MLDS 401/IEMS 404-1 (Fall 2023): Lab 9 – 11/28/2023

Question 1

Let $X_1, \ldots, X_n$ be a random sample from the following pdf:

$$f(x;\theta) = \frac{x}{\theta} e^{-x^2/(2\theta)}, \quad x > 0.$$

(c) Find the MLE of θ. How does the answer compare to part a? *Answer:*

$$L(\theta) = \prod_{i=1}^{n} \frac{x_i}{\theta} e^{-x_i^2/(2\theta)} = \theta^{-n} \prod_{i=1}^{n} x_i e^{-x_i^2/(2\theta)}$$

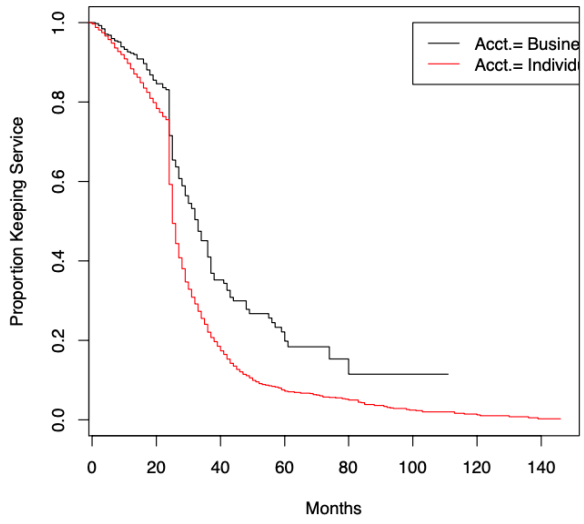
$$l(\theta) = -n \log \theta + \sum_{i=1}^{n} \log x_i - \frac{1}{2\theta} \sum_{i=1}^{n} x_i^2$$

$$\frac{dl(\theta)}{d\theta} = \frac{-n}{\theta} + 0 + \frac{1}{2\theta^2} \sum_{i=1}^{n} x_i^2 = 0$$

Solve giving $\hat{\theta} = \frac{1}{2n} \sum_i x_i^2$, *which is the same as part a.*

10.8 (Cellphone data: Kaplan-Meier curves and logrank test)

(a) The Kaplan-Meier curves are shown below.



Both plots show steep drops at approximately 24 months. The explanation is that service contracts used to be for 24 months which is

when many people changed their service. Also note that individual customers change service much more frequently than business customers. This is to be expected since businesses sign long term contracts.

(b) Here is the output for the log rank test. We see that the difference between business customers and individual customers is highly significant.

```
survdiff(formula = Surv(Months, Churn) ~ Account_Type,
        data = cell)

n=4912, 88 observations deleted due to missingness.

                   N Observed Expected (O-E)^2/E (O-E)^2/V
Account_Type=B   418      143      225     29.73      35.6
Account_Type=I 4494     2271     2189      3.05      35.6

 Chisq= 35.6  on 1 degrees of freedom, p= 2.37e-09
```

Question 18 (3 pts.): Consider a small artificial data set with a time-varying covariate Rx shown below in a short format. The variables have their usual meanings.

ID	Death	Tx	TxTime	Survival
1	1	0	N/A	6
2	0	1	6	8
3	1	1	5	7
4	0	0	N/A	8
5	1	1	5	9
6	0	1	3	10

The same data are shown in long format in the following table. Check to see if there are any errors in the long format, and if so how many and for which ID.

ID	Death	Tx	tstart	tstop
1	1	0	0	6
2	0	0	0	6
2	0	1	6	8
3	1	0	0	5
3	1	1	5	7
4	0	0	0	8
5	0	0	0	5
5	1	0	5	9
6	0	0	0	10

Answer: 3 errors, ID ♯ 3, 5 and 6. The correct long format table is shown below.

ID	Death	Tx	tstart	tstop
1	1	0	0	6
2	0	0	0	6
2	0	1	6	8
3	0	0	0	5
3	1	1	5	7
4	0	0	0	8
5	0	0	0	5
5	1	1	5	9
6	0	0	0	3
6	0	1	3	10