

Chapter 2

Introduction to Survival Analysis



Abstract This chapter provides a concise introduction to survival analysis. We review the essential tools in survival analysis, such as the survival function, Kaplan–Meier estimator, hazard function, log-rank test, Cox regression, and likelihood-based inference.

Keywords Censoring · Cox regression · Independent censoring
Kaplan–Meier estimator · Log-rank test · Overall survival
Time-to-tumor progression

2.1 Survival Time

In survival analysis, the term *survival time* refers to the time elapsed from an origin to the occurrence of an event. In many medical studies, the origin is the time at study entry which can be the start of a medical treatment, the initiation of a randomized experiment, or the operation date of surgery. In epidemiological and demographic studies, the origin is often the date of birth. The event may be the occurrence of death.

In medical research, the term *overall survival* refers to survival time measured from entry until death of a patient. For instance, to measure the effect of chemotherapy or radiotherapy in locally advanced head and neck cancer, researchers may use overall survival as the primary endpoint (Michiels et al. 2009). In this study, the origin is the start of randomization.

2.2 Kaplan–Meier Estimator and Survival Function

We shall introduce the random censorship model where we consider two *random variables*

- T : survival time
- U : censoring time

Due to censoring, either one of T or U is observed. One can observe T if death comes faster than censoring ($T \leq U$). On the other hand, one cannot exactly observe T if censoring comes faster than death ($U < T$). Even if the exact value of T is unknown for the censored case, T is known to be greater than U . What we observe are the first occurring time ($\min\{T, U\}$) and the censoring status ($\{T \leq U\}$ or $\{U < T\}$). The random censorship model typically assumes that T and U are independent, namely $\Pr(T \in A, U \in B) = \Pr(T \in A) \Pr(U \in B)$ for sets A and B .

Survival data consist of (t_i, δ_i) , $i = 1, \dots, n$, where

- t_i : survival time or censoring time whichever comes first,
- δ_i : censoring indicator ($\delta_i = 1$ if t_i is survival time, or $\delta_i = 0$ if t_i is censoring time).

Under the random censorship model, one can write $t_i = \min\{T, U\}$ and $\delta_i = \mathbf{I}(T \leq U)$, where $\mathbf{I}(\cdot)$ is the indicator function. We shall assume that all the observed times to death are distinct ($t_i \neq t_j$ whenever $i \neq j$ and $\delta_i = \delta_j = 1$), so that there is no *ties* in the death times. With the survival data, one can estimate the survival function $S(t) \equiv \Pr(T > t)$ by the following estimator:

Definition 1 *The Kaplan–Meier estimator* (Kaplan and Meier 1958) is defined as

$$\hat{S}(t) = \prod_{t_i \leq t, \delta_i=1} \left(1 - \frac{1}{n_i}\right), \quad 0 \leq t \leq \max_i(t_i)$$

where $n_i = \sum_{\ell=1}^n \mathbf{I}(t_\ell \geq t_i)$ is the *number at-risk* at time t_i ; $\hat{S}(t) = 1$ if no death occurs up to time t ; $\hat{S}(t)$ is undefined for $t > \max_i(t_i)$.

The derivation of the Kaplan–Meier estimator: Consider a survival function that is a decreasing step function with jumps only at points where a death occurs at observed times of death. Then, one can write (Exercise 1 in Sect. 2.9) the survival function in the form

$$S(t) = \Pr(T > t) = \prod_{t_i \leq t, \delta_i=1} \left(1 - \frac{\Pr(T = t_i)}{\Pr(T \geq t_i)}\right).$$

Second, suppose that T and U are independent. Then, one can write

$$\begin{aligned} S(t) &= \prod_{t_i \leq t, \delta_i=1} \left(1 - \frac{\Pr(T = t_i, U \geq t_i)}{\Pr(T \geq t_i, U \geq t_i)}\right) \\ &= \prod_{t_i \leq t, \delta_i=1} \left(1 - \frac{\Pr(\min\{T, U\} = t_i, T \leq U)}{\Pr(\min\{T, U\} \geq t_i)}\right). \end{aligned}$$

Finally, we replace the probability ratio of the last expression by its estimate to obtain

$$\hat{S}(t) = \prod_{t_i \leq t, \delta_i=1} \left(1 - \frac{\sum_{\ell=1}^n \mathbf{I}(t_\ell = t_i, \delta_\ell = 1)/n}{\sum_{\ell=1}^n \mathbf{I}(t_\ell \geq t_i)/n} \right) = \prod_{t_i \leq t, \delta_i=1} \left(1 - \frac{1}{n_i} \right).$$

It is now clear that the Kaplan–Meier estimator relies on the independence assumption between T and U . ■

The Kaplan–Meier survival curve is defined as the plot of $\hat{S}(t)$ against t , starting with $t = 0$ and ending with $t_{\max} = \max_i(t_i)$. The curve is a step function that jumps only at points where a death occurs. On the curve, censoring times are often indicated as the mark “+”.

If $t_{\max} = \max_i(t_i)$ corresponds to time-to-death of a patient, then $\hat{S}(t_{\max}) = 0$. If $t_{\max} = \max_i(t_i)$ corresponds to censoring time of a patient, then $\hat{S}(t_{\max}) > 0$. It is misleading to plot $\hat{S}(t)$ only up to the largest death time $\max_{i; \delta_i=1}(t_i)$, especially when many patients are alive beyond $\max_{i; \delta_i=1}(t_i)$.

Survival data often include *covariates*, such as gender, tumor size, and cancer stage. With covariates, survival data consist of $(t_i, \delta_i, \mathbf{x}_i)$, $i = 1, \dots, n$, where

- $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$: p -dimensional covariates

In traditional survival analysis, the data is analyzed under the following assumption:

Independent censoring assumption: *Survival time and censoring time are independent given covariates. That is, T and U are conditionally independent given \mathbf{x} .*

For a patient i , one can define the *survival function* denoted as $S(t|\mathbf{x}_i) \equiv \Pr(T > t | \mathbf{x}_i)$ for $t \geq 0$. The survival function is the probability that the patient is alive at time t . The survival function $S(t|\mathbf{x}_i)$ is, in fact, the *patient-level survival function* as it is conditionally on the patient characteristics \mathbf{x}_i . The survival function at $\mathbf{x}_i = \mathbf{0}$ is called the *baseline survival function* and denoted as $S_0(t) = S(t|\mathbf{x}_i = \mathbf{0})$.

A *parametric model* is given by a survival function that is specified by a finite number of parameters. For instance, we consider an exponential survival function

$S(t|x_i) = \exp(-\lambda t e^{\beta x_i})$, $t \geq 0$, where $\lambda > 0$ and $-\infty < \beta < \infty$ are parameters. Let x_i denote the gender with $x_i = 1$ for male and $x_i = 0$ for female. One can show that $S(t|x_i) = S_0(t)^{\exp(\beta x_i)}$ for $t \geq 0$, where $S_0(t) = S(t|x_i = 0) = \exp(-\lambda t)$ is the baseline survival function. With this model, survival difference between male and female is captured by β . The case $\beta > 0$ corresponds to poor survival prognosis for male relative to female; the case $\beta < 0$ corresponds to good survival prognosis for male relative to female. The case $\beta = 0$ corresponds to equal survival prognosis between male and female.

A *semi-parametric model* is given by a survival function that is *partially* specified by a finite number of parameters. For instance, we consider a survival function $S(t|x_i) = S_0(t)^{\exp(\beta x_i)}$, where the form of the baseline survival function $S_0(t)$ is unspecified. In terms of β , one can compare survival between males and females without assuming a specific model on the baseline survival function.

2.3 Hazard Function

Hereafter, we suppose that $S(t|\mathbf{x}_i)$ is a continuous survival function. The instantaneous death probability between t and $t + dt$ is $\Pr(t \leq T < t + dt | \mathbf{x}_i) = S(t | \mathbf{x}_i) - S(t + dt | \mathbf{x}_i)$, where dt is an infinitely small number. Since this probability is equal to zero, one can consider the *rate* by dividing by dt such that

$$\begin{aligned} f(t | \mathbf{x}_i) &= \frac{S(t | \mathbf{x}_i) - S(t + dt | \mathbf{x}_i)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{S(t | \mathbf{x}_i) - S(t + \Delta t | \mathbf{x}_i)}{\Delta t} \\ &= -\frac{dS(t | \mathbf{x}_i)}{dt}. \end{aligned}$$

This is the *density function*.

The *hazard rate* describes the instantaneous death rate between t and $t + dt$ given that the patient is at-risk at t :

Definition 2 The *hazard function (or hazard rate function)* is defined as

$$h(t|\mathbf{x}_i) \equiv \frac{\Pr(t \leq T < t + dt | T \geq t, \mathbf{x}_i)}{dt} = \frac{-\frac{d}{dt} S(t | \mathbf{x}_i)}{S(t | \mathbf{x}_i)}.$$

The hazard function at $\mathbf{x}_i = \mathbf{0}$ is called the *baseline hazard function* and denoted as $h_0(t) = h(t|\mathbf{x}_i = \mathbf{0})$. The *cumulative hazard function* is defined as $H(t|\mathbf{x}_i) = \int_0^t h(u|\mathbf{x}_i) du$. The survival function is derived from the hazard function through $S(t|\mathbf{x}_i) = \exp\{-H(t|\mathbf{x}_i)\}$.

The hazard rate is also known as the *force of mortality* in actuarial science and demography. For example, let $t = \text{“60 years old”}$, $dt = \text{“1 year”}$, and $x_i = 1$ for male or $x_i = 0$ for female. Then, the force of mortality $h(60|x_i = 1)$ is equal to the probability of death within the next one year for a 60-year-old man. The Japanese life tables show $h(60|x_i = 1) = 0.0064$ (0.64%). The value of $h(t|x_i = 1)$ monotonically increases as t grows, which represents the effect of natural aging. Eventually, it reaches $h(100|x_i = 1) = 0.3995$ (39.95%). This implies that 40% of Japanese males who have just celebrated their 100th birthday will die before their next birthday. Life tables for almost any country are available in the internet (e.g., Google “Taiwan life table”).

Unfortunately, the hazard function for cancer patients in medical studies rarely shows any simple pattern (e.g., monotonically increasing or decreasing). In many clinical trials, the time t is measured from the start of treatment, and hence, the ages are regarded as covariates. In this case, the hazard of patients may be influenced by the treatment effect, the follow-up processes, and cancer progression, so the effect of natural aging may diminish. In epidemiological studies, focusing on age-specific incidence of a particular disease, the time t is measured from birth as in the example from Japanese life tables. However, the shape of the hazard function of disease incidence may be difficult to specify.

This implies that many simple models, such as the exponential, Weibull, and lognormal models, may not fit survival data from cancer patients. This is why semi-parametric models are more useful and widely applied in medical research. One may still accept the assumptions that the hazard function is continuous, does not abruptly change over time, and smooth (continuously differentiable). Hazard models with cubic splines (Chap. 4) meet these assumptions without restricting too much the shape of the hazard function.

The semi-parametric model $S(t|x_i) = S_0(t)^{\exp(\beta x_i)}$ can alternatively be specified in terms of the hazard function

$$h(t|x_i) = h_0(t) \exp(\beta x_i) \quad (2.1)$$

where the form of $h_0(t)$ is unspecified. One can show $h_0(t) = -d\{\log S_0(t)\}/dt$ and $S_0(t) = \exp\{-H_0(t)\}$, where $H_0(t) = \int_0^t h_0(u)du$.

Let x_i be a dichotomous covariate, such as gender with $x_i = 1$ for male and $x_i = 0$ for female. Under the model (2.1), the *relative risk* (RR) is defined as

$$RR = \exp(\beta) = \frac{h(t|x_i = 1)}{h(t|x_i = 0)}.$$

For instance, the value $RR = 2$ implies that death rate for $x_i = 1$ is twice the death rate for $x_i = 0$.

Let x_i be a continuous covariate, such as a gene expression. If the scale of x_i is standardized (to be mean = 0 and SD = 1), then $RR = \exp(\beta)$ is interpreted with

respect to one SD increase. If one is interested in the effect of $x_i = 2$ relative to $x_i = -2$, then $RR = \exp(4\beta)$.

2.4 Log-Rank Test for Two-Sample Comparison

The log-rank test is a method to test the equality of the hazard rates between two groups. Specifically, we consider the null hypothesis

$$H_0 : h(t|x_i = 0) = h(t|x_i = 1), \quad t \geq 0,$$

where $x_i = 1$ for male and $x_i = 0$ for female, for instance. This null hypothesis is the same as the equality $S(t|x_i = 0) = S(t|x_i = 1)$ due to the relationship between the hazard function and survival function. We wish to test H_0 without making any model assumption, but with the assumption that there are no ties in death times. The treatment of ties shall be briefly discussed in Sect. 2.8.

Let $n_{i1} = \sum_{\ell=1}^n \mathbf{I}\{t_\ell \geq t_i, x_\ell = 1\}$ be the number of males and $n_{i0} = \sum_{\ell=1}^n \mathbf{I}\{t_\ell \geq t_i, x_\ell = 0\}$ be the number of females at-risk at time t_i . Hence, $n_{i0} + n_{i1}$ is the total number at-risk at time t_i . Each death at time t_i corresponds to either the death of male ($x_i = 1$) or the death of female ($x_i = 0$). If there is no effect of gender on survival, male and female have the same death rate. Hence, the conditional expectation of x_i given $(\delta_i = 1, n_{i0}, n_{i1})$ is

$$\begin{aligned} E[x_i | \delta_i = 1, n_{i0}, n_{i1}] &= \Pr(x_i = 1 | \delta_i = 1, n_{i0}, n_{i1}) \\ &= \frac{\Pr(x_i = 1, \delta_i = 1 | n_{i0}, n_{i1})}{\Pr(x_i = 1, \delta_i = 1 | n_{i0}, n_{i1}) + \Pr(x_i = 0, \delta_i = 1 | n_{i0}, n_{i1})} \\ &= \frac{n_{i1} h(t_i | x_i = 1)}{n_{i1} h(t_i | x_i = 1) + n_{i0} h(t_i | x_i = 0)} \\ &= \frac{n_{i1}}{n_{i0} + n_{i1}}. \end{aligned}$$

The last equation holds under the null hypothesis H_0 . The difference between x_i and its expectation leads to the *log-rank statistic*

$$S = \sum_{i=1}^n \delta_i \left(x_i - \frac{n_{i1}}{n_{i0} + n_{i1}} \right).$$

Hence, $S > 0$ is associated with higher death rate in male than that in female. Under H_0 , the mean of S is zero. If we assume that x_i 's are independent,

$$Var(S) = \sum_{i=1}^n \delta_i \frac{n_{i1}n_{i0}}{(n_{i0} + n_{i1})^2}.$$

The log-rank test for no gender effect is based on the Z-statistic $z = S/\sqrt{Var(S)}$ or the chi-square statistic z^2 . The P-value is computed as $Pr(|Z| > |z|)$, where $Z \sim N(0, 1)$.

Example 1 Consider a sample of five females and five males ($n = 10$) with $t_i = (1650, 30, 720, 450, 510, 1110, 210, 1380, 1800, 540)$, $\delta_i = (0, 1, 0, 1, 1, 0, 1, 1, 0, 1)$, and $x_i = (0, 0, 0, 0, 0, 1, 1, 1, 1, 1)$. To calculate the log-rank statistic, it is convenient to summarize the data into Table 2.1.

The log-rank statistic has the “(observed)-(expected)” form, namely

$$S = \sum_{i=1}^n \delta_i x_i - \sum_{i=1}^n \delta_i \frac{n_{i1}}{n_{i0} + n_{i1}} = 3 - \left(\frac{5}{10} + \frac{5}{9} + \frac{4}{8} + \frac{4}{7} + \frac{4}{6} + \frac{2}{3} \right) = 3 - 3.46 = -0.46.$$

The negative value of S implies that the observed mortality of male is lower than its expected value under H_0 . The variance is computed from Table 2.1 as

$$Var(S) = \sum_{i=1}^n \delta_i \frac{n_{i1}n_{i0}}{(n_{i0} + n_{i1})^2} = \frac{5 \times 5}{10^2} + \frac{5 \times 4}{9^2} + \frac{4 \times 4}{8^2} + \frac{4 \times 3}{7^2} + \frac{4 \times 2}{6^2} + \frac{2 \times 1}{3^2} = 1.436.$$

Hence, the test statistic is $z = S/\sqrt{Var(S)} = -0.46/\sqrt{1.436} = -0.384$, and the P-value is $Pr(|Z| > 0.384) = 0.70$. We see no significant evidence for gender effect on survival. ■

The log-rank test is a non-parametric test that does not employ any distributional assumption. The log-rank test simply examines the excess mortality. Software packages for survival analysis display both “observed” and “expected” numbers of deaths in their outputs, in addition to the Z-value and P-value. The log-rank test can also handle *left-truncation* (Klein and Moeschberger 2003). The log-rank test has

Table 2.1 Tabulation of the $n = 10$ samples

Death times: t_i with $\delta_i = 1$	Observed: x_i	Expected: $n_{i1}/(n_{i0} + n_{i1})$
30	0	5/10
210	1	5/9
450	0	4/8
510	0	4/7
540	1	4/6
1380	1	2/3

variants, such as multi-group tests, log-rank trend tests, and stratified log-rank tests (Collett 2003; Klein and Moeschberger 2003).

2.5 Cox Regression

Since the hazard function is the basis of the risk comparison between two groups, it is then natural to incorporate the effect of covariates into the hazard function.

Definition 3 *The Cox proportional hazards model (Cox 1972) is defined as*

$$h(t|\mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}_i),$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ are unknown coefficients and $h_0(\cdot)$ is an unknown baseline hazard function.

The Cox model states that the hazard function $h(t|\mathbf{x}_i)$ is proportional to $h_0(t)$ with the relative risk $\exp(\boldsymbol{\beta}'\mathbf{x}_i)$. This implies that all patients share the same time-trend function $h_0(t)$. The most striking feature of the Cox model is that the form of $h_0(\cdot)$ is unspecified. Hence, the Cox model is a semi-parametric model, offering greater flexibility over parametric models that specify the form of $h(t|\mathbf{x}_i)$.

One can estimate $\boldsymbol{\beta}$ without estimating $h_0(\cdot)$. Based on data $(t_i, \delta_i, \mathbf{x}_i)$, $i = 1, \dots, n$, let $R_i = \{ \ell : t_\ell \geq t_i \}$ be the *risk set* that contains patients at-risk at time t_i . The *partial likelihood estimator* $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)'$ is defined by maximizing the *partial likelihood function* (Cox 1972)

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\sum_{\ell \in R_i} \exp(\boldsymbol{\beta}'\mathbf{x}_\ell)} \right)^{\delta_i}.$$

The log-partial likelihood is

$$\ell(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \left[\boldsymbol{\beta}'\mathbf{x}_i - \log \left\{ \sum_{\ell \in R_i} \exp(\boldsymbol{\beta}'\mathbf{x}_\ell) \right\} \right]. \quad (2.2)$$

The derivatives of $\ell(\boldsymbol{\beta})$ give the *score function*,

$$\mathbf{S}(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \delta_i \left[\mathbf{x}_i - \frac{\sum_{\ell \in R_i} \mathbf{x}_\ell \exp(\boldsymbol{\beta}'\mathbf{x}_\ell)}{\sum_{\ell \in R_i} \exp(\boldsymbol{\beta}'\mathbf{x}_\ell)} \right].$$

The second-order derivatives of $\ell(\boldsymbol{\beta})$ constitute the *Hessian matrix*,

$$H(\boldsymbol{\beta}) = \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = - \sum_{i=1}^n \delta_i \left[\frac{\sum_{\ell \in R_i} \mathbf{x}_\ell \mathbf{x}_\ell' \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)}{\sum_{\ell \in R_i} \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)} - \frac{\sum_{\ell \in R_i} \mathbf{x}_\ell \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)}{\sum_{\ell \in R_i} \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)} \left\{ \frac{\sum_{\ell \in R_i} \mathbf{x}_\ell \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)}{\sum_{\ell \in R_i} \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)} \right\}' \right].$$

Since $H(\boldsymbol{\beta})$ is a negative definite matrix (see Exercise 3 in Sect. 2.9), the log-likelihood $\ell(\boldsymbol{\beta})$ is concave. This implies that $\ell(\boldsymbol{\beta})$ has a unique maxima $\hat{\boldsymbol{\beta}}$ that solves $\mathbf{S}(\boldsymbol{\beta}) = \mathbf{0}$.

Interval estimation for $\boldsymbol{\beta}$ is implemented by applying the asymptotic theory (Fleming and Harrington 1991). The *information matrix* is defined as $i(\hat{\boldsymbol{\beta}}) = -H(\hat{\boldsymbol{\beta}})$. The standard error (SE) of $\hat{\beta}_j$ is $SE(\hat{\beta}_j) = \sqrt{\{i^{-1}(\hat{\boldsymbol{\beta}})\}_{jj}}$ that uses the j -th diagonal element of the inverse information matrix. The 95% confidence interval (CI) is $\hat{\beta}_j \pm 1.96 \times SE(\hat{\beta}_j)$.

To gain more insight into Cox regression, we consider a simple case where x_i denote the gender defined as $x_i = 1$ for male and $x_i = 0$ for female. In this setting, the Cox model is written as $h(t|x_i) = h_0(t) \exp(\beta x_i)$, where the factor $\exp(\beta)$ represents the RR of male relative to female.

We shall demonstrate how the factor $\exp(\beta)$ is estimated by maximizing the log-partial likelihood in Eq. (2.2). We solve the score equation $S(\beta) = 0$ where

$$S(\beta) = \sum_{i=1}^n \delta_i \left[x_i - \frac{\sum_{\ell \in R_i} x_\ell \exp(\beta x_\ell)}{\sum_{\ell \in R_i} \exp(\beta x_\ell)} \right] = \sum_{i=1}^n \delta_i \frac{x_i n_{i0} - (1 - x_i) n_{i1} \exp(\beta)}{n_{i0} + n_{i1} \exp(\beta)}.$$

Hence, the estimate of $\exp(\beta)$ needs to satisfy the equation

$$\exp(\beta) = \frac{\sum_{i:x_i=1} \delta_i \frac{n_{i0}}{n_{i0} + n_{i1} \exp(\beta)}}{\sum_{i:x_i=0} \delta_i \frac{n_{i1}}{n_{i0} + n_{i1} \exp(\beta)}}. \quad (2.3)$$

This is the ratio of the expected number of deaths in male divided by the expected number of deaths in female, which agrees with the interpretation of $\exp(\beta)$.

Equation (2.3) can be solved by the *fixed-point iteration algorithm*. First, applying the initial value $\exp(\beta) = 1$ to the right-hand side of Eq. (2.3), we have

$$\exp(\beta) = \frac{\sum_{i:x_i=1} \delta_i \frac{n_{i0}}{n_{i0} + n_{i1}}}{\sum_{i:x_i=0} \delta_i \frac{n_{i1}}{n_{i0} + n_{i1}}}.$$

We apply this value of $\exp(\beta)$ to the right-hand side of Eq. (2.3) to give an updated value of $\exp(\beta)$. This process is repeated until the updated value does not change from the previous step. While the fixed-point iteration gives us an insight

about how $\exp(\beta)$ is estimated from data, it requires a large number of iterations until convergence.

A computationally faster algorithm is the *Newton–Raphson algorithm*, which utilizes the score function $S(\beta) = d\ell(\beta)/d\beta$ and the *Hessian* $H(\beta) = d^2\ell(\beta)/d\beta^2$. The algorithm starts with the initial value $\beta^{(0)} = 0$, and then follows the sequence

$$\beta^{(k+1)} = \beta^{(k)} - H^{-1}(\beta^{(k)})S(\beta^{(k)}), \quad k = 0, 1, \dots$$

The algorithm converges if $|\beta^{(k+1)} - \beta^{(k)}| \approx 0$. Then, the estimate is $\hat{\beta} = \beta^{(k)}$ and its standard error is $SE(\hat{\beta}) = \sqrt{-H^{-1}(\hat{\beta})}$. The score function is

$$S(\beta) = \sum_{i=1}^n \delta_i \left[x_i - \frac{\sum_{\ell \in R_i} x_\ell \exp(\beta x_\ell)}{\sum_{\ell \in R_i} \exp(\beta x_\ell)} \right] = \sum_{i=1}^n \delta_i \left[x_i - \frac{n_{i1} \exp(\beta)}{n_{i0} + n_{i1} \exp(\beta)} \right],$$

and the Hessian is

$$\begin{aligned} H(\beta) &= - \sum_{i=1}^n \delta_i \left[\frac{\sum_{\ell \in R_i} x_\ell^2 \exp(\beta x_\ell)}{\sum_{\ell \in R_i} \exp(\beta x_\ell)} - \left\{ \frac{\sum_{\ell \in R_i} x_\ell \exp(\beta x_\ell)}{\sum_{\ell \in R_i} \exp(\beta x_\ell)} \right\}^2 \right] \\ &= - \sum_{i=1}^n \delta_i \frac{n_{i0} n_{i1} \exp(\beta)}{\{ n_{i0} + n_{i1} \exp(\beta) \}^2}. \end{aligned}$$

We use Example 1 to compare the convergence between the fixed-point iteration and Newton–Raphson algorithms. Table 2.2 shows that the Newton–Raphson converges faster than the fixed-point iteration. The two algorithms reach the same value $\hat{\beta} = -0.3156$.

The Wald test for the null hypothesis $H_0 : \beta = 0$ is based on the Z-value $z = \hat{\beta}/SE(\hat{\beta})$. The P-value is computed as $\Pr(|Z| > |z|)$, where $Z \sim N(0, 1)$.

The score test for the null hypothesis $H_0 : \beta = 0$ uses the score statistic, and its variance,

$$S(0) = \sum_{i=1}^n \delta_i \left(x_i - \frac{n_{i1}}{n_{i0} + n_{i1}} \right), \quad \text{Var}\{ S(0) \} = -H(0) = \sum_{i=1}^n \delta_i \frac{n_{i1} n_{i0}}{(n_{i0} + n_{i1})^2}.$$

Table 2.2 Iteration algorithms to compute $\hat{\beta}$ using the data of Example 1

Iteration number k	Fixed-point iteration $\beta^{(k)}$	Newton–Raphson $\beta^{(k)}$
0	0	0
1	-0.3093212	-0.3204982
2	-0.3154621	-0.3155884
3	-0.3155858	–

Note The convergence criterion is $|\beta^{(k+1)} - \beta^{(k)}| \leq 10^{-5}$

The score test based on $z = S(0)/\sqrt{\text{Var}\{S(0)\}}$ is exactly the same as the log-rank test. This coincidence does not imply that the log-rank test relies on the Cox model assumption (Sect. 2.8).

The Newton–Raphson algorithm can also be applied to the multi-dimensional case ($p \geq 2$) (see Sect. 2.7). The fixed-point iteration algorithm, however, may not be easily applied to the multi-dimensional case (see Exercise 4 in Sect. 2.9).

2.6 R Survival Package

We shall briefly introduce the R package *survival* to analyze real data. After installing the package, we enter survival time t_i , censoring indicator δ_i , and covariate x_i for $n = 10$ patients. Then, we run the codes:

```
library(survival)
t.event=c(1650, 30, 720, 450, 510, 1110, 210, 1380, 1800, 540)
event=c(0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1)
x=c(0,0,0,0,0,1,1,1,1,1) ## female=0, male=1
survdifff(Surv(t.event, event) ~ x) ## log-rank test
summary( coxph(Surv(t.event,event)~x) ) ## Cox regression
fit=survfit(Surv(t.event, event)~1) ## Kaplan-Meier estimator
plot(fit,mark.time=TRUE) ## Kaplan-Meier survival curve
```

The outputs are shown below and Fig. 2.1.

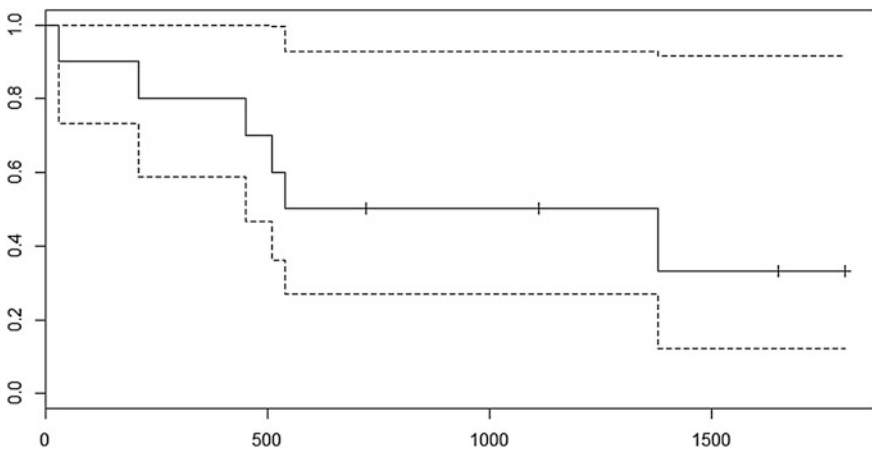


Fig. 2.1 Kaplan–Meier survival curve and the 95% CI calculated from the data of Example 1 ($n = 10$). Censoring times are indicated as the mark“+”

```

> survdiff(Surv(t.event, event) ~ x) ## log-rank test
Call:
survdiff(formula = Surv(t.event, event) ~ x)
      N Observed Expected (O-E)^2/E (O-E)^2/V
x=0   5         3     2.54   0.0834   0.148
x=1   5         3     3.46   0.0612   0.148
Chisq= 0.1 on 1 degrees of freedom, p= 0.701
> summary(coxph(Surv(t.event,event)~x)) ## Cox regression
Call:
coxph(formula = Surv(t.event, event) ~ x)
n= 10, number of events= 6
      coef exp(coef) se(coef)      z Pr(>|z|)
x -0.3156  0.7294   0.8249  -0.383  0.702

      exp(coef) exp(-coef) lower .95 upper .95
x   0.7294      1.371   0.1448   3.673

Concordance= 0.581 (se = 0.123 )
Rsquare= 0.014 (max possible= 0.898 )
Likelihood ratio test= 0.15 on 1 df, p=0.7026
Wald test              = 0.15 on 1 df, p=0.702
Score (logrank) test = 0.15 on 1 df, p=0.7009

```

The results on the log-rank test show $S = 3 - 3.46 = -0.46$ with the chi-square statistics $z^2 = 0.148$ and the P-value = 0.701 (see the row of “x = 1”). The results on Cox regression show $\hat{\beta} = -0.316$, $RR = \exp(\hat{\beta}) = 0.729$, $SE(\hat{\beta}) = 0.825$, and $z = \hat{\beta}/SE(\hat{\beta}) = -0.38$. The P-value of the Wald test is 0.702. Hence, the log-rank test and the Wald test show similar results. In addition, the log-rank test and the score test yield the identical result.

Since the difference between the two groups is not significant, we combine the two groups and then draw the Kaplan–Meier survival curve. Figure 2.1 display the Kaplan–Meier survival curve and the 95% CI.

2.7 Likelihood-Based Inference

This section considers likelihood-based methods for analyzing the data $(t_i, \delta_i, \mathbf{x}_i)$, $i = 1, \dots, n$. Recall that we defined survival time T and censoring time U such that:

- $T = t_i$ and $U > t_i$ if $\delta_i = 1$,
- $T > t_i$ and $U = t_i$ if $\delta_i = 0$.

Combining these two events, the likelihood for the i -th patient is expressed as

$$L_i = \Pr(T = t_i, U > t_i | \mathbf{x}_i)^{\delta_i} \Pr(T > t_i, U = t_i | \mathbf{x}_i)^{1-\delta_i}.$$

Under the independent censoring assumption,

$$\begin{aligned} L_i &= [\Pr(T = t_i | \mathbf{x}_i) \Pr(U > t_i | \mathbf{x}_i)]^{\delta_i} [\Pr(T > t_i | \mathbf{x}_i) \Pr(U = t_i | \mathbf{x}_i)]^{1-\delta_i} \\ &= [f_T(t_i | \mathbf{x}_i) S_U(t_i | \mathbf{x}_i)]^{\delta_i} [S_T(t_i | \mathbf{x}_i) f_U(t_i | \mathbf{x}_i)]^{1-\delta_i} \\ &= [f_T(t_i | \mathbf{x}_i)^{\delta_i} S_T(t_i | \mathbf{x}_i)^{1-\delta_i}] [f_U(t_i | \mathbf{x}_i)^{1-\delta_i} S_U(t_i | \mathbf{x}_i)^{\delta_i}] \end{aligned}$$

where $S_T(t | \mathbf{x}_i) = \Pr(T > t | \mathbf{x}_i)$, $f_T(t | \mathbf{x}_i) = -dS_T(t | \mathbf{x}_i)/dt$, $S_U(t | \mathbf{x}_i) = \Pr(U > t | \mathbf{x}_i)$, and $f_U(t | \mathbf{x}_i) = -dS_U(t | \mathbf{x}_i)/dt$. In addition to the independent censoring assumption, we further impose the following assumption:

Non-informative censoring assumption: *The censoring distribution does not involve any parameters related to the distribution of the survival times. That is, $S_U(t | \mathbf{x}_i)$ does not contain parameters related to $S_T(t | \mathbf{x}_i)$.*

Under the non-informative censoring assumption, the term $f_U(t_i | \mathbf{x}_i)^{1-\delta_i} S_U(t_i | \mathbf{x}_i)^{\delta_i}$ is unrelated to the likelihood for the survival times and can simply be ignored. Therefore, the likelihood function is re-defined as

$$L = \prod_{i=1}^n f_T(t_i | \mathbf{x}_i)^{\delta_i} S_T(t_i | \mathbf{x}_i)^{1-\delta_i} = \prod_{i=1}^n h_T(t_i | \mathbf{x}_i)^{\delta_i} \exp[-H_T(t_i | \mathbf{x}_i)],$$

where $h_T(t | \mathbf{x}_i) = f_T(t | \mathbf{x}_i) / S_T(t | \mathbf{x}_i)$ and $H_T(t | \mathbf{x}_i) = \int_0^t h_T(u | \mathbf{x}_i) du$. The log-likelihood is

$$\ell = \log L = \sum_{i=1}^n [\delta_i \log h_T(t_i | \mathbf{x}_i) - H_T(t_i | \mathbf{x}_i)]. \quad (2.4)$$

Usually, censoring is non-informative if it is independent. Only an artificial or unusual example yields informative but independent censoring (p. 150 of Andersen et al. 1993; p. 196 of Kalbfleisch and Prentice 2002). It is well-known that independent censoring is more crucial assumption than non-informative censoring that does not lead to bias in estimation. Throughout the book, we focus on dependent censoring rather than informative censoring.

If censoring is dependent, the likelihood for the i -th patient is

$$\begin{aligned} L_i &= \Pr(T = t_i, U > t_i | \mathbf{x}_i)^{\delta_i} \Pr(T > t_i, U = t_i | \mathbf{x}_i)^{1-\delta_i} \\ &= \left\{ -\frac{\partial}{\partial x} \Pr(T > x, U > t_i | \mathbf{x}_i) \Big|_{x=t_i} \right\}^{\delta_i} \left\{ -\frac{\partial}{\partial y} \Pr(T > t_i, U > y | \mathbf{x}_i) \Big|_{y=t_i} \right\}^{1-\delta_i}. \end{aligned}$$

Therefore, the log-likelihood is defined as

$$\ell = \sum_{i=1}^n [\delta_i \log h_T^\#(t_i|\mathbf{x}_i) + (1 - \delta_i) \log h_U^\#(t_i|\mathbf{x}_i) - \Phi(t_i, t_i|\mathbf{x}_i)],$$

where

$$h_T^\#(t_i|\mathbf{x}_i) = - \frac{\partial}{\partial x} \log \Pr(T > x, U > t_i|\mathbf{x}_i) \Big|_{x=t_i},$$

$$h_U^\#(t_i|\mathbf{x}_i) = - \frac{\partial}{\partial y} \log \Pr(T > t_i, U > y|\mathbf{x}_i) \Big|_{y=t_i},$$

are the *cause-specific hazard functions*, and

$$\Phi(t_i, t_i|\mathbf{x}_i) = - \log \Pr(T > t_i, U > t_i|\mathbf{x}_i) = - \log \Pr(\min\{ T, U \} > t_i|\mathbf{x}_i)$$

is the cumulative hazard function for $\min\{ T, U \}$.

Suppose that the log-likelihood is parameterized by $\boldsymbol{\varphi}$. Then, the maximum likelihood estimator (MLE) is defined by maximizing the log-likelihood, $\hat{\boldsymbol{\varphi}} = \arg \max_{\boldsymbol{\varphi}} \ell(\boldsymbol{\varphi})$. To find the MLE numerically, one can use the *score function* $\mathbf{S}(\boldsymbol{\varphi}) = \partial \ell(\boldsymbol{\varphi}) / \partial \boldsymbol{\varphi}$ and the Hessian matrix $H(\boldsymbol{\varphi}) = \partial^2 \ell(\boldsymbol{\varphi}) / \partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'$. The MLE $\hat{\boldsymbol{\varphi}}$ is obtained from the Newton–Raphson algorithm

$$\boldsymbol{\varphi}^{(k+1)} = \boldsymbol{\varphi}^{(k)} - H^{-1}(\boldsymbol{\varphi}^{(k)}) \mathbf{S}(\boldsymbol{\varphi}^{(k)}), \quad k = 0, 1, \dots$$

Interval estimates for $\boldsymbol{\varphi}$ follow from the asymptotic theory of MLEs. The *information matrix* is defined as $i(\hat{\boldsymbol{\varphi}}) = -H(\hat{\boldsymbol{\varphi}})$. The SE for $\hat{\varphi}_j$ (the j -th component of $\hat{\boldsymbol{\varphi}}$) is $SE(\hat{\varphi}_j) = \sqrt{\{i^{-1}(\hat{\boldsymbol{\varphi}})\}_{jj}}$ that uses the j -th diagonal element of the inverse information matrix. The 95% CI is $\hat{\varphi}_j \pm 1.96 \times SE(\hat{\varphi}_j)$.

For instance, the Cox model takes the form $\boldsymbol{\varphi} = (\boldsymbol{\theta}, \boldsymbol{\beta})$ and $h(t|\mathbf{x}_i) = h_0(t; \boldsymbol{\theta}) \exp(\boldsymbol{\beta}'\mathbf{x}_i)$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ is a vector of parameters related to the baseline hazard function. We assume that the baseline cumulative hazard function $H_0(t; \boldsymbol{\theta})$ is an increasing step function with jumps $dH_0(t; \boldsymbol{\theta}) = e^{\theta_j}$ at $t = t_i$ with $\delta_i = 1$. Hence, the number of parameters in $\boldsymbol{\theta}$ is equal to the number of deaths $m = \sum_{i=1}^n \delta_i$. The MLE $\hat{\boldsymbol{\varphi}} = (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$ is obtained from the Newton–Raphson algorithm. It has been shown that $\hat{\boldsymbol{\beta}}$ is equivalent to the partial likelihood estimator and $\hat{\boldsymbol{\theta}}$ is the Breslow estimator $h_0(t_j; \hat{\boldsymbol{\theta}}) = e^{\hat{\theta}_j} = \left(\sum_{t_i \geq t_j} e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_i} \right)^{-1}$ (van der Vaart 1998; van Houwelingen and Putter 2011).

2.8 Technical Notes

Readers can skip this section as it does not influence the understanding of the latter chapters of the book.

The log-rank test possesses an easy-to-understand optimality criterion. The log-rank test is *asymptotically efficient (most powerful)* to detect the constant hazard ratio $h(t|x_i = 1)/h(t|x_i = 0) = \psi$ for some $\psi \neq 1$. Any reasonable test, such as the t -test, has optimality criteria to detect some specific form. The details on the asymptotic efficiency are referred to Andersen et al. (1993) and Fleming and Harrington (1991).

If the form of $h(t|x_i = 1)/h(t|x_i = 0)$ is non-constant, then the log-rank test may be sub-optimal. For example, Gehan's generalized Wilcoxon test statistic (Gehan 1965) defined as

$$S = \sum_{i=1}^n \delta_i (n_{i0} + n_{i1}) \left(x_i - \frac{n_{i1}}{n_{i0} + n_{i1}} \right)$$

can be more powerful than the log-rank statistic if the ratio $h(t|x_i = 1)/h(t|x_i = 0)$ strongly deviates from 1 in the early stage of follow-up. The generalized Wilcoxon test statistic is a special case of the *weighted log-rank statistics* (Fleming and Harrington 1991; Klein and Moeschberger 2003). If there is a concern about the non-constant hazard ratio, the weighted log-rank statistics may be employed.

A gross misunderstanding is that the log-rank test is a test tailored to detect the effect in a proportional hazards assumption. As mentioned earlier, the log-rank statistic is a non-parametric test to detect excess mortality without any model assumption.

We have derived the Kaplan–Meier estimator and the log-rank test under the assumption that all times to death are distinct (no ties). To handle ties, it is useful to introduce counting process formulations (Andersen et al. 1993; Fleming and Harrington 1991). For $k = 0, 1$, let $\bar{Y}_k(t) = \sum_{\ell=1}^n \mathbf{I}\{t_\ell \geq t, x_\ell = k\}$ be the number at-risk at time t , and let $\bar{N}_k(t) = \sum_{\ell=1}^n \mathbf{I}\{t_\ell \leq t, \delta_\ell = 1, x_\ell = k\}$ be the number of deaths up to time t . Then, at time t , the number of deaths in male is $d\bar{N}_1(t) = \sum_{\ell=1}^n \mathbf{I}\{t_\ell = t, \delta_\ell = 1, x_\ell = 1\}$, and the total number of deaths is $d\bar{N}(t) = \sum_{\ell=1}^n \mathbf{I}\{t_\ell = t, \delta_\ell = 1\}$.

The Kaplan–Meier estimator for the group k is defined as

$$\hat{S}_k(t) = \prod_{u \leq t} \{1 - d\hat{H}_k(u)\}, \quad k = 0, 1,$$

where $d\hat{H}_k(t) = d\bar{N}_k(t)/\bar{Y}_k(t)$ is called the *Nelson–Aalen estimator*.

The conditional distribution of $d\bar{N}_1(t)$ given $(d\bar{N}(t), \bar{Y}_0(t), \bar{Y}_1(t))$ is a hypergeometric distribution with mean

$$E\{ d\bar{N}_1(t) | d\bar{N}(t), \bar{Y}_0(t), \bar{Y}_1(t) \} = \frac{d\bar{N}(t)\bar{Y}_1(t)}{\bar{Y}_0(t) + \bar{Y}_1(t)}.$$

Consequently, the aggregated differences between the observed and expected deaths is

$$S = \int_0^\infty \left[d\bar{N}_1(t) - \frac{d\bar{N}(t)\bar{Y}_1(t)}{\bar{Y}_0(t) + \bar{Y}_1(t)} \right] = \int_0^\infty d\bar{N}_1(t) - \int_0^\infty \frac{d\bar{N}(t)\bar{Y}_1(t)}{\bar{Y}_0(t) + \bar{Y}_1(t)}.$$

The univariate partial likelihood estimator as derived in Eq. (2.3) has a counting process form

$$\exp(\hat{\beta}) = \frac{\int_0^\infty W(\hat{\beta}; t) d\hat{H}_1(t)}{\int_0^\infty W(\hat{\beta}; t) d\hat{H}_0(t)}, \quad W(t; \beta) = \frac{\bar{Y}_0(t)\bar{Y}_1(t)}{\bar{Y}_0(t) + \bar{Y}_1(t) \exp(\beta)}.$$

This means that the estimator is the ratio of the expected number of deaths in male divided by the expected number of deaths in female. This way of interpreting the univariate estimator is suggested in Emura and Chen (2016) to argue the robustness of the estimator against the model misspecification. Under the independent censoring assumption, $\hat{\beta}$ is a consistent estimator for β^* that is the solution to

$$\exp(\beta) = \frac{\int_0^\infty w(\beta; t) h(t|x=1) dt}{\int_0^\infty w(\beta; t) h(t|x=0) dt}, \quad W(t; \beta) = \frac{\pi_0(t)\pi_1(t)}{\pi_0(t) + \pi_1(t) \exp(\beta)},$$

where $\pi_k(t) = \lim_{n \rightarrow \infty} \bar{Y}_k(t)/n$ and the integral is on the range of t with $\pi_0(t)\pi_1(t) > 0$. If the proportional hazards model $h(t|x_i = 1) = \exp(\beta_0)h(t|x_i = 0)$ holds for some β_0 , then $\beta^* = \beta_0$. Even if the proportional hazards model does not hold, β^* is still meaningful since $\exp(\beta^*)$ is interpreted as the RR. However, the interpretation of the partial likelihood estimator may not be robust against the violation of the independent censoring assumption (Chap. 3).

2.9 Exercises

1. *Deriving the Kaplan–Meier estimator:* Consider a survival function $S(t) = \Pr(T > t)$ that is a decreasing step function with steps at observed times of death. Assume that all the observed times to death are distinct ($t_i \neq t_j$ whenever $i \neq j$ and $\delta_i = \delta_j = 1$).

(1) Show $\Pr(T > t_i) = \Pr(T > t_i | T > t_{i-1}) \Pr(T > t_{i-1})$ if $t_i > t_{i-1}$.

(2) Show $\Pr(T > t_j) = \prod_{i=1}^j \Pr(T > t_i | T > t_{i-1})$ if $t_j > t_{j-1} > \dots > t_1 > t_0 \equiv 0$ and $S(0) = 1$.

(3) Show $\Pr(T > t_i | T > t_{i-1}) = 1 - \Pr(T = t_i | T \geq t_i)$ if there is no death in the interval (t_{i-1}, t_i) .

(4) Show $S(t_j) = \prod_{i=1}^j \left(1 - \frac{\Pr(T=t_i)}{\Pr(T \geq t_i)}\right)$.

2. *Weibull regression*: Let $\log(T_i) = \alpha_0 + \alpha' \mathbf{x}_i + \sigma \varepsilon_i$, where $\Pr(\varepsilon_i > x) = \exp(-e^x)$ for $-\infty < x < \infty$.

(1) Derive the survival function $S(t|\mathbf{x}_i)$ and the hazard function $h(t|\mathbf{x}_i)$.

(2) Show that the model can be expressed as $h(t|\mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i)$.

(3) Show $\Pr(T > t + w | T > t, \mathbf{x}_i) < \Pr(T > w | \mathbf{x}_i)$ for $0 < \sigma < 1$ and $w > 0$. What does this inequality imply?

3. Consider a discrete random vector $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$ whose distribution is given by

$$\Pr(\mathbf{X}_i = \mathbf{x}_k) = \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_k)}{\sum_{\ell \in R_i} \exp(\boldsymbol{\beta}' \mathbf{x}_\ell)}, \quad k \in R_i = \{\ell : t_\ell \geq t_i\}, \quad i = 1, \dots, n.$$

This represents the risk of the k -th patient relative to the total risk for those who are at-risk of death at time t_i . By assuming the independence of the sequence \mathbf{X}_i , $i = 1, \dots, n$, one can obtain the partial likelihood function $L(\boldsymbol{\beta}) = \prod_{i=1}^n \Pr(\mathbf{X}_i = \mathbf{x}_i)^{\delta_i}$.

(1) Express the score function $\mathbf{S}(\boldsymbol{\beta})$ using $E(\mathbf{X}_i)$.

(2) Express the Hessian matrix $H(\boldsymbol{\beta})$ using $\text{Var}(\mathbf{X}_i)$.

(3) Discuss the conditions to make $H(\boldsymbol{\beta})$ negative definite.

4. Suppose that data (t_i, δ_i, x_i) , $i = 1, \dots, n$, follow the model $S(t|x_i) = \exp(-\lambda t e^{\beta x_i})$, where $\lambda > 0$ and $-\infty < \beta < \infty$. Let $m = \sum_{i=1}^n \delta_i$ be the number of deaths.

(1) Write down the log-likelihood function $\ell(\lambda, \beta) = \log L(\lambda, \beta)$.

(2) Derive the score functions $\partial \ell(\lambda, \beta) / \partial \lambda$ and $\partial \ell(\lambda, \beta) / \partial \beta$.

(3) Derive the fixed-point iteration algorithm and apply it to the data of Example 1.

(4) Derive the Hessian matrix of $\ell(\lambda, \beta)$.

(5) Derive the Newton–Raphson algorithm and apply it to the data of Example 1.

(6) Derive the Newton–Raphson algorithm under the transformed parameter $\tilde{\lambda} = \log(\lambda)$ and apply it to the data of Example 1.

(7) Compare the numbers of iterations in all the three algorithms.

5. Use the lung cancer data available in the *compound.Cox* R package (Emura et al. 2018) to:
- (1) Perform univariate Cox regression treating the *ZNF264* gene or the *NF1* gene as a covariate. Are these genes univariately associated with survival?
 - (2) Perform multivariate Cox regression treating both the *ZNF264* and *NF1* genes as covariates. Are these genes associated with survival?
 - (3) Discuss the influence of multicollinearity between *ZNF264* and *NF1*.

References

- Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer, New York
- Collett D (2003) Modelling survival data in medical research, 2nd edn. CRC Press, London
- Cox DR (1972) Regression models and life-tables (with discussion). *J R Stat Soc Ser B Stat Methodol* 34:187–220
- Emura T, Chen YH (2016) Gene selection for survival data under dependent censoring, a copula-based approach. *Stat Methods Med Res* 25(6):2840–2857
- Emura T, Chen HY, Matsui S, Chen YH (2018) *Compound.Cox*: univariate feature selection and compound covariate for predicting survival. CRAN
- Fleming TR, Harrington DP (1991) Counting processes and survival analysis. Wiley
- Gehan EA (1965) A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52:203–224
- Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data, 2nd edn. Wiley, New York
- Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53(282):457–481
- Klein JP, Moeschberger ML (2003) Survival analysis techniques for censored and truncated data. Springer, New York
- Michiels S, Le Maître A, Buyse M, Burzykowski T, Maillard E, Bogaerts J, Pignon JP (2009) Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data. *Lancet Oncol* 10(4):341–350
- van der Vaart AW (1998) Asymptotic statistics. Cambridge series in statistics and probabilistic mathematics. Cambridge University Press, Cambridge
- van Houwelingen HC, Putter H (2011) Dynamic prediction in clinical survival analysis. CRC Press, New York