SISCER Mod 12 Survival Analysis

Problem Set 1

Due: July 22, 2021

1. Consider the life table used in the lecture:

Years	Alive at beginning	Deaths	Censored
0-1	146	27	3
1-2	116	18	10
2-3	88	21	10
3-4	57	9	3
4-5	45	1	3
5-6	41	2	11
6-7	28	3	5
7-8	20	1	8
8-9	11	2	1
9-10	8	2	6

   (a) Complete this ten-year life table by calculating survival rate for each time interval, assuming that censoring occurred at

   - the right end of each time interval,
   - immediately prior to the left end of each time interval, and
   - immediately prior to the left end of each time interval for half of the censored participants and at the right end of each time interval for the other half of the censored participants,

   respectively.

   (b) In the same figure, plot these three life table estimates of survival functions, and the point-wise confidence intervals of the third estimate. Comment on your plots.

2. This exercise is for parametric inferences of survival times by exponential distributions.

   (a) Assume that survival times follow an exponential distribution with failure rate θ. That is, the hazard function of a survival time T is $\lambda(t) = \theta$. Calculate the survival function, the density function and the median of T.

(b) Suppose that the collected data of sample size n are in the form of (X_i, Δ_i), $i = 1, 2, \ldots, n$, where

$$X_i = \min(T_i, C_i), \text{ and } \Delta_i = I(T_i \leq C_i).$$

Here, C_i are potential censoring times, X_i are censored survival times, and Δ_i are the event indicators.

Write out a likelihood function for θ, when T_i are assumed to be exponential of failure rate θ, and state your assumptions. Derive the MLE for θ and a variance estimate of the MLE in terms of (X_i, Δ_i). Does your MLE look intuitive? How is your MLE related to the usual person-year estimate of failure rate?

(c) For the PBC dataset that is distributed in the lab, use your MLE to estimate the failure rates of both arms and their associated confidence intervals. Plot their survival function estimates and the associated pointwise confidence intervals based on your MLE.

3. For the same PBC dataset, plot the Kaplan-Meier curves of both arms and their associated pointwise confidence intervals.

   (a) How do they compare with your MLE of the exponential distributions?

   (b) Can you estimate the medians from the Kaplan-Meier curves? if so, do they agree with your exponential MLE? if not, what other summary measures would you like to suggest for the Kaplan-Meier Curves?