

What to do today (2022/01/27)?

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 Function

Part IV. Advanced Topics

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

... .. *recent 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)
- ▶ *application of Martingale Limit Theory* (cf: Andersen et al, 2012)
- ▶ *application of Modern Empirical Process Theory* (cfs: R. Dudley, 1989; D. Pollard, 1990)

Part III.2. Kaplan-Meier Estimator: Motivation

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

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

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

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

Part III.2. Kaplan-Meier Estimator: Lifetable

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

Rationale?

Part III.2. Kaplan-Meier Estimator: Discrete Case

Consider a discrete event time $T \sim F(\cdot)$ with mass points at $0 \leq V_1 < \dots < V_K$

- ▶ $h_1 = P(T = V_1)$; $h_j = P(T = V_j | T > V_{j-1})$
- ▶ For $t \in [V_j, V_{j+1})$, $S(t) = P(T > t) = P(T > V_j) = P(T > V_j | T > V_{j-1})P(T > V_{j-1}) = \prod_{l=1}^j (1 - h_l)$
- ▶ $P(T = V_j) = h_j S(V_{j-1})$

$$L(F) = \prod_{i=1}^n f(u_i)^{\delta_i} S(u_i)^{1-\delta_i} = \prod_{j=1}^K h_j^{n_j} (1 - h_j)^{N_j - n_j}$$

with $n_j = \#$ who fail at $V_j = \sum \delta_i I(u_i = V_j)$,
 $N_j = \#$ who at risk at $V_j = \sum I(u_i \geq V_j)$

Part III.2. Kaplan-Meier Estimator: Discrete Case

$\implies$ the MLE $\hat{h}_j = \frac{n_j}{N_j}$. Thus

$$\hat{S}(t) = \begin{cases} 1 & t \leq V_1 \\ \prod_{l=1}^j (1 - \hat{h}_l) & V_j < t \leq V_{j+1} \\ 0 & t > V_K \end{cases}$$

left-continuous

Example. $V_j : 2, 4, 5, 7, 9, 11, 16, 18, 20$; Data ($n=10$): 2, 2, 3⁺, 5, 5⁺, 7, 9, 16, 16, 18⁺

Part III.2. Kaplan-Meier Estimator: A General Case

In general, $F \in \mathcal{F} = \{\text{all cdfs}\}$

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: $0 = V_0 \leq V_1 < \dots < V_J \leq V_{J+1}$

$\implies$ the Kaplan-Meier estimator (left-continuous)

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

Part III.2. Kaplan-Meier Estimator: More ...

Remarks

- ▶ the general case vs the discrete case
- ▶ the observed sample mean (reduced sample estm) vs KM estm
- ▶ asymptotic properties of $\hat{S}_{KM,n}(\cdot)$
 - ▶ $\hat{S}_{KM,n}(\cdot) \rightarrow S(\cdot)$ a.s. uniformly
i.e. $\sup_{0 < t < \infty} |\hat{S}_{KM,n}(t) - S(t)| \rightarrow 0$ a.s.
 - ▶ weak convergence $\sqrt{n}(\hat{S}_{KM,n}(t) - S(t)) \rightarrow$ Gaussian Process with mean zero in distribution
- ▶ interval estm
 - ▶ pointwise CI: for $t > 0$, $\hat{S}_{KM}(t) \pm 1.96\sqrt{\hat{Var}(\hat{S}_{KM}(t))}$
 - ▶ confidence band:
$$P\left(S(t) \in (\hat{S}_L(t), \hat{S}_U(t)) : t \in (0, \infty)\right) \geq 95\%$$

how to compute CB?

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

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

- ▶ $\hat{V}(\log \hat{S}_{KM}(t)) \approx \sum_{l=1}^j \text{Var}[\log(1 - \hat{h}_l)]$ for $t \in [V_j, V_{j+1})$

- ▶ $\text{Var}[\log(1 - \hat{h}_l)] \approx \text{Var}(\hat{h}_l) \frac{1}{(1 - \hat{h}_l)^2}$ by the Δ -method.

- ▶ $\text{Var}(\hat{h}_l) \approx \frac{1}{N_l} \frac{n_l}{N_l} \left(1 - \frac{n_l}{N_l}\right)$

- ▶ to obtain a CI for $S(t)$ as

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

Part III.2. Kaplan-Meier Estimator: Applications

- ▶ for comparing two populations' distn with censored data
no need to specify the population distributions into parametric models
- ▶ for justifying actuarial life table

Part III.2. Kaplan-Meier Estimator: Applications

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

Part III.2. Kaplan-Meier Estimator: More ...

Remarks

- ▶ the observed sample mean (reduced sample estm) vs KM estm
- ▶ asymptotic properties of $\hat{S}_{KM,n}(\cdot)$
 - ▶ $\hat{S}_{KM,n}(\cdot) \rightarrow S(\cdot)$ a.s. uniformly
i.e. $\sup_{0 < t < \infty} |\hat{S}_{KM,n}(t) - S(t)| \rightarrow 0$ a.s.
 - ▶ weak convergence $\sqrt{n}(\hat{S}_{KM,n}(t) - S(t)) \rightarrow$ Gaussian Process with mean zero in distribution
- ▶ interval estm
 - ▶ pointwise CI: for $t > 0$, $\hat{S}_{KM}(t) \pm 1.96\sqrt{\hat{Var}(\hat{S}_{KM}(t))}$
 - ▶ confidence band:
$$P\left(S(t) \in (\hat{S}_L(t), \hat{S}_U(t)) : t \in (0, \infty)\right) \geq 95\%$$

how to compute CB?

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

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

- ▶ $\hat{V}(\log \hat{S}_{KM}(t)) \approx \sum_{l=1}^j \text{Var}[\log(1 - \hat{h}_l)]$ for $t \in [V_j, V_{j+1})$

- ▶ $\text{Var}[\log(1 - \hat{h}_l)] \approx \text{Var}(\hat{h}_l) \frac{1}{(1 - \hat{h}_l)^2}$ by the Δ -method.

- ▶ $\text{Var}(\hat{h}_l) \approx \frac{1}{N_l} \frac{n_l}{N_l} \left(1 - \frac{n_l}{N_l}\right)$

- ▶ to obtain a CI for $S(t)$ as

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

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?

Part III.2. 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

Part III.2. Kaplan-Meier Estimator: Applications

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

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. Advanced Topics