STAT6018 Research Frontiers in Data Science

Topic I: Statistical methods for analyzing complex survival data

Yu Gu, PhD
Assistant Professor

Department of Statistics & Actuarial Science
The University of Hong Kong

Table of Contents

1. Chapter 2: Semiparametric regression analysis of interval-censored data
   - Transformation models for interval-censored data
   - Transformation models for multivariate interval-censored data
   - Marginal models for multivariate interval-censored data
   - Semiparametric regression models for panel count data

Table of Contents

1. Chapter 2: Semiparametric regression analysis of interval-censored data
    - Transformation models for interval-censored data
    - Transformation models for multivariate interval-censored data
    - Marginal models for multivariate interval-censored data
    - Semiparametric regression models for panel count data

Reference

Zeng, D., Mao, L., & Lin, D. Y. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. Biometrika, 103(2), 253-271.

Interval-censored data

- Interval-censored data arise when the failure time is only known to lie within a broad time interval.

- Commonly encountered when the disease onset can only be ascertained through a small number of examinations.
  - HIV infection: periodic blood tests
  - Alzheimer's disease onset: periodic cognitive tests
  - Tumor occurrence: biopsies at periodic clinical visits

- Types of interval-censored data:
  - Case 1: only one examination time per subject, aka current status data
  - Case k ($k \geq 2$): k examination times per subject[1]
  - Mixed case: number of examination times varies among subjects

- Theoretical and computational challenges: no exact failure time

[1] *Huang, J., & Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. In Proceedings of the first Seattle symposium in biostatistics: survival analysis (pp. 123-169). New York, NY: Springer US.*

Transformation models

Notation:

- T: failure time
- $X(t)$: potentially time-dependent covariates
- $\Lambda(t|X)$: conditional cumulative hazard function for T given $X(\cdot)$

Semiparametric transformation model:

$$\Lambda(t|X) = G\left[\int_0^t \exp\left\{\beta^{\mathsf{T}} X(s)\right\} d\Lambda(s)\right]$$

- $G(\cdot)$: strictly increasing transformation function
  - $G(x) = x \Rightarrow$ proportional hazards model
  - $G(x) = \log(1 + x) \Rightarrow$ proportional odds model
- β: unknown regression parameters
- $\Lambda(\cdot)$: unknown increasing function

Frailty-induced transformations

Log-Laplace transform:

$$G(x) = -\log \int_0^\infty e^{-x\xi} f(\xi) d\xi$$

- ξ: frailty variable with support $[0, \infty)$
- $f(\xi)$: density function of ξ
  - Gamma density with mean 1 and variance $r \Rightarrow$ logarithmic transformations $G(x) = r^{-1}\log(1 + rx)$ $(r \geq 0)$
  - Positive stable distribution with parameter $\rho < 1 \Rightarrow$ Box-Cox transformations $G(x) = \{(1 + x)^\rho - 1\}/\rho$

Data

Raw data:

- Examination times: $U = (0 = U_0, U_1, \ldots, U_M, U_{M+1} = \infty)$
- Event statuses: $\Delta = (\Delta_0, \Delta_1, \ldots, \Delta_M)$, with $\Delta_l = I(U_l < T \leq U_{l+1})$
- Covariates: $X(t)$

Independent censoring assumption: $(U, M) \perp\!\!\!\perp T$ conditional on $X(t)$

Simplified data: (L, R, X)

- $L = \max\{U_l : U_l < T\}$ and $R = \min\{U_l : U_l \geq T\}$
- $L = 0 \Rightarrow$ left censoring
- $R = \infty \Rightarrow$ right censoring

Data from *n* independent samples: $\{\mathcal{O}_i = (L_i, R_i, X_i) : i = 1, \ldots, n\}$

Likelihood

Observed-data likelihood:

$$L_n(\beta,\Lambda)=\prod_{i=1}^{n}\left(\exp\left[-G\left\{\int_0^{L_i} e^{\beta^{\mathsf T}X_i(s)}d\Lambda(s)\right\}\right]-\exp\left[-G\left\{\int_0^{R_i} e^{\beta^{\mathsf T}X_i(s)}d\Lambda(s)\right\}\right]\right)$$

NPMLE:

$$\widetilde{L}_n(\beta,\Lambda)=\prod_{i=1}^{n}\left[\exp\left\{-G\left(\sum_{t_l\le L_i}\lambda_l e^{\beta^{\mathsf T}X_{il}}\right)\right\}-I(R_i<\infty)\exp\left\{-G\left(\sum_{t_l\le R_i}\lambda_l e^{\beta^{\mathsf T}X_{il}}\right)\right\}\right]$$

- $t_1<\cdots<t_m$: distinct values of all $L_i>0$ and $R_i<\infty$ $(i=1,\ldots,n)$
- λ_l: jump size of Λ at t_l $(l=1,\ldots,m)$
- $X_{il}=X_i(t_l)$

Poisson data augmentation

- Consider the class of frailty-induced transformations, $\widetilde{L}_n(\beta, \Lambda)$ can be written as

$$\prod_{i=1}^{n} \int_{\xi_i} \underbrace{\exp\left(-\xi_i \sum_{t_l \le L_i} \lambda_l e^{\beta^\top x_{il}}\right) \left\{1 - \exp\left(-\xi_i \sum_{L_i < t_l \le R_i} \lambda_l e^{\beta^\top x_{il}}\right)\right\}^{I(R_i < \infty)}}_{p(\mathcal{O}_i | \xi_i)} f(\xi_i)\, d\xi_i$$

- Direct maximization of $\widetilde{L}_n(\beta, \Lambda)$ over β and λ_l is difficult.
  - Lack of analytical expressions for λ_l
  - Many λ_l are zero and lie on the boundary of the parameter space

- We introduce latent variables $W_{il} \overset{\text{ind}}{\sim} \text{Poisson}(\xi_i \lambda_l e^{\beta^\top x_{il}})$. Then $p(\mathcal{O}_i | \xi_i)$ is equivalent to the probability of the event

$$\widetilde{\mathcal{O}}_i = \left(\sum_{t_l \le L_i} W_{il} = 0\right) \bigcap \left(\sum_{L_i < t_l \le R_i} W_{il} > 0\right)^{I(R_i < \infty)}$$

EM algorithm

- Therefore, maximizing $\widetilde{L}_n(\beta, \Lambda)$ is equivalent to maximizing the likelihood based on $\widetilde{\mathcal{O}}_i$ $(i = 1, \ldots, n)$.
- The maximization can be solved via an EM algorithm, treating ξ_i and W_{il} as missing data.
- Define $R_i^* = L_i I(R_i = \infty) + R_i I(R_i < \infty)$. The complete-data log-likelihood is

$$\sum_{i=1}^{n} \left[\sum_{l=1}^{m} I\left(t_l \leq R_i^*\right) \left\{ W_{il} \log(\xi_i \lambda_l e^{\beta^\top X_{il}}) - \xi_i \lambda_l e^{\beta^\top X_{il}} - \log W_{il}! \right\} + \log f\left(\xi_i\right) \right]$$

E-step

- In the E-step, we evaluate the posterior means $\widehat{E}(\xi_i)$ and $\widehat{E}(W_{il})$.
- The posterior density of ξ_i is proportional to $p(\mathcal{O}_i|\xi_i)f(\xi_i)$. Simple algebra yields

$$\widehat{E}\left(\xi_i\right) = \frac{\exp\left\{-G\left(S_{i1}\right)\right\} G'\left(S_{i1}\right) - I(R_i < \infty)\exp\left\{-G\left(S_{i2}\right)\right\} G'\left(S_{i2}\right)}{\exp\left\{-G\left(S_{i1}\right)\right\} - I(R_i < \infty)\exp\left\{-G\left(S_{i2}\right)\right\}},$$

where $S_{i1} = \sum_{t_l \le L_i} \lambda_l e^{\beta^\top X_{il}}$ and $S_{i2} = \sum_{t_l \le R_i} \lambda_l e^{\beta^\top X_{il}}$.

E-step (cont.)

- Clearly, $\widehat{E}(W_{il}) = 0$ if $t_l \leq L_i$.

- For $L_i < t_l \leq R_i$ with $R_i < \infty$,

$$\begin{aligned}
\widehat{E}(W_{il}) &= E_{\xi_i}\left\{E(W_{il}|\widetilde{\mathcal{O}}_i, \xi_i)\Big|\widetilde{\mathcal{O}}_i\right\} \\
&= E_{\xi_i}\left\{E\left(W_{il}\Bigg|\sum_{L_i<t_{l'}\leq R_i} W_{il'} > 0, \xi_i\right)\Bigg|\mathcal{O}_i\right\} \\
&= \widehat{E}\left[\frac{\xi_i \lambda_l e^{\beta^\top X_{il}}}{1 - \exp\{-\xi_i(S_{i2} - S_{i1})\}}\right]
\end{aligned}$$

- The integral over ξ_i can be approximated by Gaussian–Laguerre quadrature.

M-step

- In the M-step, we first update λ_l by

$$\lambda_l = \frac{\sum_{i=1}^n I(t_l \leq R_i^*)\widehat{E}(W_{il})}{\sum_{i=1}^n I(t_l \leq R_i^*)\widehat{E}(\xi_i)e^{\beta^\top X_{il}}}, \qquad \text{for } l = 1, \ldots, m$$

- After plugging the above λ_l into the conditional expectation of the complete-data log-likelihood, we can then update β by solving the equation

$$\sum_{i=1}^n \sum_{l=1}^m I(t_l \leq R_i^*)\widehat{E}(W_{il})\left\{ X_{il} - \frac{\sum_{j=1}^n I(t_l \leq R_j^*)\widehat{E}(\xi_j)e^{\beta^\top X_{jl}} X_{jl}}{\sum_{j=1}^n I(t_l \leq R_j^*)\widehat{E}(\xi_j)e^{\beta^\top X_{jl}}} \right\},$$

which can be solved using the one-step Newton-Raphson method.

Remarks

- By introducing Poisson variables, we turn the original nonconcave likelihood function to a weighted sum of Poisson log-likelihood functions, which is strictly concave.
- In the M-step, the high-dimensional parameters λ_l ($l = 1, \ldots, m$) have closed-form solutions. This avoids the inversion of any large Hessian matrices.
- The observed-data likelihood is guaranteed to increase after each iteration of the EM algorithm.

Variance estimation

- We use profile likelihood[2] to estimate the covariance matrix of $\widehat{\beta}$.
- Define the profile likelihood as

$$\mathrm{pl}_n(\beta) = \max_{\Lambda} \log L_n(\beta, \Lambda),$$

which can be computed using the same EM algorithm but with fixed β.

- The covariance matrix of $\widehat{\beta}$ can be estimated by

$$\widehat{V} = -\left[\left\{ \frac{\mathrm{pl}_n(\widehat{\beta}) - \mathrm{pl}_n(\widehat{\beta} + h_n e_j) - \mathrm{pl}_n(\widehat{\beta} + h_n e_k) + \mathrm{pl}_n(\widehat{\beta} + h_n e_j + h_n e_k)}{h_n^2} \right\}_{(j,k)}\right]^{-1},$$

where e_j is the jth canonical vector and h_n is a constant of order $n^{-1/2}$.

[2] *Murphy, S. A., & Van der Vaart, A. W. (2000). On profile likelihood. Journal of the American Statistical Association, 95(450), 449-465.*

Asymptotic theory

Consistency:

$$\|\widehat{\beta} - \beta_0\| + \sup_{t \in [0,\tau]} |\widehat{\Lambda}(t) - \Lambda_0(t)| \xrightarrow{a.s.} 0$$

Asymptotic normality & semiparametric efficiency:

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \widetilde{\mathcal{I}}_0^{-1}),$$

where $\widetilde{\mathcal{I}}_0$ is the efficient information matrix of β.

Consistency of variance estimator: $\|n\widehat{V} - \widetilde{\mathcal{I}}_0^{-1}\|_2 = o_p(1)$.

Mixed rate of convergence:

$$E\left[\sum_{l=1}^{M}\left\{\int_0^{U_l} e^{\widehat{\beta}^{\mathsf{T}} X(s)} d\widehat{\Lambda}(s) - \int_0^{U_l} e^{\beta_0^{\mathsf{T}} X(s)} d\Lambda_0(s)\right\}^2\right]^{1/2} = O_p(n^{-1/3})$$

Table of Contents

1. Chapter 2: Semiparametric regression analysis of interval-censored data
   - Transformation models for interval-censored data
   - Transformation models for multivariate interval-censored data
   - Marginal models for multivariate interval-censored data
   - Semiparametric regression models for panel count data

Reference

Zeng, D., Gao, F., & Lin, D. Y. (2017). Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. Biometrika, 104(3), 505-525.

Motivation

- Interval-censored multiple-event data
  - Study of chronic diseases: diabetes, hypertension, Alzheimer's disease
- Interval-censored clustered data
  - Twin/family study of infectious diseases
  - Dental caries
- Challenges:
  - Event times never exactly observed
  - Dependence between events from the same subject
  - Dependence within clusters

Transformation models

Notation:

- n: number of independent clusters
- J_i: number of subjects in the ith cluster
- K: number of event types
- T_{ijk}: kth event time for the jth subject of the ith cluster
- $X_{ijk}(t)$: potentially time-dependent covariates
- $b_i \sim N(0, \Sigma_i(\gamma))$: vector of random effects

Semiparametric transformation model:

$$\Lambda_{ijk}(t|X_{ijk}, b_i) = G_k\left[\int_0^t \exp\left\{\beta^{\mathsf{T}} X_{ijk}(s) + b_i^{\mathsf{T}} Z_{ijk}(s)\right\} d\Lambda_k(s)\right]$$

- $G_k(\cdot)$: type-specific transformation function
- β, γ: unknown regression parameters
- $Z_{ijk}(\cdot)$: contains 1 and part of $X_{ijk}(\cdot)$
- $\Lambda_k(\cdot)$: arbitrary increasing function

Transformation models (cont.)

Semiparametric transformation model:

$$\Lambda_{ijk}(t|X_{ijk}, b_i) = G_k \left[\int_0^t \exp\left\{\beta^\mathsf{T} X_{ijk}(s) + b_i^\mathsf{T} Z_{ijk}(s)\right\} d\Lambda_k(s) \right] \qquad (1)$$

- By letting X_{ijk} and Z_{ijk} depend on k, model (1) allows the regression parameters and random effects to vary across the K types of events.
- The dependence of Z_{ijk} on j allows for subject-specific random effects.
- $\Sigma_i(\gamma)$ usually does not depend on i, such that γ contains the upper diagonal elements of the common covariance matrix Σ.

Data

Examination times for T_{ijk}:

$$U_{ijk} = (0 = U_{ijk0}, U_{ijk1}, \ldots, U_{ijk,M_{ijk}}, U_{ijk,M_{ijk}+1} = \infty)$$

Data:

$$\left\{\mathcal{O}_{ijk} = (L_{ijk}, R_{ijk}, X_{ijk}) : \; i = 1, \ldots, n; j = 1, \ldots, J_i; k = 1, \ldots, K\right\},$$

where $(L_{ijk}, R_{ijk}]$ is the shortest time interval induced by U_{ijk} that brackets T_{ijk}.

Independent censoring assumption:

$$\{(U_{ijk}, M_{ijk}) : j = 1, \ldots, J_i; k = 1, \ldots, K\} \text{ are independent of } \{T_{ijk} : j = 1, \ldots, J_i; k = 1, \ldots, K\} \text{ and } b_i \text{ conditional on } \{X_{ijk}(\cdot) : j = 1, \ldots, J_i; k = 1, \ldots, K\}.$$

Likelihood

Let $\theta = (\beta^{\mathsf{T}}, \gamma^{\mathsf{T}})^{\mathsf{T}}$ and $\mathcal{A} = \{\Lambda_k\}_{k=1}^{K}$. The likelihood is

$$L_n(\theta, \mathcal{A}) = \prod_{i=1}^{n} \int_{b_i} \prod_{j=1}^{J_i} \prod_{k=1}^{K} \Bigg\{ \exp\left(-G_k \left[\int_0^{L_{ijk}} \exp\left\{ \beta^{\mathsf{T}} X_{ijk}(s) + b_i^{\mathsf{T}} Z_{ijk}(s) \right\} d\Lambda_k(s) \right] \right)$$
$$- \exp\left(-G_k \left[\int_0^{R_{ijk}} \exp\left\{ \beta^{\mathsf{T}} X_{ijk}(s) + b_i^{\mathsf{T}} Z_{ijk}(s) \right\} d\Lambda_k(s) \right] \right) \Bigg\}$$
$$\times (2\pi)^{-d_i/2} \left| \Sigma_i(\gamma) \right|^{-1/2} \exp\left\{ -\frac{b_i^{\mathsf{T}} \Sigma_i(\gamma)^{-1} b_i}{2} \right\} db_i$$

NPMLE: treat each Λ_k as a step function

- $t_{k1} < t_{k2} < \cdots < t_{km_k}$: distinct values of all $L_{ijk} > 0$ and $R_{ijk} < \infty$ $(i = 1, \ldots, n; j = 1, \ldots, J_i)$
- λ_{kl}: jump size of Λ_k at t_{kl} $(l = 1, \ldots, m_k)$
- $X_{ijkl} = X_{ijk}(t_{kl})$ and $Z_{ijkl} = Z_{ijk}(t_{kl})$

Likelihood (cont.)

Consider the class of frailty-induced transformations:

$$G_k(x) = -\log \int_0^\infty e^{-x\xi} f_k(\xi) d\xi$$

The likelihood can then be written as

$$\begin{aligned}
\widetilde{L}_n(\theta, \mathcal{A}) = \prod_{i=1}^{n} \int_{b_i} \prod_{j=1}^{J_i} \prod_{k=1}^{K} \int_{\xi_{ijk}} \Bigg[&\exp\Bigg\{ -\xi_{ijk} \sum_{t_{kl} \le L_{ijk}} \exp\left(\beta^{\mathrm{T}} X_{ijkl} + b_i^{\mathrm{T}} Z_{ijkl}\right) \lambda_{kl} \Bigg\} \\
&- I\left(R_{ijk} < \infty\right) \exp\Bigg\{ -\xi_{ijk} \sum_{t_{kl} \le R_{ijk}} \exp\left(\beta^{\mathrm{T}} X_{ijkl} + b_i^{\mathrm{T}} Z_{ijkl}\right) \lambda_{kl} \Bigg\} \Bigg] f_k\left(\xi_{ijk}\right) d\xi_{ijk} \\
&\times (2\pi)^{-d_i/2} \left|\Sigma_i(\gamma)\right|^{-1/2} \exp\left\{ -\frac{b_i^{\mathrm{T}} \Sigma_i(\gamma)^{-1} b_i}{2} \right\} db_i
\end{aligned}$$

Poisson data augmentation

Latent variables: for $i = 1, \dots, n; j = 1, \dots, J_i; k = 1, \dots, K; l = 1, \dots, m_k$,

$$W_{ijkl} \overset{\text{ind}}{\sim} \text{Poisson}\left\{\lambda_{kl}\xi_{ijk}\exp\left(\beta^{\mathsf{T}} X_{ijkl} + b_i^{\mathsf{T}} Z_{ijkl}\right)\right\}$$

Equivalent likelihood: Conditional on b_i and ξ_{ijk}, the probability of the event

$$\widetilde{\mathcal{O}}_{ijk} = \left(\sum_{t_{kl} \le L_{ijk}} W_{ijkl} = 0\right) \bigcap \left(\sum_{L_{ijk} < t_{kl} \le R_{ijk}} W_{ijkl} > 0\right)^{I(R_{ijk} < \infty)}$$

is equal to

$$p(\mathcal{O}_{ijk} | b_i, \xi_{ijk}) = \exp\left\{-\xi_{ijk} \sum_{t_{kl} \le L_{ijk}} \exp\left(\beta^{\mathsf{T}} X_{ijkl} + b_i^{\mathsf{T}} Z_{ijkl}\right)\lambda_{kl}\right\}$$
$$- I\left(R_{ijk} < \infty\right)\exp\left\{-\xi_{ijk} \sum_{t_{kl} \le R_{ijk}} \exp\left(\beta^{\mathsf{T}} X_{ijkl} + b_i^{\mathsf{T}} Z_{ijkl}\right)\lambda_{kl}\right\}$$

EM algorithm

- Therefore, maximizing $\widetilde{L}_n(\theta, \mathcal{A})$ is equivalent to maximizing the likelihood arising from $\{\widetilde{\mathcal{O}}_{ijk} : i = 1, \ldots, n; j = 1, \ldots, J_i; k = 1, \ldots, K\}$.

- The maximization can be done through an EM algorithm, treating b_i, ξ_{ijk} and W_{ijkl} as missing data.

- Define $R^*_{ijk} = L_{ijk} I(R_{ijk} = \infty) + R_{ijk} I(R_{ijk} < \infty)$. The complete-data log-likelihood is

$$\sum_{i=1}^{n} \Bigg\{ \sum_{j=1}^{J_i} \sum_{k=1}^{K} \Bigg(\sum_{l=1}^{m_k} I(t_{kl} \leq R^*_{ijk}) \left[W_{ijkl} \log \left\{ \lambda_{kl} \xi_{ijk} \exp\left(\beta^{\mathrm{T}} X_{ijkl} + b_i^{\mathrm{T}} Z_{ijkl}\right) \right\} \right.$$
$$\left. - \lambda_{kl} \xi_{ijk} \exp\left(\beta^{\mathrm{T}} X_{ijkl} + b_i^{\mathrm{T}} Z_{ijkl}\right) - \log\left(W_{ijkl}!\right) \right] + \log f_k\left(\xi_{ijk}\right) \Bigg)$$
$$- \frac{d_i}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_i(\gamma)| - \frac{b_i^{\mathrm{T}} \Sigma_i(\gamma)^{-1} b_i}{2} \Bigg\}$$

E-step

- In the E-step, we evaluate $\widehat{E}(W_{ijkl})$ and some $\widehat{E}\{H(\xi_{ijk}, b_i)\}$.

- We use the fact that the joint posterior density of ξ_{ijk} ($j = 1, \ldots, J_i$ and $k = 1, \ldots, K$) and b_i is proportional to

$$\prod_{j=1}^{J_i} \prod_{k=1}^{K} p(\mathcal{O}_{ijk}|b_i, \xi_{ijk}) f_k(\xi_{ijk}) \phi(b_i; \Sigma_i(\gamma))$$

- In addition, $E(W_{ijkl}|b_i, \xi_{ijk})$ is given by

$$\frac{I\left(L_{ijk} < t_{kl} \leq R_{ijk} < \infty\right) \lambda_{kl} \xi_{ijk} \exp\left(\beta^{\mathsf{T}} X_{ijkl} + b_i^{\mathsf{T}} Z_{ijkl}\right)}{1 - \exp\left\{-\sum_{L_{ijk} < t_{kl'} \leq R_{ijk}} \lambda_{kl'} \xi_{ijk} \exp\left(\beta^{\mathsf{T}} X_{ijkl'} + b_i^{\mathsf{T}} Z_{ijkl'}\right)\right\}}.$$

- Then we integrate the above expressions over b_i and ξ_{ijk} using Gaussian quadrature approximations.

M-step

- In the M-step, we first update λ_{kl} ($k = 1, \ldots, K$ and $l = 1, \ldots, m_k$) by

$$\lambda_{kl} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{J_i} I(t_{kl} \leq R^*_{ijk}) \widehat{E}\,(W_{ijkl})}{\sum_{i=1}^{n} \sum_{j=1}^{J_i} I(t_{kl} \leq R^*_{ijk}) \widehat{E}\left\{\xi_{ijk} \exp\left(\beta^{\mathsf{T}} X_{ijkl} + b_i^{\mathsf{T}} Z_{ijkl}\right)\right\}}$$

- Then we solve the following score equation for β using the one-step Newton-Raphson method:

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \sum_{k=1}^{K} \sum_{l=1}^{m_k} I\left(t_{kl} \leq R^*_{ijk}\right) \widehat{E}\,(W_{ijkl}) \times \Bigg[X_{ijkl} - \frac{\sum_{i'=1}^{n} \sum_{j'=1}^{J_i'} I\left(t_{kl} \leq R^*_{i'j'k}\right) X_{i'j'kl} \widehat{E}\left\{\xi_{i'j'k} \exp\left(\beta^{\mathsf{T}} X_{i'j'kl} + b_{i'}^{\mathsf{T}} Z_{i'j'kl}\right)\right\}}{\sum_{i'=1}^{n} \sum_{j'=1}^{J_i'} I\left(t_{kql} \leq R^*_{i'j'k}\right) \widehat{E}\left\{\xi_{i'j'k} \exp\left(\beta^{\mathsf{T}} X_{i'j'kl} + b_{i'}^{\mathsf{T}} Z_{i'j'kl}\right)\right\}} \Bigg]$$

- Finally, we maximize $-\log|\Sigma_i(\gamma)| - \widehat{E}\left\{b_i^{\mathsf{T}} \Sigma_i^{-1}(\gamma) b_i\right\}$ to update γ.

Remarks

- The starting values of the EM algorithm can be $\beta = 0$, $\lambda_{kl} = 1/m_k$, and $\Sigma_i = I_{d_i}$.
- Due to the presence of the random effects, the conditional expectations in the E-step are more complicated than those in Zeng et al. (2016).
- The high-dimensional parameters λ_{kl} are calculated explicitly in the M-step.
- Each iteration of the EM algorithm guarantees an increase in the likelihood.

Variance estimation

- Define the profile likelihood

$$\text{pl}_n(\theta) = \max_{\mathcal{A}} \log L_n(\theta, \mathcal{A}),$$

which can be computed using the same EM algorithm but with fixed θ.

- The covariance matrix of $\widehat{\theta}$ can be estimated by

$$\widehat{V} = \left(\left[\sum_{i=1}^{n} \frac{\left\{ \text{pl}_{ni}(\widehat{\theta} + h_n e_j) - \text{pl}_{ni}(\widehat{\theta}) \right\} \left\{ \text{pl}_{ni}(\widehat{\theta} + h_n e_k) - \text{pl}_{ni}(\widehat{\theta}) \right\}}{h_n^2} \right]_{(j,k)} \right)^{-1},$$

where pl_{ni} denotes the ith cluster's contribution to pl_n, e_j is the jth canonical vector, and h_n is a constant of order $n^{-1/2}$.

- Compared to the variance estimator in Zeng et al. (2016), this $\widehat{V}$ is guaranteed to be positive semidefinite and is more robust w.r.t. the choice of h_n.

Asymptotic theory

Consistency: $\|\widehat{\theta} - \theta_0\| + \sum_{k=1}^{K} \sup_{t \in [0, \tau_k]} |\widehat{\Lambda}_k(t) - \Lambda_{0k}(t)| \xrightarrow{a.s.} 0$

Asymptotic normality & semiparametric efficiency:

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \widetilde{\mathcal{I}}_0^{-1}),$$

where $\widetilde{\mathcal{I}}_0$ is the efficient information matrix of θ.

Consistency of variance estimator: $\|n\widehat{V} - \widetilde{\mathcal{I}}_0^{-1}\|_2 = o_p(1)$.

Convergence rate for $\widehat{\Lambda}_k$:

$$E\left[\sum_{j=1}^{J_i} \sum_{k=1}^{K} \sum_{l=0}^{M_{ijk}} \left\{\widehat{\Lambda}_k\left(U_{ijkl}\right) - \Lambda_{0k}\left(U_{ijkl}\right)\right\}^2\right] = O_p\left(n^{-2/3} + \|\widehat{\beta} - \beta_0\|^2 + \|\widehat{\gamma} - \gamma_0\|^2\right).$$

Software

The methods developed in Zeng et al. (2016, 2017) have been implemented in IntCens (https://dlin.web.unc.edu/software/intcens).

Table of Contents

1. Chapter 2: Semiparametric regression analysis of interval-censored data
   - Transformation models for interval-censored data
   - Transformation models for multivariate interval-censored data
   - Marginal models for multivariate interval-censored data
   - Semiparametric regression models for panel count data

Reference

Xu, Y., Zeng, D., & Lin, D. Y. (2023). Marginal proportional hazards models for multivariate interval-censored data. Biometrika, 110(3), 815-830.

Motivation

Random-effects models have several limitations.

- Random effects may not adequately capture the dependence
- Model misspecification may lead to invalid statistical inference
- Computationally demanding
- Interpretation of β does not pertain to population-average effects

Marginal models formulate marginal distributions of multivariate event times through univariate regression models while leaving the dependence structures completely unspecified.

- More robust inference
- Faster and more stable computation
- Interpretation of population-average effects for β

Marginal models

Notation:

- n: number of independent clusters
- J_i: number of subjects in the ith cluster
- K: number of event types
- T_{ijk}: kth event time for the jth subject of the ith cluster
- $X_{ijk}(t)$: potentially time-dependent covariates
- $\lambda_{ijk}(t|X_{ijk})$: marginal hazard function for T_{ijk} conditional on X_{ijk}

Marginal Cox model:

$$\lambda_{ijk}(t|X_{ijk}) = \lambda_k(t) \exp\left\{\beta_k^{\mathsf{T}} X_{ijk}(t)\right\}$$

- β_k: type-specific regression parameters
- $\lambda_k(\cdot)$: arbitrary baseline hazard function
- $\Lambda_k(t) = \int_0^t \lambda_k(s)ds$

Remarks

Marginal Cox model:

$$\lambda_{ijk}(t|X_{ijk}) = \lambda_k(t) \exp\left\{\beta_k^{\mathsf{T}} X_{ijk}(t)\right\}$$

- The dependence structures of the event times within a cluster and between the K types of events are unspecified.
- By letting X_{ijk} depend on k, we allow different sets of covariates for different event types.

Estimation

Data: $\{(L_{ijk}, R_{ijk}, X_{ijk}) : i = 1, \ldots, n; j = 1, \ldots, J_i; k = 1, \ldots, K\}$

Independence working assumption: all event times are independent conditional on covariates

Pseudo-likelihood:

$$\widetilde{L}_k(\beta_k, \Lambda_k) = \prod_{i=1}^{n} \prod_{j=1}^{J_i} \left(\exp\left[- \int_0^{L_{ijk}} \exp\left\{ \beta_k^{\mathsf{T}} X_{ijk}(t) \right\} d\Lambda_k(t) \right] - \exp\left[- \int_0^{R_{ijk}} \exp\left\{ \beta_k^{\mathsf{T}} X_{ijk}(t) \right\} d\Lambda_k(t) \right] \right)$$

Nonparametric maximum pseudo-likelihood estimation:

- Extension of NPMLE
- (β_k, Λ_k) can be estimated using the approach developed in Zeng et al. (2016), i.e., Poisson data augmentation + EM algorithm.

Variance estimation

- Define the profile pseudo-log-likelihood for β_k as

$$\mathrm{pl}_k(\beta_k) = \max_{\Lambda_k} \widetilde{L}_k(\beta_k, \Lambda_k)$$

- We estimate $\mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_l)$ by the sandwich covariance estimator

$$\widehat{V}_{kl} = \left\{ D^2_{h_n} \mathrm{pl}_k(\widehat{\beta}_k) \right\}^{-1} \left\{ \sum_{i=1}^{n} D_{h_n} \mathrm{pl}_{ki}(\widehat{\beta}_k) D_{h_n} \mathrm{pl}_{li}(\widehat{\beta}_l)^{\mathrm{T}} \right\} \left\{ D^2_{h_n} \mathrm{pl}_l(\widehat{\beta}_l) \right\}^{-1}$$

  - pl_{ki}: ith cluster's contribution to pl_k
  - D_{h_n} and $D^2_{h_n}$: first- and second-order numerical derivatives with perturbation constant $h_n = O(n^{-1/2})$
  - account for the dependence within clusters and between event types

Asymptotic properties

Let $\theta = (\beta_1^\mathsf{T}, \ldots, \beta_K^\mathsf{T})^\mathsf{T}$.

Consistency: $\|\widehat{\theta} - \theta_0\| + \sum_{k=1}^{K} \sup_{t \in [0, \tau_k]} |\widehat{\Lambda}_k(t) - \Lambda_{0k}(t)| \stackrel{a.s.}{\rightarrow} 0$

Asymptotic normality: $\sqrt{n}(\widehat{\theta} - \theta_0) \stackrel{d}{\rightarrow} N(0, \Omega)$

Consistency of variance estimator: $\{n\widehat{V}_{kl}\}_{(k,l)}$ is consistent for Ω, regardless of the dependence structures.

Simultaneous Inference

- Parameters of interest: $\eta_k = \beta_{k1}$ $(k = 1, \ldots, K)$
- Estimators: $\widehat{\eta}_k = \widehat{\beta}_{k1}$ $(k = 1, \ldots, K)$
- Covariance estimator: $\widehat{\Psi} = \{\widehat{V}_{kl,11}\}_{(k,l)}$
- Global (Wald) test $H_0 : \eta_1 = \cdots = \eta_K = 0$

$$W = (\widehat{\eta}_1, \ldots, \widehat{\eta}_K)\widehat{\Psi}^{-1}(\widehat{\eta}_1, \ldots, \widehat{\eta}_K)^{\mathrm{T}} \xrightarrow{d} \chi^2_K \quad \text{under } H_0$$

- To make inference on an overall covariate effect, we can estimate a common parameter $\eta_1 = \cdots = \eta_K = \eta$ by

$$\widehat{\eta} = \sum_{k=1}^{K} c_k \widehat{\eta}_k$$

  - optimal weights c_k are chosen to minimize $\mathrm{Var}(\widehat{\eta})$
  - more efficient than the separate estimators $\widehat{\eta}_k$
  - powerful test for no covariate effect on the K events

Table of Contents

1. Chapter 2: Semiparametric regression analysis of interval-censored data
   - Transformation models for interval-censored data
   - Transformation models for multivariate interval-censored data
   - Marginal models for multivariate interval-censored data
   - Semiparametric regression models for panel count data

Reference

Zeng, D., & Lin, D. Y. (2021). Maximum likelihood estimation for semiparametric regression models with panel count data. Biometrika, 108(4), 947-963.

Panel count data

- Panel count data arise when only the number of recurrent events between successive examinations can be observed.
  - number of tumors in a cancer patient
  - number of damaged joints in a psoriatic arthritis patient
  - number of decayed teeth in a child

- Investigators are often interested in evaluating the effects of covariates (e.g., treatment) on the recurrent event process.

- Challenges:
  - Unknown recurrent event times
  - Within-subject correlations of recurrent event times
  - Within-subject correlations between different types of recurrent events

Notation

- n: number of subjects
- K: number of types of recurrent events
- $N_{ik}(t)$: number of the kth type of event that the ith subject has experienced by time t ($i = 1, \ldots, n$ and $k = 1, \ldots, K$); non-homogeneous Poisson process
- $X_i(t)$: potentially time-dependent covariates
- b_{ik}: random effects for the kth type of event
- ξ_i: random effects shared by the K types of events
- $\lambda_{ik}(t|X_i, b_{ik}, \xi_i)$: conditional intensity function for $N_{ik}(t)$

Proportional intensity models

$$\lambda_{ik}(t|X_i, b_{ik}, \xi_i) = \lambda_k(t) \exp\left\{\beta_k^{\mathsf{T}} X_i(t) + b_{ik}^{\mathsf{T}} Z_i(t) + \xi_i^{\mathsf{T}} \widetilde{Z}_i(t)\right\}$$

- $\lambda_k(t)$: unknown baseline intensity function
- β_k: unknown type-specific regression parameters
- $Z_i(t)$ and $\widetilde{Z}_i(t)$: contain 1 and part of $X_i(t)$
- $b_{ik} \sim N(0, \Sigma_k)$: accounts for within-subject correlations among recurrent event times of the kth type
- $\xi_i \sim N(0, \Psi)$: accounts for within-subject correlations between different recurrent event processes
- b_{ik} $(k = 1, \ldots, K)$ and ξ_i are mutually independent.
- If $K = 1$, ξ_i is omitted.

Data

Panel count data for the *i*th subject:

- Examination times:

$$U_{ik} = (0 = U_{ik0}, U_{ik1}, \ldots, U_{ikM_{ik}}), \quad \text{for } k = 1, \ldots, K$$

- Event counts:

$$\Delta_{ik} = (\Delta_{ik1}, \ldots, \Delta_{ikM_{ik}}), \quad \text{for } k = 1, \ldots, K$$

  with $\Delta_{ikj} = N_{ik}(U_{ikj}) - N_{ik}(U_{ik,j-1})$.

- Covariates: $X_i(t)$

Independent censoring assumption: $(U_{i1}, \ldots, U_{iK})$ are independent of $(N_{i1}, \ldots, N_{iK})$, $(b_{i1}, \ldots, b_{iK})$, and ξ_i conditional on $X_i(\cdot)$.

Likelihood

- $\{\Delta_{ikj}\}_{j=1}^{M_{ik}} \overset{\text{ind}}{\sim}$ Poisson with means $\int_{U_{ik,j-1}}^{U_{ikj}} \lambda_{ik}(t|X_i, b_{ik}, \xi_i)dt$
- The likelihood is proportional to

$$\prod_{i=1}^{n}\Bigg[\int_{\xi_i}\phi\left(\xi_i;\Psi\right)\prod_{k=1}^{K}\int_{b_{ik}}\phi\left(b_{ik};\Sigma_k\right)\prod_{j=1}^{M_{ik}}\frac{\left\{\int_{U_{ik,j-1}}^{U_{ikj}} e^{\beta_k^{\mathsf{T}}X_i(t)+b_{ik}^{\mathsf{T}}Z_i(t)+\xi_i^{\mathsf{T}}\widetilde{Z}_i(t)}d\Lambda_k(t)\right\}^{\Delta_{ikj}}}{\Delta_{ikj}!}$$

$$\times \exp\left\{-\int_0^{U_{ikM_{ik}}} e^{\beta_k^{\mathsf{T}}X_i(t)+b_{ik}^{\mathsf{T}}Z_i(t)+\xi_i^{\mathsf{T}}\widetilde{Z}_i(t)}d\Lambda_k(t)\right\} db_{ik}\,d\xi_i\Bigg]$$

  - $\Lambda_k(t) = \int_0^t \lambda_k(s)ds$: cumulative baseline intensity function
  - $\phi(\cdot;\Sigma)$: multivariate normal density with mean 0 and covariance matrix Σ

Estimation

NPMLE:

- $0 < t_{k1} < t_{t2} < \cdots < t_{km_k}$: unique values of U_{ik} $(i = 1, \ldots, n)$
- λ_{kl}: jump size of Λ_k at t_{kl} $(l = 1, \ldots, m_k)$
- $X_{ikl} = X_i(t_{kl})$, $Z_{ikl} = Z_i(t_{kl})$, and $\widetilde{Z}_{ikl} = \widetilde{Z}_i(t_{kl})$

New likelihood:

$$\prod_{i=1}^{n}\left\{\int_{\xi_i} \phi(\xi_i; \Psi) \prod_{k=1}^{K} \int_{b_{ik}} \phi(b_{ik}; \Sigma_k) \prod_{j=1}^{M_{ik}} \frac{\left(\sum_{l: t_{kl} \in (U_{ik,j-1}, U_{ikj}]} \lambda_{kl} e^{\beta_k^{\mathrm{T}} X_{ikl} + b_{ik}^{\mathrm{T}} Z_{ikl} + \xi_i^{\mathrm{T}} \widetilde{Z}_{ikl}}\right)^{\Delta_{ikj}}}{\Delta_{ikj}!} \right.$$
$$\left. \times \exp\left(-\sum_{l: t_{kl} \leq U_{ikM_{ik}}} \lambda_{kl} e^{\beta_k^{\mathrm{T}} X_{ikl} + b_{ik}^{\mathrm{T}} Z_{ikl} + \xi_i^{\mathrm{T}} \widetilde{Z}_{ikl}}\right) db_{ik}\, d\xi_i \right\}$$

Direct maximization is infeasible due to lack of analytic expressions for λ_{kl}.

Poissonization

- We introduce independent latent Poisson variables W_{ikl} with means $\lambda_{kl} e^{\beta_k^\top X_{ikl} + b_{ik}^\top Z_{ikl} + \xi_i^\top \widetilde{Z}_{ikl}}$, for $i = 1, \ldots, n$; $k = 1, \ldots, K$; $l = 1, \ldots, m_k$.

- It is easy to see that the likelihood for Δ_{ikj} is the same as the likelihood for $\sum_{l: t_{kl} \in (U_{ik,j-1}, U_{ikj}]} W_{ikl} = \Delta_{ikj}$.

- Thus, we can maximize the likelihood through an EM algorithm, with W_{ikl}, b_{ik}, and ξ_i as missing data.

EM algorithm

The complete-data log-likelihood is

$$\sum_{i=1}^{n}\Bigg[-\frac{1}{2}\log(2\pi)^{q}|\Psi| - \frac{1}{2}\xi_i^{\mathsf{T}}\Psi^{-1}\xi_i + \sum_{k=1}^{K}\left\{-\frac{1}{2}\log(2\pi)^{p}\left|\Sigma_k\right| - \frac{1}{2}b_{ik}^{\mathsf{T}}\Sigma_k^{-1}b_{ik}\right\}$$
$$+\sum_{k=1}^{K}\sum_{l=1}^{m_k} I\left(t_{kl} \le U_{ikM_{ik}}\right)\left\{W_{ikl}\left(\log\lambda_{kl} + \beta_k^{\mathsf{T}}X_{ikl} + b_{ik}^{\mathsf{T}}Z_{ikl} + \xi_i^{\mathsf{T}}\widetilde{Z}_{ikl}\right)\right.$$
$$\left.\left. -\lambda_{kl}e^{\beta_k^{\mathsf{T}}X_{ikl}+b_{ik}^{\mathsf{T}}Z_{ikl}+\xi_i^{\mathsf{T}}\widetilde{Z}_{ikl}} - \log W_{ikl}!\right\}\right].$$

E-step

- In the E-step, we compute $\widehat{E}(W_{ikl})$ and some $\widehat{E}\{H(b_{ik}, \xi_i)\}$.
- For any $t_{kl} \in (U_{ik,j-1}, U_{ikj}]$, conditional on Δ_{ikj}, the covariates and random effects, W_{ikl} follows a binomial distribution with success probability

$$p_{ikl} = \frac{\lambda_{kl} e^{\beta_k^{\mathsf{T}} X_{ikl} + b_{ik}^{\mathsf{T}} Z_{ikl} + \xi_i^{\mathsf{T}} \widetilde{Z}_{ikl}}}{\sum_{q: t_{kq} \in (U_{ik,j-1}, U_{ikj}]} \lambda_{kq} e^{\beta_k^{\mathsf{T}} X_{ikq} + b_{ik}^{\mathsf{T}} Z_{ikq} + \xi_i^{\mathsf{T}} \widetilde{Z}_{ikq}}}.$$

Thus, $\widehat{E}(W_{ikl}) = \Delta_{ikj} \widehat{E}(p_{ikl})$.

E-step (cont.)

- In addition, the joint posterior density of $\{b_{ik}\}_{k=1}^{K}$ and ξ_i is proportional to

$$\phi(\xi_i;\Psi)\prod_{k=1}^{K}\Bigg\{\phi(b_{ik};\Sigma_k)\prod_{j=1}^{M_{ik}}\frac{\left(\sum_{l:t_{kl}\in(U_{ik,j-1},U_{ikj}]}\lambda_{kl}e^{\beta_k^{\mathsf{T}}X_{ikl}+b_{ik}^{\mathsf{T}}Z_{ikl}+\xi_i^{\mathsf{T}}\widetilde{Z}_{ikl}}\right)^{\Delta_{ikj}}}{\Delta_{ikj}!}$$
$$\times\exp\left(-\sum_{l:t_{kl}\leq U_{ikM_{ik}}}\lambda_{kl}e^{\beta_k^{\mathsf{T}}X_{ikl}+b_{ik}^{\mathsf{T}}Z_{ikl}+\xi_i^{\mathsf{T}}\widetilde{Z}_{ikl}}\right)\Bigg\}$$

- The conditional expectations can then be calculated, with integrals over b_{ik} and ξ_i approximated by Gauss–Hermite quadrature.

M-step

- In the M-step, we first update λ_{kl} by

$$\lambda_{kl} = \frac{\sum_{i=1}^{n} I(U_{ikM_{ik}} \leq t_{kl})\widehat{E}(W_{ikl})}{\sum_{i=1}^{n} I(U_{ikM_{ik}} \leq t_{kl})\widehat{E}(e^{\beta_k^{\mathsf{T}} X_{ikl} + b_{ik}^{\mathsf{T}} Z_{ikl} + \xi_i^{\mathsf{T}} \widetilde{Z}_{ikl}})}$$

- Then we update β_k by applying the one-step Newton-Raphson method to the score equation

$$0 = \sum_{i=1}^{n} \sum_{l=1}^{m_k} I(U_{ikM_{ik}} \leq t_{kl})\widehat{E}(W_{ikl}) \Bigg\{ X_{ikl} - \frac{\sum_{i'=1}^{n} I(U_{i'kM_{i'k}} \leq t_{kl})\widehat{E}(e^{\beta_k^{\mathsf{T}} X_{i'kl} + b_{i'k}^{\mathsf{T}} Z_{i'kl} + \xi_{i'}^{\mathsf{T}} \widetilde{Z}_{i'kl}}) X_{i'kl}}{\sum_{i'=1}^{n} I(U_{i'kM_{i'k}} \leq t_{kl})\widehat{E}(e^{\beta_k^{\mathsf{T}} X_{i'kl} + b_{i'k}^{\mathsf{T}} Z_{i'kl} + \xi_{i'}^{\mathsf{T}} \widetilde{Z}_{i'kl}})} \Bigg\}$$

- Finally, we set $\Sigma_k = n^{-1} \sum_{i=1}^{n} \widehat{E}(b_{ik}^{\otimes 2})$ and $\Psi = n^{-1} \sum_{i=1}^{n} \widehat{E}(\xi_i^{\otimes 2})$.

Prediction

- We can use the event history to improve the prediction of future events.
- The event history at the current examination time t_0 consists of

$$\mathcal{H}(t_0) = \{n_k = N_k(t_0) : k = 1, \ldots, K\}$$

- The key is to update the posterior density of the random effect ξ based on $\mathcal{H}(t_0)$, which is proportional to

$$\widetilde{\phi}(\xi; \mathcal{H}(t_0)) = \phi(\xi; \widehat{\Psi}) \prod_{k=1}^{K} \int_{b_k} \left\{ \int_0^{t_0} e^{\widehat{\beta}_k^{\mathsf{T}} X(t) + b_k^{\mathsf{T}} Z(t) + \xi^{\mathsf{T}} \widetilde{Z}(t)} d\widehat{\Lambda}_k(t) \right\}^{n_k}$$
$$\times \exp\left\{ - \int_0^{t_0} e^{\widehat{\beta}_k^{\mathsf{T}} X(t) + b_k^{\mathsf{T}} Z(t) + \xi^{\mathsf{T}} \widetilde{Z}(t)} d\widehat{\Lambda}_k(t) \right\} \phi(b_k; \widehat{\Sigma}_k) db_k$$

- Then, the new event count of the kth type at $t_1 > t_0$ can be predicted by

$$\frac{\int_{\xi} \widetilde{\phi}(\xi; \mathcal{H}(t_0)) \int_{t_0}^{t_1} e^{\widehat{\beta}_k^{\mathsf{T}} X(t) + b_k^{\mathsf{T}} Z(t) + \xi^{\mathsf{T}} \widetilde{Z}(t)} d\widehat{\Lambda}_k(t) d\xi}{\int_{\xi} \widetilde{\phi}(\xi; \mathcal{H}(t_0)) d\xi}$$

Asymptotic properties

Let θ contain β_k and the upper triangular elements of Σ_k and Ψ $(k = 1, \ldots, K)$.

Consistency: $\|\widehat{\theta} - \theta_0\| + \sum_{k=1}^{K} \sup_{t \in [0,\tau_k]} |\widehat{\Lambda}_k(t) - \Lambda_{0k}(t)| \overset{a.s.}{\to} 0$

Asymptotic normality & semiparametric efficiency: $\sqrt{n}(\widehat{\theta} - \theta_0)$ converges weakly to a mean-zero normal random vector whose covariance matrix attains the semiparametric efficiency bound.

Variance estimation: The limiting covariance matrix of $\sqrt{n}(\widehat{\theta} - \theta_0)$ can be consistently estimated by the inverse of the matrix whose (j, l)th element is

$$n^{-1} \sum_{i=1}^{n} \left\{ \frac{\mathrm{pl}_{ni}(\widehat{\theta} + h_n e_j) - \mathrm{pl}_{ni}(\widehat{\theta})}{h_n} \right\} \left\{ \frac{\mathrm{pl}_{ni}(\widehat{\theta} + h_n e_l) - \mathrm{pl}_{ni}(\widehat{\theta})}{h_n} \right\}$$

- $\mathrm{pl}_{ni}(\theta)$: ith subject's contribution to the profile likelihood for θ
- $h_n = O(n^{-1/2})$

Concluding remarks

- The rationale behind Poisson data augmentation is that conditional on latent variables, the counting process $N(t)$ is a non-homogeneous Poisson process with intensity function the same as the hazard/intensity function for the failure time.

- With interval-censored data, the convergence rate of $\widehat{\Lambda}$ is usually slower than $\sqrt{n}$. In all the papers discussed, $\widehat{\Lambda}$ converges at a $n^{1/3}$ rate.

- However, the finite-dimensional component of the estimators is still asymptotically normal and efficient, and the limiting variance can be consistently estimated using profile likelihood.

Related work

- Mao, L., Lin, D. Y., & Zeng, D. (2017). Semiparametric regression analysis of interval-censored competing risks data. Biometrics, 73(3), 857-865.
- Gao, F., Zeng, D., & Lin, D. Y. (2018). Semiparametric regression analysis of interval-censored data with informative dropout. Biometrics, 74(4), 1213-1222.
- Gao, F., Zeng, D., Couper, D., & Lin, D. Y. (2019). Semiparametric regression analysis of multiple right-and interval-censored events. Journal of the American Statistical Association, 114(527), 1232-1240.
- Xu, Y., Zeng, D., & Lin, D. Y. (2024). Proportional rates models for multivariate panel count data. Biometrics, 80(1), ujad011.
- Gu, Y., Zeng, D., Heiss, G., & Lin, D. Y. (2024). Maximum likelihood estimation for semiparametric regression models with interval-censored multistate data. Biometrika, 111(3), 971-988.