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:
$\frac{1}{0.033} \approx 30.303$ months Mean survival time:
$\frac{1}{0.013} \approx 76.923$ months

ii. Median is $0.5 = e^{-\lambda\tau} \Rightarrow \tau = \frac{-log(0.5)}{\lambda}$. By $\hat{\lambda}$,
Median time to relapse:
$\frac{-log(0.5)}{0.033} \approx 21.004$ months Median survival time:
$\frac{-log(0.5)}{0.013} \approx 53.319$ months

iii. The survival function of exponential distribution is $S(t) = e^{-\lambda t}$. For relapse:
$S_R(12) = e^{-0.033*12} = 0.67$
$S_R(24) = e^{-0.033*24} = 0.449$
For death:
$S_D(12) = e^{-0.013*12} = 0.852$
$S_D(24) = e^{-0.013*24} = 0.726$

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:
$F_D(12) = 0.013 \times 12 = 0.16$
$F_D(24) = 0.013 \times 24 = 0.32$

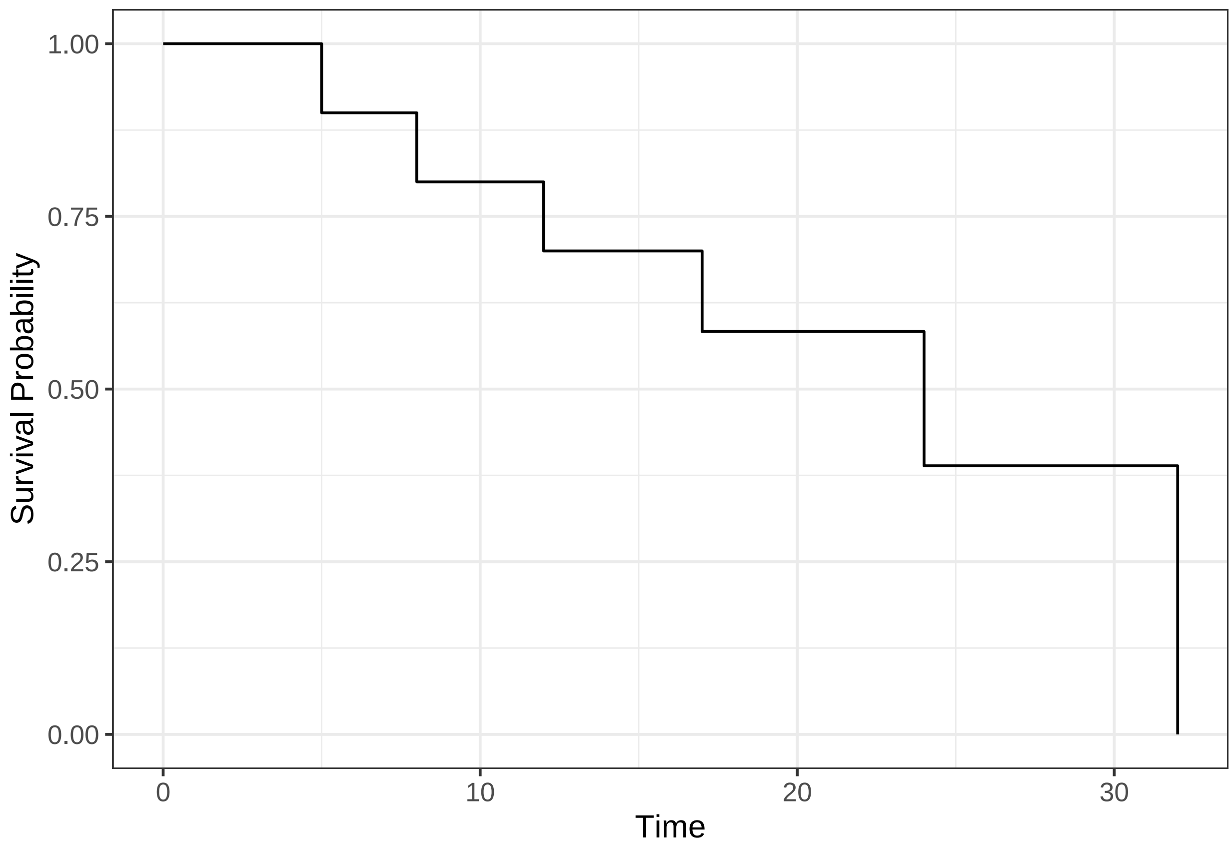
v. The conditional probability can be expressed as $P(T > 24|T > 12) = \frac{S_R(24)}{S_R(12)} = 0.67$
It is the same as what we observed in (iii) $S_R(12)$. This means that the conditional probability of being relapse-free after 2 years given that one has remained relapse-free for at least 1 year simplifies 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.

```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

(a) I will make a table with a row for every death or censoring time.
t_j: distinct death or censoring times
d_j: the number of death at t_j

r_j: the number of individuals at risk right before the j-th death time
c_j: the number of censored observations between the j-th and $(j+1)$-st death time

t_j	d_j	c_j	r_j	$1-(d_j/r_j)$	$\hat{S}(t_j)$
2	1	0	17	0.941	0.941
3	1	0	16	0.938	0.882
4	1	0	15	0.933	0.824
12	1	0	14	0.929	0.765
22	1	0	13	0.923	0.706
48	1	0	12	0.917	0.647
51	0	1	11	1	0.647
56	0	1	10	1	0.647
80	1	0	9	0.889	0.575
85	1	0	8	0.875	0.503
90	1	0	7	0.857	0.431
94	0	1	6	1	0.431
160	1	0	5	0.8	0.345
171	1	0	4	0.75	0.259
180	1	1	3	0.667	0.173
238	1	0	1	0	0