STAT3655 Survival Analysis

Yu Gu, PhD
Assistant Professor

Department of Statistics & Actuarial Science
The University of Hong Kong

Table of Contents

Table of Contents

Table of Contents

Multivariate Failure Time Data

In general, there are two types of multivariate failure time data:

Multiple events data: each study subject may experience several events/failures.

Clustered data: there exists natural/artificial clustering of subjects which induces dependence among failure times of the same cluster.

Multiple Events Data

Recurrent events: repetitions of a phenomenon (e.g., illness)

- tumor recurrences
- infection episodes
- repeated breakdowns of machinery

Multiple types of events: combination of multiple types of phenomena

- Ordered events: natural ordering of successive events
  - HIV-infection $\rightarrow$ AIDS $\rightarrow$ death
  - randomization $\rightarrow$ cancer recurrence $\rightarrow$ death
  - birth $\rightarrow$ marriage $\rightarrow$ child birth
- Unordered events: several concurrent failure processes
  - physical symptoms or diseases in several organ systems (cardiovascular disease, cancer, Alzheimer's disease, etc.)
  - purchases of various products

Clustered Data

- family (twin) studies
- litter matched carcinogeneicity experiments
- group randomized studies

Real Example: Bladder Tumor Study

A randomized clinical trial was conducted to assess the efficacy of thiotepa in reducing cancer recurrences in patients with superficial bladder tumors.

		Cancer recurrences						
Group	n	0	1	2	3	4	> 4	Total
Thiotepa	38	20	8	3	2	2	3	45
Placebo	48	19	10	4	6	2	7	87

[Recurrent events]

Real Example: Colon Cancer Study

A randomized clinical trial was conducted in the 1980's to study the drugs Lev and 5-FU for adjuvant therapy of resected colon carcinoma. Patients with Stage C disease were randomly assigned to observation, Lev alone, or Lev+5-FU. The time to cancer recurrence and the survival time were both considered important outcome measures.

Group	#Patients	#Recurrences	#Deaths
observation	315	155	114
Lev	310	144	109
Lev+5-FU	304	103	78

[Multiple types of events]

Real Example: Diabetic Retinopathy Study

The Diabetic Retinopathy Study was conducted by the National Eye Institute to assess the effectiveness of laser photocoagulation in delaying visual loss in patients with diabetic retinopathy. One eye of each patient was randomly selected for photocoagulation and the other eye was observed without treatment. The patients were followed over several years for the occurrence of visual loss in the left and right eyes.

Treatment	Patients (eyes)	Visual loss
Yes	1727	242
No	1727	535

[Clustered data]

Analysis of Multivariate Failure Time Data

Questions:

- Distributions of multivariate failure times (joint, marginal and conditional distributions)
- Effects of covariates (e.g., treatment) on multivariate failure times

Challenges:

- Dependence of failure times within the same subject/cluster
- Censoring due to patient withdrawal/study termination
- Multiplicity of outcome measures

Table of Contents

Intensity Model

Notation:

- $N^*(t)$: counting process recording the number of events by time t
- $X(t)$: potentially time-dependent covariates
- $\mathcal{F}_t = \{N^*(s), X(s) : 0 \le s \le t\}$: history up to time t
- $dN^*(t)$: increment of N^* (i.e., number of events) over $[t, t + dt)$
- $\lambda(t; X) = \lim_{dt \downarrow 0} E\{dN^*(t) \mid \mathcal{F}_{t-}\}/dt$: intensity function

Intensity Model:

$$\lambda(t; X) = \lambda_0(t) \exp\left\{\beta^{\mathsf{T}} X(t)\right\}$$

- $\lambda_0(t)$: arbitrary baseline intensity function
- β: unknown regression parameters
- Also known as Andersen–Gill model

Recurrent Events Data

Notation:

- T_{ij}: jth event time of the ith subject
- $N_i^*(t) = \sum_{j=1}^{\infty} I(T_{ij} \le t)$: counting process of the ith subject
- C_i: censoring time of the ith subject
- $N_i(t) = N_i^*\{\min(t, C_i)\}$: observed counting process
- $\delta_{ij} = I(T_{ij} \le C_i)$: failure indicator
- $R_i(t) = I(C_i \ge t)$: at-risk indicator
- For $r = 0, 1, 2$, define

$$S^{(r)}(t; \beta) = \sum_{i=1}^{n} R_i(t) \exp\{\beta^{\mathsf{T}} X_i(t)\} X_i(t)^{\otimes r}$$

Independent censoring assumption: $C_i \perp\!\!\!\perp N^*(t)$ given $X(t)$

Maximum Partial Likelihood Estimation

$$L_n(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{\infty} \left[\frac{\exp\{\beta^{\mathsf{T}} X_i(T_{ij})\}}{S^{(0)}(T_{ij};\beta)} \right]^{\delta_{ij}}$$

$$\ell_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{\infty} \delta_{ij} \left\{ \beta^{\mathsf{T}} X_i(T_{ij}) - \log S^{(0)}(T_{ij};\beta) \right\}$$

$$U_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{\infty} \delta_{ij} \left\{ X_i(T_{ij}) - \frac{S^{(1)}(T_{ij};\beta)}{S^{(0)}(T_{ij};\beta)} \right\}$$

$$\mathcal{I}_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{\infty} \delta_{ij} \left\{ \frac{S^{(2)}(T_{ij};\beta)}{S^{(0)}(T_{ij};\beta)} - \frac{\left\{S^{(1)}(T_{ij};\beta)\right\}^{\otimes 2}}{\left\{S^{(0)}(T_{ij};\beta)\right\}^{2}} \right\}$$

Unified Counting Process Representation

In fact, using the counting process notation, partial likelihood for single and recurrent events data can be expressed uniformly:

$$L_n(\beta) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{\exp\{\beta^\mathsf{T} X_i(t)\}}{S^{(0)}(t;\beta)} \right\}^{dN_i(t)}$$

$$\ell_n(\beta) = \sum_{i=1}^{n} \int_0^\infty \left\{ \beta^\mathsf{T} X_i(t) - \log S^{(0)}(t;\beta) \right\} dN_i(t)$$

$$U_n(\beta) = \sum_{i=1}^{n} \int_0^\infty \left\{ X_i(t) - \frac{S^{(1)}(t;\beta)}{S^{(0)}(t;\beta)} \right\} dN_i(t)$$

$$\mathcal{I}_n(\beta) = \sum_{i=1}^{n} \int_0^\infty \left\{ \frac{S^{(2)}(t;\beta)}{S^{(0)}(t;\beta)} - \frac{\left\{S^{(1)}(t;\beta)\right\}^{\otimes 2}}{\left\{S^{(0)}(t;\beta)\right\}^2} \right\} dN_i(t)$$

For single event, $R_i(t) = I(Y_i \geq t)$. For recurrent events, $R_i(t) = I(C_i \geq t)$.

Asymptotic Properties

(i) $U_n(\beta) \dot{\sim} N\big(0, \mathcal{I}_n(\beta)\big)$.

(ii) $\widehat{\beta} \dot{\sim} N\big(\beta, \mathcal{I}_n^{-1}(\widehat{\beta})\big)$.

For hypothesis testing on β, score and Wald test statistics can be derived based on the asymptotic normality of $U_n(\beta)$ and $\widehat{\beta}$, respectively.

- $H_0 : \beta = \beta^*$. $SC_n = U_n(\beta^*)^{\mathsf{T}} \mathcal{I}_n(\beta^*)^{-1} U_n(\beta^*) \xrightarrow{d} \chi_p^2$.
- $H_0 : C\beta = C\beta^*$. $W_n = (C\widehat{\beta} - C\beta^*)^{\mathsf{T}} \big\{ C\mathcal{I}_n^{-1}(\widehat{\beta}) C^{\mathsf{T}} \big\}^{-1} (C\widehat{\beta} - C\beta^*) \xrightarrow{d} \chi_q^2$ for a $q \times p$ matrix C.

Implementation in R

Data input for the ith subject:

sid	start	stop	status	X1	X2
i	0	T_{i1}	1	X_{i1}	X_{i2}
i	T_{i1}	T_{i2}	1	X_{i1}	X_{i2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	T_{i,J_i-1}	T_{i,J_i}	1	X_{i1}	X_{i2}
i	T_{i,J_i}	C_i	0	X_{i1}	X_{i2}

The ith subject contributes $(J_i + 1)$ records, where J_i is the number of observed events. For the jth record of the ith subject, **start** is the time of the $(j-1)$th event (or 0 if $j = 1$), **stop** is the time of the jth event (or censoring time if $j = J_i + 1$), and **status** indidates whether there is an event at the **stop** time.

```
coxph(Surv(start, stop, status) ~ covariates, ties = "breslow")
```

Table of Contents

Multiple Events Data

Setup: n subjects, K types of events

- T_{ki}: failure time of the kth event and ith subject
- C_{ki}: censoring time of the kth event and ith subject
- $Y_{ki} = \min(T_{ki}, C_{ki})$: observation time of the kth event and ith subject
- $\delta_{ki} = I(T_{ki} \leq C_{ki})$: failure indicator of the kth event and ith subject
- $X_{ki}(t)$: covariates of the kth event and ith subject

Observed data: $(Y_{ki}, \delta_{ki}, X_{ki})$, for $k = 1, \ldots, K$ and $i = 1, \ldots, n$

Independent censoring assumption: $T_{ki} \perp\!\!\!\perp C_{ki}$ given $X_{ki}(t)$

Notation:

- $R_{ki}(t) = I(Y_{ki} \geq t)$: at-risk indicator
- $S_k^{(r)}(t; \beta) = \sum_{i=1}^n R_{ki}(t) \exp\{\beta^{\mathsf{T}} X_{ki}(t)\} X_{ki}(t)^{\otimes r}$, $r = 0, 1, 2$

Marginal Cox model

We specify a Cox model for each type of event:

$$\lambda_k(t; X_{ki}) = \lambda_{0k}(t) \exp\left\{\beta_k^{\mathsf{T}} X_{ki}(t)\right\}, \qquad \text{for } k = 1, \ldots, K \text{ and } i = 1, \ldots, n$$

- $\lambda_{0k}(t)$: event-specific baseline hazard functions
- β_k: event-specific regression parameters
- modeling the marginal distributions for each event without specifying the dependence structures

Maximum Partial Likelihood Estimation

For the kth event, we obtain $\widehat{\beta}_k$ via maximum partial likelihood estimation.

$$\widehat{\beta}_k = \arg\max_{\beta} L_k(\beta)$$

$$L_k(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta^{\mathsf{T}} X_{ki}(Y_{ki})\}}{S_k^{(0)}(Y_{ki};\beta)} \right\}^{\delta_{ki}}$$

$$\ell_k(\beta) = \sum_{i=1}^{n} \delta_{ki} \left\{ \beta^{\mathsf{T}} X_{ki}(Y_{ki}) - \log S_k^{(0)}(Y_{ki};\beta) \right\}$$

$$U_k(\beta) = \sum_{i=1}^{n} \delta_{ki} \left\{ X_{ki}(Y_{ki}) - \frac{S_k^{(1)}(Y_{ki};\beta)}{S_k^{(0)}(Y_{ki};\beta)} \right\}$$

$$\mathcal{I}_k(\beta) = \sum_{i=1}^{n} \delta_{ki} \left\{ \frac{S_k^{(2)}(Y_{ki};\beta)}{S_k^{(0)}(Y_{ki};\beta)} - \frac{S_k^{(1)}(Y_{ki};\beta)^{\otimes 2}}{S_k^{(0)}(Y_{ki};\beta)^2} \right\}$$

Asymptotic Properties

$$\begin{bmatrix} \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_K \end{bmatrix} \dot{\sim} N\left(\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \begin{bmatrix} D_{11} & \dots & D_{1K} \\ \vdots & \vdots & \vdots \\ D_{K1} & \dots & D_{KK} \end{bmatrix} \right),$$

where

$$D_{kl} = \mathcal{I}_k^{-1}(\widehat{\beta}_k) B_{kl} \mathcal{I}_l^{-1}(\widehat{\beta}_l),$$

$$B_{kl} = \sum_{i=1}^{n} W_{ki} W_{li}^{\mathsf{T}},$$

$$W_{ki} = \delta_{ki} \left\{ X_{ki}(Y_{ki}) - \frac{S_k^{(1)}(Y_{ki}; \widehat{\beta}_k)}{S_k^{(0)}(Y_{ki}; \widehat{\beta}_k)} \right\}$$

$$- \sum_{i'=1}^{n} \frac{\delta_{ki'} R_{ki}(Y_{ki'}) \exp\{\widehat{\beta}_k^{\mathsf{T}} X_{ki}(Y_{ki'})\}}{S_k^{(0)}(Y_{ki'}; \widehat{\beta}_k)} \left\{ X_{ki}(Y_{ki'}) - \frac{S_k^{(1)}(Y_{ki'}; \widehat{\beta}_k)}{S_k^{(0)}(Y_{ki'}; \widehat{\beta}_k)} \right\}$$

Remarks

- Robust sandwich covariance matrix estimators D_{kl} $(k \neq l)$ account for the dependence of the multiple failure times.

- For the kth event, D_{kk} is the robust variance estimator, while $\mathcal{I}_k^{-1}(\widehat{\beta}_k)$ is the model-based variance estimator.

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: $\widehat{\Psi} = \widehat{\text{Cov}}(\widehat{\eta}_1, \ldots, \widehat{\eta}_K)$
- Global (Wald) test $H_0 : \eta_1 = \cdots = \eta_K = 0$

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

- To estimate a common parameter $\eta_1 = \cdots = \eta_K = \eta$, we let

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

  where the weights c_k are chosen to minimize the variance of $\widehat{\eta}$.
- Define $e = (1, \ldots, 1)^{\mathsf{T}}$. It can be shown that

$$(c_1, \ldots, c_K)^{\mathsf{T}} = (e^{\mathsf{T}}\widehat{\Psi}^{-1}e)^{-1}\widehat{\Psi}^{-1}e$$

Event-Specific Breslow Estimator

- The Breslow estimator for the kth event is

$$\widehat{\Lambda}_{0k}(t) = \sum_{i=1}^{n} \frac{I(Y_{ki} \leq t)\delta_{ki}}{S_k^{(0)}(Y_{ki}; \widehat{\beta}_k)}$$

- Asymptotic normality:

$$\left\{\widehat{\Lambda}_{01}(t), \ldots, \widehat{\Lambda}_{0K}(t)\right\} \dot{\sim} N_K\Big(\{\Lambda_{01}(t), \ldots, \Lambda_{0K}(t)\}, V(t)\Big),$$

where $V(t)$ is the robust variance-covariance estimator that accounts for the dependence of the multiple failure times.

Implementation

Data input for the ith subject:

sid	enum	time	status	X1	X2
i	1	Y_{1i}	δ_{1i}	X_{1i1}	X_{1i2}
i	2	Y_{2i}	δ_{2i}	X_{2i1}	X_{2i2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	K	Y_{Ki}	δ_{Ki}	X_{Ki1}	X_{Ki2}

Each subject has K records, one for each event type (identified by **enum**).

Implementation (Cont.)

R code:

```
coxph(Surv(time, status) ~ covariates * strata(enum) +
      cluster(sid) + strata(enum), ties = "breslow")
```

Note: additional calculation is needed since the output estimates for enum=2–K are $\widehat{\beta}_k - \widehat{\beta}_1$!

SAS code:

```
proc phreg covs(aggregate);
    model time*status(0)=covariates/cov ties=breslow;
    strata enum;
    id sid;
```

Note: covs(aggregate) is specified to compute the robust sandwich covariance matrix estimator.

Table of Contents

Clustered Data

Setup: n clusters, J_i subjects in the ith cluster

- T_{ij}: failure time of the jth subject in the ith cluster
- C_{ij}: censoring time of the jth subject in the ith cluster
- $Y_{ij} = \min(T_{ij}, C_{ij})$: observation time of the jth subject in the ith cluster
- $\delta_{ij} = I(T_{ij} \leq C_{ij})$: failure indicator of the jth subject in the ith cluster
- $X_{ij}(t)$: covariates of the jth subject in the ith cluster

Observed data: $(Y_{ij}, \delta_{ij}, X_{ij})$, for $i = 1, \ldots, n$ and $j = 1, \ldots, J_i$

Independent censoring assumption: $T_{ij} \perp\!\!\!\perp C_{ij}$ given $X_{ij}(t)$

Notation:

- $R_{ij}(t) = I(Y_{ij} \geq t)$: at-risk indicator
- $S^{(r)}(t; \beta) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} R_{ij}(t) \exp\{\beta^{\mathsf{T}} X_{ij}(t)\} X_{ij}(t)^{\otimes r}$, $r = 0, 1, 2$

Marginal Cox model

We consider the marginal Cox model:

$$\lambda(t; X_{ij}) = \lambda_0(t) \exp\left\{\beta^{\mathsf{T}} X_{ij}(t)\right\}, \qquad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, J_i$$

- $\lambda_0(t)$: arbitrary baseline hazard function
- β: unknown regression parameters
- modeling the marginal distributions without specifying the within-cluster dependence structures

To enable likelihood-based inference, we impose the independence working assumption that failure times within the same cluster are independent.

This unrealistic independence working assumption will be fixed later by the robust variance estimator.

Maximum Pesudo Partial Likelihood Estimation

Pesudo Partial Likelihood:

$$L_n(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{J_i} \left\{ \frac{\exp\{\beta^{\mathsf{T}} X_{ij}(Y_{ij})\}}{S^{(0)}(Y_{ij};\beta)} \right\}^{\delta_{ij}}$$

Compute $\widehat{\beta} = \arg\max_{\beta} L_n(\beta)$:

$$\ell_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \delta_{ij} \left\{ \beta^{\mathsf{T}} X_{ij}(Y_{ij}) - \log S^{(0)}(Y_{ij};\beta) \right\}$$

$$U_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \delta_{ij} \left\{ X_{ij}(Y_{ij}) - \frac{S^{(1)}(Y_{ij};\beta)}{S^{(0)}(Y_{ij};\beta)} \right\}$$

$$\mathcal{I}_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \delta_{ij} \left\{ \frac{S^{(2)}(Y_{ij};\beta)}{S^{(0)}(Y_{ij};\beta)} - \frac{S^{(1)}(Y_{ij};\beta)^{\otimes 2}}{S^{(0)}(Y_{ij};\beta)^2} \right\}$$

Asymptotic Properties

(i) $U_n(\beta) \dot{\sim} N(0, B(\beta))$, where $B(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \sum_{j'=1}^{J_i} W_{ij}(\beta) W_{ij'}(\beta)^{\mathsf{T}}$ and

$$W_{ij}(\beta) = \delta_{ij} \left\{ X_{ij}\left(Y_{ij}\right) - \frac{S^{(1)}\left(Y_{ij}; \beta\right)}{S^{(0)}\left(Y_{ij}; \beta\right)} \right\}$$

$$- \sum_{i'=1}^{n} \sum_{j'=1}^{J_{i'}} \frac{\delta_{i'j'} R_{ij}\left(Y_{i'j'}\right) \exp\{\beta^{\mathsf{T}} X_{ij}\left(Y_{i'j'}\right)\}}{S^{(0)}\left(Y_{i'j'}; \beta\right)} \left\{ X_{ij}\left(Y_{i'j'}\right) - \frac{S^{(1)}\left(Y_{i'j'}; \beta\right)}{S^{(0)}\left(Y_{i'j'}; \beta\right)} \right\}$$

(ii) $\widehat{\beta} \dot{\sim} N(\beta, D)$, where $D = \mathcal{I}_n^{-1}(\widehat{\beta}) B(\widehat{\beta}) \mathcal{I}_n^{-1}(\widehat{\beta})$.

Remarks

- Robust variance estimators $B(\beta)$ and D account for within-cluster dependence, while naive variance estimators $\mathcal{I}_n(\beta)$ and $\mathcal{I}_n^{-1}(\widehat{\beta})$ do not.
- If $J_i \equiv 1$, then D becomes the robust variance estimator for misspecified univariate Cox models.
- Score and Wald tests:
  - $H_0 : \beta = \beta^*$. $SC_n = U_n(\beta^*)^\mathsf{T} B(\beta^*)^{-1} U_n(\beta^*) \xrightarrow{d} \chi_p^2$.
  - $H_0 : C\beta = C\beta^*$. $W_n = (C\widehat{\beta} - C\beta^*)^\mathsf{T}(CDC^\mathsf{T})^{-1}(C\widehat{\beta} - C\beta^*) \xrightarrow{d} \chi_q^2$ for a $q \times p$ matrix C.

Estimation for Λ_0

- Breslow estimator:

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \frac{I(Y_{ij} \leq t)\delta_{ij}}{S^{(0)}(Y_{ij}; \widehat{\beta})}$$

- Asymptotic normality:

$$\widehat{\Lambda}_0(t) \dot{\sim} N\big(\Lambda_0(t), V(t)\big),$$

where $V(t)$ is the robust variance estimator that accounts for the within-cluster dependence.

Implementation in R

Data input for the ith cluster:

cid	**time**	**status**	**X1**	**X2**
i	Y_{i1}	δ_{i1}	X_{i11}	X_{i12}
i	Y_{i2}	δ_{i2}	X_{i21}	X_{i22}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	Y_{iJ_i}	δ_{iJ_i}	X_{iJ_i1}	X_{iJ_i2}

The ith cluster contributes J_i records, one for each subject within the cluster.

```
coxph(Surv(time, status) ~ covariates + cluster(cid),
      ties = "breslow")
```

Table of Contents

Frailty Models

- Another way to address the dependence of multivariate failure time data is through frailty models, also known as random effects models.
- This is simply by adding a frailty (or random effects) term to the original Cox-type models.
- For multiple events data, all events from the same subject share the same frailty. For clustered data, all subjects within the same cluster share the same frailty.
- This shared frailty accounts for the within-subject or within-cluster dependence.

Frailty Models

Recurrent events: for the ith subject ($i = 1, \ldots, n$), we specify the intensity function as

$$\lambda(t; X_i, b_i) = \lambda_0(t) \exp\{\beta^{\mathsf{T}} X_i(t) + b_i^{\mathsf{T}} \widetilde{X}_i(t)\}$$

Multiple events: for the kth event and ith subject ($k = 1, \ldots, K$ and $i = 1, \ldots, n$), we specify the hazard function as

$$\lambda(t; X_{ki}, b_i) = \lambda_{0k}(t) \exp\{\beta^{\mathsf{T}} X_{ki}(t) + b_i^{\mathsf{T}} \widetilde{X}_{ki}(t)\}$$

Clustered data: for the jth subject in the ith cluster ($i = 1, \ldots, n$ and $j = 1, \ldots, J_i$), we specify the hazard function as

$$\lambda(t; X_{ij}, b_i) = \lambda_0(t) \exp\{\beta^{\mathsf{T}} X_{ij}(t) + b_i^{\mathsf{T}} \widetilde{X}_{ij}(t)\}$$

Remarks

- The frailty $b_i \overset{iid}{\sim} f(b;\gamma)$, which accounts for the within-subject or within-cluster dependence.
- The set of covariates $\widetilde{X}$ contains 1 and part of X.
- Under frailty models, the regression coefficient β is no longer the population average effect of the covariates as under marginal models.
- Instead, β should be interpreted as the subject-specific effect which can vary from person to person.
- A special case of frailty models is the Cox model with frailty:

$$\lambda(t; X_i, Z_i) = Z_i \lambda_0(t) \exp\left\{\beta^{\mathsf{T}} X_i(t)\right\}, \qquad \text{for } Z_i > 0,$$

  where a larger Z_i value indicates a higher risk of failure (i.e., the ith subject is more frail than others).

Other Advantages of Frailty Models

- The frailty b_i also addresses unobservable heterogeneity. For example, immunocompromised subjects relapse into COVID-19 infection more rapidly than others, but this cannot be fully described by the covariates.
- Thus, prediction based on frailty models is usually more accurate than that based on marginal models.
- Frailty models also enable dynamic prediction based on the evolving disease history, by dynamically updating the distribution of the frailty b_i given the observed data.
- Finally, we can adjust for dependent censoring by jointly modelling the failure and censoring times with shared frailty.

Maximum Likelihood Estimation

The likelihood function for clustered data is

$$\prod_{i=1}^{n} \int_{b_i} \prod_{j=1}^{J_i} \left[\lambda_0(Y_{ij}) \exp\left\{ \beta^{\mathrm{T}} X_{ij}(Y_{ij}) + b_i^{\mathrm{T}} \widetilde{X}_{ij}(Y_{ij}) \right\} \right]^{\delta_{ij}}$$

$$\times \exp\left[- \int_0^{Y_{ij}} \exp\left\{ \beta^{\mathrm{T}} X_{ij}(t) + b_i^{\mathrm{T}} \widetilde{X}_{ij}(t) \right\} d\Lambda_0(t) \right] \times f(b_i;\gamma) db_i$$

- Apply the nonparametric maximum likelihood estimation (NPMLE) approach and treat $\Lambda_0(t)$ as a step function with nonnegative jumps at observed event times.
- Replace $\lambda_0(t)$ by the jump size of Λ_0 at t.
- Maximize the discretized likelihood function over β, γ and jump sizes of Λ_0 using standard MLE procedures.
- Likelihoods for multiple events and recurrent events are similar.

Computation

- Compared to marginal models, the computation for frailty models is much more challenging due to the integrals of frailty.

- One can use numerical methods to approximate the integrals of frailty. For example, when b_i is normally distributed, Gauss-Hermite quadratures can be used.

- More accurate computation is achieved through EM algorithm, where frailty b_i $(i = 1, \ldots, n)$ are treated as missing data.
  - E-step: compute the conditional expectation of the complete-data log-likelihood given the observed data and the current parameter estimates.
  - M-step: update the parameter estimates by maximizing the conditional expectation of the complete-data log-likelihood obtained in the E-step.
  - Iterate between E-step and M-step until convergence.

Asymptotic Properties

Let $\theta^{\mathsf{T}} = (\beta^{\mathsf{T}}, \gamma)$ be the vector of all finite-dimensional parameters. Denote the nonparametric MLEs by $\widehat{\theta}$ and $\widehat{\Lambda}_0(t)$.

(i) Consistency: $\|\widehat{\theta} - \theta\| + \sup_{t \in [0,\tau]} |\widehat{\Lambda}_0(t) - \Lambda_0(t)| \xrightarrow{p} 0$.

(ii) Asymptotic normality: $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, V)$.

(iii) Semiparametric efficiency: the covariance matrix V attains the semiparametric efficiency bound.

Implementation in R

Recurrent events:

```
coxph(Surv(start, stop, status) ~ covariates +
      frailty(sid, dist = "gaussian"), ties = "breslow")
```

Multiple events:

```
coxph(Surv(time, status) ~ covariates + strata(enum) +
      frailty(sid, dist = "gaussian"), ties = "breslow")
```

Clustered data:

```
coxph(Surv(time, status) ~ covariates +
      frailty(cid, dist = "gaussian"), ties = "breslow")
```