STAT 855. Lifetime Data Analysis

X. Joan Hu

Department of Statistics and Actuarial Science
Simon Fraser University

Spring 2022

What to do today (2022/01/18)?

Part II. Parametric Inference in LIDA

Part II.1 Some Basic Concepts

Consider continuous r.v. $T \geq 0$, time to an event

- probability density function (pdf) $f(t), t \geq 0$
- cumulative distribution function (cdf) $F(t) = P(T \leq t), t \geq 0$
- survivor function $S(t) = P(T \geq t), t \geq 0$

Part II.1 Some Basic Concepts

Consider continuous r.v. $T \geq 0$, time to an event

- hazard function

$$h(t) = \lim_{\Delta t \to 0+} \frac{P\big(T \in [t, t + \Delta t) \big| T \geq t\big)}{\Delta t}$$

  interpretation, analytic properties, ...

- cumulative hazard function $H(t) = \int_0^t h(u)du, t \geq 0$

Part II.2 Commonly Used Parametric Distributions

Exponential distribution $T \sim NE(\lambda), \lambda > 0$

$$f(t;\lambda) = \lambda \exp(-\lambda t), \quad t > 0$$

(or $f(t;\theta) = \frac{1}{\theta}\exp(-t/\theta),\ t > 0$ with $\theta > 0$)

- $E(T)$
- $S(t)$
- $h(t)$
- memoryless property: $P(T > a + b | T > a) = P(T > b)$
- the central role

Part II.2 Commonly Used Parametric Distributions

Weibull distribution $T \sim \mathit{Weibull}(k, \theta)$:

$$f(t;\theta,k) = \frac{k}{\theta}\Big(\frac{t}{\theta}\Big)^{k-1} \exp\big(-(t/\theta)^k\big), \quad t > 0$$

with $\theta > 0$ and $k > 0$.

- $E(T) = \theta\Gamma(1 + 1/k)$, $\mathit{Var}(T) = \theta^2\Big[\Gamma(1+\frac{2}{k}) - \Gamma(1+\frac{1}{k})^2\Big]$
- $S(t)$
- $h(t) = \frac{k}{\theta}\Big(\frac{t}{\theta}\Big)^{k-1}$
- $T^k \sim NE(1/\theta)$

A translated Weibull distribution: $T \sim \mathit{Weibull}(k, \theta, \alpha)$:

$$f(t;\theta,k) = \frac{k}{\theta}\Big(\frac{t-\alpha}{\theta}\Big)^{k-1} \exp\Big\{-\Big(\frac{t-\alpha}{\theta}\Big)^k\Big\}, \quad t > \alpha$$

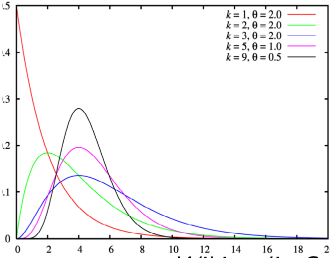
Part II.2 Commonly Used Parametric Distributions

Gamma distribution $T \sim \Gamma(k, \theta)$:

$$f(t; \theta, k) = \frac{1}{\theta\Gamma(k)}\left(\frac{t}{\theta}\right)^{k-1} \exp\left(-t/\theta\right), \quad t > 0$$

with $\theta > 0$ and $k > 0$.

- $E(T) = k\theta,\ Var(T) = k\theta^2$
- $S(t)$
- $h(t)$
- T_1, T_2 indpt and $T_j \sim \Gamma(\alpha_j, \theta)$: $T_1 + T_2 \sim \Gamma(\alpha_1 + \alpha_2, \theta)$
- $\frac{2}{\theta} T \sim \chi^2(2k)$



– *Wikimedia Commons*

Part II.2 Commonly Used Parametric Distributions

Log-normal distribution $T \sim logN(\mu, \sigma^2)$, ie $logT \sim N(\mu, \sigma^2)$:

$$T \sim \frac{1}{t\sigma\sqrt{2\pi}} \exp\big(-\frac{(logt-\mu)^2}{2\sigma^2}\big),\ \ t > 0$$

with $\sigma > 0$.

- $E(T) = \exp(\mu + \sigma^2/2)$
- $S(t)$
- $h(t)$

Part II.2 Commonly Used Parametric Distributions

- **Extreme value distribution**
- **Gumbel distribution**
- **... ...**

See books on reliability, such as Lawless (2003), for more examples of parametric models for event time

Part II.3 Incomplete Data Structures – Censoring

Consider event time r.v. T: to make inference on its distn

- When there are iid observations on T,
  $\Rightarrow$ various approaches such as likelihood-based methods, especially when $n >> 1$

- What if it is not allowed to collect iid observations from the population?

A Reliability Example: at a lab to conduct an experiment to assess the quality of a certain kind of light bulb ... ... (the distn of T, the lifetime of such light bulb?)

Part II.3 Incomplete Data Structures – Censoring

... select n such light bulbs randomly, plug in them at the same time,

- wait till all of them burned out: record the lifetimes $T_1, \ldots, T_n$; take them as iid observations on T

- alternatively, choosing a time c before the experiment, stop the experiment after time c elapses: only available are T_i if $T_i \leq c$, $i = 1, \ldots, n \Rightarrow$ **type I censoring**

- or, choosing an interger $r < n$ before the experiment, stop the experiment after r number of light bulbs burn out: only available are $T_{(1)} < T_{(2)} < \ldots < T_{(r)}$. $\Rightarrow$ **type II censoring**

Part II.3 Incomplete Data Structures – Censoring

Consider event time r.v. T: to make inference on its distn
Let $\{T_1, \ldots, T_n\}$ be iid observations on T with the order statistics $T_{(1)} < \ldots < T_{(n)}$

- **type I censoring** with a predetermined time c, T_i is observed only if $T_i \le c$ for $i = 1, \ldots, n$.
- **type II censoring** with a predetermined number r, T_i is observed only if its one of $T_{(1)}, \ldots, T_{(r)}$ for $i = 1, \ldots, n$.

What if it is in a clinical trial ...

... staggered entries of the study subjects, predetermined study duration of A

Part II.3 Incomplete Data Structures – Censoring

- **right-censoring** Let C_i be the censoring time associated with unit i. T_i is observed if $T_i \leq C_i$ for $i = 1, \ldots, n$

In general, the right-censored data are presented as follows:

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

with $U_i = \min(T_i, C_i)$ (or denoted by $T_i \wedge C_i$),
and $\delta_i = I(T_i \leq C_i) = \begin{cases} 1, & T_i \leq C_i \\ 0, & otherwise \end{cases}$

Example. $n = 3$ and $\{(4, 1), (10, 0), (9, 1)\}$

Special Cases

- **type I censoring.** $C_i \equiv c$ for all i;
- **type II censoring.** $C_i = T_{(r)}$ for all i

How to make inference with right-censored data?

What to study next?

Part I. Preliminaries

Part II. Parametric Inference

- *II.1 Some basic concepts*
- *II.2 Commonly used parametric distributions*
- *II.3 Incomplete data structures*
- **II.4 Parametric analysis with right-censored data**

Part III. Nonparametric/Semi-parametric Inference

Part IV. Further Topics