Statistical Inference
R: Assessment 1

István Z. Kiss and Francesco Di Lauro

1. The deadline for this assessment is 14:00 Thursday, Week 6, 4th of March 2021.
2. Please submit online and use your corresponding Canvas site for submission, i.e. if you are taking this course at Level 7, please submit via the Canvas site corresponding to Level 7 (same for Level 6).
3. Your work must be submitted as ONE SINGLE PDF file. This may contain R code, figures produced in R to support your answers, handwritten calculations that you typed up or photographed and inserted it into your word or other editor.
4. You can use a text editor of your choice.
5. Each figure must be annotated with title, axes labels and legends, when multiple lines are needed.
6. Please stick to the methods used in class!

1 Identify distributions from data.

The file *R_assessment1_data.csv* (downloaded from Canvas) contains 6 columns, each of which is a sample of 100 i.i.d. realisations from the distributions given below:

Binomial, Beta, Exponential, Uniform, Log-normal, Poisson.

Your task is to identify which distribution generated which column as well as estimating the parameters of each distributions (e.g. Maximum Likelihood estimation). You may want to use different plotting techniques and manipulation previously seen in class.

Hint: The MLE estimation can be done using the package *fitdistrplus* available on CRAN servers (probably needs to be setup as it is not part of standard packages). In the case of a binomial distribution, the parameter p is estimated once n is given. Since n is not known, try different values of n and pick the most likely one (highest likelihood).

2 Likelihood ratio test.

A sample $x = (x_1, \ldots, x_n) \in (\mathbb{R}^+)^n$, of i.i.d. realisations from a normal distribution with parameter μ and σ^2 is provided. Suppose that for practical reasons, experts agree that only two values of μ are possible. This is represented by the two competing simple hypotheses:

$$H_0 := \{\mu = \mu_0\} \text{ and } H_1 := \{\mu = \mu_1 > \mu_0\}.$$

1. Provide an analytical expression for $\Lambda(x) = \frac{\mathcal{L}(\mu_0;x)}{\mathcal{L}(\mu_1;x)}$, the likelihood ratio test statistic of H_0 against H_1.

2. Prove the following statement:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sigma^2}{n(\mu_0 - \mu_1)} \left(\ln(\Lambda(x)) + \frac{n\left(\mu_0^2 - \mu_1^2\right)}{2\sigma^2} \right) \sim N\left(\mu_0, \frac{\sigma^2}{n}\right). \tag{1}$$

3. By manipulating $\Lambda(x)$, as in the lectures, show that the rejection of the null hypothesis comes down to the condition below

$$\bar{X} > \kappa.$$

   Work out κ (i.e. give its numerical value) such that

$$\mathbb{P}(\bar{X} > \kappa | H_0) = 0.05,$$

   when $n = 100$, $\mu_0 = 10$ and $\sigma^2 = 100$.

4. Knowing the value of κ, work out and give the numerical value for κ_Λ such that

$$\mathbb{P}(\Lambda(x) < k_\Lambda | H_0) = 0.05. \tag{2}$$

5. Use this test with the two sample sets, provided as columns with headers x and y in the *R_assessment1_norm.xlxs* file, to conduct a statistical test of significance 5%. Each sample contains $n = 100$ numbers. The other parameters are $\mu_0 = 10$, $\mu_1 = 15$ and $\sigma^2 = 100$. For both sample sets report whether you reject H_0.

6. Work out the p-value associated to this test, and do this for both samples. Please report the numerical values for p.

7. Work out and plot the power function of the test. You should include a plot of the power function in your report.

8. Could you come up with a more, powerful test with the same α?