SPA 2403: SURVIVAL ANALYSIS

Dr. Mutua Kilai

Department of Pure and Applied Sciences

2024-01-22

Kirinyaga University

Introduction

- Survival analysis is the study of survival times and of the factors that influence them.
- Examples of studies with survival outcomes include the following:

    i. Clinical trials

    ii. Prospective Studies

    iii. Retrospective Observational Studies
- Example of survival times include:
    - Time from birth until death
    - Time from birth to development of lung cancer
    - Time to relapse

Contents of Survival Analysis

a. Estimation of the survival distribution

b. Comparisons of the survival distributions of various treatments

c. Examining the factors that influence survival times (Regression models in survival context)

Features of Survival Data

The key features of survival data include:

- The response variable is a non-negative discrete or continuous random variable that represents the time from a well defined origin to a well defined event
- Censoring that arises when starting or ending events are not precisely defined.

Types of Censoring

- Censoring occurs when the event of interest is not observed for some subjects before the study is terminated.
- We present three types of censoring models. Let $T_1, T_2, ..., T_n$ be independent and identically distributed with distribution function F

i. Right Censoring

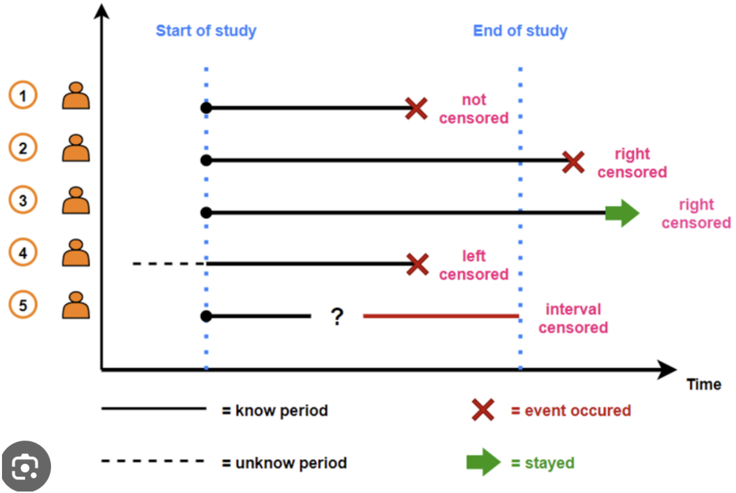
   a. Type I censoring

   b. Type II censoring

ii. Left Censoring

iii. Interval censoring

Pictorial Representation



Right Censoring

- Right censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred.
- For example, we consider patients in a clinical trial to study the effect of treatments on stroke occurrence.
- The study ends after 5 years. Those patients who have had no strokes by the end of the year are censored.

Type I Censoring

- In Type I censoring the censoring times are pre-specified.
- An example is a smoking cessation study where by design each subject is followed until relapse (return to smoking) or 180 days or whichever comes first. The subjects that did not relapse within the 180 days were censored at that time.

Type II Censoring

- This type of censoring occurs when the experimental objects are followed until a pre-specified fraction have failed.
- Such design is rare in bio medical studies but are used in engineering set ups where time to failure is of primary interest.
- An example is where a study stops after for example 25 out of 100 devices are observed to fail. The rest 75 are censored.

Left Censoring

- Left censoring is when the event of interest has already occurred before enrollment. This is very rarely encountered.
- For example, in a medical study someone dies before the drug trial begins.

Interval Censoring

- Interval-censoring occurs in survival analysis when the time until an event of interest is not known precisely (and instead, only is known to fall into a particular interval).

Censoring Occurence

- *Loss to Follow Up:* Patient moves away. We never see him again. We only know he has survived from entry date until he left. So his survival time is greater than the observed value.
- *Drop Out:* Bad side effects forces termination of treatment. Or patient refuses to continue treatment for whatever reasons.
- *Termination of Study:* Patient is still "alive" at end of study.

Goals of survival data

- The goals of survival analysis are to:

i. Estimate the survival distribution

ii. To compare two or more survival distributions

iii. To assess the effects of a number of survival factors on survival

Examples of Survival data

- In this entire unit Survival Analysis we use the asaur package.
- To install the package use:

```
install.packages("asaur")
```

Example 1

- This is a Phase II (single sample) clinical trial of Xeloda and oxaliplatin (XELOX) chemotherapy given before surgery to 48 advanced gastric cancer patients with paraaortic lymph node metastasis.

Table 1: Example 1

time	Weeks	delta
23	42	1
24	43	1
25	43	0
26	46	1

Basic Principles of Survival Analysis

- Let T denote a non-negative random variable representing the lifetimes of individuals in some population.
- Let $F(.)$ denote the cumulative **distribution function** of T with corresponding **probability density function** $f(.)$. Then

$$F(t) = P(T \leq t) = \int_0^t f(x)dx \tag{1}$$

Survival Function

- The probability that an individual survives to time t is given by the **survivor function**.

$$S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(x)dx \qquad (2)$$

- The function is also called **reliability function**.
- $S(t)$ is a monotone decreasing function with $S(0) = 1$ and $S(\infty) = 0$
- We can express the pdf as:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \qquad (3)$$

Hazard Function

- The **hazard function** specifies the instantaneous rate of failure at $T = t$ given that the individual survived up to time t and is defined as:

$$h(t) = lim_{\Delta t \to 0^+} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{4}$$

- The hazard rate is a rate, rather than a probability. It can assume values in $[0, \infty)$.

- The hazard function is related to the PDF and the survival function by:

$$h(t) = \frac{f(t)}{S(t)}$$

- Integrating $h(t)$ over $(0, t)$ gives the cumulative hazard function $H(t)$ given as

$$H(t) = \int_0^t h(u)du = -\log\Big(S(t)\Big) \tag{5}$$

- Thus $S(t)$ can be expressed as:

$$S(t) = \exp\Big(-H(t)\Big)$$

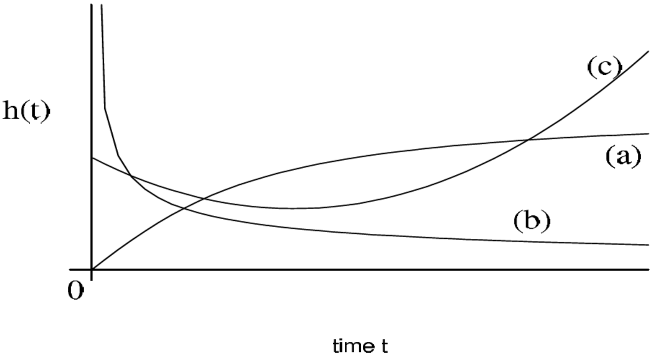
- The pdf of T can be written as:

$$f(t) = h(t)\exp\Big(-\int_0^t h(u)du\Big) \tag{6}$$

Types of Hazard Functions

- Model (a) has an increasing hazard rate. It arises when there is a natural aging war.
- Model (b) has a decreasing hazard rate. Arises in patients experiencing certain types of organ transplant.
- Model (c) has a bathtub-shaped hazard rate. Appropriate for populations followed from birth.

Figure 1: Types of hazard functions



Mean Value

- For a non-negative random variable T the **mean value** written

$$E(T) = \int_0^{\infty} tf(t)dt$$

can be shown to be:

$$E(T) = \int_0^{\infty} S(t)dt \qquad (7)$$

Mean Residual Life

- The **mean residual life** at time u denoted by $mrl(u)$.
- For individuals of age u, this parameter measures their expected remaining lifetime. It is defined as:

$$mrl(u) = E(T - u | T > u)$$

- For a continuous random variable it can be verified that

$$mrl(u) = \frac{\int_u^\infty S(t)dt}{S(u)} \tag{8}$$

Parametric Survival Distributions

- Several survival distributions exists for modeling survival data.
  - Exponential Distribution
  - Weibull Distribution

Exponential Distribution

- The p.d.f is given as:

$$f(t) = \lambda e^{\lambda t}$$

- The hazard function

$$h(t) = \lambda$$

- The cumulative hazard function is given as:

$$H(t) = \int_0^t = h(u)du = \int_0^t \lambda du = \lambda t|_0^t = \lambda t$$

- The survival function:

$$S(t) = \exp^{-H(t)} = e^{-\lambda t}$$

- The mean of an exponential distribution is given by:

$$E(T) = \int_0^\infty S(t)dt = \int_0^\infty e^{-\lambda t}dt = \frac{1}{\lambda}$$

- The median is the value of t that satisfies

$$0.5 = e^{-lambdat}$$

so that

$$t_{med} = \frac{\log 2}{\lambda}$$

Weibull Distribution

- The constant hazard of the exponential distribution makes it difficult to work with as it is not appropriate for describing lifetimes of humans or animals.
- The weibull distribution which has a decreasing and increasing hazard rate is more appropriate.
- The hazard function is given as:

$$h(t) = \alpha\lambda^{\alpha} t^{\alpha-1} \tag{9}$$

- The cumulative hazard and survival function are given respectively by:

$$H(t) = (\lambda t)^{\alpha}$$

```r
weibHaz <- function(x, shape, scale) dweibull(x, shape=shape,
scale=scale)/pweibull(x, shape=shape, scale=scale,
lower.tail=F)
curve(weibHaz(x, shape=1.5, scale=1/0.03), from=0, to=80,
ylab='Hazard', xlab='Time', col="red")
```

Maximum Likelihood Estimation

- If we know that a random variable T has an exponential distribution with parameter $\lambda = 0.03$ we can directly compute the probability that T exceeds a particular value.
- Suppose that we have a series of observations $t_1, t_2, ..., t_n$ from an exponential distribution with unknown parameter λ, then we can estimate the parameter λ via maximum likelihood estimation.
- Assuming that there is no censoring, the likelihood takes the general form

$$L(\lambda; t_1, t_2, ..., t_n) = f(t_1, \lambda) \times f(t_1, \lambda) \times ... \times f(t_n, \lambda) = \prod_{i=1}^{n} f(t_i, \lambda) \tag{10}$$

- If some observations are censored, we have to make adjustments.

- For an observation of an observed death, we put in the pdf.
- But for a right-censored observation we put in the survivor function indicating that the observations is known only to exceed a particular value.
- The likelihood in general takes the form:

$$L(\lambda; t_1, .., t_n) = \prod_{i=1}^{n} f(t_i, \lambda)^{\delta_i} S(t_i, \lambda)^{(1-\delta_i)} = \prod_{i=1}^{n} h(t_i, \lambda)^{\delta_i} S(t_i, \lambda)$$

- The expression means that when t_i is an observed death, the censoring indicator is $\delta_i = 1$ and we enter the pdf factor.
- When t_i is a censored observation we have $\delta_i = 0$ and we enter the survival function.
- For the case of exponential distribution we substitute the expressions for the pdf and survival functions and simplfy as follows:

$$L(\lambda) = \prod_{i=1}^{n} \left[\lambda e^{-\lambda t_i}\right]^{\delta_i} \left[e^{-\lambda t_i}\right]^{1-\delta_i} = \lambda^d e^{-\lambda V} \tag{11}$$

- We have the total number of deaths

$$d = \sum_{i=1}^{n} \delta_i$$

- and the total amount of time of patients on the study

$$V = \sum_{i=i}^{n} t_i$$

- We can maximize the log-likelihood as follows:

$$l(\lambda) = d \log \lambda - \lambda V$$

- Getting first derivative and equating to zero we get

$$\hat{\lambda} = \frac{d}{V}$$

Using Standard mathematical statistics theory, the inverse of the information matrix is approximately the variance of the m.l.e

$$Var(\hat{\lambda}) \approx I^{-1}(\lambda) = \frac{\lambda^2}{d} = \frac{d}{V^2}$$

Thank You!