TD n° 5

Exercice 1. *Gamma distribution*

A continuous random variable X is said to have a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$, shown as $X \sim Gamma(\alpha, \lambda)$, if its PDF is given by :

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

with $\Gamma(\alpha)$ denotes the gamma function, and is defined as :

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{for } \alpha > 0$$

1. What do you observe in case $\alpha = 1$?

2. Find the mean and the moment generating function of $X \sim Gamma(\alpha, \lambda)$

3. Show that when α is a positive integer, we have $\Gamma(\alpha) = (\alpha - 1)!$

4. Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables and $X_i \sim Exp(\lambda)$, and $X = \sum_{i=1}^n X_i$. What can you say about the distribution of X ?

Exercice 2. *Asymptotic level vs exact level of confidence*

A hardware company investigates again the mean lifetime $\theta > 0$ of a device, modelized by an exponential law (of parameter $1/\theta$). They have access to large samples of real lifetimes $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$ for their devices, and assume that those values are sampled from random variables $T_1, \ldots, T_n$ i.i.d. with exponential law of parameter $1/\theta$. They wish to estimate with intervals the parameter θ, with a confidence level of $\alpha = 95\%$. We will denote $\theta_n = \frac{1}{n}\sum_{i=1}^n t_i$ the empirical mean estimator of θ, and $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n T_i$.

1. Using Bienaymé-Chebychev inequality, provide a first way to compute a confidence interval for θ from $T_1, \ldots, T_n$, of exact level 95%.

Slutky's theorem states that, given $(X_n)_{n\in\mathbb{N}}$ a sequence of real random variables converging in law to a random variable X and (Y_n) a sequence of real random variables converging almost surely to a constant $c \in \mathbb{R}$, then (X_n, Y_n) converges in law to (X, c).

2. Show that $\sqrt{n}(\hat{\theta}_n - \theta)/\hat{\theta}_n \overset{\text{law}}{\to} \mathcal{N}(0,1)$. Deduce from this a way to compute a confidence interval for θ from $T_1, \ldots, T_n$, of asymptotic level 95%.

3. In Exercise 1, you have shown the sum of n i.i.d. exponential random variables of parameter λ follows the Gamma law $Gamma(n, \lambda)$. Explain why $n\hat{\theta}_n/\theta$ follows the Gamma law $\Gamma(n, 1)$. Deduce from this a way to construct different confidence intervals for θ from $T_1, \ldots, T_n$, of exact level 95%. Compare with confidence intervals produced in Question 1.

4. Reconsider the asymptotic approximation of Question 2 in light of Question 3 : for $n = 1, 5, 10, 100, 1000$, compute the exact probability α_n that θ falls in your confidence interval of Question 2. Does it depend on θ ? Discuss the use of your confidence intervals of asymptotic level 95%.

5. There exists another way to estimate the quality of your confidence intervals of asymptotic level : Berry-Esséen theorem provides the following bound where $C = 0,8$:

$$\beta_n = |\mathbb{P}\big(\theta \in \big[\hat{\theta}_n \pm \frac{a\theta}{\sqrt{n}}\big]\big) - \int_{-a}^{a} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx| \leq 2C \frac{\mathbb{E}|T_1 - \theta|^3}{\mathrm{Var}(T_1)^{3/2}\sqrt{n}}$$

Compute the value of β_n for $n = 1, 5, 10, 100$ and a the threshold used for the asymptotic confidence interval. Study the Berry-Esséen upper bound for $n = 1, 5, 10, 100$: does it depend on θ ? Draw a comparison between $|\alpha - 0,95|, \beta_n$ and this upper bound.

6. To conclude, discuss the pros and cons of the different approaches.