Homework I

(Due on 02/18/2014 Tuesday by 11:59pm)

Instructions: While discussion with classmates are allowed and encouraged, please try to work on the project independently and direct your questions to me. Please interpret your analysis results using concise and clear language and focusing on interesting findings. Remember to include your R codes in an Appendix. Either an e-copy or a hard copy for submission is acceptable.

PART I Exercises

1. Suppose that the lifetime (in hours) of light bulbs is a random variable $X$ that follows an exponential distribution with a hazard rate of 0.005 failures per hour of use.

   (a) Find the mean and median lifetime of a randomly selected light bulb.

   (b) What is the probability a light bulb will still function after 1,000 hours of use?

   (c) Find the mean residual life after $x_0 = 2,000$ hours. How does it compare to the mean survival time in (a)?

2. For the following small data set of survival time: 3, 4, 5+, 6, 6+, 8+, 11, 14, 15, 16+, where ‘+’ means a right censored survival time.

   (a) Find the Kaplan-Meier estimate of the survival function using the distributed-to-the-right approach. Show your worksheet.

   (b) Find the Kaplan-Meier estimate of the survival function using the product-limit formula form, as well as its variance.

   (c) Find an estimate and its variance of the survival function using the Nelson-Aalen estimate and compare is to the KM curves in part (b).

PART II Computer Project

Consider the lung cancer data available from http://www.mayo.edu/research/documents/lunghtml/DOC-10027247

The data set contains the following variables:

- Enrolling institution
- Survival time
- Status 1=alive, 2=dead
- Age
• Sex 1=male 2=female
• ECOG performance score, as judged by physician: 0,1,2,3
• Karnofsky performance score, as judged by physician: 100, 90, ..., 30
• Karnofsky performance score, as judged by the patient (self)
• Daily calories consumed at meals
• Weight loss in the last 30 days (negative number = weight gain)

Detailed information about the data set can be found in Loprinzi et al. (1994, *J. Clinical Oncology*).

To bring in the data, you may use

```r
lung.cancer <- read.table(
    file="http://www.mayo.edu/research/documents/lungdat/DOC-10027697",
    sep=" ", header=F,
    col.names=c("inst", "time", "status", "age", "sex", "ECOG",
    "Karnofsky.physician", "Karnofsky.patient", "calories",
    "weight.loss"))
head(lung.cancer)
```

With this lung cancer data set, perform the analysis by following the steps listed below.

1. Find and plot the Kaplan-Meier estimate and the 90% pointwise confidence interval of the survival function. You may use a transformation option of your choice.

2. Find the estimate and its 90% CIs of the median survival time from the above KM survival curve.

3. Find the Nelson-Aalen estimate of the survival function and add the curve to the Kaplan-Meier survival plot for a comparison.