Exercise 1 (Unbiased estimation). Let $(X_1, \ldots, X_n)$ i.i.d random variables drawn from a Poisson distribution with parameter $\theta > 0$. We are interested in estimating $\mathsf{P}_\theta(X_1 = 0)$ from $(X_1, \ldots, X_n)$.

1) Prove that this model belongs to the exponential family and determine the canonic statistic S. Give the important properties of S, and derive its distribution.

2) By using the Law of Large Numbers, propose a naturel estimator for $\mathsf{P}_\theta(X_1 = 0)$. Is this empirical estimator unbiased ?

3) Prove that the conditional distribution of X_k conditionally to $S = s$ is a binomial distribution with parameters $(\frac{1}{n}, s)$, for all $k \in \{1, \ldots, n\}$.

4) Deduce from previous question that $\delta_S = (1 - \frac{1}{n})^S$ is the optimal estimator of $\mathsf{P}_\theta(X_1 = 0)$.

5) Prove that δ_S converges (almost surely) towards $\mathsf{P}_\theta(X_1 = 0)$ when n tends to infinity.

6*) Compute the score function and the Fisher information.

7*) Deduce the Cramér-Rao bound when estimating $\mathsf{P}_\theta(X_1 = 0)$. Does δ_S achieve this bound ?

Exercise 2 (Confidence Interval – Estimation). é Let $(X_n)_{n\geq 1}$ a sequence of random variables, i.i.d that follows a Gaussian distribution $\mathcal{N}(\theta, \theta)$, with $\theta > 0$. The objective of this exercise is to present a method to estimate θ as well as to give a (good) confidence interval for this estimation. Let us denote

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \text{and} \qquad V_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2.$$

1) What is the distribution of $\bar{X}_n$, its expectation and its variance ? Give the limiting distribution of $(\sqrt{n}(\bar{X}_n - \theta))_{n\geq 1}$.

2) What is the distribution of V_n, its expectation and its variance ? Give the limiting distribution of $(V_n)_{n\geq 2}$.

3) What is the distribution of $(\bar{X}_n, V_n)$. Give the limit of $(\bar{X}_n, V_n)_{n\geq 2}$.

4) Let us consider the class of random variables T_n^λ defined as

$$T_n^\lambda = \lambda \bar{X}_n + (1 - \lambda)V_n, \quad \lambda \in \mathbb{R}.$$

Compute the expectation and the variance of T_n^λ and derive the (almost sure) convergence of $(T_n^\lambda)_{n\geq 2}$.

5) Analyze the convergence (in distribution) of $(\sqrt{n}(\bar{X}_n - \theta))_{n\geq 1}$.

6) Analyze the convergence (in distribution) of $(\sqrt{n}(V_n - \theta))_{n\geq 2}$.

7) Analyze the convergence (in distribution) of $(\sqrt{n}(\bar{X}_n - \theta, V_n - \theta))_{n\geq 2}$.

8)Analyze the convergence (in distribution) of $(\sqrt{n}(T_n^\lambda - \theta))_{n\geq 2}$.

9) Let $\sigma = \sqrt{\lambda^2\theta + 2(1-\lambda)^2\theta^2}$. From T_n^λ, σ and n, build an asymptotic 95%-confidence interval for θ. In other words, find a random interval, depending on T_n^λ, σ and n, that contains θ with an asymptotic probability of 95%.

10) Since σ is unknown, we estimate it as follows : $\tilde{\sigma}_n = \sqrt{\lambda^2 T_n^\lambda + 2(1-\lambda)^2(T_n^\lambda)^2}$ and we replace σ by $\tilde{\sigma}_n$ in I_n. Prove that the resulting interval $\tilde{I}_n$ is still 95%-confidence interval for θ. Compute $\tilde{I}_n$ for the following observations $\lambda = 0,5$, $n = 100$, $\bar{x}_n = 4,18$ et $v_n = 3,84$.

11*) Verifies that there exists a unique real number $\lambda^\star \in [0,1]$, depending on θ, that minimizes the length of I_n. Now, let us consider the random variable $\lambda_n^\star = \frac{2V_n}{1+2V_n}$. Prove that $(\lambda_n^\star)_{n\geq 2}$ converges (almost surely) towards $\lambda^\star$.

12*) Analyze the convergence (in distribution) of $(\sqrt{n}(T_n^{\lambda_n^\star} - \theta))_{n\geq 2}$. Then, deduce a 95%-confidence interval denoted $\tilde{I}_n^\star$ for θ. Compute $\tilde{I}_n^\star$ for the previous values of the observations.