2018/12/7　計量生物セミナー@京都

ランドマークモデルによる動的予測

北海道大学　医学統計学
横田　勲

内容

2

- 経時データの組入
- ランドマーク Cox モデル
- 柔軟なモデリングに向けた擬似値の利用

生存時間アウトカムに対する予測

3

- がんや心血管疾患のような
  重篤なイベント(死亡や再発)を伴う慢性疾患
  - 臨床家が病態を知るうえで参考
  - 患者が治療選択する際に有用な情報
- ベースライン情報を用いた予測
  - Cox モデルを基に個別に生存関数を予測可能
    - "3年後の予測イベント発生確率"のような
      表現も可能

治療途中における予測

4

- ある時点sまで生存した下での条件付き推測
  - 例）診断後2年生存した対象者の3年後の死亡確率
- 動的予測 (dynamic prediction)
  - 条件付き生存(conditional survival)とも
  - $s=0$にて通常のw年生存割合と同様

$$\boxed{F(s+w \mid X>s)=\Pr\{X\in(s,s+w] \mid X>s\}}$$

- $F(\cdot)$：イベント発生確率
- X：イベント時点を表す確率変数

動的予測の指す範囲

5

- 狭い意味で、
  ある時点sまで生存下での
  $s+w$までにイベントを発生する確率
  - wを予測幅 (window) と呼ぶ
- 広い意味で、
  ある時点まで生存下での条件付き推測
  - 生存関数の形で表現

経時データ$Z(t)$の利用

6

- 予測開始時点sまでに観察された
  患者の経過をせっかくなので取り入れたい
- 時間依存性共変量Coxモデル
  - 外生 external である必要
    $\Pr\{u \le X < u + \Delta u | Z(u), X \ge u\}$
    $= \Pr\{u \le X < u + \Delta u | Z(t), X \ge u\}$
    - $0 < u \le t$を満たす任意のu, t
    - 時点uでのハザードはuまでの共変量に従属でよい
    - $u \le X < u + \Delta u$でのイベント発生と
      将来の共変量$Z(t)$は独立

内生 internal 変数への対処

7

- 外生でない場合をさす
- $Z(t)$が観察されたということは、
  対象者が生存していることが確定
  - ハザードλとの対応が不成立

$$\Pr\{T > t \mid Z(t)\} \neq \exp\left[-\int_0^t \lambda\{u \mid Z(u)\} du\right]$$

- 同時モデル joint model
- ランドマークモデル landmarking model

ランドマークモデル

8

- ランドマーク時点 s における
  at risk例で条件付け
- 時点 s までに得られた情報を利用
- 単一の生存時間である場合

(vanHouwelingen 2007, *Scand Stat Theory Appl.*)

$$\boxed{\lambda(t|s, Z(t)) = \lambda_0(t|s)\exp\{Z(s)^\top \beta(s)\}}$$

- $Z(s)$はこのモデルにおいて外生変数

適用方法

9

- ランドマーク時点 s 以前の
  イベント・打ち切り例は左側切断(truncation)
- 予測する時点 $s+w$ で生存例は
  時点 $s+w$ で強制打ち切り
- 以上の加工を行ったデータセットに
  通常のCox回帰を適用するだけ

ランドマークスーパーモデル

10

- ランドマーク時点以外からの予測
  - モデルがないので、予測できない
    - 例：ランドマーク時点が3ヶ月ごとであれば、
      2ヶ月目での予測はできない
- スーパーモデル
  - ランドマークモデルを複数時点設定し、
    それらのパラメータを時点に対し平滑化
  - 経時データのハザードへの影響は
    時点変化とともになめらかに変わるはず

スーパーモデルの構築

11

1. ランドマーク時点の組$\{s_1, \dots, s_L\}$を用意
2. 各時点ごとに左側切断と強制打ち切り
3. 各時点でのデータセットを統合し、解析
   - 同一対象者のデータを複数回利用するため、
     対象者IDをクラスターとした
     推定方程式に基づく回帰係数の推定
   - 標準誤差についてロバスト分散を利用

(Lin, Wei 1989, *JASA*)

スーパーモデル①
（経時データのパラメータのみ平滑化）

12

- $\beta(s)$に時点の多項式関数$f(s)$を導入
- $\lambda(t|s, Z(s)) = \lambda_0(t|s)\exp\{Z(s)^\top \beta(s)\}$,

$$\text{for } s < t \le s+w,$$

$$\beta(s) = \sum_j \gamma_j f_j(s)$$

  - γは経時データの次元の回帰係数ベクトル
  - jは多項式の次数を表す
- 時点sを層とした層別Coxモデルのあてはめ
  - $Z(s)$と$f_j(s)$の交互作用項を時間依存性共変量に

スーパーモデル②
（ベースラインハザードも平滑化）

- 時点を主効果に
- $\lambda(t|s, Z(s))$

$$= \lambda_0(t)\exp\left\{\sum_k \eta_k g_k(s)\right\}\exp\{Z(s)^{\mathsf{T}}\beta(s)\},$$

for $s < t \leq s + w$

  - ηはベースラインハザードに関する係数
  - kは多項式の次数を表す
  - $g_k(s)$は定数を含めず、
    s_1において$g_1(s_1) = g_2(s_1) = \cdots = 0$と標準化

Dutch Gastric Cancer Trial

- 胃がん患者に対する手術法を比較する
  ランダム化研究
  - D1 リンパ節郭清術 v.s. D2 リンパ節郭清術
- 1078名の対象者を1:1ランダム化
- 711 名が実際に手術を受けた
  - D1 リンパ節郭清術:380 名
  - D2 リンパ節郭清術:331 名

長期予後の結果

Lancet Oncol. 2010; 439-449（本体論文は*N Engl J Med.* 1999; 908-914）

手術法ごとに予後予測

- 生存関数がクロスした！
  - 異なるリスクを持つ人が集まると、
    それぞれハザード比一定でも起こりうる現象
- リスク分類を行ってから、
  各手術法を行った場合の予後を動的予測
  - リスクスコア
    - リンパ節転移、年齢、T分類、全摘or部分摘出
  - 残存腫瘍

スーパーモデル②を適用

17

4年間の死亡確率を予測

18

リスクスコアが
全体の90%目の人

リスクスコアが
全体の10%目の人

競合リスクデータの場合

19

- Coxモデルの代わりに、
  Fine-Grayモデルを利用 (Nicolaie et al. 2013 *Stat Med*)
- 動的擬似値 (dynamic pseudo-observations)の利用
  (Nicolaie et al. 2013 *Biometrics*)

$s+w$におけるイベント確率θに興味

20

- 対象者iのイベント時点T_iが全員観察
  - 途中での打ち切りがないデータ

$$\theta = E\{I(X \le s+w)\} = \frac{1}{n}\sum_{i}^{n} \boxed{I(X_i \le s+w)}$$

  - イベント時点を表す確率変数T
  - $I(\cdot)$は指示関数

$s+w$までに打ち切り
となった対象者では不明

打ち切りを含むデータ

21

- 全員のアウトカムを擬似値 $\hat{\theta}_i$ で置換
  - 打ち切りがなかった場合に観察される予測値

$$\theta = E\{I(X \le s + w)\} = \frac{1}{n}\sum_{i}^{n} \boxed{\hat{\theta}_i}$$

- θの推定量$\hat{\theta}$とleave-one-out推定量$\hat{\theta}^{(-i)}$で定義
  - 打ち切りを考慮するための(ex. Kaplan-Meier)推定量

$$\boxed{\hat{\theta}_i = n \cdot \hat{\theta} - (n-1) \cdot \hat{\theta}^{(-i)}}$$

擬似値の定義

22

当該対象者 i を除いた
$n-1$人のデータで推定した
$s+w$におけるイベント確率

$$\boxed{\hat{\theta}_i = \underline{n \cdot \hat{\theta}} - \underline{(n-1) \cdot \hat{\theta}^{(-i)}}}$$

当該対象者 i を含めた
n人のデータにおける
イベント発生人数の予測値

当該対象者 i を除いた
$n-1$人のデータにおける
イベント発生人数の予測値

- $\hat{\theta}_i$ は打ち切りがなかった場合に観察される予測値
  - 元々打ち切りがない場合は $\hat{\theta}_i = I(X_i \le s + w)$

ランドマークごとに擬似値を導入

23

- 推定したいイベント発生確率そのものを直接的にモデル化可能
  - 一般化線形モデルを導入できるため、関数形をより自由に設定できる
- 多変量生存時間アウトカムへ拡張可能
  - 競合リスクイベント
  - 繰り返し（再発）イベント・・etc

ランドマークモデルの利点

24

- データセットの加工だけで実行が簡便
- モデルにおく仮定が少ない
  - 比例ハザード性
  - 治療や中間イベント発生について時間とともに影響が変わることも考慮可能
- Coxモデルのみならず、一般化線形モデルにも応用可能

ランドマークモデルの弱点

25

- 正しく特定された同時モデルに比べ、予測精度は悪い
- 経時データの過程にモデルをおかない
  - 妥当な予測モデルにするためには、
    - 経時データの過程（分布）を正しく特定
    - 経時データとイベントの関係を正しく特定

  をいかなる時点においても達成すべき

(Jewell, Nielsen 1993, *Biometrika*. Suresh et al. 2017, *Biom J*.)

文献

26

- Andersen PK, Klein JP, Rosthoj S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* 2003. 90: 15-27
- Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983; 1: 710-9.
- Jewell NP, Nielsen JP. A framework for consistent prediction rules based on markers. *Biometrika* 1993; 80: 153–164.
- Lin DY, Wei LJ. The robust inference for the cox proportional hazards model. *J Am Statist Assoc*. 1989; 84: 1074–1078.
- Nicolaie MA, van Houwelingen JC, deWitte TM, Putter H. Dynamic prediction by landmarking in competing risks. Stat Med. 2013; 32: 2031-2047.
- Nicolaie MA, van Houwelingen JC, deWitte TM, Putter H. Dynamic pseudo-observations: A robust approach to dynamic prediction in competing risks. *Biometrics*. 2013; 69: 1043-1052.
- Suresh K, Taylor JMG, Spratt DE, Daignault S, Tsodikov A. Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biom J*. 2017; 59: 1277-1300.
- Taylor JMG, Park Y, Ankerst DP, Bae K, Pickles T, Sandler H. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* 2013; 69: 206-213.
- van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scand Stat Theory Appl*. 2007; 34: 70–85.
- van Houwelingen HC, Putter, H. *Dynamic prediction in clinical survival analysis*. 2011. CRC Press, Boca Raton(FL).
- Zheng Y, Heagerty, PJ. Partly conditional survival models for longitudinal data. *Biometrics* 2005; 61: 379–391.