What to do today (2022/03/24)?

Part IV. Advanced Topics

- *Part IV.1 Counting Process Formulation* (Revisits to KM estm, Logrank test, and Cox PH model)
  - *IV.1.1 Theoretical Preparation*
  - *IV.1.2 Counting Process Formulation in LIDA and Applications: Revisits to KM, Logrank, Cox PH*

- **Part IV.2 Selected Recent Topics in LIDA**
  - **IV.2.1 Alternatives to Cox PH model**
  - **IV.2.2 Multivariate event times**
  - *IV.2.3 More incomplete data structures*
  - *IV.2.4 Missing covariates in regression*

- *Part IV.3 Beyond Lifetime Data Analysis**

Part IV.2.1 Alternatives to Cox PH (regression) model

Recall Cox PH: event time T's conditional hazard function with covariate $Z = z$ is

$$h(t|z) = h_0(t)e^{\beta z}$$

- *interpretation*
- *partial likelihood approach*
- *robustness*

Part IV.2.1 Alternatives to Cox PH (regression) model

Recall Cox PH: event time T's conditional hazard function with covariate $Z = z$ is

$$h(t|z) = h_0(t)e^{\beta z}$$

rather restrictive ... ...

Extensions:

- $h(t|z) = h_0(t)e^{\beta z(t)}$; $h(t|z) = h_0(t)e^{\beta(t) z}$
- $h(t|z, s) = h_{s0}(t)e^{\beta_s z}$
- $h(t|z) = h_0(t)g(\beta; z)$

Any other alternatives to Cox PH model?

Part IV.2.1A Alternatives to Cox PH (regression) model: AFT

Accelerated Failure Time Model (AFT) Assume event time T follows

$$T = T_0 e^{-\beta Z},$$

where T_0 is another event time and Z is the covarate.

- *direct physical interpretation* (time-scale change)
- AFT $\Longleftrightarrow$ the regression model

  $$\log T = -\beta Z + \epsilon$$

  with $\epsilon = \log T_0$, and thus $T|Z = z \sim S(t|z)$: if $T_0 \sim S_0(\cdot)$, $S(t|z) = S_0(te^{\beta z})$.
  - If $\epsilon \sim N(0, \sigma^2)$, $T \sim$ *lognormal*
  - If ϵ's distn unspecified, a semi-parametric model.

Part IV.2.1A Alternatives to Cox PH (regression) model: AFT

Statistical Inference under AFT with right-censored data

- When $\{(T_i, Z_i) : i = 1, \ldots, n\}$ available:
- When $\{(U_i, \delta_i, Z_i) : i = 1, \ldots, n\}$ available:
  $e_i(\beta) = \min(\log T_i + \beta Z_i, \log C_i + \beta Z_i)$
  - *Buckley and James (1979)* with $\epsilon_i \sim S^*(\cdot)$

$$E\Big(\log T_i | U_i, \delta, Z_i \Big) = \log U_i + (1 - \delta_i) \frac{\int_{e_i}^{\infty} S^*(s) ds}{S^*(e_i)}$$

  - *partial-score type estimating function*

$$U^*(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \Big(Z_i - \frac{\sum_j I(e_j(\beta) \geq u) Z_j}{\sum_j I(e_j(\beta) \geq u)} \Big) d\tilde{N}_i(u)$$

with $\tilde{N}_i(t) = I(e_i(\beta) \leq u, \delta_i = 1)$.

Part IV.2.1B Alternatives to Cox PH (regression) model: Chen and Jewell (2001)

Consider $h(t|z) = h_0(te^{\beta_1 z})e^{\beta_2 z}$

- $\beta_1 = 0 \Longrightarrow$ Cox PH
- $\beta_1 = \beta_2 \Longrightarrow$ AFT

Statistical Inference under Chen-Jewell Model?

- likelihood-based
- estimating equation

Part IV.2.1C Alternatives to Cox PH (regression) model: linear transformation models (LTM) (Cheng, Wei and Ying, 1996)

Consider $g\Big(S(t|z)\Big) = g\Big(S_0(t)\Big) + \beta z$

- $g(u) = \log(-\log(u)) \Longrightarrow$ Cox PH
- $g(u) = \log(S_0^{-1}(u)) \Longrightarrow$ AFT
- $g(u) = logit(u) \Longrightarrow$ the log-odds model:

$$\log\left(\frac{S(t|z)}{1 - S(t|z)}\right) = logit(S_0(t)) + \beta z$$

Statistical Inference under LTM?

Part IV.2.2A Multivariate event times

Examples for multiple events (associated with an individual)

- Automobile repairs/failures
- Times to AIDS events

Part IV.2.2A Multivariate event times

Approaches to analyzing the data

- Focus on the time to an event, say, the 1st event
- Consider all the times to events $(T_1, \ldots, T_J)$ jointly
  - if $T_1, \ldots, T_J \perp\!\!\!\perp$
  - if $T_1, \ldots, T_J \not\perp\!\!\!\perp$

Part IV.2.2B Multivariate event times: nonparametric estimation (e.g. Prentice et al, 2004)

Consider $\underline{T} = (T_1, T_2)'$:

$$S(\underline{t}) = P(\underline{T} >> \underline{t}) = P(T_1 \geq t_1, T_2 \geq t_2)$$

with marginal survivor functions
$S_1(t) = P(T_1 \geq t_1), S_2(t) = P(T_2 \geq t_2)$
$d\Lambda(\underline{t}) = dS(\underline{t})/S(\underline{t}-)$ and

$$S(\underline{t}) = [S_1(t) + S_2(t) - 1] + \int_0^{\underline{t}} S(\underline{u}-)\Lambda(d\underline{s})$$

Data: $\underline{U}_i = (U_{1i}, U_{2i})', \delta_i = (\delta_{1i}, \delta_{2i})'$ for $i = 1, \ldots, n$

Part IV.2.2C Multivariate event times: marginal approach (regression) (e.g. Wei et al, 1989)

Consider $i = 1, \ldots, n \perp\!\!\!\perp$ subjects with event times $T_{ki} \sim \lambda_{k0}(t)e^{\beta_k Z_{ki}}$ for $k = 1, \ldots, K$.

Suppose the observations are subject to censoring: the available data are

$$\bigcup_{i=1}^{n} \left\{(U_{ki}, \delta_{ki}, Z_{ki}) : k = 1, \ldots, K\right\}$$

Maximize $L_k(\beta_k)$ to obtain $\hat{\beta}_k$:

$$L_k(\beta_k) = \prod_{i=1}^{n} \left(\frac{e^{\beta_k Z_{ki}}}{\sum_{l \in \mathcal{R}_k(u_{ki})} e^{\beta_k Z_{kl}}} \right)^{\delta_{ki}}$$

Properties of $(\hat{\beta}_1, \ldots, \hat{\beta}_K)$ as estimator for $(\beta_1, \ldots, \beta_k)$?

Part IV.2.2C Multivariate event times: marginal approach (regression) (e.g. Wei et al, 1989)

In many situations, $\hat{\beta}_k$'s are the solutions to

$$U_k(\beta_k) = \sum_{i=1}^{n} \int_0^\infty \left(Z_{ki} - \frac{\sum_l Y_{kl}(u) Z_{kl} e^{\beta_k Z_{kl}}}{\sum_l Y_{kl}(u) e^{\beta_k Z_{kl}}} \right) dN_{ki}(u) = 0$$

with $N_{ki}(t) = I(U_{ki} \le t, \delta_{ki} = 1)$ and $Y_{ki}(t) = I(U_{ki} \ge t)$

The (asymptotic) joint distribution of $(\hat{\beta}_1, \ldots, \hat{\beta}_K)$ can be derived based on the (asymptotic) distribution of $(U_1(\beta_1), \ldots, U_K(\beta_K))$.

Part IV.2.2D Multivariate event times: Poisson process analysis (e.g. Andersen and Gill, 1982)

Consider $N(t) = \#$ of events till t together with covariate Z and $Y(t) = I(C \geq t)$.

Assume the conditional intensity of $N(\cdot)|Z$ is

$$\lim_{\Delta t \to 0} P(N(t) - N(t - \Delta t) > 0 | \mathcal{H}_{t-}) \big/ \Delta t = \lambda(t|Z) = \lambda_0(t) e^{\beta Z}$$

The likelihood function based on the right-censored Poisson process data $\left\{(\{N_i(t) : 0 \leq t \leq C_i\}, Z_i) : i = 1, \ldots, n\right\}$ is

$$L(\beta, \lambda_0(\cdot)) = \prod_{i=1}^{n} L_i(\beta, \lambda_0(\cdot)) = \prod_{i=1}^{n} \Big(\prod_{k=1}^{N_i(C_i)} \lambda_0(t_{ki}) e^{\beta Z_i} \Big) \exp\{- \int_0^{C_i} e^{\beta Z_i} d\Lambda_0(u)\}$$

Part IV.2.2D Multivariate event times: Poisson process analysis (e.g. Andersen and Gill, 1982)

Solve the equation to obtain $\hat{\beta}$:

$$U(\beta) = \sum_{i=1}^{n} \int_0^{C_i} \left(Z_i - \frac{\sum_l Y_l(u) Z_l e^{\beta Z_l}}{\sum_l Y_l(u) e^{\beta Z_l}} \right) dN_i(u) = 0$$

Plug in $\hat{\beta}$ to obtain an estimator for $\Lambda_0(t)$:

$$\hat{\Lambda}(t; \beta) = \sum_{i=1}^{n} \int_0^{t} \frac{Y_i(u) dN_i(u)}{\sum_l Y_l(u) e^{\beta Z_l}}$$

Part IV.2.2D Multivariate event times: Poisson process analysis (e.g. Andersen and Gill, 1982)

Remarks:

- What if the Poisson assumption is not plausible?
  - Prentice et al (1981): if stratification variable $s = \mathcal{S}(\mathcal{N}(t-), \mathcal{Z}(t-))$,

    $$\lim_{\Delta t \to 0} P(N(t) - N(t - \Delta t) > 0 | \mathcal{H}_{t-})/\Delta t = \lambda_{0s}(t) e^{\beta_s Z}$$

    or

    $$\lim_{\Delta t \to 0} P(N(t) - N(t - \Delta t) > 0 | \mathcal{H}_{t-})/\Delta t = \lambda_{0s}(t - t_{N(t-)}) e^{\beta_s Z}$$

    renewal process; semi-Markov process
  - Lawless and Nadeau (1995, Technometrics)

- What if the covariates are time-dept: $Z_i(t)$?

- Other variations?
  e.g. Hu et al (2010)

What to study next?

Part IV. Advanced Topics

- *Part IV.1 Counting Process Formulation* (Revisits to KM estm, Logrank test, and Cox PH model)

- **Part IV.2 Selected Recent Topics in LIDA**
  - *Part IV.2.1 Alternatives to Cox PH model*
  - *Part IV.2.2 Multivariate event times*
  - **Part IV.2.3 More incomplete data structures in LIDA**
  - **Part IV.2.4 Analysis of incomplete data**

- *Part IV.3 Beyond Lifetime Data Analysis**