Mei-Cheng Wang
Survival Analysis
Biostatistics 140.641

Problem Set 2
Due date: Monday, 9/26/2022

1. Suppose the failure time T has the Exponential(λ) distribution, where $\lambda > 0$ is the hazard parameter. Suppose the observed data are observed subject to censoring: $0.24, 0.47, 0.81, 1.22^+, 1.75, 2.53^+$.

   (a) Write out the likelihood function.

   (b) Derive the maximum likelihood estimate of λ.

2. The following data are the remission times of 42 patients with acute leukemia: 6MP (mercaptopurine) vs placebo. Reference: Freireich et al. (1963) Blood, p699.

   Placebo: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

   6MP: 6, 6, 6, 7, 10, 13, 16, 22, 23, 6^+, 9^+, 10^+, 11^+, 17^+, 19^+, 20^+, 25^+, 32^+, 32^+, 34^+, 35^+.

   (a) Calculate both the Kaplan-Meier estimate and the empirical survival distribution on the basis of the placebo data. What is the relationship between the Kaplan-Meier estimate and the empirical survival distribution from your calculation?

   (b) Do you observe a similar phenomenon which also holds for other uncensored survival data? Explain.

   (c) Calculate the Kaplan-Meier estimate on the basis of the 6MP data.

   (d) Use the variance estimates that you learn from Biostatistics 140.641 (or other courses) to produce a 95% confidence interval at each observed failure time for the empirical survival distribution on the basis of the placebo data. Comment on your results.

   (e) Use the variance estimates that you learn from Biostatistics 140.641 to produce a 95% confidence interval, at each uncensored time, for the

the Kaplan-Meier estimate on the basis of the 6MP data. Comment on your results.

(f) Compare the survival function estimates from the placebo and 6MP groups. Interpret your results.

3. The Kaplan-Meier estimate is a well known nonparametric method for estimating survival function. It generally performs well for (reasonably) large sample size, but also has a few undesirable properties when the sample size is small. Answer the questions based on the following two data sets.

   Data A:
   $0 \;\; 1 \;\; 1 \;\; 2^+ \;\; 3 \;\; 5 \;\; 6^+ \;\; 9 \;\; 10$

   Data B:
   $0 \;\; 1 \;\; 1 \;\; 2^+ \;\; 3 \;\; 6^+ \;\; 7 \;\; 9 \;\; 10$

   (a) Use your "eyes" to compare the survival patterns from the two data sets and comment on what you have observed.

   (b) Calculate the Kaplan-Meier estimate based on Data A.

   (c) Calculate the Kaplan-Meier estimate based on Data B.

   (d) Compare the two estimates from (b) and (c). Comment on the results.

4. Consider the following models for a failure time variable T, $0 < T < \infty$. These models could be regular models (for a regular failure time T) or cure models (for an irregular failure time T).

   M1. The hazard function of T is $h(t) = \theta \times I(t > 0)$, where $\theta > 0$.

   M2. The hazard function of T is $f(t) = \theta e^{\theta t} \times I(t > 0)$, where $\theta > 0$.

   M3. The hazard function of T is $h(t) = 5e^{-\theta t} \times I(t > 0)$, where $\theta > 0$.

   (a) Identify which models are regular models and which models are cure models. Please provide detailed explanation.
   (b) Derive the pdf of T for each model.