

Question 1

Let X_1, \dots, 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?

10.8 (Cellphone data: Kaplan-Meier curves and logrank test) A cellular service provider keeps data on how many months their customers maintained service with the company before they switched to another company. If a customer did not switch to another service provider then the service time on the customer is censored. The file `cellphone data.csv` contains 4912 records of customers (88 have missing data on at least one variable) with data on the following variables: Months = the number of months of service, Account-Type = Business (B) or Individual (I), Churn = censor indicator (0 if no, 1 if yes), Line-Count = number of phone lines served).

- a) Make Kaplan-Meier curves for business and individual customers. Which customers maintain their service longer? How long do most customers maintain the service? Does this seem to be related to the standard service contract period of two years that the cellphone providers used to have before this restriction was removed?
- b) Do the logrank test to check if there is a significant difference between the two survival curves.

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