What to do this week (March 14 and 16, 2023)?

Part I. Introduction
Part II. Epidemiologic Concepts and Designs
Part III. Clinical Trials

Part IV. Modern Biostatistical Approaches

Part IV.1 Incomplete Data Analysis

Part IV.2 Some Other Important Topics (Chp 8 - 18, Koepsell and Weiss, 2003)

Part IV.3 Selected Widely-Used Algorithms

Part IV.4 Lifetime Data Analysis

IV.4.1 Parametric Inference
IV.4.2 Nonparametric Inference: Estimation
IV.4.3 Nonparametric Inference: Testing
IV.4.4 Semiparametric Inference

Part IV.4.1 Lifetime Data Analysis: Parametric Inference

Parametric Analysis with Right-Censored Data

Consider event time r.v. $T \sim f(\cdot;\theta)$: to make inference on θ with a set of right-censored data $\{(U_i, \delta_i) : i = 1, \ldots, n\}$, arising from n indpt individuals

Independent Censoring. the situations with indpt T_i and C_i for $i = 1, \ldots, n$, denoted by $T_i \perp\!\!\!\perp C_i$.

What is the likelihood function $L(\theta|data)$?

$$L(\theta|data) \propto \prod_{i=1}^{n} f(u_i;\theta)^{\delta_i} S(u_i;\theta)^{1-\delta_i} = \prod_{i=1}^{n} h(u_i;\theta)^{\delta_i} S(u_i;\theta)$$

... **applications of MLE/likelihood-based testing proceudres**

Example. r.v. $T \sim NE(1/\theta)$ with a set of right-censored data from n indpt individuals: $\{(u_i, \delta_i) : i = 1, \ldots, n\}$, assuming indpt censoring

- Can we use $\bar{T} = \frac{\sum_{i=1}^n t_i}{n}$, the sample mean to estimate the population mean of T, $E(T) = \theta$?
  $\bar{T} \sim AN(\theta, \theta^2/n)$
- What is the MLE of θ with the censored data?

$$L(\theta|data) = \prod_{i=1}^{n} \left(\frac{1}{\theta} e^{-u_i/\theta}\right)^{\delta_i} \left(e^{-u_i/\theta}\right)^{1-\delta_i} = \frac{1}{\theta}^{\sum \delta_i} \exp(-\sum u_i/\theta)$$

$\frac{\partial log L(\theta)}{\partial \theta} = 0 \Longrightarrow$ the MLE $\hat{\theta} = \frac{\sum_{i=1}^n u_i}{\sum_{i=1}^n \delta_i}$:

$$\sqrt{n}(\hat{\theta} - \theta) \sim AN(0, FI^{-1}(\theta)), \quad n >> 1$$

How to compare the efficiency of MLE $\hat{\theta}$ with $\bar{T} \sim AN(\theta, \theta^2/n)$?

Moreover,

- (i) how to compute the MLE $\hat{\theta}$ in general?
- (ii) how to estimate the variance of the MLE $\hat{\theta}$?

Part IV.4.2-4 Lifetime Data Analysis: Nonparametric/Semiparametric Inference

Motivation

Consider event time $T \sim f(\cdot)$, or $T|X = x \sim f(\cdot|x)$
Goal: to make inference on $f(\cdot)$ or $f(\cdot|x)$

Recall parametric inference in LIDA ...

- What if the parametric model is not plausible?
- What if not to take any risks?

$\Longrightarrow$ the demand of approaches requiring less model assumptions:

- nonparametric inference procedures
- semiparametric inference procedures

... ... *Modern Survival/Failure Time/Lifetime Analysis*

Overview

- Kaplan and Meier (1958, JASA)

  product-limit (Kaplan-Meier) estimator for $S(t)$ with right-censored event times – nonparametric estimator

- Mantel (1966, Cancer Chem); Gehan (1965, Biometrika)

  logrank test (extended Wilcoxon test) with right-censored event times – nonparametric test

- Cox (1972, JRSSB; 1975, Biometrika)

  Cox's proportional hazards model and partial likelihood approach – semiparametric inference

Part IV.4.3 Lifetime Data Analysis: Kaplan-Meier Estimator

Motivation

$T_1, \ldots, T_n \sim F(\cdot)$ iid

the empirical distribution $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n I(T_i \leq t)$, the nonparametric MLE (Kiefer's version)

the empirical distribution $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n I(T_i \leq t)$

- $\forall t \in [0, \infty)$,
  - $E\{\hat{F}_n(t)\} = F(t)$
  - $Var\{\hat{F}_n(t)\} = F(t)[1 - F(t)]/n$
  - $\sqrt{n}\{\hat{F}_n(t) - F(t)\} \to N(0, F(t)[1 - F(t)])$ in distn, as $n \to \infty$
- $\sup_{t \geq 0} \left|\hat{F}_n(t) - F(t)\right| \to 0$ a.s.
- $\sqrt{n}\{\hat{F}_n(t) - F(t)\} \to$ Gaussian Process with mean zero and variance function $F(t)[1 - F(t)]$ in distribution (weak convergence)

What if $\{(U_i, \delta_i) : i = 1, \ldots, n\}$**?**

Recall "Actuarial Life Table"

time interval	number of death	number of withdrawal	number at risk	$\hat{q}_j$	$\hat{p}_j$	$\hat{P}_j$
I_1			... ...			
			... ...			
I_j	D_j	W_j	N_j	$\frac{D_j}{N_j - \frac{1}{2} W_j}$	$1 - \hat{q}_j$	
			... ...			
I_K			... ...			

p_j=P(an individual survives beyond I_j|beyond I_{j-1})
$q_j = 1 - p_j$=P(an individual dies in I_j|beyond I_{j-1})
P_j=P(an individual survives beyond I_j)

Rationale?

Kaplan-Meier Estimator

In general, $F \in \mathcal{F} = \{all\ cdfs\}$
With the right-censored data, the likelihood function

$$L(F) = \prod_{i=1}^{n} dF(u_i)^{\delta_i}[1 - F(u_i)]^{1-\delta_i}$$

Maximize $L(F)$ as $F(\cdot)$ having only masses at the distinct observed event times: $0 = V_0 \leq V_1 < \ldots < V_J \leq V_{J+1}$
$\Longrightarrow$ the Kaplan-Meier estimator (left-continuous)

$$\hat{S}(t) = \prod_{j:V_j<t} \left(1 - \frac{n_j}{N_j}\right) = \begin{cases} 1 & t \leq V_1 \\ \prod_{l=1}^{j}(1 - \hat{h}_j) & V_j < t \leq V_{j+1} \\ ? & t > V_{J+1} \end{cases}$$

Recall a pointwise CI: for $t > 0$, $\hat{S}_{KM}(t) \pm 1.96\sqrt{\hat{Var}(\hat{S}_{KM}(t))}$

an alternative way to construct a CI for $S(t)$:

- to obtain a CI for $\log S(t)$ first

$$\log \hat{S}_{KM}(t) \pm 1.96\sqrt{\hat{V}\big(\log \hat{S}_{KM}(t)\big)}$$

  - $\hat{V}\big(\log \hat{S}_{KM}(t)\big) \approx \sum_{l=1}^{j} Var\big[\log(1 - \hat{h}_l)\big]$ for $t \in [V_j, V_{j+1})$
  - $Var\big[\log(1 - \hat{h}_l)\big] \approx Var(\hat{h}_l)\frac{1}{(1-\hat{h}_l)^2}$ by the Δ-method.
  - $Var(\hat{h}_l) \approx \frac{1}{N_l}\frac{n_l}{N_l}\left(1 - \frac{n_l}{N_l}\right)$

- to obtain a CI for $S(t)$ as

$$\exp\left\{\log \hat{S}_{KM}(t) \pm 1.96\sqrt{\hat{V}\big(\log \hat{S}_{KM}(t)\big)}\right\} = \hat{S}_{KM}(t)e^{\pm 1.96\sqrt{\hat{V}\left(\log \hat{S}_{KM}(t)\right)}}$$

Recall the alternative pointwise CI: for $t > 0$,
$(\hat{S}_{KM}(t)e^{-1.96\sqrt{\hat{Var}(\hat{S}_{KM}(t))}}, \hat{S}_{KM}(t)e^{1.96\sqrt{\hat{Var}(\hat{S}_{KM}(t))}})$

any other alternative constructions for a CI of $S(t)$?

- the logit transformation?

  (*proportional odds failure time models*)

- the probit transformation?

Kaplan-Meier Estimator: Applications

- for comparing two populations' distn with censored data
  e.g. $\sup_{t>0} |\hat{S}_{1,KM}(t) - \hat{S}_{2,KM}(t)|$? an extension of the Kolmogorov-Smirnov test statistic $\sup_{t>0} |F_{1,n}(t) - F_{2,m}(t)|$
  no need to specify the population distributions into parametric models

- for justifying actuarial life table

Kaplan-Meier Estimator: Applications

- for assessing parametric goodness-of-fit with censored data
  - e.g. is $T \sim NE(\lambda)$ $(H(t) = \lambda t)$?
    $\Longrightarrow$ to check if $\log S(t) = -\lambda t$?
    using the scatter plot of $\log \hat{S}(t)$ vs t: is $\log \hat{S}(t)$ linear function of t?
  - e.g. is $T \sim Weibull(\lambda, \rho)$ $(H(t) = \lambda t^{\rho})$?
    $\Longrightarrow$ to check if $\log\big(-\log S(t)\big) = \log \lambda + \rho \log t$?
    using the scatter plot of $\log\big(-\log \hat{S}(t)\big)$ vs $\log t$: look for linearity?

Part IV.4.3 Lifetime Data Analysis: Nonparametric Tests

Introduction

Consider to compare two groups wrt the event time distns ... ...

For example,

- in the placebo group, iid $T_{0i} \sim F_0(\cdot)$: $i = 1, \ldots, n$
- in the treatment group, iid $T_{1j} \sim F_1(\cdot)$: $j = 1, \ldots, m$

$\Longrightarrow H_0 : F_0(\cdot) = F_1(\cdot)$

... Many different ways to differ: any UMP?

- *directional tests*: designated/oriented to a specific type of difference between the two population distns
  e.g. $S_1(t) = S_0(t)^c$
- *omnibus tests*: there is power to detect all or most types of differences but not with great power for a specific difference

Early work with censored data ... ...

- Gehan (1965, Biometrika): modifying rank tests to allow censoring
- Mantel (1966, Cancer Chem): adapting data to use methods for several 2×2 tables
- Application of the Cox partial likelihood approach (Cox, 1975)*

Nonparametric Tests: Gehan Test

Recall that, without censoring and provided the two populations are indpt, the Wilcoxon sign test:

$$\Phi(T_{1i}, T_{0j}) = \begin{cases} 1 & T_{0j} > T_{1i} \\ -1 & T_{0j} < T_{1i} \\ 0 & T_{0j} = T_{1i} \end{cases}$$

$W = \sum_{i=1}^{n} \sum_{j=1}^{m} \Phi(T_{1i}, T_{0j})$

- $E_{H_0}(W) = 0$
- $W/SE(W) \sim N(0,1)$ in distn as $n, m \to \infty$

$\Longrightarrow$ the Wilcoxon sign test

Nonparametric Tests: Gehan Test

- efficiency?
- what if $T_1 \not\perp T_0$?
- what if the data are right-censored?
  $\{(U_{1i}, \delta_{1i}) : i = 1, \ldots, n\} \bigcup \{(U_{0j}, \delta_{0j}) : j = 1, \ldots, m\}$

Nonparametric Tests: Gehan Test

With the right-cenosred data
$\{(U_{1i}, \delta_{1i}) : i = 1, \ldots, n\} \bigcup \{(U_{0j}, \delta_{0j}) : j = 1, \ldots, m\}$:

$$\Phi(U_{1i}, \delta_{1i}; U_{0j}, \delta_{0j}) = \begin{cases} 1, & U_{0j} > U_{1i}; \delta_{1i} = 1 \\ -1, & U_{0j} < U_{1i}; \delta_{0j} = 1 \\ 0, & otherwise \end{cases}$$

$GW = \sum_i \sum_j \Phi(U_{1i}, \delta_{1i}; U_{0j}, \delta_{0j})$ provided indpt censoring,

- $E_{H_0}(GW) = 0$
- $GW/SE(GW) \sim N(0, 1)$ in distn as $n, m \to \infty$

$\Longrightarrow$ Wilcoxon-Gehan testing procedure ...

Nonparametric Tests: Gehan Test

- efficiency?
- what if $T_1 \not\perp T_0$?
- how is it compared to the extended Kolmogorov-Smirnov test based on the KM estm?

Nonparametric Tests: Mantel Test

with all observed distinct event times: $0 < V_1 < \ldots, V_K$

First, consider what happens at time $t = V_l$ ...

	at $t = V_l$		
Group	failure	not	at risk
placebo	n_{0l}	–	N_{0l}
treatment	n_{1l}	–	N_{1l}
total	$n_{.l}$	–	$N_{.l}$

The number of observed failures at time V_l from the treatment group
$O_l = n_{1l} \sim$ Hypergeometric distn under $H_0 : S_0(\cdot) = S_1(\cdot)$

$$P(O_l = a) = \frac{\begin{pmatrix} N_{1l} \\ a \end{pmatrix} \begin{pmatrix} N_{0l} \\ n_{.l} - a \end{pmatrix}}{\begin{pmatrix} N_{.l} \\ n_l \end{pmatrix}}$$

Nonparametric Tests: Mantel Test

with all observed distinct event times: $0 < V_1 < \ldots, V_K$

First, consider what happens at time $t = V_l$ ...

	at $t = V_l$		
Group	failure	not	at risk
placebo	n_{0l}	–	N_{0l}
treatment	n_{1l}	–	N_{1l}
total	$n_{.l}$	–	$N_{.l}$

- the expected number of failures from treatment group $E_l = E(O_l) = n_l \frac{N_{1l}}{N_l}$ under H_0
- $V(O_l) = \frac{N_l - N_{1l}}{N_l - 1} N_{1l} \left(\frac{n_l}{N_l}\right)\left(1 - \frac{n_l}{N_l}\right)$ under H_0

Nonparametric Tests: Mantel Test

Now, pull together the information at all the observed failure times ...

$$Z = \frac{\sum_{l=1}^{K}(O_l - E_l)}{\sqrt{\sum_{l=1}^{K} V(O_l)}} \sim N(0,1)$$

approximately under H_0
$\Longrightarrow$ the Mantel (logrank) testing procedure ...

- $Z^2 \sim \chi^2(1)$ under H_0
- $Z = (\sum_l O_l - \sum_l E_l)/SE(O)$

Nonparametric Tests: Mantel Test

Example. Group 0: $3.1,\ 6.8^+,\ 9,\ 9,\ 11.3^+,\ 16.2$
Group 1: $8.7,\ 9,\ 10.1^+,\ 12.1^+,\ 18.7,\ 23.1^+$

Nonparametric Tests: Mantel Test

Remarks

- similar to the techniques for combining 2×2 tables across trata to test for independence
- efficiency?
- Mantel (logrank) test vs Gehan test?
- oriented towards $S_1(t) = S_0(t)^c$, a directional test

Nonparametric Tests: Variants of Logrank Test

What if the subjects are stratified according to a factor, say, gender?

Stratified Logrank Test with the factor of K levels

$$Z = \frac{\sum_{k=1}^{K}(O^{(k)} - E^{(k)})}{\left(\sum_k V^{(k)}\right)^{1/2}} \sim N(0,1)$$

approximately under H_0.

$\Longrightarrow$ the testing procedure ... ...

Nonparametric Tests: Variants of Logrank Test

What if there is a need to weight the information at different times differently?

Weighted Logrank Test

$$Z_W = \frac{\sum_{l=1}^{L} w_l(O_l - E_l)}{\left(\sum_l w_l^2 V_l\right)^{1/2}} \sim N(0,1)$$

approximately under H_0.

$\Longrightarrow$ the testing procedure ... ...

How to choose the weights in general?

- If $w_l = N_{.l}$, the test is similar to Gehan test.

Nonparametric Tests: Variants of Logrank Test

What if to compare p treatment groups with the placebo group?
$H_0 : S_0(\cdot) = S_1(\cdot) = \ldots = S_p(\cdot)$
Given all the distinct failure times are $0 < V_1 < \ldots < V_L < \infty$,

	at $t = V_l$		
Group	failure	not	at risk
placebo	n_{0l}	...	N_{0l}
treatment 1	n_{1l}	...	N_{1l}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
treatment p	n_{pl}	...	N_{pl}
total	$n_{.l}$	...	$N_{.l}$

$$\mathbf{O}_l = \begin{pmatrix} n_{1l} \\ \vdots \\ n_{pl} \end{pmatrix}; \; \mathbf{E}_l = E\{\mathbf{O}_l\} = \begin{pmatrix} N_{1l} \\ \vdots \\ N_{pl} \end{pmatrix} \frac{n.l}{N_{.l}}; \; \mathbf{V}_l = Var\{\mathbf{O}\}$$

Nonparametric Tests: Variants of Logrank Test

$\tilde{\mathbf{O}} = \sum_{l=1}^{L} \mathbf{O}_l$, $\tilde{\mathbf{E}} = \sum_{l=1}^{L} \mathbf{E}_l$, $\tilde{\mathbf{V}} = \sum_{l=1}^{L} \mathbf{V}_l$

$$\left(\tilde{\mathbf{O}} - \tilde{\mathbf{E}}\right)' \tilde{\mathbf{V}}^{-1} \left(\tilde{\mathbf{O}} - \tilde{\mathbf{E}}\right) \sim \chi^2(p)$$

approximately under H_0, provided the sample size is large.

$\Longrightarrow$ the testing procedure ... ...

- The test is *omnibus*.
- If a trend test is intended?
  to consider $\mathbf{c}' \left(\tilde{\mathbf{O}} - \tilde{\mathbf{E}}\right) \sim N(0, \mathbf{c}'\tilde{\mathbf{V}}\mathbf{c})$?

IV.4.4 Semiparametric Inference: Cox Proportional Hazards Model

- Recall the two-sample problem $\rightarrow$ testing on $H_0 : h_1(\cdot) = h_0(\cdot)$
  - $Z = \begin{cases} 1 & \textit{treatment} \\ 0 & \textit{placebo} \end{cases}$,
    to study event time $T|Z = z$?
  - with general covariates Z, to explore event time $T|Z = z$?
    $\Longrightarrow$ regression modeling?

- Feigl and Zelen (1965)
  $T|Z = z \sim NE(\lambda_z)$: $h(t|z) = \lambda_z = \lambda_0 e^{\beta z}$
  $\beta = 0 \rightarrow$ no effect of Z

$\Longrightarrow$ **Cox Proportional Hazards Model** (Cox, JRSSB 1972)

Cox Proportional Hazards Model: (Cox, JRSSB 1972)
The hazard function of event time $T|Z = z$ is

$$h(t|z) = h_0(t)e^{\beta z}, \quad t > 0$$

The conditional survivor function is

$$S(t|z) = \exp(-\int_0^t h_0(u)e^{\beta z}du) = \exp(-H_0(t)e^{\beta z}), \quad t > 0$$

Remarks

- the hazard ratio $h(t|Z=z_1)/h(t|Z=z_0) = e^{\beta(z_1-z_0)}$ for all $t>0$ *proportional*!
- $Z = \begin{cases} 1 & \textit{treatment} \\ 0 & \textit{placebo} \end{cases}$, e^{β}: treatment effect
- $Z_1 = \begin{cases} 1 & \textit{treatment} \\ 0 & \textit{placebo} \end{cases}$, $Z_2 = \begin{cases} 1 & \textit{male} \\ 0 & \textit{female} \end{cases}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, $\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, $h(t|\mathbf{Z}) = h_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}$: relative impacts of the treatment to female and male are the same.

Cox Proportional Hazards Model: Estimation of β

Often is interested to estm β in the Cox PH model, for comparison/evaluate/assess effect ... ...

With right-censored event times along with the covariates

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

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

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

$$L(\beta, h_0(\cdot)\big|data) = L_1(\beta\big|data)L_2(\beta, h_0(\cdot)\big|data)$$

$\Longrightarrow$ **the Cox partial likelihood function** (Cox, Biometrika 1975)

the Cox partial likelihood function (Cox, Biometrika 1975)

$$L_1(\beta \big| data) = \prod_{i=1}^{n} \Big(\frac{e^{\beta z_i}}{\sum_{l \in \mathcal{R}_i} e^{\beta z_l}} \Big)^{\delta_i}$$

the risk set at time u_i: $\mathcal{R}_i = \{j : u_j \geq u_i\}$

$\Longrightarrow$ the MPLE (maximum partial likelihood estimator) of β:

$$\hat{\beta} = argmax_{all\ \beta} L_1(\beta \big| data)$$

With some conditions, as $n \to \infty$

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

Example. $n = 5$ indpt subjects and $Z = \begin{cases} 1 & \textit{treatment} \\ 0 & \textit{placebo} \end{cases}$

$(u_i, \delta_i, z_i):\ (16,1,1),\ (13,0,0),\ (21,1,1),\ (11,1,0),\ (12,1,1)$

$$L_1(\beta) \propto \frac{e^{\beta}}{(3e^{\beta}+2)(3e^{\beta}+1)},\ \ \partial \log L_1(\beta)/\partial\beta = 1 - \frac{9e^{\beta}(2e^{\beta}+1)}{(3e^{\beta}+2)(3e^{\beta}+1)}$$

$\Longrightarrow \hat{\beta} = \frac{1}{2}\log 2 - \log 3$

Remarks

- *implementation*
  - to use $\log L_1(\beta) = \sum_{i=1}^{n} \delta_i \{ \beta z_i - \log(\sum_{l \in \mathcal{R}_i} e^{\beta z_l}) \}$

    or $U(\beta) = \partial \log L_1(\beta) / \partial \beta = \sum_{i=1}^{n} \delta_i \{ z_i - \frac{\sum_{l \in \mathcal{R}_i} z_l e^{\beta z_l}}{\sum_{l \in \mathcal{R}_i} e^{\beta z_l}} \} = 0$
  - e.g. *R: coxph*

Remarks (cont'd)

- *interpretation*
    - recall *likelihood, marginal likelihood, conditional likelihood, partial likelihood*
    - **the Cox partial likelihood function** of β
        - conditional arguments
        - the marginal distn of the rank statistic when no tie, no censored observation, cfs: Kalbfleisch and Prentice (1980, 2011)

Cox Proportional Hazards Model: Testing on β

Consider $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$

the partial score test

$$U(\beta) = \partial \log L_1(\beta)/\partial\beta = \sum_{i=1}^{n} \delta_i \Big[z_i - \frac{\sum_{l \in \mathcal{R}_i} z_l e^{\beta z_l}}{\sum_{l \in \mathcal{R}_i} e^{\beta z_l}} \Big]$$

Based on $U(\beta)/\sqrt{n} \sim AN(0, ??)$ as $n \to \infty$ with some conditions,

$\Longrightarrow$ the partial score testing procedure ...

Remarks.

- e.g. when $Z = \begin{cases} 1 & treatment \\ 0 & placebo \end{cases}$

  $U(\beta)\big|_{\beta=0} = \sum_{l=1}^{L} \left(O_l - n_{.l} \frac{N_{1l}}{N_{.l}} \right) = O - E$, the numerator of the logrank test statistic

- the Wald-type, using the MPLE of β and its asymptotic normality?

- the PLRT, using the structure of LRT?