.7		

1. 単純な解決策

問題点

- 6 以上のサンプルサイズだと、全通り計算するにはコストが高すぎる。
 - 3^n のオーダーで計算量が増えていく
- 限られたサンプルからランダム抽出すると不確実性を十分に考慮できない。
 - 上の例だと、 Y_2^{mis} , Y_4^{mis} , Y_6^{mis} のみからランダム抽出するのでは、現実を反映しきれなかった。
 - --> 現実の値が Y_2^{mis} , Y_4^{mis} , Y_6^{mis} のどれかに必ずなるわけがない

1. 単純な解決策

- どうすればより柔軟に Y_i^{mis} の不確実性を知れるのか
- --> 観測値で条件付けた Y_i^{mis} の分布を知る
- --> ベイズ！

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4ステップで ATE 推測
4. 具体例
5. シミュレーションベースで再計算
6. 潜在的結果変数に従属関係があった場合
7. 無限超母集団における ATE の推測
8. 実データで求めてみよう

2. \mathbf{Y}^{mis} の条件付き分布

- このベイジアンアプローチのゴールは、以下をモデル化すること

$$f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}).$$

- > これさえ分かれば、処置効果でも、各群における潜在的変数の分散比でも、導出できる！

2. \mathbf{Y}^{mis} の条件付き分布

1.で挙げた一つ目の例を確率分布で表現するなら

$$\Pr(\mathbf{Y}_i^{mis} = \mathbf{y} | \mathbf{Y}^{obs}, \mathbf{W}) = \begin{cases} 1 & \text{if } \mathbf{y} = 12.8, \text{ and } W_i = 0, \\ 1 & \text{if } \mathbf{y} = 4.1, \text{ and } W_i = 1, \\ 0 & \text{それ以外のとき} \end{cases}$$

二つ目の例は

$$\Pr(\mathbf{Y}_i^{mis} = \mathbf{y} | \mathbf{Y}^{obs}, \mathbf{W}) = \begin{cases} 1/3 & \text{if } \mathbf{y} \in \{3.6, 9.9, 24.9\}, \text{ and } W_i = 0, \\ 1/3 & \text{if } \mathbf{y} = 12.4, \text{ and } W_i = 1, \\ 2/3 & \text{if } \mathbf{y} = 0, \text{ and } W_i = 1, \\ 0 & \text{それ以外のとき} \end{cases}$$

2. \mathbf{Y}^{mis} の条件付き分布

- どちらのモデルも、欠測した潜在的結果変数の不確実性を捉えきれていなかった。
 - より柔軟な $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ を知りたい。
- しかし直接 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ を求めるのは非常に困難
 - $\mathbf{Y}(0), \mathbf{Y}(1)$ の同時分布と、割り付けメカニズム $f(\mathbf{W} | \mathbf{Y}(0), \mathbf{Y}(1))$ の複雑な関係を考慮しなければならないため。
 - しかし RCT では、割り付けメカニズムははっきりしている
 - --> 関係は比較的単純
- $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ を用いて、4 ステップで τ (= 今回の Estimand)を求める

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4 ステップで ATE 推測
4. 具体例
5. シミュレーションベースで再計算
6. 潜在的結果変数に従属関係があった場合
7. 無限超母集団における ATE の推測
8. 実データで求めてみよう

3. ATE 推測

4ステップの前に...

モデルに投入する要素の説明。

- $f(\mathbf{Y}(0), \mathbf{Y}(1))$
 - $= \int \prod_{i=1}^N f(Y_i(0), Y_i(1)|\theta) \cdot p(\theta) d\theta$ *¹
(デ・フィネッティの定理)
- $p(\theta)$ *²
- 処置割り当てメカニズム

$$f(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1} *^3$$

ここで N_t は、処置群に割り当てられるサンプルの総数。

3. ATE 推測

- 1: $f(Y_i(0), Y_i(1)|\theta)$ の特定はむずいが、完全な RCT ができていれば、誤特定をしても推論は大抵頑健である
- 2: 事前分布の選択はむずいが、結論は事前分布にそこまで左右されない
- 3: 観察研究であれば、 $f(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1))$ (割り付けメカニズム) もモデルの投入要素として自分で仮定しなければならないが、完全な RCT であれば $\binom{N}{N_t}^{-1}$ である

再掲：2. \mathbf{Y}^{mis} の条件付き分布

- どちらのモデルも、欠測した潜在的結果変数の不確実性を捉えきれていなかった。
 - より柔軟な $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ を知りたい。
- しかし直接 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ を求めるのは非常に困難
 - $\mathbf{Y}(0), \mathbf{Y}(1)$ の同時分布と、割り付けメカニズム $f(\mathbf{W} | \mathbf{Y}(0), \mathbf{Y}(1))$ の複雑な関係を考慮しなければならないため。
 - しかし RCT では、割り付けメカニズムははっきりしている
 - --> 関係は比較的単純
- **$f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ を用いて、4ステップで τ (= 今回の Estimand)を求める**

3. ATE 推測

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta)$ の導出

$$\begin{aligned} f(\mathbf{Y}(0), \mathbf{Y}(1) | \mathbf{W}, \theta) & (= f(\mathbf{Y}(0), \mathbf{Y}(1) | \theta)) \\ & = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} | \theta)}{\Pr(\mathbf{W} | \theta)} \\ & = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} | \theta)}{\int f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} | \theta) d\mathbf{Y}(0)d\mathbf{Y}(1)} \end{aligned}$$

3. ATE 推測

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta)$ の導出

ここで

$$\mathbf{Y}_i^{obs} = \begin{cases} \mathbf{Y}_i(0) & \text{if } \mathbf{W}_i = 0, \\ \mathbf{Y}_i(1) & \text{if } \mathbf{W}_i = 1, \end{cases}$$
$$\mathbf{Y}_i^{mis} = \begin{cases} \mathbf{Y}_i(0) & \text{if } \mathbf{W}_i = 1, \\ \mathbf{Y}_i(1) & \text{if } \mathbf{W}_i = 0, \end{cases}$$

あることに注意すると、

$$(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}) = g(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$$

3. ATE 推測

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta)$ の導出

$(\mathbf{Y}(0), \mathbf{Y}(1))$ は θ で表せるから、上記の $(\mathbf{Y}^{mis}, \mathbf{Y}^{obs})$ の変換を用いると、

$$f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs} | \mathbf{W}, \theta)$$

のように、分布を \mathbf{W}, θ の関数として表せる。

3. ATE 推測

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta)$ の導出

よって、以下が導出できる

$$\begin{aligned} f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta) &= \frac{f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs} | \mathbf{W}, \theta)}{f(\mathbf{Y}^{obs} | \mathbf{W}, \theta)} \\ &= \frac{f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs} | \mathbf{W}, \theta)}{\int_{\mathbf{y}^{mis}} f(\mathbf{y}^{mis}, \mathbf{Y}^{obs} | \mathbf{W}, \theta) d\mathbf{y}^{mis}} \end{aligned}$$

$f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta)$ は、 \mathbf{Y}^{mis} の事後予測分布ともいう

3. ATE 推測

ステップ2: θ の事後分布 $p(\theta | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

$$\begin{aligned} p(\theta | \mathbf{Y}^{obs}, \mathbf{W}) &= \frac{p(\theta) \cdot f(\mathbf{Y}^{obs}, \mathbf{W} | \theta)}{f(\mathbf{Y}^{obs}, \mathbf{W})} \\ &= \frac{p(\theta) \cdot f(\mathbf{Y}^{obs}, \mathbf{W} | \theta)}{\int_{\theta} p(\theta) \cdot f(\mathbf{Y}^{obs}, \mathbf{W} | \theta) d\theta} \\ &= \frac{p(\theta) \cdot \int_{\mathbf{y}^{mis}} f(\mathbf{y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W} | \theta) d\mathbf{y}^{mis}}{\int_{\theta} p(\theta) \cdot \int_{\mathbf{y}^{mis}} f(\mathbf{y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W} | \theta) d\mathbf{y}^{mis} d\theta} \end{aligned}$$

3. ATE 推測

ステップ3: \mathbf{Y}^{mis} の事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

ステップ1で求めた $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta)$ と、

ステップ2で求めた $p(\theta | \mathbf{Y}^{obs}, \mathbf{W})$ を用いて、

$$f(\mathbf{Y}^{mis}, \theta | \mathbf{Y}^{obs}, \mathbf{W}) = f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta) \cdot p(\theta | \mathbf{Y}^{obs}, \mathbf{W})$$

$$f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{mis}, \theta | \mathbf{Y}^{obs}, \mathbf{W}) d\theta$$

3. ATE 推測

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

$\mathbf{Y}(0), \mathbf{Y}(1)$ は以下のように、

$$Y_i(0) = \begin{cases} Y_i^{mis} & \text{if } W_i = 1, \\ Y_i^{obs} & \text{if } W_i = 0, \end{cases}$$

$$Y_i(1) = \begin{cases} Y_i^{mis} & \text{if } W_i = 0, \\ Y_i^{obs} & \text{if } W_i = 1, \end{cases}$$

となるので、

$$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W})$$

と表せる

3. ATE 推測

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

$(\mathbf{Y}(0), \mathbf{Y}(1))$ が分かってしまえば、求めたい Estimand (例えば ATE) は求められる。

- 例 : ATE $\Rightarrow E[\mathbf{Y}(1) - \mathbf{Y}(0)]$ で求まる

よって、 $\mathbf{Y}(0), \mathbf{Y}(1)$ の関数であり、任意の Estimand を一般に表した $\tilde{\tau}$ は、

$\mathbf{Y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W}$ の関数、 $\tilde{\tau}(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W})$ として求められる。

($\because (\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W})$ より)

3. ATE 推測

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

ステップ3で求めた $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ と組み合わせることにより、
 $\mathbf{Y}^{obs}, \mathbf{W}$ で条件付けた τ の事後分布

$$f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$$

を求めることができ、 τ の事後平均や、標準偏差などを求められる。

3. ATE 推測

- 単純な解決策との違いは？
- i. $\mathbf{Y}(0), \mathbf{Y}(1)$ のモデルを特定しなければならない
 - 未知母数 θ で条件付けたユニット単位の同時分布 $f(Y_i(0), Y_i(1) | \theta)$ が要特定
 - 比較: 単純な解決策は、条件付き分布が確実に分かると仮定していた
 - <-- どのパラメトリックな特定よりも重い仮定
- ii. θ の事前分布を選ばなければならない
 - RCT では、モデルが柔軟で、事前分布がいい加減なものでなく、データが十分にあれば、事後分布の推論は頑強である。

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4 ステップで ATE 推測
4. **具体例**
5. シミュレーションベースで再計算
6. 潜在的結果変数に従属関係があった場合
7. 無限超母集団における ATE の推測
8. 実データで求めてみよう

4. 具体例

- ・サンプルサイズ 6 の就職支援プログラムの例を再び考える
- ・前掲のモデルに投入する 3 要素を考える。
- ・potential outcomes を、平均は未知だが共分散行列は既知の二変量正規分布として設定する

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} | \theta \sim \mathcal{N}\left(\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix}\right)$$

- ・つまり、 $f(Y_i(0), Y_i(1)|\theta)$ は、

$$f(Y_i(0), Y_i(1)|\theta) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2 \cdot 100}(Y_i(0) - \mu_c)^2 - \frac{1}{2 \cdot 64}(Y_i(1) - \mu_t)^2\right)$$

再掲：3. ATE 推測

4ステップの前に...

モデルに投入する要素の説明。

- $f(\mathbf{Y}(0), \mathbf{Y}(1))$
 - $= \int \prod_{i=1}^N f(Y_i(0), Y_i(1)|\theta) \cdot p(\theta) d\theta$ *1
(デ・フィネッティの定理)
- $p(\theta)$ *2
- 処置割り当てメカニズム

$$f(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1} *3$$

4. 具体例

- $Y_i(0), Y_i(1)$ の値は、それぞれの平均から $2 * 10, 2 * 8$ の範囲に約 95 % の確率で入るというモデルになっている。
(正規分布より、 1.96σ 範囲に値を取る確率が 95% であるため)
- 分散共分散行列が既知だという仮定を緩めることもできる。
 - 後ほど 6 で紹介する
- もちろん別の事前分布を置くこともできる。
(非 0 の平均や、共分散を考慮など)
 - 今回で言うと、収入が負の値を取ることは考えられないで、本当は対数正規分布の方が良いかも？

4. 具体例

- 二つ目のモデル投入要素 θ の事前分布 $p(\theta)$ を考える
 - 今回だと、 $p(\theta) = p(\mu_c, \mu_t)$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10,000 & 0 \\ 0 & 10,000 \end{pmatrix}\right)$$

- もちろん別の事前分布を置くこともできる。
(非 0 の平均や、共分散を考慮など)
 - 今回で言うと、収入が負の値を取ることは考えられないで、本当は対数正規分布の方が良いかも？
- 完全な RCT の場合、データサイズが十分であれば、推論結果は事前分布の選択に左右されないことが分かっている

4. 具体例

- 三つ目のモデル投入要素 Assignment Mechanism は、RCT の場合、定義的に以下である。

$$\Pr(\mathbf{W} = \mathbf{w} | \mathbf{Y}(0), \mathbf{Y}(1), \mu_c, \mu_t) = \binom{N}{N_t}^{-1}$$

なお、 N_t は、処置群に割り当たられるサンプルの総数である。

- N 個のサンプルから処置群に割り当てるサンプル N_t 個を選ぶ組み合わせの数

- 以下、4ステップで具体的に ATE を求めていく

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

- potential outcomes は、 $\theta = (\mu_c, \mu_t)$ で条件付けた時にユニット単位で独立なので、

$$f(\mathbf{Y}(0), \mathbf{Y}(1) | \mu_c, \mu_t) = \prod_{i=1}^N f(Y_i(0), Y_i(1) | \mu_c, \mu_t).$$

- i_N, I_N をそれぞれ、要素が全て 1 の N 次元ベクトル、N 次元単位行列とすると、

$$\begin{pmatrix} \mathbf{Y}(0) \\ \mathbf{Y}(1) \end{pmatrix} | \mu_c, \mu_t \sim \mathcal{N}\left(\begin{pmatrix} \mu_c \cdot i_N \\ \mu_t \cdot i_N \end{pmatrix}, \begin{pmatrix} 100 \cdot I_N & 0 \cdot I_N \\ 0 \cdot I_N & 64 \cdot I_N \end{pmatrix}\right)$$

のように表せる（ユニット間で独立なため）

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

- 次に、データは完全な RCT より取得しているため、potential outcomes と θ で条件付けた W の分布は、

$$\Pr(\mathbf{W} = \mathbf{w} | \mathbf{Y}(0), \mathbf{Y}(1), \mu_c, \mu_t) = \binom{N}{N_t}^{-1}$$

$$s.t. \quad \sum_i W_i = N_t$$

- となるが、 θ で条件付けた \mathbf{W} と $(\mathbf{Y}(0), \mathbf{Y}(1))$ は独立なため、

$$\begin{pmatrix} \mathbf{Y}(0) \\ \mathbf{Y}(1) \end{pmatrix} | \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left(\begin{pmatrix} \mu_c \cdot i_N \\ \mu_t \cdot i_N \end{pmatrix}, \begin{pmatrix} 100 \cdot I_N & 0 \cdot I_N \\ 0 \cdot I_N & 64 \cdot I_N \end{pmatrix}\right)$$

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

- ここで $(\mathbf{Y}^{mis}, \mathbf{Y}^{obs})$ の条件付き分布に変換することを考える。

$$Y_i^{obs} = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1, \end{cases}$$

$$Y_i^{mis} = \begin{cases} Y_i(0) & \text{if } W_i = 1, \\ Y_i(1) & \text{if } W_i = 0, \end{cases}$$

であった。

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

- また、 $(Y_i(0), Y_i(1))$ と $(Y_{i'}(0), Y_{i'}(1))$ は ($i \neq i'$)、条件付き独立であったので、 (Y_i^{mis}, Y_i^{obs}) と $(Y_{i'}^{mis}, Y_{i'}^{obs})$ も条件付き独立であることに注意すると、

$$f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs} | \mathbf{W}, \mu_c, \mu_t) = \prod_{i=1}^N f(Y_i^{mis}, Y_i^{obs} | \mathbf{W}, \mu_c, \mu_t).$$

- であり、 $f(Y_i^{mis}, Y_i^{obs} | \mathbf{W}, \mu_c, \mu_t)$ は、

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

$$\begin{pmatrix} Y_i^{mis} \\ Y_i^{obs} \end{pmatrix} | \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left(\begin{pmatrix} W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t \\ (1 - W_i) \cdot \mu_c + W_i \cdot \mu_t \end{pmatrix}, \begin{pmatrix} W_i \cdot 100 + (1 - W_i) \cdot 64 & 0 \\ 0 & (1 - W_i) \cdot 100 + W_i \cdot 64 \end{pmatrix}\right)$$

- (前述の $Y_i(0), Y_i(1)$ の条件付き分布より)

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

- $f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs} | \mathbf{W}, \mu_c, \mu_t)$ より、 $\mathbf{Y}^{mis}, \mathbf{Y}^{obs}$ の共分散は 0 なので、 $(Y_i^{obs}, \mu_c, \mu_t)$ が与えられた上で Y_i^{mis} の条件付き分布は、 (μ_c, μ_t) を与えた上で Y_i^{mis} の周辺分布に等しい

$$Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}(W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t, \\ W_i \cdot 100 + (1 - W_i) \cdot 64)$$

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

- よって、 N 次元の \mathbf{Y}^{mis} の条件付き同時分布は、

$$\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N} \left(\begin{pmatrix} W_1 \cdot \mu_c + (1 - W_1) \cdot \mu_t \\ W_2 \cdot \mu_c + (1 - W_2) \cdot \mu_t \\ \vdots \\ W_N \cdot \mu_c + (1 - W_N) \cdot \mu_t \end{pmatrix}, \begin{pmatrix} W_1 \cdot 100 + (1 - W_1) \cdot 64 & 0 & \dots & 0 \\ 0 & W_2 \cdot 100 + (1 - W_2) \cdot 64 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_N \cdot 100 + (1 - W_N) \cdot 64 \end{pmatrix} \right).$$

4. 具体例

ステップ1: $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ の導出

- 今回の 6 例だと、

$$\left(\begin{array}{c} Y_1^{mis} \\ Y_2^{mis} \\ Y_3^{mis} \\ Y_4^{mis} \\ Y_5^{mis} \\ Y_6^{mis} \end{array} \right) \mid \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N} \left(\left(\begin{array}{c} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{array} \right), \left(\begin{array}{cccccc} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{array} \right) \right).$$

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- ベイズの定理より、

$$\begin{aligned} p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W}) &\propto p(\mu_c, \mu_t) \cdot f(\mathbf{Y}^{obs}, \mathbf{W} | \mu_c, \mu_t) \\ &= p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W}) \propto p(\mu_c, \mu_t) \cdot \mathcal{L}(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W}) \end{aligned}$$

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- $f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs} | \mathbf{W}, \mu_c, \mu_t)$ より \mathbf{Y}_i^{obs} の周辺分布を求める

$$Y_i^{obs} | \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}((1 - W_i) \cdot \mu_c + W_i \cdot \mu_t, (1 - W_i) \cdot 100 + W_i \cdot 64)$$

- よって尤度関数 \mathcal{L}_i は以下に比例する

$$\begin{aligned} \mathcal{L}_i &\propto \frac{1}{\sqrt{2\pi \cdot ((1 - W_i) \cdot 100 + W_i \cdot 64)}} \\ &\times \exp \left[-\frac{1}{2} \left(\frac{1}{(1 - W_i) \cdot 100 + W_i \cdot 64} (Y_i^{obs} - (1 - W_i) \cdot \mu_c - W_i \cdot \mu_t)^2 \right) \right] \end{aligned}$$

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- よって、

$$\mathcal{L}(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$$

$$\begin{aligned} & \propto \prod_{i=1}^6 \left\{ \frac{1}{\sqrt{2\pi \cdot ((1 - W_i) \cdot 100 + W_i \cdot 64)}} \right. \\ & \quad \times \exp \left[-\frac{1}{2} \left(\frac{1}{(1 - W_i) \cdot 100 + W_i \cdot 64} (Y_i^{obs} - (1 - W_i) \cdot \mu_c - W_i \cdot \mu_t)^2 \right) \right] \Big\} \\ & \propto \prod_{i:W_i=0} \frac{1}{\sqrt{2\pi \cdot 100}} \exp \left[-\frac{1}{2} \left(\frac{1}{100} (Y_i^{obs} - \mu_c)^2 \right) \right] \\ & \quad \times \prod_{i:W_i=1} \frac{1}{\sqrt{2\pi \cdot 64}} \exp \left[-\frac{1}{2} \left(\frac{1}{64} (Y_i^{obs} - \mu_t)^2 \right) \right]. \end{aligned}$$

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- 上で $p(\mathbf{W} | \mu_c, \mu_t)$ は既知の定数なので、無視している
 - $\mathcal{L} = f(\mathbf{Y}^{obs}, \mathbf{W} | \theta) = f(\mathbf{Y}^{obs} | \mathbf{W}, \theta) \cdot p(\mathbf{W} | \theta)$

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- これを用いると、 (μ_c, μ_t) の事後分布は、

$$p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W}) \propto$$

$$\exp \left[-\frac{1}{2} \left(\frac{\mu_c^2}{10,000} \right) \right] \cdot \prod_{i:W_i=0} \frac{1}{\sqrt{2\pi \cdot 100}} \exp \left[-\frac{1}{2} \left(\frac{(Y_i^{obs} - \mu_c)^2}{100} \right) \right]$$
$$\times \exp \left[-\frac{1}{2} \left(\frac{\mu_t^2}{10,000} \right) \right] \cdot \prod_{i:W_i=1} \frac{1}{\sqrt{2\pi \cdot 64}} \exp \left[-\frac{1}{2} \left(\frac{(Y_i^{obs} - \mu_t)^2}{64} \right) \right].$$

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- 頑張って計算すると、以下のような正規分布に比例することが分かる

$$\begin{aligned} & \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \mathbf{Y}^{obs}, \mathbf{W} \\ & \sim \mathcal{N} \left(\begin{pmatrix} \bar{Y}_c^{obs} \cdot \frac{N_c \cdot 10,000}{N_c \cdot 10,000 + 100} \\ \bar{Y}_t^{obs} \cdot \frac{N_t \cdot 10,000}{N_t \cdot 10,000 + 64} \end{pmatrix}, \begin{pmatrix} \frac{1}{N_c/100 + 1/10,000} & 0 \\ 0 & \frac{1}{N_t/64 + 1/10,000} \end{pmatrix} \right) \end{aligned}$$

- この計算はいつもこのように解析的に綺麗な形になるとは限らない

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- 例を再掲する。

Table 8.2. First Six Observations from NSW Program Data

Unit	Potential Outcomes		Treatment W_i	Observed Outcome Y_i^{obs}
	$Y_i(0)$	$Y_i(1)$		
1	0	?	0	0
2	?	9.9	1	9.9
3	12.4	?	0	12.4
4	?	3.6	1	3.6
5	0	?	0	0
6	?	24.9	1	24.9

Note: Question marks represent missing potential outcomes.

4. 具体例

ステップ2: 事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- $\bar{Y}_c^{obs} = 4.1, \bar{Y}_t^{obs} = 12.8, N_c = 3, N_t = 3$ を代入すると、

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \mathbf{Y}^{obs}, \mathbf{W} \sim \mathcal{N} \left(\begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

- 見ての通り、推論は μ の事前分布にほぼ左右されていない
 - (μ の平均に関しては、少数第一位までなら観測平均と等しい)
 - potential outcomes の分散の逆数で、観測データを weighting しているため

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- 前のステップで求めた $\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t$ と
 $(\mu_c, \mu_t) | \mathbf{Y}^{obs}, \mathbf{W}$ を組み合わせる
 - $\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}$ を得たい
- 両者はガウス分布なので、その同時分布 $\mathbf{Y}^{mis}, \mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W}$ もガウス分布になる
 - --> 周辺分布の $\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}$ もガウス分布になる
 - --> 一次と二次のモーメントを求めれば $\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}$ の分布が求まる

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- ステップ 1 で求めた $\mathbf{Y}_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t$ より、

$$\mathbb{E} [\mathbf{Y}_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t] = W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t$$

- ステップ 2 より、

$$\mathbb{E} \left[\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \mathbf{Y}^{obs}, \mathbf{W} \right] = \begin{pmatrix} \bar{Y}_c^{obs} \cdot \frac{N_c \cdot 10,000}{N_c \cdot 10,000 + 100} \\ \bar{Y}_t^{obs} \cdot \frac{N_t \cdot 10,000}{N_t \cdot 10,000 + 64} \end{pmatrix}$$

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- よって、

$$\begin{aligned}\mathbb{E} [\mathbf{Y}_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}] &= W_i \cdot \left(\bar{Y}_c^{obs} \cdot \frac{N_c \cdot 10,000}{N_c \cdot 10,000 + 100} \right) \\ &\quad + (1 - W_i) \cdot \left(\bar{Y}_t^{obs} \cdot \frac{N_t \cdot 10,000}{N_t \cdot 10,000 + 64} \right)\end{aligned}$$

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- 次に二次のモーメントを求める。繰り返し期待値の法則 (tower rule) より、

$$\begin{aligned}\mathbb{V}(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}) &= \mathbb{E} \left[\mathbb{V}(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t) \mid \mathbf{Y}^{obs}, \mathbf{W} \right] \\ &\quad + \mathbb{V} \left(\mathbb{E} \left[Y_i^{mis} \mid \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t \right] \mid \mathbf{Y}^{obs}, \mathbf{W} \right) \\ &= \mathbb{E} \left[W_i \cdot 100 + (1 - W_i) \cdot 64 \mid \mathbf{Y}^{obs}, \mathbf{W} \right] + \mathbb{V} \left(W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t \mid \mathbf{Y}^{obs}, \mathbf{W} \right) \\ &= W_i \cdot 100 + (1 - W_i) \cdot 64 + W_i \cdot \frac{1}{N_c/100 + 1/10,000} + (1 - W_i) \cdot \frac{1}{N_t/64 + 1/10,000} \\ &= W_i \cdot \left(100 + \frac{1}{N_c/100 + 1/10,000} \right) + (1 - W_i) \cdot \left(64 + \frac{1}{N_t/64 + 1/10,000} \right).\end{aligned}$$

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

軽く以下を証明する

$$V(X) = E[V(X|Y)] + V[E(X|Y)]$$

右辺の一項目と二項目は、

$$\begin{aligned} E[V(X|Y)] &= E[E[X^2|Y] - E[X|Y]^2] \\ &= E[E[X^2|Y]] - E[E[X|Y]^2] \\ &= E[X^2] - E[E[X|Y]^2] \end{aligned}$$

$$\begin{aligned} V[E[X|Y]] &= E[E[X|Y]^2] - E[E[X|Y]]^2 \\ &= E[E[X|Y]^2] - E[X]^2 \end{aligned}$$

であり、後はこれらを足し合わせると、 X の分散になる(左辺と一致する)。

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

$Y_i^{mis}, Y_{i'}^{mis}$ の共分散も考慮しなければならない ($i \neq i'$)

$$\begin{aligned}\mathbb{C}(Y_i^{mis}, Y_{i'}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}) &= \mathbb{E} \left[\mathbb{C}(Y_i^{mis}, Y_{i'}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t) \mid \mathbf{Y}^{obs}, \mathbf{W} \right] \\ &\quad + \mathbb{C} \left(\mathbb{E}[Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t], \mathbb{E}[Y_{i'}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t] \mid \mathbf{Y}^{obs}, \mathbf{W} \right) \\ &= 0 + \mathbb{C}(W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t, W_{i'} \cdot \mu_c + (1 - W_{i'}) \cdot \mu_t \mid \mathbf{Y}^{obs}, \mathbf{W}) \\ &= W_i \cdot W_j \cdot \frac{1}{N_c/100 + 1/10,000} + (1 - W_i) \cdot (1 - W_j) \cdot \frac{1}{N_t/64 + 1/10,000}.\end{aligned}$$

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

一次と二次のモーメント、そして共分散を踏まえると、 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ は、

$$\left(\begin{array}{c} Y_1^{mis} \\ Y_2^{mis} \\ Y_3^{mis} \\ Y_4^{mis} \\ Y_5^{mis} \\ Y_6^{mis} \end{array} \right) \mid \mathbf{Y}^{obs}, \mathbf{W} \sim \mathcal{N} \left(\begin{pmatrix} 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \end{pmatrix}, \begin{pmatrix} 85.3 & 0 & 21.3 & 0 & 21.3 & 0 \\ 0 & 133.2 & 0 & 33.2 & 0 & 33.2 \\ 21.3 & 0 & 85.3 & 0 & 21.3 & 0 \\ 0 & 0 & 0 & 133.2 & 0 & 33.2 \\ 21.3 & 0 & 21.3 & 0 & 85.3 & 0 \\ 0 & 33.2 & 0 & 33.2 & 0 & 133.2 \end{pmatrix} \right).$$

4. 具体例

ステップ3: 事後分布 $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- 欠損値はもはや独立でないことに注意する（共分散がある）。
 - パラメータ(μ_c, μ_t)で条件付けると、独立
 - しかし、データが共通のパラメータに依存するという構造が、データ間自体に依存関係をもたらしている

4. 具体例

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- 今回は SATE (Sample Average Treatment Effect) に興味がある。
 - つまり標本内での ATE に興味がある。

$$\tau_{fs} = \tau(\mathbf{Y}(0), \mathbf{Y}(1)) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)).$$

- $(Y_i(0), Y_i(1))$ を (Y_i^{mis}, Y_i^{obs}) で書き換えると、

$$\tau_{fs} = \tau(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W}) = \frac{1}{N} \sum_{i=1}^N (1 - 2 \cdot W_i) \cdot Y_i^{mis} + \sum_{i=1}^N (2 \cdot W_i - 1) \cdot Y_i^{obs}$$

4. 具体例

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- $(\mathbf{Y}^{obs}, \mathbf{W})$ で条件付けた時、確率的な要素は Y_i^{mis} のみになる。
- τ_{fs} は $Y_1^{mis}, \dots, Y_6^{mis}$ の線形結合であり、 Y_i^{mis} は正規分布なので、 τ_{fs} も正規分布に従う。
 - なので、期待値と分散が分かれば分布が分かる。

4. 具体例

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

$$\begin{aligned}\mathbb{E}[\tau_{fs} | \mathbf{Y}^{obs}, \mathbf{W}] &= \frac{1}{N} \sum_{i=1}^N (2 \cdot W_i - 1) \cdot Y_i^{obs} + \frac{1}{N} \sum_{i=1}^N (1 - 2 \cdot W_i) \cdot \mathbb{E}[Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}] \\ &= \frac{N_t}{N} \cdot \bar{Y}_t^{obs} - \frac{N_c}{N} \cdot \bar{Y}_c^{obs} \\ &\quad + \frac{1}{N} \sum_{i=1}^N (1 - 2 \cdot W_i) \cdot \left(W_i \cdot \left(\bar{Y}_c^{obs} \cdot \frac{N_c \cdot 10,000}{N_c \cdot 10,000 + 100} \right) \right. \\ &\quad \left. + (1 - W_i) \cdot \left(\bar{Y}_t^{obs} \cdot \frac{N_t \cdot 10,000}{N_t \cdot 10,000 + 64} \right) \right) \\ &= \bar{Y}_t^{obs} \cdot \frac{N_t \cdot 10,000 + 64 \cdot N_t/N}{N_t \cdot 10,000 + 64} - \bar{Y}_c^{obs} \cdot \frac{N_c \cdot 10,000 + 100 \cdot N_c/N}{N_c \cdot 10,000 + 100}.\end{aligned}$$

4. 具体例

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

$$\begin{aligned}\mathbb{V}(\tau_{fs} | \mathbf{Y}^{obs}, \mathbf{W}) &= \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}((1 - 2 \cdot W_i) \cdot Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}) \\ &\quad + \frac{1}{N^2} \sum_{i=1}^N \sum_{i' \neq i} \mathbb{C}((1 - 2 \cdot W_i) \cdot Y_i^{mis}, (1 - 2 \cdot W_{i'}) \cdot Y_{i'}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}) \\ &= \frac{1}{N^2} \left(N_t \cdot \left(100 + \frac{1}{N_c/100 + 1/10,000} \right) + N_c \cdot \left(64 + \frac{1}{N_t/64 + 1/10,000} \right) \right) \\ &\quad + \frac{1}{N^2} \left(N_t \cdot (N_t - 1) \cdot \frac{1}{N_c/100 + 1/10,000} + N_c \cdot (N_c - 1) \cdot \frac{1}{N_t/64 + 1/10,000} \right).\end{aligned}$$

4. 具体例

ステップ4: Estimand の事後分布 $f(\tau | \mathbf{Y}^{obs}, \mathbf{W})$ の導出

- $(N = 6, N_c = 3, N_t = 3)$ を代入すると、

$$\tau_{fs} | \mathbf{Y}^{obs}, \mathbf{W} \sim \mathcal{N}(8.7, 5.2^2)$$

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4 ステップで ATE 推測
4. 具体例
5. シミュレーションベースで再計算
6. 潜在的結果変数に従属関係があった場合
7. 無限超母集団における ATE の推測
8. 実データで求めてみよう

5. シミュレーションベース

- いつも事後分布の解析解が弾き出せるわけではない。
- もしくは、弾き出せても非現実的な場合もある。
- --> シミュレーションで求めてみよう

5. シミュレーションベース

- シミュレーションをするには、二つの重要な要素がある
 - 観測値とパラメータで条件付けた欠測値の分布（ステップ 1 で導出） $f(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$
 - 観測値が与えられた上でのパラメータの事後分布（ステップ 2 で導出） $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$

5. シミュレーションベース

- 以下の分布から、 $(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ の組、 $(\mu_c^{(1)}, \mu_t^{(1)})$ を引く。

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \mathbf{Y}^{obs}, \mathbf{W} \sim \mathcal{N} \left(\begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

- $(\mu_c^{(1)}, \mu_t^{(1)})$ を用いて、 $f(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ に代入する

$$\left(\begin{array}{c} Y_1^{mis} \\ Y_2^{mis} \\ Y_3^{mis} \\ Y_4^{mis} \\ Y_5^{mis} \\ Y_6^{mis} \end{array} \right) \mid \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N} \left(\begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right).$$

5. シミュレーションベース

- 今回 $(\mu_c^{(1)}, \mu_t^{(1)}) = (1.63, 5.09)$ だったとすると、

$$\left(\begin{array}{c} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{array} \right) \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}, \boldsymbol{\theta} \sim \mathcal{N} \left(\begin{pmatrix} 5.09 \\ 1.63 \\ 5.09 \\ 1.63 \\ 5.09 \\ 1.63 \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

- この条件付き分布から Y_i^{mis} を引いて、それらを用いて $\hat{\tau}$ を求める

5. シミュレーションベース

- 今回は $Y_1^{mis}, \dots, Y_6^{mis}$ は以下の表のようになつた。

Unit	Potential Outcomes		Treatment W_i	Observed Outcome Y_i^{obs}
	$Y_i(0)$	$Y_i(1)$		
Panel A: First Parameter Draw $(\mu_c^{(1)}, \mu_t^{(1)}) = (1.63, 5.09)$				
1	0	(6.1)	0	0
2	(13.5)	9.9	1	9.9
3	12.4	(7.4)	0	12.4
4	(13.5)	3.6	1	3.6
5	0	(-4.1)	0	0
6	(1.3)	24.9	1	24.9
Average	6.8	8.0		
$\tau_{fs}^{(1)}$		1.2		

- この表から $\hat{\tau}_{fs}^{(1)}$ を求めると、1.2 となる。

5. シミュレーションベース

- まとめると、二段階の抽出をしている
 - $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ から抽出
 - $f(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ から抽出
- 同じことをもう一度行うと次の表のようになる。

5. シミュレーションベース

Panel B: Second Parameter Draw $(\mu_c^{(2)}, \mu_t^{(2)}) = (6.01, 13.58)$

1	0	(12.1)	0	0
2	(27.8)	9.9	1	9.9
3	12.4	(19.4)	0	12.4
4	(4.6)	3.6	1	3.6
5	0	(8.9)	0	0
6	(7.1)	24.9	1	24.9
Average	8.7	13.1		
$\tau_{fs}^{(2)}$		4.5		

- ここでは $(\mu_c^{(2)}, \mu_t^{(2)}) = (6.01, 13.58)$, Y_i^{mis} は以上のようになった。
- $\hat{\tau}_{fs}^{(2)}$ を求めると、4.5 となる。

5. シミュレーションベース

- この二段階の抽出を N_R 回繰り返すと、 N_R 個の $\hat{\tau}_{fs}^{(r)}$ が求まる。
 - その平均を取れば、 τ_{fs} の期待値の推定量 $\hat{\tau}$ が求まる
- 同様に、推定量 τ_{fs} の標準偏差の推定量 $\hat{\sigma}_{\hat{\tau}_{fs}}$ も求まる
- 今回は、それぞれ 8.6, 5.3 だった。
 - 解析解が 8.7, 5.2 であったことを踏まえると悪くない近似である。

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4 ステップで ATE 推測
4. 具体例
5. シミュレーションベースで再計算
6. **潜在的結果変数に従属関係があった場合**
7. 無限超母集団における ATE の推測
8. 実データで求めてみよう

6. 潜在的結果変数に従属関係があった場合

- モデルベースのアプローチで最も重大な仮定は、 $f(Y_i(0), Y_i(1)|\theta)$ の特定である。
- 今まででは、二変量正規分布の共分散が既知($= 0$)であると仮定していたが、共分散も考慮しよう。

$$f(Y_i(0), Y_i(1)|\theta) \sim \mathcal{N} \left(\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & \rho\sigma_c\sigma_t \\ \rho\sigma_c\sigma_t & \sigma_t^2 \end{pmatrix} \right)$$

- 今までと違い、パラメータベクトルは $\theta = (\mu_c, \mu_t, \sigma_c^2, \sigma_t^2, \rho)'$

6. 潜在的結果変数に従属関係があった場合

- この状況下で、 (\mathbf{W}, θ) を与えた時の Y_i^{obs} の条件付き分布は、

$$f(Y_i^{obs} | \mathbf{W}, \theta) = \frac{1}{\sqrt{2\pi \cdot ((1 - W_i) \cdot \sigma_c^2 + W_i \cdot \sigma_t^2)}} \\ \times \exp \left[-\frac{1}{2} \left(\frac{(Y_i^{obs} - (1 - W_i) \cdot \mu_c - W_i \cdot \mu_t)^2}{(1 - W_i) \cdot \sigma_c^2 + W_i \cdot \sigma_t^2} \right) \right],$$

6. 潜在的結果変数に従属関係があった場合

- よって、尤度関数は、

$$\begin{aligned}\mathcal{L}(\mu_c, \mu_t, \sigma_c^2, \sigma_t^2, \rho | \mathbf{Y}^{\text{obs}}, \mathbf{W}) &= \prod_{i=1}^6 \frac{1}{\sqrt{2\pi \cdot ((1 - W_i) \cdot \sigma_c^2 + W_i \cdot \sigma_t^2)}} \\ &\times \exp \left[-\frac{1}{2} \left(\frac{1}{(1 - W_i) \cdot \sigma_c^2 + W_i \cdot \sigma_t^2} (Y_i^{\text{obs}} - (1 - W_i) \cdot \mu_c - W_i \cdot \mu_t)^2 \right) \right].\end{aligned}$$

- --> 尤度関数は今までと変わらない
- --> つまり、データは潜在的変数間の相関について何の情報も持たない

6. 潜在的結果変数に従属関係があった場合

- ここで、事前分布が以下のように分解できるとする

$$\theta = p(\rho)p(\mu_c, \mu_t, \sigma_c^2, \sigma_t^2).$$

- 尤度関数が ρ に依存しなかったことから、 ρ の事後分布は事前分布と同一になることが示唆される。
 - 二つの潜在的変数を同時に観測することはあり得ないのだから、依存関係の経験的情報が無いのは、さして不思議なことではないはず。

6. 潜在的結果変数に従属関係があった場合

- 仮定の差異が生む帰結を確認するために、3, 4章と同じ手順で処置効果を求める。
- 新たに $\rho = 1$ と仮定し、

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} | \theta \sim \mathcal{N} \left(\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 80 \\ 80 & 64 \end{pmatrix} \right)$$

- 3, 4 章の手順と同様に $f(Y_i^{mis}, Y_i^{obs} | \mathbf{W}, \mu_c, \mu_t)$ を求めると、

$$\begin{pmatrix} Y_i^{mis} \\ Y_i^{obs} \end{pmatrix} | \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N} \left(\begin{pmatrix} W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t \\ (1 - W_i) \cdot \mu_c + W_i \cdot \mu_t \end{pmatrix}, \begin{pmatrix} W_i \cdot 100 + (1 - W_i) \cdot 64 & 80 \\ 80 & (1 - W_i) \cdot 100 + W_i \cdot 64 \end{pmatrix} \right)$$

6. 潜在的結果変数に従属関係があった場合

- よって、条件付き分布 $f(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ は、

$$Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N} \left(W_i \cdot \left(\mu_c + \frac{80}{64} \cdot (Y_i^{obs} - \mu_t) \right) + (1 - W_i) \cdot \left(\mu_t + \frac{80}{100} \cdot (Y_i^{obs} - \mu_c) \right), 0 \right)$$

- 分散が 0 なのは、 $\rho = 1$ としているため Y_i^{obs} が分かれば Y_i^{mis} が 100% 分かるため

6. 潜在的結果変数に従属関係があった場合

- 3, 4 章と同じく、知りたいのは $f(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t)$ ではなく $f(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W})$ であった。
- (μ_c, μ_t) の事後分布 $p(\mu_c, \mu_t | \mathbf{Y}^{obs}, \mathbf{W})$ が必要
- Y_i^{obs} の条件付き分布を思い出すと、

$$f(Y_i^{obs} | \mathbf{W}, \theta) = \frac{1}{\sqrt{2\pi \cdot ((1 - W_i) \cdot \sigma_c^2 + W_i \cdot \sigma_t^2)}} \\ \times \exp \left[-\frac{1}{2} \left(\frac{(Y_i^{obs} - (1 - W_i) \cdot \mu_c - W_i \cdot \mu_t)^2}{(1 - W_i) \cdot \sigma_c^2 + W_i \cdot \sigma_t^2} \right) \right],$$

- ρ に関係なく決定される。
- --> 尤度関数は ρ に関係なく決定される

6. 潜在的結果変数に従属関係があった場合

- もし $\theta (= \mu_c, \mu_t, \sigma_c, \sigma_t)$ に同じ事前分布を置いたとしたら (σ は恣意的に定める) 、

$$\begin{aligned} & \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \mathbf{Y}^{obs}, \mathbf{W} \\ & \sim \mathcal{N} \left(\begin{pmatrix} \bar{Y}_c^{obs} \cdot \frac{N_c \cdot 10,000}{N_c \cdot 10,000 + 100} \\ \bar{Y}_t^{obs} \cdot \frac{N_t \cdot 10,000}{N_t \cdot 10,000 + 64} \end{pmatrix}, \begin{pmatrix} \frac{1}{N_c/100 + 1/10,000} & 0 \\ 0 & \frac{1}{N_t/64 + 1/10,000} \end{pmatrix} \right) \end{aligned}$$

- 上記のように 3, 4 章と同様の事後分布が導かれる。

6. 潜在的結果変数に従属関係があった場合

- 今回、 Y_i^{mis} は μ_c, μ_t の線形関数なので、

$$Y_i^{mis} | Y_i^{obs}, W_i, \mu_c, \mu_t = W_i \cdot \left(\mu_c + \frac{80}{64} \cdot (Y_i^{obs} - \mu_t) \right) + (1 - W_i) \cdot \left(\mu_t + \frac{80}{100} \cdot (Y_i^{obs} - \mu_c) \right)$$

- (μ_c, μ_t) の正規性は、 Y_i^{mis} の正規性を示唆する
 - > 期待値、分散、共分散が分かれば、分布が分かる

6. 潜在的結果変数に従属関係があった場合

$$\begin{aligned}\mathbb{E} \left[Y_i^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right] &= W_i \cdot \left\{ \bar{Y}_{\text{c}}^{\text{obs}} \cdot \frac{N_{\text{c}} \cdot 10,000}{N_{\text{c}} \cdot 10,000 + 100} + \frac{80}{64} \right. \\ &\quad \left. \cdot \left(Y_i^{\text{obs}} - \bar{Y}_{\text{t}}^{\text{obs}} \cdot \frac{N_{\text{t}} \cdot 10,000}{N_{\text{t}} \cdot 10,000 + 64} \right) \right\} \\ &+ (1 - W_i) \cdot \left\{ \bar{Y}_{\text{t}}^{\text{obs}} \cdot \frac{N_{\text{t}} \cdot 10,000}{N_{\text{t}} \cdot 10,000 + 64} + \frac{80}{100} \cdot \left(Y_i^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}} \cdot \frac{N_{\text{c}} \cdot 10,000}{N_{\text{c}} \cdot 10,000 + 100} \right) \right\}\end{aligned}$$

6. 潜在的結果変数に従属関係があった場合

$$\begin{aligned} \mathbb{V}(Y_i^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}) &= W_i \cdot \left\{ \mathbb{V}(\mu_c) + \left(\frac{80}{64} \right)^2 \cdot \mathbb{V}(\mu_t) \right\} \\ &\quad + (1 - W_i) \cdot \left\{ \mathbb{V}(\mu_t) + \left(\frac{80}{100} \right)^2 \cdot \mathbb{V}(\mu_c) \right\} \\ &= W_i \cdot \left\{ \frac{1}{N_c/100 + 1/10,000} + \left(\frac{80}{64} \right)^2 \cdot \frac{1}{N_t/64 + 1/10,000} \right\} \\ &\quad + (1 - W_i) \cdot \left\{ \frac{1}{N_t/64 + 1/10,000} + \left(\frac{80}{100} \right)^2 \cdot \frac{1}{N_c/100 + 1/10,000} \right\} \end{aligned}$$

6. 潜在的結果変数に従属関係があった場合

$$\begin{aligned} \mathbb{C}(Y_i^{\text{mis}}, Y_{i'}^{\text{mis}} \mid \mathbf{Y}^{\text{obs}}, \mathbf{W}) &= W_i \cdot W_{i'} \\ &\cdot \left(\frac{1}{N_c/100 + 1/10,000} + \left(\frac{80}{64} \right)^2 \cdot \frac{1}{N_t/64 + 1/10,000} \right) \\ &- W_i \cdot (1 - W_{i'}) \cdot \left(\frac{80}{100} \cdot \frac{1}{N_c/100 + 1/10,000} + \frac{80}{64} \cdot \frac{1}{N_t/64 + 1/10,000} \right) \\ &- (1 - W_i) \cdot W_{i'} \cdot \left(\frac{80}{100} \cdot \frac{1}{N_c/100 + 1/10,000} + \frac{80}{64} \cdot \frac{1}{N_t/64 + 1/10,000} \right) \\ &+ (1 - W_i) \cdot (1 - W_{i'}) \cdot \left(\frac{1}{N_t/64 + 1/10,000} + \left(\frac{80}{100} \right)^2 \cdot \frac{1}{N_c/100 + 1/10,000} \right). \end{aligned}$$

6. 潜在的結果変数に従属関係があった場合

- 最終的に知りたいのは、この条件付き分布ではなく、Estimand の条件付き分布であった。
- 今回の我々の Estimand は SATE なので（サンプルでの ATE）、

$$\begin{aligned}\tau_{fs} &= \frac{1}{N} \sum_{i=1}^N (2 \cdot W_i - 1) \cdot (Y_i^{obs} - Y_i^{mis}) \\ &= \frac{1}{N} \sum_{i=1}^N (2 \cdot W_i - 1) \cdot Y_i^{obs} - \frac{1}{N} \sum_{i=1}^N (2 \cdot W_i - 1) \cdot Y_i^{mis}\end{aligned}$$

- > $\tau_{fs} | \mathbf{Y}^{obs}, \mathbf{W}$ はガウス分布になる

6. 潜在的結果変数に従属関係があった場合

- $f(\tau_{\text{fs}} | \mathbf{Y}^{\text{obs}}, \mathbf{W})$ の平均は、

$$\begin{aligned}\mathbb{E} [\tau_{\text{fs}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}] &= \frac{1}{N} \sum_{i=1}^N (2 \cdot W_i - 1) \cdot Y_i^{\text{obs}} + \frac{1}{N} \sum_{i=1}^N (1 - 2 \cdot W_i) \cdot \mathbb{E} [Y_i^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}] \\ &= \bar{Y}_{\text{t}}^{\text{obs}} \cdot \frac{N_{\text{t}} \cdot 1000 - 16 \cdot N_t/N}{N_{\text{t}} \cdot 1000 + 64} - \bar{Y}_{\text{c}}^{\text{obs}} \cdot \frac{N_{\text{c}} \cdot 1000 + 20 \cdot N_c/N}{N_{\text{c}} \cdot 1000 + 100}.\end{aligned}$$

$$\begin{aligned}
\mathbb{V}(\tau_{\text{fs}} | \mathbf{Y}^{obs}, \mathbf{W}) &= \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}(Y_i^{mis} | \mathbf{Y}^{obs}, \mathbf{W}) + \frac{1}{N^2} \sum_{i=1}^N \sum_{i' \neq i} \mathbb{C}(Y_i^{mis}, Y_{i'}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}) \\
&= \frac{N_t}{N^2} \cdot \left\{ \frac{1}{N_c/100 + 1/10,000} + \left(\frac{80}{64} \right)^2 \cdot \frac{1}{N_t/64 + 1/10,000} \right\} \\
&\quad + \frac{N_c}{N^2} \cdot \left\{ \frac{1}{N_t/64 + 1/10,000} + \left(\frac{80}{100} \right)^2 \cdot \frac{1}{N_c/100 + 1/10,000} \right\} \\
&\quad + \frac{N_t \cdot (N_t - 1)}{N^2} \cdot \left\{ \frac{1}{N_c/100 + 1/10,000} + \left(\frac{80}{64} \right)^2 \cdot \frac{1}{N_t/64 + 1/10,000} \right\} \\
&\quad - \frac{2 \cdot N_c \cdot N_t}{N^2} \cdot \left(\frac{80}{100} \cdot \frac{1}{N_c/100 + 1/10,000} + \frac{80}{64} \cdot \frac{1}{N_t/64 + 1/10,000} \right) \\
&\quad + \frac{N_c \cdot (N_c - 1)}{N^2} \cdot \left\{ \frac{1}{N_t/64 + 1/10,000} + \left(\frac{80}{100} \right)^2 \cdot \frac{1}{N_c/100 + 1/10,000} \right\}
\end{aligned}$$

6. 潜在的結果変数に従属関係があった場合

- 今回のデータセットの値を挿入すると、

$$\tau_{\text{fs}} | \mathbf{Y}^{obs}, \mathbf{W} \sim \mathcal{N}(8.7, 7.7^2)$$

- 潜在的結果変数間の独立性の仮定と比べると、平均はほぼ同じだが、標準偏差は大きくなつた。
- 今回の章のメインポイントは、潜在的変数間の相関のパラメータは、モデル内の他のパラメータと異なるという点
 - 欠測値という性質上、データが経験的情報を何ら持たないため。
 - <-- \mathbf{X} を与えた時の $\mathbf{Y}(0), \mathbf{Y}(1)$ の条件付き関係性を制御するパラメータに関する経験的情報

6. 潜在的結果変数に従属関係があった場合

- 相関パラメータが異なるという点は、一つの疑問を残す
- どのようにモデルに組み込むべきか?
 - 保守的に独立である ($\rho = 0$) とモデルを組むと悪いケースも
 - 一方、今回のように $\rho = 1$ とモデルを組んでも最悪な仮定であるケースも
 - 事後分布の分散の観点から鑑みて
- 欠測値を統制群の値から代入した値を見た後に、処置を考えるというふうに、処置構造を動的に変更して汚染することを恐れる結果、やむなし PO 間の独立を仮定するケースもある。

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4 ステップで ATE 推測
4. 具体例
5. シミュレーションベースで再計算
6. 潜在的結果変数に従属関係があった場合
7. **無限超母集団における ATE の推測**
8. 実データで求めてみよう

7. 無限超母集団における ATE

- 今まで SATE を考えてきた
 - 有限母集団の中の ATE
 - $\tau_{fs} = \sum_{i=1}^N (Y_i(1) - Y_i(0))/N$
- ここからは観測が無限超母集団からのランダムサンプルだとし、興味のある ATE は、その無限母集団の ATE (PATE) だとする。
 - $\tau_{sp} = \mathbb{E}_{sp}[Y_i(1) - Y_i(0)]$

7. 無限超母集団における ATE

- τ_{sp} は単純にパラメータで表せることもある
 - 例：普通の線形モデルの場合
 - $\tau_{\text{sp}} = \tau(\theta) = \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0)|\theta] = \mu_t - \mu_c$
- 一般的には、PATE は、潜在的結果変数の同時分布のモデル化を通して定義できる
 - $\tau(\theta) = \int \int (y(1) - y(0)) f(y(1), y(0)|\theta) dy(1) dy(0)$

7. 無限超母集団における ATE

- 線形モデルは PATE をストレートに表せる
 - シミュレーションでやったように、 $\theta^{(1)}$ を引いてくる
 - $f(\mathbf{Y}^{mis} | \mathbf{Y}^{obs}, \mathbf{W}, \theta^{(1)})$ から引く代わりに、単純に PATE を求める
 - $\tau^{(1)} = \tau(\theta^{(1)})$
 - N_R 回繰り返し、 $\{\hat{\tau}_{sp}^{(r)}, r = 1, \dots, N_R\}$ を得る
 - Estimand の性質を調べられる（平均、標準偏差等）

7. 無限超母集団における ATE

- 6 件の例を用いて PATE を求めてみる

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \mathbf{Y}^{obs}, \mathbf{W} \sim \mathcal{N} \left(\begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

- だったので、

$$\begin{aligned} \mu_t - \mu_c | \mathbf{Y}^{obs}, \mathbf{W} &\sim \mathcal{N} ((12.8 - 4.1), (33.2 + 21.3 + 2 \cdot 0)) \\ &\sim \mathcal{N}(8.7, 7.4^2) \end{aligned}$$

7. 無限超母集団における ATE

- 期待値は τ_{fs} と一致する
- PO 間で独立を仮定した時の SATE の標準偏差 ($= 5.2$) よりも大きく、PO が完全に相関していると仮定を置いた時の SATE の標準偏差 ($= 7.7$) に近い。
- PATE を求めるときは、 ρ の仮定は全く重要でないことに注意

目次

1. 因果推論の根本問題の単純な解決策
2. Y^{mis} の条件付き分布が知りたい
3. 4 ステップで ATE 推測
4. 具体例
5. シミュレーションベースで再計算
6. 潜在的結果変数に従属関係があった場合
7. 無限超母集団における ATE の推測
8. 実データで求めてみよう

8. データで求める

- notebook へ移動