Homework 6

Put your name and student ID here

2020-10-13

Censored data analysis

Suppose that Y_i are iid sample with PDF $f(y;\theta)$ and CDF $F(y;\theta)$. But if $Y_i > t_i$ then we don't see Y_i we only learn that $Y_i > t_i$. Let $\delta_i = 1$ if Y_i was observed and $\delta_i = 0$ otherwise. The likelihood function is given by

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\theta)^{\delta_i}(1 - F(t_i;\theta))^{1-\delta_i} = \prod_{i=1}^{n} f(x_i;\theta)^{\delta_i} S(x_i;\theta)^{1-\delta_i},$$

where $x_i = \min(y_i, t_i)$ denote the observed data, and $S(t;\theta) = 1 - F(t;\theta)$ is called the survival function in the context of survival analysis.

In our class, we have derived MLE for exponential population, i.e., $Y_i \overset{iid}{\sim} Exp(\lambda)$. We now consider a more flexible distribution – **Weibull distribution**. The Weibull distribution with shape parameter $\gamma > 0$ and rate parameter $\lambda > 0$ has a density given by

$$f(y;\gamma,\lambda) = \lambda\gamma y^{\gamma-1}\exp(-\lambda y^{\gamma})$$

for $y > 0$. The CDF is $F(y;\gamma,\lambda) = 1 - \exp(-\lambda y^{\gamma})$ on $y > 0$. Particularly, if $\gamma = 1$, the Weibull distribution turns out to be an Exponential distribution, and thus it is more flexible. Now suppose $Y_i \overset{iid}{\sim} f(y;\gamma,\lambda)$. Please answer the following questions:

Q1. Derive MLEs for the parameters γ and λ.

Q2. Show the estimated parameters for the two real datasets `aml` and `lung` in R package `survival`, respectively. You may use some numerical algorithm, such as Newton-Raphson algorithm.

Q3. Show the estimated average survival time and plot the estimated survival function $\hat{S}(t;\theta)$ as a function of t using the results in Q2.

Q4. Compare the results in Q3 with the associated results for Exponential population, which was done in our class.