BST 222 - Homework 1

Xiaowei Zeng

October 10, 2023

Problem 1

The log-logistic distribution has survival function

$$S(t) = \frac{1}{1 + \lambda t^\alpha},$$

where $\lambda, \alpha > 0$ and $t \geq 0$. Here, the distribution of the logarithm of the failure times is logistic with mean μ and scale parameter σ, and $\alpha = 1/\sigma$, $\lambda = \exp(-\mu/\sigma)$.

a. Find the hazard rate for this distribution.

b. For what values of α is the hazard function monotone. When it is monotone, is it increasing or decreasing?

c. When the hazard function is not monotone, where does it change slope? Is it decreasing to increasing or increasing to decreasing?

d. Describe a practical situation where the hazard function might change slopes in this manner.

Solution:

a. Hazard function:

$$\begin{aligned} h(t) = \frac{f(t)}{S(t)} &= -\frac{d}{dt} \log S(t) \\ &= \frac{d}{dt} \log\left(1 + \lambda t^\alpha\right) \\ &= \frac{d \log\left(1 + \lambda t^\alpha\right)}{d\left(1 + \lambda t^\alpha\right)} \cdot \frac{d\left(1 + \lambda t^\alpha\right)}{dt} \\ &= \frac{\alpha \lambda t^{\alpha - 1}}{1 + \lambda t^\alpha} \end{aligned}$$

b&c. Compute the first derivative of $h(t)$:

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

where $\alpha\lambda t^{\alpha-2} \geq 0$ and $(1+\lambda t^{\alpha-1})^2 \geq 1$ are always true.

Case 1: $0 < \alpha \leq 1$, then $\alpha - 1 - \lambda t^{\alpha} \leq 0$ is always true. So $h'(t) \leq 0$ $(\forall t \geq 0)$, the hazard function is monotone decreasing.

Case 2: $\alpha > 1$, then $\begin{cases} h'(t) > 0 & \text{for } t < \left(\frac{\alpha-1}{\lambda}\right)^{1/a} \\ h'(t) = 0 & \text{for } t = \left(\frac{\alpha-1}{\lambda}\right)^{1/a} \\ h'(t) < 0 & \text{for } t > \left(\frac{\alpha-1}{\lambda}\right)^{1/a} \end{cases}$. So the hazard function is monotone increasing in $[0, \left(\frac{\alpha-1}{\lambda}\right)^{1/a}]$ and monotone decreasing in $[\left(\frac{\alpha-1}{\lambda}\right)^{1/a}, +\infty]$.

In conclusion, when $0 < \alpha \leq 1$, the hazard function is monotone decreasing; when $\alpha > 1$, it is not monotone and changes slope at $\left(\frac{\alpha-1}{\lambda}\right)^{1/a}$, from increasing to decreasing.

d. The changing of slope means that the hazard rate initially increases and then decreases over time. The assessment of postoperative risk is a common scenario with this characteristic. Postoperative risk represents the risk encountered by patients who have just taken a medical surgery or treatment.

Increasing period: Patients' hazard rate may experience a significant increase after treatment because the treatment itself can introduce certain inherent risks, such as surgical complications or treatment side effects.

Decreasing period: Later, patients' survival conditions may gradually improve due to their steady progress in rehabilitation, or the reduction of postoperative risk factors.

Exercise 2.18

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

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

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.

Solution:

a. Survival function:

$$\begin{aligned} S(x) &= \int_x^{+\infty} f(t)dt \\ &= \int_x^{100} \frac{1}{100} dt \\ &= 1 - \frac{x}{100} \end{aligned}$$

So we have $S(25) = 0.75, S(50) = 0.5, S(75) = 0.25$.

b. Mean residual lifetime: the mean residual life function evaluated at time t is the expected remaining life given survival up to time t.

$$\begin{aligned} \text{mrl}(x) &= \frac{1}{S(x)} \int_x^{+\infty} (u - x) f(u) du \\ &= \frac{100}{100 - x} \int_x^{100} (u - x) \frac{1}{100} du \\ &= \frac{1}{100 - x} \int_x^{100} (u - x) du \\ &= \frac{1}{100 - x} \left(\frac{1}{2} u^2 - xu \right) \Big|_x^{100} \\ &= \frac{1}{100 - x} \left(\frac{1}{2} 100^2 - 100x - \frac{1}{2} x^2 + x^2 \right) \\ &= \frac{1}{2(100 - x)} (100 - x)^2 \\ &= \frac{1}{2} (100 - x) \\ &= 50 - \frac{1}{2} x \end{aligned}$$

So we have $\text{mrl}(25) = 37.5, \text{mrl}(50) = 25, \text{mrl}(75) = 12.5$.

c. Median residual lifetime: y=mdrl(x)

$$\because \int_{x}^{x+\mathrm{mdrl}(x)} f(u)du = \int_{x+\mathrm{mdrl}(x)}^{+\infty} f(u)du = \frac{1}{2}\int_{x}^{+\infty} f(u)du$$

$$\therefore S(x+\mathrm{mdrl}(x)) = \frac{1}{2}S(x)$$

$$\therefore 1 - \frac{x+\mathrm{mdrl}(x)}{100} = \frac{1}{2}\left(1-\frac{x}{100}\right)$$

$$\therefore \mathrm{mdrl}(x) = 50 - \frac{1}{2}x$$

So we have $\mathrm{mdrl}(25) = 37.5, \mathrm{mdrl}(50) = 25, \mathrm{mdrl}(75) = 12.5$.

Insights:

If $X \sim$ Uniform, the median residual lifetime is just the same as the mean residual lifetime! This is because the uniform distribution is symmetric and not skewed.

Problem 3

Use the data from Exercise 3.6 in Klein and Moeschberger.

Patient	Relapse Time (months)	Death Time (months)
1	5	11
2	8	12
3	12	15
4	24	33^+
5	32	45
6	17	28^+
7	16^+	16^+
8	17^+	17^+
9	19^+	19^+
10	30^+	30^+

$^+$ Censored observation

The data consists of the times to relapse and the times to death following relapse of 10 patients. In the sample patients 4 and 6 were alive in relapse at the end of the study and patients 7–10 were alive, free of relapse at the end of the study.

a. Suppose that time from the start of the study to relapse is assumed to be exponential with parameter λ_1 and time from start of the study to death is assumed to be exponential with parameter λ_2. Find the likelihoods for λ_1 and λ_2 and find the MLE for each of these parameters.

b. Suppose that we could not observe the death times of patients who had not relapsed. Find the estimate of λ_2 from the truncated sample and discuss the difference.

Solution:

a. Let $t_i, i = 1, 2, ..., m$ denote the observed survival time, and $u_j, j = 1, 2, ..., k$ denote the right censored values.

Likelihood for λ:

$$\begin{aligned}
L(\lambda) &= \left[\prod_i \lambda \exp\left(-\lambda t_i\right)\right]\left[\prod_j \exp\left(-\lambda u_j\right)\right] \\
&= \lambda^m \exp\left\{-\lambda\left(\sum_i t_i + \sum_j u_j\right)\right\} \\
&(\text{Let } T = \sum_i t_i, U = \sum_j u_j) \\
&= \lambda^m \exp\left\{-\lambda\left(T+U\right)\right\}
\end{aligned}$$

Log Likelihood for λ:

$$l(\lambda) = m\log\lambda - \lambda(T+U)$$

First derivative of log likelihood and the MLE of λ:

$$\begin{aligned}
&l'(\lambda) = \frac{m}{\lambda} - (T+U) \\
&\text{Let } l'(\lambda) = 0, \\
&\hat{\lambda}_{MLE} = \frac{m}{T+U}
\end{aligned}$$

Note that $l''(\lambda) = -m/\lambda^2 < 0,\ \forall m, \lambda > 0$. So $m/(T+U)$ is the MLE for λ.

Since we have $m_1 = 6, k_1 = 4, T_1 = 98, U_1 = 82$ for relapse and $m_2 = 4, k_2 = 6, T_2 = 83, U_2 = 143$ for death, substitute these values in the equations above.

Likelihood and MLE for λ_1:

$$\begin{aligned}
&L(\lambda_1) = {\lambda_1}^6 \exp\left(-180\lambda_1\right) \\
&\hat{\lambda_1}_{MLE} = \frac{6}{180} = \frac{1}{30} \approx 0.0333
\end{aligned}$$

Likelihood and MLE for λ_2:

$$\begin{aligned}
&L(\lambda_2) = {\lambda_2}^4 \exp\left(-226\lambda_2\right) \\
&\hat{\lambda_2}_{MLE} = \frac{4}{226} = \frac{2}{113} \approx 0.0177
\end{aligned}$$

b. In this case, the patients who had not relapsed are truncated.

So we have $m_3 = 4, k_3 = 4, T_3 = 83, U_3 = 61$ for death.

New likelihood and MLE for λ_2:

$$L(\lambda_2) = {\lambda_2}^4 \exp\left(-144\lambda_2\right)$$
$$\hat{\lambda_2}'_{MLE} = \frac{4}{144} = \frac{1}{36} \approx 0.0278$$

Compared with question (a), the MLE of λ_2 for truncated data is larger (overestimated), indicating that hazard rate are larger. This difference can be attributed not only to the reduction of sample size, but also to the sample selection bias.

Actually, the likelihood for the truncated data is to some extent equivalent to conditional likelihood. In this context, it is conditional on the patients who had relapsed, and thus the likelihood may be biased.

Problem 4

Suppose that the failure distribution of a particular type of disk drive is exponential with a mean time between failures (MTBF) of 4 years (so that $\lambda = 0.25$). In this model we suppose that after each failure, the time to the next failure is still exponential with the same parameter λ. This means that the failures form a Poisson process with the mean number of failures per year of $\lambda = 0.25$.

If we have n disks on test, then the failure process is still Poisson with MTBF $1/(n\lambda)$ and mean number of failures per year of $n\lambda$.

Suppose that we put n disks on test and need to have a 95% chance of at least 30 failures in the first year to have sufficient accuracy in estimating λ. How many disks need to be put on test?

Hint: We can attack this either using the Poisson distribution (`qpois`, `ppois` in R) or by using the gamma distribution (`qgamma`, `pgamma`).

- For the Poisson calculation, you need to compute the mean number of failures in the first year as a function of n, and then you can use `qpoiss` with `lower=F` and trial and error in choosing n until you find the minimum value of n that results in 30 as the point with upper tail equal to 0.95.
- For the gamma calculation, the waiting time to the kth failure with exponential waiting times with parameter λ_0 has rate λ_0 and shape k. You can use `pgamma` to find the least value of n that results in a time to 30 failures of one year that has a probability of 0.95 or greater. Do this calculation both ways until you get the same answer from both methods.

Solution 1: Poisson calculation.

Suppose $N \sim \text{Poi}(n\lambda)$.

$$\because \text{ want to find } n \ s.t. \ P(N \geq 30) \geq 95\%,$$

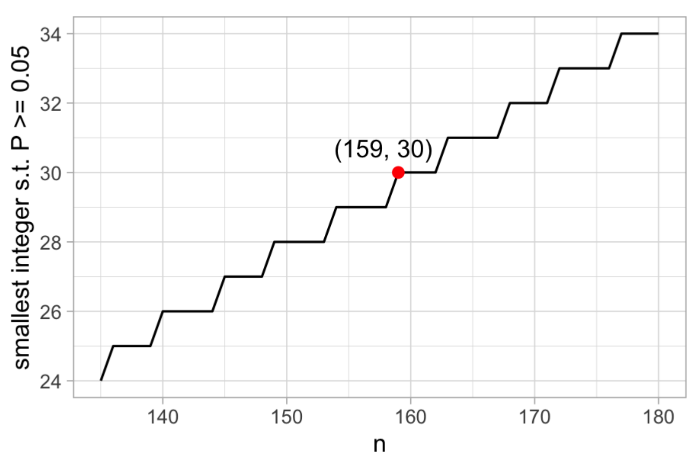
$$\therefore P(N \leq 29) < 5\%.$$

So we can use `qpois(p,lambda)` to find n such that 30 is the smallest integer x such that $P(X \leq x) \geq 0.05$.

```
# Define the function of n to find the proper n.
find_n = function(n){qpois(p = 0.05, lambda = 0.25 * n)}

# Set the range of n and compute the corresponding x.
n_list = seq(135, 180)
x_list = find_n(n_list)

# Plot the graph and annotate the smallest appropriate number.
ggplot() +
  geom_line(aes(y = x_list, x = n_list)) +
  geom_point(aes(y = 30, x = n_list[x_list == 30][1]), color = 'red') +
  annotate('text', x = n_list[x_list == 30][1], y = 30, label = '(159, 30)') +
  labs(x = 'n', y = 'smallest integer s.t. P >= 0.05') +
  theme_light()
```



Therefore, at least 159 disks need to be put on test.

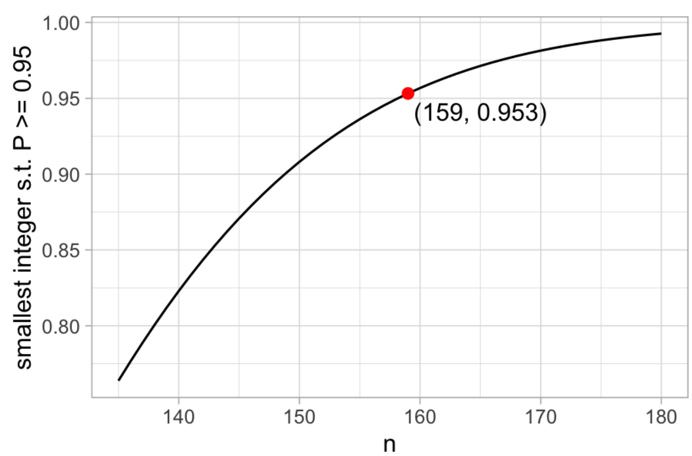
Solution 2: Gamma calculation.

Suppose $S_i \sim \text{Gamma}(i, n\lambda)$. We can use `pgamma(q,shape,rate)` to find the smallest integer n s.t. $P(S_{30} \leq 1) \geq 0.95$.

```r
# Define the function of n to find the proper n.
find_n = function(n){pgamma(q = 1, shape = 30, rate = 0.25 * n)}

# Set the range of n and compute the corresponding possibility p.
n_list = seq(135, 180)
p_list = find_n(n_list)

# Plot the graph and annotate the smallest appropriate number.
ggplot() +
  geom_line(aes(y = p_list, x = n_list)) +
  geom_point(aes(y = p_list[p_list >= 0.95][1],
                 x = n_list[p_list >= 0.95][1]),
                 color = 'red') +
  annotate('text', x = n_list[p_list >= 0.95][1],
           y = p_list[p_list >= 0.95][1], label = '(159, 0.953)') +
  labs(x = 'n', y = 'smallest integer s.t. P >= 0.95') +
  theme_light()
```



Therefore, at least 159 disks need to be put on test, which is the same to the result of poisson calculation.