Survival Analysis HW 1

David Coomes

Due 1/23/2020

Problem 1

(a) Individual A is administratively (right) censored at age 60 (30 years after the beginning of the study), and contributes 30 years to left truncation.

(b) Individual B experiences interval censoring from the time they developed breast cancer after the fourth exam until they were diagnosed with breast cancer at the fifth exam. They contribute 40 years to left truncation in this study.

(c) Individual C is right censored at age 61 due to death, and they contribute 50 years to left truncation.

(d) Individual D is right censored at age 54 (the time of their last clinical exam) due to loss to follow up. They contribute 42 years to left truncation.

If we were interested in studying the time from enrollment into study until onset of breast cancer instead of age at onset, this would change my answers to when each individual contributes to left truncation, but it would not change the answers to when they were censored. In that case (time from enrollment), there would be no left truncation present.

Problem 2

(a) This data set is affected by left truncation. By sampling a set of current Crohn's disease patients at a specific time, we are more likely to sample those individuals who will have a longer period between diagnosis of disease and the outcome of interest. For censoring, if we are able to follow all recruited patients for 10 years, and there is no loss to follow up, then we will likely have some administrative censoring at the end of the study. If, by some chance, all those who were selected had the outcome, then there would be no censoring.

(b) I do not necessarily agree with this statement. If A and T are totally independent, then I would agree that the observed ages at diagnosis are not affected by selection bias. But, if there is an association between these two, then I would argue that there is some selection bias on age.

Problem 3

(a) The study population was 312 primary biliary cirrhosis patients enrolled in two RCTs at the Mayo Clinic.

(b) The initiating event was the date the patient was determined eligible for the clinical trials.

(c) The terminating event was death from any cause.

(d) The time scale for this study was in months.

(e) The causes of censoring were loss to follow up, liver transplant, and end of the study.

(f) The end of the study was not likely related to the outcome in this study. It's difficult to know whether loss to follow up is related to death, but I can't think of a strong reason why it would be so in this case. Liver transplantation may be related to the outcome because those that were worse off (closer to dying) may have been more likely to receive a transplant, but it is difficult to know without knowing the rules regarding transplant priority.

(g) For those in the 'Low' category, approximately 15% die within the first five years. This number is almost 50% for those in the 'Medium' category, and about 80% for those in the 'High' category.

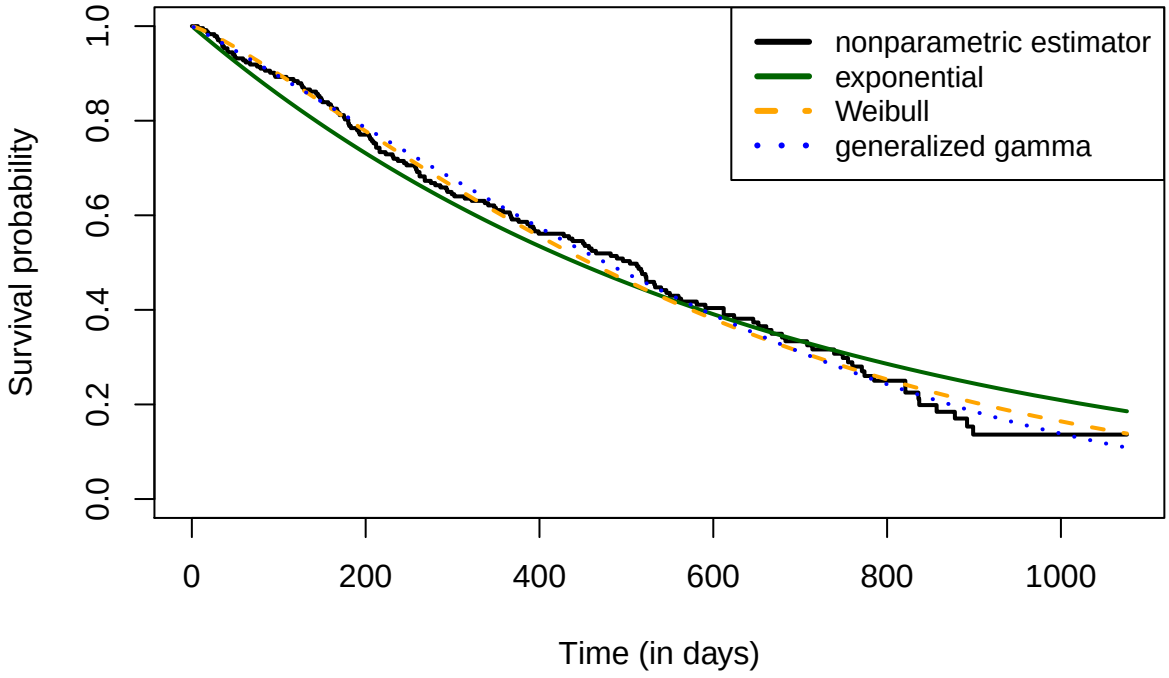
Problem 4

(a) The mean follow up time was 402.6 days and the median follow up time was 367.5 days. The proportion of censored individuals was 37.0%.

(b)

Model	Parameter	Estimate	95% Lower	95% Upper	Maximum loglikelihood
Exponential	λ	0.00157	0.00132	0.00182	-1118.93
Weibull	λ	0.00162	0.00141	0.00183	-1114.92
	P	1.226	1.060	1.392	
Generalized Gamma	μ	6.550	6.269	6.831	-1114.36
	Σ	0.660	0.326	0.993	
	Q	1.468	0.385	2.552	

(c) The exponential model does not fit the data particularly well. It is off by close to 10% at times, and perhaps even more towards the end of the study. The Weibull and the generalized gamma both fit the data fairly well. They seem to only be off from the nonparametric estimator by a few percentage points, and maybe as much as 5% towards the end of the study. The generalized gamma (unsurprisingly) seems to fit the data slightly better, however, it is difficult to tell just by looking at the curves.



(d) The Weibull model is an appropriate simplification of the generalized gamma model in this example. By comparing the two models using a Likelihood Ratio Test, we get a p-value of 0.29 which does not allow us to reject our null hypothesis that these two models are the same.

(e)

i. The median time until exit is 68.99 [95% CI: 58.56-79.42] days
ii. The probability that no exit will occur by one year is 59.15% [95% CI: 53.76-64.54]

iii. The probability that no exit will occur by two years given that no exit has occurred by one year is 49.48% [95% CI:42.22-56.74]

(f) The exponential model is not an appropriate simplification of the Weibull model. Using a Likelihood Ratio Test to compare the two models we get a p-value of 0.0046, allowing us to reject the null hypothesis that these models are the same.

(g) Lambda for clinic 1 = 0.00205 [95% CI:0.00168-0.00241] Lambda for clinic 2 = 0.00077 [95% CI:0.00049-0.00106]

By comparing these two models using Wald's test, we get a p-value of 6.40*10^-8, allowing us to reject the null hypothesis that the distribution of time to recurrence is the same for clinics 1 & 2.

(h) Lambda for no incarceration = 0.00145 [95% CI: 0.00114-0.00177] Lambda for incarceration = 0.00172 [95% CI: 0.00132-0.00213]

By comparing these two models using Wald's test, we get a p-value of 0.306, which does not allow us to reject the null hypothesis that the distribution of time to recurrence is the same for those who were incarcerated vs. those that were not.