動的治療計画のためのポリシー・ロバストなQ学習

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**においてダブルロバストな価値関数の推定やSVMを使った**アウトカム重み付け学習**などが提案されて以来，強化学習の内容で生物統計において積極的に被験者の個人の履歴データに基づく，より良い治療の計画を提供する斬新な方法論が開発されつつある．治療計画の主要な目的の一つは履歴に基づく最適な治療を導くことである．最適治療とはQ関数に個人の共変量履歴を代入して，治療に関するQ関数の最大化によって定義される．しかし被験者の治療の感受性に異質性があるときには，尤度に基づくQ関数の推定は問題を含む．この問題に対して**治療の割付確率とアウトカムを観測した事後割付確率の比**で定められる重み付け尤度方程式による推定を提案する．従来のロバスト回帰法と比較して，特に**ポリシー・ロバスト性**の観点から提案手法の良さが主張される．

問題設定と提案推定量

単一ステージの場合について考察する．$(\boldsymbol{X}, A, Y)$ を共変量ベクトル，治療ラベル，アウトカムの組とする．Q関数 $Q_0(\boldsymbol{x}, a) = \mathbb{E}[Y|\boldsymbol{x}, a]$ が既知ならば最適ポリシーは

$$d_0(\boldsymbol{x}) = \mathrm{argmax}_a\, Q_0(\boldsymbol{x}, a)$$

と得られる．Q関数が未知ならば，例えば，**一般化線形モデル**

$$p_{\boldsymbol{\beta}}(y|\boldsymbol{x}, a) = \exp\{\frac{y\nu_{\boldsymbol{\beta}}(\boldsymbol{x}, a) - \psi(\nu_{\boldsymbol{\beta}}(\boldsymbol{x}, a))}{\phi} + c(\boldsymbol{x}, a, \phi)\}. \qquad (1)$$

を仮定すると，Q関数は　$Q_{\boldsymbol{\beta}}(\boldsymbol{x}, a) = \frac{\partial\psi}{\partial\nu}(\nu_{\boldsymbol{\beta}}(\boldsymbol{x}, a))$　と書ける．これより，

パラメータ $\boldsymbol{\beta}$ の推定がキーとなる．標準的な推定は全尤度を使ったスコア関数

$$\mathcal{E}_{\mathrm{F}}(\boldsymbol{\beta}, \boldsymbol{x}, a, y) = \frac{y - Q_{\boldsymbol{\beta}}(\boldsymbol{x}, a)}{\phi}\frac{\partial}{\partial\boldsymbol{\beta}}\nu_{\boldsymbol{\beta}}(\boldsymbol{x}, a) \qquad (2)$$

が考えられ，最尤推定量 $\hat{\boldsymbol{\beta}}_{\mathrm{F}}$ はデータ $\{(\boldsymbol{x}_i, a_i, y_i) : i = 1, \ldots, n\}$ に基づき

$$\sum_{i=1}^{n} \mathcal{E}_{\mathrm{F}}(\boldsymbol{\beta}; \boldsymbol{x}_i, a_i, y_i) = 0$$

の解として求められる．これより最適ポリシーは

$$d_{\mathrm{F}}(\boldsymbol{x}) = \operatorname*{argmax}_a Q_{\hat{\boldsymbol{\beta}}_{\mathrm{F}}}(\boldsymbol{x}, a)$$

と推定される．しかし，あとで導入するポリシー・ロバスト性の観点から深刻な問題が懸念される．　データの条件付き分布が**モデルと乖離**して

$$p_0(y|\boldsymbol{x}, a) = (1 - \epsilon)p_{\boldsymbol{\beta}}(y|\boldsymbol{x}, a) + \epsilon\, p_{\mathrm{h}}(y|\boldsymbol{x}, a) \qquad (3)$$

と書けたとする．　例えば，主要なグループの Y の条件付き分布はモデル (1) に従うが，少数のグループが存在して $p_{\mathrm{h}}(y|\boldsymbol{x}, a)$ に従っていると考えられる．

このように真のQ関数 $Q_0(\boldsymbol{x}, a) = \mathbb{E}[Y|\boldsymbol{x}, a]$ はモデルQ関数 $Q_{\boldsymbol{\beta}}(\boldsymbol{x}, a)$ と異なってしまう．とくに $p_{\mathrm{h}}(y|\boldsymbol{x}, a)$ と $p_{\boldsymbol{\beta}}(y|\boldsymbol{x}, a)$ の平均がかけ離れると最尤推定量に大きなバイアスが生じるので推定された ポリシー $\hat{d}_{\mathrm{F}}(\boldsymbol{x})$ は判断を大きく誤る可能性がある．最適ポリシー $d_0(\boldsymbol{x}) = \mathrm{argmax}_a\, Q_0(\boldsymbol{x}, a)$ に対して i 番目の観測値 $(\boldsymbol{x}_i, a_i, y_i)$ が $a_i = d_0(\boldsymbol{x}_i)$ となるときアウトカム y_i が想定よりも低い値であった時，**尤度方程式 (2) にダメージを与える**．逆に y_i が想定よりも大きな値であればQ関数の治療に関する最大値を過大推定するだけなのでポリシーの推定には影響しないことになる．

このような考察から，既存のロバスト法ではなく，**重み付け推定関数**

$$\mathcal{E}_{\mathrm{W}}(\boldsymbol{\beta}, \boldsymbol{x}, a, y) = \frac{p_{\boldsymbol{\beta}}(a|\boldsymbol{x}, y)}{p(a|\boldsymbol{x})}\mathcal{E}_{\mathrm{F}}(\boldsymbol{\beta}, \boldsymbol{x}, a, y) - \boldsymbol{B}(\boldsymbol{\beta}, \boldsymbol{x}, a) \qquad (4)$$

を使った推定量を提案する．ここで

$$\boldsymbol{B}(\boldsymbol{\beta}, \boldsymbol{x}, a) = \mathbb{E}\Big[\frac{p_{\boldsymbol{\beta}}(a|\boldsymbol{x}, Y)}{p(a|\boldsymbol{x})}\mathcal{E}_{\mathrm{F}}(\boldsymbol{\beta}, \boldsymbol{x}, a, Y)\Big]$$

提案推定量 $\hat{\boldsymbol{\beta}}_{\mathrm{W}}$ は，**推定方程式**

$$\sum_{i=1}^{n} \mathcal{E}_{\mathrm{W}}(\boldsymbol{\beta}, \boldsymbol{x}_i, a_i, y_i) = \boldsymbol{0}$$

の解と定義される．

ポリシー・ロバスト性

つぎに，最適ポリシーの推定量のロバスト性について考察する．簡単のため，治療空間を $\{1, -1\}$ とする．このとき最適ポリシーは

$$d_0(\boldsymbol{x}) = \mathrm{sign}\{Q_0(\boldsymbol{x}, 1) - Q_0(\boldsymbol{x}, -1)\}$$

と書ける．データの分布が (3) のように誤特定されるときでも，推定されたポリシー $d_{\hat{\boldsymbol{\beta}}}(\boldsymbol{x})$ が最適ポリシーと一致する確率が高いとき，**ポリシー・ロバスト性**を持つという．

このとき，重み付け推定関数 (4) の重み関数は**ロジスティック関数形**

$$p_{\boldsymbol{\beta}}(a|\boldsymbol{x}, y) = \begin{cases} \dfrac{1}{1 + \exp(-\alpha_1(\boldsymbol{x})y - \alpha_0(\boldsymbol{x}))} & \text{if } a = 1 \\ \dfrac{1}{1 + \exp(\alpha_1(\boldsymbol{x})y + \alpha_0(\boldsymbol{x}))} & \text{if } a = -1 \end{cases}$$

で表せる．ここで，

$$\alpha_1(\boldsymbol{x}) = \frac{\nu_{\boldsymbol{\beta}}(\boldsymbol{x}, 1) - \nu_{\boldsymbol{\beta}}(\boldsymbol{x}, -1)}{\phi}$$

$$\alpha_0(\boldsymbol{x}) = \frac{\psi(\nu_{\boldsymbol{\beta}}(\boldsymbol{x}, -1)) - \psi(\nu_{\boldsymbol{\beta}}(\boldsymbol{x}, 1))}{\phi} + \log\frac{p(1|\boldsymbol{x})}{p(-1|\boldsymbol{x})}$$

従って，この表現と一般化線形モデルの基本性質から

$$\alpha_1(\boldsymbol{x}) > 0 \text{ if and only if } Q(\boldsymbol{x}, 1) > Q(\boldsymbol{x}, -1)$$

が成立することが分かるので $Q(\boldsymbol{x},1) > Q(\boldsymbol{x},-1)$ ならば Y が大きな値を持つときは**重み付き推定関数のノルム**は大きくなるが Y が小さな値を持つときは有界なノルムを持つことが分かる．具体的には以下の例題で示される．

例題.　Y の条件付き分布が正規分布 $\mathcal{N}(Q_{\boldsymbol{\beta}}(\boldsymbol{x}, a), \sigma^2)$ として，線形モデル

$$Q_{\boldsymbol{\beta}}(\boldsymbol{x}, a) = \beta_0 + a\beta_1 + \boldsymbol{x}^\top\boldsymbol{\beta}_2 + a\boldsymbol{x}^\top\boldsymbol{\beta}_3$$

を仮定する．重み付き推定関数 (3) は

$$\lim_{y\to-\infty} \|\mathcal{E}_{\mathrm{W}}(\boldsymbol{\beta}, \boldsymbol{x}, a, y)\| = 0, \qquad \lim_{y\to\infty} \|\mathcal{E}_{\mathrm{W}}(\boldsymbol{\beta}, \boldsymbol{x}, a, y)\| = \infty$$

となり，切片項に対する成分は下の図ような挙動を示す．ガンマ・べき推定関数は Y の両側で有界になり，極限で 0 になるが，この重み付き推定関数 は片側だけが有界になる．このように**ポリシーが影響を受ける Y の外れ値にのみ**にロバスト性があることが分かる．

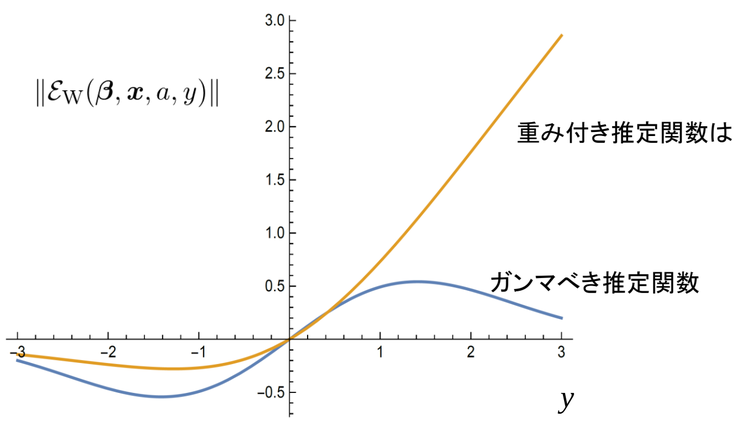

図1. 最適ポリシー $d_0(x) = 1$ のときの推定関数グラフ