Chapter 9 Refinements of Semi-parametric PHM

PubH 7450

©Wei Pan
Email: weip@biostat.umn.edu
Http: www.biostat.umn.edu/∼weip

§9.2 Time-dependent covariates

- Z: baseline/fixed, not changing over time; most common.
  $Z(t)$: time-dependent or time-varying covariate, e.g. weight, BMI, blood pressure, health or trt status at time t.

- PHM and PL:

$$h(t|Z(t)) = h_0(t)\exp\left(Z(t)'\beta\right).$$

$$L(\beta) = \prod_{i=1}^{D} L_i = \prod_{i=1}^{D} \frac{\exp\left(Z_{(i)}(t)'\beta\right)}{\sum_{j\in R(t_i)} \exp\left(Z_{(j)}(t)'\beta\right)}.$$

- Example: $n = 4$; $Z(t)$: smoking status at t.
  obs 1: $Z(t) = 1$ for $0 \le t < 2$; $Z(t) = 0$ for $t \ge 2$; event at $t = 5$.
  obs 2: $Z(t) = 0$ for $0 \le t < 4$; $Z(t) = 1$ for $t \ge 4$; event at $t = 6$.
  obs 3: $Z(t) = 0$ up to an event at $t = 1$.
  obs 4: $Z(t) = 1$ up to an event at $t = 3$.
  $L_1 = \frac{\exp(0*\beta)}{\exp(0*\beta)+\exp(1*\beta)+\exp(0*\beta)+\exp(1*\beta)}$.

$L_2 = \frac{\exp(1*\beta)}{\exp(1*\beta)+\exp(0*\beta)+\exp(1*\beta)}$.
$L_3 = \frac{\exp(0*\beta)}{\exp(0*\beta)+\exp(1*\beta)}$.
$L_4 = \frac{\exp(1*\beta)}{\exp(1*\beta)} = 1$.

- Example 9.1: SAS
  DFS: disease-free survival, min(disease recurrence, death).
  Dan Sargent's paper: in colon-cancer trials, the outcome is
  1) 5-yr survival as the current practice;
  2) 3-yr DFS as the new proposed one.
  $Z_a = 0$ if $t <$ time at which aGVHD occurs; $= 1$ o/w.
  $Z_c = 0$ if $t <$ time at which cGVHD occurs; $= 1$ o/w.
  $Z_p = 0$ if $t <$ time at the platelets recovered; $= 1$ o/w.
- An application: check the PH assumption.
- Consider a simple example: a binary $Z = 0$ or 1, indicating one of the two groups.
  PHM: $h(t|Z) = h_0(t)\exp(Z\beta_1) \Longrightarrow$

$\frac{h(t|\text{Grp } 1)}{h(t|\text{Grp } 0)} = \exp(\beta_1)$, constant!

- Now, consider an alternative model:
  $h(t|Z) = h_0(t) \exp(Z\beta_1 + Zg(t)\beta_2) \Longrightarrow$
  $\frac{h(t|\text{Grp } 1)}{h(t|\text{Grp } 0)} = \exp(\beta_1 + g(t)\beta_2)$, time-varying (if $\beta_2 = 0$)!
- Checking a PHM: add a time-dependent covariate, e.g. $Z \times \log(t)$, and see whether its coefficient is significant!
  not a general GOF test: depending on the true and specified $g(t)$; you used this type of tests often...
- Example 9.2: SAS

§9.3 Stratified PHM

- Consider two binary covariates: Z_1 indicating trt grp, Z_2 indicating male grp.
  PHM: $h(t|Z_1, Z_2) = h_0(t)\exp(Z_1\beta_1 + Z_2\beta_2) \Longrightarrow$
  $\frac{h(t|trt,Z_2)}{h(t|ctl,Z_2)} = \exp(\beta_1)$,
  $\frac{h(t|Z_1,M)}{h(t|Z_1,F)} = \exp(\beta_2)$. –again constant and PH!

- How about if 1) PH assumption may not hold for Z_2 while holds for Z_1 and 2) only β_1 is of interest?

- 1) fix the problem, e.g., by ...; 2) use a *stratified PHM*:
  $h(t|Z_1, Z_2) = h_0(t|Z_2)\exp(Z_1\beta_1)$

- Implications:
  $h(t|Z_1, M) = h_{0,M}(t)\exp(Z_1\beta_1)$,
  $h(t|Z_1, F) = h_{0,F}(t)\exp(Z_1\beta_1)$,
  and there is **no** restriction/assumption on the relationship between the two baseline hazards; e.g.,

$h(t|Z_1, M)/h(t|Z_1, F) = h_{0,M}(t)/h_{0,F}(t)$ may or may *not* be a constant!

On the other hand,

$\frac{h(t|trt, Z_2)}{h(t|ctl, Z_2)} = \exp(\beta_1)$, still a PHM!

- Summary: *to adjust for a confounder*, 1) put it in a regression model; or 2) treat it as a stratifier.

  A key difference: 1) may require a stronger modeling assumption.

- How to estimate β?

- Method 1: analogous to M-H method,

  1) fit a PHM for each stratum: using $L^{(h)}(\beta)$ based on the data in stratum $h \Longrightarrow \hat{\beta}^{(h)}$;

  2) combine $\hat{\beta}^{(h)}$'s, e.g., by ...

  Prior to 2), may conduct a homogeneity test on H_0: $\beta^{(1)} = \beta^{(2)} = ...$; why and how?

- Method 2: under the homogeneity assumption, use $L(\beta) = \prod_h L^{(h)}(\beta)$.
- Example 9.1b: SAS

§9.4 Left-truncated and right-censored data

- Channing House data:
  Age 1: age at the entry;
  Age 2: age at the end of study or death.
  Covariate: gender.

- Recall how we dealt with LT-RC'ed data:
  One sample problem: modify K-M or N-A estimator; how?
  K-sample problem: modify generalized rank tests; how?
  Regression with PHM: modify PL; how?

- LT-RT'ed data: $(Y_j, T_j, \delta_j, Z_j)$, $j = 1, ..., n$.

- PHM: $h(t|Z) = h_0(t)\exp(Z'\beta)$.

- Modification: $R(t_i) = \{j : Y_j < t_i \le T_j\}$;
  PL remains the same with the new $R(t_i)$.

- MPLE

$\hat{\beta} = \text{argmax}_\beta \prod_{i=1}^{D} \frac{\exp(Z'_{(i)}\beta)}{\sum_{j \in R(t_i)} \exp(Z'_j\beta)}$,

$\hat{H}_0(t) = \sum_{t_i \le t} \frac{d_i}{\sum_{j \in R(t_i)} \exp(Z'_j\beta)}$.

- Example 9.4: SAS
- Another use of the counting process notation:
  Code time-dependent covariates: e.g., how to input the data with time-dependnet smoking status in the example on p.2? Any issue with multiple "observations" from the same subject?