

Homework1

Yuki Joyama

1. Exponential Density and Survival-related Functions

- (a) Let $\hat{\lambda}$ be the maximum likelihood estimator of the parameter λ .

For relapse time:

$$\hat{\lambda} = \frac{6}{5+8+12+24+32+17+16+17+19+30} \approx 0.033$$

This indicates that the rate of relapse is about 3.33% per month.

For relapse time:

$$\hat{\lambda} = \frac{3}{10+12+15+33+45+28+16+17+19+30} \approx 0.013$$

This indicates that the rate of death is about 1.33% per month.

- (b)

- i. Mean is $\int_0^\infty t\lambda e^{-\lambda t} dt = \frac{1}{\lambda}$ and I will use $\hat{\lambda}$ to derive the following values.

Mean time to relapse:

$$\frac{1}{0.033} \approx 30.303 \text{ months}$$

$$\frac{1}{0.013} \approx 76.923 \text{ months}$$

- ii. Median is $0.5 = e^{-\lambda\tau} \Rightarrow \tau = \frac{-\log(0.5)}{\lambda}$. By $\hat{\lambda}$,

Median time to relapse:

$$\frac{-\log(0.5)}{0.033} \approx 21.004 \text{ months}$$

$$\frac{-\log(0.5)}{0.013} \approx 53.319 \text{ months}$$

- iii. The survival function of exponential distribution is $S(t) = e^{-\lambda t}$. For relapse:

$$S_R(12) = e^{-0.033 \times 12} = 0.67$$

$$S_R(24) = e^{-0.033 \times 24} = 0.449$$

For death:

$$S_D(12) = e^{-0.013 \times 12} = 0.852$$

$$S_D(24) = e^{-0.013 \times 24} = 0.726$$

- iv. The cumulative probabilities can be calculated as: $F(t) = \int_0^t \lambda(u) du = \lambda t$

For relapse:

$$F_R(12) = 0.033 \times 12 = 0.4$$

$$F_R(24) = 0.033 \times 24 = 0.8$$

For death:

$$F_D(12) = 0.013 \times 12 = 0.16$$

$$F_D(24) = 0.013 \times 24 = 0.32$$

- v. The conditional probability can be expressed as $P(T > 24 | T > 12) = \frac{S_R(24)}{S_R(12)} = 0.67$

It is the same as what we observed in (iii) $S_R(12)$. This means that the conditional probability of being relapse-free after 2 years given that one has remained relapse-free for at least 1 year simplifies to the survival function for the remaining time period (memoryless property of the exponential distribution).

- (c) We can use Kaplan-Meier estimator to estimate median time to relapse. I will calculate them using R.

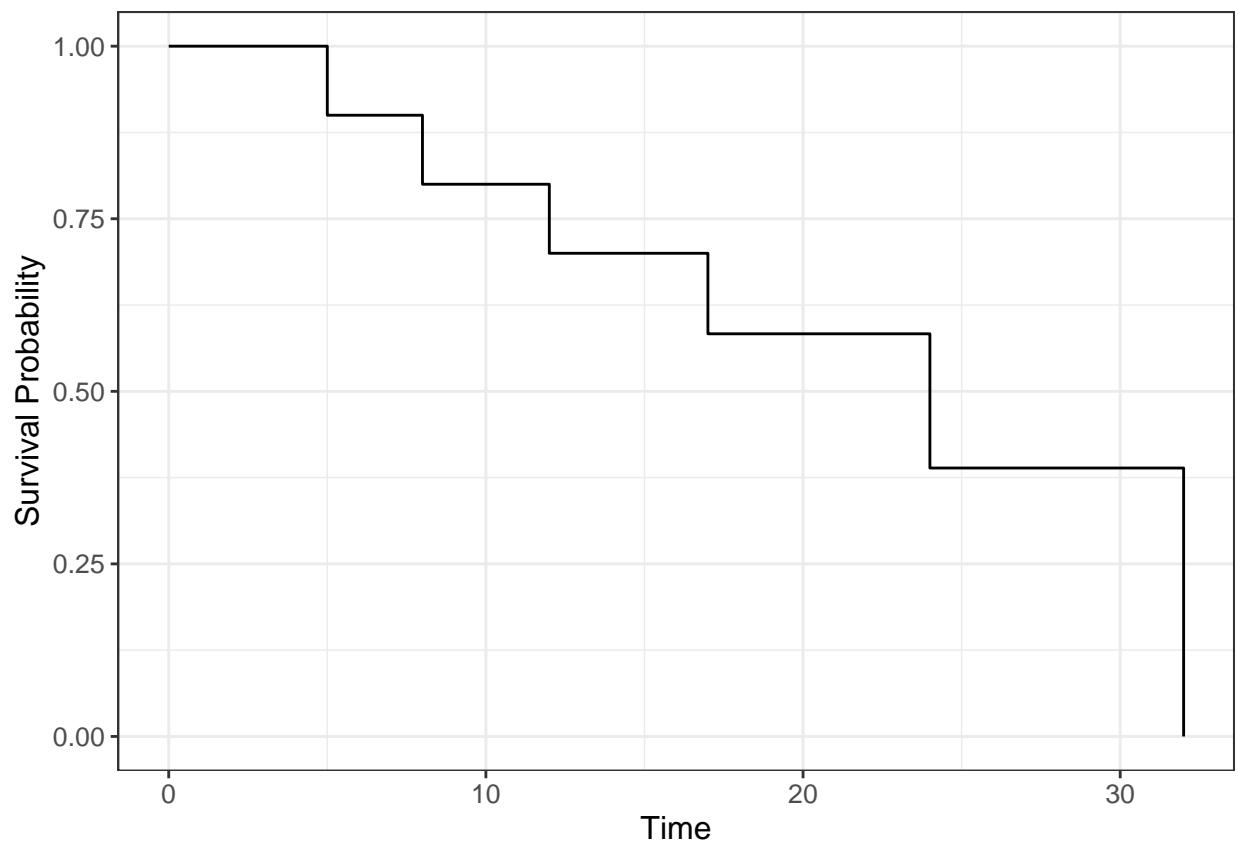
```

library(survival)
library(ggsurvfit)

# set up data
df1 <- data.frame(
  relapse_time = c(5, 8, 12, 24, 32, 17, 16, 17, 19, 30),
  relapse_censored = c(1, 1, 1, 1, 1, 1, 0, 0, 0, 0) # 1: event, 0: censored
)

# fit KM curve
relapse_surv <- Surv(df1$relapse_time, df1$relapse_censored)
relapse_km <- survfit(relapse_surv ~ 1)
relapse_km |>
  ggsurvfit()

```



This tells us that the median time τ s.t. $\hat{S}_R(\tau) \leq 0.50$ is 24 months. Meanwhile, we cannot estimate median time to death because the survival probability does not go below 0.50.

2. Kaplan-Meier Survival Estimate