

Survival Analysis

1 Introduction

Definition: A failure time (survival time, lifetime), T , is a nonnegative-valued random variable.

For most of the applications, the value of T is the time from a certain event to a failure event. For example,

- in a clinical trial, time from start of treatment to a failure event
- Time from birth to death = age at death
- to study an infectious disease, time from onset of infection to onset of disease
- to study a genetic disease, time from birth to onset of a disease = onset age

Definition: Cumulative distribution function

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

Definition: Survival function $S(t)$.

$$S(t) = Pr(T > t) = 1 - P(T \leq t)$$

Characteristics of $S(t)$:

- $S(t) = 1$ if $t < 0$
- $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$
- $S(t)$ is non-increasing in t

Density: density function $f(t)$

- if T is discrete then $f(t) = P(T = t)$.
- b) If T is (absolutely) continuous, the density function is

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

Definition. Hazard function $\lambda(t)$ is

- If T is discrete

$$\begin{aligned}\lambda(t) &= P(T = t | T \geq t) = \frac{P(T = t)}{P(T \geq t)} \\ &= \frac{f(t)}{S(t-)}\end{aligned}$$

If $t_1 < t_2 < t_3 < \dots$ are the possible values of T and $t_j \leq t < t_{j+1}$, then

$$S(t) = \prod_{i=1}^j (1 - \lambda(t_i))$$

this because

$$\begin{aligned}S(t) &= P(T \geq t) \\ &= P(T \geq t_{j+1}) \\ &= \frac{P(T \geq t_2) P(T \geq t_3)}{P(T \geq t_1) P(T \geq t_2)} \cdots \frac{P(T \geq t_{j+1})}{P(T \geq t_j)} \\ &= \left(1 - \frac{P(T = t_1)}{P(T \geq t_1)}\right) \left(1 - \frac{P(T = t_2)}{P(T \geq t_2)}\right) \cdots \left(1 - \frac{P(T = t_j)}{P(T \geq t_j)}\right) \\ &= \prod_{i=1}^j (1 - \lambda(t_i))\end{aligned}$$

- If T is (absolutely) continuous

$$\begin{aligned}\lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \text{Instantaneous failure rate at } t \text{ given survival up to } t \\ &= \frac{f(t)}{S(t)}\end{aligned}$$

Definition: Cumulative hazard function (chf) $\Lambda(t)$

- If T is discrete

$$\Lambda(t) = \sum_{t_i \leq t} \lambda(t_i)$$

- It T is (absolutely) continuous,

$$\Lambda(t) = \int_0^t \lambda(u) du$$

therefore

$$\lambda(t) = \frac{d\Lambda(t)}{dt}.$$

Also

$$\begin{aligned} \Lambda(t) &= \int_0^t \lambda(u) du \\ &= \int_0^t \frac{-dS(u)/du}{S(u)} \\ &= -\log(S(t)) \end{aligned}$$

This implies that

$$S(t) = e^{-\Lambda(t)} = e^{-\int_0^t \lambda(u) du}.$$

in addition, since $S(\infty) = 0$,

$$\int_0^t \lambda(u) du = \infty.$$

- Examples

- Suppose T has a survival function given by

$$S(t) = \begin{cases} e^{-\lambda t}, & t > 0 \\ 0, & t \leq 0 \end{cases}$$

In this case T has exponential distribution with parameter λ and

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

- Suppose T has a survival function given by

$$S(t) = \begin{cases} e^{-(\lambda t)^\alpha}, & t > 0 \\ 0, & t \leq 0 \end{cases}$$

$\alpha > 0$ and $\lambda > 0$. In this case T has a Weibull distribution with parameters λ and α . and

$$\lambda(t) = \frac{f(t)}{S(t)} = \lambda^\alpha \alpha t^{\alpha-1} = \alpha \lambda (\lambda t)^{\alpha-1}.$$

If $\alpha = 1$, we have exponential (λ).

Type I censoring: Let T_1, T_2, \dots, T_n be independent, identically distributed random variables. Assume that t_c is some fixed time. Instead of observing T_1, T_2, \dots, T_n , we only observe

$$X_i = \min(T_i, t_c) = \begin{cases} X_i, & \text{if } T_i \leq t_c \\ t_c, & \text{if } T_i > t_c \end{cases}.$$

Type II Censoring: Let $r < n$ and let $T_{(1)} < T_{(2)} < \dots < T_{(n)}$ be the order statistics of T_1, T_2, \dots, T_n . Suppose observations cease after the r th failure and we only observe $T_{(1)}, T_{(2)}, \dots, T_{(r)}$

In Type II Censoring model we have instead t_c the random time $T_{(r)}$

Both the Type I and the Type II censoring arise in engineering applications.

Random Censoring: Let C_1, C_2, \dots, C_n be iid random variables with cdf G . Here C_i is the censoring time associated with T_i . We observe the pairs

$$(X_1, \delta_1), (X_2, \delta_2), \dots, (X_n, \delta_n)$$

where

$$X_i = \min(T_i, C_i) = \begin{cases} X_i, & \text{if } T_i \leq C_i \\ C_i, & \text{if } T_i > C_i \end{cases}.$$

and

$$\delta_i = \begin{cases} 1, & \text{if } T_i \leq C_i \\ 0, & \text{if } T_i > C_i \end{cases}.$$

Example: A set of observed survival data is

x_i	25	18	17	22	27
δ_i	1	0	1	0	1

The data can also be presented as

$$25 \quad 18^+ \quad 17 \quad 22^+ \quad 27$$

Notice that

$$\begin{aligned}
S_X(x) = P(X > x) &= P(\min(T, C) > x) \\
&= P(T > x, C > x) \\
&= P(T > x)P(C > x) \\
&= S(x)\bar{G}(x) \leq S(x)
\end{aligned}$$

where $\bar{G} = 1 - G$ is the survival function corresponding to C .

2 Estimation of S:

- Complete Failure Times: Nonparametric Models Recall that

$$\begin{aligned}
S(t) &= P(T > t) \\
&= \text{population fraction surviving beyond } t
\end{aligned}$$

The set of the complete data t_1, t_2, \dots, t_n reflects the structure of population failure times. Thus, we estimate $S(t)$ by the sample fraction surviving beyond t :

$$\hat{S}(t) = \frac{\#t_i > t}{n} = \frac{1}{n} \sum_{i=1}^n I(t_i > t).$$

\hat{S} is also called the empirical survival distribution. How to derive confidence interval for $S(t)$?

We have

$$\frac{\sqrt{n}[\hat{S}(t) - S(t)]}{\sqrt{\hat{S}(t)(1 - \hat{S}(t))}} \xrightarrow{d} N(0, 1)$$

that is

$$\frac{\sqrt{n}[\hat{S}(t) - S(t)]}{\sqrt{\hat{S}(t)(1 - \hat{S}(t))}} \overset{App}{\approx} N(0, 1)$$

or

$$\lim_{n \rightarrow \infty} P \left(\frac{\sqrt{n}[\hat{S}(t) - S(t)]}{\sqrt{\hat{S}(t)(1 - \hat{S}(t))}} \leq x \right) = \Phi(x).$$

An approximate $100(1 - \alpha)\%$ confidence interval for $S(t)$ is

$$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\hat{S}(t)(1 - \hat{S}(t))/n}$$

What do we do when we have right censoring?

Kaplan-Meier Estimator: The Kaplan-Meier estimator is a nonparametric estimator for the survival function S . Consider now either random censoring or type I censoring. The data are

$$(x_1, \delta_1), (x_2, \delta_2), \dots, (x_n, \delta_n)$$

Let $y_{(1)} < y_{(2)} < \dots < y_{(k)}, k \leq n$, be the distinct, uncensored and ordered failure times.

Example. Data: 3, 2, 0, 1, 5⁺, 3, 5 then

$$(y_{(1)}, y_{(2)}, y_{(3)}, y_{(4)}, y_{(5)}) = (0, 1, 2, 3, 5)$$

Suppose $y_{(i-1)} \leq t < y_{(i)}$. A principle of nonparametric estimation of S is to assign positive probability to and only to uncensored failure time. Therefore, we try to estimate

$$S(t) \approx \frac{P(T \geq y_{(2)})}{P(T \geq y_{(1)})} \frac{P(T \geq y_{(3)})}{P(T \geq y_{(2)})} \cdots \frac{P(T \geq y_{(i)})}{P(T \geq y_{(i-1)})}$$

How to estimate $S(t)$? Define

$$\begin{aligned} R_{(j)} &= \{y_k : y_k \geq y_{(j)}\} \\ d_{(j)} &= \# \text{ of failures at } y_{(j)} \\ N_{(j)} &= \# \text{ of individuals at risk at } y_{(j)} = \text{cardinal of } R_{(j)} \end{aligned}$$

Example: Using the previous data, we have

$$\begin{aligned} N_{(1)} &= 7, N_{(2)} = 6, N_{(3)} = 4, N_{(4)} = 2 \\ d_{(1)} &= 1, d_{(2)} = 1, d_{(3)} = 2, d_{(4)} = 1 \end{aligned}$$

We estimate

$$\frac{P(T \geq y_{(j+1)})}{P(T \geq y_{(j)})}$$

by

$$\frac{N_{(j)} - d_{(j)}}{N_{(j)}} = 1 - \frac{d_{(j)}}{N_{(j)}}$$

$j = 1, 2, \dots, i - 1$. The Kaplan-Meier estimator of $S(t)$ is that

$$\begin{aligned}\hat{S}(t) &= \left(1 - \frac{d_{(1)}}{N_{(1)}}\right) \left(1 - \frac{d_{(2)}}{N_{(2)}}\right) \dots \left(1 - \frac{d_{(i-1)}}{N_{(i-1)}}\right) \\ &= \prod_{y_{(j)} \leq t} \left(1 - \frac{d_{(j)}}{N_{(j)}}\right)\end{aligned}$$

Example:

Uncensored Times	0	1	3	5
$d_{(i)}$	1	2	2	1
$N_{(i)}$	7	6	4	2

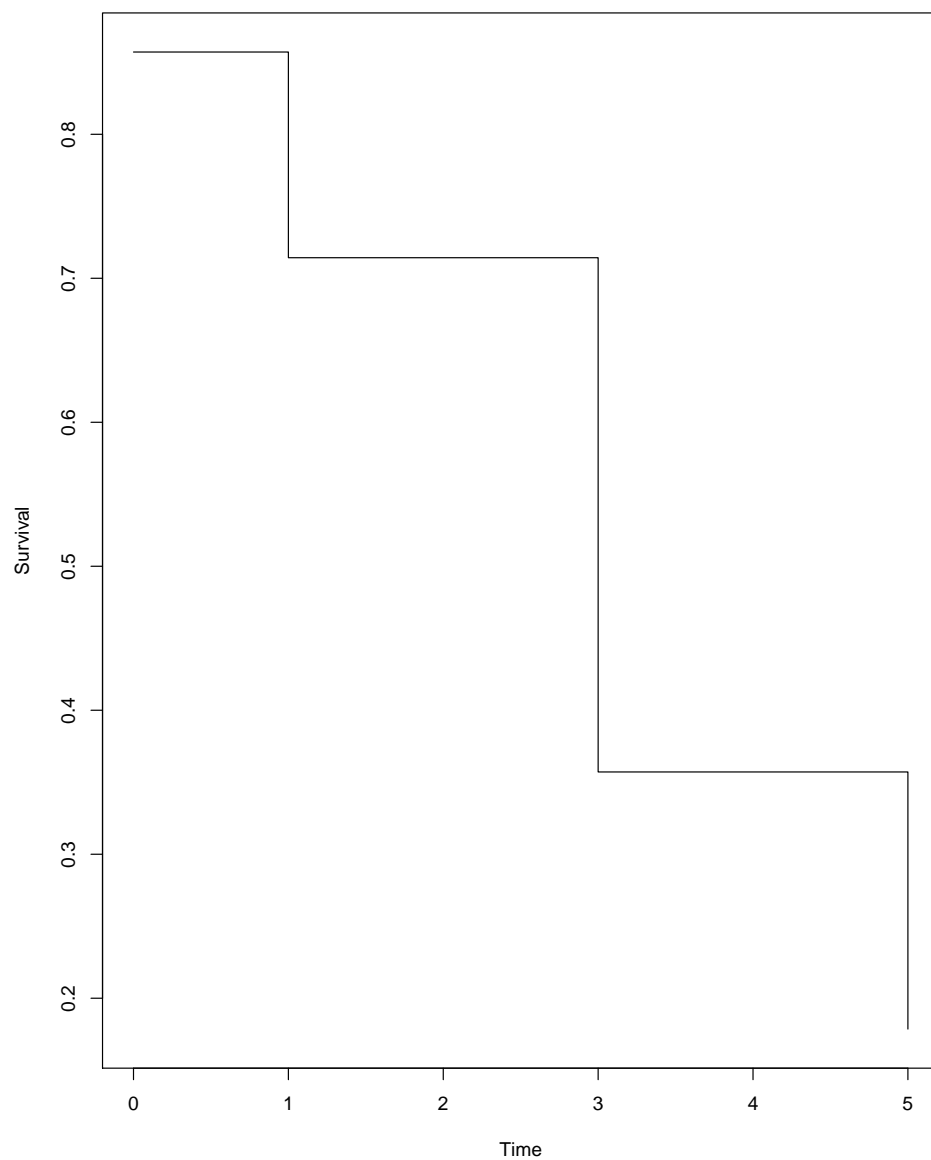
$$\hat{S}(0) = \left(1 - \frac{1}{7}\right) = \frac{6}{7} = 0.86$$

$$\hat{S}(1) = \left(1 - \frac{1}{7}\right) \left(1 - \frac{1}{6}\right) = \frac{5}{7} = 0.71$$

$$\hat{S}(3) = \left(1 - \frac{1}{7}\right) \left(1 - \frac{1}{6}\right) \left(1 - \frac{2}{4}\right) = \frac{5}{14} = 0.36$$

$$\hat{S}(5) = \left(1 - \frac{1}{7}\right) \left(1 - \frac{1}{6}\right) \left(1 - \frac{2}{4}\right) \left(1 - \frac{1}{2}\right) = \frac{5}{28} = 0.18$$

```
> library(survival)
> time<-c(3, 2,0, 1,5, 3, 5)
> delta<-c(1,0,1, 1, 0, 1, 1)
> library(survival)
> km<-survfit(Surv(time, delta)~1,type="kaplan-meier")
> km$urv
[1] 0.8571429 0.7142857 0.7142857 0.3571429 0.1785714
> plot(km$time,km$urv, type="s",xlab="Time",ylab="Survival")
```



It turns out that

$$\sqrt{n}[\hat{S}(t) - S(t)] \xrightarrow{d} N(0, \sigma^2(t))$$

where

$$\sigma^2(t) = S(t)^2 \int_0^t \frac{d\Lambda(u)}{\pi(u)}$$

where $\pi(u) = S(u)\bar{G}(u)$

An estimate of $\sigma^2(t)$ is

$$\hat{\sigma}^2(t) = \hat{S}(t)^2 \sum_{y_{(j)} \leq t} \frac{d_{(j)}}{N_{(j)}(N_{(j)} - d_{(j)})}$$

An approximate $100(1 - \alpha)\%$ confidence interval for $S(t)$ is

$$\hat{S}(t) \pm z_{\alpha/2} \hat{\sigma}(t)$$

```
> summary(km)
```

```
Call: survfit(formula = Surv(time, delta) ~ 1, type = "kaplan-meier")
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0	7	1	0.857	0.132	0.6334	1
1	6	1	0.714	0.171	0.4471	1
3	4	2	0.357	0.198	0.1205	1
5	2	1	0.179	0.160	0.0307	1

Comparing two survival distributions: the logrank statistic

Let $T_1^0 < T_2^0 < \dots < T_L^0$ denote the ordered observed distinct failure times in the sample made by combining the two groups. Let d_{ik} and N_{ik} , $k = 1, 2, \dots, L$ denote the number of failures and the number at risk, respectively, in the i th sample at time T_k^0 . Let d_k and N_k the corresponding values in the combined sample. The data at T_k^0 can be summarized as follows

	Sample		
Failure	Group 1	Group 2	Total
Yes	d_{1k}	d_{2k}	d_k
No	$N_{1k} - d_{1k}$	$N_{2k} - d_{2k}$	$N_k - d_k$
Total	N_{1k}	N_{2k}	N_k

Given N_{ik} , d_{ik} has a binomial distribution with the number of trials N_{ik} and under the hypothesis of common failure rate λ is the two groups, approximate probability

Under H_0 , d_{1k} given d_k has a hypergeometric with mean and variance

$$E_{1k} = d_k \frac{N_{1k}}{N_k}$$

$$V_{1k} = d_k \frac{N_{1k} N_{2k}}{N_k^2} \frac{N_k - d_k}{N_k - 1}$$

Given the margins in each of the L tables at the observed death time

$$\{d_{11} - E_{11}, d_{12} - E_{12}, \dots, d_{1L} - E_{1L}\}$$

is a vector of observed-minus-conditionally-expected number of failures across the observed failure times, and if we assume these difference are independent

$$Q = \frac{\sum_{k=1}^L (d_{1k} - E_{1k})}{\sqrt{\sum_{k=1}^L V_{1k}}}$$

has approximately a standard normal distribution. So Q^2 has approximately χ_1^2 .

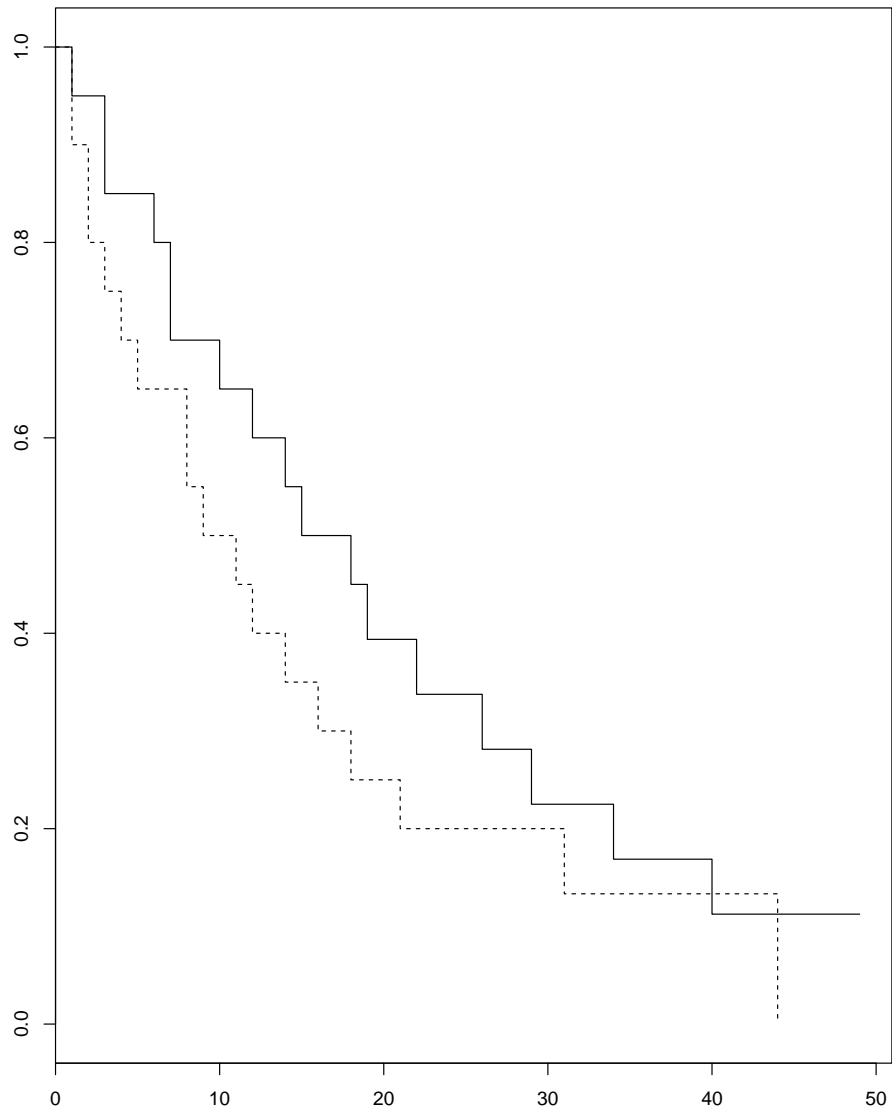
Example: The data below shows remission times, in weeks, for leukemia patients given two types of treatments. In the study 20 patients were given treatment A and 20 treatment B. Starred observation are censoring time.

Treatment A : 1, 3, 3, 6, 7, 7, 10, 12, 14, 15, 18, 19, 22, 26, 18+, 29, 34, 40, 48+, 49.

Treatment B : 1, 1, 2, 2, 3, 4, 5, 8, 8, 9, 11, 12, 14, 16, 18, 21, 27+, 31, 38+, 44.

First we create estimate the survival functions. To do this in R we use the package survival

```
> head(data)
  time status group
1    1      1     1
2    3      1     1
3    3      1     1
4    6      1     1
5    7      1     1
> library(survival)
> fit<-survfit(Surv(time,status)~group)
> plot(fit,lty=1:2)
```



To implement the test we use

```
surdiff(Surv(time,status) ~ group, rho=0)

> survdiff(Surv(time,status)~group, rho=0)
Call:
survdiff(formula = Surv(time, status) ~ group, rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
group=1	20	17	20.2	0.506	1.24
group=2	20	18	14.8	0.690	1.24

Chisq= 1.2 on 1 degrees of freedom, p= 0.265

The p-value is 0.265, we fail to reject the null hypothesis that the two survival functions are equal.

Example (in detail)

Treatment A : 3, 5, 7, 9+, 18
 Treatment B : 12, 19, 20, 20+, 33+

Here $k = 7$ and $T_1^0 = 2, T_2^0 = 5, T_3^0 = 7, T_4^0 = 12, T_5^0 = 18, T_6^0 = 19, T_7^0 = 19$.
 at $T_1^0 = 3$, we need to look at the table

Sample			
Failure	Group 1	Group 2	Total
Yes	1	0	1
No	4	5	9
Total	5	5	10

$d_1 = 1, N_{11} = 5$ and $N_1 = 10$. this gives $E_{11} = 1/2$ and $V_{11} = 1/4$
 At $T_1^0 = 5$, we need to look at the table

Sample			
Failure	Group 1	Group 2	Total
Yes	1	0	1
No	3	5	8
Total	4	5	9

$d_2 = 1, N_{12} = 4$ and $N_2 = 9$. this gives $E_{12} = 4/9$ and $V_{12} = 20/81$
 at $T_3^0 = 7$, we need to look at the table

Sample			
Failure	Group 1	Group 2	Total
Yes	1	0	1
No	2	5	7
Total	3	5	8

$d_3 = 1, N_{13} = 3$ and $N_3 = 8$. this gives $E_{13} = 3/8$ and $V_{13} = 15/64$
at $T_4^0 = 12$, we need to look at the table

Sample			
Failure	Group 1	Group 2	Total
Yes	0	1	1
No	1	4	5
Total	1	5	6

$d_4 = 1, N_{14} = 1$ and $N_4 = 6$. this gives $E_{14} = 1/6$ and $V_{14} = 5/36$
at $T_5^0 = 18$, we need to look at the table

Sample			
Failure	Group 1	Group 2	Total
Yes	1	0	1
No	0	4	4
Total	1	4	5

$d_5 = 1, N_{15} = 1$ and $N_5 = 5$. this gives $E_{15} = 1/5$ and $V_{15} = 4/25$
at $T_6^0 = 19$, we need to look at the table

Sample			
Failure	Group 1	Group 2	Total
Yes	0	1	1
No	0	3	3
Total	0	4	4

$d_6 = 1, N_{16} = 0$ and $N_6 = 4$. this gives $E_{16} = 0$ and $V_{16} = 0$
at $T_7^0 = 20$, we need to look at the table

Sample			
Failure	Group 1	Group 2	Total
Yes	0	1	1
No	0	2	2
Total	0	3	3

$d_7 = 1, N_{17} = 0$ and $N_6 = 3$. this gives $E_{17} = 0$ and $V_{17} = 0$

The log-rank statistics is

$$Q = \frac{\sum_{k=1}^L (d_{1k} - E_{1k})}{\sqrt{\sum_{k=1}^L V_{1k}}} = \frac{(1 - 1/2) + (1 - 4/8) + (1 - 3/8) + (0 - 1/6) + (1 - 1/5) + (0 - 0) + (0 - 0)}{\sqrt{1/4 + 20/81 + 15/64 + 5/36 + 4/25}} = 2.26$$

and $Q^2 = 5.108$ the p-value $= 1 - pchisq(5.108, 1) = 0.02381576$. **Generalizations** Use a weighted sum with weights w_1, w_2, \dots, w_k . The test statistic is

$$Q^2(w_1, w_2, \dots, w_k) = \frac{[\sum_{\ell=1}^k w_{\ell}(d_{1\ell} - E_{1\ell})]^2}{\sum_{\ell=1}^k w_{\ell}^2 V_{\ell}}$$

Note that if

- $w_1 = w_2 = \dots = w_k = 1$ we get the logrank test
- $w_{\ell} = n_{\ell}, \ell = 1, 2, \dots, \ell$, we get the Gehan test statistics
- $w_{\ell} = (n_{\ell})^{\alpha}, \ell = 1, 2, \dots, \ell, \alpha \in [0, 1]$, we get the Tarone-Ware test statistics

Regression Approach to Survival Analysis

In the presence of covariates, the standard linear regression model formulation is not appropriate for survival time, due to censoring and the skewed nature of the distributions. Before we present the model, we review some results on the types of test that we use in this part.

Cox Proportional Hazard Model (phm)

Let $\mathbf{x} = (x_1, x_2, \dots, x_p)$ be a vector of covariates and let T be time until failure (failure time)

Define the hazard function for a given individual by

$$\lambda(t, \mathbf{x}) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}$$

- $\lambda_0(t)$ is some baseline hazard function.
- $\beta_1, \beta_2, \dots, \beta_p$ are coefficients and they do not include the intercept (the intercept is absorbed in $\lambda_0(t)$)
- $\lambda_0(t)$ does not need to be specified in order to carry out the analysis
- The exponential guarantees that $\lambda(t)$ is positive for any $\beta_1, \beta_2, \dots, \beta_p$
- The beauty of this model, as observed by Cox, is that if you use a model of this form, and you are interested in the effects of the covariates on survival, then you do not need to specify the form of $\lambda_0(t)$.
- Even without doing so you may estimate β
- The Cox phm is thus called a semi-parametric model, as some assumptions are made but no form is pre-specified for $\lambda_0(t)$.
- To see why it is called the phm, suppose $p = 1$ and consider two individuals with covariates x_1 and x_2 . Then the ratio of their failure rates (or hazard rates) at time t is

$$\frac{\lambda(t, x_1)}{\lambda(t, x_2)} = \frac{e^{\beta_1 x_1}}{e^{\beta_1 x_2}} = e^{\beta_1(x_1 - x_2)}$$

that is

$$\lambda(t, x_1) \propto \lambda(t, x_2)$$

- The hazards are proportional to each other and do not depend on time. In particular, the hazard for the individual with covariate x_1 is $e^{\beta_1(x_1 - x_2)}$ times that of the individual with covariate x_2
- The term $e^{\beta_1(x_1 - x_2)}$ is called the hazard ratio comparing x_1 to x_2 .
- If $\beta_1 = 0$ then the hazard ratio for that covariate is equal to 1, i.e. that covariate doesn't affect survival. Thus we can use the notion of hazard ratios to test if covariates influence survival.

- The hazard ratio also tells us how much more likely one individual is to die than another at any particular point in time.
- If the hazard ratio comparing men to women were 2, say, it would mean that, at any instant in time, men are twice as likely to die than women.

We need to estimate the beta's in order to assess the effect of the covariates on the survival time.

Let t_1, t_2, \dots, t_n be the survival time for n individuals and $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be the corresponding covariates. Let

$$Y_i(t) = \begin{cases} 1, & \text{if the } i\text{th person is alive at time } t \\ 0, & \text{otherwise} \end{cases}$$

Suppose a death is observed at time t , the conditional probability that it is subject j is

$$\begin{aligned} L_j(\beta_1, \beta_2, \dots, \beta_p) &= \frac{\sum_{\ell=1}^p x_{\ell j} \beta_j}{\sum_{i=1}^n Y_i(t) \lambda_0(t) e^{\sum_{\ell=1}^p x_{\ell i} \beta_{\ell}}} \\ &= \frac{\sum_{\ell=1}^p x_{\ell j} \beta_j}{\sum_{i=1}^n Y_i(t) e^{\sum_{\ell=1}^p x_{\ell i} \beta_{\ell}}} \end{aligned}$$

and this does not depend on $\lambda_0(t)$.

Notice that

- The model depends only on the ranks of the survival times. Therefore, it is nonparametric in nature
- The parameters may be sensitive to outliers in the covariates

The partial likelihood is defined as

$$L(\beta_1, \beta_2, \dots, \beta_p) = \prod_{i=1}^n L_j(\beta_1, \beta_2, \dots, \beta_p)$$

The estimates of $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ of $\beta_1, \beta_2, \dots, \beta_p$ are obtained by maximizing $L(\beta_1, \beta_2, \dots, \beta_p)$

To test $H_0 : \beta_\ell = \beta_{\ell 0}, \ell = 1, 2, \dots, p$ we can use one of the following test statistics

1. Likelihood ratio test

$$LR = -2 \ln \frac{L(\beta_{10}, \beta_{20}, \dots, \beta_{p0})}{L(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)}$$

and we reject H_0 if $LR > \chi_p^2(\alpha)$

2. Wald Statistic

$$W = (\hat{\beta} - \beta_0)^T \hat{\Sigma}^{-1} (\hat{\beta} - \beta_0)$$

where $\hat{\Sigma}$ is an estimate of the covariance of $\hat{\beta}$ and is given by

$$\hat{\Sigma} = I^{-1}(\hat{\beta})$$

where

$$I(\hat{\beta}) = -\frac{\partial^2}{\partial \beta^2} \ln L(\beta)$$

Under H_0 , W has approximately χ_p^2 and we reject H_0 if $W > \chi_p^2(\alpha)$. For individual hypotheses, say we want to test $H_0 : \beta_\ell = \beta_{\ell 0}$ against $H_a : \beta_\ell \neq \beta_{\ell 0}$, we use

$$Z = \frac{\hat{\beta}_\ell - \beta_{\ell 0}}{SE(\hat{\beta}_\ell)}$$

and reject H_0 if $|Z| > Z_{\alpha/2}$.

3. Rao's Score Test

$$RS = U'(\beta_0) I(\beta_0) U(\beta_0)$$

where

$$U(\beta_0) = \frac{\partial}{\partial \beta} \ln L(\beta_0)$$

Under H_0 , RS has approximately χ_p^2 and we reject H_0 if $S > \chi_p^2(\alpha)$.

```

Example (cont)
x<-c(rep("1",20),rep("2",20))
> fitphm<-coxph(Surv(time,status)~factor(x))
> fitphm
Call:
coxph(formula = Surv(time, status) ~ factor(x))

```

	coef	exp(coef)	se(coef)	z	p
factor(x)2	0.377	1.458	0.340	1.11	0.27

```

Likelihood ratio test=1.22 on 1 df, p=0.269
n= 40, number of events= 35

```

```

> summary(fitphm)
Call:
coxph(formula = Surv(time, status) ~ factor(x))

```

```

n= 40, number of events= 35

```

	coef	exp(coef)	se(coef)	z	Pr(> z)
factor(x)2	0.3769	1.4577	0.3403	1.107	0.268

	exp(coef)	exp(-coef)	lower .95	upper .95
factor(x)2	1.458	0.686	0.7482	2.84

```

Concordance= 0.563 (se = 0.05 )
Rsquare= 0.03 (max possible= 0.994 )
Likelihood ratio test=1.22 on 1 df, p=0.2686
Wald test = 1.23 on 1 df, p=0.2681
Score (logrank) test = 1.24 on 1 df, p=0.2654

```

Example: The data contains information from an experimental study of recidivism of 432 male prisoners, who were observed for a year after being released from prison. The following variables are included in the data; the variable names are those used by Allison (1995), from whom this example and variable descriptions are adapted:

- week : week of first arrest after release, or censoring time.

- arrest : the event indicator, equal to 1 for those arrested during the period of the study and 0 for those who were not arrested.
- fin : a factor, with levels yes if the individual received financial aid after release from prison, and no if he did not; financial aid was a randomly assigned factor manipulated by the researchers.
- age : in years at the time of release.
- race: a factor with levels black and other
- wexp: a factor with levels yes if the individual had full-time work experience prior to incarceration and no if he did not. mar: a factor with levels married if the individual was married at the time of release and not married if he was not.
- paro : a factor coded yes if the individual was released on parole and no if he was not.
- prio : number of prior convictions.
- educ: education, a categorical variable coded numerically, with codes 2 (grade 6 or less), 3 (grades 6 through 9), 4 (grades 10 and 11), 5 (grade 12), or 6 (some post-secondary).

```
> head(data)
```

	week	arrest	fin	age	race	wexp	mar	paro	prio	educ
1	20	1	no	27	black	no not	married	yes	3	3
2	17	1	no	18	black	no not	married	yes	8	4
3	25	1	no	19	other	yes not	married	yes	13	3
4	52	0	yes	23	black	yes	married	yes	1	5
5	52	0	no	19	other	yes not	married	yes	3	3
6	52	0	no	24	black	yes not	married	no	2	4

```
> fit<-coxph(Surv(week, arrest) ~ fin + age + race + wexp + mar + paro + prio,data)
```

```
> summary(fit)
```

```
Call:
```

```
coxph(formula = Surv(week, arrest) ~ +fin + age + race + wexp +  
      mar + paro + prio, data = data)
```

n= 432, number of events= 114

	coef	exp(coef)	se(coef)	z	Pr(> z)	
finyes	-0.37942	0.68426	0.19138	-1.983	0.04742	*
age	-0.05744	0.94418	0.02200	-2.611	0.00903	**
raceother	-0.31390	0.73059	0.30799	-1.019	0.30812	
wexpyes	-0.14980	0.86088	0.21222	-0.706	0.48029	
marnot married	0.43370	1.54296	0.38187	1.136	0.25606	
paroyes	-0.08487	0.91863	0.19576	-0.434	0.66461	
prio	0.09150	1.09581	0.02865	3.194	0.00140	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
finyes	0.6843	1.4614	0.4702	0.9957
age	0.9442	1.0591	0.9043	0.9858
raceother	0.7306	1.3688	0.3995	1.3361
wexpyes	0.8609	1.1616	0.5679	1.3049
marnot married	1.5430	0.6481	0.7300	3.2614
paroyes	0.9186	1.0886	0.6259	1.3482
prio	1.0958	0.9126	1.0360	1.1591

Concordance= 0.64 (se = 0.027)
Rsquare= 0.074 (max possible= 0.956)
Likelihood ratio test= 33.27 on 7 df, p=2.362e-05
Wald test = 32.11 on 7 df, p=3.871e-05
Score (logrank) test = 33.53 on 7 df, p=2.11e-05

Reduced model

```
> fit<-coxph(Surv(week, arrest) ~ fin + age + prio,data=Rossi)
> fit
Call:
coxph(formula = Surv(week, arrest) ~ fin + age + prio, data = Rossi)
```

	coef	exp(coef)	se(coef)	z	p
finyes	-0.3470	0.7068	0.1902	-1.82	0.06820
age	-0.0671	0.9351	0.0209	-3.22	0.00129

prio 0.0969 1.1017 0.0273 3.56 0.00038

Likelihood ratio test=29.1 on 3 df, p=2.19e-06
n= 432, number of events= 114

Test stat = 33.3-29.1= 4.2. The p-value based on a chi-square with 4 df
is 0.3796

Asymptotic Likelihood Theory

3 Likelihood function and the maximum likelihood estimator

In general, we have y_1, y_2, \dots, y_n , independent observations, their distribution depending on the parameter $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \dots, \theta_p)$. Frequently it is the case that $\boldsymbol{\theta}$ is partitioned into two sub-vectors $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$ where $\boldsymbol{\theta}_1$ is the parameter of interest and $\boldsymbol{\theta}_2$ is a nuisance parameter.

- The likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n L_i(\boldsymbol{\theta}).$$

1. Uncensored data:

$$L_i(\boldsymbol{\theta}) = f(t_i|\boldsymbol{\theta}, x_i)$$

where $y_i = (t_i, x_i)$

2. Random censorship data with noninformative censoring:

$$L_i(\boldsymbol{\theta}) = [\lambda(t_i|\boldsymbol{\theta}, x_i)]^{\delta_i} S(t_i|\boldsymbol{\theta}, x_i)$$

where $y_i = (t_i, \delta_i, x_i)$

- The score function

$$U(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial}{\partial \theta_1} \ln L(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_2} \ln L(\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial}{\partial \theta_p} \ln L(\boldsymbol{\theta}) \end{pmatrix}$$

- Under some regularity conditions

$$E(U(\boldsymbol{\theta})) = \mathbf{0} \quad \text{and} \quad \text{Var}(U(\boldsymbol{\theta})) = \mathcal{J}(\boldsymbol{\theta})$$

where $\mathcal{I}(\boldsymbol{\theta})$ is the information matrix

$$\begin{aligned}\mathcal{J}(\boldsymbol{\theta}) = E[U(\boldsymbol{\theta})U^T(\boldsymbol{\theta})] &= -E \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ln L(\boldsymbol{\theta}) \right] \\ &= -E \begin{pmatrix} \frac{\partial^2}{\partial \theta_1^2} \ln L(\boldsymbol{\theta}) & \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \ln L(\boldsymbol{\theta}) & \dots & \frac{\partial^2}{\partial \theta_1 \partial \theta_p} \ln L(\boldsymbol{\theta}) \\ \frac{\partial^2}{\partial \theta_2 \partial \theta_1} \ln L(\boldsymbol{\theta}) & \frac{\partial^2}{\partial \theta_2^2} \ln L(\boldsymbol{\theta}) & \dots & \frac{\partial^2}{\partial \theta_2 \partial \theta_p} \ln L(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \theta_p \partial \theta_1} \ln L(\boldsymbol{\theta}) & \frac{\partial^2}{\partial \theta_p \partial \theta_2} \ln L(\boldsymbol{\theta}) & \dots & \frac{\partial^2}{\partial \theta_p^2} \ln L(\boldsymbol{\theta}) \end{pmatrix}\end{aligned}$$

- Likelihood Equations

$$U(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial}{\partial \theta_1} \ln L(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_2} \ln L(\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial}{\partial \theta_p} \ln L(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$$

- The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{MLE}$ has the property

$$U(\hat{\boldsymbol{\theta}}_{MLE}) = \mathbf{0}.$$

- Under regularity conditions

$$n^{-1/2}U(\boldsymbol{\theta}) \stackrel{Approx}{\sim} N(\mathbf{0}, \mathcal{J}(\boldsymbol{\theta}))$$

as $n \rightarrow \infty$.

4 Hypotheses tests

4.1 The score test

- We can use the result above to test $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.
- The test is the score test and is give by

$$RS = U^T(\boldsymbol{\theta}_0)[\mathcal{J}(\boldsymbol{\theta}_0)]^{-1}U(\boldsymbol{\theta}_0)/n$$

Under H_0 ,

$$RS \xrightarrow{d} \chi_p^2$$

as $n \rightarrow \infty$. We reject H_0 if

$$RS > \chi_p^2(\alpha).$$

- Approximate $100(1 - \alpha)$ confidence interval for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{\theta}, U^T(\boldsymbol{\theta})[\mathcal{J}(\boldsymbol{\theta})]^{-1}U(\boldsymbol{\theta}) \leq \chi_p^2(\alpha)\}$$

- To handle the situation with nuisance parameters, partition $\boldsymbol{\theta}^T$ in $(\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$ where $\boldsymbol{\theta}_1^T = (\theta_1, \theta_2, \dots, \theta_k)^T$ and $\boldsymbol{\theta}_2 = (\theta_{k+1}, \theta_{k+2}, \dots, \theta_p)^T$ and similarly

$$[\mathcal{J}(\boldsymbol{\theta})]^{-1} = \begin{pmatrix} \mathcal{J}^{11}(\boldsymbol{\theta}) & \mathcal{J}^{12}(\boldsymbol{\theta}) \\ \mathcal{J}^{21}(\boldsymbol{\theta}) & \mathcal{J}^{22}(\boldsymbol{\theta}) \end{pmatrix}$$

Where $\mathcal{J}^{11}(\boldsymbol{\theta})$ is a k by k matrix. Suppose we want to test

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0 \quad (\text{nothing is hypothesized about } \boldsymbol{\theta}_2)$$

against

$$H_0 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0.$$

- Let $\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1^0)$ be the mle of $\boldsymbol{\theta}_2$ given $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$, i.e.

$$\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1^0) = \arg \max L(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2)$$

- Define

$$U(\boldsymbol{\theta}_1^0) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) |_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1^0)}$$

- The score statistic for testing $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ is

$$RS = U^T(\boldsymbol{\theta}_1^0) \mathcal{J}^{11}(\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1^0)) U(\boldsymbol{\theta}_1^0) / n$$

- This test statistic has approximately under H_0 a chi-square distribution with k degrees of freedom. We reject H_0 if

$$RS > \chi_k^2(\alpha)$$

where k is the dimension of $\boldsymbol{\theta}_1$.

4.2 Wald test

- Under regularity conditions

$$\hat{\boldsymbol{\theta}}_{MLE} \overset{Approx}{\sim} N(\boldsymbol{\theta}, \mathcal{J}^{-1}(\boldsymbol{\theta})).$$

- This gives another test called the Wald test for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. The test statistic is

$$W = (\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}_0)^T \mathcal{J}(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}_0).$$

- Under H_0

$$W \xrightarrow{d} \chi_p^2$$

and we reject H_0 if $W > \chi_p^2(\alpha)$

- Approximate $100(1 - \alpha)$ confidence interval for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{\theta}, (\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta})^T \mathcal{J}(\boldsymbol{\theta}) ((\hat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}) \leq \chi_p^2(\alpha)\}$$

- To test $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$, use

$$W = (\hat{\boldsymbol{\theta}}_{1,MLE} - \boldsymbol{\theta}_1^0)^T [\mathcal{J}^{11}(\hat{\boldsymbol{\theta}}_{MLE})]^{-1} (\hat{\boldsymbol{\theta}}_{1,MLE} - \boldsymbol{\theta}_1^0)$$

and reject H_0 if $W > \chi_k^2(\alpha)$.

- Practical Example ($k = 1$). Suppose we want to test

$$H_0 : \theta_j = \theta_j^0$$

Here $\theta_1 = \theta_j$ and $\boldsymbol{\theta}_2 = \text{rest}$. Let $\mathcal{J}^{jj}(\boldsymbol{\theta})$ be the (j, j) th element of $[\mathcal{J}(\boldsymbol{\theta})]^{-1}$. Then

$$\frac{(\hat{\theta}_j - \theta_j^0)}{\sqrt{\mathcal{J}^{jj}(\hat{\boldsymbol{\theta}})}} \overset{Approx}{\sim} N(0, 1)$$

- An approximate 95% confidence interval for θ_j is

$$\hat{\theta}_j \pm 1.96 \sqrt{\mathcal{J}^{jj}(\hat{\boldsymbol{\theta}})}$$

4.3 Likelihood Ratio Test (LRT)

- We have always

$$\frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})} \leq 1$$

- When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the likelihood ratio $\Lambda = \frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})}$ is close to 1.
- For this reason, the LRT for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 for small value of Λ
- It turns out that under H_0

$$-2 \ln \frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})} \xrightarrow{d} \chi_p^2$$

- We reject H_0 if

$$-2 \ln \frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})} > \chi_p^2(\alpha).$$

- For testing $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$, again let $\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1^0)$ be the mle of $\boldsymbol{\theta}_2$ given that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ the test statistic is

$$-2 \ln \frac{L(\boldsymbol{\theta}_1^0, \hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1^0))}{L(\hat{\boldsymbol{\theta}})}$$

and we reject H_0 if this test statistics is greater than $\chi_k^2(\alpha)$.

4.4 Information Matrix

$I(\theta)$ is called the "Fisher information" or "expected information", But how should one calculate expectation (i.e $-E \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln L(\boldsymbol{\theta}) \right)$.) when there censoring.

In survival analysis, we typically use "observed information"

$$I(\boldsymbol{\theta}) = - \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln L(\boldsymbol{\theta}) \right)_{p \times p}$$

5 Examples

Example 1: Suppose we observe $(t_i, \delta_i), i = 1, 2, \dots, n$.

$$L(\lambda) = \prod_{i=1}^n L_i(\lambda) = \prod_{i=1}^n [\lambda e^{-\lambda t_i}]^{\delta_i} [e^{-\lambda t_i}]^{1-\delta_i} = \prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda t_i}$$

then

$$\ln L(\lambda) = \sum_{i=1}^n \delta_i \ln(\lambda) - \lambda \sum_{i=1}^n t_i$$

Then

$$U(\lambda) = \frac{\partial}{\partial \lambda} \ln L(\lambda) = \frac{\sum_{i=1}^n \delta_i}{\lambda} - \sum_{i=1}^n t_i = 0 \Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i}$$

The observed Fisher information is

$$I(\lambda) = -\frac{\partial^2}{\partial \lambda^2} \ln L(\lambda) = \frac{\sum_{i=1}^n \delta_i}{\lambda^2}$$

and

$$\widehat{\text{Var}}(\hat{\lambda}) = I^{-1}(\lambda)|_{\lambda=\hat{\lambda}} = \frac{\hat{\lambda}}{\sum_{i=1}^n \delta_i} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i^2}$$

and

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, I^{-1}(\lambda))$$

An approximate 95% confidence interval for λ is

$$\hat{\lambda} \pm 1.96 \sqrt{\widehat{\text{Var}}(\hat{\lambda})}$$

that is

$$\frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i} \pm 1.96 \frac{\sqrt{\sum_{i=1}^n \delta_i}}{\sum_{i=1}^n t_i}$$

Example 2 : Exponential Regression

Data : $(t_i, \delta_i, x_i), \mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T, i = 1, 2, \dots, n$

Hazard Model : $\lambda(t|x) = \lambda e^{\beta^T \mathbf{x}}$ where $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$

Likelihood:

$$\begin{aligned} L(\lambda, \beta) &= \prod_{i=1}^n [\lambda(t_i|\mathbf{x}_i)]^{\delta_i} S(t_i|\mathbf{x}_i) \\ &= \prod_{i=1}^n \lambda^{\delta_i} e^{\delta_i \beta^T \mathbf{x}_i} e^{-\lambda e^{\beta^T \mathbf{x}_i} t_i} \\ &= \lambda^{\sum \delta_i} e^{\beta^T \sum \mathbf{x}_i \delta_i} e^{-\lambda \sum e^{\beta^T \mathbf{x}_i} t_i} \end{aligned}$$

log-likelihood

$$\sum \delta_i \lambda + \beta^T \sum \mathbf{x}_i \delta_i - \lambda \sum e^{\beta^T \mathbf{x}_i} t_i$$

Likelihood equations:

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ln L &= \frac{\sum \delta_i}{\lambda} - \sum e^{\beta^T \mathbf{x}_i} t_i = 0 \\ \frac{\partial}{\partial \beta_j} \ln L &= \sum \mathbf{x}_{ji} \delta_i - \lambda \sum x_{ji} e^{\beta^T \mathbf{x}_i} t_i = 0, j = 1, 2, \dots, p. \end{aligned}$$

How do we solve the likelihood equations for $\hat{\lambda}$ and $\hat{\beta}$?

Special Case: Two sample problem

$$p = 1, x_i = \begin{cases} 0, & \text{if } i \text{ is in group 1} \\ 1, & \text{if } i \text{ is in group 2} \end{cases}$$

so the hazard rate for group 1 is λ and for group 2 the hazard rate is λe^{β} .

Suppose

$$\begin{aligned} d_j &= \text{number of individuals who failed in group } j, j = 1, 2 \\ V_j &= \text{total observed time under study in group } j, j = 1, 2 \end{aligned}$$

that is

$$\begin{aligned} d_1 &= \sum_{i=1}^n \delta_i (1 - x_i), \quad d_2 = \sum_{i=1}^n \delta_i x_i \quad \text{this implies that} \quad \sum \delta_i = d_1 + d_2 \\ V_1 &= \sum_{i=1}^n t_i (1 - x_i), \quad V_2 = \sum_{i=1}^n t_i x_i \end{aligned}$$

Then from above, the likelihood equations are

$$\frac{\partial}{\partial \lambda} \ln L = \frac{d_1 + d_2}{\lambda} - V_1 - e^\beta V_2 = 0 \quad (1)$$

$$\frac{\partial}{\partial \beta} \ln L = d_2 - \lambda e^\beta V_2 = 0 \quad (2)$$

$$(2) \Rightarrow \hat{\lambda} = e^{-\hat{\beta}} \frac{d_2}{V_2}$$

Substitution into (1) gives

$$(d_1 + d_2) e^{\hat{\beta}} \frac{d_2}{V_1} - V_1 - e^{\hat{\beta}} V_2 = 0$$

this implies that

$$e^{\hat{\beta}} = \frac{V_1/d_1}{V_2/d_2} \quad \text{and} \quad \hat{\lambda} = d_1/V_1$$

Notice that $e^{\hat{\beta}}$ is the ratio of failure rates.

The information matrix in this case is

$$I(\lambda, \beta) = \begin{pmatrix} \lambda^{-2} \sum_i \delta_i & \sum_i x_i t_i e^{\beta x_i} \\ \sum_i x_i t_i e^{\beta x_i} & \lambda \sum_i x_i^2 t_i e^{\beta x_i} \end{pmatrix}$$

Note that

$$\begin{aligned} \sum_i \delta_i / \hat{\lambda}^2 &= \frac{d_1 + d_2}{d_1^2} V_1^2 \\ \sum_i x_i t_i e^{\hat{\beta} x_i} &= e^{\hat{\beta}} V_2 = \frac{V_1 d_2}{d_1 V_2} V_2 \\ \sum_i x_i^2 t_i e^{\beta x_i} &= \sum_i x_i t_i e^{\beta x_i} \end{aligned}$$

This implies that

$$I(\hat{\lambda}, \hat{\beta}) = \begin{pmatrix} \frac{d_1 + d_2}{d_1^2} V_1^2 & \frac{V_1 d_2}{d_1} \\ \frac{V_1 d_2}{d_1} & d_2 \end{pmatrix}$$

Therefore

$$\widehat{Var}(\hat{\lambda}, \hat{\beta}) = I^{-1}(\hat{\lambda}, \hat{\beta}) = \begin{pmatrix} \frac{d_1}{V_1^2} & -V_1^{-1} \\ -V_1^{-1} & \frac{d_1 + d_2}{d_1 d_2} \end{pmatrix}$$

We will use the following data to illustrate the mle based procedures. the data is time measured in 100 days

Group 1 : 43, 64, 88, 88, 90, 92, 106, 109, 113, 116, 120, 127, 130, 134, 146, 165, 204, 116+, 144 +
Group 2 : 42, 56, 73, 98, 105, 132, 132, 133, 133, 133, 133, 139, 140, 161, 180, 196, 223, 104+, 244+

It is easy to see that $d_1 = 17, V_1 = 2195, d_2 = 19, V_2 = 2923$. This implies that

$$\hat{\lambda} = \frac{17}{2195} = 0.007745, \quad \hat{\beta} = \ln \left(\frac{V_1 d_2}{V_2 d_1} \right) = \ln(0.839) = -0.175.$$

To test $H_0 : \beta = 0$, the Wald test is

$$\frac{\hat{\beta}}{\sqrt{\widehat{Var}(\hat{\beta})}} = \frac{\hat{\beta}}{\sqrt{\frac{d_1 + d_2}{d_1 d_2}}} = \ln \left(\frac{V_1 d_2}{V_2 d_1} \right) \sqrt{\frac{d_1 d_2}{d_1 + d_2}}$$

This is equal to -0.524 (not significant)

A 95% confidence interval for β is

$$\hat{\beta} \pm 1.96 \sqrt{\frac{d_1 + d_2}{d_1 d_2}} = (-0.830, 0.480)$$