

What to do today (2022/01/25)?

Part I. Preliminaries

Part II. Parametric Inference

Part II.1 Some Basic Concepts

Part II.2 Commonly Used Parametric Distributions

Part II.3 Incomplete Data Structures

Part II.4 Parametric Analysis with Right-Censored Data

Part III. Nonparametric/Semi-parametric Inference

Part III.1. Introduction and Overview: Motivation

Part III.2. Kaplan-Meier Estimator

Part III.3. Nonparametric Tests

Part III.4. Cox Proportional Hazards Model

Part IV. Further Topics

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 : $\{T_1, \dots, T_n\}$

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

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

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 ...:

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

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

- ▶ \implies Issue 1. how to obtain MLE $\hat{\theta}$?
- ▶ \implies 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 $Q(\theta, \theta^*) = E\{\log L_0(\theta|\mathbf{T})|\mathbf{U}, \delta; \theta^*\}$

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

► E-step. $Q(\theta, \theta^{(j-1)}) = E\{\log L_0(\theta|\mathbf{T})|\mathbf{U}, \delta; \theta^{(j-1)}\}$

► M-step. Obtain $\theta^{(j)}$ such that

$$Q(\theta^{(j)}, \theta^{(j-1)}) = \max_{\text{all } \theta} Q(\theta, \theta^{(j-1)})$$

iterating ... $\Rightarrow \{\theta^{(j)} : j = 1, 2, \dots\}$

The sequence converges to $\hat{\theta}$, the maximum point of $\log L(\theta|\mathbf{U}, \delta)$,
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, \dots, 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

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

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

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

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

Use $\theta^{(1)}$ to update $\theta^{(0)}$ and repeat E-step and M-step.

$$\implies \{\theta^{(j)} : j = 1, \dots\}$$

The limit is $\hat{\theta}$.

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

- ▶ Recall $\hat{\theta} \sim AN(\theta, AV(\hat{\theta}))$ when $n >> 1$

- ▶ if iid case, $AV(\hat{\theta}) = \frac{1}{n} FI(\theta)^{-1}$;
in general,

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

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

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

- ▶ 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 $\mathbf{X}_1^*, \dots, \mathbf{X}_B^*$ $\perp \Rightarrow \hat{\theta}_1^*, \dots, \hat{\theta}_B^*$:

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

Part III. Nonparametric/Semi-parametric Inference

Part III.1. Introduction and Overview: 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?

⇒ the demand of approaches requiring less model assumptions:

- ▶ nonparametric inference procedures
- ▶ semiparametric inference procedures

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

Part III.1. Introduction and Overview: Overview I

- ▶ 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 III.1. Introduction and Overview: Overview II

- ▶ Kaplan and Meier (1958, JASA)

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

- ▶ Breslow and Crowley (1974, Annals)
- ▶ Aalen (1975, UC-Berkley PhD thesis)

Counting Process Formulation and Martingale Results

Part III.1. Introduction and Overview: Overview II

- ▶ Mantel (1966, Cancer Chem); Gehan (1965, Biometrika)
 - logrank test (extended Wilcoxon test) with right-censored event times – nonparametric test
 - ▶ Gill (1980, *Censoring and Stochastic Integrals*)

Part III.1. Introduction and Overview: Overview II

- ▶ Cox (1972, JRSSB; 1975, Biometrika)

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

- ▶ Tsiatis (1981, Annals)
- ▶ Anderson and Gill (1982, Annals)

Modern Empirical Process Theory (cfs: R. Dudley, 1989; D. Pollard, 1990)



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 Model*

Part IV. Further Topics