MAP 531: Homework

Paul-Antoine GIRARD & Adrien TOULOUSE

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 $\mathbb{N}$:

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

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?

The poisson distribution is a discrete distribution since it has a countable number of possible values ($\mathbb{N}$).

In statistics, we use this distribution to compute the probability of a given number of (rare) events in a time period or the probability of a discrete waiting time until the next event (eg. number of minutes).

For example a poisson distribution can model:

- The number of patients arriving in an emergency room between 9 and 10am.
- The number of minutes we wait a bus at the bus stop.
- In quality control, the number of manufacturing defect.

Question 2: Compute the mean and the variance of this distribution.

We assume that $\mathbb{X}$ follows a Poisson distribution with parameter $\theta > 0$.

$$\mathbb{E}[\mathbb{X}] = \sum_{i=0}^{\infty}(i * p(\theta, i)) = \sum_{i=0}^{\infty}(i * e^{-\theta}\frac{\theta^i}{i!}) = \theta * e^{-\theta}\sum_{i=1}^{\infty}(\frac{\theta^{i-1}}{(i-1)!}) = \theta * e^{-\theta}\sum_{i=0}^{\infty}(\frac{\theta^i}{i!}) = \theta * e^{-\theta} * e^{\theta} = \theta$$

$$\mathbb{E}[\mathbb{X}^2] = \sum_{i=0}^{\infty}(i^2 * p(\theta, i)) = \sum_{i=0}^{\infty}(i^2 * e^{-\theta}\frac{\theta^i}{i!}) = \theta * e^{-\theta}\sum_{i=1}^{\infty}(i\frac{\theta^{i-1}}{(i-1)!}) = \theta * e^{-\theta}\sum_{i=0}^{\infty}((i+1)\frac{\theta^i}{i!})$$

$$= \theta * e^{-\theta}[\sum_{i=0}^{\infty}(i\frac{\theta^i}{i!}) + \sum_{i=0}^{\infty}(\frac{\theta^i}{i!})] = \theta * e^{-\theta}[\theta * e^{\theta} + e^{\theta}] = \theta(\theta + 1)$$

Question 3: What are our observations? What distribution do they follow? Write the corresponding statistical model. What parameter are we trying to estimate?

We are provided with n independent observations of a Poisson random variable of parameter $\theta \in \Theta = \mathbb{R}_+^*$.
Our observations are $X_k \sim Pois(\theta), \forall k \in 1, ..., n$.
The corresponding statistical model is

$$\mathbb{M} = \{p(. \mid \theta),\ \theta \in \Theta\}$$

We are trying to estimate the parameter θ.

Question 4: What is the likelihood function? Compute the Maximum Likelihood Estimator.

The likelihood function is the function on θ that makes our n observations most likely.

$$l(\theta) = \prod_{k=1}^{n} p(\theta, x_k) = \prod_{k=1}^{n} e^{-\theta} \frac{\theta^{x_k}}{x_k!}, with\ x_k \in \mathbb{N}, \forall k \in 1, ..., n$$

$$L(\theta) = log(l(\theta)) = \sum_{k=1}^{n}(-\theta + x_k log(\theta) - log(x_k!)) = -n\theta + log(\theta)\sum_{k=1}^{n} x_k - \sum_{k=1}^{n} log(x_k!)$$

By derivating with respect to θ, we have:

$$L'(\theta) = -n + \frac{\sum_{k=1}^{n} x_k}{\theta}$$

Then, we set this derivative equal to zero to obtaine a critical point:

$$L'(\theta) = 0 \Leftrightarrow -n + \frac{\sum_{k=1}^{n} x_k}{\theta} = 0 \Leftrightarrow \hat{\theta} = \overline{x}$$

and this critical point is a local maximum, and we will assume that it is also a global maximum of the likelihood function:

$$L''(\theta) = -\frac{\sum_{k=1}^{n} x_k}{\theta^2} < 0$$

So, the maximum likelihood estimator is:

$$\hat{\theta}_{MLE} = \overline{x}$$

Question 5: Prove that ** converges in distribution as n.

We have that:

$$\mathbb{E}[\overline{x}] = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[x_k] = \mathbb{E}[x_1] = \theta$$

$$\mathbb{V}(\overline{x}) = \frac{1}{n^2}\sum_{k=1}^{n} \mathbb{V}(x_k) = \frac{1}{n}\mathbb{V}[x_1] = \frac{\theta}{n}$$

Applying the central limit theorem, we have that $\sqrt{n}(\hat{\theta}_{MLE} - \theta)$ converges towards a Gaussian $\mathcal{N}(0, \theta)$.

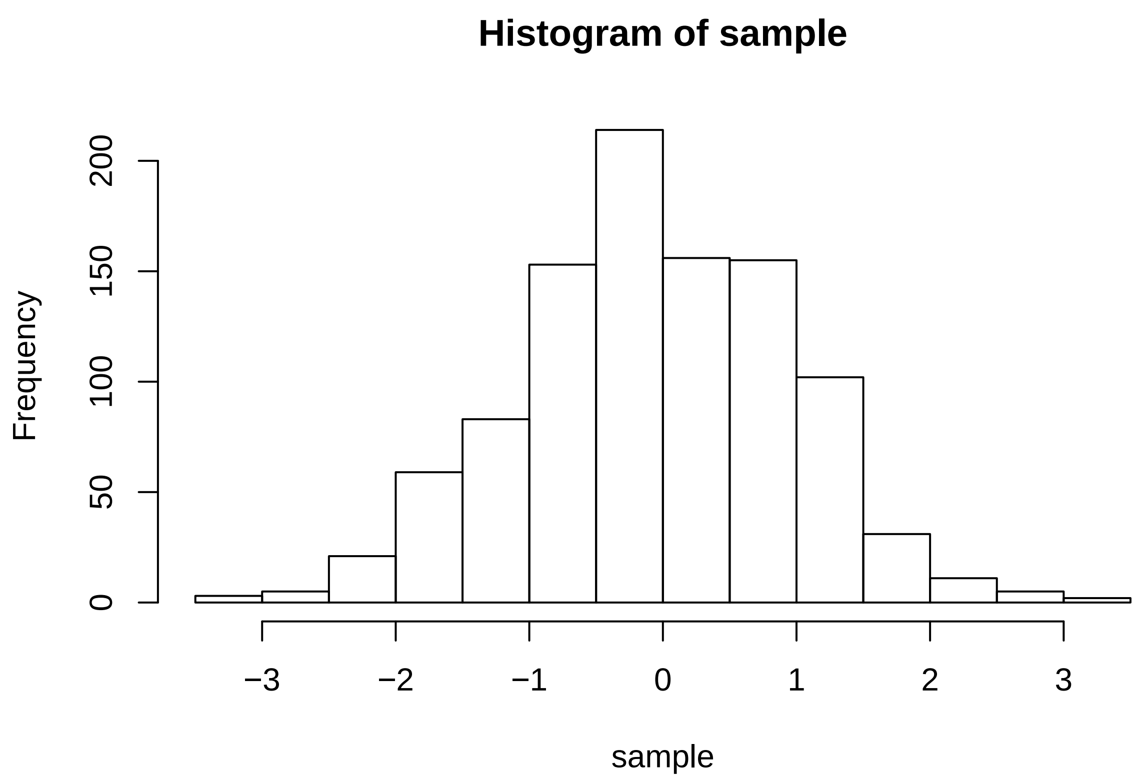
Question 6:

By continuous mapping, $\sqrt{\hat{\theta}_{MLE}}$ converges in probability towards $\sqrt{\theta}$. Then, by Slutsky's theorem, we have that $\sqrt{n}\frac{(\hat{\theta}_{MLE} - \theta)}{\sqrt{\hat{\theta}_{MLE}}}$ converges in law towards a gaussian $\mathcal{N}(0, 1)$.

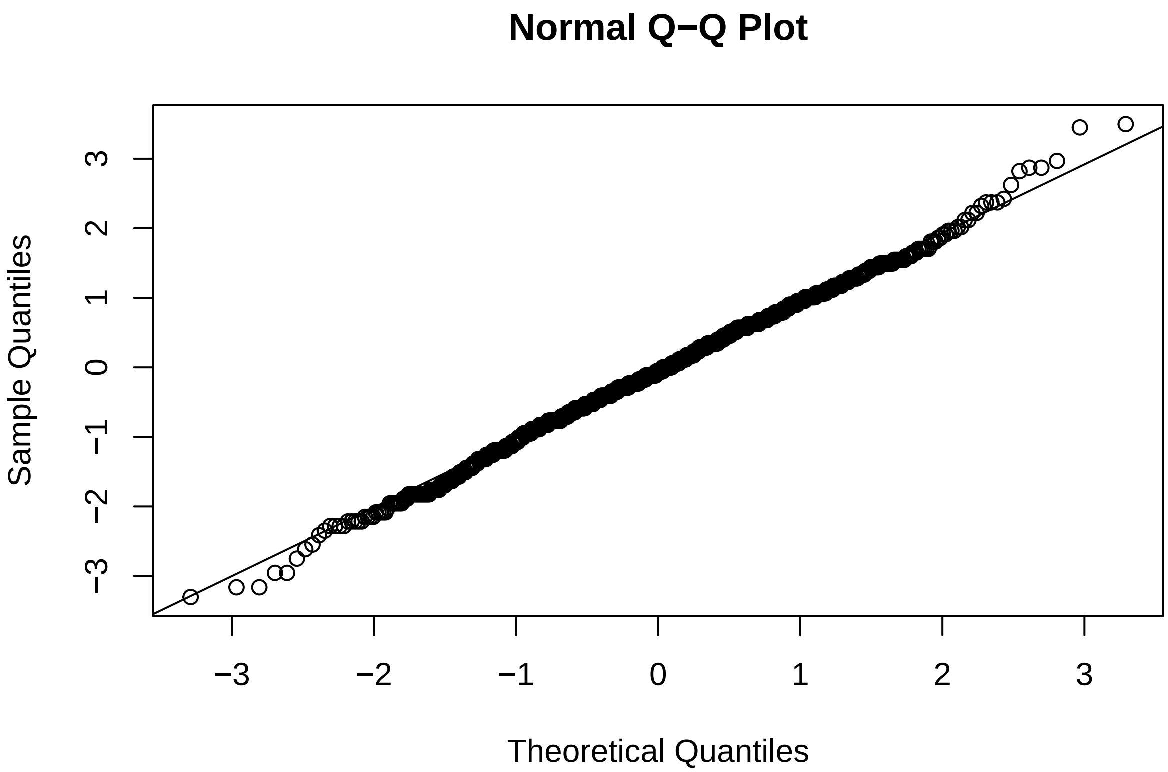
Let's check this result in R by simulating 1000 times our random variable $\sqrt{n}\frac{(\hat{\theta}_{MLE} - \theta)}{\sqrt{\hat{\theta}_{MLE}}}$ with a sample size of 100:

```r
Nattempts = 1000
nsample = 100
lambda = 3
sample = rep(0, 1000)
for (i in 1:Nattempts)  # can be written without the for loop (nicer) !
{poisson_sample = rpois(nsample, lambda)
  sample[i] = sqrt(nsample) * (mean(poisson_sample) - lambda) / sqrt(mean(poisson_sample))
}

hist(sample)
```




```r
qqnorm(sample)
qqline(sample)
```



Question 7:

Let Z_n be our random variable, so that $Z_n = \sqrt{n}\frac{(\hat{\theta}_{MLE}-\theta)}{\sqrt{\hat{\theta}_{MLE}}}$

$$\mathbb{P}(-z_{1-\alpha/2} \leq Z_n \leq z_{1-\alpha/2}) = 1-\alpha \Leftrightarrow P(-z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}_{MLE}}{n}} \leq \hat{\theta}_{MLE} - \theta \leq z_{1-\alpha/2}\sqrt{\frac{\hat{\theta}_{MLE}}{n}}) = 1-\alpha$$

For $\alpha \in (0,1)$, an asymptotic confidence interval for θ of level α is therefore :

$$[\hat{\theta}_{MLE} - z_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_{MLE}}}{\sqrt{n}}; \hat{\theta}_{MLE} + z_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_{MLE}}}{\sqrt{n}}]$$