What to do today (2022/03/17)?

Part IV. Advanced Topics

- **Part IV.1 Counting Process Formulation** (Revisits to KM estm, Logrank test, and Cox PH model)
  - *Part IV.1.1 Theoretical Preparation*
  - **Part IV.1.2 Counting Process Formulation in LIDA and Applications: Revisits to KM, Logrank, Cox PH**
    - *Part IV.1.2A Formulation*
    - *Part IV.1.2B Revisit to Logrank Test*
    - **Part IV.1.2C Revisit to Cox's Partial Likelihood Approach**
    - **Part IV.1.2D Revisit to Nelson-Aalen Estimator and Breslow Estimator**
    - *Part IV.1.2E Revisit to Kaplan-Meier Estimator*
    - *Part IV.1.2F Others*
- *Part IV.2 Selected Recent Topics in LIDA*
- *Part IV.3 Beyond Lifetime Data Analysis*

Review: Part IV.1.2A Counting Process Formulation in LIDA and Applications: Revisits to KM, Logrank, Cox PH

Consider T, a lifetime with hazard function $h(\cdot)$. Let C be a censoring time, $U = T \wedge C$ and $\delta = 1, 0$ for $U = T, C$, respectively.
Define $N(t) = I(U \leq t, \delta = 1)$ and $Y(t) = I(U \geq t)$. If $T \perp\!\!\!\perp C$ and $T \sim f(\cdot)$, $C \sim g(\cdot)$,

- $E\big[N(t)\big] = E\big[A(t)\big]$ with $A(t) = \int_0^t h(x)Y(x)dx$.
- $A(\cdot)$ is the continuous compensator of $N(\cdot)$ wrt $\{\mathcal{H}_t\}$, and denote $M(\cdot) = N(\cdot) - A(\cdot)$.

Recall *right-censored* lifetimes $\{(U_i, \delta_i) : i = 1, \ldots, n\}$ from n indpt individuals.
Define $N_i(t) = \int_0^t dN_i(v) = I(U_i \leq t, \delta_i = 1)$ and $Y_i(t) = I(U_i \geq t)$.

Part IV.1.2B Revisit to Logrank Test

Logrank Test.

- Consider T_1, T_0 event times in the treatment, control group, with hazard functions $h_1(\cdot), h_0(\cdot)$. Assume $T_1 \perp\!\!\!\perp T_0$ and indepdent censoring time C.
- $H_0 : h_1(\cdot) = h_0(\cdot)$ vs H_1 : *otherwise*.
- Data are $\{(U_{1i}, \delta_{1i}) : i = 1, \ldots, n_1\} \bigcup \{U_{0j}, \delta_{0j}) : j = 1, \ldots, n_0\}$
- Test statistic
$$Z = \frac{O - E}{\sqrt{V}} \to N(0, 1) \text{ in distn}$$

under H_0 as $n \to \infty$ with $n = n_1 + n_0$ and $n_1/n \to p \in (0, 1)$.

to verify the asymptotic normality?

Introducing $N_{1i}(t) = I(U_{1i} \leq t, \delta_{1i} = 1)$ and $Y_{1i}(t) = I(U_{1i} \geq t)$, $N_{0j}(t) = I(U_{0j} \leq t, \delta_{0j} = 1)$ and $Y_{0j}(t) = I(U_{0j} \geq t)$, and $N_{1.}(t), Y_{1.}(t), N_{0.}(t), Y_{0.}(t)$.
With $0 < V_1 < \ldots < V_K < \infty$ all the distinct observed failure times,

$$O = N_{1.}(\infty) = \sum_{i=1}^{n_1} \int_0^\infty dN_{1i}(t) = \sum_k d_{1k}, \quad E = \sum_{k=1}^{K} E_k$$

$E_k = d_k \frac{N_{k1}}{N_k} = \int_{V_{k-1}}^{V_k} \frac{Y_{1.}(v)}{Y_{1.}(v)+Y_{0.}(v)} d[N_{1.}(v) + N_{0.}(v)]$.

$$O - E = \int_0^\infty [1 - H_n(t)] dN_{1.}(t) - \int_0^\infty H_n(t) dN_{0.}(t)$$

Further, introducing $M_{li}(t) = N_{li}(t) - A_{li}(t)$ and $A_{li} = \int_0^t h_l(v) Y_{li}(v) dv$ for $i = 1, \ldots, n_l$, $l = 1, 0$. Under H_0, $h_1(\cdot) = h_0(\cdot)$, and let $U(\infty)$ be

$$O - E = \sum_i \int_0^\infty [1 - H_n(t)] dM_{1i}(t) - \sum_j \int_0^\infty H_n(t) dM_{0j}(t).$$

By the Martingale CLT, $U_n(t) \to Gaussian\big(0, \alpha(t)\big)$ in distn as $n \to \infty$. Thus $\frac{1}{\sqrt{n\alpha(\infty)}} U(\infty) \to N(0,1)$ in distn as $n \to \infty$.

Example IV.1 (cont'd) $n = 5$ indpt subjects and $Z = \begin{cases} 1 & \textit{treatment} \\ 0 & \textit{placebo} \end{cases}$

(u_i, δ_i, z_i) : $(16, 1, 1)$, $(13, 0, 0)$, $(21, 1, 1)$, $(11, 1, 0)$, $(12, 1, 1)$
V_l : 11, 12, 13, 16, 21; $Z_{(l)}$: 0, 1, 0, 1, 1;
$\mathcal{R}_l$: $\{1, 2, 3, 4, 5\}, \{1, 2, 3, 5\}, \{1, 2, 3\}, \{1, 3\}, \{3\}$,

Introducing $N_{1i}(t) = I(U_{1i} \leq t, \delta_{1i} = 1)$ and $Y_{1i}(t) = I(U_{1i} \geq t)$,
$N_{0j}(t) = I(U_{0j} \leq t, \delta_{0j} = 1)$ and $Y_{0j}(t) = I(U_{0j} \geq t)$, and
$N_{1.}(t), Y_{1.}(t), N_{0.}(t), Y_{0.}(t)$.

Part IV.1.2B Revisit to Logrank Test

Remarks

- Nelson-Aalen Estimator for the cumulative hazards $H_l(\cdot)$:

$$\hat{H}_1(t) = \int_0^t \frac{1}{Y_{1.}(v)} dN_{1.}(v), \quad \hat{H}_0(t) = \int_0^t \frac{1}{Y_{0.}(v)} dN_{0.}(v)$$

(Nelson, 1969; Aalen, 1972)

$\Longrightarrow \hat{S}_l(t) = \exp\{-\hat{H}_l(t)\}$,

an estimator for the survivor function $S_l(\cdot)$: not the same as KM estimator but asymptotically equivalent, with higher finite sample efficiency.

Part IV.1.2B Revisit to Logrank Test

- A Class of Rank Tests: $H_0 : H_0(\cdot) = H_1(\cdot)$

$$\int_0^\infty W_n(t)[d\hat{H}_1(t) - d\hat{H}_0(t)],$$

with $W_n(\cdot)$ predictable, non-negative and order of n.

  - e.g. $W_n(t) = \frac{Y_{0.}(t)Y_{1.}(t)}{Y_{1.}(t)+Y_{0.}(t)} \Longrightarrow$ *Logrank Test*
  - e.g. $W_n(t) = \frac{Y_{0.}(t)Y_{1.}(t)}{n} \Longrightarrow$ *Gehan-Wilcoxon Test*
  - e.g. $W_n(t) = \hat{S}^-(t)\frac{Y_{0.}(t)Y_{1.}(t)}{Y_{1.}(t)+Y_{0.}(t)} \Longrightarrow$ *Prentice-Wicoxon Test* ($\hat{S}^-(t)$ the left-cont's version of SM estm)
  - e.g. $W_n(t) = [\hat{S}^-(t)]^\alpha \frac{Y_{0.}(t)Y_{1.}(t)}{Y_{1.}(t)+Y_{0.}(t)} \Longrightarrow$ *Fleming-Hurrington Test*

Part IV.1.2B Revisit to Logrank Test

- A Class of Rank Tests: $H_0 : H_0(\cdot) = H_1(\cdot)$

$$\int_0^\infty W_n^*(t) \frac{Y_{0.}(t) Y_{1.}(t)}{Y_{1.}(t) + Y_{0.}(t)} [d\hat{H}_1(t) - d\hat{H}_0(t)],$$

  with $W_n^*(\cdot)$ predictable, bounded, and non-negative. *How to choose $W_n^*(\cdot)$ to achieve a powerful test?*

  - If $H_1 : \log\left[h_1(t)/h_0(t)\right] = \alpha g(t)$, optimal weight?

e.g. Lagakos and Schoenfeld (1984).

Part IV.1.2C Revisit to Cox's Partial Likelihood Approach

Consider T following the Cox's proportional harzards model:

$$T|Z \sim h(t|Z) = h_0(t)e^{\beta Z}$$

Let C be an independent censoring time, $U = T \wedge C$ and $\delta = 1, 0$ for $U = T, C$, respectively.
Define $N(t) = I(U \leq t, \delta = 1)$, $Y(t) = I(U \geq t)$ and $A(t|Z) = \int_0^t Y(u)h(u|Z)du = \int_0^t Y(u)e^{\beta Z}dH_0(u)$.

Suppose the available data are

$$\{(U_i, \delta_i, Z_i) : i = 1, \ldots, n\}$$

from n indpt subjects and indpt censoring $T_i \perp\!\!\!\perp C_i$:

$$L(\beta, h_0(\cdot)\big|data) = \prod_{i=1}^{n} \left(h_0(u_i)e^{\beta z_i}\right)^{\delta_i} \exp(-H_0(u_i)e^{\beta z_i}) = L_p(\beta\big|data)L_1(\beta, h_0(\cdot)\big|data)$$

Part IV.1.2C Revisit to Cox's Partial Likelihood Approach

Cox partial likelihood function (Cox, Biometrika 1975)

$$L_p(\beta \big| data) = \prod_{i=1}^{n} \Big(\frac{e^{\beta z_i}}{\sum_{l \in \mathcal{R}_i} e^{\beta z_l}} \Big)^{\delta_i} = \prod_{i=1}^{n} \Big(\frac{e^{\beta z_i}}{\sum_{l=1}^{n} Y_l(u_i) e^{\beta z_l}} \Big)^{\delta_i}$$

the risk set at time u_i: $\mathcal{R}_i = \{j : u_j \leq u_i\}$
$\Longrightarrow$ the MPLE (maximum partial likelihood estimator) of β:

$$\hat{\beta} = argmax_{all\ \beta} L_p(\beta \big| data)$$

the solution of $U(\beta) = \frac{\partial \log L_p(\beta)}{\partial \beta} = 0$, where

$$\sum_{i=1}^{n} \delta_i \{ z_i - \frac{\sum_{l=1}^{n} Y_l(u_i) z_l e^{\beta z_l}}{\sum_{l=1}^{n} Y_l(u_i) e^{\beta z_l}} \} = \sum_{i=1}^{n} \int_0^{\infty} \{ z_i - \frac{\sum_{l=1}^{n} Y_l(u) z_l e^{\beta z_l}}{\sum_{l=1}^{n} Y_l(u) e^{\beta z_l}} \} dN_i(u)$$

Verify $\hat{\beta} \to \beta$ a.s. and $\sqrt{n}(\hat{\beta} - \beta) \to N(0, *)$ in distn?

Part IV.1.2C Revisit to Cox's Partial Likelihood Approach

Firstly, to check if $\frac{1}{\sqrt{n}} U(\beta) \to N(0, \alpha(\infty))$ *in distn as* $n \to \infty$

Note that

$$
\begin{aligned}
U_n(\beta; t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \{z_i - \frac{\sum_{l=1}^n Y_l(u) z_l e^{\beta z_l}}{\sum_{l=1}^n Y_l(u) e^{\beta z_l}}\} dN_i(u) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \{z_i - \frac{\sum_{l=1}^n Y_l(u) z_l e^{\beta z_l}}{\sum_{l=1}^n Y_l(u) e^{\beta z_l}}\} dM_i(u).
\end{aligned}
$$

It converges to the Gaussian process with mean zero and variance function $\alpha(t)$ by martingale CLT as $n \to \infty$, where

$$
< U_n, U_n > (\beta; t) = \sum_{i=1}^{n} \int_0^t H_{i,n}^2(u) h_0(u) e^{\beta z_i} Y_i(u) du \to \alpha(t).
$$

Part IV.1.2C Revisit to Cox's Partial Likelihood Approach

Secondly, to check if $\sqrt{n}(\hat{\beta}-\beta) \rightarrow N(0, i(\beta)^{-1})$ *in distn*

Note that

$$\begin{aligned} \frac{1}{\sqrt{n}}[U(\hat{\beta}) - U(\beta)] = -\frac{1}{\sqrt{n}}U(\beta) \\ = \quad \frac{1}{n}\frac{\partial U(\beta)}{\partial \beta}\sqrt{n}(\hat{\beta}-\beta) + o(\hat{\beta}-\beta)\frac{1}{\sqrt{n}}. \end{aligned}$$

Plus $-\frac{1}{n}\frac{\partial U(\beta)}{\partial \beta} \rightarrow \alpha(\infty) = i(\beta)$.

Use $\left(\hat{I}_p(\beta)\right)^{-1} = \left(-\frac{\partial U(\beta)}{\partial \beta}\right)^{-1}$ to estm $Var(\hat{\beta})$.

Part IV.1.2D Revisit to Nelson-Aalen Estimator and Breslow Estimator

Recall the notation:

$$N_\cdot(t),\ Y_\cdot(t),\ A_\cdot(t),\ M_\cdot(t) = N_\cdot(t) - A_\cdot(t)$$

Thus $dM_\cdot(t) = dN_\cdot(t) - Y_\cdot(t)dH(t)$.

Assign $dM_\cdot(t) = 0 \Longrightarrow d\hat{H}(t) = \frac{dN_\cdot(t)}{Y_\cdot(t)}$:

Nelson-Aalen Estimator for the cumulative hazard function

$$\hat{H}(t) = \int_0^t \frac{1}{Y_\cdot(u)} dN_\cdot(u)$$

Part IV.1.2D Revisit to Nelson-Aalen Estimator and Breslow Estimator

Remarks:

- The observed failure times $0 \leq V_1 < \ldots < V_J$,

$$\hat{H}(t) = \sum_{V_j \leq t} \frac{\#\ failures\ at\ V_j}{\#at\ risk\ at\ V_j} = \sum_{V_j \leq t} \frac{d_j}{N_j}.$$

- Define $0/0 = 0$ or replace $1/Y_.(u) = I(Y_.(u) > 0)/Y_.(u)$.

- $E\left[dN_i(t) | Y_i(t)\right] = Y_i(t)h(t)dt$

Nelson-Aalen Estimator's Asymptotics

$\left[\hat{H}(t) - H(t)\right]$ is

$$
\begin{aligned}
& \int_0^t \frac{I(Y_\cdot(u) > 0)}{Y_\cdot(u)} dN_\cdot(u) - \int_0^t \frac{I(Y_\cdot(u) > 0) Y_\cdot(u)}{Y_\cdot(u)} dH(u) - \int_0^t I(Y_\cdot(u) = 0) dH(u) \\
= \quad & \int_0^t \frac{I(Y_\cdot(u) > 0)}{Y_\cdot(u)} dM_\cdot(u) - \int_0^t I(Y_\cdot(u) = 0) dH(u)
\end{aligned}
$$

By Martingale CLT, in distn

$$
\sqrt{n} \int_0^t \frac{I(Y_\cdot(u) > 0)}{Y_\cdot(u)} dM_\cdot(u) \to Gaussian(0, \sigma^2(t))
$$

and $\sqrt{n} \int_0^t I(Y_\cdot(u) = 0) dH(u) \to 0$ in Pr 1.

Thus $\sqrt{n}\left[\hat{H}(t) - H(t)\right] \to Gaussian(0, \sigma^2(t))$.
Plus $\sup_{t>0} |\hat{H}(t) - H(t)| \to 0$ in Pr 1.

Remarks:

- $\sigma^2(t) = \int_0^t \frac{1}{P(U \geq u)} dH(u)$ and $\hat{\sigma}^2(t) = \int_0^t \frac{nI(Y_.(u)>0)}{Y_.(u)} d\hat{H}_{NA}(u)$ with $n >> 1$.
- **Fleming-Harrington Estimator** for survivor function $\hat{S}(t) = \exp\{-\hat{H}(t)\}$:

$$\sqrt{n}\big[\hat{S}(t) - S(t)\big] \approx S(t)\sqrt{n}\big[\hat{H}(t) - H(t)\big]$$

- **Breslow Estimator** for the baseline hazard function in the Cox PH model:

$$\hat{H}_0(t;\beta) - H_0(t) = \int_0^t \frac{1}{\sum_{l=1}^n Y_l(t) e^{\beta z_l}} dN_.(u)$$

  with β replaced by the PMLE $\hat{\beta}$
  - Note that $\hat{H}_0(t;\hat{\beta}) = \big[\hat{H}_0(t;\hat{\beta}) - \hat{H}_0(t;\beta)\big] + \big[\hat{H}_0(t;\beta) - H_0(t)\big]$

Part IV.1.2D Revisit to Nelson-Aalen Estimator and Breslow Estimator: Confidence Band of $H(\cdot)$

Recall that $\hat{H}_{NA}(t) = \int_0^t \frac{I(Y_\cdot(u))}{Y_\cdot(u)} dN_\cdot(u) \sim Gaussian(H(t), \frac{\sigma^2(t)}{n})$ approximately.

Pointwise 95% CI: $\forall t > 0$, $\hat{H}_{NA}(t) \pm 1.96\sqrt{\frac{\hat{\sigma}^2(t)}{n}}$

95% Confidence Band: the lower, upper boundaries $L(t), U(t)$ satisfy $P\big(L(t) \leq H(t) \leq U(t) : t \in (0, \infty)\big) = 95\%$.

e.g. $\hat{H}_{NA}(t) \pm c\sqrt{\frac{\hat{\sigma}^2(t)}{n}}$ with c determined by

$$P\Big(\sup_{t>0} \Big|\frac{\hat{H}_{NA}(t) - H(t)}{\sqrt{\hat{\sigma}^2(t)/n}}\Big| \leq c\Big) = 95\%.$$

How to compute the critical value c?

- **Approach I** (based on the Brownian Motion)

Recall $\sqrt{n}\frac{\hat{H}_{NA}(\cdot)-H(\cdot)}{\hat{\sigma}(t)} \to W(\frac{\sigma^2(\cdot)}{\sigma^2(t)})$ in distn: $W(\cdot)$ the standard Brownian motion.
Thus, by the continuous mapping theorem,

$$\sup_{0\le s\le t} \sqrt{n}\Big|\frac{\hat{H}_{NA}(\cdot)-H(\cdot)}{\hat{\sigma}(t)}\Big| \to \sup_{0\le s\le t}\Big|W(\frac{\sigma^2(\cdot)}{\sigma^2(t)})\Big| = \sup_{0\le u\le 1}|W(u)|$$

in distribution.
From the table of Brownian Motion, find c such that $P(\sup_{0\le u\le 1}|W(u)| \le c) = 95\%$.

- **Approach II** (resampling) (Lin et al, 1993, Biometrika)

Recall to choose c such that $P(\sup_{0\le t<\infty}|Q_n(t)|\le c)=95\%$:
$Q_n(t)=\sqrt{n}\frac{\hat{H}_{NA}(\cdot)-H(\cdot)}{\hat{\sigma}(t)}$ is about

$$
\begin{aligned}
& \frac{\sqrt{n}}{\hat{\sigma}(t)}\Big[\int_0^t \frac{I(Y_\cdot(u)>0)}{Y_\cdot(u)}(dN_\cdot(u)-Y_\cdot(u)dH(u))\Big] \\
= \quad & \frac{\sqrt{n}}{\hat{\sigma}(t)}\sum_{i=1}^n\int_0^t \frac{I(Y_\cdot(u)>0)}{Y_\cdot(u)}dM_i(u)
\end{aligned}
$$

Define $\tilde{Q}_n(t)$ as follows

$$
\frac{\sqrt{n}}{\hat{\sigma}(t)}\sum_{i=1}^n\int_0^t \frac{I(Y_\cdot(u)>0)}{Y_\cdot(u)}dN_i(u)Z_i
$$

with $Z_1,\ldots,Z_n$ iid from $N(0,1)$ and $\perp\!\!\!\perp$ the data:
$\tilde{Q}_n(\cdot)$ converges weakly to $\mathit{Gaussian}(0,\sigma^2(t))$ (i.e. asymptotically equivalent to $\sigma(t)Q_n(\cdot)$).

Algorithm. At lth time, $l = 1, \ldots, M$,

- Step A. Generate $(Z_1, \ldots, Z_n)^{(l)}$ as an iid sample from $N(0, 1)$ and $\perp\!\!\!\perp$ the data.
- Step B. Evaluate $\tilde{Q}_n^{(l)}(\cdot)$ over $[0, T^*]$.
- Setp C. Calculate $W^{(l)} = \sup_{0 \leq s \leq T^*} |\tilde{Q}_n^{(l)}(s)|$.

Choose the critical c to be the 95% quantile of $W^{(1)}, \ldots, W^{(M)}$.

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.1.1 Theoretical Preparation*
  - **Part IV.1.2 Counting Process Formulation in LIDA and Applications: Revisits to KM, Logrank, Cox PH**
    - *Part IV.1.2A Formulation*
    - *Part IV.1.2B Revisit to Logrank Test*
    - *Part IV.1.2C Revisit to Cox's Partial Likelihood Approach*
    - *Part IV.1.2D Revisit to Nelson-Aalen Estimator and Breslow Estimator*
    - **Part IV.1.2E Revisit to Kaplan-Meier Estimator**
- **Part IV.2 Selected Recent Topics in LIDA**
- *Part IV.3 Beyond Lifetime Data Analysis*