

# Data Analysis

## *Class Exercises - Solutions*

Adaloglou Lazaros

*Aristotle University of Thessaloniki*

# Contents

<b>1</b>	<b>Intro</b>	<b>1</b>
<b>2</b>	<b>Probability and Random Variables</b>	<b>1</b>
2.1	Exercise 1 . . . . .	1
2.2	Exercise 2 . . . . .	2
2.3	Exercise 3 . . . . .	2
2.4	Exercise 4 . . . . .	2
2.5	Exercise 5 . . . . .	2
2.6	Exercise 6 . . . . .	2
<b>3</b>	<b>Elements of Statistics</b>	<b>2</b>
3.1	Exercise 1 . . . . .	2

## 1 Intro

Data Analysis in Matlab.

## 2 Probability and Random Variables

### 2.1 Exercise 1

Confirm the Definition of Probability as the limit of relative frequency for repetition count tending to infinity. Simulate the toss of a coin  $n$  times using the random number generator, either from a uniform discrete distribution (two-value for *heads* and *tails*), or from a uniform continuous distribution in the interval  $[0,1]$  using a threshold of 0.5 (e.g. a number smaller than 0.5 is *heads* and a number bigger than 0.5 is *tails*). Repeat the experiment for increasing  $n$  and calculate every time the ratio of *tails* in  $n$  repetitions. Plot the respective graph of the ratio for different  $n$ . For the creation of random numbers the *rand* or *unidrnd* functions can be used.

## 2.2 Exercise 2

## 2.3 Exercise 3

## 2.4 Exercise 4

## 2.5 Exercise 5

## 2.6 Exercise 6

# 3 Elements of Statistics

## 3.1 Exercise 1

### Solution

a) As stated,  $\{x_1, x_2, \dots, x_n\}$  are independent observations. Also the Cumulative Distribution Function (cdf) is known

$$F_X(x; \lambda) = \int_{-\infty}^x f_X(u; \lambda) du = \int_{-\infty}^x e^{-\lambda} \frac{\lambda^u}{u!} du.$$

The Likelihood Function is

$$L(x_1, x_2, \dots, x_n; \lambda) = f_X(x_1; \lambda) f_X(x_2; \lambda) \dots f_X(x_n; \lambda)$$

So the Maximum Likelihood Estimator for lambda is the  $\hat{\lambda}$  that maximizes the Likelihood Function, thus

$$\hat{\lambda} = \arg \max_{\lambda} L(x_1, x_2, \dots, x_n; \lambda)$$

Now we take the natural logarithm ( $\log_e$ ) of the Likelihood Function. We do this because the natural logarithm is a monotonically increasing function. This ensures that the maximum value of the log of the probability occurs at the same point as the original probability function. Now this can be differentiated to find the maximum.

$$\frac{\partial \log_e L(x_1, x_2, \dots, x_n; \lambda)}{\partial \lambda} = 0 \tag{1}$$

First we elaborate on  $I = \log_e L(x_1, x_2, \dots, x_n; \lambda)$

$$I = \log_e (f_X(x_1; \lambda) f_X(x_2; \lambda) \dots f_X(x_n; \lambda))$$

$$= \log_e \left( e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \dots e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} \right)$$

$$= \log_e \left( e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \right)$$

$$\begin{aligned}
&= \log_e(e^{-n\lambda}) + \log_e(\lambda^{\sum_{i=1}^n x_1}) - \log_e \prod_{i=1}^n x_1! \\
&= -n\lambda + \left( \sum_{i=1}^n x_1 \right) \log_e(\lambda) - \log_e \prod_{i=1}^n x_1!
\end{aligned}$$

Substituting in (1) we have

$$\frac{\partial (-n\lambda + (\sum_{i=1}^n x_1) \log_e(\lambda) - \log_e \prod_{i=1}^n x_1!)}{\partial \lambda} = 0$$

$$-n + \frac{1}{\hat{\lambda} \log_e e} \sum_{i=1}^n x_1 = 0$$

$$\sum_{i=1}^n x_1 = n\hat{\lambda}$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_1$$

So the Maximum Likelihood Estimator of  $\lambda$  is the sample mean

$$\hat{\lambda} = \bar{x}$$