

School of Statistics, University of the Philippines (Diliman) Linangan ng Estadistika, Unibersidad ng Pilipinas (Diliman)

12 of November 2015

Al-Ahmadgaid B. Asaad

Problem Set 5 | Stat 245

- 2.3 Table 1.6 of section 1.11 gives data on the times in weeks from diagnosis to death of 28 patients with diploid cancers of the tongue.
 - (a) Using a Dirichlet prior for S(t) with $\alpha(t, \infty) = 4/(1 + 0.15t^{0.5})$, find the Bayes estimate of the survival function under squared-error loss.
 - (b) Compare the estimates found in parts (a) to the usual Kaplan-Meier estimate of the survival function.

Solution

(a) The following is the output of the out_data1 variable in the R script.

	time	SD	KM
1	0	1.0000000	1.00000000
2	1	0.9524457	0.96428571
3	3	0.9117215	0.92857143
4	4	0.8774038	0.89285714
5	5	0.8123542	0.82142857
6	8	0.7752646	0.78571429
7	12	0.7369638	0.74829932
8	13	0.7032210	0.71088435
9	18	0.6657218	0.67346939
10	23	0.6293179	0.63605442
11	26	0.5948032	0.59863946
12	27	0.5616364	0.56122449
13	30	0.5273797	0.52380952
14	42	0.4893593	0.48639456
15	56	0.4521172	0.44897959
16	62	0.4179058	0.41156463
17	67	0.4166539	0.41156463
18	69	0.3808171	0.37040816
19	76	0.3791358	0.37040816

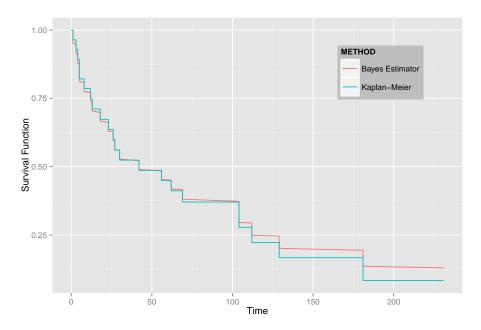


Figure 1: Surival Function using Bayesian and Kaplan-Meier Estimators

```
20 104 0.2953051 0.27780612
21 112 0.2488508 0.22224490
22 129 0.2009980 0.16668367
23 176 0.1946413 0.16668367
24 181 0.1356016 0.08334184
25 231 NA 0.08334184
```

SD column is the Bayesian estimate of the survival function using Dirichlet as a priori. While column KM stands for Kaplan-Meier estimate, this is used to answer part (b). The last column of SD is NA (stands for Not Available) since there is no value for Y_{26} (out of range), from Equation 6.4.1, in the data.

(b) The plot of the survival functions is in Figure 1. The two survival lines indicate that there is no significant departure of the Bayesian estimator from the Product-Limit estimator. The Bayesian survival function is not a step function since it is continuous between distinct death times and has jumps at these death times, this difference is not pronounced but can be observed in the plot.