MAP531: Homework

You are asked to provided answers to all these exercises as both Rmd and pdf files. The two files should be uploaded on Moodle on the 15th of November (23h59 Paris time).

This homework should be done by groups of 2. Only one submission per group on moodle, with both names indicated in the file.

This homework is composed of 2 independent problems.

Some of the questions are a bit more technical: they are marked by a * and are optional.

Problem 1: Estimating parameters of a Poisson distribution to model the number of goals scored in football

We recall that the Poisson distribution with parameter $\theta > 0$ has a pdf given by $(p(\theta, k), k \in \mathbb{N})$ w.r.t the counting measure on $\mathbb{N}$:

$$p(\theta, k) = \exp(-\theta)\frac{\theta^k}{k!}\,.$$

Question 1

Is it a discrete or continuous distribution? Can you give 3 examples of phenomenons that could be modeled by such a distribution in statistics?

Question 2

Compute the mean and the variance of this distribution as a function of λ.

Remark: that if X_1 and X_2 are two independent random variables following a Poisson distribution with respective parameters $\lambda_1 > 0$ and $\lambda_2 > 0$, then $X_1 + X_2$ has a Poisson distribution of parameter $\lambda_1 + \lambda_2$. You do not need to prove this result.

We are provided with n independent observations of a Poisson random variable of parameter $\theta \in \Theta = \mathbb{R}_+^*$.

Question 3

- What are our observations? What distribution do they follow?
- Write the corresponding statistical model.
- What parameter are we trying to estimate?

Question 4

- What is the likelihood function?
- Compute the Maximum Likelihood Estimator $\hat{\theta}_{ML}$.

Question 5

Prove that $\sqrt{n}(\hat{\theta}_{ML} - \theta)$ converges in distribution as $n \to \infty$.

Question 6

- Prove that $\sqrt{n}\frac{\hat{\theta}_{ML}-\theta}{\sqrt{\theta_{ML}}}$ converges in distribution as $n \to \infty$.
- On R, verify that the distribution of the random variable $\sqrt{n}\frac{\hat{\theta}_{ML}-\theta}{\sqrt{\theta_{ML}}}$ is what you found theoretically, through a histogram and a QQ-plot (compute $Nattempts = 1000$ times the random variable $\sqrt{n}_{sample}\frac{\hat{\theta}_{ML}-\theta}{\sqrt{\theta_{ML}}}$ from a sample of size n_{sample} of simulated Poisson data, with $\theta = 3$, like in PC2).

Question 7

For $\alpha \in (0, 1)$, give an asymptotic confidence interval of level α, that is an interval $[a_n(\alpha, (X_i)_{i\in\{1\ldots,n\}}); b_n(\alpha, (X_i)_{i\in\{1\ldots,n\}})]$, such that:

$$\lim_{n\to+\infty} \mathbb{P}\Big(\theta \in \Big[a_n\left(\alpha, (X_i)_{i\in\{1\ldots,n\}}\right); b_n\left(\alpha, (X_i)_{i\in\{1\ldots,n\}}\right)\Big]\Big) \geq 1 - \alpha.$$

Question 8*

Using δ-method (seen during refreshers), prove that $\sqrt{n}(2\sqrt{\hat{\theta}_{ML}} - 2\sqrt{\theta})$ converges in distribution as $n \to \infty$.

Question 9* (another CI)

Give another asymptotic confidence interval of level α, based on question 9, that is an interval $[c_n(\alpha, (X_i)_{i\in\{1\ldots,n\}}); d_n(\alpha, (X_i)_{i\in\{1\ldots,n\}})]$, such that:

$$\lim_{n\to+\infty} \mathbb{P}\Big(\theta \in \Big[c_n\left(\alpha, (X_i)_{i\in\{1\ldots,n\}}\right); d_n\left(\alpha, (X_i)_{i\in\{1\ldots,n\}}\right)\Big]\Big) \geq 1 - \alpha.$$

Question 10

- Propose two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ based on the first and second moments of a Poisson distribution.
- What can you say about $\hat{\theta}_1$?

Question 11

Compute the Bias, the Variance, and the quadratic risk of $\hat{\theta}_{ML}$.

Question 12*

Compute the Cramer Rao bound. What do you conclude about $\hat{\theta}_{ML}$?

Question 13

Let $\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, with $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Show that:

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \theta)^2 - (\theta - \bar{X}_n)^2.$$

Question 14

- Compute $\mathbb{E}(\theta - \bar{X}_n)^2$.
- Prove that $\hat{\theta}_2$ is an biased estimator of θ and give the bias. How can we get an unbiased estimator?

Question 15

- Using the decomposition in Question 13, prove that $\sqrt{n}(\hat{\theta}_2 - \theta)$ converges in distribution, and give a third asymptotic confidence interval centered in $\hat{\theta}_2$. Comment on this third interval.
- You may use: $Var[(X_i - \theta)^2] = 2\theta^2 + \theta$.
- Compare the asymptotic variance to the one of $\hat{\theta}_{ML}$ and to the Cramer Rao bound. What can you say?

Question 16*

Compute the probability generating function of the Poisson distribution given by $G_X(s) = \mathbb{E}[\exp(sX)]$ for $s \in \mathbb{R}$. Recover the result of question 2 and prove $Var[(X_i - \theta)^2] = 2\theta^2 + \theta$.

Problem 2: Analysis of the USJudgeRatings data set

This exercise is open. You are asked to use the tools we have seen together to analyze the USJudgeRatings data set. This data set is provided in the package *datasets*. Your analysis should be reported here and include:

- an introduction
- a general description of the data
- the use of descriptive statistics
- the use of all techniques we have seen together that might be relevant
- a conclusion

Overall, your analysis, including the graphs and the codes should not exceed 15 pages in pdf.