

Survival Data Analysis (BIOS 7210)  
Breheny

Assignment 3

Due: Thursday, September 17

1. [KP 1.8] Consider solving for the nonparametric MLE of  $S(t)$  with an imposed constraint that for some time  $b > t_1$ ,  $S(b) = c$ .

(a) Show that, under this constraint, the nonparametric MLE of  $S(t)$  is given by

$$\tilde{S}(t) = \prod_{t_j \leq t} (1 - \tilde{\lambda}_j),$$

where  $\tilde{\lambda}_j = d_j/(n_j + a)$  if  $t_j \leq b$  and  $\tilde{\lambda}_j = d_j/n_j$  otherwise. The value  $a$  is chosen to satisfy the constraint. Hint: The technique of Lagrange multipliers is a useful method for solving optimization problems in the presence of constraints.

(b) Let  $R$  denote the likelihood ratio comparing the likelihood of  $\tilde{S}(t)$  to the MLE  $\hat{S}(t)$  that we derived in class. Show that

$$R = \prod_{t_j \leq b} \left( \frac{n_j}{n_j + a} \right)^{n_j} \left( \frac{n_j + a - d_j}{n_j - d_j} \right)^{n_j - d_j}$$

(c) Consider using the likelihood ratio in (b) to construct a confidence interval for  $S(b)$  by including in the interval all values for  $S(b)$  such that  $R > 0.15$ ; i.e., that have at least 15% of the likelihood of the MLE. It has been shown (Thomas and Grunkemeier, 1975) that this interval has approximate 95% coverage. Construct this interval at 9 days for the MTX+CSP group in the GVHD data and briefly compare it to the intervals we constructed in class.

2. Carry out a simulation to investigate the coverage of the three Kaplan-Meier confidence interval methods we discussed in class. Generate  $n = 100$  true failure times from an  $\text{Exp}(1)$  distribution and censoring times from an  $\text{Exp}(0.5)$  distribution. At time points  $0.1, 0.2, \dots, 2$ , calculate the true survival function and then for each independently generated data set, calculate the three confidence intervals and record whether the CI contain the true  $S(t)$ . Repeat the simulation at least  $N = 10,000$  times to get an accurate estimate of coverage.

(a) Plot the coverage of the three intervals as a function of time.

(b) Comment on the performance of the three CI methods relative to each other; which method seems to perform the best overall?

(c) Comment on the performance of the three CI methods over time; which times are they most/least accurate; why?

3. From 1974 to 1984, the Mayo Clinic conducted a study of patients with primary biliary cirrhosis of the liver. Their data is available from the `survival` package; see `?pbc` for full details. For the purposes of this question, we will restrict ourselves to three variables:

- `time`: The time until either censoring, death, or transplant (in days)

- **status:** What happened at the end of the patient's time on study. Either the patient's failure time was censored (0), the patient required a liver transplant (1), or the patient died (2)
- **stage:** A categorization of the severity of cirrhosis based on a liver biopsy. Stage 1 is the least severe, while Stage 4 denotes the most severe progression of the disease.

The outcome of interest is the time until the patient either dies or requires a liver transplant. This is often referred to as *progression-free survival*; the time that the patient is both alive and in which their health has not deteriorated to a more critical condition.

For this problem, you may use whichever CI method you like, but state which one you used.

- (a) Plot Kaplan-Meier curves, with confidence intervals displayed, for each stage of cirrhosis patient. Briefly comment on the conclusion you would draw from the figure.
- (b) Give estimates, along with confidence intervals, for 5-year progression-free survival for each stage.
- (c) Give estimates, along with confidence intervals, for the median survival time for each stage. If you are unable to provide estimates or intervals for some groups, briefly explain why.
- (d) For patients with Stage 4 cirrhosis, give an estimate and 95% CI for the 25th percentile of survival time.
- (e) For patients with Stage 4 cirrhosis, what is the estimated cumulative hazard at 3 years (using the Nelson-Aalen estimator)?