

# Assignment 1

Zhiwen Tan

1/24/2022

## Question 1

The survival function is given by

$$\begin{aligned} S(t) &= P(T > t) \\ &= P(\log(T) > \log(t)) \\ &= P(3 + 0.8Z + 2W > \log(t)) \\ &= P(W > \frac{1}{2}(\log(t) - 3 - 0.8Z)) \\ &= 1 - P(W \leq \frac{1}{2}(\log(t) - 3 - 0.8Z)) \\ &= 1 - \Phi(\frac{1}{2}(\log(t) - 3 - 0.8Z)) \end{aligned}$$

Since we know  $W$  has a standard extreme value distribution, so we know  $\Phi$  is CDF of the standard extreme value distribution. Therefore, we can calculate the survival probabilities for treatment A and B.

For a standard extreme value distribution, we know  $W \sim EV(\mu = 0, \sigma = 1)$ , thus, CDF  $G(t) = 1 - e^{-e^t}$ .

Now we have survival function

$$\begin{aligned} S(t) &= 1 - (1 - (e^{-e^{\frac{1}{2}(\log(t) - 3 - 0.8Z)}})) \\ &= e^{-e^{\frac{1}{2}(\log(t) - 3 - 0.8Z)}} \end{aligned}$$

To find the survival probabilities of the two treatments at 1 year and 5 years, we need to know the time unit. In this question, we have time to relapse in years, so we don't need to convert it month. Then we can plug the value  $t$  and  $z$  in this function. The results are shown as below.

We can see the survival probability for new treatment is greater than standard treatment, and this is because the survival function is an exponential function with a negative power. The result means the probability of relapse in 1 year for standard treatment is  $1 - 80\% = 20\%$ , the probability of relapse in 1 year for new treatment is  $1 - 86\% = 14\%$ , the probability of relapse in 5 years for standard treatment is  $1 - 61\% = 39\%$ , the probability of relapse in 5 years for new treatment is  $1 - 72\% = 28\%$ . Now we can conclude that the new treatment is better than the standard treatment because the proportion of people relapses is less than the standard treatment.

##	standard treatment	new treatment
## year 1	0.8000107	0.8610793
## year 5	0.6071774	0.7157348

## Question 2

- (a) From lecture notes, we know the survival function for lognormal distribution T is

$$S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$$

Then we can use the survival function to calculate the hazard function. The Hazard is given by

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{-S'(t)}{S(t)} \\ &= \frac{\frac{d}{dt}(1 - \Phi(\frac{\log(t) - \mu}{\sigma}))}{1 - \Phi(\frac{\log(t) - \mu}{\sigma})} \\ &= \frac{-(0 - \frac{1}{t\sigma}\phi(\frac{\log(t) - \mu}{\sigma}))}{1 - \Phi(\frac{\log(t) - \mu}{\sigma})} \\ &= \frac{\frac{1}{t\sigma}\phi(\frac{\log(t) - \mu}{\sigma})}{1 - \Phi(\frac{\log(t) - \mu}{\sigma})} \end{aligned}$$

- (b) The standard normal distribution has a pdf

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)$$

Since we know for a log normal distribution T,  $Y = \log T$  where  $Y$  in  $Normal(\mu, \sigma^2)$ . so we could derive the pdf of T from here

$$\begin{aligned} Y &= \log T \\ T &= e^Y \\ F_T(t) &= P[T \leq t] \\ &= P[e^Y \leq t] \\ &= P[Y \leq \log t] \\ &= F_Y(\log t) \\ f_T(t) &= \frac{d}{dt} F_Y(\log t) \\ &= 1/t f_Y(\log t) \\ &= \frac{1}{t} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right) \end{aligned}$$

Now we have the pdf for T, then we need to get the CDF for T

$$\begin{aligned} S(t) &= 1 - F(t) \\ F(t) &= 1 - S(t) \\ &= 1 - (1 - \Phi(\frac{\log t - \mu}{\sigma})) \\ &= \Phi(\frac{\log t - \mu}{\sigma}) \end{aligned}$$

We also know the survival function for T is

$$S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$$

we know the median = 100 and 95 percentile = 300, so we can calculate  $\mu$  and  $\sigma$ . Since  $\log(t) = \mu_{\lognormal} + \sigma * N(0, 1)$ , and  $\mu_{\lognormal} = e^{\mu_{normal}}$ , so we have  $\mu_{\lognormal} = \log(\mu_{normal}) = \log(100)$ , take this in, we have  $\log(300) = \log(100) + \sigma * 1.65$  which is  $\sigma = \frac{\log(300) - \log(100)}{1.65} = 0.668$

$$f_T(t) = \frac{1}{t} \frac{1}{0.668\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\log t - \log(100)}{0.668}\right)^2\right) \quad (1)$$

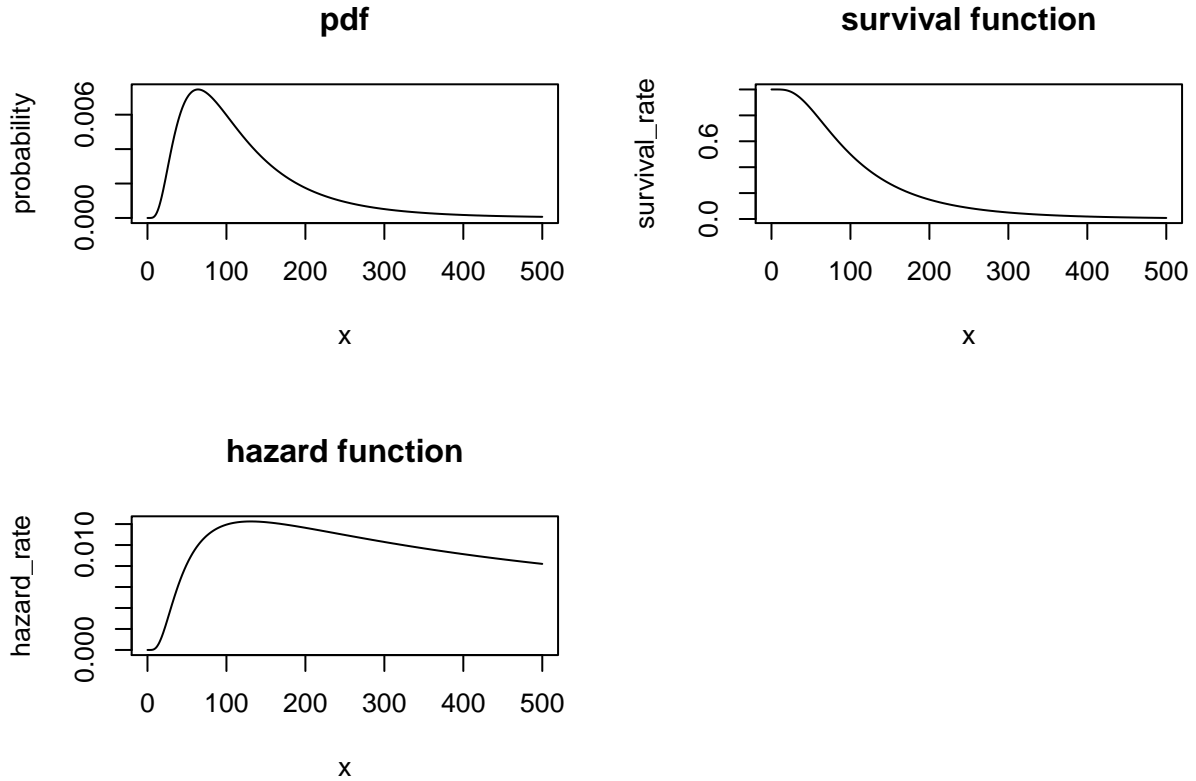
$$F_T(t) = \Phi\left(\frac{\log t - \log(100)}{0.668}\right) \quad (2)$$

$$S(t) = 1 - \Phi\left(\frac{\log t - \log(100)}{0.668}\right) \quad (3)$$

$$h(t) = \frac{\phi\left(\frac{\log(t) - \log(100)}{0.668}\right)}{1 - \Phi\left(\frac{\log(t) - \log(100)}{0.668}\right)} \quad (4)$$

$$T \sim \lognormal(\log 100, 0.668) \quad (5)$$

(c) The pdf, survival function and hazard function are shown as below



### Question 3

- (a) Let  $W \sim \text{logistic}(0, 1)$ , then  $G(w) \sim \frac{e^w}{1+e^w}$ .  
 let  $Y \sim \text{logistic}(\mu, \sigma)$ , then

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(\frac{Y - \mu}{\sigma} \leq \frac{y - \mu}{\sigma}\right) \end{aligned} \tag{1}$$

now let  $W = \frac{Y - \mu}{\sigma}$ , then we have

$$\begin{aligned} F_W(w) &= P(W \leq w) \\ &= P\left(\frac{Y - \mu}{\sigma} \leq w\right) \\ &= P(Y \leq \mu + \sigma w) \\ &= F_Y(\mu + \sigma w) \\ &= \frac{e^w}{1 + e^w} \end{aligned}$$

we can see this is the cdf of  $\text{logistic}(0, 1)$ . therefore, we have

$$\begin{aligned} F_Y(y) &= P(W \leq \frac{y - \mu}{\sigma}) \\ &= G\left(\frac{y - \mu}{\sigma}\right) \end{aligned} \tag{continue from (1)}$$

Therefore,  $Y \sim \text{logistic}(\mu, \sigma)$  is a location-scale distribution. Since  $Y = \log T$ , we can conclude that  $T$  has a log location-scale distribution

- (b) We could first calculate the survival function using  $\mu$  and  $\sigma$

$$\begin{aligned} Y &= \log T \\ T &= e^Y \\ F_T(t) &= P[T \leq t] \\ &= P[e^Y \leq t] \\ &= P[Y \leq \log t] \\ &= F_Y(\log t) \\ &= \frac{e^{\frac{\log t - \mu}{\sigma}}}{1 + e^{\frac{\log t - \mu}{\sigma}}} \end{aligned}$$

Since we know  $S(t) = 1 - F(t)$ , so we have

$$\begin{aligned} S(t) &= 1 - \frac{e^{\frac{\log t - \mu}{\sigma}}}{1 + e^{\frac{\log t - \mu}{\sigma}}} \\ &= \frac{1 + e^{\frac{\log t - \mu}{\sigma}}}{1 + e^{\frac{\log t - \mu}{\sigma}}} - \frac{e^{\frac{\log t - \mu}{\sigma}}}{1 + e^{\frac{\log t - \mu}{\sigma}}} \\ &= \frac{1}{1 + e^{\frac{\log t - \mu}{\sigma}}} \end{aligned}$$

Then let  $\lambda = e^{\frac{-\mu}{\sigma}}$ , and  $\alpha = \frac{1}{\sigma}$ , then we have

$$\begin{aligned} S(t) &= \frac{1}{1 + e^{\frac{\log t}{\sigma}} e^{\frac{-\mu}{\sigma}}} \\ &= \frac{1}{1 + t^{\frac{1}{\sigma}} e^{\frac{-\mu}{\sigma}}} \\ &= \frac{1}{1 + \lambda t^\alpha} \end{aligned}$$

Now we have the desired  $S(t)$ . For  $h(t)$ , we only need to use the formula  $h(t) = \frac{-S'(t)}{S(t)}$ .

$$\begin{aligned} S'(t) &= \frac{d}{dt} \frac{1}{1 + \lambda t^\alpha} \\ &= -\frac{\frac{d}{dt}(1 + \lambda t^\alpha)}{(1 + \lambda t^\alpha)^2} \\ &= -\frac{\lambda \frac{d}{dt} t^\alpha}{(1 + \lambda t^\alpha)^2} \\ &= -\frac{\alpha \lambda t^{\alpha-1}}{(1 + \lambda t^\alpha)^2} \\ h(t) &= \frac{-S'(t)}{S(t)} \\ &= \frac{\frac{\alpha \lambda t^{\alpha-1}}{(1 + \lambda t^\alpha)^2}}{\frac{1}{1 + \lambda t^\alpha}} \\ &= \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^\alpha} \end{aligned}$$

Now we have shown  $T$  can be written as  $h(t) = \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^\alpha}$ , and  $S(t) = \frac{1}{1 + \lambda t^\alpha}$

#### Question 4

Since  $T$  is a continuous random variable, then we will have the pdf and cdf for this  $T$  which denote as  $f_T(t)$  and  $F_T(t)$ , From here, we can get calculate  $E(T)$ .

$$\begin{aligned} E(T) &= \int_0^\infty t f_T(t) dt \\ &= \int_0^\infty -t \frac{d}{dt} S(t) dt && (\text{since } f_T(t) = -\frac{d}{dt} S(t)) \\ &= -t S(t) \Big|_0^\infty - \int_0^\infty S(t) \frac{d}{dt} (-t) dt && (\text{integral by part}) \\ &= 0 + \int_0^\infty S(t) dt && (\text{we know } \lim_{t \rightarrow \infty} t S(t) = 0) \\ &= \int_0^\infty S(t) dt \end{aligned}$$

Therefore, we have proved,  $E(T) = \int_0^\infty S(t) dt$ .

## Question 5

- (a) To find survival function  $S(t)$ , we could use the formula  $S(t) = e^{-H(t)}$ . Now, we need to find  $H(t)$ , since this is a piece-wise function, we need to sum up all intervals to get  $H(t)$ . Then for each interval  $\tau_{i-1} \leq t \leq \tau_i$

$$\begin{aligned}
 H(t) &= \int_0^t h(t)dt \\
 &= \int_0^{\tau_1} \theta_1 dt + \int_{\tau_1}^{\tau_2} \theta_2 dt + \dots + \int_{\tau_{k-2}}^{\tau_{k-1}} \theta_{k-1} dt + \int_{\tau_{k-1}}^t \theta_k dt \\
 &= \sum_{j=1}^{i-1} \int_{\tau_{j-1}}^{\tau_j} \theta_j dt + \int_{\tau_{i-1}}^t \theta_i dt \\
 &= \sum_{j=1}^{i-1} \theta_j (\tau_j - (\tau_{j-1})) + \theta_i (t - (\tau_{i-1}))
 \end{aligned}$$

Then we can calculate  $S(t)$

$$\begin{aligned}
 S(t) &= e^{-H(t)} \\
 &= \exp\left[-\sum_{j=1}^{i-1} \theta_j (\tau_j - (\tau_{j-1})) - \theta_i (t - (\tau_{i-1}))\right]
 \end{aligned}$$

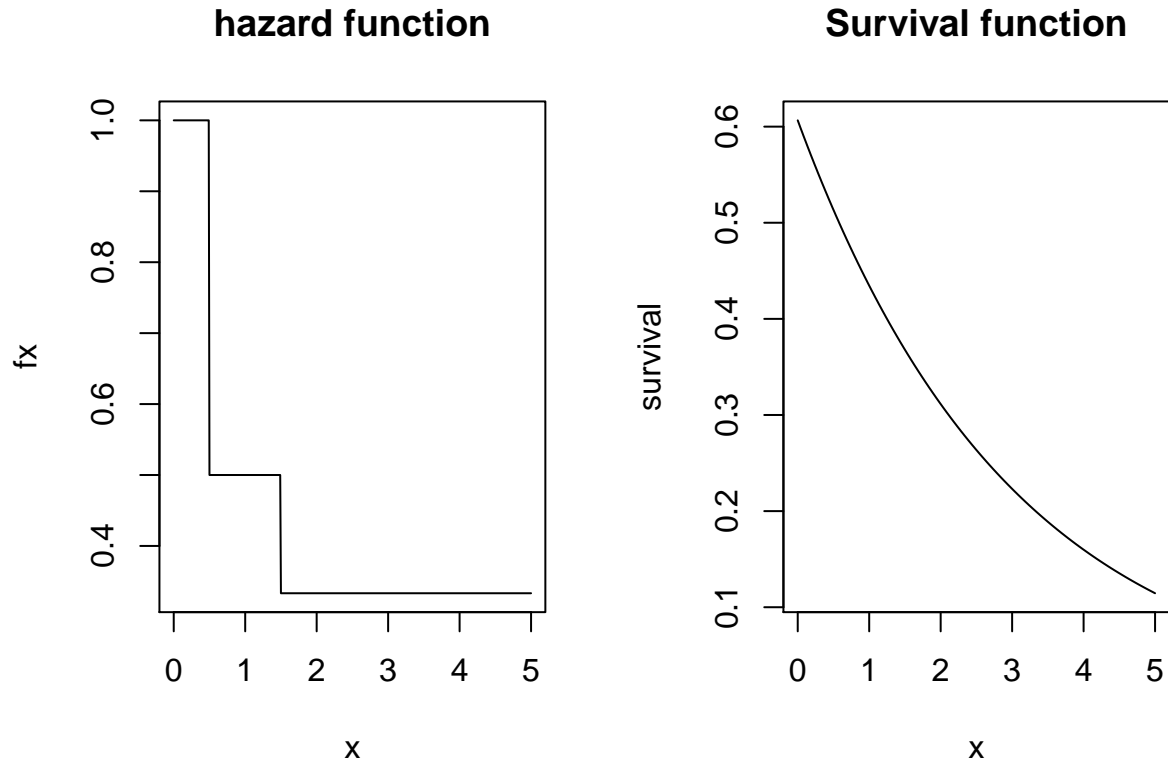
Now we have the desired survival function.

- (b) We need to find the actual survival curve first

$$\begin{aligned}
 H(t) &= \int_0^{0.5} 1 dt + \int_{0.5}^{1.5} \frac{1}{2} dt + \int_{1.5}^t \frac{1}{3} dt \\
 &= t \Big|_0^{0.5} + \frac{1}{2} t \Big|_{0.5}^{1.5} + \frac{1}{3} t \Big|_{1.5}^t \\
 &= 0.5 + 0.5 + \frac{t}{3} - 0.5 \\
 &= \frac{1}{2} + \frac{t}{3} \\
 S(t) &= e^{-(\frac{1}{2} + \frac{t}{3})}
 \end{aligned}$$

for hazard function, we could just plug numbers in to the given  $h(t)$

Now we have two plots below shown hazard function and survival function



from the above plots, we can see even though the hazard function is peicewise,the survival function is still continuous.

- (c) If  $T$  is a piece wise exponential distribution, then we know the exact hazard rate ( $h(t)$ ) on each interval. From part (a), we know the cumulative hazard ratio is

$$H(t) = \sum_{j=1}^{i-1} \theta_j(\tau_j - (\tau_{j-1})) + \theta_i(t - (\tau_{i-1}))$$

because the first part

$$\sum_{j=1}^{i-1} \theta_j(\tau_j - (\tau_{j-1}))$$

is a function that not related to  $x$ , so we could see this part as a constant  $c_1$ . Then we have

$$H(t) = c_1 + \theta_i(t - (\tau_{i-1}))$$

From here, we can see the second term can be rewritten to

$$\theta_i t - \theta_i(\tau_{i-1})$$

since the  $\theta_i(\tau_{i-1})$  is not a function of  $t$ , so we could let this part equals to  $c_2$ .  
Then we have

$$H(t) = c_1 + \theta_i t - c_2$$

combine  $c_1$  and  $c_2$  to get a constant  $c$ , then we have

$$H(t) = c + \theta_i t$$

Now, the survival function is just  $S(t) = e^{-H(t)}$ , so in our case, the survival function can be written as

$$\begin{aligned} S(t) &= e^{-(c+\theta_i t)} \\ &= e^{-c} e^{-\theta_i t} \\ &= a e^s \end{aligned} \quad (\text{let } a = e^{-c}, \text{ and } s = -\theta_i t)$$

From here we can see  $S(t)$  is just an exponential function. Since an exponential function is continuous, so know  $S(t)$  is continuous, this means  $T$  is a continuous random variable.

## Appendix

### Question 1

survival function calculation

```
survival <- function(t, z){
  temp <- -exp(0.5*(log(t)-3-0.8*z))
  s <- 1 - (1 - exp(temp))
  return(s)
}
x <- matrix(NA,2,2)
colnames(x) <- c("standard treatment", "new treatment")
rownames(x) <- c("year 1", "year 5")
x[1,1] <- survival(12, 0)
x[1,2] <- survival(12, 1)
x[2,1] <- survival(60, 0)
x[2,2] <- survival(60, 1)
print(x)
```

### Question 2

plots for question 2

```
# plot pdf
# plot pdf
par(mfrow=c(2,2))
x<-seq(0,500, by=1)
probability <- dlnorm(x, log(100), 0.668)
plot(x, probability, type = "l", main = "pdf")
# plot survival function
```



```

survival_rate <- 1 - pnorm((log(x)-log(100))/0.668)
plot(x, survival_rate, type = "l", main = "survival function")
# harzrd function
hazard_rate <- probability/survival_rate
plot(x,hazard_rate, type = "l", main = "hazard function")

```

## Question 5

harzard function and survival function plot

```

x <- seq(0, 5, by=0.01)
fx <- ifelse(x >= 0 & x < 0.5, 1,
  ifelse(x >= 0.5 & x < 1.5, 0.5,
    ifelse(x >= 1.5, 1/3, 0)))
par(mfrow=c(1,2))
# hazard function
plot(x, fx, main = "hazard function", type = "l")
# survival function
survival <- exp(-(1 + (x^2)/3 - x/2))
plot(x, survival, main = "Survival function", type = "l")

```