yc4384_Yangyang_Chen_HW1

Yangyang Chen

2024-09-20

Question 1

a) Using the data above, calculate the maximum likelihood estimator of the parameter λ for time to relapse and time to death assuming an exponential distribution:

$$f(t) = \lambda e^{-\lambda t}$$

Write a brief sentence interpreting this parameter.

Solution:

For an exponential distribution, the probability density function is:

$$f(t; \lambda) = \lambda e^{-\lambda t}$$

The likelihood function and log-likelihood function for $T_1, T_2, \ldots, T_n$ are:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda T_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} T_i}$$

$$\log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} T_i$$

To maximize the likelihood, differentiate $\log L(\lambda)$ with respect to λ and set it equal to zero:

$$\frac{d}{d\lambda} \log L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} T_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} T_i}$$

Given relapse data:

- $n = 10$
- Relapse times: $5, 8, 12, 24, 32, 17, 16, 17, 19, 30$
- $\sum T_i = 180$

Substitute into the MLE formula, we obtain:

$$\hat{\lambda}_{\text{relapse}} = \frac{10}{180} = 0.0556$$

Similarly, given:

- Death times: $10, 12, 15, 33, 45, 28, 16, 17, 19, 30$
- $\sum T_i = 225$

we have:

$$\hat{\lambda}_{\text{death}} = \frac{10}{225} = 0.0444$$

Interpretation:

- The MLE for λ for the time to relapse is $\hat{\lambda}_{\text{relapse}} = 0.0556$, indicating that approximately 5.56% of patients relapse each month, and the expected time to relapse is $\frac{1}{0.0556} \approx 18$ months.
- The MLE for λ for the time to death is $\hat{\lambda}_{\text{death}} = 0.0444$, meaning that approximately 4.44% of patients die each month, and the expected time to death is $\frac{1}{0.0444} \approx 22.5$ months.

b) Now you will see how powerful this single parameter can be! Using this parameter estimate (round to 3 decimal places), estimate the following quantities:

(i) The mean time to relapse and mean survival time after bone marrow transplant.

Solution:

The mean of an exponential distribution is given by:

$$E(T) = \frac{1}{\lambda}$$

Using the MLE for the relapse rate $\hat{\lambda}_{\text{relapse}} = 0.0556$, the mean time to relapse is:

$$E(T_{\text{relapse}}) = \frac{1}{0.0556} \approx 18.007 \text{ months}$$

Using the MLE for the death rate $\hat{\lambda}_{\text{death}} = 0.0444$, the mean survival time is:

$$E(T_{\text{death}}) = \frac{1}{0.0444} \approx 22.523 \text{ months}$$

(ii) The median time to relapse and median survival time after bone marrow transplant.

Solution:

Using the CDF of the exponential distribution:

$$1 - e^{-\lambda t_{\text{median}}} = 0.5$$

Therefore, the median of the exponential distribution is:

$$t_{\text{median}} = \frac{\log(2)}{\lambda}$$

Since the MLE for the relapse rate $\hat{\lambda}_{\text{relapse}} = 0.0556$, the median time to relapse is:

$$\text{Median}(T_{\text{relapse}}) = \frac{\log(2)}{0.0556} \approx 12.47 \text{ months}$$

Similarly, the median survival time is:

$$\text{Median}(T_{\text{death}}) = \frac{\log(2)}{0.0444} \approx 15.62 \text{ months}$$

(iii) The one-year and two-year probabilities of remaining relapse-free and surviving: $S_R(12)$ and $S_R(24)$ for relapse, and $S_D(12)$ and $S_D(24)$ for death.

Solution:

Given:

$$S(t) = P(T > t) = e^{-\lambda t}$$

Using the MLE for the relapse rate $\hat{\lambda}_{\text{relapse}} = 0.0556$, we calculate the probabilities of remaining relapse-free:

$$S_R(12) = e^{-\hat{\lambda}_{\text{relapse}} \cdot 12} = e^{-0.0556 \cdot 12}$$

$$S_R(24) = e^{-\hat{\lambda}_{\text{relapse}} \cdot 24} = e^{-0.0556 \cdot 24}$$

Similarly, using the MLE for the death rate $\hat{\lambda}_{\text{death}} = 0.0444$, we obtain the survival probabilities:

$$S_D(12) = e^{-\hat{\lambda}_{\text{death}} \cdot 12} = e^{-0.0444 \cdot 12}$$

$$S_D(24) = e^{-\hat{\lambda}_{\text{death}} \cdot 24} = e^{-0.0444 \cdot 24}$$

(iv) The cumulative probabilities of relapse and death by one and two years (based on the CDF, $F(t)$).

Solution:

Since the CDF of an exponential distribution is:

$$F(t) = 1 - e^{-\lambda t}$$

Using the MLE for the relapse rate $\hat{\lambda}_{\text{relapse}} = 0.0556$, the cumulative probability of relapse is:

$$F_R(12) = 1 - e^{-\hat{\lambda}_{\text{relapse}} \cdot 12} = 1 - e^{-0.0556 \cdot 12}$$

$$F_R(24) = 1 - e^{-\hat{\lambda}_{\text{relapse}} \cdot 24} = 1 - e^{-0.0556 \cdot 24}$$

Similarly, using the MLE for the death rate $\hat{\lambda}_{\text{death}} = 0.0444$, the cumulative probability of death is:

$$F_D(12) = 1 - e^{-\hat{\lambda}_{\text{death}} \cdot 12} = 1 - e^{-0.0444 \cdot 12}$$

$$F_D(24) = 1 - e^{-\hat{\lambda}_{\text{death}} \cdot 24} = 1 - e^{-0.0444 \cdot 24}$$

(iv) Based on the exponential distribution with $\hat{\lambda}$ as calculated in (a), calculate the conditional probability of being relapse-free after 2 years given that one has remained relapse-free for at least one year. How does this compare with the probability of remaining relapse-free one year after bone marrow transplant calculated in part (iii)?

Solution:

Simplify the conditional probability:

$$P(T > 24 \mid T > 12) = \frac{P(T > 24)}{P(T > 12)} = \frac{S(24)}{S(12)} = \frac{e^{-\lambda \cdot 24}}{e^{-\lambda \cdot 12}} = e^{-\lambda \cdot (24-12)} = e^{-\lambda \cdot 12}$$

Since the MLE for the relapse rate $\hat{\lambda}_{\text{relapse}} = 0.0556$, we have:

$$P(T > 24 \mid T > 12) = e^{-0.0556 \cdot 12}$$

From part (iii), the probability of remaining relapse-free one year after bone marrow transplant is:

$$S(12) = e^{-0.0556 \cdot 12} = P(T > 24 \mid T > 12)$$

Therefore, the conditional probability of remaining relapse-free after 2 years, given relapse-free for 1 year, is exactly the same as the probability of remaining relapse-free after 1 year. This is due to the memoryless property of the exponential distribution.

(c) If we decide that an exponential distribution is not appropriate and want to estimate the survival distribution non-parametrically, is it possible to estimate the median time to relapse? Is it possible to estimate the median time to death? If so, provide the appropriate estimates.

Solution:

Yes, it is possible. The Kaplan-Meier estimator is widely used to estimate survival functions when censoring is present and does not rely on any specific parametric distribution.

To compute the Kaplan-Meier estimates and the median times, we can use the `survival` package in R. Here's the code:

```
library(survival)

relapse_times <- c(5, 8, 12, 24, 32, 17, 16, 17, 19, 30)
relapse_status <- c(1, 1, 1, 1, 1, 1, 0, 0, 0, 0)  # 1 = event (relapse), 0 = censored

death_times <- c(10, 12, 15, 33, 45, 28, 16, 17, 19, 30)
death_status <- c(1, 1, 1, 1, 0, 0, 0, 0, 0, 0)  # 1 = event (death), 0 = censored

# Kaplan-Meier estimate for relapse
```

```r
km_relapse <- survfit(Surv(relapse_times, relapse_status) ~ 1)
median_time_relapse <- summary(km_relapse)$table["median"]

# Kaplan-Meier estimate for death
km_death <- survfit(Surv(death_times, death_status) ~ 1)
median_time_death <- summary(km_death)$table["median"]

median_time_relapse
```

```
## median
##     24
```

```r
median_time_death
```

```
## median
##     33
```