What to do today (2022/01/20)?

Part II.4 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 $\big\{(U_i, \delta_i) : i = 1, \ldots, n\big\}$, arising from n indpt individuals

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

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 2.1. 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, $E(T) = \theta$? $\bar{T} \sim AN(\theta, \theta^2/n)$
- Can it be a 'good estimator' $\tilde{\theta} = \frac{\sum_{i=1}^n U_i}{n}$?
- What is the MLE of θ with the censored data?

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

$\frac{\partial logL(\theta)}{\partial\theta} = 0 \Longrightarrow$ the MLE $\hat{\theta} = \frac{\sum_{i=1}^n u_i}{\sum_{i=1}^n \delta_i}$ and $\hat{\theta} \sim AN(\theta, 1/nFI(\theta))$, $n >> 1$

$-\frac{1}{n}\frac{\partial^2 logL(\theta)}{\partial\theta^2} \to \frac{1}{\theta^2}\left(\frac{2E(U)}{\theta} - p\right)$, the Fisher information $FI(\theta)$
$E(U)$ the expectation of $U = \min(T, C)$ $p = \lim_{n\to\infty} \frac{\sum_{i=1}^n \delta_i}{n} = P(T \le C)$.

Work on HW-Q1.1 to check those out via simulation!

Example 2.2. Suppose an event time $T \sim f(t;\theta) = \exp(\theta^{'} D(t) + A(t) + B(\theta))$, a member of exponential family in its natural parameterization. Consider a study with n indpt individuals ...

if available are iid realizations of T: $\{T_1, \ldots, T_n\}$

$L_0(\theta|\mathbf{T}) = \prod_{i=1}^{n} f(T_i;\theta) = \exp\{\theta^{'} \sum_{i=1}^{n} D(T_i) + \sum_{i=1}^{n} A(T_i) + nB(\theta)\}$

$\frac{\partial log L_0(\theta|\mathbf{T})}{\partial\theta} = \sum_{i=1}^{n} D(T_i) + n\frac{\partial B(\theta)}{\partial\theta}$

- Sufficient Statistic of θ: $\sum_{i=1}^{n} D(T_i)$
- $E_\theta\big[D(T)\big] = -\partial B(\theta)/\partial\theta$;
- $\partial^2 \log L_0(\theta|\mathbf{T})/\partial\theta^2 = n\partial^2 B(\theta)/\partial\theta^2 = -n\partial E_\theta\big[D(T)\big]/\partial\theta$
  $\Longrightarrow FI(\theta) = -\partial^2 B(\theta)/\partial\theta^2 = \partial E_\theta\big[D(T)\big]/\partial\theta$
- MLE of θ: $\hat{\theta}$ is the solution to $\sum_{i=1}^{n} D(T_i) - nE_\theta\big[D(T)\big] = 0$
- $\hat{\theta} \sim AN(\theta, 1/nFI(\theta))$

Example 2.2. Suppose an event time $T \sim f(t;\theta) = \exp(\theta^{'} D(t) + A(t) + B(\theta))$, a member of exponential family in its natural parameterization. Consider a study with n indpt individuals ...

if the data collection subject to indpt right-censoring ...: $\{(U_i, \delta_i) : i = 1, \ldots, n\}$

$L(\theta|\mathbf{U}, \boldsymbol{\delta}) = \prod_{i=1}^{n} f(U_i;\theta)^{\delta_i} S(U_i;\theta)^{1-\delta_i}$

- $\Longrightarrow$ Issue 1. how to obtain MLE $\hat{\theta}$?
- $\Longrightarrow$ Issue 2. how to estimate $V(\hat{\theta})$?

Issue 1. EM (Expectation-Maximization) Algorithm (cf: Dumpster, Laird and Rubin, 1977; Self-Consistency Algorithm, cf: Turnbull, 1976) an iterative procedure for computing MLE

e.g. in the setting with right-censored data ... ...
Define $Q(\theta, \theta^*) = E\{\log L_0(\theta|\mathbf{T})|\mathbf{U}, \boldsymbol{\delta}; \theta^*\}$

Given $\theta^{(j-1)}$, $j \geq 1$,

- E-step. $Q(\theta, \theta^{(j-1)}) = E\{\log L_0(\theta|\mathbf{T})|\mathbf{U}, \boldsymbol{\delta}; \theta^{(j-1)}\}$
- M-step. Obtain $\theta^{(j)}$ such that
  $Q(\theta^{(j)}, \theta^{(j-1)}) = \max_{all\ \theta} Q(\theta, \theta^{(j-1)})$

iterating ... $\Longrightarrow \{\theta^{(j)} : j = 1, 2, \ldots\}$
The sequence converges to $\hat{\theta}$, the maximum point of $\log L(\theta|\mathbf{U}, \boldsymbol{\delta})$, provided convergence.

Remarks:

- Why does it work? $\log L(\theta^{(j)}|\mathbf{U}, \boldsymbol{\delta}) \nearrow$ as $j \nearrow$

- When $\log L(\theta|\mathbf{T})$ is a linear function of $T_1, \ldots, T_n$, "E-step" is to get $E(T_i|U_i, \delta_i)$.

- "M-step" is replaced with an "S-step" when to max $Q(\theta, \theta^*)$ with fixed θ^* can be achieved by solving the equation $\partial Q(\theta, \theta^*)/\partial\theta = 0$.

- Why is it so popular? intuitive;
not very efficient, though

Example 2.2. (cont'd) an application of EM-algorithm

$$Q(\theta, \theta^*) = \theta' E\Big\{ \sum_{i=1}^{n} D(T_i) \Big| \mathbf{U}, \boldsymbol{\delta}; \theta^* \Big\} + nB(\theta)$$

$$\partial Q(\theta, \theta^*)/\partial \theta = E\Big\{ \sum_{i=1}^{n} D(T_i) \Big| \mathbf{U}, \boldsymbol{\delta}; \theta^* \Big\} + n\partial B(\theta)/\partial \theta$$

Given an initial value $\theta^{(0)}$,

- E-step. Calculate $E\big\{ \sum_{i=1}^{n} D(T_i) \big| \mathbf{U}, \boldsymbol{\delta}; \theta^{(0)} \big\}$.
- M-step. Solve $E\big\{ \sum_{i=1}^{n} D(T_i) \big| \mathbf{U}, \boldsymbol{\delta}; \theta^{(0)} \big\} = -n\partial B(\theta)/\partial \theta$ $\Longrightarrow \theta^{(1)}$.

Use $\theta^{(1)}$ to update $\theta^{(0)}$ and repeat E-step and M-step. .. ...
$\Longrightarrow \{\theta^{(j)} : j = 1, \ldots\}$
The limit is $\hat{\theta}$.

Issue 2. Variance Estimation for MLE $\hat{\theta}$

- Recall $\hat{\theta} \sim AN\big(\theta, AV(\hat{\theta})\big)$ when $n >> 1$
  - if iid case, $AV(\hat{\theta}) = \frac{1}{n} FI(\theta)^{-1}$;
    in general,

$$AV(\hat{\theta}) = E\Big(- \frac{\partial^2 \log L(\theta)}{\partial \theta^2}\Big)^{-1} = V\Big(\frac{\partial \log L(\theta)}{\partial \theta}\Big)^{-1}$$

- Estimating $AV(\hat{\theta})$ by $-\frac{\partial^2 \log L(\theta)}{\partial \theta^2}^{-1}\Big|_{\hat{\theta}}$
- Robust Variance Estimator: the Huber sandwich estimator is based on

$$E\Big(- \frac{\partial^2 \log L(\theta)}{\partial \theta^2}\Big)^{-1} V\Big(\frac{\partial \log L(\theta)}{\partial \theta}\Big) E\Big(- \frac{\partial^2 \log L(\theta)}{\partial \theta^2}\Big)^{-1}$$

- Alternative variance estimator?

  Bootstrap, Jackknife resampling variance estimation

  - e.g. Bootstrap variance estm (cf. Efron and Tibshirani, 1993)

Viewing $\theta = \theta(F)$ and thus $\hat{\theta} = \theta(\hat{F})$ ...

data $\mathbf{X} \Rightarrow \hat{\theta}$: $V(\hat{\theta}) =$?

bootstrap samples $\mathbf{X}_1^*, \ldots, \mathbf{X}_B^* \perp\!\!\!\perp \Rightarrow \hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$:

$$\bar{\hat{\theta}}^* = \sum_{b=1}^{B} \hat{\theta}_b^*/B; \quad \hat{V}(\hat{\theta}) = \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2/(B-1)$$

What to study next?

Part I. Preliminaries

Part II. Parametric Inference

Part III. Nonparametric/Semi-parametric Inference

- **Part III.1. Introduction and Overview**
- **Part III.2. Kaplan-Meier Estimator**
- *Part III.3. Nonparametric Tests*
- *Part III.4. Cox Proportional Hazards Function*

Part IV. Further Topics