SURVIVAL ANALYSIS

K. E. Gneyou

University of Lomé - Faculty of Sciences
BP-1515 Lomé-Togo

École Mathématiques Africaine

UAC, Cotonou, Republic of Benin

January 22 –29, 2018

ORGANIZATION OF THE COURSE

1. Introduction
2. Chapter 1 : Censoring and truncation
3. Chapter 2 : Nonparametric methods
4. Chapter 3 : Parametric methods
5. Chapter 4 : Regression methods

1 INTRODUCTION

Survival analysis is the modern name given to the collection of statistical procedures which accommodate time-to-event censored data.

Aim of survival analysis : model and analyse time-to-event data, i.e., data that has as a principal endpoint the time when an event occurs.

A survival time is the time to occurrence of some event of interest. It is called a lifetime or failure time when it is the duration from the origin of times to the moment of death of a patient or failure of an electronic component in an industrial life-testing experiment.

Survival times arise especially in medical follow-up as well as in industrial reliability.

Models of survival data has become a particular field of statistical methods because of the following reasons :

- The data observed is rarely "complete" in the sense that it may be subject to observations that complicate seriously the description of the phenomenon. This kind of observations are called **censoring** or **truncation** variables.
- The usual parametric distributions are rarely adapted to the observed times of the phenomenon (they are often centred or have positive skewness whereas the life times have negative skewness). That is why, parametric and semi-parametric methods are generally used to estimate the lifetime distributions.

There are globally two ways which aim to resolve the same problems in survival analysis but use different approaches:

- Approach with original formulations of the models using the methods of the traditional statistical inference,
- Approach with punctual processes using the powerful results of martingales theory.

The objective of this course is to provide an introduction to the field of survival analysis in a coherent manner which captures the spirit of the methods of statistical modelling and analysis of lifetimes without getting too embroiled in the theoretical technicalities.

1.1 Basic concepts and tools

Let T denote the lifetime. T is a continuous random variable defined on a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, with probability density function f and distribution function $F(t) = \int_0^t f(x)dx$.
In survival analysis the distribution of T is usually characterized by the following other functions

- the survivor function : $S(t) = \mathbb{P}[T \geqslant t] = 1 - F(t)$,
- the hazard function : $\lambda(t) = \dfrac{f(t)}{S(t)} = \dfrac{f(t)}{1 - F(t)}$,
- the cumulative hazard function : $\Lambda(t) = \int_0^t \lambda(s)ds$.

The function $S(t)$ is also referred to as the reliability function. The hazard function $\lambda(t)$ specifies the instantaneous rate of failure at $T = t$ given that the individual survived up to time t, that is

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(t \leqslant T < t + \Delta t | T \geqslant t). \tag{1}$$

The p-th quantile of the distribution of T is the value t_p of T such that $F(t_p) = \mathbb{P}[T \leqslant t_p] = p$. It is also referred to as the $100 \times p$-th percentile of the distribution.

The **mean residual life** at a time t :

$$\mathrm{mrl}(t) = \mathbb{E}[T - t | T > t] = \frac{\int_t^{+\infty} S(x) dx}{S(t)}.$$

For individuals of age t, it measures their expected remaining lifetime.

The quantity $\lambda(t)\Delta t$ is approximatively the probability of a death in the interval $[t, t + \Delta t)$, given survival up to time t. $\lambda(t)$ is also referred to as the **risk** or **mortality rate** viewed as a measure of intensity or a measure of the potential of failure at time t.

NB: The hazard is a rate, rather than a probability. It is a probability per unit time and depends on whether time is measured in days, weeks, months or years, etc. It can assume values in $[0, +\infty)$.

For example, if $\mathbb{P}[t \leqslant T \leqslant t + \Delta t | T \geqslant t] = \frac{1}{4}$, then we have

$$\Delta t = \frac{1}{3}\text{day}, \Rightarrow \lambda = 0.75 \text{ per } \text{ day}$$

$$\Delta t = \frac{1}{21}\text{week}, \Rightarrow \lambda = 5.25 \text{ per } \text{ week}$$

It is easy to check the following fondamental relations

$$\begin{aligned} \Lambda(t) &= \int_0^t \lambda(u)du = -\log S(t), && (2) \\ S(t) &= \exp(-\Lambda(t)) = \exp(-\int_0^t \lambda(u)du) && (3) \\ f(t) &= \lambda(t)S(t) = \lambda(t)\exp(-\Lambda(t)). && (4) \end{aligned}$$

Note that the hazard function is usually more informative about the underlying mechanism of failure than the survivor function. For this reason, modeling the hazard function is an important method for summarizing survival data.

1.2 Censoring and truncation models

A distinctive feature of survival data is that some observations may be **censored**. That is, often the event of interest (e.g. death of patient, failure of component, recovery of patient) has not occurred by the time of recording so that, the lifetime of that subject is at least some value which is referred to as censoring time. Such censoring observations cannot be ignored since they carry important information about the effectiveness of the treatment. The necessity of obtaining methods of analysis that accommodate censoring is the primary reason for developing specialized models and procedures for failure time data.

We present here three types of censoring models and two truncation models. $T_1, T_2, \cdots, T_n$ will denote independent and identically distributed random variables, representing the lifetimes of n individuals under a study.

1.2.1 Type I censoring

Put items considered on test at t=0 and record their times to failure. Some items may take a long time to "fail" and one will not want to wait that long time to terminate the experiment. Therefore, one **terminates experiment at a pre-specified time** t_c. Hence the number of observed failure times is random. Instead of observing the T_is, one observes $Y_1, Y_2, \cdots, Y_n$ where

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

t_c is called the fixed censoring time. Let δ be the censoring variable which indicates if a failure time is observed or censored,

$$\delta = 1\!\!\mathrm{I}_{\{T \leqslant t_c\}} = \begin{cases} 1 & \text{if } T \leqslant t_c \\ 0 & \text{otherwise.} \end{cases}$$

One then observes the i.i.d. random pairs (Y_i, δ_i).

1.2.2 Type II censoring

In this type, the experiment is run until a pre-specified fraction $\frac{r}{n}$ of the n items has failed. Then by plan, **observations terminate after the r-th failure occurs**. So only one observes the r smallest observations in the sample. Hence one has $n - r$ censored observations.

Remarks 1.1 Note that

1. In Type I censoring, the end point t_c is a fixed value and the number of observed failure times is a r.v. which assumes a value in the set $\{0, 1, 2, \cdots, n\}$.
2. In Type II censoring, the number of failures times r is a fixed value whereas the endpoint T_r is a random observation. Hence one could wait possibly a very long time to observe the r failures or, vice versa, see all r relatively early on.

1.2.3 Type III or Random censoring

We present only Right censoring. Left censoring is analogous. Random censoring occurs frequently in medical studies. In clinical trials, patients typically enter a study at different times. Then each is treated with one of several possible therapies. We want to observe their "failure" time but censoring can occur in one of the following ways:

1. *Loss to Follow-up*. Patients moves away. We never see him again. We only know he has survived from entry date until he left. So his survival time is greater than the observed value.

2. *Drop out*. Bad side effects forces termination of treatment. Or patient refuses to continue treatment for whatever reasons.

3. *Termination of study*. Patient is still alive at end of study.

There are other types of censored data such as Left-censored, Interval censored, doubly-censored or truncation data.

In Left-censored case, event had already occurred before the study started. Subject cannot be included in study.

In Interval censoring case, each event time T_i is only known to fall in an interval $(L_i, R_i]$ where L_i and R_i denote respectively the left and right endpoints of the censoring interval.

Doubly censored data are observations which are randomly left and right censored.

Truncation is a procedure where a condition other than the main event of interest is used to screen individuals, that is, only if the individual has the truncation condition prior to the event of interest will s/he be observed by the investigator.

1.3 Examples

The following examples illustrate studies where different types of censored observations could occur.

Exemple 1.1 (AML study)

The data in Table 1 are preliminary results from a clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukemia (AML). The study was conducted by Embury et al. (1977) at Stanford University. After reaching a status of remission through treatment by chemotherapy, the patients who entered the study were assigned randomly to two groups. The first group received maintenance chemotherapy; the second, or control group, did not. The objective of the trial was to see if maintenance chemotherapy prolonged the time until relapse.

Table 1: *Data for the AML maintenance study. A+ indicates a censored value.*

Group	Length of complete remission (in weeks)
Maintained	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48,161+
Non-maintained	5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45

This example illustrate a right random censored data.

Exemple 1.2

A child psychiatrist visits a togolese village to study the age at which children first learn to perform a particular task. Let T denote the age a child learns to perform a specified task. The time values which can be recorded are, T: exact age is observed (uncensored), T^-: age is left-censored as the child already knew the task when s/he was initially tested in the study and T^+: age is right-censored since the child did not learn the task during the study period.

This example and the next illustrate a left and right censored data model. And when all these can occur, this is also referred to as a model of doubly censored data.

Exemple 1.3

Extracted from Klein and Moeschberger (1997). High school boys are interviewed to determine the distribution of the age of boys when they first used marijuana. The question stated was "When did you first use marijuana ?". The three possible answers and respective recorded values are given as follow:

a. I used it but cannot recall just when the first time was. (Recorded value: T^-: age at interview as exact age was earlier but unknown).

b. I first used it when I was x old. (Recorded value: T: exact age since it is known (uncensored)).

c. I never used it. (Recorded value: T^+: age at interview since exact age occurs sometime in the future).

Exemple 1.4

Age in months when members of a retirement community died or left the center (right-censored) and age when the members entered the community (the truncation event) are recorded. Individuals must survive to a sufficient age to enter the retirement community. Individuals who die at an early age are excluded from the study. Hence, the life lengths in this data set are *left-truncated*.

Exemple 1.5

Measurement of interest is the waiting time in years from HIV infection to development of AIDS. In the sampling scheme, only individuals who have developed AIDS prior to the end of the study are included in the study. Infected individuals who have yet to develop AIDS are excluded from the sample; hence, unknown to the investigator. This is a case of *right truncation*.

- *The three basic goals of survival analysis* :

Goal 1 Estimate and interpret survivor and/or hazard functions from survival data.

Goal 2 Compare survivor and/or hazard functions.

Goal 3 Assess the relationship of explanatory variables to survival time, especially through the use of formal mathematical modeling.

2 NONPARAMETRIC METHODS

Nonparametric methods are procedures of inference that are valid simultaneously for many different types of underlying distributions of the life-time T. The inference will concern the survivor function $S(t) = \mathbb{P}[T > t]$ and, hence, functions of it.

2.1 Complete Failure Times

Let n be the number of individuals in the sample. Assume that there is no censored observation. Then the set of the complete data $t_1, t_2, \cdots, t_n$ reflects the structure of population failure times and $S(t)$ can be estimated by

$$S_n(t) = \frac{\#\{t_i > t\}}{n} = \frac{1}{n}\sum_{i=1}^{n} 1\!\!\mathrm{I}_{\{t_i > t\}}$$

called the empirical survival distribution function.

- *Confidence interval for $S(t)$:*

$S_n(t)$ is the proportion of individuals still alive after time t and hence $nS_n(t)$ is a binomial random variable with parameters n and $p = S(t)$. It follows that,

$$\mathbb{E}(S_n(t)) = S(t) \quad \text{and} \quad \mathrm{var}(S_n(t)) = \frac{S(t)(1 - S(t))}{n}.$$

By the central limit theorem (CLT), $S_n(t) \rightsquigarrow \mathcal{N}(S(t), \sqrt{\frac{S(t)(1-S(t))}{n}})$ for n large. Hence a 95% confidence interval for $S(t)$ is given by

$$\left[S_n(t) - 1.96\sqrt{\frac{S_n(t)(1 - S_n(t))}{n}}, S_n(t) + 1.96\sqrt{\frac{S_n(t)(1 - S_n(t))}{n}} \right].$$

2.2 Right Censored Failure Times

We consider only random censoring. When there are right-censored observations, we use the product-limit (PL) estimator to estimate $S(t)$. This is commonly called the Kaplan-Meier (KM) estimator.

For each of the n individuals, instead of T_i, one observes the pair (Y_i, δ_i) where

$$Y_i = \min(T_i, C_i) \qquad \text{and} \qquad \delta_i = 1\!\!\mathrm{I}_{\{T_i \leqslant C_i\}} = \begin{cases} 1, & \text{if } T_i \leqslant C_i \\ 0, & \text{if } T_i > C_i \end{cases}$$

Let $y_{(i)}$ denotes the i-th distinct ordered censored or uncensored observation and be the right endpoint of the interval $I_i = (y_{(i-1)}, y_{(i)}]$, $i = 1, 2 \cdots, n' \leqslant n$ with $y_{(0)} = 0$.

In what follows

- death is the generic word for the event of interest.
- A cohort is a group of people who are followed throughout the course of the study.
- The people at risk at the beginning of the interval I_i are those people who survived (not dead, lost, or withdrawn) the previous interval I_{i-1}.

$\mathcal{R}(t)$ denotes the risk set just before time t and let

$n_i =$ Number of individuals in $\mathcal{R}(y_{(i)}) =$ Number of alive (and not censored) just before $y_{(i)}$.

$d_i =$ Number of died at time $y_{(i)}$.

$$\begin{aligned} p_i &= P(\text{surviving trhough } I_i \mid \text{alive at beginning of } I_i) \\ &= \mathbb{P}[T > y_{(i)} \, |T > y_{(i-1)}]. \end{aligned}$$

$q_i = 1 - p_i = P(\text{die in } I_i \mid \text{alive at beginning of } I_i)$.

Recall the multiplication rule for joint events A_1 and A_2 :

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1).$$

From repeated application of this product rule, the survivor function can be expressed as

$$S(t) = \mathbb{P}[T > t] = \prod_{y_{(i)} \leqslant t} p_i.$$

The estimates of p_i and q_i are

$$\hat{q}_i = \frac{d_i}{n_i} \qquad \text{and} \qquad \hat{p}_i = 1 - \hat{q}_i = 1 - \frac{d_i}{n_i} = \frac{n_i - d_i}{n_i}.$$

The KM estimator $\hat{S}(t)$ of the survivor function $S(t)$ is then

$$\hat{S}(t) = \prod_{y_{(i)} \leqslant t} \hat{p}_i = \prod_{y_{(i)} \leqslant t} (\frac{n_i - d_i}{n_i}) = \prod_{i=1}^{k} (\frac{n_i - d_i}{n_i}), \tag{5}$$

for $y_{(k)} \leqslant t < y_{(k+1)}$.

Remarks 2.1 If there is no tie then

a) $d_i = 1$ if $\delta_{(i)} = 1$ and $d_i = 0$ if $\delta_{(i)} = 0$. Hence

$$\hat{p}_i = \begin{cases} 1 - \frac{1}{n_i}, & \text{if } \delta_{(i)} = 1 \text{ i.e. the event is a death at time } t_{(i)} \\ 1, & \text{if } \delta_{(i)} = 0 \text{ i.e. the event is a censoring at time } t_{(i)} \,. \end{cases}$$

b) $n_i = n - i + 1$ and formula (5) becomes

$$\hat{S}(t) = \prod_{y_{(i)} \leqslant t} \left(1 - \frac{1}{n-i+1}\right)^{\delta_{(i)}} = \prod_{y_{(i)} \leqslant t} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}. \qquad (6)$$

c) $\hat{S}(t_{(i)}) = \hat{p}_i \times \hat{S}(t_{(i-1)})$.

d) These estimates are subject to sampling error. Greenwood showed that for $y_{(k)} \leqslant t < y_{(k+1)}$ one has approximately

$$\mathrm{var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{i/y_{(i)} \leqslant t} \frac{d_i}{n_i(n_i - d_i)} = (\hat{S}(t))^2 \sum_{i=1}^{k} \frac{d_i}{n_i(n_i - d_i)}, \qquad (7)$$

Exemple 2.1 Let's consider AML1 the maintained group of AML example (Example 1.1) where a "+" denotes a right-censored observed value.

Maintained : 9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+

We have

Table 2: *KM Survival probabilities for AML1 data.*

i	Times	Intervals	Deaths d_i	At-risk $n_i = n-i+1$	$\hat{q}_i$	$\hat{p}_i$	$\hat{S}(t_{(i)})$	$\hat{F}(t_{(i)})$
1	9	$(0,9]$	1	11	1/11	10/11	0.91	0.09
2	13	$(9,13]$	1	10	1/10	9/10	0.82	0.18
3	13+	$(13,13+]$	0	9	0/9	1	0.82	0.18
4	18	$(13+,18]$	1	8	1/8	7/8	0.72	0.28
5	23	$(18,23]$	1	7	1/7	6/7	0.61	0.38
6	28+	$(23,28+]$	0	6	0	1	0.61	0.38
7	31	$(28+,31]$	1	5	1/5	4/5	0.49	0.51
8	34	$(31,34]$	1	4	1/4	3/4	0.37	0.63
9	45+	$(34,45+]$	0	3	1/3	1	0.37	0.63
10	48	$(45+,48]$	1	2	1/2	1/2	0.18	0.82
11	161+	$(48,161+]$	0	1	0	1	0.18	0.82

e.g.

$$\begin{aligned}\hat{S}(0) &= 1 \\ \hat{S}(9) &= \hat{S}(0) \times \frac{11-1}{11} = 0.91\end{aligned}$$

$$\begin{aligned}
\hat{S}(13) &= \hat{S}(9) \times \frac{10-1}{10} = 0.82; \quad \hat{S}(13+) = \hat{S}(13) \times \frac{9-0}{9} = \hat{S}(13) = 0.82 \\
\hat{S}(18) &= \hat{S}(13) \times \frac{8-1}{8} = 0.72 \\
\hat{S}(23) &= \hat{S}(18) \times \frac{7-1}{7} = 0.61 \\
\hat{S}(28+) &= \hat{S}(23) \times \frac{6-0}{6} = \hat{S}(23) = 0.61 \\
\hat{S}(31) &= \hat{S}(23) \times \frac{5-1}{5} = 0.49 \\
\hat{S}(34) &= \hat{S}(31) \times \frac{4-1}{4} = 0.37; \quad \hat{S}(45+) = \hat{S}(34) \times \frac{3-0}{3} = \hat{S}(34) = 0.37 \\
\hat{S}(48) &= \hat{S}(34) \times \frac{2-1}{2} = 0.18; \quad \hat{S}(161+) = \hat{S}(48) \times \frac{1-0}{1} = \hat{S}(48) = 0.18.
\end{aligned}$$

Example of Greenwood-estimate of the variance of $\hat{S}(13)$:
$\mathrm{var}(\hat{S}(13)) = (0.82)^2 \left(\frac{1}{11(11-1)} + \frac{1}{10(10-1)} \right) = 0.0136.$

- *Estimates of hazard (risk) :*

Let t_i, $i = 1, \cdots, K$ denote a distinct ordered death time and let $\lambda(t)$ denote the hazard function as in the introduction. $\lambda(t)$ can be estimated by :

1. estimate at an observed death time t_i: $\tilde{\lambda}(t_i) = \dfrac{d_i}{n_i}$,

2. estimate in the interval $[t_i, t_{i+1})$: $\hat{\lambda}(t) = \dfrac{d_i}{n_i(t_{i+1} - t_i)}$.

   $\hat{\lambda}(t)$ estimates the hazard rate of death per unit time in the interval $[t_i, t_{i+1})$. It is referred to as the KM type estimate.

Examples with AML1 data:

$\tilde{\lambda}(23) = \frac{1}{7} = 0.143$

$\hat{\lambda}(26) = \frac{1}{7(31-23)} = 0.018.$

- *Cumulative hazard Estimates :*

1. Breslow-estimate

$$\hat{\Lambda}(t) \;=\; -\log(\hat{S}(t)) = -\log \prod_{y_{(i)}\leqslant t} \left(\frac{n_i - d_i}{n_i}\right), \tag{8}$$

$$\mathrm{var}(\hat{\Lambda}(t)) \;=\; \sum_{y_{(i)}\leqslant t} \frac{d_i}{n_i(n_i - d_i)} = \sum_{i=1}^{k} \frac{d_i}{n_i(n_i - d_i)}, \text{ for } y_{(k)} \leqslant t < y_{(k+1)}.$$

2. Nelson-Aalen -estimate (1972, 1978)

$$\tilde{\Lambda}(t) \;=\; \sum_{y_{(i)}\leqslant t} \frac{d_i}{n_i}, \quad \mathrm{var}(\tilde{\Lambda}(t)) = \sum_{y_{(i)}\leqslant t} \frac{d_i}{n_i^2}. \tag{9}$$

3. Harrington-Fleming estimator of S :

$$\hat{S}_{\mathrm{HF}}(t) \;=\; \exp(-\tilde{\Lambda}(t)) = \prod_{y_{(i)}\leqslant t} \exp(-\frac{d_i}{n_i}). \tag{10}$$

Examples with AML1 data:

$$\begin{aligned}\hat{\Lambda}(26) &= -\log \hat{S}(26) = -\log(0.614) = 0.488,\\ \tilde{\Lambda}(26) &= \frac{1}{11} + \frac{1}{10} + \frac{1}{8} + \frac{1}{7} = 0.4588.\end{aligned}$$

Note that if there were no censored observations, S is estimated by the e.s.d.f. $S_n(t)$ which is a right continuous step function with steps down at each $t_{(i)}$. The KM estimator $\hat{S}(t)$ is also a right continuous step function which steps down only at an uncensored observation. When there are no censored data values, KM reduces to the empirical survival d.f.

- *Confidence bounds of the hazard function :*

Confidence intervals for $S(t)$ depends upon theoretical results.

1. Using Greenwood's standard error, one get the confidence interval

$$\exp\left(\log \hat{S}(t) \pm 1.96\sigma_{\hat{\Lambda}(t)}\right).$$

   This confidence interval is the default one in the R package **survfit** and is obtained by using the delta-method. The log-transform on $\hat{S}(t)$ gives other more efficient intervals. These intervals are called "**log**" and in the **survfit** function, you must specify **conf.int="plain"**.

2. Kalbfleisch and Printice (1980), using the transform $W = \log(-\log(\hat{S}(t)))$ which estimates $\log(-\log(S(t)))$ and the delta-method, suggests an approximate $(1-\alpha) \times 100\%$ C.I. given by

$$\left(\hat{S}(t)\right)^{\exp(z_{\alpha/2}\sigma_W)} \leqslant S(t) \leqslant \left(\hat{S}(t)\right)^{\exp(-z_{\alpha/2}\sigma_W)}$$

   where $\sigma_W = \sqrt{\operatorname{var} W}$. To get these intervals in R commands, specify **conf.int="log-log"** in the **survfit** function.

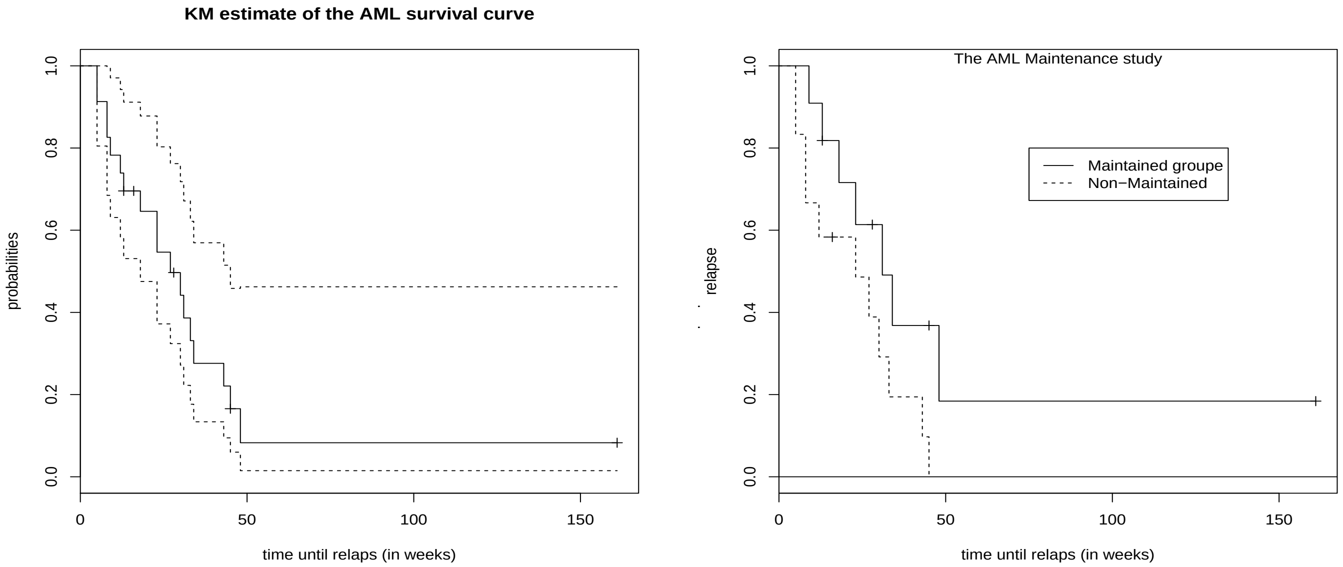

Figure 1: Kaplan-Meier product-limit estimates of AML data

2.3 Comparison of two binomial populations : Fisher's exact test

Suppose we have two populations, and an individual in either population can have one of two characteristics. For example, Population 1 might be cancer patients under a certain treatment and Population 2 cancer patients under a different treatment. The patients in either group may either die within a year or survive beyond a year.

The data are summarized in the following 2×2 contingency table. Our interest here is to compare the two binomial populations, which is common in medical studies.

Populations	Dead	Alive	Totals
Population 1	a	b	n_1
Population 2	c	d	n_2
Totals	m_1	m_2	n

Denote

$$\begin{aligned} p_1 &= P\{\text{Dead} \mid \text{Population}1\} \\ p_2 &= P\{\text{Dead} \mid \text{Population}2\}. \end{aligned}$$

The null hypothesis of the test is $H_0 : p_1 = p_2$.

If the margins of this 2×2 table are considered fixed, the random variable A, which is the entry in the $(1,1)$ cell, has a hypergeometric distribution under H_0 with :

$$\mathbb{P}\{A = a\} = \frac{C(n_1, a).C(n_2, m_1 - a)}{C(n, m_1)}$$

where $C(n,k) = \binom{n}{k}$.

Let $t_1 < t_2 < \cdots , < t_K$ be the ordered times of deaths. At all time t_i set

. $n_i = n$ the number of subjects at risk in the two populations together;

. $d_i = m_1$ the number of observed deaths in the two populations together;

. n_{li} and d_{li} the analogous of n_i and d_i in the population l, $l = 1, 2$;

. $e_{1i} = \mathbb{E}_{H_0}(A_i)) = n_i p_i = d_i \frac{n_{1i}}{n_i}$.

The statistic

$$U = \sum_{i=1}^{K} W_i (d_{1i} - e_{1i}). \tag{11}$$

is a centred r.v. and by independence of the variables $A_i = d_{1i}$, $i = 1, \cdots , K$, we have

$$V = \mathrm{var}(U) = \sum \text{of variances} = \sum_{i=1}^{K} W_i^2 v_i. \tag{12}$$

where $v_i = \mathrm{Var}_{H_0}[A_i] = \mathrm{var}(d_{1i}) = \frac{n_i - d_i}{n_i - 1} d_i \frac{n_{1i} n_{2i}}{n_i^2}$.

If K is big enough or if the margins of each table are big, then U is approximatively normally distributed. 3 choices of weighting W_i lead to the following test statistics :

a) $W_i = 1$ (that is, all deaths have the same weight) yields the **Mantel-Haenszel or log-rank** test given by

$$\chi_1^2 = \frac{(O_1 - E_1)^2}{V}.$$

where $O_1 = \sum_{i=1}^{K} d_{1i}$ is the number of the deaths in the population 1 and $E_1 = \sum_{i=1}^{K} e_{1i}$ is the number of the expected deaths in population 1 under H_0. Similarly one defines considering the population 2

$$\chi_1^2 = \frac{(O_2 - E_2)^2}{V}.$$

b) $W_i = n_i$ (that is, the first deaths have larger weights than the next deaths) yields **Gehan**'s test statistic;

c) $W_i = S_i^* = \sum_{j=1}^{i} \frac{n_j}{n_j+d_j}$ yields **Peto and Prentice** test statistic.

Considering that the sum of observed deaths in the two population together equals to the expected sum of deaths in the two populations under H_0, i.e.

$$O_1 + O_2 = E_1 + E_2 \quad \text{and} \quad V = \mathrm{var}(O_1 - E_1) = \mathrm{var}(O_2 - E_2),$$

the log-rank test statistic is

$$\chi_1^2 = \frac{(O_1 - E_1)^2}{V} = \frac{(O_2 - E_2)^2}{V}$$

and one can show that

$$\chi_a^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

is always less than χ_1^2. So the rejection of H_0 with χ_a^2 leads to the rejection of H_0 based on χ_1^2.

Exemple 2.2

12 brain tumour patients randomized to radiation (group 1) or radiation + chemotherapy (Group 2):

Group	Survival times (in weeks)
Group 1	10 12^+ 26 28 30 41
Group 2	15^+ 24 30 42 40^+ 42^+

Here $K = 6$. Recalling that for all $i = 1, \cdots,$

$$e_{1i} = d_i \frac{n_{1i}}{n_i}, \qquad O_1 = \sum d_{1i}, \qquad E_1 = \sum e_{1i}$$

$$e_{2i} = d_i \frac{n_{2i}}{n_i}, \qquad O_2 = \sum d_{2i}, \qquad E_2 = \sum e_{1i}$$

we have

Table 3: *Calculus for MH statistic.*

i	t_i	n_{1i}	n_{2i}	n_i	d_{1i}	d_{2i}	d_i	e_{1i}	e_{2i}
1	10	6	6	12	1	0	1	1/2	1/2
2	24	4	5	9	0	1	1	4/9	5/9
3	26	4	4	8	1	0	1	1/2	1/2
4	28	3	4	6	1	0	1	3/7	4/7
5	30	2	4	6	1	1	2	2/3	4/3
6	41	1	2	3	1	0	1	1/3	2/3
7	42	0	2	2	0	1	1	0	1
Totaux					5	3		2.87	5.13

Hence

$$\chi^2_C = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} = 2.46$$

and $\chi^2_1(0.95) = 3.84$. We conclude that there is no significant difference in survivor in the two groups al the level $\alpha = 5\%$.

Remarks 2.2

The S/R function **survdiff** provides the log-rank MH test by default. Its first argument takes a **Surv** object. It gives the square of the MH statistic which is then an approximate chi-square statistic with 1 degree of freedom. This is a two-tailed test. Hence, the *p*-value is twice that of the MH above.

The **survdiff** function contains a ”rho” parameter. The default value, rho = 0, gives the log-rank test. When rho = 1, this gives the Peto test. This test was suggested as an alternative to the log-rank test by Prentice and Marek (1979). The Peto test emphasizes the beginning of the survival curve in that earlier failures receive larger weights. The log-rank test emphasizes the tail of the survival curve in that it gives equal weight to each failure time. Thus, choose between the two according to the interests of the study. The choice of emphasizing earlier failure times may rest on clinical features of one’s study.

Hazard ratio as a measure of effect:

The hazard ratio is a descriptive measure of the treatment effect on survival. e.g. if x is a covariate with values in $\{x_1, x_2\}$, one can compute

$$\frac{\hat{\lambda}(t|x = x_2)}{\hat{\lambda}(t|x = x_1)} \qquad \text{or} \qquad \frac{\tilde{\lambda}(t|x = x_2)}{\tilde{\lambda}(t|x = x_1)}$$

and then, interpret them in the context of the study.

Example: Consider in AML study, $x = 1$ if "maintained" and $x = 0$ if nonmaintained. The two hazard ratios of nonmaintained to maintained are given by :

$$\frac{\hat{\lambda}(15|x = 0)}{\hat{\lambda}(15|x = 1)} = \frac{0.011}{0.020} = 0.55 \qquad \text{and} \qquad \frac{\hat{\lambda}(25|x = 0)}{\hat{\lambda}(25|x = 1)} = \frac{0.042}{0.018} = 2.33.$$

This means that, the nonmaintained group has 55% of the risk of those maintained of relapsing at 15 weeks. However, on the average, those nonmaintained have 2.33 times the risk of those maintained of relapsing at 25 weeks.

3 PARAMETRIC METHODS

Parametric models assume that the failure time has the density function $f(t;\theta)$, where $\theta = (\theta_1, \theta_2, \cdots, \theta_m)$ is the unknown vector of parameters. The density and survival functions are completely specified if θ is known.

3.1 Frequently used continuous models

Frequently used parametric models assume that T follows

1. exponential distribution, in which case $m = 1$
2. Weibull distribution, in which case $m = 2$
3. log-normal distribution, in which case $m = 2$
4. log-logistic distribution, in which case $m = 2$ and
5. gamma distribution, in which case $m = 2$.

Let us recall the properties of each of those distributions.

Let T be the lifetime of interest. Denote by

. $F(t)$ the probability distribution function (d.f.),

. $S(t) = 1 - F(t)$ the survivor function,

. $f(t) = F'(t)$ the probability density function (p.d.f.),

. $\lambda(t) = \frac{f(t)}{1-F(t)}$ the hazard rate function,

. $t_p = \inf\{t : S(t) \leq 1 - p\}$ the p-th quantile,

. $\Lambda(t) = \int_0^t \lambda(s)ds$ the cumulative hazard function,

. $\mathbb{E}(T) = \int_0^{+\infty} S(t)dt$ the mean and

. $\mathrm{var}(T) = \mathbb{E}(T - \mathbb{E}(T))^2$ the variance of T.

3.1.1 Exponential distribution

The lifetime T is exponentially distributed with parameter $\lambda > 0$ if we have

density	survivor	hazard rate	mean	variance	pth quantile
$f(t) = \lambda e^{-\lambda t}$	$S(t) = e^{-\lambda t}$	$\lambda(t) = \lambda$	$\mathbb{E}(T) = \frac{1}{\lambda}$	$\mathrm{var}(T) = \frac{1}{\lambda^2}$	$t_p = -\lambda^{-1} \log(1 - p)$

By the relationship $\log(\Lambda(t)) = \log(-\log(S(t))) = \log(\lambda) + \log(t)$ or, equivalently expressed with $\log(t)$ on the vertical axis,

$$\log(t) = -\log(\lambda) + \log(-\log(S(t))) \tag{13}$$

the plot of $y = \log(t)$ versus $x = \log(-\log(S(t))$ is a straight line with slope 1 and y-intercept $-\log(\lambda)$. The exponential is a special case of both the Weibull and gamma models, each with their shape parameter equal to 1.

3.1.2 Weibull distribution

T follows the Weibull distribution with parameters $\lambda > 0$ and $\alpha > 0$ if we have the following table

density	survivor	hazard rate	mean
$f(t) = \lambda\alpha(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}}$	$S(t) = e^{-(\lambda t)^{\alpha}}$	$\lambda(t) = \lambda\alpha(\lambda t)^{\alpha-1}$	$\mathbb{E}(T) = \frac{1}{\lambda}\Gamma(1+\frac{1}{\alpha})$

variance	p-th quantile
$\mathrm{var}(T) = \frac{1}{\lambda^2}\Gamma(1+\frac{2}{\alpha}) - \frac{1}{\lambda^2}(\Gamma(1+\frac{1}{\alpha}))^2$	$t_p = \lambda^{-1}(-\log(1-p))^{\frac{1}{\alpha}}$

where $\Gamma(t)$ denotes the gamma function defined by $\Gamma(t) = \int_0^{+\infty} x^{t-1}e^{-x}dx$, $t > 0$. The parameter α is called the shape parameter and λ is a scale parameter. The effect of different values of λ is just to change the scale on the horizontal (t) axis.

By the relationship $\log(\Lambda(t)) = \log(-\log(S(t))) = \alpha(\log(\lambda) + \log(t))$, equivalently expressed with $\log(t)$ on the vertical axis,

$$\log(t) = -\log(\lambda) + \sigma \log(-\log(S(t))) \tag{14}$$

where $\sigma = \frac{1}{\alpha}$, the plot of $y = \log(t)$ versus $x = \log(-\log(S(t))$ is a straight line with slope $\sigma = \frac{1}{\alpha}$ and y-intercept $-\log(\lambda)$.

The Weibull distribution is intrinsically related to the extreme value distribution. The natural log transform of a Weibull random variable produces an extreme value random variable. This relationship is exploited quite frequently, particularly in the statistical computing packages and in diagnostic plots.

T follows the extreme (minimum) value distribution with parameters μ and $\sigma > 0$ if T has

density	survivor	mean	variance	p-th quantile
$f(t) = \frac{1}{\sigma} e^{\frac{t-\mu}{\sigma} - e^{\frac{t-\mu}{\sigma}}}$	$S(t) = e^{-e^{\frac{t-\mu}{\sigma}}}$	$\mu - \gamma\sigma$	$\frac{\pi^2}{6}\sigma^2$	$t_p = \mu + \sigma.a(p)$

where $a(p) = \log(-\log(1-p))$, $\gamma = 0.5772...$ denotes Euler's constant, μ is the location parameter (it is 0.632-th quantile) and t can also be negative so that $-\infty < t < +\infty$. The standard extreme value distribution has $\mu = 0$ and $\sigma = 1$.

Further the following relationship can be easily shown :

If T is a Weibull random variable with parameters α and λ, then $Y = \log(T)$ follows an extreme value distribution with $\mu = -\log(\lambda)$ and $\sigma = \alpha^{-1}$. The r.v. Y can be represented as $Y = \mu + \sigma Z$, where Z is a standard extreme value r.v., as the extreme value distribution is a location and scale family of distributions.

3.1.3 log-normal distribution

T is log-normally distributed with parameters μ and $\sigma > 0$ denoted by $T \rightsquigarrow LN(\mu, \sigma)$ if $Y = \log(T)$ is normally distributed with mean and variance specified by μ and σ^2 respectively. Hence Y is of the form $Y = \mu + \sigma Z$, where Z is a standard normal r.v. We have the following table for T with $\alpha > 0$ and $\lambda > 0$ and where $\Phi(t)$ denotes the standard normal d.f.

density	survivor	hazard rate	mean	variance
$f(t) = \frac{\alpha}{\sqrt{2\pi}t} e^{-\frac{\alpha^2(\log(\lambda t))^2}{2}}$	$S(t) = 1 - \Phi(\alpha \log(\lambda t))$	$\lambda(t) = \frac{f(t)}{S(t)}$	$e^{\mu + \frac{\sigma^2}{2}}$	$e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$

with $\mu = -\log(\lambda)$ and $\sigma = \frac{1}{\alpha}$.

The hazard function has value 0 at $t = 0$, increases to a maximum, and then decreases, approaching zero as t becomes large.

3.1.4 log-logistic distribution

T is log-logistically distributed with parameters μ and $\sigma > 0$ denoted by $T \rightsquigarrow$ Llogist(μ, σ) if $Y = \log(T)$ is logistically distributed with location parameter μ and scale parameter σ. Hence Y is also of the form $Y = \mu + \sigma Z$, where Z is a standard logistic r.v. with density function $f(z) = \frac{e^z}{(1+e^z)^2}$, $z \in \mathbb{R}$. Z is a symmetric r.v. with mean 0 and variance $\frac{\pi^2}{3}$, and with slightly heavier tails than the standard normal, the excess in kurtosis being 1.2. We have the following table for T with $\alpha > 0$ and $\lambda > 0$:

density	survivor	hazard rate	pth quantile
$f(t) = \frac{\lambda\alpha(\lambda t)^{\alpha-1}}{(1+(\lambda t)^\alpha)^2}$	$S(t) = \frac{1}{1+(\lambda t)^\alpha}$	$\lambda(t) = \frac{\lambda\alpha(\lambda t)^{\alpha-1}}{1+(\lambda t)^\alpha}$	$\lambda^{-1}(\frac{p}{1-p})^{\frac{1}{\alpha}}$

with $\mu = -\log(\lambda)$ and $\sigma = \alpha^{-1}$.

Note that the hazard function of log-logistic distribution is identical to the Weibull hazard aside from the denominator factor $1 + (\lambda t)^{\alpha}$. For $\alpha < 1$ it is monotone decreasing from ∞ and is monotone decreasing from λ if $\alpha = 1$. If $\alpha > 1$, the hazard resembles the log-normal hazard as it increases from zero to a maximum at $t = \frac{1}{\lambda}(\alpha - 1)^{1/\alpha}$ and decreases toward zero thereafter.

Note also that $\dfrac{S(t)}{1 - S(t)} = (\lambda t)^{-\alpha}$. It easily follows that $\log(t)$ is a linear function of the log-odds of the survival beyond t. That is

$$\log(t) = \mu + \sigma \left(- \log(\frac{S(t)}{1 - S(t)}) \right), \tag{15}$$

where $\mu = -\log(\lambda)$ and $\sigma = \frac{1}{\alpha}$. Thus the plot of $y = \log(t)$ against $x = -\log(\frac{S(t)}{1-S(t)})$ is a straight line with slop σ and y-intercept μ.

Summary:

Almost all distributions of lifetime T we work with, have the property that the distribution of the log-transform $\log(T)$ is a **member of the location and scale family** of distributions. The common features are :

- The time T distributions have two parameters : scale λ and shape α.
- In log-time, $Y = \log(T)$, the distributions have two parameters : location $\mu = -\log(\lambda)$ and scale $\sigma = \frac{1}{\alpha}$.
- Each can be expressed in the form

$$Y = \log(T) = \mu + \sigma Z \tag{16}$$

  where Z is the standard member; that is $\mu = 0$ $(\lambda = 1)$ and $\sigma = 1$ $(\alpha = 1)$.
- They are log-linear models.

The three distributions previously considered are summarized as follow

Law of T	Weibull	Log-normal	Log-logistic
Law of $Y = \log(T)$	Extreme value distribution	Normal	Logistic

If the true distribution of $Y = \log(T)$ is one of the above, then the p-th quantile y_p is a linear function of z_p, the p-th quantile of the standard member of the specified distribution. The straight line has slope σ and y-intercept μ. Let t_p denote an arbitrary p-th quantile. The linear relationships for $y_p = \log(t_p)$ reported in expressions (14), (15), (16) are summarized in the following table

Table 4: *Relationships between quantiles and transformed quantiles.*

t_p quantile	$y_p = \log(t_p)$ quantile	form of standard quantile z_p
Weibull	extreme value	$\log(-\log(S(t_p))) = \log(-\log(1-p))$
log-normal	normal	$\Phi^{-1}(p)$ where Φ denotes the standard normal d.f.
log-logistic	logistic	$-\log\left(\frac{S(t_p)}{1-S(t_p)}\right) = -\log(\frac{1-p}{p})$

Construction of the quantile-quantile (Q-Q) plot

Let $\hat{S}(t)$ denote the K-M estimator of survival probability $S(t)$. Let $t_i,\ i = 1, \cdots, r \leqslant n$, denote the ordered uncensored observed failure times. For each uncensored sample quantile $y_i = \log(t_i)$, the estimated failure probability is $\hat{p}_i = 1 - \hat{S}(t_i)$. Use $\hat{p}_i$ to obtain the parametric standard quantile z_i as in Table 4. As the K-M estimator is distribution free and consistently estimates the "true" survival function, for large sample sizes n, the z_i should reflect the "true" standard quantiles. Hence, if the proposed model fits the data adequately, the points (z_i, y_i) should lie close to a straight line with slope σ and y-intercept μ. The plot of the points (z_i, y_i) is called a quantile-quantile (Q-Q) plot.

An appropriate line to compare the plot pattern to is $y_p = \hat{\mu} + \hat{\sigma} z_p$, where $\hat{\mu}$ and $\hat{\sigma}$ denote the maximum likelihood estimates of μ and σ to be discussed in the next section. The more closely the plot pattern follows this line, the more evidence there is in support of the proposed model. The Q-Q plot is a major diagnostic tool for checking model adequacy.

3.2 Maximum likelihood estimation

Recall the definition of censored data (Y, δ). We need to calculate the joint likelihood of the pair (Y, δ).

Type I censoring case

We check that the likelihood function for the n i.i.d. random pairs (Y_i, δ_i) is given by

$$L = \prod_{i=1}^{n} f(y_i)^{\delta_i} S(t_c)^{1-\delta_i} \tag{17}$$

Type II censoring case

In this case, denoting by $T_{(1)}, \cdots, T_{(r)}$ the r smallest lifetimes out of the n i.i.d. lifetimes $T_1, \cdots, T_n$, we can check that the likelihood function of $T_{(1)}, \cdots, T_{(r)}$ is

$$L = \frac{n!}{(n-r)!} f(t_{(1)}) \cdots f(t_{(r)}) (S(t_{(r)}))^{n-r}. \tag{18}$$

Type III or Random censoring case

Denote here by F, f and S_f (resp. G, g and S_g) the probability distribution function, the probability density function and the survivor function of the life time T (resp. the random censor time C). For each individual the lifetime T_i and a censor time C_i are usually assumed to be independent and the observation is the pair (Y_i, δ_i) where

$$Y_i = \min(T_i, C_i) \qquad \text{and} \qquad \delta_i = 1\!\!\mathrm{I}_{\{T_i \leqslant C_i\}} = \begin{cases} 1 & \text{if } T_i \leqslant C_i \\ 0 & \text{if } T_i > C_i \end{cases}$$

We check also that the likelihood function of the n pairs (Y_i, δ_i) is given by

$$\begin{aligned} L &= \prod_{i=1}^{n} (f(y_i)S_g(y_i))^{\delta_i}(g(y_i)S_f(y_i))^{1-\delta_i} \\ &= \left(\prod_{i=1}^{n} (S_g(y_i))^{\delta_i}(g(y_i))^{1-\delta_i}\right) \cdot \left(\prod_{i=1}^{n} (f(y_i))^{\delta_i}(S_f(y_i))^{1-\delta_i}\right) \qquad (19) \end{aligned}$$

But if the distribution of C does not involve any parameters of interest, then the first factor plays no role in the maximization process. Hence, the likelihood function can be taken to be

$$L = \prod_{i=1}^{n} (f(y_i))^{\delta_i}(S_f(y_i))^{1-\delta_i} \qquad (20)$$

In the complete data setting (all $\delta_i = 1$) that is, there is no censoring the likelihood has the usual form

$$L \;=\; \prod_{i=1}^{n} f(y_i) \tag{21}$$

If the p.d.f. is $f(t|\theta)$, where θ belongs to some parameter space $\Theta \subset \mathbb{R}^d$, $d \geqslant 1$, the likelihood function L of the sample is regarded as a function of θ denoted by $L(\theta)$ given by

$$L = L(\theta) \;=\; \prod_{i=1}^{n} f(t_i|\theta). \tag{22}$$

The **maximum likelihood estimator** (MLE), denoted by $\hat{\theta}$, is the value of $\theta \in \Theta$ that maximizes $L(\theta)$ or, equivalently, maximizes the log-likelihood

$$\log L(\theta) = \sum_{i=1}^{n} \log f(t_i|\theta).$$

MLE's possess the *invariance property* : the MLE of a function of θ, say $\phi(\theta)$, is $\phi(\hat{\theta})$.

Under all types of random censoring models, we see that the log-likelihood for the maximization process can be taken to be of the general form

$$\begin{aligned}\log L(\theta) &= \log \prod_{i=1}^{n} (f(y_i|\theta))^{\delta_i} (S_f(y_i|\theta))^{1-\delta_i} \\ &= \sum_{u} \log f(y_i|\theta) + \sum_{c} \log S_f(y_i|\theta) \qquad (23)\end{aligned}$$

where u and c mean sums over the uncensored and censored observations, respectively.

Let $I(\theta)$ denote the Fisher information matrix with elements

$$I_{j,k}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log L(\theta)\right].$$

As we are working with random samples (i.i.d. observations), we point out that $I(\theta)$ can be expressed as

$$I(\theta) = nI_1(\theta)$$

where $I_1(\theta) = \left[-\mathbb{E}(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(y_1|\theta))\right]$, $1 \leq i, j \leq d$ is the Fisher information matrix of any one of the observations.

The MLE $\hat{\theta}$ has the following large sample distribution:

$$\hat{\theta} \Rightarrow MVN(\theta, I^{-1}(\theta)) \tag{24}$$

where $\Rightarrow$ means "**asymptotically distributed**" and MVN denotes multivariate normal.

The i-th diagonal element of $I^{-1}(\theta)$ is the asymptotic variance of the i-th component of θ. The off-diagonal elements are the asymptotic covariances of the corresponding components of θ. If θ is a scalar (real valued), then the asymptotic variance, denoted var_a, of θ is $\mathrm{var}_a(\hat{\theta}) = \dfrac{1}{I(\hat{\theta})}$, where

$I(\theta) = -\mathbb{E}(\frac{\partial^2 \log L(\theta)}{\partial\theta^2})$.

For censored data, this expectation is a function of the censoring distribution G as well as the survival time distribution F. Hence, it is necessary to approximate $I(\theta)$ by the **observed information matrix** $i(\theta)$ evaluated at the MLE $\hat{\theta}$, where

$$i(\theta) = \left[-\frac{\partial^2}{\partial\theta_j \partial\theta_k} \log L(\theta) \right]. \tag{25}$$

For univariate case, $i(\theta) = -\frac{\partial^2 \log L(\theta)}{\partial\theta^2}$ and $\mathrm{var}_a(\hat{\theta})$ is approximated by $(i(\hat{\theta}))^{-1}$.

Delta method:

The **delta method** is useful for obtaining limiting distributions of smooth functions of a MLE. When variance of a MLE includes the parameter of interest, the delta method can be used to remove the parameter in the variance. This is called the variance-stabilization. We describe it for the univariate case.

Let Z be a r.v. with mean μ and variance σ^2 and suppose we want to approximate the distribution of some function $g(Z)$. Take a first order Taylor expansion of $g(Z)$ about μ and ignore the higher order terms to get

$$g(Z) \approx g(\mu) + g'(\mu)(Z - \mu).$$

Then $\mathbb{E}(g(Z)) \approx g(\mu)$ and $\mathrm{var}(g(Z)) \approx (g'(\mu))^2\sigma^2$. The delta method tells us that, if $Z \Rightarrow \mathcal{N}(\mu, \sigma^2)$, then

$$g(Z) \Rightarrow \mathcal{N}(g(\mu), (g'(\mu))^2\sigma^2). \tag{26}$$

Example:

Let $X_1, \cdots, X_n$ be i.i.d. from a Poisson distribution with mean μ. Then the MLE of μ is $\hat{\mu} = \overline{X}_n$. We know that the mean and the variance of $Z = \overline{X}_n$ are μ and $\dfrac{\mu}{n}$. Take $g(Z) = Z^{\frac{1}{2}}$. Then $g(\mu) = \mu^{\frac{1}{2}}$ and $\overline{X}_n^{\frac{1}{2}} \Rightarrow \mathcal{N}(a, b^2)$ with mean $a = \mu^{\frac{1}{2}}$ and variance $b^2 = \frac{1}{4n}$.

Bivariate version of the delta method :

Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \Rightarrow MVN\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right)$$

and suppose we want the asymptotic distribution of $g(x, y)$ where $g : \mathbb{R}^2 \to \mathbb{R}$ is a bivariate function that yields a scalar.

Then the 1^{st} order Taylor approximation of $g(x, y)$ about (μ_x, μ_y) is

$$g(x, y) = g(\mu_x, \mu_y) + (x - \mu_x)\frac{\partial}{\partial x}g(\mu_x, \mu_y) + (y - \mu_y)\frac{\partial}{\partial y}g(\mu_x, \mu_y).$$

Then

$$g(X, Y) \Rightarrow \mathcal{N}(\mu, \sigma^2)),$$

where $\mu \approx g(\mu_x, \mu_y)$ and

$$\sigma^2 \approx \sigma_x^2(\frac{\partial}{\partial x}g(\mu_x, \mu_y))^2 + \sigma_y^2(\frac{\partial}{\partial y}g(\mu_x, \mu_y))^2 + 2\sigma_{xy}\frac{\partial}{\partial x}g(\mu_x, \mu_y)\frac{\partial}{\partial y}g(\mu_x, \mu_y)$$

are the mean and asymptotic variance respectively.

The delta method for a bivariate vector field:

Let Σ denote the asymptotic covariance matrix of the random vector $(X, Y)'$ given above. We want the asymptotic distribution of $g(X, Y)$ where $g : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $(x, y) \mapsto (g_1(x, y), g_2(x, y))$. We apply a 1^{st} order Taylor approximation for vector fields of $\underline{g}$ about $\underline{\mu} = (\mu_x, \mu_y)'$.

Let A denote the Jacobian matrix of $\underline{g}$ evaluated at $\underline{\mu}$. That is,

$$A = \begin{pmatrix} \frac{\partial g_1(\underline{\mu})}{\partial x} & \frac{\partial g_1(\underline{\mu})}{\partial y} \\ \frac{\partial g_2(\underline{\mu})}{\partial x} & \frac{\partial g_2(\underline{\mu})}{\partial y} \end{pmatrix}.$$

Then the first order Taylor approximation is

$$\underline{g}(x, y) = \begin{pmatrix} g_1(\mu_x, \mu_y) \\ g_2(\mu_x, \mu_y) \end{pmatrix} + A' \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}.$$

The delta method now yields the following asymptotic distribution :

$$\underline{g}(X, Y) \Rightarrow MVN(M, \Omega),$$

where $M = \begin{pmatrix} g_1(\mu_x, \mu_y) \\ g_2(\mu_x, \mu_y) \end{pmatrix}$ and $\Omega = A'\Sigma A$ are the asymptotic mean vector and covariance matrix respectively.

3.3 Confidence intervals and tests

For confidence intervals or for testing $H_0: \; \theta = \theta_0$, we can construct the asymptotic z-intervals with the standard errors (s.e.) taken from the diagonal of the asymptotic covariance matrix which is the inverse of the information matrix $I(\theta)$ evaluated at the MLE $\hat{\theta}$ if necessary. The s.e.'s are the square roots of these diagonal values. In summary:

An approximate $(1-\alpha)\times 100\%$ confidence interval for the parameter θ is given by

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} s.e.(\hat{\theta}) \tag{27}$$

where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ quantile of the standard normal distribution and by (25), s.e. is the square root of $\mathrm{var}(\hat{\theta}) \approx (i(\hat{\theta}))^{-1} = -(\frac{\partial^2 \log L(\theta)}{\partial \theta^2})^{-1}$.

If we are testing a vector-valued θ, we have three well known procedures. Assume θ_0 has d-components, $d \geqslant 1$ and $\hat{\theta}$ denotes the MLE.

- The **Wald** statistic

$$(\hat{\theta} - \theta_0)' I(\theta_0)(\hat{\theta} - \theta_0) \Rightarrow \ \chi^2_d \quad \text{under } H_0.$$

- The **Rao** statistic

$$\frac{\partial}{\partial \theta} \log L(\theta_0)' I^{-1} \frac{\partial}{\partial \theta} \log L(\theta_0) \Rightarrow \ \chi^2_d \quad \text{under } H_0.$$

  Note that Rao's method does not use the MLE. Hence, no iterative calculation is necessary.

- The Neyman-Pearson/Wilks **likelihood ratio test** (LRT):
  Let t represent the vector of n observed values; that is, $\underline{t}' = (t_1, \cdots, t_n)$.
  The LRT statistic is given by

$$r^*(\underline{t}) = -2 \log \left(\frac{L(\theta_0)}{L(\hat{\theta})} \right) \Rightarrow \ \chi^2_d \quad \text{under } H_0. \tag{28}$$

To test $H_0 : \theta = \theta_0$ against $H_A : \theta \neq \theta_0$, we reject for small values of $\frac{L(\theta_0)}{L(\hat{\theta})}$. Equivalently, we reject for large values of $r^*(t)$.

For **joint confidence regions** we simply take the region of values that satisfy the elliptical region formed with either the Wald or Rao statistic with $I(\theta)$ or $i(\theta)$ evaluated at the MLE $\hat{\theta}$.

For example, an approximate $(1-\alpha) \times 100\%$ joint confidence region for θ is given by

$$\left\{\theta : \text{Wald} \leqslant \chi^2_\alpha\right\},$$

where χ^2_α is the chi-square upper αth-quantile with d degrees of freedom.

3.4 One sample problem

3.4.1 Fitting data to the exponential model

Case 1 : No censoring

All "failures" are observed, the $T_1, \cdots, T_n$ are iid.

- **Likelihood :** $L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda t_i} = \lambda^n \exp(-\lambda \sum_{i=1}^{n} t_i)$.

- **MLE :** Setting $\dfrac{\partial \log L(\lambda)}{\partial \lambda} = 0$ gives

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{1}{\bar{T}}$$

The likelihood estimator of the mean $\theta = \frac{1}{\lambda}$ is $\hat{\theta} = \bar{T}$.

- **Exact distribution theory :**
  Since the T_i are iid exponential(λ), the sum $\sum_{i=1}^{n} T_i$ has gamma distribution with parameters $k = n$ and λ. From basic theory, we know that

$$2\lambda \sum_{i=1}^{n} T_i = \frac{2n\lambda}{\hat{\lambda}} \rightsquigarrow \chi^2_{2n}. \tag{29}$$

  This can be used as a pivotal statistic to construct both test and confidence interval.

- **Confidence interval (C.I.) for** λ.
  By an observation of a picture of the Chi-square distribution, it is easy to see that, with probability $1 - \alpha$,

$$\chi^2_{1-\alpha/2} \leqslant \frac{2n\lambda}{\hat{\lambda}} \leqslant \chi^2_{\alpha/2}.$$

  It then follows from simple algebra that : a $(1 - \alpha) \times 100\%$ C.I. for λ is given by

$$\frac{\hat{\lambda}}{2n}\chi^2_{1-\alpha/2} \leqslant \lambda \leqslant \frac{\hat{\lambda}}{2n}\chi^2_{\alpha/2}.$$

Let $\theta = \frac{1}{\lambda}$ be the mean of the distribution. It follows that, a $(1-\alpha) \times 100\%$ C.I. for θ is given by

$$\frac{2n\bar{T}}{\chi^2_{\alpha/2}} \leqslant \theta \leqslant \frac{2n\bar{T}}{\chi^2_{1-\alpha/2}}.$$

- **Confidence interval (C.I.) for the p-th quantile.**
  The p-th quantile is the t_p such that $F(t_p) = p$. Thus t_p is such that $1 - e^{-\lambda t_p} = p$. Therefore, $t_p = -\frac{\log(1-p)}{\lambda}$. By the invariance property of MLE's, the MLE of t_p is $\hat{t}_p = -\hat{\lambda}^{-1}\log(1-p)$.

  e.g. the MLE of the median is median $= -\bar{T}\log(0.5) = \bar{T}\log(2)$.

Example with AML1 data : pretending that there is no censored points, we have $n = 11$, $\sum_{i=1}^{n} t_i = 423$, degrees of freedom$=2 \times 11 = 22$, MLE's : $\hat{\theta} = \bar{t} = 38.4545$, $\hat{\lambda} = 0.026$.

For a 95% confidence interval, $\chi^2_{0.02522} = 36.78$, $\chi^2_{0.97522} = 10.98$.

- **A** 95% **C.I. for** λ is $\frac{0.026}{2\times 11} \times 10.98 \leqslant \lambda \leqslant \frac{0.026}{2\times 11} \times 36.78$ or $0.01298 \leqslant \lambda \leqslant 0.04347$.
- **A** 95% **C.I. for** θ the mean survival (in weeks) is $\frac{2.423}{36.78} \leqslant \theta \leqslant \frac{2.423}{10.98}$ or $23 \leqslant \theta \leqslant 77.05$.
- **The MLE of the median :**

$$\text{median} = -\bar{T}\log(0.5) = -38.4545\log(0.5) = 26.6546 \text{ weeks } < \bar{T}.$$

- Test of H_0 : mean $\theta = 30$ weeks against $H_A : \theta \neq 30$ or equivalently $H_0 : \lambda = 1/30 = 0.033$ weeks against $H_A : \lambda \neq 0.033$.

At the 5% level of significance, we can use the exact C.I. for θ obtained above. We reject H_0 if the 95% C.I. does not contain 30. Therefore, we do not reject H_0. That is, the mean survival is not significantly different from 30 weeks.

For one-sided test, the significance-level would be 2.5%. We can base a test on the test-statistic

$$T^* = 2\lambda_0 \sum_{i=1}^{n} T_i \rightsquigarrow \chi^2_{2n} \qquad \text{under } H_0 : \lambda = \lambda_0.$$

To test against $H_A : \lambda \neq \lambda_0$, construct a two-tailed size α critical region. Here

$$T^* = 20033 \times 423 = 28.2$$

At $\alpha = 0.05$, $df = 22$, $\chi^2_{0.975} = 10.98$ and $\chi^2_{0.025} = 36.78$. We fail to reject H_0.

This is a flexible test as you can test one-sided alternatives. For example, to test $H_A : \lambda < \lambda_0$ $(\theta > \theta_0)$, the computed p-value is,

$$p - \text{value} = \mathbb{P}(T^* \geqslant 28.2) = 0.17.$$

Again, we fail to reject H_0. The p-value for the two-sided alternative is then 0.34.

- **The likelihood ratio test (LRT)**
  The LRT can be shown to be equivalent to the two-sided test on the test statistic T^* just above. Therefore, we use the asymptotic distribution and then compare. The test statistic is

$$r^*(\underline{t}) = -2\log\left(\frac{L(\lambda_0)}{L(\hat{\lambda})}\right) \Rightarrow \chi_1^2.$$

  We reject $H_0 : \theta = 30$ when $r^*(\underline{t})$ is large.

$$\begin{aligned} r^*(\underline{t}) &= -2\log L(\lambda_0) + 2\log L(\hat{\lambda}) \\ &= -2n\log(\lambda_0) + 2\lambda_0 n\bar{t} + 2n\log(1/\bar{t}) - 2n \\ &= -2 \times 11 \times \log(1/30) + \frac{2}{30} \times 423 + 2 \times 11 \times \log(11/423) - 2 \times 11 \\ &= 0.7378. \end{aligned}$$

  The p-value=$\mathbb{P}(r^*(\underline{t}) \geqslant 0.7378) \approx 0.39$. Therefore, we fail to reject H_0. This p-value is very close to the exact p-value 0.34 computed above.

Case 2: Random censoring

Let u, c, and n_u denote uncensored, censored, and number of uncensored observations, respectively. The n observed values are now represented by the vectors $\underline{y}$ and $\underline{\delta}$, where $\underline{y}' = (y_1, \cdots, y_n)$ and $\underline{\delta}' = (\delta_1, \cdots, \delta_n)$. Then

- **Likelihood:** By (19) and (23), we have

$$L(\lambda) = \lambda^{n_u} \exp(-\lambda \sum_u y_i) \exp(-\lambda \sum_c y_i) = \lambda^{n_u} \exp(-\lambda \sum_{i=1}^n y_i).$$

$$\log L(\lambda) = n_u \log(\lambda) - \lambda \sum_{i=1}^n y_i, \qquad \frac{\partial^2 \log L(\lambda)}{\partial \lambda^2} = -\frac{n_u}{\lambda^2} = -i(\lambda),$$

We deduce the maximum likelihood estimator

- **MLE**

$$\hat{\lambda} = \frac{n_u}{\sum_{i=1}^n y_i} \qquad \text{and} \qquad \mathrm{var}_a(\hat{\lambda}) = (-\mathbb{E}(\frac{-n_u}{\lambda^2}))^{-1} = \frac{\lambda^2}{\mathbb{E}(n_u)},$$

where $\mathbb{E}(n_u) = n.\mathbb{P}(T \leqslant C)$.

From (24), $\dfrac{\hat{\lambda} - \lambda}{\sqrt{\lambda^2/\mathbb{E}(n_u)}} \Rightarrow \mathcal{N}(0,1)$.

We replace $\mathbb{E}(n_u)$ by n_u since we don't usually know the censoring distribution $G(.)$. We substitute the unknown parameter λ in the asymptotic variance by $\hat{\lambda}$ and obtain $\mathrm{var}_a(\hat{\lambda}) = \dfrac{\hat{\lambda}^2}{n_u} = \dfrac{1}{i(\hat{\lambda})}$, where $i(\lambda)$ is just above. The MLE for the mean $\theta = 1/\lambda$ is simply $\hat{\theta} = \dfrac{\sum_{i=1}^n y_i}{n_u}$.

Example, on the AML data, $n_u = 7$,

$$\hat{\lambda} = \frac{7}{423} = 0.0165 \qquad \text{and} \qquad \mathrm{var}_a(\hat{\lambda}) = \frac{\hat{\lambda}^2}{7} = \frac{0.0165^2}{7}.$$

- **A** 95% **C.I. for** λ is by (27) given by

$$\hat{\lambda} \pm z_{0.025} s.e.(\hat{\lambda}) =: 0.0165 \pm 1.96 \frac{0.0165}{\sqrt{7}} =: [0.004277, 0.0287].$$

- **A 95% C.I. for θ**, the mean survival, can be obtained by inverting the previous interval for λ. This interval is: $[34.8, 233.808]$ weeks.

  Both intervals are very skewed. However, as $\hat{\theta} = 1/\hat{\lambda} = 60.42856$ weeks, we have $\theta = g(\lambda) = 1/\lambda$ and we can use the delta method to obtain the asymptotic variance of θ. As $g'(\lambda) = -\lambda^{-2}$, the asymptotic variance is

$$\mathrm{var}_a(\hat{\theta}) = \frac{1}{\lambda^2 \mathbb{E}(n_u)} \approx \frac{1}{\hat{\lambda}^2 n_u} = \frac{\hat{\theta}^2}{n_u}. \tag{30}$$

  Hence a second 95% C.I. for θ, the mean survival, is given by

$$\hat{\theta} \pm z_{0.025} s.e.(\hat{\theta}) =: 60.42856 \pm 1.96 \frac{1}{\sqrt{70.0165}} =: [15.66246, 105.1947] \text{ weeks.}$$

  Notice this is still skewed, but much less so, and it is much narrower. Here we use the asymptotic variance of θ directly, and hence, eliminate one source of variation. However, the asymptotic variance still depends on λ.

- **The MLE of the p-th quantile**

$$\hat{t}_p = -\frac{1}{\hat{\lambda}} \log(1-p) = -\frac{\sum_{i=1}^n y_i}{n_u} \log(1-p).$$

Thus, the MLE of the median is

$$\text{median} = -\frac{423}{7} \log(0.5) = 41.88 \text{ weeks.}$$

Notice how much smaller the median is compared to the estimate $\hat{\theta} = 60.43$. The median reflects a more typical survival time. The mean is greatly influenced by the one large value 161$+$. Note that

$$\mathrm{var}_a(\hat{t}_p) = (\log(1-p))^2 \, \mathrm{var}_a(\hat{\lambda}^{-1}) \approx (\log(1-p))^2 \frac{1}{\hat{\lambda}^2 . n_u}.$$

The $\mathrm{var}_a(\hat{\lambda}^{-1})$ is given in expression (25). Thus, a 95% C.I. for the median is given by

$$\hat{t}_{0.5} \pm 1.96 \frac{-\log(0.5)}{\hat{\lambda}\sqrt{n_u}} =: 41.88 \pm 1.96 \frac{-\log(0.5)}{0.0165\sqrt{7}} =: [10.76, 73] \text{ weeks.}$$

- **The MLE of the survivor function** $S(t) = e^{-\lambda t}$

$$\hat{S}(t) = e^{-\hat{\lambda} t} = e^{-0.0165t}.$$

For any fixed t, $\hat{S}(t)$ is a function of $\hat{\lambda}$. We can take a log-log transformation and have

$$\log(-\log(\hat{S}(t))) = \log(\hat{\lambda}) + \log(t).$$

Hence,

$$\text{var}_a\{\log(-\log(\hat{S}(t)))\} = \text{var}_a(\log(\hat{\lambda})) \approx \frac{1}{n_u}.$$

It follows from the delta method that for each t,

$$\log(\hat{\lambda}) \Rightarrow \mathcal{N}(\log(\lambda t), \frac{1}{n_u}).$$

Then, with some algebraic manipulation, a $(1-\alpha)\times 100\%$ C.I. for the true probability of survival beyond time t, $S(t)$, is given by

$$\exp\left\{\log(\hat{S}(t)) \exp\left(\frac{z_{\alpha/2}}{\sqrt{n_u}}\right)\right\} \leqslant S(t) \leqslant \exp\left\{\log(\hat{S}(t)) \exp\left(\frac{-z_{\alpha/2}}{\sqrt{n_u}}\right)\right\}.$$

- **The likelihood ratio test:** see (28)

$$
\begin{aligned}
r^*(\underline{t}) &= -2\log L(\lambda_0) + 2\log L(\hat{\lambda}) \\
&= -2n_u \log(\lambda_0) + 2\lambda_0 \sum_{i=1}^{n} y_i + 2n_u \log(\frac{n_u}{\sum_{i=1}^{n} y_i}) - 2n_u \\
&= -2 \times 7 \times \log(1/30) + \frac{2}{30}.423 + 2 \times 7 \times \log(7/423) - 2 \times 7 \\
&= 4.396.
\end{aligned}
$$

The p-value $= \mathbb{P}(r^*(\underline{y}) \geqslant 4.396) \approx 0.036$. Therefore, here we reject $H_0 : \theta = 1/\lambda = 30$ and conclude that mean survival is > 30 weeks.

Computer application

To fit parametric models (with the MLE approach) for censored data, use the S or R function **survReg**. It fits an exponential model to the data, yields point and 95% C.I. estimates for both the mean and the median, and provides a Q-Q plot for diagnostic purposes.

3.4.2 Fitting data to the Weibull and log-logistic models

The S or R **survReg** function is used to fit data to the Weibull, log-logistic or log-normal models as for exponential model which is just a Weibull with shape $\alpha = 1$ and $\theta = 1/\lambda$. **survReg** uses by default a log link function which transforms the problem into estimating location $\mu = -\log(\lambda)$ and scale $\sigma = 1/\alpha$.

Using the function **summary(fit)** resulting from **survReg** evaluated at the "**Weibull**", "**log-logistic**", or "**log-normal**", we get the MLE's $\hat{\mu}$ and $\hat{\sigma}$. Once the parameters are estimated via **survReg**, we can use S functions **pweibull(**q**,**α**,** λ^{-1}**)**, **plogis(**q**,**μ**,** σ**)**, **pnorm(**q**,** μ**,** σ**)**, for the distributions $F(t)$ or **qweibull(**p**,**α**,** λ^{-1}**)**, **qlogis(**p**,**μ**,** σ**)**, **qnorm(**p**,** μ**,** σ**)** for the quantiles t_p to compute estimated survival probabilities and quantiles.

3.5 Two-sample problem

In this section we compare two survival curves from the same parametric family. It is equivalent to compare the two scale parameters $\lambda = (\lambda_1, \lambda_2)$. In the log-transformed problem, this compares the two location parameters $\mu = -\log(\lambda) = (\mu_1, \mu_2)$.

The nonparametric **log-rank test** was used to detect a significant difference between the two K-M survival curves for two groups. We now explore if any of the log-transform distributions, which belong to the location and scale family (16), fit this data adequately. The full model can be expressed as a log-linear model as follows:

$$Y = \log(T) = \tilde{\mu} + \text{error} = \theta + \beta^{*}\text{group} + \text{error}$$

The $\tilde{\mu}$ is called the linear predictor.

In case of two groups model (groupe=1 and groupe =0), $\tilde{\mu}$ has two values

$$\mu_1 = \theta + \beta^{*} \quad \text{and} \quad \mu_2 = \theta.$$

Since $\tilde{\mu} = -\log(\tilde{\lambda})$, we have $\tilde{\lambda} = \exp(-\theta - \beta^{*}\text{group})$ and the two values of $\tilde{\lambda}$ are

$$\lambda_1 = \exp(-\theta - \beta^{*}) \quad \text{and} \quad \lambda_2 = \exp(-\theta).$$

Hence the **null hypothesis** is:

$$H_0 : \lambda_1 = \lambda_2 \iff \mu_1 = \mu_2 \iff \beta^{*} = 0.$$

Recall that the scale parameter in the log-transform model is the reciprocal of the shape parameter in the original model; that is, $\sigma = 1/\alpha$. We test H_0 under each of the following cases:

. **Case 1 :** Assume equal shapes (α); that is, we assume equal scales $\sigma_1 = \sigma_2 = \sigma$. Hence, error$=\sigma Z$, where the random variable Z has either the standard extreme value, standard logistic, or the standard normal distribution. Recall by standard, we mean $\mu = 0$ and $\sigma = 1$.

. **Case 2 :** Assume different shapes; that is, $\sigma_1 \neq \sigma_2$.

We can use the **S** or **R** program to fit the data to the Weibull model and conduct formal tests.

Model 1: Data come from a same distribution. The Null Model is

$$Y = \log(T) = \theta + \sigma Z,$$

where Z is a standard extreme value random variable. Example of S code:

```
> attach(aml)
> weib.fit0 <- survReg(Surv(weeks,status)~1,dist="weib")
> summary(weib.fit0)
```

Model 2 / Case 1: Data come from distributions with different locations and equal scales σ. Express this model by

$$Y = \log(T) = \theta + \beta^{*}\textsf{group} + \sigma Z.$$

Example of S code.

```
> weib.fit1 <- survReg(Surv(weeks,status)~group,dist="weib")
> summary(weib.fit1)
```

Model 2 / Case 2: Data come from distributions with different locations and different scales. Express this model by

$$Y = \log(T) = \theta + \beta^{*}\text{group} + \text{error}.$$

Fit each group separately. On each group run a **survReg** to fit the data **weib.fit2.0** and **weib.fit2.1**. This gives the **MLE**'s of the two locations μ_1 and μ_2 and the two scales σ_1 and σ_2. Example of S code.

```
> weib.fit2.0 <- survReg(Surv(weeks,status)~1,data=aml[aml$group==0,],
dist="weib")
> weib.fit2.1 <- survReg(Surv(weeks,status)~1,data=aml[aml$group==1,],
dist="weib")
> summary(weib.fit2.0)

> summary(weib.fit2.1)
```

To test the reduced model against the full model, use the **LRT** e.g.

```
> loglik3 <- weib.fit20$loglik[2]+weib.fit21$loglik[2]
> loglik3
[1] -79.84817
> lrt23 <- -2*(weib.fit1$loglik[2]-loglik3)
> lrt23
[1] 1.346954
> 1 - pchisq(lrt23,1)
[1] 0.2458114 # Retain Model 2.
```

The following table summarizes the three models weib.fit0, 1, and 2:

Model	Calculated parameters	the Picture
1 (0)	$\theta,\ \sigma$	same location, same scale
2 (1)	$\theta,\ \beta^*,\ \mu_1,\ \mu_2,\ \sigma \equiv \mu_1$	different locations, same scale
3 (2)	$\mu_1,\ \mu_2,\ \sigma_1,\ \sigma_2$	different locations, different scales

If we use the log-logistic and log-normal distribution to estimate Model 2, the form of the log-linear model is the same. The distribution of error terms is what changes.

Prelude to parametric regression models

As a prelude to parametric regression models presented in the next chapter, we continue to explore Model 2 under the assumption that $T \rightsquigarrow$ Weibull. That is, we explore

$$\begin{aligned} Y &= \log(T) \\ &= \theta + \beta^* \textsf{group} + \sigma Z \\ &= \tilde{\mu} + \sigma Z \end{aligned}$$

where Z is a standard extreme minimum value random variable. Let the linear predictor $\tilde{\mu} = -\log(\tilde{\lambda})$ and $\sigma = 1/\alpha$. It follows from page 46 that the hazard function for the Weibull in this context is expressed as

$$\begin{aligned} \lambda(t|\textsf{group}) &= \alpha \tilde{\lambda}^{\alpha} t^{\alpha-1} \\ &= \alpha \lambda^{\alpha} t^{\alpha-1} \exp(\beta \textsf{group}) \\ &= \lambda_0(t) \exp(\beta \textsf{group}), \end{aligned} \tag{31}$$

when we set $\lambda = \exp(-\theta)$ and $\beta = -\beta^*/\sigma$.

The $\lambda_0(t) = \alpha\lambda^\alpha t^{\alpha-1}$ denotes the baseline hazard; that is, when group $=$ 0 or $\beta = 0$. Thus, $\lambda_0(t)$ is the hazard function for the Weibull with scale parameter λ, which is free of any covariate.

The hazard ratio (**HR**) of group 1 to group 0 is

$$HR = \frac{\lambda(t|1)}{\lambda(t|0)} = \frac{\exp(\beta)}{\exp(0)} = \exp(\beta).$$

If we believe the Weibull model is appropriate, the **HR** is constant over follow-up time t. That is, the graph of **HR** is a horizontal line with height $\exp(\beta)$.

We say the Weibull enjoys the *proportional hazards property* to be formally introduced in the next Chapter.

On the AML data, we have $\hat{\beta} = \dfrac{-\hat{\beta}^*}{\hat{\sigma}} = \dfrac{-0.929}{0.791} = -1.1745$. Therefore, the estimated **HR** is $\widehat{HR} = \dfrac{\hat{\lambda}(t|1)}{\hat{\lambda}(t|0)} = \exp(-1.1745) \approx 0.31$.

That is, the maintained group has 31% of the risk of the control group's risk of relapse. Or, the control group has (1/0.31)=3.23 times the risk of the maintained group of relapse at any given time t. The **HR** is a measure of effect that describes the relationship between time to relapse and group.

If we consider the ratio of the estimated survival probabilities, say at $t = 31$ weeks, since $\hat{\hat{\lambda}} = \exp(-\hat{\hat{\mu}})$, we obtain $\dfrac{\hat{S}(31|1)}{\hat{S}(31|0)} = \dfrac{0.652}{0.252} \approx 2.59$.

The maintained group is 2.59 times more likely to stay in remission at least 31 weeks. The Weibull survivor function $S(t)$ is given in a table on page 46.

4 REGRESSION MODELS

Let T denote failure time and $x = (x_1, \cdots, x_p)'$ represent a vector of available covariates. We are interested in modelling and determining the relationship between T and x. The primary question is: Do any subsets of the d covariates help to explain survival time ? If so, how and by what estimated quantity ?

Exemple 4.1 Let

- x_1 denote the sex ($x_1 = 1$ for males and $x_1 = 0$ for females),
- x_2=Age at diagnosis,
- $x_3 = x_1.x_2$ (interaction),
- T=survival time.

Here we introduce four models: the exponential, the Weibull, the Cox proportional hazards, and the accelerated failure time model.

4.1 Exponential regression model

We first generalize the exponential distribution. Recall that for the exponential distribution, the hazard function $\lambda(t) = \lambda$ is constant with respect to time and that $\mathbb{E}(T) = \frac{1}{\lambda}$. We model the hazard rate λ as a function of the covariate vector x.

We assume the hazard function at time t for an individual has the form

$$\lambda(t|x) = \lambda_0(t).k(x'\beta) = \lambda.k(x'\beta),$$

where $\lambda > 0$ is a constant, $\beta = (\beta_1, \beta_2, \cdots, \beta_p)'$ is a vector of regression parameters (coefficients), $x'\beta = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$ and k is a specified **link function**.

The function $\lambda_0(t)$ is called the baseline hazard. It's the value of the hazard function when the covariate vector $x = 0$ or $\beta = 0$.

The most natural choice for k is $k(x) = \exp(x)$, which implies

$$\begin{aligned} \lambda(t|x) &= \lambda. \exp(x'\beta) \\ &= \lambda. \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) \\ &= \lambda. \exp(\beta_1 x_1) \times \exp(\beta_2 x_2) \times \cdots \times \exp(\beta_p x_p). \end{aligned}$$

This says that the covariates act multiplicatively on the hazard rate. Equivalently, this specifies

$$\log(\lambda(t|x)) = \log(\lambda) + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p = \log(\lambda) + (x'\beta) = \log(\lambda) + \eta.$$

That is, the covariates act additively on the log failure rate. The quantity $\eta = x'\beta$ is called the **linear predictor of the log-hazard**. Then the survivor function of T given x is

$$S(t|x) = \exp(-\lambda(t|x)t) = \exp(-\lambda \exp(x'\beta)t)$$

and thus, the p.d.f. of T given x is

$$f(t|x) = \lambda(t|x)S(t|x) = \lambda \exp(x'\beta) \exp(-\lambda \exp(x'\beta)t).$$

Recall from Section 4, subsection 1, page 48, that if T is distributed exponentially, $Y = \log(T)$ is distributed as the extreme (minimum) value distribution with scale parameter $\sigma = 1$. Here, given x, we have

$$\tilde{\mu} = -\log(\lambda(t|x)) = -\log(\lambda \exp(x'\beta)) = -\log(\lambda) - x'\beta \quad \text{and} \quad \sigma = 1.$$

Therefore, given x,

$$Y = \log(T) = \tilde{\mu} + \sigma Z = \beta_0^* + x'\beta^* + Z,$$

where $\beta_0^* = -\log(\lambda)$, $\beta^* = -\beta$ and $Z \rightsquigarrow f(z) = \exp(z - e^z)$, $-\infty < z < +\infty$, the standard extreme (minimum) value distribution. The quantity $\tilde{\mu} = \beta_0^* + x'\beta^*$ is called the *linear predictor of the log-time*.

In summary, $\lambda(t|x) = \lambda \exp(x'\beta)$ is a log-linear model for the failure rate and transforms into a linear model for $Y = \log(T)$ in that the covariates act additively on Y.

4.2 Weibull regression model

We generalize the Weibull distribution to regression in a similar fashion. Recall that its hazard function is $\lambda(t) = \alpha\lambda^{\alpha}t^{\alpha-1}$.

To include the covariate vector x write the hazard for a given x as

$$\begin{aligned}\lambda(t|x) &= \lambda_0(t).\exp(x'\beta)\\ &= \alpha\lambda^{\alpha}t^{\alpha-1}\exp(x'\beta) = \alpha\left(\lambda(\exp(x'\beta))^{\frac{1}{\alpha}}\right)^{\alpha}t^{\alpha-1}\\ &= \alpha(\tilde{\lambda})^{\alpha}t^{\alpha-1}. \qquad (32)\end{aligned}$$

where $\tilde{\lambda} = \lambda.\exp(x'\beta)^{\frac{1}{\alpha}}$.

Again notice that

$$\begin{aligned}\log(\lambda(t|x)) &= \log(\alpha) + \alpha\log(\tilde{\lambda}) + (\alpha-1)\log(t)\\ &= \log(\alpha) + \alpha\log(\lambda) + x'\beta + (\alpha-1)\log(t).\end{aligned}$$

We know that if $T \rightsquigarrow$ Weibull, then given x, $Y = \log(T) = \tilde{\mu} + \sigma Z$ with $Z \rightsquigarrow$ standard extreme value distribution, where $\sigma = \frac{1}{\alpha}$ and

$$\tilde{\mu} = -\log(\tilde{\lambda}) = -\log(\lambda(\exp(x'\beta))^{\frac{1}{\alpha}}) = -\log(\lambda) - \frac{1}{\alpha}x'\beta. \tag{33}$$

Therefore,

$$\begin{aligned} Y &= \beta_0^* + x'\beta^* + \sigma Z \\ &= \tilde{\mu} + \sigma Z, \end{aligned} \tag{34}$$

where $\beta_0^* = -\log(\lambda)$, $\beta^* = -\sigma\beta$ and $\tilde{\mu} = \beta_0^* + x'\beta^*$. It then follows from the table on page 46 that the survivor function of T given x is

$$S(t|x) = \exp(-(\tilde{\lambda}t)^{\alpha}). \tag{35}$$

From the relationship $\Lambda(t|x) = -\log(S(t|x))$ for a given x and from expression (33) for $\log(\tilde{\lambda})$, an expression for the log-cumulative hazard function follows :

$$\begin{aligned} \log(\Lambda(t|x)) &= \alpha \log(\tilde{\lambda}) + \alpha \log(t) \\ &= \alpha \log(\lambda) + \alpha \log(t) + x'\beta \\ &= \log(\Lambda_0(t)) + x'\beta, \end{aligned} \tag{36}$$

where $\Lambda_0(t) = -\log(S_0(t)) = (\lambda t)^\alpha$.

The log of the cumulative hazard function is linear in $\log(t)$ and in the β coefficients. Thus, for a fixed x value, the plot of $\Lambda(t|x)$ against t on a log-log scale is a straight line with slope α and intercept $x'\beta + \alpha \log(\lambda)$. Expression (36) in conjunction with expression (32) and relation between $\Lambda(t|x)$ and $\lambda(t|x)$ give

$$\Lambda(t|x) = \Lambda_0(t) \exp(x'\beta) = (\lambda t)^\alpha \exp(x'\beta). \tag{37}$$

In summary, for both the exponential and Weibull regression model, the effects of the covariates x act multiplicatively on the hazard function $\lambda(t|x)$ which is clear from the form

$$\begin{aligned} \lambda(t|x) &= \lambda_0(t).\exp(x'\beta) \\ &= \lambda_0(t).\exp(\beta_1 x_1 + \cdots + \beta_p x_p) \\ &= \lambda_0(t).\exp(\beta_1 x_1).\exp(\beta_2 x_2)\cdots\exp(\beta_p x_p). \end{aligned} \tag{38}$$

This suggests the more general **Cox proportional hazards model**, presented in the next section. Further, both are log-linear models for T in that these models transform into a linear model for $Y = \log(T)$. That is, the covariates x act additively on $\log(T)$ (multiplicatively on T), which is clear from the form

$$Y = \log(T) = \tilde{\mu} + \sigma Z = \beta_0^* + x'\beta^* + \sigma Z.$$

This suggests a more general class of log-linear models called **accelerated failure time models** discussed in a further Section below.

4.3 Cox proportional hazards (PH) model

For the Cox (1972) PH model, the hazard function is

$$\lambda(t|x) = \lambda_0(t).\exp(x'\beta), \tag{39}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function free of the covariates x. The covariates act multiplicatively on the hazard. Clearly, the exponential and Weibull are special cases. At two different points x and y, the proportion

$$\frac{\lambda(t|x)}{\lambda(t|y)} = \exp((x' - y')\beta), \tag{40}$$

called the hazard ratio, is constant with respect to time t. This defines the **proportional hazards property**.

For any PH model, which includes the Weibull model as well as the Cox model, the **survivor function of** T given x is

$$\begin{aligned} S(t|x) &= \exp(-\int_0^t \lambda(u|x)du) = \exp\left(-\exp(x'\beta)\int_0^t \lambda_0(u)du\right) \\ &= \left(\exp(-\int_0^t \lambda_0(u)du)\right)^{\exp(x'\beta)} = (S_0(t))^{\exp(x'\beta)}. \end{aligned}$$

where $S_0(t)$ denotes the baseline survivor function.

The p.d.f. of T given x then is

$$f(t|x) = \lambda_0(t)\exp(x'\beta)(S_0(t))^{\exp(x'\beta)}.$$

There are two important generalizations :

(1) The baseline hazard $\lambda_0(t)$ can be allowed to vary in specified subsets of the data.

(2) The regression variables x can be allowed to depend on time; that is, $x = x(t)$.

4.4 Cox's partial likelihood

The hazard function defined in (39) depends on the baseline hazard $\lambda_0()$. Hence, so does the p.d.f. Cox (1975) defines a likelihood based on conditional probabilities which are free of the baseline hazard. His estimate is obtained from maximizing this likelihood. In this way he avoids having to specify $\lambda_0(.)$ at all. This likelihood is derived heuristically.

Let $t_{(1)}, \cdots, t_{(r)}$ denote the $r \leqslant n$ distinct ordered (uncensored) death times, so that $t_{(j)}$ is the j-th ordered death time. Let $x_{(j)}$ denote the vector of covariates associated with the individual who dies at $t_{(j)}$. Then, the **Cox partial likelihood function**, denoted by $L_c(\beta)$, is defined by

$$L_c(\beta) = \prod_{j=1}^{r} L_j(\beta) = \prod_{j=1}^{r} \frac{\exp(x'_{(j)}\beta)}{\sum_{l \in \mathcal{R}(t_{(j)})} \exp(x'_l \beta)}.$$

Recall that in the random censoring model we observe the times $y_1, \cdots, y_n$ along with the associated $\delta_1, \cdots, \delta_n$ where $\delta_i = 1$ if the y_i is uncensored (i.e., the actual death time was observed) and $\delta_i = 0$ if the y_i is censored. We can now give an equivalent expression for the partial likelihood function in terms of all n observed times :

$$L_c(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(x_i'\beta)}{\sum_{l \in \mathcal{R}(y_i)} \exp(x_l'\beta)} \right)^{\delta_i} . \tag{41}$$

Remarks 4.1

1. *Cox's estimates maximize the log-partial likelihood.*
2. *To analyze the effect of covariates, there is no need to estimate the nuisance parameter $\lambda_0(t)$, the baseline hazard function.*
3. *This partial likelihood is not a true likelihood in that it does not integrate out to 1 over $\{0,1\}^n \times \mathbb{R}_+^n$.*

4.5 Model Diagnostics

The Cox proportional-hazards regression model is fit in **S** with the **coxph** function (located in the survival library in **R**).

As is the case for a linear or generalized linear model, it is desirable to determine whether a fitted Cox regression model adequately describes the data. Three kinds of diagnostics are considered :

1. for violation of the assumption of proportional hazards;
2. for influential data and
3. for nonlinearity in the relationship between the log hazard and the covariates.

All of these diagnostics use the residuals method for **coxph** objects, which calculates several kinds of residuals.

Cox-Snell residual

The i-th Cox-Snell residual is defined as

$$r_i = -\log(\hat{S}_0(t_i))e^{[\sum \hat{b}_j x_{ij}]}.$$

These r_i-values apply to survival distribution models in general. For the Weibull proportional hazard model, the Cox-Snell residual values are

$$r_i = (\hat{\lambda}_0 t_i^{\hat{\gamma}})e^{[\sum \hat{b}_j x_{ij}]}$$

using the hazards model estimates λ_0, γ and the k estimated regression coefficients $\hat{b}_j$.

Modified Cox-Snell residual

For survival models such as the Weibull proportional hazards model, a residual value can be calculated for all observations (complete and censored). A residual value for each observation is defined as $m_i = r_i$ when the observation t_i is complete and $m_i = r_i + 1$ when the observation t_i is censored.

Martingale residual

The i-th martingale residual is defined as $\hat{M}_i = \delta_i - r_i$, where $\delta_i = 1$ for complete observation and $\delta_i = 0$ for censored observation. The $\hat{M}_i$ take values in $(-\infty, 1]$ and are always negative for censored observations. In large samples, the martingale residuals are uncorrelated and have expected value equal to zero. But they are not symmetrically distributed about zero.

Deviance residual

The deviance residuals are useful in detecting outliers. The ith deviance residual **under the extreme value model**, is given by

$$D_i = \text{sign}(r_i)\sqrt{-2(r_i + \delta_i \log[\delta_i - r_i])}.$$

where $\text{sign}(r_i) = -1$ for $r_i < 0$ and $+1$ otherwise.

4.6 Accelerated failure time model

This model is a log-linear regression model for T. $Y = \log(T)$ is modelled as a linear function of the covariate x :

$$Y = x'\beta + Z^*,$$

where Z^* has a certain distribution. Then

$$T = \exp(Y) = \exp(x'\beta^*). \exp(Z^*) = \exp(x'\beta^*).T^*$$

where $T^* = \exp(Z^*)$.

Suppose that T^* has hazard function $\lambda_0^*(t^*)$ which is free of the covariate vector x. The hazard function of T for a given x can be written in terms of the baseline function λ_0^* according to

$$\lambda(t|x) = \lambda_0^*(\exp(-x'\beta^*)t). \exp(-x'\beta^*)). \tag{42}$$

The survivor function of T given x is

$$S(t|x) = \exp\left(-\exp(-x'\beta)\int_0^t \lambda_0^*(\exp(-x'\beta^*)u)du\right).$$

Change the integration variable to $v = \exp(-x'\beta^*)u$. We have $dv = \exp(-x'\beta^*)du$ and $0 < v < \exp(-x'\beta^*)t$. Then for the accelerated failure time model,

$$\begin{aligned} S(t|x) &= \exp\left(-\exp(-x'\beta)\int_0^{\exp(-x'\beta)t} \lambda_0^*(v)dv\right) \\ &= S_0^*(\exp(-x'\beta)t) = S_0^*(t^*). \end{aligned} \tag{43}$$

where $S_0^*(t)$ denotes the baseline survivor function. We notice that the covariate x changes the scale of the horizontal (t) axis. For example, if $x'\beta^*$ increases, then the last term in expression (43) increases. In this case it has decelerated the time to failure. This is why the log-linear model defined here is called the **accelerated (decelerated) failure time model**.

Remark 4.1

1. *We have seen that the Weibull regression model, which includes the exponential, is a special case of both the Cox PH model and the accelerated failure time model. It can be shown that the only log-linear models that are also PH models are the Weibull regression models.*

2. *Through the* **partial likelihood** *(Cox, 1975) we obtain estimates of the coefficients β that require no restriction on the baseline hazard $\lambda_0(t)$. The* **S** *function* **coxph** *implements this.*

3. *For the accelerated failure time models we specify the baseline hazard function $\lambda_0(t)$ by specifying the distribution function of Z^*.*

4.7 AIC procedure for variable selection

Comparisons between a number of possible models, can be made on the basis of the statistic

$$\text{AIC} = -2 \times \log(\text{maximum likelihood}) + k \times p,$$

where p is the number of parameters in each model under consideration and k a predetermined constant. This statistic is called **Akaike's (1974) information criterion (AIC)**; the smaller the value of this statistic, the better the model. Here we shall take $k = 2$. For other choice of values for k, see the remarks at the end of this section.

For the parametric models discussed, the **AIC** is given by

$$\text{AIC} = -2 \times \log(\text{maximum likelihood}) + 2 \times (a + b), \qquad (44)$$

where a is the number of parameters in the specific model and b the number of one-dimensional covariates. For example, $a = 1$ for the exponential model, $a = 2$ for the Weibull, log-logistic, and log-normal models.

Motorette data example :

The data set given in Table 5 below was obtained by Nelson and Hahn (1972) and discussed again in Kalbfleisch and Prentice (1980). Hours to failure of motorettes are given as a function of operating temperatures 1500C, 1700C, 1900C, or 2200C. There is severe (Type I) censoring, with only 17 out of 40 motorettes failing. The primary purpose of the experiment was to estimate certain percentiles of the failure time distribution at a design temperature of 1300C. We see that this is an accelerated process. The experiment is conducted at higher temperatures to speed up failure time.

The authors use the single regressor variable x = 1000/(273.2+Temperature). They also omit all ten data points at temperature level of 1500 C. The data is entered into a data frame called motorette. It contains

Table 5: *Hours to failure of Motorettes.*

Temperature	Hours to failure
150C	8064+, 8064+, 8064+, 8064+, 8064+, 8064+, 8064+, 8064+, 8064+, 8064+
170C	1764, 2772, 3444, 3542, 3780, 4860, 5196, 5448+, 5448+, 5448+
190C	408, 408, 1344, 1344, 1440, 1680+, 1680+, 1680+, 1680+, 1680+
220C	408, 408, 504, 504, 504, 528+, 528+, 528+, 528+, 528+
$n = 40$ n_u = number of uncensored =17	

Table 6: *Results of fitting parametric models to the Motorette data.*

Model	log-likelihood		AIC
exponential	intercept only	-155.875	$311.750 + 2(1) = 313.750$
	both	-151.803	$303.606 + 2(1 + 1) = 307.606$
Weibull	intercept only	-155.681	$311.363 + 2(2) = 315.363$
	both	-144.345	$288.690 + 2(2 + 1) = 294.690$
Log-logistic	intercept only	-155.732	$311.464 + 2(2) = 315.464$
	both	-144.838	$289.676 + 2(2 + 1) = 295.676$
Log-normal	intercept only	-155.681	$310.036 + 2(2)= 314.036$
	both	-145.867	$291.735 + 2(2 + 1) = 297.735$

We fit the exponential, Weibull, log-logistic, and log-normal models. The log likelihood and the AIC for each model are reported in Table 6.

$$\begin{aligned} \text{intercept only:} \quad y &= \log(t) = \hat{\beta}_0^* + \sigma Z \\ \text{both:} \quad y &= \log(t) = \hat{\beta}_0^* + \beta_1^* + \sigma Z \end{aligned}$$

where the distributions of Z are standard extreme value, standard logistic, and standard normal, respectively.

The S code for computing the AIC for a number of specified distributions

```
> attach(motorette)              # attach the data frame motorette
                                 # to avoid continually referring to it.
# Weibull fit
> weib.fit <- survReg(Surv(time,status)~x,dist="weibull")
> weib.fit$loglik                  # the first component for intercept
                                   # only and the second for both
[1] -155.6817   -144.3449

> -2*weib.fit$loglik               # -2 times maximum log-likelihood
[1] 311.3634     288.6898
```

```
# exponential fit
> exp.fit <- survReg(Surv(time,status)~x,dist="exp")
> -2*exp.fit$loglik
[1] 311.7501    303.6064

# log-normal fit
> lognormal.fit <- survReg(Surv(time,status)~x,dist="lognormal")
> -2*lognormal.fit$loglik
[1] 310.0359    291.7345

# log-logistic fit
> loglogistic.fit <- survReg(Surv(time,status)~x,dist="loglogistic")
> -2*loglogistic.fit$loglik
[1] 311.4636    289.6762
> detach()    # Use this to detach the data frame when no
              # longer in use.
```

The Weibull model is to some extent preferable to the log-normal on account of the larger maximized log likelihood. From Table 6, we find that the Weibull distribution provides the best fit to this data, the log-logistic distribution is a close second, and the log-normal distribution is the third. When there are no subject matter grounds for model choice, the model chosen for initial consideration from a set of alternatives might be the one for which the value of AIC is a minimum. It will then be important to confirm that the model does fit the data using the methods for model checking.

Estimation and testing : the Weibull model

The **S** function **survReg** fits the times T as log-failure times $Y = \log(T)$ to model

$$Y = \beta_0^* + x'\beta^* + \sigma Z,$$

where Z has the standard extreme value distribution.

Further, when we re-express Y as

$$Y = x'\beta^* + Z^*,$$

where $Z^* = \beta_0^* + \sigma Z$, and this model is an accelerated failure time model. Here $Z^* \rightsquigarrow$ extreme value with location β_0^* and scale σ. The linear predictor $\tilde{\mu}$ given on page 96 is

$$\tilde{\mu} = -\log(\tilde{\lambda}) = \beta_0^* + x'\beta^* \qquad (45)$$

with $\beta_0^* = -\log(\lambda)$ and $\beta^* = -\sigma\beta$, the vector β denoting the coefficients in the Weibull hazard on page 96 and, $\sigma = 1/\alpha$, where α denotes the Weibull shape parameter.

Let $\hat{\beta}_0^*$, $\hat{\beta}^*$ and $\hat{\sigma}$ denote the MLE's of the parameters. To test $H_0: \ \beta_j^* = \beta_j^{*0}, \ j = 1, \cdots, m,$ use

$$\frac{\hat{\beta}_j^* - \beta_j^{*0}}{s.e.(\hat{\beta}_j^*)} \rightsquigarrow \mathcal{N}(0,1) \quad \text{under } H_0.$$

An approximate $(1-\alpha) \times 100\%$ confidence interval for β_j^* is given by

$$\hat{\beta}_j^* \pm z_{\frac{\alpha}{2}} s.e.(\hat{\beta}_j^*),$$

where $z_{\frac{\alpha}{2}}$ is taken from the $\mathcal{N}(0,1)$ table. Inferences concerning the intercept β_0 follow analogously.

At the point $x = x_0$, the MLE of the $(p\times 100)$th percentile of the distribution of $Y = \log(T)$ is

$$\hat{Y}_p = \hat{\beta}_0^* + x_0'\hat{\beta}^* + \hat{\sigma} z_p = (1, x_0', z_p) \begin{pmatrix} \hat{\beta}_0^* \\ \hat{\beta}^* \\ \hat{\sigma} \end{pmatrix}$$

where z_p is the $(p \times 100)$th percentile of the error distribution, which, in this case, is standard extreme value. The estimated variance of $\hat{Y}_p$ is

$$\operatorname{var}(\hat{Y}_p) = (1, x_0', z_p)\hat{\Sigma} \begin{pmatrix} 1 \\ x_0 \\ z_p \end{pmatrix} \tag{46}$$

where $\hat{\Sigma}$ is the estimated variance-covariance matrix of $\hat{\beta}_0^*$, $\hat{\beta}^*$ and $\hat{\sigma}$. Then an approximate $(1-\alpha) \times 100\%$ confidence interval for the $(p \times 100)$th percentile of the log-failure time distribution is given by

$$\hat{Y}_p \pm z_{\frac{\alpha}{2}} s.e.(\hat{Y}_p),$$

where $z_{\frac{\alpha}{2}}$ is taken from the $\mathcal{N}(0,1)$ table. These are referred to as the **uquantile** type in the S function **predict**. The MLE of t_p is $\exp(\hat{Y}_p)$. To obtain confidence limits for t_p, take the exponential of the endpoints of the above confidence interval.

The function **predict**, a companion function to **survReg**, conveniently reports both the quantiles in time and the uquantiles in log(time) along with their respective s.e.'s. We often find the confidence intervals based on uquantiles are shorter than those based on quantiles. See, for example, the results at the end of this section.

Doing the analysis using S:

In **S**, we fit the model

$$Y = \log(\text{time}) = \beta_0^* + \beta_1^* x + \sigma Z,$$

where Z has the standard extreme value distribution. the $(p \times 100)$th percentile of the standard extreme (minimum) value distribution (see Table 4) is

$$z_p = \log(-\log(1-p)).$$

The function **survReg** outputs the estimated variance-covariance matrix $\hat{V}$ for the MLE's $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and $\hat{\tau} = \log \hat{\sigma}$. However, internally it computes $\hat{\Sigma}$ to estimate the $\mathrm{var}(\hat{Y}_p)$.

The following is an **S** program along with modified output. The function **survReg** is used to fit a Weibull regression model. Then the 15th and 85th percentiles as well as the median failure time are estimated with corresponding standard errors. We also predict the failure time in hours at $x_0 = 2.480159$, which corresponds to the design temperature of 1300C.

```
> attach(motorette)
> weib.fit <- survReg(Surv(time,status)~x,dist="weibull")
> summary(weib.fit)

             Value   Std. Error          z     p
(Intercept) -11.89        1.966      -6.05  1.45e-009
          x  9.04         0.906       9.98  1.94e-023
Log(scale)  -1.02         0.220      -4.63  3.72e-006
```

```
> weib.fit$var   # The estimated covariance matrix of the
                 # coefficients and log(sigmahat).

              (Intercept)            x   Log(scale)
(Intercept)    3.86321759  -1.77877653   0.09543695
          x   -1.77877653   0.82082391  -0.04119436
Log(scale)     0.09543695  -0.04119436   0.04842333

> predict(weib.fit,newdata=list(x),se.fit=T,type="uquantile",
p=c(0.15,0.5,0.85))

# newdata is required whenever
# uquantile is used as a type whereas quantile
# uses the regression variables as default.
# This returns the estimated quantiles in log(t)
# along with standard error as an option.
# Estimated quantiles in log(hours) and standard errors in
# parentheses. The output is edited because of redundancy.

x=2.256318     7.845713      8.369733      8.733489
               (0.1806513)   (0.12339772)   (0.1370423)
x=2.158895     6.965171      7.489190       7.852947
               (0.1445048)   (0.08763456)   (0.1189669)
x=2.027575     5.778259      6.302279       6.666035
               (0.1723232)   (0.14887233)   (0.1804767)

>predict(weib.fit,newdata=data.frame(x=2.480159),se.fit=T,
type="uquantile",p=c(0.15,0.5,0.85))
# Estimated quantiles in log(hours) at the new x value = 2.480159;
# i.e., the design temperature of 130 degrees Celsius.
x=2.480159     9.868867     10.392887     10.756643
               (0.3444804)  (0.3026464)  (0.2973887)
```

```
>sigmahat <- weib.fit$scale
>alphahat <- 1/sigmahat #       estimate of shape
>coef <- weib.fit$coef
>lambdatildehat <- exp(- coef[1] - coef[2]*2.480159)
# estimate of scale
> pweibull(25000,alphahat,1/lambdatildehat) # Computes the
# estimated probability that a motorette failure time
# is less than or equal to 25,000 hours. pweibull is
# the Weibull distribution function in S.
[1] 0.2783054  # estimated probability
> Shat <- 1 - 0.2783054  # survival probability at 25,000
# hours. About 72% of motorettes are still working
# after 25,000 hours at x=2.480159; i.e., the design
# temperature of 130 degrees Celsius.
> xl <- levels(factor(x)) # Creates levels out of the distinct
                          # x-values.
```

```
> ts.1 <- Surv(time[as.factor(x)==xl[1]],status[as.factor(x)==xl[1]])
# The first group of data
> ts.2 <- Surv(time[as.factor(x)==xl[2]],status[as.factor(x)==xl[2]])
                                                     # The second
> ts.3 <- Surv(time[as.factor(x)==xl[3]],status[as.factor(x)==xl[3]])
                                                     # The third
> par(mfrow=c(2,2))  # divides a screen into 2 by 2 pieces.
> Svobj <- list(ts.1,ts.2,ts.3)  # Surv object
> qq.weibreg(Svobj,weib.fit)  # The first argument takes
# a Surv object and the second a survReg object.
# Produces a Weibull Q-Q plot.
> qq.loglogisreg(Svobj,loglogistic.fit) # log-logistic
# Q-Q plot
> qq.lognormreg(Svobj,lognormal.fit)  # log-normal Q-Q plot
> detach()
```

Results:

- From **summary(weib.fit)**, we learn that $\hat{\sigma} = \exp(-1.02) = 0.3605949$ and $\hat{\hat{\mu}} = -\log(\hat{\hat{\lambda}}) = \hat{\beta}_0^* + \hat{\beta}_1^* x = -11.89 + 9.04x$.

  Thus, we obtain $\hat{\alpha} = \dfrac{1}{0.3605949} = 2.773195$ and $\hat{\hat{\lambda}} = \exp(11.89 - 9.04 \times 2.480159) = 0.0000267056$ at $x = 2.480159$. Note also that both the intercept and covariate x are highly significant with p-values 1.4510^{-9} and 1.9410^{-23} respectively.

- It follows from Section 2 that the estimated hazard function is

$$\hat{\lambda}(t|x) = \frac{1}{\hat{\sigma}} t^{\frac{1}{\hat{\sigma}}-1} (\exp(-\hat{\hat{\mu}}))^{\frac{1}{\hat{\sigma}}}$$

  and the estimated survivor function is

$$\hat{S}(t|x) = \exp\left[-\exp(-\hat{\hat{\mu}})t)^{\frac{1}{\hat{\sigma}}}\right].$$

- The point estimate $\hat{\beta}_1$ of β_1 is $-\hat{\sigma}^{-1}\hat{\beta}_1^*$. A 95% C.I. for β_1 based on the delta method is given by $[-37.84342, -12.29594]$. Whereas the one based on the common approach is given by
$$[-\hat{\sigma}^{-1}(10.82), -\hat{\sigma}^{-1}(7.26)] = [-29.92, -20.09]$$
where $\hat{\sigma} = 0.3605949$ and the 95% C.I. for β_1^* is $[9.04 - 1.96 \times 0.906, 9.04 + 1.96 \times 0.906] = [7.26, 10.81]$. It is clear that the latter interval is much shorter than the former as it ignores the variability of $\hat{\sigma}$.

- A 95% C.I. for λ based on the delta method is given by $[-416023.7, 707626.3]$. But this includes negative values, which is not appropriate because $\lambda > 0$. Therefore, we report the truncated interval [0, 707626.3]. The one based on the common approach is given by
$$[\exp(8.04), \exp(15.74)] = [3102.61, 6851649.6],$$
where the 95% C.I. for β_0^* is $[-11.89 - 1.96 \times 1.966, -11.89 + 1.96 \times 1.966] = [-15.74, -8.04]$. Although the common approach ends up with an unreasonably wide confidence interval compared to the one based on the delta method, this approach always yields limits within the legal range of λ.

- At $x = 2.480159$, the design temperature of 1300C, the estimated 15th, 50th, and 85th percentiles in log(hours) and hours, respectively based on uquantile and quantile, along with their corresponding 90% C.I.'s in hours are reported in the following table.

Type	percentile	estimate	std. err	90% LCL	90% UCL
uquantile	15	9.868867	0.3444804	10962.07	34048.36
	50	10.392887	0.3026464	19831.64	53677.02
	85	10.756643	0.2973887	28780.08	76561.33
quantile	15	19319.44	6655.168	9937.174	37560.17
	50	32626.72	9874.361	19668.762	54121.65
	85	46940.83	13959.673	28636.931	76944.21

The 90% C.I.'s based on uquantile, $\exp($estimate $\pm 1.645 \times$ std.err$)$, are shorter than those based on quantile at each x value. However, we also suspect there is a minor bug in predict in that there appears to be a discrepancy between the standard error estimate for the 15th percentile resulting from uquantile and ours based on the delta method which follows. The other two standard error estimates are arbitrarily close to ours.

Our standard error estimates are 0.3174246, 0.2982668, and 0.3011561 for the 15th, 50th, and 85th percentiles, respectively. Applying the trivariate delta method, we obtain the following expression:

$$\begin{aligned}\operatorname{var}(\hat{y}_p) &= \operatorname{var}(\hat{\beta}_0^*) + \operatorname{var}(\hat{\beta}_1^*)x_0^2 + z_p^2\hat{\sigma}^2 \operatorname{var}(\log\hat{\sigma}) \\ &+ 2x_0\operatorname{Cov}(\hat{\beta}_0^*, \hat{\beta}_1^*) + 2z_p\hat{\sigma}\operatorname{Cov}(\hat{\beta}_0^*, \log\hat{\sigma}) + 2x_0 z_p\hat{\sigma}\ \operatorname{Cov}(\hat{\beta}_1^*, \log\hat{\sigma})\end{aligned} \tag{47}$$

- At the design temperature 1300C, by 25,000 hours about 28% of the motorettes have failed. That is, after 25,000 hours, about 72% are still working.
- As $\hat{\alpha} = \frac{1}{\hat{\sigma}} = \frac{1}{0.3605949} = 2.773195$, then for fixed x the hazard function increases as time increases. The covariate x is fixed at 2.480159 which corresponds to the design temperature 1300C.
- The estimated coefficient $\hat{\beta}_1 = -\frac{1}{\hat{\sigma}}\hat{\beta}_1^* = -\frac{1}{0.3605949}(9.04) = -25.06968 < 0$. Thus, for time fixed, as x increases, the hazard decreases and survival increases.

- For $x_1 < x_2$,
$$\frac{\lambda(t|x_2)}{\lambda(t|x_1)} = \exp((x_1 - x_2)(-25.06968)).$$
For example, for $x = 2.1$ and $x = 2.2$
$$\frac{\lambda(t|2.2)}{\lambda(t|2.1)} = \exp(0.1(-25.06968)) = 0.08151502.$$
Thus, for 0.1 unit increase in x, the hazard becomes about 8.2% of the hazard before the increase. In terms of Celsius temperature, for 21.645 degree decrease from 202.99050C to 181.34550C, the hazard becomes about 8.2% of the hazard before the decrease.

- The Q-Q plots show that the Weibull fit looks slightly better than the log-logistic fit at the temperature 1700C, but overall they are the same. On the other hand, the Weibull fit looks noticeably better than the log-normal fit at the temperature 1700C and is about the same at the other two temperatures. This result coincides with our finding from **AIC** in Table 6; that is, among these three accelerated failure time models, the Weibull best describes the motorette data.