BIOSTAT M215 HW1

Hanbei Xiong

Reading assignment: (Klein and Moeschberger) Chapters 1; Chapter 2: 2.1-2.5; Chapter 3: 3.1-3.5; Appendix A, B

Exercise 2.1

The lifetime of light bulbs follows an exponential distribution with a hazard rate of 0.001 failures per hour of use. (a) Find the mean lifetime of a randomly selected light bulb. (b) Find the median lifetime of a randomly selected light bulb. (c) What is the probability a light bulb will still function after 2,000 hours of use?

(a) Answer:

E(X) = 1 / 0.001 = 1000 hours.

(b) Answer:

Median = ln(2) / 0.001 = 693.1472 hours.

(c) Answer:

P(X > 2000) = exp(-0.001 * 2000) = 0.1353353.

Exercise 2.4

A model for lifetimes, with a bathtub-shaped hazard rate, is the ex- ponential power distribution with survival function $S(x) = exp\{1 - exp[(\lambda x)^{\alpha}]\}$. (a) If $\alpha = 0.5$, show that the hazard rate has a bathtub shape and find the time at which the hazard rate changes from decreasing to increasing. (b) If $\alpha = 2$, show that the hazard rate of x is monotone increasing.

(a) Answer:

$b(x) = -\frac{dln[S(x)]}{dx} = -\frac{d}{dx}[1 - exp(-(\lambda x)^{\alpha})] = \lambda x^{\alpha}\alpha(\lambda x)^{\alpha-1}exp(-(\lambda x)^{\alpha}) = \lambda x^{2\alpha-1}\,\alpha exp(-(\lambda x)^{\alpha})$

$b'(x) = \lambda x^{2\alpha-1}\,\alpha exp(-(\lambda x)^{\alpha}) + \lambda x^{2\alpha-1}\,\alpha^2(\lambda x)^{\alpha-1}exp(-(\lambda x)^{\alpha}) = \lambda x^{2\alpha-1}\,\alpha exp(-(\lambda x)^{\alpha})(1 + \alpha(\lambda x)^{\alpha-1})$

$b'(x) = 0$ when $1 + \alpha(\lambda x)^{\alpha-1} = 0$

$\lambda x = \left(\frac{-1}{\alpha}\right)^{\frac{1}{\alpha-1}}$

Exercise 2.11

In some applications, a third parameter, called a guarantee time, is included in the models discussed in this chapter. This parameter ϕ is the smallest time at which a failure could occur. The survival function of the three-parameter Weibull distribution is given by:

$$S(x) = \begin{cases} 1 & \text{if } x < \phi, \\ \exp[-\lambda(x-\phi)^{\alpha}] & \text{if } x \geq \phi. \end{cases}$$

a. Find the hazard rate and the density function of the three-parameter Weibull distribution.

b. Suppose that the survival time X follows a three-parameter Weibull distribution with $\alpha = 1$, $\lambda = 0.0075$, and $\phi = 100$. Find the mean and median lifetimes.

(a) Answer:

When $x < \phi$, $b(x) = 0$ and $f(x) = 0$.

When $x \geq \phi$,

$f(x) = -\frac{dS(x)}{dx} = \lambda\alpha(x-\phi)^{\alpha-1}\exp[-\lambda(x-\phi)^{\alpha}]$.

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

(b) Answer:

$\mu = E(X) = \int_{100}^{\infty} xf(x)dx = \int_{100}^{\infty} x\lambda\exp[-\lambda(x-100)]dx = \frac{1}{\lambda} + \phi = 100 + \frac{1}{0.0075} = 233.3333.$

Set S(m) = 0.5, we have $\exp[-0.0075(m-100)] = 0.5$, so
$m = 100 + ln(2)/0.0075 = 100 + 92.4196 = 192.4196.$

Exercise 2.18

Let X have a uniform distribution on 0 to 100 days with probability density function:

$$f(x) = \begin{cases} \frac{1}{100}, & \text{for } 0 < x < 100, \\ 0, & \text{elsewhere.} \end{cases}$$

a. Find the survival function at 25, 50, and 75 days.

b. Find the mean residual lifetime at 25, 50, and 75 days.

c. Find the median residual lifetime at 25, 50, and 75 days.

(a) Answer:

$S(x) = \int_x^{100} f(t)dt = \int_x^{100} \frac{1}{100}dt = \frac{100-x}{100}$.

$S(25) = \frac{75}{100} = 0.75$.

$S(50) = \frac{50}{100} = 0.5$.

$S(75) = \frac{25}{100} = 0.25$.

(b) Answer:

$mrl(x) = \frac{\int_x^{100}(t-x)f(t)dt}{S(x)} = \frac{\int_x^{100}(t-x)\frac{1}{100}dt}{\frac{100-x}{100}} = \frac{\frac{1}{200}(100^2-x^2-100x+x^2)}{\frac{100-x}{100}} = \frac{100-x}{2}$.

$$mrl(25) = \frac{100-25}{2} = 37.5.$$

$$mrl(50) = \frac{100-50}{2} = 25.$$

$$mrl(75) = \frac{100-75}{2} = 12.5.$$

(c) Answer:

$$S(x + M(x)) = 0.5S(x)$$

$$\frac{100-(x+M(x))}{100} = 0.5 * \frac{100-x}{100}$$

$$M(x) = \frac{100-x}{2}$$

$$M(25) = \frac{100-25}{2} = 37.5.$$

$$M(50) = \frac{100-50}{2} = 25.$$

$$M(75) = \frac{100-75}{2} = 12.5.$$

Exercise 3.1

Describe, in detail, the types of censoring which are present in the following studies:

a. The example dealing with remission duration in a clinical trial for acute leukemia described in section 1.2.

b. The example studying the time to death for breast cancer patients described in section 1.5.

(a) Answer: Type I censoring. The study was conducted over a fixed duration, so participants who had not relapsed by the end of the study were censored at that point.

(b) Answer: Type I censoring. The censoring was caused by the fixed study period of 10 years. Patients who were alive at the end of the study were censored at that time.

Exercise 3.2

A large number of disease-free individuals were enrolled in a study beginning January 1, 1970, and were followed for 30 years to assess the age at which they developed breast cancer. Individuals had clinical exams every 3 years after enrollment. For four selected individuals described below, discuss in detail the types of censoring and truncation that are represented.

a. A healthy individual, enrolled in the study at age 30, never developed breast cancer during the study.

b. A healthy individual, enrolled in the study at age 40, was diagnosed with breast cancer at the fifth exam after enrollment (i.e., the disease started sometime between 12 and 15 years after enrollment).

c. A healthy individual, enrolled in the study at age 50, died from a cause unrelated to the disease (i.e., not diagnosed with breast cancer at any time during the study) at age 61.

d. An individual, enrolled in the study at age 42, moved away from the community at age 55 and was never diagnosed with breast cancer during the period of observation.

e. Confine your attention to the four individuals described above, and write down the likelihood for this portion of the study.

(a) Answer: Type I censoring. No truncation applies to this individual. They were disease-free at age 30, enrolled in the study, and remained under observation for 30 years.

(b) Answer: Interval censoring. No truncation applies here either. The individual was disease-free at age 40 and enrolled, meaning they met the inclusion criteria.

(c) Answer: Competing risk censoring. No truncation applies. The individual was alive and disease-free at enrollment (age 50) and observed until their death from unrelated causes.

(d) Answer: Progressive Type I censoring. No truncation applies. The individual was disease-free at enrollment and included in the study until they moved away.

(e) Answer:

L=S(30)[S(12)-S(15)]S(11)S(13)

Exercise 3.4

In section 1.2, a clinical trial for acute leukemia is discussed. In this trial, the event of interest is the time from treatment to leukemia relapse. Using the data for the 6-MP group and assuming that the time to relapse distribution is exponential with hazard rate λ, Construct the likelihood function. Using this likelihood function, find the maximum likelihood estimator of λ by finding the value of λ which maximizes this likelihood.

a. Estimate S(12) = P (X > 12), the probability that a 6-MP patient will be relapse free for more than 12 months, with a 95% confidence interval.

b. Instead of the exponential distribution assumption, assume that the time to relapse for 6-MP patients follows a log-normal distribution. Find the maximum likelihood estimate of the unknown parameters and repeat question 1.

c. Using the exponential AFT model to test if 6-MP is effective in pro- longing remission duration in comparison with the placebo group at $\alpha = 0.05$. Be specific about the model and hypotheses, and find the p-value.

d. Repeat part (c) using the Weibull AFT model. Do you observe any differences between the results from the two models? If so, try to explain why they are different?

Given the data in section 1.2:

Remission Data Table

Pair	Randomization	Placebo Patients	6-MP Patients
1	Partial Remission	1	10
2	Complete Remission	22	7
3	Complete Remission	3	32+
4	Complete Remission	12	23
5	Complete Remission	8	22
6	Partial Remission	17	6
7	Complete Remission	2	16

Pair	Randomization	Placebo Patients	6-MP Patients
8	Complete Remission	11	34+
9	Complete Remission	8	32+
10	Complete Remission	12	25+
11	Complete Remission	2	11+
12	Partial Remission	5	20+
13	Complete Remission	4	19+
14	Complete Remission	15	6
15	Complete Remission	8	17+
16	Partial Remission	23	35+
17	Partial Remission	5	6
18	Complete Remission	11	13
19	Complete Remission	4	9+
20	Complete Remission	1	6+
21	Complete Remission	8	10+

(a) Answer:

$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \prod_{j=1}^{m} e^{-\lambda x_j}$

$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} e^{-\lambda \sum_{j=1}^{m} x_j}$

$L(\lambda) = \lambda^n e^{-\lambda(\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} x_j)}$

Since $\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} x_j$ is the total observed time, we denote it as T.

$L(\lambda) = \lambda^n e^{-\lambda T}$

$ln(L(\lambda)) = nln(\lambda) - \lambda T$

$\frac{dln(L(\lambda))}{d\lambda} = \frac{n}{\lambda} - T$

$\hat{\lambda} = \frac{n}{T}$

Since number of exact lifetime patients is 9. The total observed time is 359.

$\hat{\lambda} = \frac{9}{359} = 0.0251$

$S(12) = e^{-0.0251*12} = 0.740$

$Var(\hat{\lambda}) = \frac{\hat{\lambda^2}}{n} = \frac{0.0251^2}{9} = 0.00007$

$SE(\hat{\lambda}) = \sqrt{Var(\hat{\lambda})} = \sqrt{0.00007} = 0.0084$

LB: $\hat{\lambda} - 1.96 * SE(\hat{\lambda}) = 0.0251 - 1.96 * 0.0084 = 0.0087$ UB: $\hat{\lambda} + 1.96 * SE(\hat{\lambda}) = 0.0251 + 1.96 * 0.0084 = 0.0415$

LB for S(12): $e^{-0.0415*12} = 0.608$ UB for S(12): $e^{-0.0087*12} = 0.901$

95% CI for S(12): (0.608, 0.901)

(b) Answer:

The full log-likelihood is:

$$l(\mu, \sigma^2) = \sum_{i=1}^{n} ln(\frac{1}{x_i \sigma \sqrt{2\pi}} e^{-\frac{(ln(x_i)-\mu)^2}{2\sigma^2}}) + \sum_{j=1}^{m} ln[1 - \Phi(\frac{ln(x_j)-\mu}{\sigma})]$$

```
library(survival)
```

```
## Warning: package 'survival' was built under R version 4.3.3
```

```
remission_data <- data.frame(
  time = c(10, 7, 32, 23, 22, 6, 16, 34, 32, 25, 11, 20, 19, 6, 17, 35, 6, 13, 9, 6, 1
0),
  notcensor = c(1,1,0,1,1,1,1,0,0,0,0,0,0,1,0,0,1,1,0,0,0)
)

lognormal_model <- survreg(Surv(time, notcensor) ~ 1, data = remission_data, dist = "log
normal")
summary(lognormal_model)
```

```
##
## Call:
## survreg(formula = Surv(time, notcensor) ~ 1, data = remission_data,
##     dist = "lognormal")
##               Value Std. Error     z      p
## (Intercept)  3.2031     0.2861 11.19 <2e-16
## Log(scale)  -0.0215     0.2560 -0.08   0.93
##
## Scale= 0.979
##
## Log Normal distribution
## Loglik(model)= -40.7   Loglik(intercept only)= -40.7
## Number of Newton-Raphson Iterations: 4
## n= 21
```

$$\hat{\mu} = intercept = 3.2031$$

$$\hat{\sigma} = scale = exp(-0.0215) = 0.9787295$$

```
mu_hat <- lognormal_model$coefficients[1]
sigma_hat <- lognormal_model$scale
vcov <- lognormal_model$var

var_mu <- vcov[1,1]
var_sigma <- vcov[2,2]
cov_mu_sigma <- vcov[1,2]

# Compute survival probability S(12)
log_12 <- log(12)
z <- (log_12 - mu_hat) / sigma_hat
S_12 <- 1 - pnorm(z)  # Survival probability

print(paste0("S(12) is ", S_12))
```

```
## [1] "S(12) is 0.768456193722297"
```

```
# Partial derivatives
phi <- dnorm(z)  # PDF of the standard normal at z
derivative_mu <- phi / sigma_hat
derivative_sigma <- phi * (log_12 - mu_hat) / sigma_hat^2

# Variance of S(12)
var_S_12 <- derivative_mu^2 * var_mu + derivative_sigma^2 * var_sigma + 2 * derivative_m
u * derivative_sigma * cov_mu_sigma
se_S_12 <- sqrt(var_S_12)  # Standard error

# 95% Confidence Interval for S(12)
lower <- S_12 - 1.96 * se_S_12
upper <- S_12 + 1.96 * se_S_12

print(paste0("95% CI for S(12): (", lower, ",", upper, ")"))
```

```
## [1] "95% CI for S(12): (0.614524028292476,0.922388359152118)"
```

(c) Answer:

$H_0 : \beta_1 = 0$: No difference in remission duration between 6-MP and placebo. $H_A : \beta_1 \neq 0$: Significant difference in remission duration between 6-MP and placebo.

The form of the model is:

$$\log(T) = \beta_0 + \beta_1 * treatment + \epsilon$$

The Exponential AFT model assumes that the time to event follows an exponential distribution.

```
remission_data <- data.frame(
  time = c(1, 22, 3, 12, 8, 17, 2, 11, 8, 12, 2, 5, 4, 15, 8, 23, 5, 11, 4, 1, 8,
           10, 7, 32, 23, 22, 6, 16, 34, 32, 25, 11, 20, 19, 6, 17, 35, 6, 13, 9, 6, 1
0),
  status = c(rep(1, 21), 1,1,0,1,1,1,1,0,0,0,0,0,0,1,0,0,1,1,0,0,0),
  group = c(rep(0, 21), rep(1, 21))
)

exp_aft_model <- survreg(Surv(time, status) ~ group, data = remission_data, dist = "expo
nential")

summary(exp_aft_model)
```

```
##
## Call:
## survreg(formula = Surv(time, status) ~ group, data = remission_data,
##     dist = "exponential")
##             Value Std. Error    z       p
## (Intercept) 2.159      0.218 9.90 < 2e-16
## group       1.527      0.398 3.83 0.00013
##
## Scale fixed at 1
##
## Exponential distribution
## Loglik(model)= -108.5   Loglik(intercept only)= -116.8
##  Chisq= 16.49 on 1 degrees of freedom, p= 4.9e-05
## Number of Newton-Raphson Iterations: 4
## n= 42
```

Since p is 0.00013 smaller than 0.05, we reject the null hypothesis and conclude that there is a significant difference in remission duration between 6-MP and placebo.

(d) Answer:

$H_0 : \beta_1 = 0$: No difference in remission duration between 6-MP and placebo. $H_A : \beta_1 \neq 0$: Significant difference in remission duration between 6-MP and placebo.

The form of the model is:

$$\log(T) = \beta_0 + \beta_1 * treatment + \epsilon$$

The Weibull AFT model generalizes the Exponential AFT model by allowing a non-constant hazard rate, characterized by the shape parameter k.

```
weibull_aft_model <- survreg(Surv(time, status) ~ group, data = remission_data, dist =
"weibull")

summary(weibull_aft_model)
```

```
##
## Call:
## survreg(formula = Surv(time, status) ~ group, data = remission_data,
##     dist = "weibull")
##              Value Std. Error     z       p
## (Intercept)  2.248      0.166 13.55 < 2e-16
## group        1.267      0.311  4.08 4.5e-05
## Log(scale)  -0.312      0.147 -2.12   0.034
##
## Scale= 0.732
##
## Weibull distribution
## Loglik(model)= -106.6   Loglik(intercept only)= -116.4
##  Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06
## Number of Newton-Raphson Iterations: 5
## n= 42
```

Since p is 4.5e-05 less than 0.05, we reject the null hypothesis and conclude that there is a significant difference in remission duration between 6-MP and placebo.

Compare to the exponential AFT model:

The Weibull model detects an increasing hazard rate (k = 1/scale = 1.366), meaning the likelihood of relapse grows over time. The exponential model assumes a constant hazard rate (k = 1), which may oversimplify the data. Also, the Weibull model has a higher log-likelihood and a more significant improvement in fit (p = 4.5e-05).