Revisit Partial Likelihood and Tie Corrections for the Cox Model Using Poisson Binomial Distributions

Youngjin Cho, Yili Hong, and Pang Du

Department of Statistics

pangdu@vt.edu

Motivation

- Cox model [3] : use partial likelihood (PL) [4].
- Original PL idea is buried : due to computational challenges, especially in tied data.
- Existing PL computing methods : use approximation.
- Revisit original PL idea : by efficient computation.

- n : the number of observations.
- $\tilde{T}_i \in (0, \infty)$: underlying event time from a continuous distribution.
- $x_i = (x_{i1}, \ldots, x_{id})^\mathsf{T}$: covariate.
- The hazard function (HF) is

$$d\Lambda(t; x_i) = \mathrm{pr}(\tilde{T}_i \in [t, t + dt) | x_i, \tilde{T}_i \geq t) = \exp(x_i^\mathsf{T} \beta) d\Lambda_0(t).$$

- $\Lambda(t; x_i) = \int_0^t d\Lambda(s; x_i) = \exp(x_i^\mathsf{T} \beta) \Lambda_0(t)$: cumulative HF (CHF).
- $d\Lambda_0(t) = \lambda_0(t) dt$: baseline HF.
- $\zeta \in (0, \infty)$: ending time of the study.
- $C_i \in (0, \zeta]$: right censoring time.

- Ties : generated by the grouping of an underlying continuous random variable.
- $\lceil a \rceil$: smallest integer which is not smaller than a.
- $\tau \in (0, \infty)$: grouping parameter.
- Discretization :

$$\begin{aligned} \tilde{T}_i^* &= \tau \lceil \tilde{T}_i / \tau \rceil \in \Omega_{\mathsf{G}} = \{\tau, 2\tau, \dots\}, \\ C_i^* &= \tau \lceil C_i / \tau \rceil \in \Omega = \{\tau, 2\tau, \dots, \zeta\}. \end{aligned}$$

- $T_i^* = \text{min}(\tilde{T}_i^*, C_i^*) \in \Omega$: observed time.
- $\delta_i = \mathbb{1}(\tilde{T}_i^* \leq C_i^*) = \begin{cases} 1, & \text{if } i\text{th subject had an event.} \\ 0, & \text{if } i\text{th subject was censored.} \end{cases}$

For $t \in \Omega_{\mathsf{G}}$, the HF of $\tilde{T}_i^*$ is

$$\begin{aligned} d\Lambda^*(t; x_i) &= \mathrm{pr}(\tilde{T}_i^* \in [t, t+dt) | x_i, \tilde{T}_i^* \geq t) \\ &= 1 - \exp\left(-\exp(x_i^{\mathsf{T}}\beta) d\Lambda_0^*(t)\right). \end{aligned}$$

- Baseline HF : $d\Lambda_0^*(t) = \begin{cases} \Lambda_0(t) - \Lambda_0(t-\tau), & \text{if } t \in \Omega_{\mathsf{G}}. \\ 0, & \text{if } t \notin \Omega_{\mathsf{G}}. \end{cases}$
- Baseline CHF : $\Lambda_0^*(t) = \int_0^t d\Lambda_0^*(s) = \sum_{s \leq t, s \in \Omega_{\mathsf{G}}} d\Lambda_0^*(s)$.

We call the model continuous model with grouping (CMG).

- Assume CMG.
- Data : $\{T_i^*, \delta_i, x_i\}_{i=1}^n$.
- $\{t_j\}_{j=1}^k$: the distinct ordered event times.
- $\mathcal{R}(t_j)$: at-risk set at time t_j, $n_j = |\mathcal{R}(t_j)|$.
- $\mathcal{D}(t_j)$: event set at time t_j, $d_j = |\mathcal{D}(t_j)|$.
- CMG allows $d_j > 1$.
- $p_{ij} = p_{ij}(\beta, \lambda_j) \equiv d\Lambda^*(t_j; x_i)$.
- $\lambda_j \equiv d\Lambda_0^*(t_j)$.

The accurate PL :

$$L(\beta, \Lambda) = \prod_{j=1}^{k} L_j(\beta, \lambda_j) = \prod_{j=1}^{k} \frac{A_j(\beta, \lambda_j)}{B_j(\beta, \lambda_j)},$$

$$L_j(\beta, \lambda_j) = \frac{\mathrm{pr}(\text{item } j_1, \ldots, j_{d_j} \text{ had event at } t_j \,\big|\, n_j \text{ units survived up to } t_j)}{\mathrm{pr}(d_j \text{ out of } n_j \text{ units had event at } t_j \,\big|\, n_j \text{ units survived up to } t_j)}.$$

- $\Lambda = (\lambda_1, \ldots, \lambda_k)^\mathsf{T}$.
- $j_1, \ldots, j_{d_j}$: d_j individuals in $\mathcal{D}(t_j)$.

$$A_j(\beta,\lambda_j) = \prod_{i \in \mathcal{D}(t_j)} p_{ij}(\beta,\lambda_j) \prod_{i \in \mathcal{R}(t_j)\setminus\mathcal{D}(t_j)} (1 - p_{ij}(\beta,\lambda_j))\,.$$

$$B_j(\beta,\lambda_j) = \sum_{\mathcal{A} \in \mathfrak{F}_{d_j}} \left\{ \prod_{i \in \mathcal{A}} p_{ij} \prod_{i \in \mathcal{R}(t_j)\setminus\mathcal{A}} (1 - p_{ij}) \right\}. \tag{1}$$

- $\mathfrak{F}_{d_j}$: set of all subsets of d_j individuals that can be selected from $\mathcal{R}(t_j)$.
- (1) is the form of probability mass function (PMF) of a poisson binomial distribution (PBD).
- PBD : sum of independent Bernoulli random variables.
- Computing PBD by enumeration has high computational cost.

Partial Likelihood

Existing Methods

Data without ties :

$$L_j(\beta, \lambda_j) \approx \frac{\exp(x_{j_1}^{\mathsf{T}}\beta)}{\sum_{i \in \mathcal{R}(t_j)} \exp(x_i^{\mathsf{T}}\beta)}.$$

Data with ties :

1. Breslow correction [1]

$$L_j(\beta, \lambda_j) \propto \frac{\exp\left(\sum_{i \in \mathcal{D}(t_j)} x_i^{\mathsf{T}}\beta\right)}{\left\{\sum_{i \in \mathcal{R}(t_j)} \exp(x_i^{\mathsf{T}}\beta)\right\}^{d_j}}.$$

2. Efron correction [5]

$$L_j(\beta, \lambda_j) \propto \frac{\exp\left(\sum_{i \in \mathcal{D}(t_j)} x_i^{\mathsf{T}}\beta\right)}{\prod_{\ell=0}^{d_j-1}\left\{\sum_{i \in \mathcal{R}(t_j)} \exp(x_i^{\mathsf{T}}\beta) - \ell \overline{A}\,(\beta, t_j)\right\}}. \quad (2)$$

- Existing methods use approximations to both $A_j(\beta, \lambda_j)$ and $B_j(\beta, \lambda_j)$.
- Discrete Fourier transform of the characteristic function (DFT-CF) method [6] can efficiently compute $B_j(\beta, \lambda_j)$: PMF of PBD.
- However, p_{ij}s depend on Λ.
- $\widehat{\Lambda}_{\mathsf{e}} = (\widehat{\lambda}_{\mathsf{e}1}, \ldots, \widehat{\lambda}_{\mathsf{e}k})^{\mathsf{T}}$ from (2) substitutes Λ.

Propose accurate PL based on PBD :

$$L(\beta, \widehat{\Lambda}_{\mathsf{e}}) = \prod_{j=1}^{k} \frac{A_j(\beta, \widehat{\lambda}_{\mathsf{e}j})}{B_j(\beta, \widehat{\lambda}_{\mathsf{e}j})}.$$

Partial Likelihood

Estimation Based on Poisson Binomial Distribution

PBD Estimation Procedure

Data : $\{T_i^*, \delta_i, x_i\}_{i=1}^n$.

1. Get $\widehat{\Lambda}_{\mathsf{e}}$ from (2).
2. Get $\widehat{\beta}_{\mathsf{pb}} = \arg\max_{\tilde{\beta}} L(\tilde{\beta}, \Lambda = \widehat{\Lambda}_{\mathsf{e}})$.
3. For $j = 1, \cdots, k$, get $\widehat{\lambda}_{\mathsf{pb},j} = \arg\max_{\tilde{\lambda}_j} A_j(\beta = \widehat{\beta}_{\mathsf{pb}}, \tilde{\lambda}_j)$.

Return : $\widehat{\beta}_{\mathsf{pb}}$ and $\widehat{\Lambda}_{\mathsf{pb}}(t) = \sum_{j=1}^k \widehat{\lambda}_{\mathsf{pb},j} \mathbb{1}(t_j \le t)$.

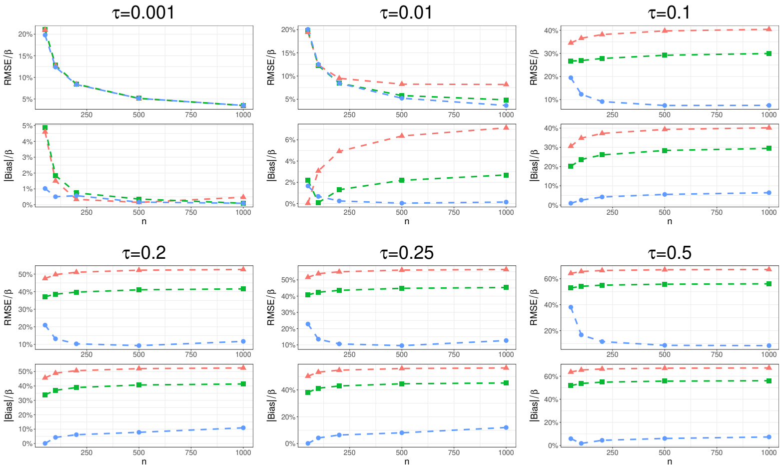
Statistical Properties

- Under some conditions, consistency and asymptotic normality for $\widehat{\beta}_{\mathsf{pb}}$ are satisfied under CMG.
- These conditions include $\sup_{j\in\{1,\cdots,k\}} d_j = \mathcal{O}_{\mathsf{P}}(1)$: non-diverging ties as $n \to \infty$.
- Under some conditions, consistency and asymptotic normality for $\widehat{\beta}_{\mathsf{pb}}$ are satisfied under continuous data without ties.

- $\tilde{T}_i$ has Weibull baseline HF : $\lambda_0(t) = \gamma t^{\gamma-1}/\eta^\gamma$.
- $C_i = \max\{\tilde{C}_i, \zeta\}$, $\tilde{C}_i$ has Weibull HF : $\lambda_c(t) = \gamma_c t^{\gamma_c - 1}/\eta_c^{\gamma_c}$.
- $x_i \sim \mathsf{N}(0, \sigma_x^2)$.
- Ties are generated under CMG.
- $\zeta = 1$, $\eta = \eta_c = 1.31$ and $\gamma = \gamma_c = 1.5$.
- $\tau \in \{0.001, 0.01, 0.1, 0.2, 0.25, 0.5\}$, $\beta \in \{1, 1.5\}$, $n \in \{50, 100, 200, 500, 1000\}$, and $\sigma_x \in \{1, 1.5, 2\}$.
- Repeat the simulations $B = 10000$ times for all the simulation cases.
- Compare PBD estimator $\widehat{\beta}_{\mathrm{pb}}$, Breslow estimator $\widehat{\beta}_{\mathrm{b}}$, and Efron estimator $\widehat{\beta}_{\mathrm{e}}$ by the root mean square error (RMSE) and absolute bias (|Bias|).

Simulation Studies

Simulation Results : $\beta = 1.5$ and $x_i \sim \mathrm{N}(0, 2^2)$



- Breslow (red) vs Efron (green) vs PBD (blue).

- In simulation, large τ or large σ_x increases |Bias| and RMSE of $\widehat{\beta}_{\mathrm{b}}$ and $\widehat{\beta}_{\mathrm{e}}$.
- By Le Cam [2], large value of

$$\frac{1}{k}\sum_{j=1}^{k}\frac{1}{n_j}\sum_{i\in\mathcal{R}(t_j)}p_{ij}^2 \tag{3}$$

  makes approximated PL used in $\widehat{\beta}_{\mathrm{b}}$ and $\widehat{\beta}_{\mathrm{e}}$ less accurate, resulting in their bad performances.
- This can be caused by large τ (since $p_{ij} = \mathcal{O}_{\mathrm{P}}(\tau)$) or large σ_x.
- We expect large τ or large (3) increases RMSE of $\widehat{\beta}_{\mathrm{b}}$ and $\widehat{\beta}_{\mathrm{e}}$ in real datasets, showing superiority of $\widehat{\beta}_{\mathrm{pb}}$.

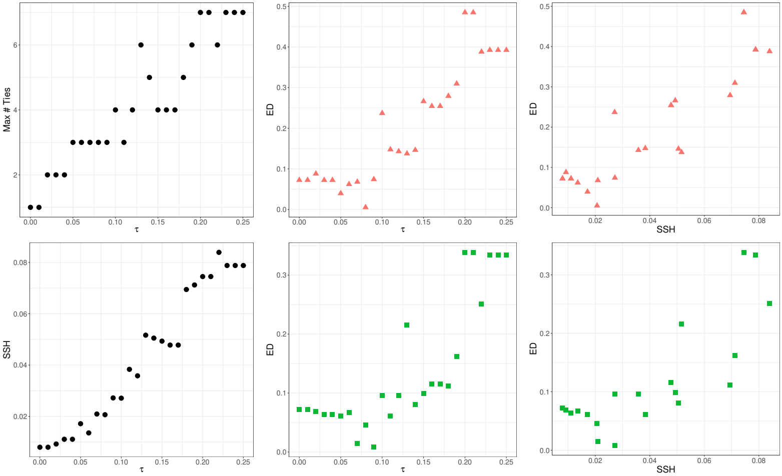
- $\{t_i, \delta_i, x_i\}_{i=1}^n$: observed data.
- $t_i^* = \lceil t_i/\tau \rceil \tau$ for $\tau \in \{0.01, 0.02, \cdots, 0.25\}$.
- For each τ, fit $\widehat{\beta}_{\mathsf{b}}$, $\widehat{\beta}_{\mathsf{e}}$, and $\widehat{\beta}_{\mathsf{pb}}$ with $\{t_i^*, \delta_i, x_i\}_{i=1}^n$.
- β is unknown, use $\widehat{\beta}_{\mathsf{pb}}$ as benchmark : theoretically accurate and has strong performance in simulation.

For each τ, record the followings :

1. Sum of Squared Hazards (SSH) : $\frac{1}{k}\sum_{j=1}^{k}\frac{1}{n_j}\sum_{i\in\mathcal{R}(t_j)}\hat{p}_{ij}^2$.
2. Estimation Discrepancy (ED) : $\max_{l\in\{1,\cdots,d\}}\left\{\exp\left(|\widehat{\beta}_l - \widehat{\beta}_{\mathsf{pb},l}|\right) - 1\right\}$.

- $\hat{p}_{ij}$: p_{ij} evaluated at $\{\widehat{\beta}_{\mathsf{pb}}, \widehat{\lambda}_{\mathsf{pb},j}\}$ and $\widehat{\beta}$: $\widehat{\beta}_{\mathsf{b}}$ or $\widehat{\beta}_{\mathsf{e}}$.
- If τ or SSH increases, we expect larger ED, showing superiority of $\widehat{\beta}_{\mathsf{pb}}$.

- Many ties compared to small sample size : $n = 26$.



- Breslow (red) vs Efron (green).

- Accurate computation using DFT-CF : exact partial likelihood method for the Cox model based on PBD.
- Asymptotics : for both grouped data with ties and continuous data without ties.
- RMSE reduction by less $|\text{Bias}|$: useful for data with many ties or high variation among covariate.
- Future research : accurate PL for competing risk cox model using a Poisson multinomial distribution (PMD).
- Lin et al. [7] : efficient calculation for PMD.

References I

[1] N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–99, 1974. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2529620.

[2] Lucien Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics*, 10:1181 – 1197, 1960.

[3] D. R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.

[4] D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

[5] Bradley Efron. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72:557–565, 1977. ISSN 01621459. URL http://www.jstor.org/stable/2286217.

[6] Y. Hong. On computing the distribution function for the Poisson bionomial distribution. *Computational Statistics and Data Analysis*, 59:41–51, 2013.

[7] Zhengzhi Lin, Yueyao Wang, and Yili Hong. The computing of the Poisson multinomial distribution and applications in ecological inference and machine learning. *Computational Statistics*, 38:1851–1877, 2023. doi: 10.1007/s00180-022-01299-0.