Chapter 1

General Introduction

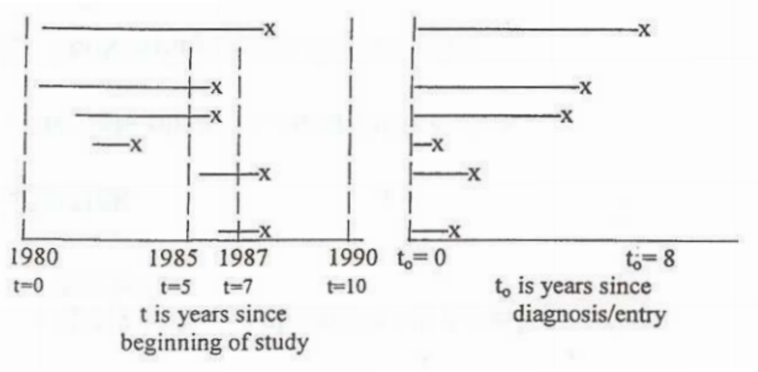
Fall, 2017 at WHU

1. What is Survival Analysis?

- Outcome variable:
  Time until an event occurs
- Time origin:
  Precisely defined, comparable across subjects

Example 1

1. date of randomization in clinical trial
2. date of enrollment in observational study
3. date of birth



Remark:

Typically not the same calendar date

Event:

- Typically death, disease incidence, relapse
- Usually assume at most one failure per individual
- If more than one failure per subject need advanced methods
- If more than one failure type need competing risks methods
- For this course, assume at most one failure and no competing risks

Time:

- Measured in years, months, days, etc.
- Positive, distribution typically skewed
- "failure time" or "survival time"

Our interests and goals:

- Characterize distribution of failure times
- Compare distributions of failure times between groups
- Association of explanatory variables and survival

2. Censoring

- If we know the exact failure time

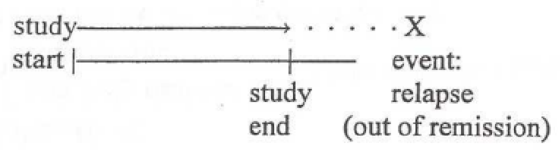
  $\Longrightarrow$ apply all the techniques we have learned in statistics.

- usually can not be obtained

  $\Longleftarrow$ incomplete observations of failure time

Censoring

Example 2



Source of censoring:

1. study end before any event happens
2. lost to follow-up
3. withdraw from the study

Example 3

- <u>Ex</u> Hypothetical data (+ denotes right censoring)

Treatment: $1, 1, 1, 1^+, 4^+, 5$
Control: $1^+, 2^+, 3, 3^+$ (in years)

1) Ignore +'s: $\frac{1+1+1+1+4+5}{6} = \frac{13}{6} = \bar{X}$
biased; define "better" estimator $> \bar{X}$

2) delete +'s:
inefficient, also possibly biased

Difficulties of censoring scheme:

1. ignore censoring, treat each subject's observed time as failure time
   $\Longrightarrow$
   - biased
   - underestimate

2. delete all subjects with censoring
   $\Longrightarrow$
   - inefficient
   - possible biased

Need methods allowing for censoring!

Survival Analysis

3. General Introduction

Some notations:

- T_i: the potential failure time
- C_i: the censoring time
- the observed time

$$\cancel{T_i} = \min(T_i, C_i)$$

- the censoring indicator

$$\delta_i = \begin{cases} 1, & \text{if failed;} \\ 0, & \text{if censored.} \end{cases}$$

- When $\Delta_i = 1$

$$X_i = T_i$$

$\Longrightarrow$ failure time is observed

- When $\Delta_i = 0$

$$X_i = C_i$$

$\Longrightarrow$ censoring time is observed

Remark:

$T \sim f(t,\theta)$ probability frequency or density function, $T_1, \cdots, T_n \overset{iid}{\sim} f(t,\theta)$.

$$L(\theta; T_1, \cdots, T_n) = \prod_{i=1}^{n} f(T_i, \theta)$$

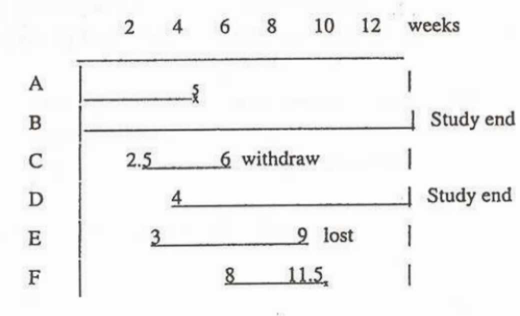
$$\ell(\theta; T_1, \cdots, T_n) = \log L(\theta) = \sum_{i=1}^{n} \log f(T_i, \theta)$$

$$U(\theta) = \frac{\partial \ell(\theta; T_1, \cdots, T_n)}{\partial \theta} = 0 \Longrightarrow \hat{\theta} = \hat{\theta}(T_1, \cdots, T_n)$$

Question:

$T_1, \cdots, T_n$ can not be observed!

Example 4



2 events
A,F

4 censored:
B,C,D,E

person	X_i Observed Time	δ_i Failed (1) Censored (0)
A	5	1
B	12	0
C	3.5	0
D	8	0
E	6	0
F	3.5	1

Type of data to be analyzed in survival analysis.

- Type I Censoring:

  Censoring time fixed in advance, known

Example 5

Study ends when a certain time point is reached

$$C_i = c = 1 \text{ year}$$

- Type II Censoring:

  Study ends when a certain number of failures occur

Example 6

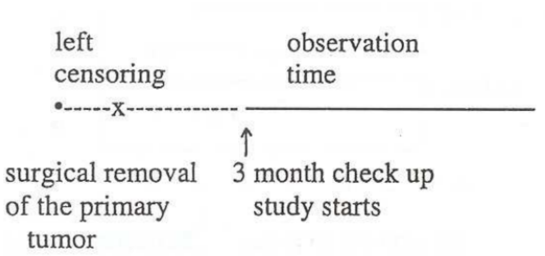
Study ends when 50 infections occur

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

- Left Censoring:
  Actual survival time is less than observed

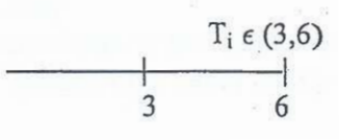
Example 7

Suppose event is recurrence of cancer
At first follow-up visit, cancer already recurred

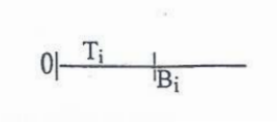


- Interval Censoring:

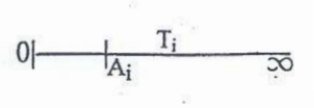
$$T_i \in (A_i, B_i)$$



$$\left.\begin{array}{l} A_i = 0 \\ B_i \neq \infty \end{array}\right. \Rightarrow \text{left censoring}$$



$$\left.\begin{array}{l} B_i = \infty \\ A_i \neq 0 \end{array}\right. \Rightarrow \text{right censoring}$$



Example 8

- In HIV incidence studies, endpoint of interest is time until HIV infection
- Tested for HIV periodically, eg, every 6 months
- Therefore, T only known to be between last negative and first positive test

4. Survival Function and Hazard Function

(1) Notations:

- T: potential failure time (≥ 0);
- C: potential censoring time;
- $X = \min(T, C)$: observed time;
- $\Delta = I(T \leq C)$: censoring indicator;

$$\Delta = \begin{cases} 1 & \text{failed} \\ 0 & \text{censored} = \begin{cases} \text{study ends} \\ \text{lost} \\ \text{withdraw} \end{cases} \end{cases}$$

(2) Survival Function

- Cumulative distribution function (cdf):

$$T \sim F(t) = P(T \leq t)$$

- Probability density function (pdf):

$$f(t) = \lim_{\Delta t \longrightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}$$

Definition 9 (Survival Function)

T is the subject's failure time.

$$S(t) = P(T > t)$$

is called the survival function of T.

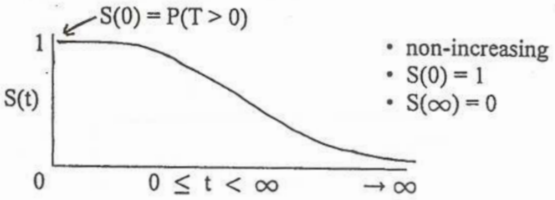
Remarks:

1. $S(t) = 1 - F(t)$
2. $S(t)$ is a probability
   - a subject survives longer than some specific time t
   - in survival analysis, we tend to focus on $S(t)$ in stand of $F(t)$.

Prop1:

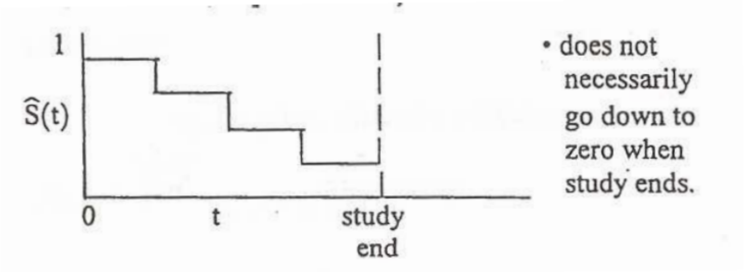
$$S(t) = P(T > t)$$

1. non-increasing;
2. $S(0) = 1$;
3. $S(\infty) = 0$.



Prop2:

In practice, $\hat{S}(t)$ is usually a step function.



(3) Hazard Function

Definition 10 (Hazard Function)

$$
\begin{aligned}
\lambda(t) &= \lim_{\Delta t \longrightarrow 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\
&= \lim_{\Delta t \longrightarrow 0^+} \frac{\text{Conditional probability}}{\text{Unit time}}
\end{aligned}
$$

Remark:

$\lambda(t)$ is the instantaneous potential per unit time for the event to occur given that the individual has survived up to time t.

Remarks:

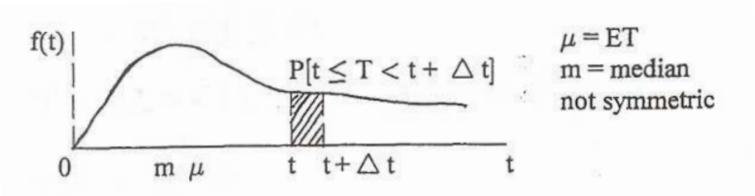
1. $\lambda(t)$ is not a probability but a rate
2. $\lambda(t)$ is an instantaneous potential given $T \geq t$

   conditional failure rate
3. $\lambda(t)$
   - $\lambda(t) \geq 0$
   - $\lambda(t)$ has no upper bound
4. $S(t)$
   - survive.
   - probability.
   - directly descriptive.

5. Survival Distributions

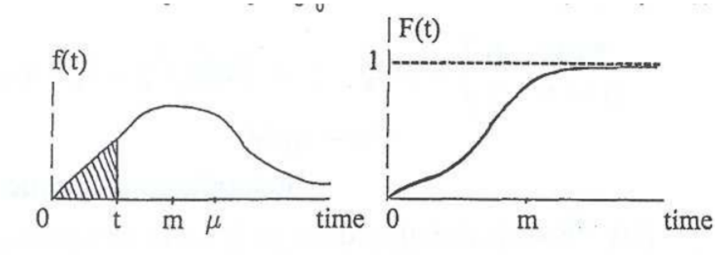
(1) probability density function (pdf)

$$f(t) = \lim_{\triangle t \to 0} \frac{P(t \leq T < t + \triangle t)}{\triangle t} = F'(t)$$



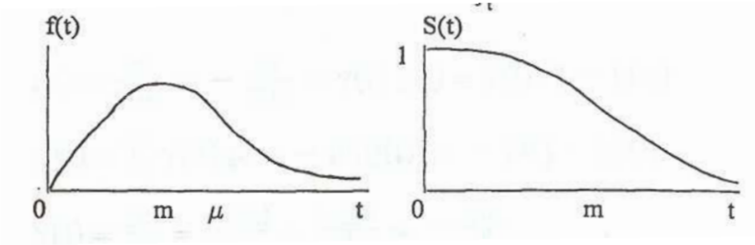
(2) cumulative distribution function (cdf)

$$F(t) = P(T \leq t) = \int_0^t f(s)ds = 1 - S(t)$$



(3) survival function

$$S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(s)ds$$



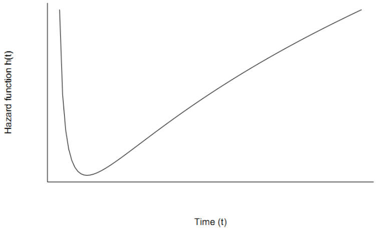
(4) hazard function

$$\lambda(t) = \lim_{\triangle t \to 0} \frac{1}{\triangle t} P(t \le T < t + \triangle t | T \ge t) = \frac{f(t)}{S(t)}$$

Prop1: $\lambda(t) = \frac{f(t)}{S(t)}$

Prop2: $S(t) = \exp\{-\int_0^t \lambda(s)ds\}$

Prop3: $\lambda(t) = [-\log(S(t))]'_t$

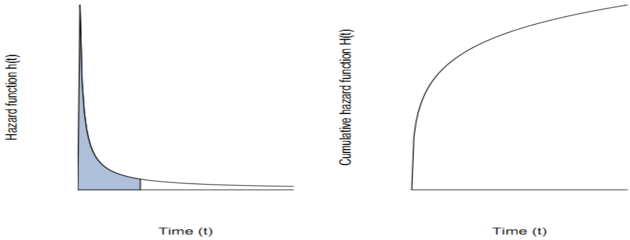


(5) cumulative hazard function

$$\Lambda(t) = \int_0^t \lambda(s)\, ds$$

Prop1: $\Lambda(t) = -\log S(t)$

Prop2: $S(t) = e^{-\Lambda(t)}$



Summary:

1. $F(t) = \int_0^t f(s)\,ds = 1 - S(t)$

2. $S(t) = 1 - F(t) = \int_t^\infty f(s)\,ds = e^{-\int_0^t \lambda(s)ds} = e^{-\Lambda(t)}$

3. $\lambda(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)} = -\frac{S'(t)}{S(t)} = [-\log S(t)]_t'$

4. $\Lambda(t) = \int_0^t \lambda(s)ds = -\log S(t) = -\log(1 - F(t))$

5. $f(t) = F'(t) = -S'(t) = \lambda(t)S(t) = \lambda(t)(1 - F(t))$

(6) discrete distribution

- $$f_i = f(a_i), \qquad i = 1, \cdots, p$$

- $$T \sim \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ f_1 & f_2 & \cdots & f_p \end{pmatrix}.$$

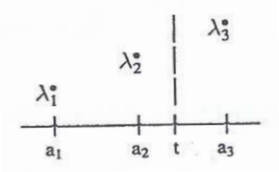
-

$$
\begin{aligned}
\lambda_i &= \lim_{\triangle x \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\
&= P(T = a_i | T \geq a_i) \\
&= \frac{P(T = a_i)}{P(T \geq a_i)} \\
&= \frac{f_i}{S(a_i)} \\
&= \frac{f_i}{f_i + f_{i+1} + \cdots + f_p}.
\end{aligned}
$$

- $$\Lambda(t) = \sum_{j:\ a_j \le t} \lambda_j$$

Example 11

$$\Lambda(t) = \lambda_1 + \lambda_2$$

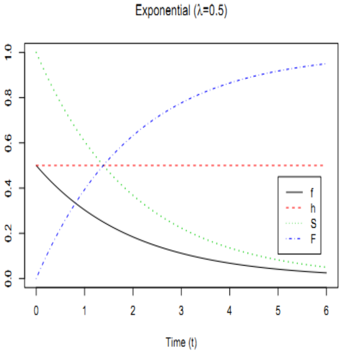


$$\begin{cases} \text{grouping of continuous data(imprecise measurements)} \\ \text{discrete time scale} \end{cases}$$

6. Some Special Distributions

(1) exponential distribution ($\rho > 0$)

- $f(t) = \rho e^{-\rho t}$ $\quad E(T) = \frac{1}{\rho}$
- $S(t) = e^{-\rho t}$ $\quad F(t) = (1 - e^{-\rho t})$
- $\lambda(t) = \rho$ (constant hazard) $\quad \Lambda(t) = \rho t$



Prop1: lack of memory;

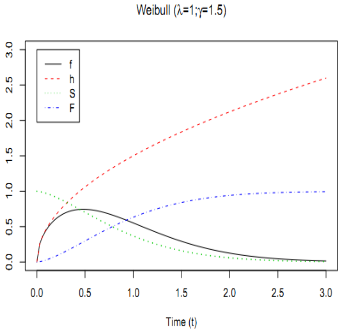
Prop2: coefficency of variation $= \frac{\text{sd}}{\text{mean}} = 1$

Prop3: empirical check of the data plot $\log\{S(t)\}$ vs t

Prop4: If T has an arbitrary continuous distribution, the $\Lambda(T)$ has an exponential distribution with unit parameter, that is, $\Lambda(T) \sim E(1)$.

(2) Weibull distribution ($\rho > 0, \ \gamma > 0$)

- $f(t) = \rho\gamma t^{\gamma-1}e^{-\rho t^{\gamma}}$
- $S(t) = \exp(-\rho t^{\gamma})$ $\qquad F(t) = 1 - \exp(-\rho t^{\gamma})$
- $\lambda(t) = \rho\gamma t^{\gamma-1}$ $\qquad \Lambda(t) = \rho t^{\gamma}$



Prop1: important generalization of the exponential distribution, allows for a power dependence of the hazard on time.

Prop2: $\lambda(t)$ is monotone decreasing for $\gamma < 1$; $\lambda(t)$ is monotone increasing for $\gamma > 1$; reduces to the constant exponential hazard if $\gamma = 1$.

Prop3: empirical check of the data

- plot $\log(-\log \hat{S}(t))$ vs $\log(t)$
- plot should give approximately a straight line with slope γ and intercept $\log \rho$.

(3) Gamma distribution ($\rho,\ \kappa > 0$)

$$f(t) = \rho(\rho t)^{\kappa-1}e^{\rho t}/\Gamma(\kappa)$$

where $\Gamma(t) = \int_0^{+\infty} x^{t-1}e^{-x}\,dx$

- $E(T) = \frac{\kappa}{\rho}$
- $S(t) = 1 - F(t)$
- $\lambda(t) = \frac{f(t)}{S(t)}$

Prop1: $\lambda(t)$ is monotone increasing from 0 if $\kappa > 1$; monotone decreasing from ∞ if $\kappa < 1$; in either case approaches ρ as $t \to \infty$.

Prop2: The gamma distribution reduces to the exponential distribution if $\kappa = 1$.

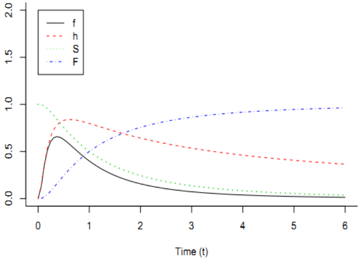
Prop3: The gamma distribution with integer κ can be derived as the distribution of the waiting time to the κth emission from a Poisson source with intensity parameter ρ.

Prop4: $\sum_{i=1}^{\kappa} T_i \sim \Gamma(\rho, \kappa)$.

(4) log-normal distribution

- $f(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp(-\frac{(\log t - \mu)^2}{2\sigma^2})$
- $F(t) = \Phi(\frac{(\log(t) - \mu)}{\sigma})$
- $S(t) = 1 - \Phi(\frac{(\log(t) - \mu)}{\sigma})$
- $\lambda(t) = \frac{f(t)}{S(t)}$

Log-normal distribution ($\mu = 0$, $\sigma = 1$)



Prop1: $\log(T) \sim N(\mu, \sigma^2)$

Prop2: simple to apply if no censoring.

Prop3: sensitive to the small failure times

Prop4: log-logistic distribution provides a good approximation to the log-normal distribution (may frequently be a preferable survival time model)

(5) log-logistic distribution

logistic density with location parameter ν and scale τ:

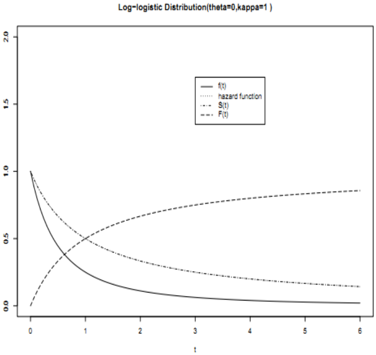
$$f(x) = \tau^{-1} e^{\frac{x-\nu}{\tau}} / (1 + e^{\frac{x-\nu}{\tau}})^2$$

let $\theta = -\frac{\nu}{\tau}, \kappa = \frac{1}{\tau}$, then

$$f(x) = \frac{\kappa e^{\theta + \kappa x}}{(1 + e^{\theta + \kappa x})^2}$$

Next, do a change of variable from x to $\log(t)$, then

- $f(t) = \frac{e^{\theta}\kappa t^{\kappa-1}}{(1+e^{\theta}t^{\kappa})^2}$
- $S(t) = \frac{1}{1+e^{\theta}t^{\kappa}}$
- $\lambda(t) = \frac{e^{\theta}\kappa t^{\kappa-1}}{1+e^{\theta}t^{\kappa}}$



Prop1: $\log(T) \sim$ logistic distribution

Prop2: relatively simple explicit forms for $S(t),\ f(t),\ \lambda(t)$ (vs.log-normal)

Prop3: $\lambda(t)$ has a single maximum if $\kappa > 1$; is decreasing if $\kappa < 1$ (from ∞); is decreasing if $\kappa = 1$ (from e^{θ}).

Prop4: more convenient in handling censored data than the log-normal distribution.

Prop5: provides a good approximation to the log-normal except in the extreme tails.

(6) Gompertz distribution

Two parameters θ and λ where $\lambda > 0$

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

- exponential special case $\theta = 0$
- log hazard linear in t; in contrast, for Weibull log hazard linear in $\log t$
- hazard increases or decreases monotonically
- survival function $S(t) = \exp(\frac{\lambda}{\theta}(1 - e^{\theta t}))$ if $\theta \neq 0$
- Gompertz-Makeham distribution (three parameters) $\lambda(t) = \lambda_0 + \lambda_1 e^{\theta t}$

Utility of parametric modeling

1. explicit forms for $f(t),\ \lambda(t),\ S(t)$
2. convenient for statistical inference; not computationally intensive
3. usually more efficient than non-parametric inference when model is correct
4. permit extrapolation (e.g., for t small or large)

Disadvantage

1. lack of robustness to violation of modeling assumptions