調査観察データの統計科学ゼミ-第 16 回-

Masamichi Ito

Osaka University Graduate School of Human Sciences
Adachi Lab M1

February 6, 2020

担当章

5.4 共変量シフト

選択バイアス (selectivity bias, selction bias)-復習-

観察するサンプリング（抽出）された集団が、全体の様子を反映していないことによって起こる偏りのこと

(東北電力, https://www.tohoku-epco.co.jp/denjikai/research/epidemiology/bias.html から抜粋)

同様の問題が機械学習などの情報工学の分野にも存在する.

共変量シフト説明する前に機械学習の大まかな目的を

- 機械学習の研究関心：入力 $\boldsymbol{w}$ から出力 y をなるべく精度よく予測した〜い !! (e.g. 重回帰分析，ロジスティック回帰分析, NN etc.)

- 入力から出力を予測する問題
  ＝条件付き分布 $p(y|\boldsymbol{w})$(回帰関数モデル) の推定と選択の問題！

じゃあ，機械学習分野での選択バイアス (= 共変量シフト) ってどういうこと？

共変量シフト

共変量シフトは，選択バイアスの下位概念として定義可能

共変量シフト (covariate shift)

入力 $\boldsymbol{w}$ を所与とした時の出力 y の条件付き分布 $p(y|\boldsymbol{w})$ は訓練データ A とテストデータ B で共通であるが，入力 $\boldsymbol{w}$ の分布は異なるという状況 (本書)
入出力規則 (与えられた入力 $\boldsymbol{w}$ に対する出力 y の生成規則) は訓練時とテスト時で変わらないが，入力 $\boldsymbol{w}$(共変量) の分布が訓練時とテスト時で異なるという状況 (杉山, 2000)

-その心は-
入力 $\rightarrow$ 出力の関係 (条件付き分布) は共通やけど，入力の分布は違うで〜〜

共変量シフトの例

訓練データ，テストデータに対応するデータをそれぞれ，データ A, データ B と表す．

共変量シフトの例

- 過去のデータ A を訓練データとしてモデルのパラメータを推定して，データ B による出力を予測したい
  (パネルデータとか？)
- データ B を入力とした時の出力を予測したいが，データ B のサンプル数は少ない．反面，多少異なるデータ発生メカニズムだが，データ B と類似しているデータ A のサンプルを大量に利用可能である場合．(外挿とか？)

→2 つのデータ発生メカニズムが近いが同じではないデータが存在する状況はかなり普遍的

共変量シフト

本来関心があるのは，テストデータ B
$\rightarrow$ 統計学的には，データ B を母集団からの無作為抽出標本と考える.
$z = 1$ を訓練データへの割り当て，$z = 0$ をテストデータへの割り当てと考えると，共変量シフトは，

$$p(\boldsymbol{w}|z = 1) \neq p(\boldsymbol{w}|z = 0)$$
$$p(y|\boldsymbol{w}, z = 1) = p(y|\boldsymbol{w}, z = 0)$$

のようにも表すことができる.

共変量シフト

訓練データでの $(y, \boldsymbol{w})$ のペアから関心のあるテストデータでの$p(y|\boldsymbol{w})$の推定を行う場合，本来，$(y, \boldsymbol{w})$ は，同時分布

$$p(y, \boldsymbol{w}|z = 0) = p(y|\boldsymbol{w})p(\boldsymbol{w}|z = 0)$$

からの無作為標本であり，そのうち，訓練データ $(z = 1)$ として，

$$\begin{aligned} p(z = 1|\boldsymbol{w}) &= \frac{p(\boldsymbol{w}|z = 1)p(z = 1)}{p(\boldsymbol{w})} \\ &= \frac{p(\boldsymbol{w}|z = 1)p(z = 1)}{p(\boldsymbol{w}|z = 1)p(z = 1) + p(\boldsymbol{w}|z = 0)p(z = 0)} \end{aligned} \tag{5.25}$$

の確率で $(y, \boldsymbol{w})$ が観測されていると考える．

共変量シフト (続き)

-> 共変量 $\boldsymbol{w}$ を条件付ければ，y は" ランダムな欠測" または，" 観測値による選択" が行われている (=欠測値に欠測するかどうかが依存しない) と考えることができる.

条件付き分布 $p(y|\boldsymbol{w})$ の母数推定のみに興味があるなら，訓練データを用いて単純に最尤推定や最小二乗推定などを行えば良い !!

この問題設定は，3.6 節の「独立変数を条件付けた時の結果変数の分布 (ここでいう $p(y|\boldsymbol{w})p(\boldsymbol{w}|z=0)$) の母数推定」と同じ (続く)

共変量シフト (続き)

(5.25) 式が傾向スコアであり，これを用いると重みつき M 推定の目的関数 (3.24) の重みは，

$$\frac{1 - p(z = 1|\boldsymbol{w})}{p(z = 1|\boldsymbol{w})} = \frac{p(\boldsymbol{w}|z = 0)p(z = 0)}{p(\boldsymbol{w}|z = 1)p(z = 1)}$$

となる．ここで，$p(z = 0), p(z = 1)$ は $\boldsymbol{w}$ の関数でないため，これを除いた

$$\frac{p(\boldsymbol{w}|z = 0)}{p(\boldsymbol{w}|z = 1)}$$

を重みとする重み付き尤度を利用しても良い．

共変量シフト (続き)

予測値 $\hat{y}$ と実データ y の損失関数を $l(\hat{y}, y)$ とすると，訓練データにおける「重みをつけた期待損失」はテストデータの期待損失と等しい．

$$
\begin{aligned}
&\because E_{p(y|\boldsymbol{w})p(\boldsymbol{w}|z=1)}\left[\frac{p(\boldsymbol{w}|z=0)}{p(\boldsymbol{w}|z=1)} l(\hat{y}, y)\right] \\
&= \int p(y|\boldsymbol{w}) p(\boldsymbol{w}|z=1) \frac{p(\boldsymbol{w}|z=0)}{p(\boldsymbol{w}|z=1)} l(\hat{y}, y) dy d\boldsymbol{w} \\
&= \int p(y|\boldsymbol{w}) p(\boldsymbol{w}|z=0) l(\hat{y}, y) dy d\boldsymbol{w} = E_{p(y|\boldsymbol{w})p(\boldsymbol{w}|z=0)}[l(\hat{y}, y)]
\end{aligned}
$$

このことから，条件付き分布 $p(y|\boldsymbol{w})$ を正しく指定できない場合には，重みをつけて推定した方が予測がうまく行くことが予想される．数理的な詳細は Shimodaira(2000) を参照．

共変量シフト (図 5.4 の例)

(図 5.4 を参照しながら)

- 真の回帰関数は 2 次以上 (= 曲線) のデータを，(正しくはないが，近似の) 線形回帰モデルを用いて予測を行う問題
- 訓練データから単純な最小二乗法によって得られた回帰直線は，右に行くほどテストデータでの回帰直線との乖離がひどなる
- $p(\boldsymbol{w}|z=0)/p(\boldsymbol{w}|z=1)$ の重みづけをして得られた回帰直線は，単純な LS の推定よりも，テストデータの回帰直線と近い.

共変量シフト-結びとして-

-関連するトピック-

- Shimodaira(2000): $p(\boldsymbol{w}|z=1)$，$p(\boldsymbol{w}|z=0)$ が既知の場合，もしくは，別々に推定する場合
- Bickel et al.(2007): 分布の比 $p(\boldsymbol{w}|z=0)/p(\boldsymbol{w}|z=1)$ をデータから直接推定

共変量シフト-結びとして-

- " 共変量シフト" の問題は，選択バイアス，欠測データの問題としても考えることが可能
- " 共変量シフト" の問題設定では，通常，テストデータ B には正解がない(= 出力 y が存在しない) 場合を考える
- 機械学習分野での関心: 高次元の入力と出力の関係を明確化し，特定の入力値に対応する出力値を予測すること
- 人文社会科学での関心:
  1. $\boldsymbol{w}$ の一部を独立変数 z として y と z の関係を調べたい
  2. z 以外の $\boldsymbol{w}$(= 共変量 $\boldsymbol{x}$) と y の関数関係はなるべく仮定したくない

機械学習分野と，人文社会科学では研究目的が若干異なることに留意

5.5 パネル調査における脱落と無回答

脱落 (dropout)

同じ対象者を追跡し，繰り返し調査や測定を行うパネル調査において，途中で一部の対象者が調査に応じなくなること．

-例-

- 消費生活に関するパネル調査 (1993-2003) において，結婚予定者，新婚者などのライフイベントの前後で脱落が起こる (坂本，2005)
- 治療やアドバイスを継続的に行っていても，お医者さんの指示に不信感を抱くと，通院しなくなる (= 脱落, 治療アドヒアランス)

脱落のメカニズムも，" 完全にランダムな脱落"，" ランダムな脱落"，" ランダムでない脱落" の 3 種類を考えられる．

5.5 パネル調査における脱落と無回答

以下，y_t を t 時点目での従属変数の測定値，$\boldsymbol{x}$ を共変量，m_t を t 時点での脱落インディケータ ($m_t = 0$ なら脱落) とする．

- 完全にランダムな脱落 $\rightarrow$ 観測されているデータだけを用いた解析を行って良い

- ランダムな欠測: 以下の二つの対処法を利用
  1. 脱落する確率に影響を与える変数と従属変数との回帰関係が明確な場合
     $\rightarrow$ y_t を $y_1, \ldots, y_{t-1}, \boldsymbol{x}$ で説明する回帰分析モデルを最尤推定し，そのモデルを用いて，y_t の周辺分布の母数 (期待値など) を推定する
  2. 脱落するかどうかのモデリングが明確な場合
     $\rightarrow$ $p(m_t|y_1, \ldots, y_{t-1}, \boldsymbol{x})$ を傾向スコアとして利用し，IPW 推定量や二重にロバストな推定量を計算する

- ランダムでない脱落 $\rightarrow$ 医学分野では Diggle & Kenward(1994) のモデルがよく利用される

Diggle & Kenward(1994) のモデル

このモデルでは，$(y_1, \ldots, y_t) \sim N_t(\boldsymbol{\mu}, \Sigma)$ とする (e.g. 従属変数が時間の関数と多変量正規分布に従う誤差から構成されている:$\boldsymbol{y_t} = \phi(t) + \boldsymbol{e}$)

- 「m_t が 1 となるか 0 となるか (=t 時点で脱落するかどうか) の」の確率が，t 時点での測定値 y_t にも依存する場合も扱える!!
  =" 観測されないものによる選択"(= ランダムでない欠測) も扱える

ここで，$m_t = 0$ の対象者では y_t は欠測しているが，

$$\frac{p(m_t = 0|y_1, \ldots, y_t, \boldsymbol{x}, m_{t-t} = 1)}{1 - p(m_t = 0|y_1, \ldots, y_t, \boldsymbol{x}, m_{t-t} = 1)} = \beta_0 + \sum_{k=1}^{t} \beta_k y_k + \boldsymbol{x}^t \boldsymbol{\beta}_x$$

$$p(m_t = 0|y_1, \ldots, y_{t-1}, \boldsymbol{x}, m_{t-1} = 1)$$

$$= \int p(m_t = 0|y_1, \ldots, y_t, \boldsymbol{x}, m_{t-1} = 1) p(y_t|y_1, \ldots, y_{t-1}, \boldsymbol{x}) dy_t$$

(↑ 選択モデル，次のページへ続く)

Diggle & Kenward(1994) のモデル (続き)

のように考えると，$p(y_t|y_1, \ldots, y_{t-1}, \boldsymbol{x})$ は正規分布であることから，ロジスティック回帰モデルの説明変数に「正規分布に従う変量効果」が入っていると考えればよいため，数値計算も簡単になる

- 「m_t が 1 となるか 0 となるかの確率が y_t には依存しない」時には，" ランダムな脱落" となる
- モデル仮定が強いという点で批判が多く，現在は感度分析がよく行われるようになっている

(p.167 の例で脱落の補正の具体例を見ましょう)

References

- 坂本 (2005). サンプル脱落に関する分析-「消費生活に関するパネル調査」を用いた脱落の規定要因と推計バイアスの検証-, 日本労働研究雑誌
- Shimodaira(2000). Improving predictive inference under covariate shift by weighting the log-likelihood function, Journal of Statistical Planning and Inference, 90, 227-244
- 杉山 (2006). 共変量シフト下での教師付き学習, 日本精神神経学会誌，13(3), 111-118
- 高井, 星野, 野間 (2016)「調査観察データ解析の実際 欠測データの統計科学-医学と社会科学への応用-」
- 星野 (2009)「調査観察データの統計科学-因果推論・選択バイアス・データ融合-」

Thank you for your attention!!

Any Questions?

メモに使ってね!!