STAT3655 Survival Analysis

Yu Gu, PhD
Assistant Professor

Department of Statistics & Actuarial Science
The University of Hong Kong

Table of Contents

Table of Contents

Table of Contents

Survival Data

- Survival (failure) times: times to the occurrence of a given event (failure) measured from a well-defined starting point.
  - death times of patients enrolled in a clinical trial
  - times to the occurrence of a disease after exposing to a hazard material
  - ages at the onset of a genetically related disease
  - life times of machine components in industrial reliability
  - times to complete specified tasks in psychological experimentation

- Censoring: some subjects are not observed for the full time to failure.

- Sources of (right) censoring:
  - loss to follow-up
  - alive at the end of the study
  - death from other causes

Right-Censored Survival Data

- Full data of the ith subject:
  - failure time: T_i
  - censoring time: C_i
  - covariates: $X_i = (X_{i1}, \ldots, X_{ip})^{\mathsf{T}}$
- Observed data of the ith subject:
  - observation time: $Y_i = \min(T_i, C_i)$
  - failure indicator: $\delta_i = I(T_i \leq C_i)$
  - covariates: $X_i = (X_{i1}, \ldots, X_{ip})^{\mathsf{T}}$

Survival Analysis

- Problems
  - Estimating the failure time distribution
  - Testing the equality of two failure time distributions
  - Estimating the effects of covariates on the failure time
- Naive methods
  - Ignore censoring statuses: the failure times associated with censored observations are underrepresented.
  - Delete censored cases: estimates are biased towards smaller failure times because larger failure times are more likely to be censored.
- The need to accommodate censoring has been a key motivating factor for the development of specialized statistical methods for analyzing failure time data.

Example: Carcinogenesis

Table 1.1: Days to vaginal cancer mortality in rats

Group 1	143	164	188	188	190	192	206
	209	213	216	220	227	230	234
	246	265	304	216*	244*		
Group 2	142	156	163	198	205	232	232
	233	233	233	233	239	240	261
	280	280	296	296	323	204*	344*

* Right-censored failure times.

In this example, the (right) censoring may have arisen because these four rats died of causes unrelated to carcinogen, or they may not have died by the time of data analysis.

Example: Randomized Clinical Trial

Table 1.2: Days to severe AGVHD in bone marrow transplant patients

CSP + MTX						MTX					
Time	LAF	Age	Time	LAF	Age	Time	LAF	Age	Time	LAF	Age
3*	0	40	324*	0	23	9	1	35	104*	1	27
8	1	21	356*	1	13	11	1	27	106*	1	19
10	1	18	378*	1	34	12	0	22	156*	1	15
12*	0	42	408*	1	27	20	1	21	218*	1	26
16	0	23	411*	1	5	20	1	30	230*	0	11
17	0	21	420*	1	23	22	0	7	231*	1	14
22	1	13	449*	1	37	25	1	36	316*	1	15
64*	0	20	490*	1	37	25	1	38	393*	1	27
65*	1	15	528*	1	32	25*	0	20	395*	0	2
77*	1	34	547*	1	32	28	0	25	428*	0	3
82*	1	14	691*	1	38	28	0	28	469*	1	14
98*	1	10	769*	0	18	31	1	17	602*	1	18
155*	0	27	1111*	0	20	35	1	21	681*	0	23
189*	1	9	1173*	0	12	35	1	25	690*	1	9
199*	1	19	1213*	0	12	46	1	35	1112*	1	11
247*	1	14	1357*	0	29	49	0	19	1180*	0	11

[a] Asterisks indicate that time to severe AGVHD is right censored; that is, the patient died without severe AGVHD or was without severe AGVHD at last contact.

Failure Time Distributions

The failure time T is a continuous nonnegative random variable. We are interested in the following quantities:

Distribution function: $F(t) = \Pr(T \leq t)$

Density function: $f(t) = dF(t)/dt$

Survival function: $S(t) = \Pr(T > t) = 1 - F(t)$

Hazard function: instantaneous risk or rate of failure

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t \mid T \geq t) = f(t)/S(t)$$

Cumulative hazard function: $\Lambda(t) = \int_0^t \lambda(u)du$

Failure Time Distributions (Cont.)

We can easily verify the following relationships:

(i) $S(t) = \int_t^\infty f(u)du$

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

(iii) $\lambda(t) = -\frac{d \log S(t)}{dt}$

(iv) $S(t) = \exp\{-\Lambda(t)\}$

(v) $f(t) = \lambda(t) \exp\{-\Lambda(t)\}$

Independent Censoring

- Recall that T_i and C_i are the ith subject's failure time and censoring time, respectively, and $Y_i = \min(T_i, C_i)$ is the observation time.
- A right-censoring mechanism is said to be independent if subjects censored at each time $t > 0$ are "representative" of the subjects under observation. In other words, subjects cannot be censored because they appear to be at an unusually high or low risk of failure.
- Usually, this means that T_i and C_i are statistically independent given the covariates X_i.
- Let F, G and H be the distribution functions of T_i, C_i and Y_i, respectively. Under the independent censoring assumption, they satisfy $1 - H = (1 - F)(1 - G)$.

Other Types of Censoring

Although the main focus of this course is right-censored failure time data, it's good to know other types of censoring.

Left censoring: a subject is observed to fail prior to some time t, but the actual failure time is unknown.

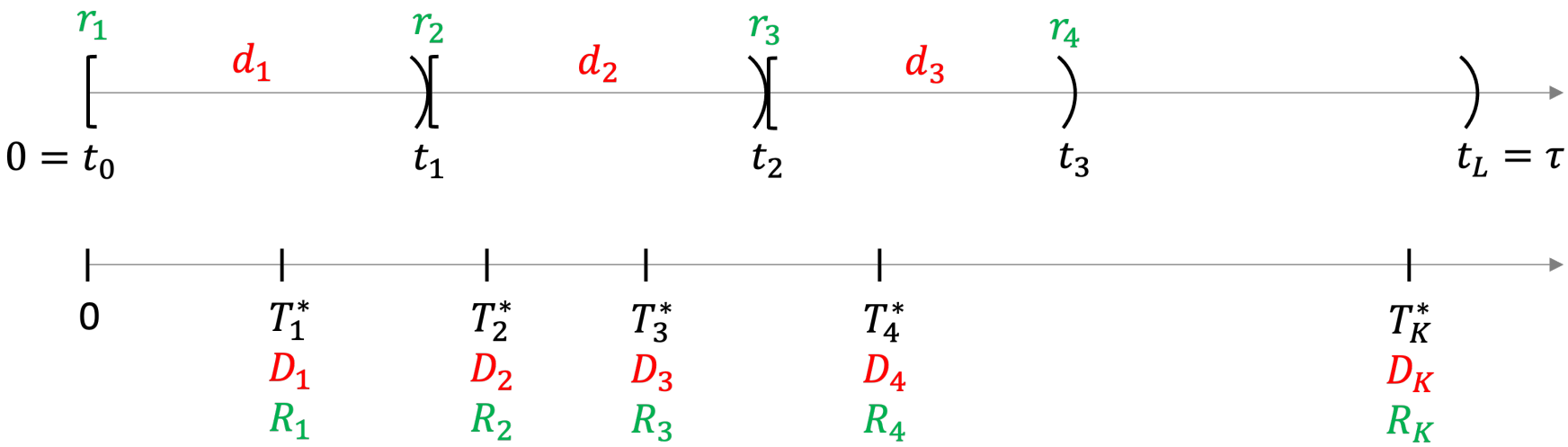
Interval censoring: failure time is only known to fall within some interval (a, b).

Both left censoring and right censoring are special cases of interval censoring.

- For right censoring, a is the observation time while $b = \infty$.
- For left censoring, $a = 0$ while b is the observation time.

Table of Contents

Notation



- $0 = t_0 < t_1 < \cdots < t_L = \tau$: partition of the study period $[0, \tau]$
- d_l: number of failures within the interval $[t_{l-1}, t_l)$, $l = 1, \ldots, L$
- r_l: number of subjects at risk (i.e., under observation at the previous instant) at t_{l-1}
- $T_1^* < T_2^* < \cdots < T_K^*$: distinct time points of observed failures
- D_k: number of failures at T_k^*, $k = 1, \ldots, K$
- R_k: number of subjects at risk at T_k^*, $k = 1, \ldots, K$

Nelson-Aalen Estimator

- When Δt is small enough, we have

$$\begin{aligned}\Lambda(t+\Delta t)-\Lambda(t) &\approx \lambda(t)\Delta t \\ &\approx \Pr(t \le T < t+\Delta t \mid T \ge t) \\ &= \Pr(t \le T < t+\Delta t \mid \underbrace{T \ge t,\, C \ge t}_{Y \ge t})\end{aligned}$$

- Thus,

$$\Lambda(t_l)-\Lambda(t_{l-1}) \approx \Pr(t_{l-1} \le T < t_l \mid Y \ge t_{l-1}) \approx d_l/r_l$$

- A natural nonparametric estimator of $\Lambda(t)$ is

$$\widehat{\Lambda}(t) = \sum_{l:t_l \le t} d_l/r_l \to \sum_{k:T_k^* \le t} D_k/R_k \qquad \text{(Nelson-Aalen Estimator)}$$

as $L \to \infty$ and $\max_{1 \le l \le L} |t_l - t_{l-1}| \to 0$.

Nelson-Aalen Estimator (Cont.)

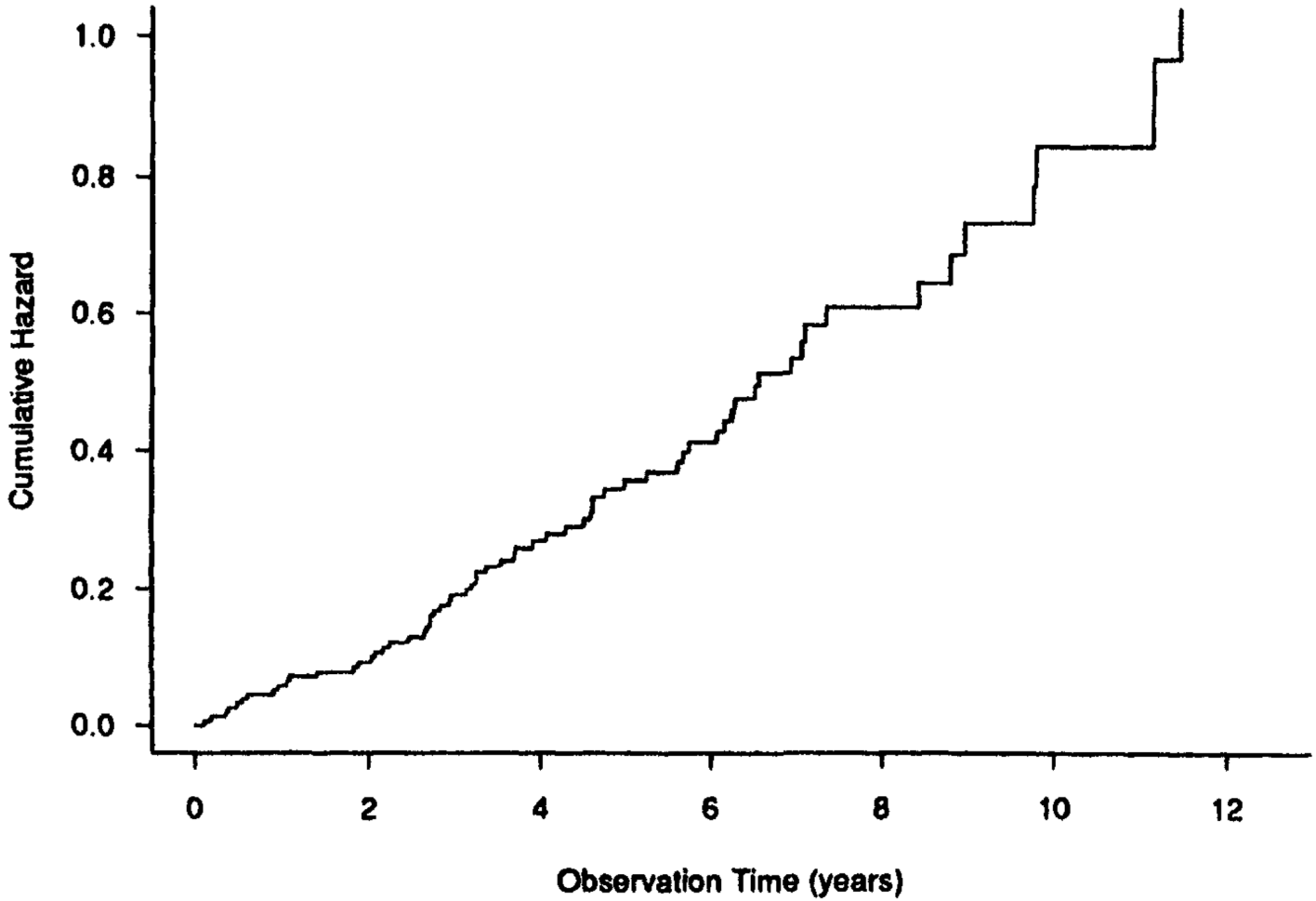


Figure 1.1: Example of the Nelson-Aalen estimator

Kaplan-Meier Estimator

- To estimate the survival function, note that

$$\begin{aligned} & \Pr(T > t_l) \\ = & \Pr(T > t_1 \mid T > t_0)\Pr(T > t_2 \mid T > t_1) \cdots \Pr(T > t_l \mid T > t_{l-1}) \end{aligned}$$

- Thus,

$$\begin{aligned} S(t) &\approx \prod_{l:t_l \le t} \Pr(T \ge t_l \mid T \ge t_{l-1}) \\ &= \prod_{l:t_l \le t} \{1 - \Pr(T < t_l \mid T \ge t_{l-1})\} \end{aligned}$$

- A natural nonparametric estimator of $S(t)$ is

$$\widehat{S}(t) = \prod_{l:t_l \le t} (1 - d_l/r_l) \to \prod_{k:T_k^* \le t} (1 - D_k/R_k) \quad \text{(Kaplan-Meier Estimator)}$$

as $L \to \infty$ and $\max_{1 \le l \le L} |t_l - t_{l-1}| \to 0$.

Kaplan-Meier Estimator (Cont.)

- We can easily observe that $\widehat{S}(T_k^*) = \widehat{S}(T_{k-1}^*)(1 - D_k/R_k)$.
- For uncensored data, $\widehat{S}$ reduces to one minus the empirical distribution function.
- An alternative estimator for $S(t)$ is $\widetilde{S}(t) = e^{-\widehat{\Lambda}(t)}$, which is close to $\widehat{S}(t)$ under certain condition.

Kaplan-Meier Estimator (Cont.)

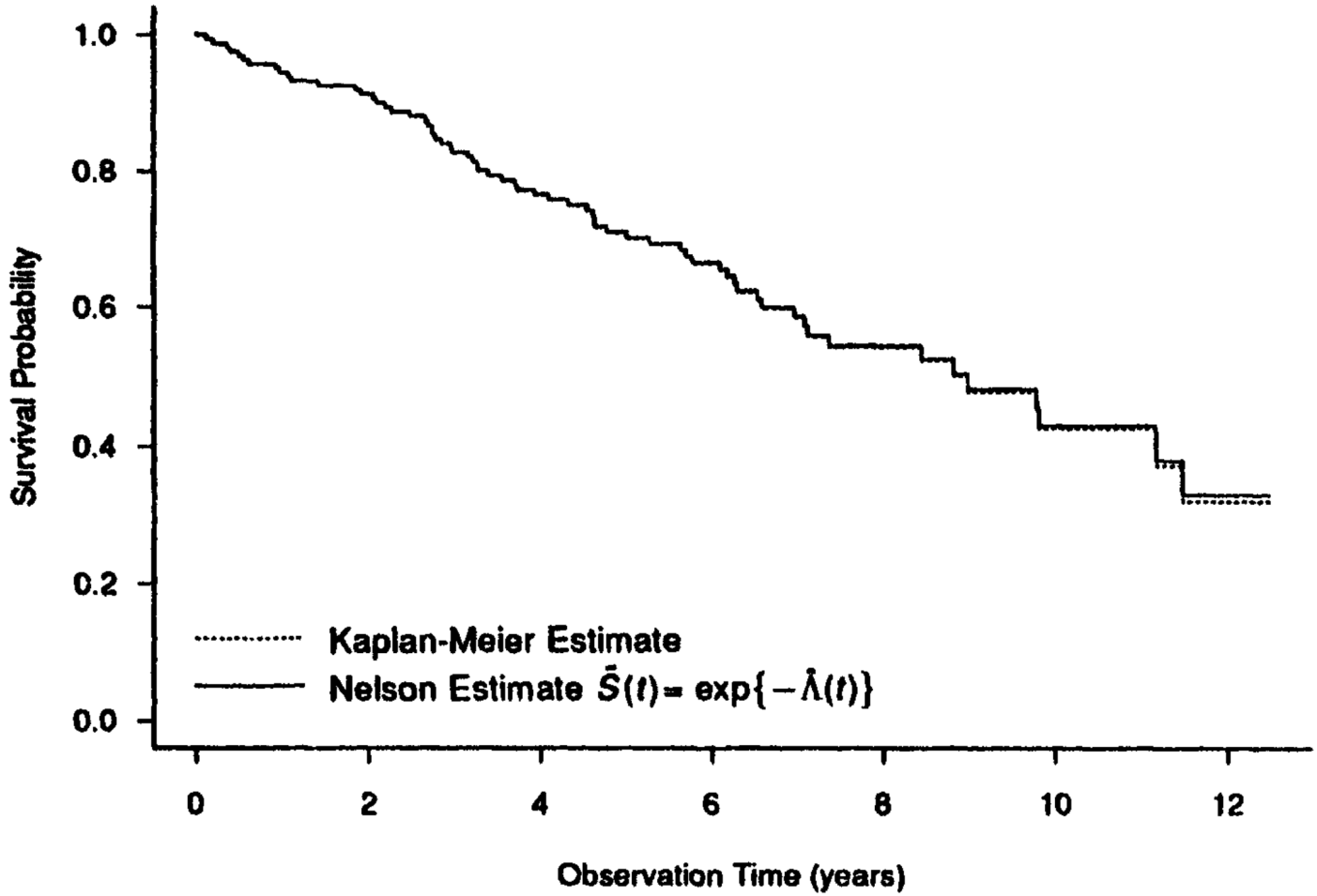


Figure 1.2: Example of the Kaplan-Meier estimator

Exercise

Calculate the Kaplan-Meier estimator for the following data:

$$1 \quad 2 \quad 2 \quad 4^* \quad 5^* \quad 6 \quad 7^* \quad 8^* \quad 9^* \quad 10^*$$

Exercise

Calculate the Kaplan-Meier estimator for the following data:

$$1 \quad 2 \quad 2 \quad 4^* \quad 5^* \quad 6 \quad 7^* \quad 8^* \quad 9^* \quad 10^*$$

T_k^*	D_k	R_k	$1 - \frac{D_k}{R_k}$	$\widehat{S}(T_k^*)$
1	1	10	$1 - \frac{1}{10} = \frac{9}{10}$	$\frac{9}{10}$
2	2	9	$1 - \frac{2}{9} = \frac{7}{9}$	$\frac{9}{10} \times \frac{7}{9} = \frac{7}{10}$
6	1	5	$1 - \frac{1}{5} = \frac{4}{5}$	$\frac{7}{10} \times \frac{4}{5} = \frac{14}{25}$

Exercise

Calculate the Kaplan-Meier estimator for the following data:

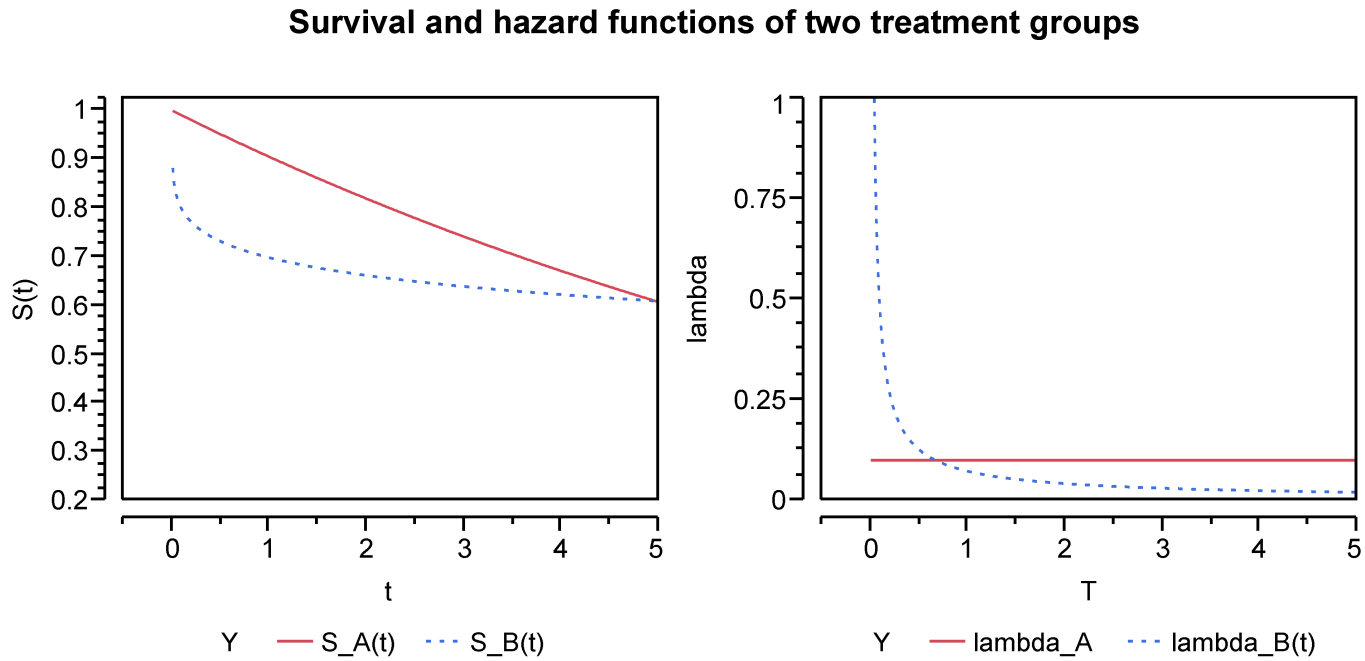
$$1 \quad 2 \quad 2 \quad 4^* \quad 5^* \quad 6 \quad 7^* \quad 8^* \quad 9^* \quad 10^*$$

T_k^*	D_k	R_k	$1 - \frac{D_k}{R_k}$	$\widehat{S}(T_k^*)$
1	1	10	$1 - \frac{1}{10} = \frac{9}{10}$	$\frac{9}{10}$
2	2	9	$1 - \frac{2}{9} = \frac{7}{9}$	$\frac{9}{10} \times \frac{7}{9} = \frac{7}{10}$
6	1	5	$1 - \frac{1}{5} = \frac{4}{5}$	$\frac{7}{10} \times \frac{4}{5} = \frac{14}{25}$

$$\widehat{S}(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1 \\ \frac{9}{10} & \text{if } 1 \leq t < 2 \\ \frac{7}{10} & \text{if } 2 \leq t < 6 \\ \frac{14}{25} & \text{if } t \geq 6 \end{cases}$$

Table of Contents

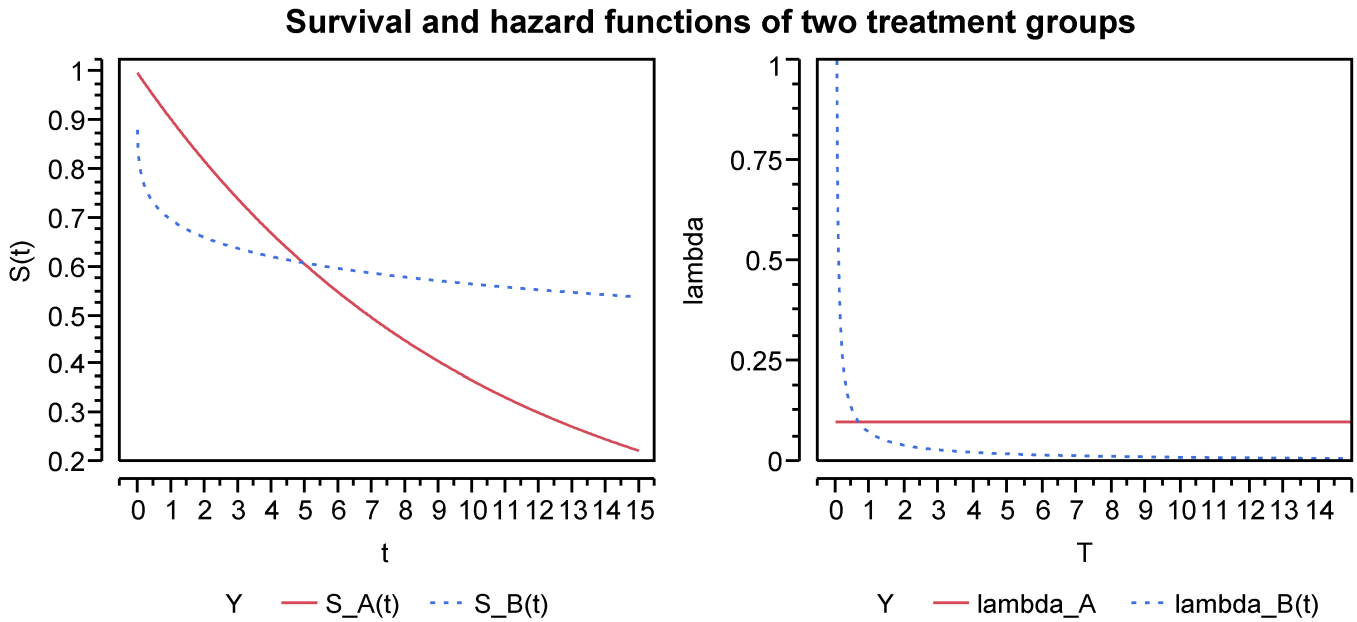
Simple Graphical Comparison



Based solely on the above plots, we would conclude that treatment A yields superior survival compared to treatment B. The hazard rate of group A is initially lower than that of group B, but subsequently becomes higher than that of group B.

Simple Graphical Comparison (Cont.)

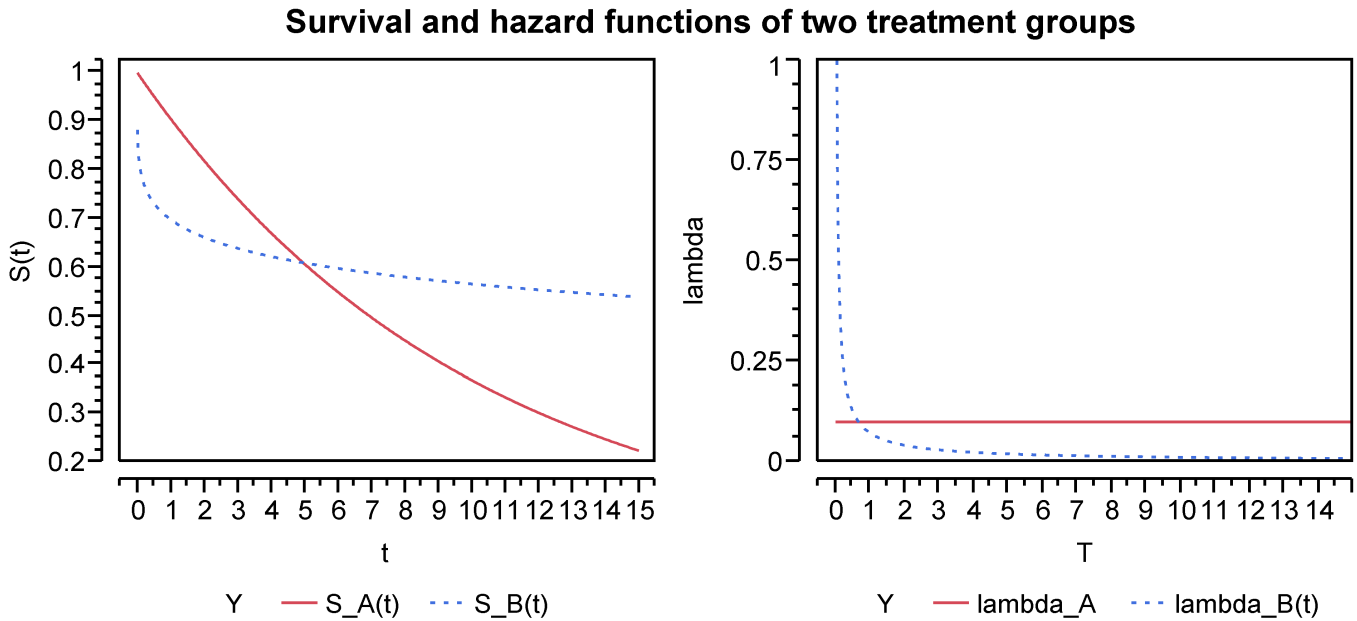
What if we extend the follow-up period?



It turns out that at 15 time units, treatment B yields a much higher survival probability compared to treatment A.

Simple Graphical Comparison (Cont.)

What if we extend the follow-up period?



It turns out that at 15 time units, treatment B yields a much higher survival probability compared to treatment A.

The failure time distribution seems to differ between the two treatment groups, but how can we formally test for this difference?

Log-Rank Test

- $H_0 : S_1(t) \equiv S_2(t)$
- $T_1^* < \cdots < T_K^*$: distinct time points of observed failures in combined sample
- Contingency table at T_k^* $(k = 1, \ldots, K)$:

	Failures	Non-failures	At risk
Group 1	D_{1k}	$R_{1k} - D_{1k}$	R_{1k}
Group 2	D_{2k}	$R_{2k} - D_{2k}$	R_{2k}
Total	D_k	$R_k - D_k$	R_k

- Under H_0, the conditional distribution of D_{1k} given (R_{1k}, R_{2k}, D_k) is Hypergeometric(R_k, R_{1k}, D_k), with conditional mean and variance

$$E_{1k} = \frac{D_k R_{1k}}{R_k}, \quad V_{1k} = \frac{D_k (R_k - D_k) R_{1k} R_{2k}}{R_k^2 (R_k - 1)}$$

Log-Rank Test (Cont.)

- Thus, the statistic $D_{1k} - E_{1k}$ has conditional mean 0 and variance V_{1k}.
- Summing over the K failure times yields the log-rank statistic

$$Q = \frac{\sum_{k=1}^{K}(D_{1k} - E_{1k})}{\sqrt{\sum_{k=1}^{K} V_{1k}}} \sim N(0,1)$$

- For the carcinogenesis data, $Q^2 = 3.12$, p-value $= 0.08$.

Group 1	143	164	188	188	190	192	206
	209	213	216	220	227	230	234
	246	265	304	216*	244*		
Group 2	142	156	163	198	205	232	232
	233	233	233	233	239	240	261
	280	280	296	296	323	204*	344*

Table of Contents

Exponential Distribution

- As before, $T \geq 0$ is a random variable representing failure time, and t represents a point in its range.

- The one-parameter exponential distribution is obtained by taking the hazard function to be constant, i.e., $\lambda(t) \equiv \lambda > 0$, for all $t \geq 0$.

- Properties:
  - (i) Survival function: $S(t) = e^{-\lambda t}$
  - (ii) Density function: $f(t) = \lambda e^{-\lambda t}$
  - (iii) Memoryless property: $\Pr(T \geq t_1 + t_2 \mid T \geq t_1) = \Pr(T \geq t_2)$
  - (iv) Expectation: $E(T) = 1/\lambda$
  - (v) Moment-generating function: $M_T(s) = E(e^{sT}) = \frac{\lambda}{\lambda - s}$ for $s < \lambda$
  - (vi) If $T \sim \text{Exp}(\lambda)$, then for $c > 0$, $cT \sim \text{Exp}(\lambda/c)$.
  - (vii) If T_1 and T_2 are independently distributed according to $\text{Exp}(\lambda_1)$ and $\text{Exp}(\lambda_2)$, respectively, then $\min(T_1, T_2) \sim \text{Exp}(\lambda_1 + \lambda_2)$.

Weibull Distribution

- An important generalization of the exponential distribution allows for a power dependence of the hazard on time. This yields the two-parameter Weibull distribution with hazard function

$$\lambda(t) = \lambda p(\lambda t)^{p-1}, \qquad \text{for } \lambda,\ p > 0.$$

- Properties:
  - (i) Survival function: $S(t) = \exp\{-(\lambda t)^p\}$
  - (ii) Density function: $f(t) = \lambda p(\lambda t)^{p-1} \exp\{-(\lambda t)^p\}$
  - (iii) $\lambda(t)$ is decreasing if $p < 1$, constant if $p = 1$, and increasing if $p > 1$.
  - (iv) If $T \sim \text{Weibull}(\lambda, p)$ and $U = T^p$, then $U \sim \text{Exp}(\lambda^p)$.

Weibull Distribution (Cont.)

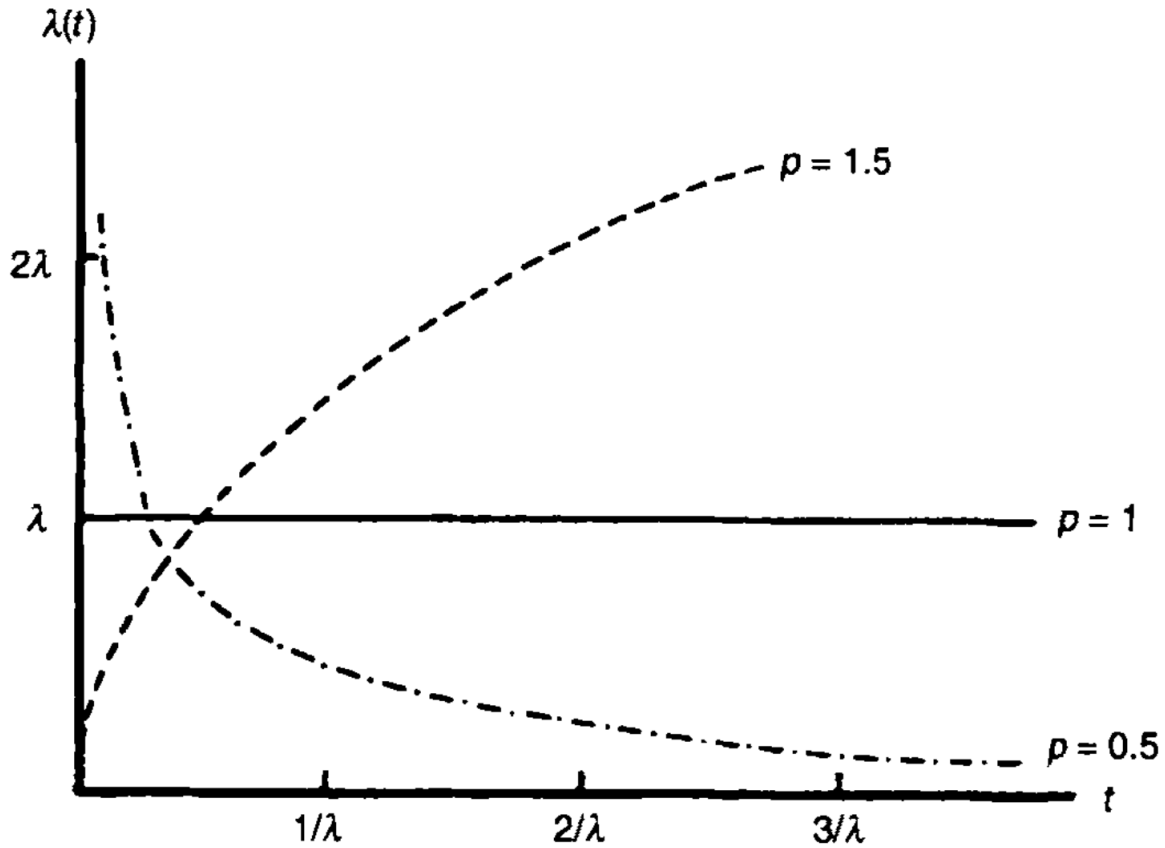


Figure 1.3: Hazard functions for Weibull models with different shape parameters

Gamma Distribution

- Another two-parameter generalization of the exponential model is the Gamma distribution with density function

$$f(t) = \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)}, \qquad \text{for } \lambda,\ k > 0.$$

- When $k = 1$, the Gamma distribution reduces to the exponential distribution.

- The moment-generating function is $M_T(s) = \{\lambda/(\lambda - s)\}^k$.

- Both the survival and hazard functions involve incomplete gamma functions and cannot be written down explicitly. This restricts the use of the Gamma distribution in practice.

Log-Normal Distribution

- The failure time T has a log-normal distribution if and only if

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

- The survival and hazard functions have no closed forms which involve $\Phi(\cdot)$, the distribution function of a standard normal random variable.

- The log-normal model is particularly simple to apply if there is no censoring, but with censoring the computations become more difficult.

Log-Logistic Distribution

- The failure time T has a log-logistic distribution if and only if $\log T = \mu + \sigma Z$, where Z follows the standard logistic distribution, with density and distribution functions

$$f_Z(z) = \frac{e^z}{(1+e^z)^2}, \quad F_Z(z) = \frac{e^z}{1+e^z}, \quad \text{for } -\infty < z < \infty.$$

- Let $\lambda = e^{-\mu}$ and $p = 1/\sigma$, then the density, survival and hazard functions of T are

$$f(t) = \frac{\lambda p(\lambda t)^{p-1}}{\{1+(\lambda t)^p\}^2}, \quad S(t) = \frac{1}{1+(\lambda t)^p}, \quad \lambda(t) = \frac{\lambda p(\lambda t)^{p-1}}{1+(\lambda t)^p}.$$

- Like the exponential and Weibull models, this model has simple algebraic expressions for the survival and hazard functions, thus is more convenient than the log-normal distribution in handling censored data, while providing a good approximation to it except in the extreme tails.

Gompertz Distribution

- The relationship in which the hazard function is an exponential function of the failure time T has been found to be descriptive in many investigations.
- This leads to the Gompertz hazard $\lambda(t) = \lambda e^{\beta t}$, for $\lambda, \beta > 0$.
- The survival and density functions of the Gompertz distribution are

$$S(t) = \exp\left\{\frac{\lambda}{\beta}(1 - e^{\beta t})\right\},$$

$$f(t) = \lambda e^{\beta t} \exp\left\{\frac{\lambda}{\beta}(1 - e^{\beta t})\right\}.$$

Table of Contents

Covariates

- In the previous section, we consider modeling the survival distribution of a homogeneous population.
- However, the failure time T usually depends on some covariates X. For example,
  - COVID-19 vaccination status may affect the time of infection.
  - Smoking may increase the risk of lung cancer.
  - Some studies show that patients with Type 2 diabetes are at an increased risk of dementia.
- Regression models allow us to evaluate the associations between T and X.

Exponential Regression Models

- The exponential distribution can be generalized to obtain a regression model by allowing the hazard rate to be a function of the covariates:

$$\lambda(t; X) = \lambda(X).$$

- Thus the hazard function for a given X is a constant characterizing an exponential failure time distribution.

- The function $\lambda(\cdot)$ can be parameterized in many ways, say

$$\lambda(t; X) = \lambda g(\beta^{\mathsf{T}} X),$$

where $\beta = (\beta_1, \ldots, \beta_p)^{\mathsf{T}}$ is a vector of regression parameters, λ is a positive constant, and $g(\cdot)$ is a specified non-negative functional form.

Exponential Regression Models (Cont.)

- The choice of g may depend on the particular data being considered. Some examples are:
  - (i) $g(u) = 1 + u$, which can be interpreted as the hazard rate being a linear function of X.
  - (ii) $g(u) = (1 + u)^{-1}$, which can be interpreted as the mean survival time being a linear function of X.
  - (iii) $g(u) = e^u$, which specifies that the log hazard rate is a linear function of X.

- Both (i) and (ii) suffer from the disadvantage that the set of β values considered must be restricted to guarantee that $g(\beta^{\mathsf{T}} X) > 0$.

- In many ways, (iii) is the most natural form since it takes only positive values. In this case, the hazard function of T given X is

$$\lambda(t; X) = \lambda \exp(\beta^{\mathsf{T}} X) \tag{1}$$

Weibull Regression Models

- The Weibull distribution can be generalized to the regression situation in essentially the same way. For example, the hazard function can be specified as

$$\lambda(t; X) = \lambda p(\lambda t)^{p-1} \exp(\beta^{\mathsf{T}} X) \tag{2}$$

- Under both models (1) and (2), the effect of the covariates is to act multiplicatively on the hazard function.

- A more general regression model of this feature is the Cox model.

Cox Model

- The Cox model specifies

$$\lambda(t; X) = \lambda_0(t) \exp(\beta^{\mathsf{T}} X), \tag{3}$$

  where $\lambda_0(\cdot)$ is an arbitrary unspecified baseline hazard function for continuous T. Thus, this model is semiparametric.

- It is easy to see that models (1) and (2) are both special cases of the Cox model by setting $\lambda_0(t) = \lambda$ and $\lambda_0(t) = \lambda p(\lambda t)^{p-1}$, respectively.

- The conditional survival and density functions of T given X are

$$S(t; X) = \exp\left\{-\exp(\beta^{\mathsf{T}} X) \int_0^t \lambda_0(u) du\right\} = \{S_0(t)\}^{\exp(\beta^{\mathsf{T}} X)},$$

$$f(t; X) = \lambda_0(t) \exp(\beta^{\mathsf{T}} X) \exp\left\{-\exp(\beta^{\mathsf{T}} X) \int_0^t \lambda_0(u) du\right\},$$

  where $S_0(t) = S(t; X = 0)$ is the baseline survival function.

Accelerated Failure Time Model

- To postulate a direct relationship between T and X, we consider another semiparametric class of log-linear models for T:

$$\log T = \beta^{\mathsf{T}} X + W,$$

where W is an error variable with unspecified density f.

- Exponentiation of the above model gives $T = \exp(\beta^{\mathsf{T}} X) Z$, where $Z = \exp(W) > 0$ has hazard function $\lambda_0(s)$ that is independent of β.

- It follows that the survival and hazard functions for T can be written in terms of this baseline hazard $\lambda_0(\cdot)$ according to

$$S(t; X) = \exp\left\{-\Lambda_0(te^{-\beta^{\mathsf{T}} X})\right\},$$
$$\lambda(t; X) = \lambda_0(te^{-\beta^{\mathsf{T}} X}) \exp(-\beta^{\mathsf{T}} X),$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$.

Cox versus AFT Models

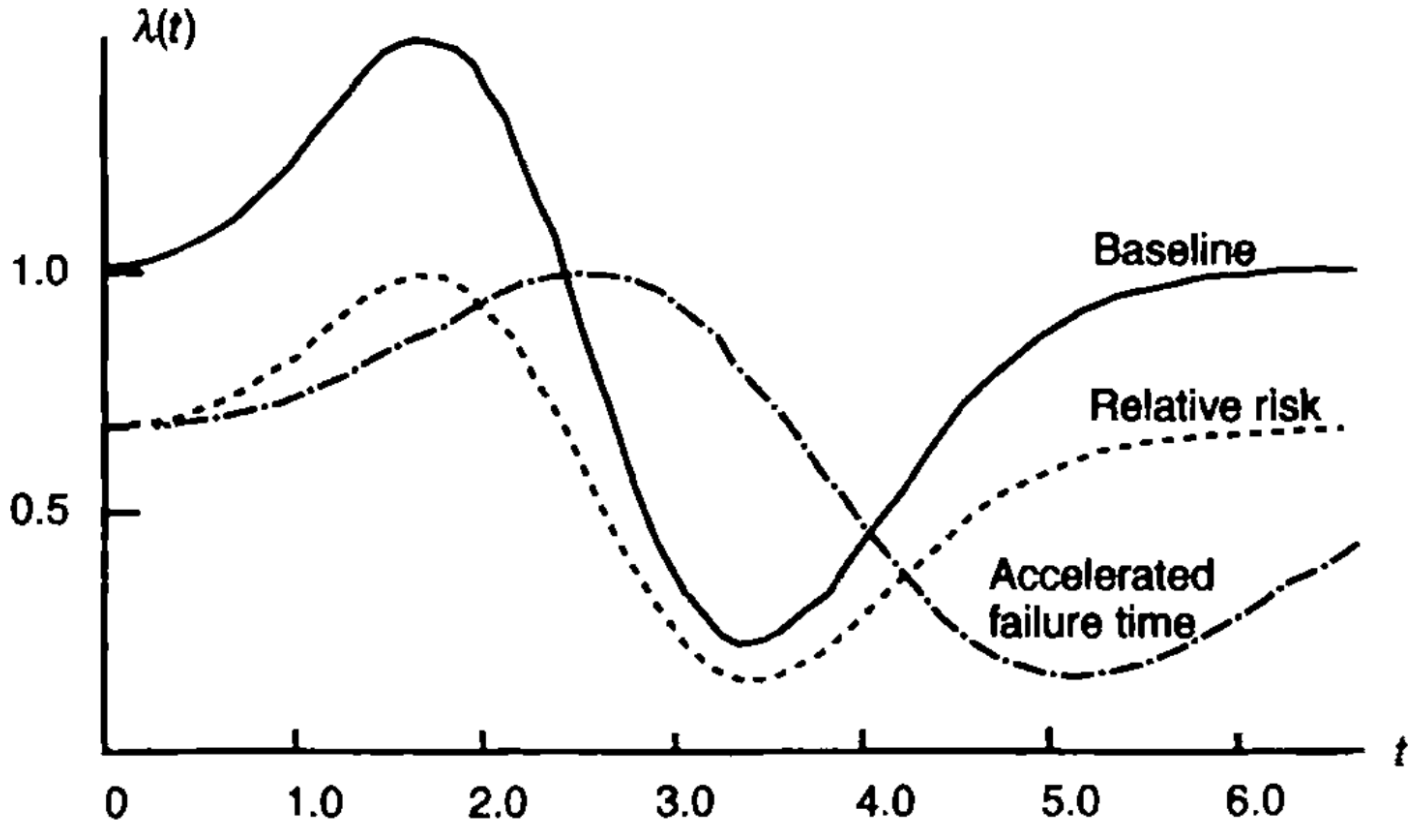


Figure 1.4: Baseline hazard function ($X = 0$) and hazard functions for $X = 1$ under the Cox and AFT models

The figure shows the proportional hazards (PH) property of the Cox model.

Table of Contents

Data and Likelihood

- Observed data:

$$\Big\{ Y_i = \min(T_i, C_i),\ \delta_i = I(T_i \leq C_i),\ X_i \Big\}, \quad i = 1, \ldots, n$$

- Model: any parametric (regression) model previously discussed
- Unknown parameters: θ
- Assumption: $T_i \perp\!\!\!\perp C_i \mid X_i$ (independent censoring)
- Likelihood contribution from the ith subject:
  - $f(Y_i; \theta)$ if $\delta_i = 1$
  - $S(Y_i; \theta)$ if $\delta_i = 0$
- Likelihood:

$$L_n(\theta) = \prod_{i=1}^{n} f(Y_i; \theta)^{\delta_i} S(Y_i; \theta)^{1-\delta_i}$$

- Log-likelihood:

$$\ell_n(\theta) = \sum_{i=1}^{n} \Big\{ \delta_i \log f(Y_i; \theta) + (1 - \delta_i) \log S(Y_i; \theta) \Big\}$$

Maximum Likelihood Estimation

- Let $\dot{\ell}_n(\theta)$ and $\ddot{\ell}_n(\theta)$ be the gradient and Hessian of $\ell_n(\theta)$ with respect to θ, respectively.

- The maximum likelihood estimator (MLE) for θ solves the score equation $\dot{\ell}_n(\theta) = 0$.

- If the solution to the score equation does not have an explicit expression, we can approximate it using numerical methods such as the Newton-Raphson algorithm, which iteratively updates the estimator $\widehat{\theta}^{(k)}$ at the kth iteration by

$$\widehat{\theta}^{(k+1)} = \widehat{\theta}^{(k)} - \{\ddot{\ell}_n(\widehat{\theta}^{(k)})\}^{-1}\dot{\ell}_n(\widehat{\theta}^{(k)})$$

  until convergence.

Inference and Hypothesis Testing

- Under some mild regularity conditions, the MLE $\widehat{\theta}_n$ satisfies
  - consistency: $\widehat{\theta}_n \xrightarrow{p} \theta$
  - asymptotic normality: $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \mathcal{I}(\theta)^{-1})$, where $\mathcal{I}(\theta)$ is the Fisher information matrix.
  - Delta method: $\sqrt{n}\{g(\widehat{\theta}_n) - g(\theta)\} \xrightarrow{d} N(0, \dot{g}(\theta)^{\mathsf{T}}\mathcal{I}(\theta)^{-1}\dot{g}(\theta))$, for any continuous function g of θ.

- Three commonly used tests for the null hypothesis $H_0 : \theta = \theta^*$ (suppose θ is an r-dimensional vector):
  - Wald test: $W_n = (\widehat{\theta}_n - \theta^*)^{\mathsf{T}}\{n\mathcal{I}(\theta^*)\}(\widehat{\theta}_n - \theta^*) \xrightarrow{d} \chi^2_r$
  - Score test: $SC_n = \dot{\ell}_n(\theta^*)^{\mathsf{T}}\{n\mathcal{I}(\theta^*)\}^{-1}\dot{\ell}_n(\theta^*) \xrightarrow{d} \chi^2_r$
  - Likelihood ratio test: $LRC_n = 2\{\ell_n(\widehat{\theta}_n) - \ell_n(\theta^*)\} \xrightarrow{d} \chi^2_r$

- Inference for semiparametric models such as the Cox and AFT models will be studied in future chapters.