動的治療計画のための2重ロバスト推定

江口 真透　　医療健康データ科学研究センター 特任教授

概要： 動的治療計画のための統計モデルの基本形は治療デザインモデルとアウトカムモデルの **2** つからなる．標準的なアウトカムモデルは治療共変量とベースライン共変量の線形モデルに分解される．この複合モデルの下で主要な目的は動的治療計画の最適ポリシーを決定することにある．従って，興味あるパラメータは治療共変量の回帰係数パラメータとなる．その他の治療割り付けパラメータ，ベースライン共変量の回帰パラメータは必ずしも正しく特定されてなくても，治療効果パラメータが正しく特定されていれば正確な推定が可能になる推定法の構築が要請されている．この発表では，治療割り付けパラメータ，ベースラインモデル・パラメータのどちらかが正しく特定されていれば治療効果パラメータの一致推定量のクラスを提案する．その中で **2** つのモデルが正しいときに漸近有効になる推定法が示される．簡単な数値実験が理論的な考察を支持することが分かった．

1 確率モデル

動的治療計画の標準的な確率モデルを導入する．簡単のため単一ステージの設定で，最適な方策を求める統計方法のクラスについて議論する．多重ステージでも同様な議論ができる．

共変数 X は p 次元ベクトル，治療変数 A は $\{1, \cdots, m\}$，結果変数 Y は実数値をとるとする．(X, A, Y) の確率密度関数のパラメトリックモデルを

$$p(x, a, y) = p(y|x, a; \theta)p(a|x; \alpha)p(x)$$

と表す．ここで (θ, α) はパラメータとする．また $X = x$ と $A = a$ を与えた Y の条件付き密度関数が一般化線形モデル：

$$p(y|x, a; \theta) = \exp\left\{\frac{yf(x, a; \theta) - \kappa(f(x, a; \theta))}{\phi} - c(y, \phi)\right\} \quad (1)$$

によって与えられるとする． ここで，線形モデル

$$f(x, a; \theta) = \beta^\top x + a\psi^\top x \quad (2)$$

を仮定する．ただし $\theta = (\psi, \beta)$．モデル (1) の性質より，Q-関数は

$$Q(x, a; \theta) = \kappa'(f(x, a; \theta))$$

と書ける．

2 2重ロバスト推定

研究の主要な目的は，線形モデル (2) の治療効果パラメータ ψ の推定である．一方で，ベースラインモデルのパラメータ β と治療デザインモデルのパラメータ α の推定は必ずしも正確でなくても良い．このような問題設定の下でパラメータ β または α が誤特定されているときでも，興味あるパラメータ ψ が正しく特定化されていれば一致推定が得られる方法を提案したい．

データ $\{(X_i, A_i, Y_i) : 1 \le i \le n\}$ に対して，ψ の推定方程式を

$$\sum_{i=1}^{n}\{Y_i - Q(X_i, A_i; \psi, \hat\beta)\}w(X_i, A_i)X_i = 0 \quad (3)$$

と提案する．ここで，重み関数は

$$w(X, A) = \frac{\eta(X, A)}{p(A|X; \hat\alpha)}A - \sum_{a\in\mathcal{A}}\eta(X, a)a \quad (4)$$

とする．また $\hat\alpha$ と $\hat\beta$ は α と β の適当な推定量とする．通常の尤度方程式は

$$\sum_{i=1}^{n}\{Y_i - Q(X_i, A_i; \psi, \hat\beta)\}A_iX_i = 0$$

と書かれる．推定方程式 (3) はこの尤度方程式を (4) によって中心化したものと見れる．例えば，(4) において $\eta(X, A) = 1$ と選択すると，(3) は逆確率重み付き最小 **2** 乗の推定方程式となる．Cf. **Wallace & Moodie (2015)**．このとき，次の性質が示される．

命題 1. 推定方程式 (3) によって求まる推定量 $\hat\psi_w$ は次の条件のどちらかが成立していればパラメータ ψ の一致推定量である．

(C_1) $E[Y|x, a] = Q(x, a; \psi, \beta)$

(C_2) $P[A = a|x] = p(a|x; \alpha), E[Y|x, a] = g(x) + Q(x, a; \psi, \beta)$

このように，アウトカムモデルが正しい仮説 (C_1)，または治療割り付けモデルと治療効果モデルが正しい仮説 (C_2) のどちらかが正しければ，推定方程式が (3) と書けるクラスはパラメータ ψ の一致推定を導く推定が導かれる（**2** 重ロバスト性）ことが分かった．このような推定量の中で最適なものを見つけたい．以下では漸近分散の意味で最良な推定量が導かれる．

命題 2. モデルの強い条件

(C_0) $E[Y|x, a] = Q_\theta(x, a)$ かつ， $P[A = a|x] = p(a|x; \alpha)$

を仮定する．重み関数 (4) を

$$w_0(X, A) = A - \sum_{a\in\mathcal{A}} ap(a|X; \hat\alpha)$$

と定めた推定関数 (3) による推定量を $\hat\psi_0$ とする．このとき，任意の重み関数 w による推定量 $\hat\psi_w$ に対して，漸近分散はつぎの関係：

$$V_{\mathrm{asym}}(\hat\psi_0) \le V_{\mathrm{asym}}(\hat\psi_w) \quad (5)$$

を満足する．ここで $\le$ は行列不等式を表す．

このように **2** 重ロバスト推定の中で推定量 $\hat\psi_0$ が漸近有効であることが分かる．実は，不等式 (5) は，正定値行列の空間上でピタゴラスの **3** 平方の定理：

$$V_{\mathrm{asym}}(\hat\psi_0) + V_{\mathrm{asym}}(\hat\psi_w - \hat\psi_0) = V_{\mathrm{asym}}(\hat\psi_w)$$

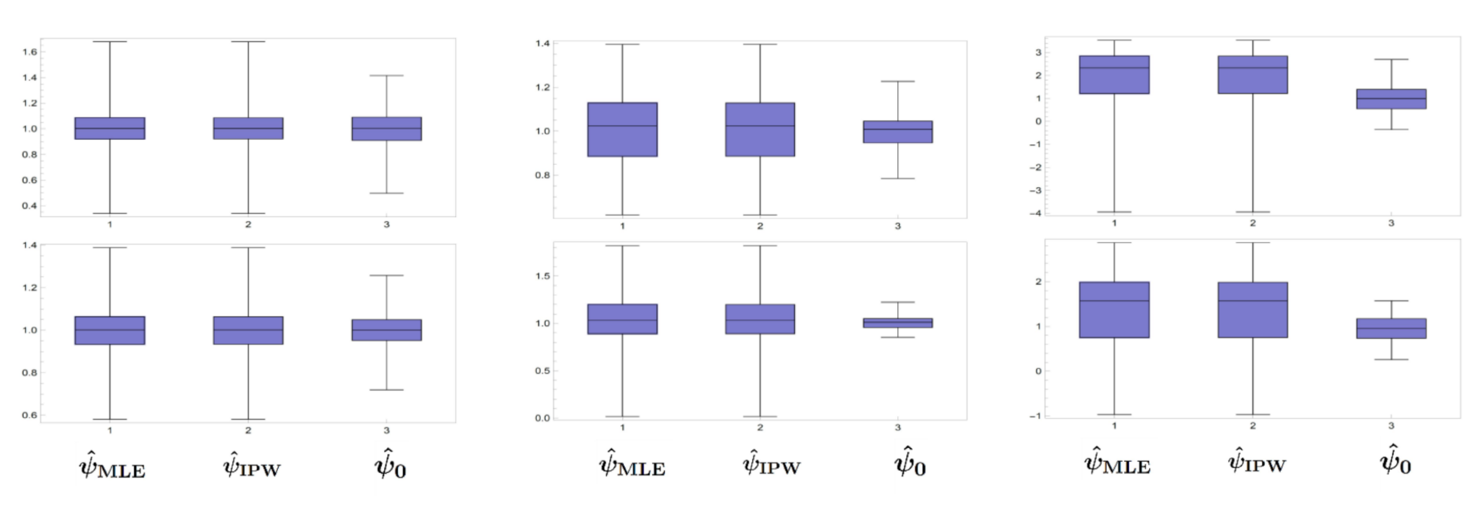
から直ちに示される．

以下に簡単な設定で数値実験を行った．

アウトカムモデル： $Y \sim \mathrm{Normal}(\beta_1X + \beta_0 + \psi_1XA + \psi_0A, 1)$

治療割り付けモデル： $A \sim \mathrm{Multinomial}(\{p(A|X; \alpha)\}, 1)$

ここで $X \sim \mathrm{Normal}(0, 1)$，$p(A|X; \alpha)$ は **3** 値ロジットモデルとする．3つシナリオは上で考えた条件 (C_0)，(C_1)，(C_2) に対応させた．治療効果パラメータの真値は $\psi_0 = 1, \psi_1 = 1$ とすると下図になった．



(C_0): Both correct models　　(C_1): Correct outcome mode Incorrect allocation model　　(C_2): Correct treatment model Correct allocation model