0217report

概要

誤分類を含む二値アウトカムに対する、因果効果の推定量に関する手法のシュミレーションを行った結果である。

モデル化の設定

ターゲット分布の潜在結果を以下のように周辺構造モデルを用いて考える。

$E(Y|T,Z)$ を，以下のようにモデル化する

$$E[Y^{(t)} \mid Z] = E\left[\frac{T}{e^{(t)}} Y \mid Z\right] = \pi(T,Z;\beta) = \mathbf{expit}\{t\beta_t + z^T\beta_z\} \tag{1}$$

ここで，$X = (Z, \tilde{Z})$ とし，Z を共変量全体の中でモデルに必要な共変量とする。

また $e^{(t)}$ は傾向スコア $\Pr(T = t \mid X)$ である。

データ生成過程

ここでは、共変量 X として考える。Z に変更した場合の仮定は以前の仮定の整理を確認。

ベルヌーイ分布に従う Y のコンタミのないの分布

$$f(y^{(t)}|x,t;\beta) = \{\pi(t,x)\}^{y^{(t)}}\{1-\pi(t,x)\}^{1-y^{(t)}} \tag{2}$$

コンタミがある場合の，データ生成過程の設定

$$g(Y=y|X=x,T=t) = c(x)f(y|x,t;\beta) + (1-c(x))h(y|x,t)) \tag{3}$$

$$\eta_0(x) = P(Y=1|Y_{true}=0, X=x) \tag{4}$$
$$\eta_1(x) = P(Y=0|Y_{true}=1, X=x) \tag{5}$$

- 真の分布：$f(y|x,t)$
- コンタミの割合：$1-c(x) = \eta_0(x) + \eta_1(x)$
- コンタミの分布：$h(y|x,t)$

シュミレーション設定

共変量

$$XX_i \sim N(0_9, \Sigma) \tag{6}$$

where,

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{7}$$

処置

$$TT_i \sim Bin(p) \tag{8}$$

$$\mathbf{lm} = 0.2 + XX_{i1} + XX_{i2} + XX_{i3} + XX_{i4} + XX_{i5} + XX_{i6} + XX_{i7} + XX_{i8} + XX_{i9} \tag{9}$$

$$p = \mathbf{expit}(\mathbf{lm}) \tag{10}$$

潜在アウトカム

$$YY_i^{(1)} \sim Bin(p1), \quad YY_i^{(0)} \sim Bin(p0) \tag{11}$$

$$\mathbf{lm1} = 0.2 + 1 + 1 * XX_{i1} - 1 * XX_{i2} + 1 * XX_{i3} \tag{12}$$

$$\mathbf{lm0} = 0.2 + 1 * XX_{i1} - 1 * XX_{i2} + 1 * XX_{i3} \tag{13}$$

$$p1 = \mathbf{expit}(\mathbf{lm1}) \tag{14}$$

$$p0 = \mathbf{expit}(\mathbf{lm0}) \tag{15}$$

つまり、$Z = (X_1, X_2, X_3)$

観測アウトカム

$$YY_i = TT_i * YY_i^{(1)} + (1 - TT_i) * YY_i^{(0)} \tag{16}$$

誤判別の割合

$$\eta_0 = 0.1,\ \eta_1 = 0.05 \tag{17}$$

シュミレーション結果

サンプルサイズ、シュミレーション回数はともに 500 である。

表1: 真の値

t	b0	b1	b2	b3
1	0.2	1	-1	1

誤判別がないとき

- $\gamma = 0.00001$ のとき

```
##        t                b0                   b1                b2
##  Min.   :0.1460   Min.   :-0.29423   Min.   :0.5282   Min.   :-1.4793
##  1st Qu.:0.7952   1st Qu.: 0.09464   1st Qu.:0.9202   1st Qu.:-1.0999
##  Median :1.0039   Median : 0.20551   Median :1.0130   Median :-0.9970
##  Mean   :0.9921   Mean   : 0.21005   Mean   :1.0175   Mean   :-1.0081
##  3rd Qu.:1.1666   3rd Qu.: 0.32282   3rd Qu.:1.1092   3rd Qu.:-0.9197
##  Max.   :1.9920   Max.   : 0.75937   Max.   :1.4117   Max.   :-0.6754
##        b3
##  Min.   :0.5848
##  1st Qu.:0.9198
##  Median :1.0185
##  Mean   :1.0194
##  3rd Qu.:1.1059
##  Max.   :1.5137
```

- $\gamma = 1$ のとき

```
##        t                b0                   b1                b2
##  Min.   :0.1873   Min.   :-0.40779   Min.   :0.5394   Min.   :-1.4899
##  1st Qu.:0.7925   1st Qu.: 0.07247   1st Qu.:0.8857   1st Qu.:-1.0946
##  Median :0.9849   Median : 0.19884   Median :0.9977   Median :-0.9947
##  Mean   :0.9790   Mean   : 0.21253   Mean   :1.0039   Mean   :-0.9991
##  3rd Qu.:1.1600   3rd Qu.: 0.33101   3rd Qu.:1.1128   3rd Qu.:-0.8829
##  Max.   :1.8881   Max.   : 0.84838   Max.   :1.6355   Max.   :-0.5618
##        b3
##  Min.   :0.4886
##  1st Qu.:0.8996
##  Median :1.0038
##  Mean   :1.0097
##  3rd Qu.:1.1101
##  Max.   :1.7208
```

誤判別があるとき

誤判別の設定は、$\eta_0 = 0.1$, $\eta_1 = 0.05$ である

- $\gamma = 0.00001$ のとき

```
##          t                   b0                 b1                 b2
##  Min.   :-0.09453   Min.   :-0.2390   Min.   :0.3618   Min.   :-1.0743
##  1st Qu.: 0.59615   1st Qu.: 0.1699   1st Qu.:0.6569   1st Qu.:-0.8240
##  Median : 0.76300   Median : 0.2850   Median :0.7403   Median :-0.7390
##  Mean   : 0.75771   Mean   : 0.2756   Mean   :0.7404   Mean   :-0.7447
##  3rd Qu.: 0.90572   3rd Qu.: 0.3812   3rd Qu.:0.8177   3rd Qu.:-0.6564
##  Max.   : 1.37592   Max.   : 0.7541   Max.   :1.2435   Max.   :-0.3625
##          b3
##  Min.   :0.3337
##  1st Qu.:0.6602
##  Median :0.7367
##  Mean   :0.7451
##  3rd Qu.:0.8283
##  Max.   :1.2643
```

- $\gamma = 1$ のとき

```
##          t                  b0                 b1                 b2
##  Min.   :0.02884   Min.   :-0.1692   Min.   :0.3246   Min.   :-1.2461
##  1st Qu.:0.63203   1st Qu.: 0.1481   1st Qu.:0.7122   1st Qu.:-0.8820
##  Median :0.80964   Median : 0.2854   Median :0.8028   Median :-0.7920
##  Mean   :0.81239   Mean   : 0.2742   Mean   :0.8105   Mean   :-0.7967
##  3rd Qu.:0.99609   3rd Qu.: 0.3898   3rd Qu.:0.8977   3rd Qu.:-0.6940
##  Max.   :1.74617   Max.   : 0.8643   Max.   :1.2832   Max.   :-0.4523
##          b3
##  Min.   :0.4801
##  1st Qu.:0.7076
##  Median :0.7954
##  Mean   :0.8116
##  3rd Qu.:0.9117
##  Max.   :1.2935
```

- $\gamma = 2$ のとき

```
##          t                  b0                 b1                 b2
```

```
##  Min.   :0.04313   Min.   :-0.2801   Min.   :0.4078   Min.   :-1.7998
##  1st Qu.:0.62166   1st Qu.: 0.1590   1st Qu.:0.7419   1st Qu.:-0.9575
##  Median :0.83875   Median : 0.2953   Median :0.8470   Median :-0.8263
##  Mean   :0.83601   Mean   : 0.3000   Mean   :0.8606   Mean   :-0.8453
##  3rd Qu.:1.03258   3rd Qu.: 0.4374   3rd Qu.:0.9770   3rd Qu.:-0.7151
##  Max.   :1.83847   Max.   : 0.8835   Max.   :1.4504   Max.   :-0.2914
##        b3
##  Min.   :0.4063
##  1st Qu.:0.7196
##  Median :0.8357
##  Mean   :0.8438
##  3rd Qu.:0.9446
##  Max.   :1.5976
```

- $\gamma = 3$ のとき

```
##        t                  b0                 b1                b2
##  Min.   :-0.2463   Min.   :-0.4780   Min.   :0.3717   Min.   :-1.5092
##  1st Qu.: 0.6309   1st Qu.: 0.1260   1st Qu.:0.7364   1st Qu.:-1.0165
##  Median : 0.8597   Median : 0.2888   Median :0.8733   Median :-0.8629
##  Mean   : 0.8818   Mean   : 0.2828   Mean   :0.8950   Mean   :-0.8814
##  3rd Qu.: 1.1215   3rd Qu.: 0.4351   3rd Qu.:1.0149   3rd Qu.:-0.7416
##  Max.   : 2.2905   Max.   : 1.0109   Max.   :1.5522   Max.   :-0.3581
##        b3
##  Min.   :0.3616
##  1st Qu.:0.7255
##  Median :0.8509
##  Mean   :0.8756
##  3rd Qu.:1.0138
##  Max.   :1.7459
```

- $\gamma = 4$ のとき

```
##        t                  b0                 b1                b2
##  Min.   :-0.1147   Min.   :-0.3576   Min.   :0.3728   Min.   :-1.5910
##  1st Qu.: 0.5821   1st Qu.: 0.1619   1st Qu.:0.7231   1st Qu.:-1.0469
##  Median : 0.8419   Median : 0.3001   Median :0.8755   Median :-0.8800
##  Mean   : 0.8440   Mean   : 0.3229   Mean   :0.8856   Mean   :-0.8878
##  3rd Qu.: 1.0622   3rd Qu.: 0.4652   3rd Qu.:1.0044   3rd Qu.:-0.7378
##  Max.   : 1.8513   Max.   : 1.3680   Max.   :1.6595   Max.   :-0.3215
##        b3
##  Min.   :0.3621
```

```
##  1st Qu.:0.7337
##  Median :0.8696
##  Mean   :0.8884
##  3rd Qu.:1.0165
##  Max.   :1.7908
```

結果と考察

- 誤判別がないとき、すべてのパラメータで $\gamma \approx 0$ のときに、ほとんどバイアスなく推定できており、求めている挙動通りである
- 誤判別がある時、求めたい対象である治療変数のパラメータ t は $\gamma \approx 0$ のとき、平均 0.75 であるが、$\gamma = 2$ で平均 0.83、$\gamma = 3$ で平均 0.88 であり、バイアスの減少は見られるが完全には取り除けていない
- Hung のロバストなロジスティック回帰のシュミレーションを行った際も、論文ではバイアスなく推定できると結果が報告されていたが、自分の再現では 0.1 程度バイアスが残ってしまっている
- シュミレーションの実装コードのバグが原因なのではないかと考えているが、原因はわかっていない
- 平均ではなく第三四分位数の値が、真の値に近くなっている結果も、ロバストなロジスティック回帰の再現を行ったときと同じ現象である。