Sorbonne Université Année 2020/2021

L3 Deuxième semestre

Feuille de TD 3 : Intervalles de confiance

Exercice 1 : Dans un centre avicole, des études antérieures ont montré que le poids d'un oeuf choisi au hasard peut être considéré comme la réalisation d'une variable aléatoire gaussienne X, d'espérance μ et de variance σ^2. On admet que les poids des oeufs sont indépendants les uns des autres. On prend un échantillon de $n = 36$ oeufs que l'on pèse. Les mesures obtenues (exprimées en g) sont données (par ordre croissant) dans le tableau suivant :

50.34	52.62	53.79	54.99	55.82	57.67	51.41	53.13	53.89
55.04	55.91	57.99	51.51	53.28	54.63	55.12	55.95	58.10
52.07	53.30	54.76	55.24	57.05	59.30	52.22	53.32	54.78
55.28	57.18	60.58	52.38	53.39	54.93	55.56	57.31	63.15

Tab. 1: Mesure des Poids des oeufs (en g).

1. La figure ci-dessous présente l'histogramme en fréquences des poids des oeufs sur lequel on a superposé une version lissée de l'histogramme. Quelles conclusions peut-on tirer de cet histogramme sur la distribution des oeufs ?

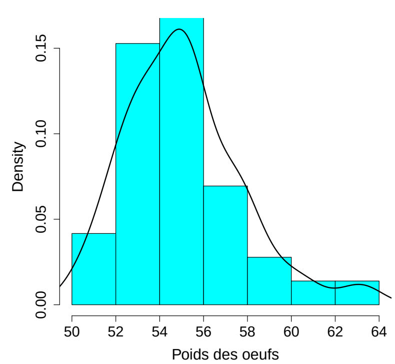

Fig. 1: Histogramme en fréquence des poids des oeufs.

Il faut faire du blabla, c'est chaint mais il faut qu'il le fasse de temps en temps.

2. Donner une estimation de la moyenne μ et de la variance σ^2. On les notera m et s^2.

Pour vous aider, on fournit les informations suivantes : si $x_1, x_2, \ldots, x_{36}$ désignent les poids mesurés, alors

$$\sum_{j=1}^{36} x_j = 1982,99 \quad \text{et} \quad \sum_{j=1}^{36} x_j^2 = 109481,1$$

On a

$$m = \frac{1}{36}\sum_{j=1}^{36} x_j = \frac{1982.99}{36} = 55.08$$

On a

$$s^2 = \frac{1}{35}\sum_{j=1}^{36}\left(x_j - m\right)^2 = \frac{1}{35}\left(\sum_{j=1}^{36} x_j^2 - 36m^2\right) = 7.54.$$

3. Construire un intervalle de confiance au niveau 95% pour le poids moyen des oeufs.

   On a

$$\sqrt{n}\frac{\overline{X}_n - \mu}{S} \sim T_{35}$$

   On note $t_{35,0.975}$ le quantile d'ordre 0.975 de la loi de Student à 35 degrés de liberté, et on obtient

$$\begin{aligned} 0.95 &= \mathbb{P}\left[-t_{35,0.975} \leq \sqrt{n}\frac{\overline{X}_n - \mu}{S} \leq t_{35,0.975}\right] \\ &= \mathbb{P}\left[-t_{35,0.975}\frac{S}{\sqrt{n}} \leq \overline{X}_n - \mu \leq t_{35,0.975}\frac{S}{\sqrt{n}}\right] \\ &= \mathbb{P}\left[\overline{X}_n - t_{35,0.975}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t_{35,0.975}\frac{S}{\sqrt{n}}\right] \end{aligned}$$

   De plus $t_{35,0.975} = 2.030$. On obtient donc l'intervalle de confiance

$$IC_{0.95} = \left[m - t_{35,0.975}\frac{s}{\sqrt{n}}; m + t_{35,0.975}\frac{s}{\sqrt{n}}\right] = [54.15; 56.01].$$

4. Construire un intervalle de confiance au niveau 95% pour la variance.

   On a

$$35\frac{S^2}{\sigma^2} \sim \chi^2_{35},$$

   et $k_{35,0.025} = 20.57$ et $k_{35,0.975} = 53.2$. On a donc

$$\begin{aligned} 0.95 &= \mathbb{P}\left[k_{35,0.025} \leq 35\frac{S^2}{\sigma^2} \leq k_{35,0.975}\right] \\ &= \mathbb{P}\left[\frac{k_{35,0.025}}{35S^2} \leq \frac{1}{\sigma^2} \leq \frac{k_{35,0.975}}{35S^2}\right] \\ &= \mathbb{P}\left[\frac{35S^2}{k_{35,0.975}} \leq \sigma^2 \leq \frac{35S^2}{k_{35,0.025}}\right] \end{aligned}$$

On obtient donc l'intervalle de confiance

$$IC_{0.95} = \left[\frac{35s^2}{k_{35,0.975}}; \frac{35S^2}{k_{35,0.025}}\right] = [4.96; 12.83].$$

5. A quel niveau de confiance correspondrait un intervalle centré en m et de demi-longueur 0.76 ?

   On a

$$\begin{aligned} 2t_{35,1-\alpha/2}\frac{s}{\sqrt{n}} = 0.76 &\Leftrightarrow t_{1-\alpha/2} = \frac{0.76\sqrt{n}}{2s} \\ &\Leftrightarrow 1-\alpha/2 = F\left(\frac{0.76\sqrt{n}}{2s}\right) \\ &\Leftrightarrow \alpha = 2\left(1-F\left(\frac{0.76\sqrt{n}}{2s}\right)\right) \\ &\Leftrightarrow \alpha = 2(1-F(0.42)) = 0.68 \end{aligned}$$

   où F est la fonction de répartition de la loi de Student à 35 degrés de liberté.

Exercice 2 : On considère une variable aléatoire X dont la loi dépend d'un paramètre inconnu $p > 0$ et telle que $\mathbb{E}\left[X^2\right] = \frac{1}{p^2}$ et $\mathbb{E}\left[X^4\right] = \frac{1}{p^6} + \frac{1}{p^4}$, et on pose $\theta = p^{-2}$.

1. Proposer un estimateur $\hat{\theta}_n$ de θ. Est-il consistant ? Fortement consistant ?

   On a $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i^2$. Par la loi forte des grands nombres, il converge presque sûrement vers θ.

2. Est-il sans biais ?

   On a

$$\mathbb{E}\left[\hat{\theta}_n\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[X_i^2\right] = \theta$$

   et l'estimateur est donc sans biais.

3. Donner son erreur quadratique moyenne.

   A noter que l'on a $\mathbb{V}\left[X^2\right] = \frac{1}{p^6} = \theta^3$. On a par indépendance des X_i (et donc des X_i^2)

$$\mathbb{V}\left[\hat{\theta}_n\right] = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\left[X_i^2\right] = \frac{\theta^3}{n}.$$

4. Donner sa normalité asymptotique.

   On a le TLC

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta^3\right).$$

5. En déduire un intervalle de confiance asymptotique de niveau $1-\alpha$ pour θ.

On a

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\theta^{3/2}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Comme on ne connait pas $\theta^{3/2}$ on s'intéresse à la variable

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n^{3/2}} = \frac{\theta^{3/2}}{\hat{\theta}_n^{3/2}}\sqrt{n}\frac{\hat{\theta}_n - \theta}{\theta^{3/2}}$$

Et comme $\hat{\theta}_n$ converge presque sûrement vers θ, par le théorème de continuité,

$$\frac{\theta^{3/2}}{\hat{\theta}_n^{3/2}} \xrightarrow[n\to\infty]{p.s} 1,$$

et on a en particulier la convergence en probabilité, et on obtient en appliquant le Théorème de Slutsky,

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n^{3/2}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Soit $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite, on a alors

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2} \leq \sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n^{3/2}} \leq q_{1-\alpha/2}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}} \leq \hat{\theta}_n - \theta \leq q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}}\right]
\end{aligned}$$

On obtient donc l'intervalle de confiance

$$IC_{1-\alpha} = \left[\theta_n - q_{1-\alpha/2}\frac{\theta_n^{3/2}}{\sqrt{n}}; \theta_n + q_{1-\alpha/2}\frac{\theta_n^{3/2}}{\sqrt{n}}\right]$$

6. En déduire un intervalle de confiance asymptotique de niveau $1-\alpha$ pour p.

   On a

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left[\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}} \leq \frac{1}{p^2} \leq \hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sqrt{\hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}}}} \leq p \leq \frac{1}{\sqrt{\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}}}}\right]
\end{aligned}$$

   Attention, l'inégalité précédente est vrai seulement parce que presque sûrement, à partir d'un certain rang, $\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n^{3/2}}{\sqrt{n}} > 0$. On obtient donc l'intervalle de confiance

$$\left[\frac{1}{\sqrt{\theta_n + q_{1-\alpha/2}\frac{\theta_n^{3/2}}{\sqrt{n}}}} ; \frac{1}{\sqrt{\theta_n - q_{1-\alpha/2}\frac{\theta_n^{3/2}}{\sqrt{n}}}}\right]$$

7. Proposer un estimateur $\hat{p}_n$ de p et montrer sa forte consistance.
   On propose $\hat{p}_n = \frac{1}{\sqrt{\hat{\theta}_n}}$. Comme $\hat{\theta}_n$ converge presque sûrement vers $p^{-2} > 0$, par le théorème de continuité, $\hat{p}_n$ converge presque sûrement vers p.
8. Donner sa normalité asymptotique
   On pose $g : x \longmapsto x^{-1/2}$. g est dérivable en p^{-2} et $g'\left(p^{-2}\right) = -\frac{1}{2}p^3$. Comme

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, p^{-6}\right)$$

   par la delta-méthode, on obtient

$$\sqrt{n}\left(\hat{p}_n - p\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4}\right)$$

9. En déduire un nouvel intervalle de confiance asymptotique de niveau $1-\alpha$ pour p.
   Le TLC précédent peut s'écrire

$$2\sqrt{n}\left(\hat{p}_n - p\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

   On note alors $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale, et on obtient

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2} \leq 2\sqrt{n}\left(\hat{p}_n - p\right) \leq q_{1-\alpha/2}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2}\frac{1}{2\sqrt{n}} \leq \hat{p}_n - p \leq q_{1-\alpha/2}\frac{1}{2\sqrt{n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\hat{p}_n - q_{1-\alpha/2}\frac{1}{2\sqrt{n}} \leq p \leq \hat{p}_n + q_{1-\alpha/2}\frac{1}{2\sqrt{n}}\right]
\end{aligned}$$

   On obtient donc l'intervalle de confiance

$$IC_{1-\alpha} = \left[p_n + q_{1-\alpha/2}\frac{1}{2\sqrt{n}} ; p_n + q_{1-\alpha/2}\frac{1}{2\sqrt{n}}\right].$$

Exercice 3 : Soit $\theta > 0$ et Y une variable aléatoire suivant une loi exponentielle de paramètre θ^{-1}.

Soit $X = Y + \theta$, sa densité f_θ définie pour tout x par

$$f_\theta(x) = C_\theta \exp\left(-\frac{x-\theta}{\theta}\right) \mathbf{1}_{[\theta,+\infty[}(x).$$

Dans ce qui suit, on considère $x_1, \ldots, x_n$ qui sont des réalisations des variables aléatoires indépendantes $X_1, \ldots, X_n$ de même loi que X.

1. Que vaut C_θ ?

   On a

   $$\int_\theta^\infty \exp\left(-\frac{x-\theta}{\theta}\right) = \left[-\theta \exp\left(-\frac{x-\theta}{\theta}\right)\right]_\theta^\infty = \theta,$$

   et donc $C_\theta = \theta^{-1}$.

2. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.

   On peut soit calculer comme des bourrins, soit voir que $\mathbb{E}[Y] = \theta$ et donc $\mathbb{E}[X] = \mathbb{E}[Y] + \theta = 2\theta$. De la même façon, on a $\mathbb{V}[Y] = \mathbb{V}[X] = \theta^2$.

3. Par la méthode des moments, en déduire un estimateur de θ.

   On obtient l'estimateur $\hat{\theta}_n = \frac{\overline{X}_n}{2}$.

4. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?

   Comme $\overline{X}_n$ converge presque sûrement vers 2θ (LFGN), par le théorème de continuité on obtient la convergence presque sûre de $\hat{\theta}_n$ vers θ. De plus le TLC nous donne

   $$\sqrt{n}\left(\overline{X}_n - 2\theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} Z \sim \mathcal{N}\left(0, \theta^2\right).$$

   Par le théorème de continuité, on obtient donc

   $$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \frac{1}{2} Z \sim \mathcal{N}\left(0, \frac{\theta^2}{4}\right).$$

5. Soit $\alpha \in (0,1)$, construire un intervalle de confiance asymptotique de niveau $1-\alpha$ du paramètre θ.

   Version 1 : D'après la question précédente, on a en particulier

   $$2\sqrt{n}\frac{\hat{\theta}_n - \theta}{\theta} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

   On va donc en particulier s'intéresser à la variable aléatoire

   $$2\sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} = \frac{\theta}{\hat{\theta}_n} 2\sqrt{n}\frac{\hat{\theta}_n - \theta}{\theta}$$

   Or par le théorème de continuité, on a $\frac{\theta}{\hat{\theta}}$ qui converge en probabilité vers 1, et donc, en appliquant le Théorème de Slutsky,

$$2\sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0,1\right).$$

On a donc, si on note $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite,

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2} \le 2\sqrt{n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} \le q_{1-\alpha/2}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2}\frac{2\hat{\theta}_n}{2\sqrt{n}} \le \hat{\theta}_n - \theta \le q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[-\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{n}} \le -\theta \le -\hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{n}} \ge \theta \ge \hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{n}}\right].
\end{aligned}$$

On obtient donc l'intervalle de confiance asymptotique :

$$IC_{1-\alpha} = \left[\theta_n - q_{1-\alpha/2}\frac{\theta_n}{2\sqrt{n}}; \theta_n + q_{1-\alpha/2}\frac{\theta_n}{2\sqrt{n}}\right]$$

Version 2 : Le TLC précédent peut s'écrire

$$2\sqrt{n}\frac{\hat{\theta}_n - \theta}{\theta} = 2\sqrt{n}\left(\frac{\hat{\theta}_n}{\theta} - 1\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

On note alors $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite, et on obtient

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2} \le 2\sqrt{n}\left(\frac{\hat{\theta}_n}{\theta} - 1\right) \le q_{1-\alpha/2}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[1 - \frac{q_{1-\alpha/2}}{2\sqrt{n}} \le \frac{\hat{\theta}_n}{\theta} \le 1 + \frac{q_{1-\alpha/2}}{2\sqrt{n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\frac{1 - \frac{q_{1-\alpha/2}}{2\sqrt{n}}}{\hat{\theta}_n} \le \frac{1}{\theta} \le \frac{1 + \frac{q_{1-\alpha/2}}{2\sqrt{n}}}{\hat{\theta}_n}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\frac{\hat{\theta}_n}{1 + \frac{q_{1-\alpha/2}}{2\sqrt{n}}} \le \theta \le \frac{\hat{\theta}_n}{1 - \frac{q_{1-\alpha/2}}{2\sqrt{n}}}\right]
\end{aligned}$$

La dernière inégalité est vraie car à partir d'un certaine rang, $1 - \frac{q_{1-\alpha/2}}{\sqrt{n}} > 0$. On obtient donc l'intervalle de confiance

$$IC_{1-\alpha} = \left[\frac{\theta_n}{1 + \frac{q_{1-\alpha/2}}{2\sqrt{n}}} ; \frac{\theta_n}{1 - \frac{q_{1-\alpha/2}}{2\sqrt{n}}} \right]$$

6. Donner l'estimateur du maximum de vraisemblance.

   On a la vraisemblance qui est définie par

$$L_n(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left(-\frac{X_i - \theta}{\theta} \right) \mathbf{1}_{[\theta, +\infty[}(X_i) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left(-\frac{X_i - \theta}{\theta} \right) \mathbf{1}_{]0, x_i]}(\theta)$$

   et la vraisemblance est donc nulle si il existe i tel que $\theta > X_i$ et donc non nulle si $0 < \theta \leq X_i$ pour tout i, i.e $0 < \theta \leq \min_i X_i = X_{(1)}$. Or pour tout i, la fonction $g : \theta \longmapsto \frac{1}{\theta} \exp\left(-\frac{X_i - \theta}{\theta} \right)$ est strictement croissante sur $]0, X_1]$. En effet, on a

$$g'(\theta) = \left(-\frac{1}{\theta^2} + \frac{X_i}{\theta^3} \right) \exp\left(-\frac{X_i - \theta}{\theta} \right) = (-\theta + X_i) \frac{1}{\theta^3} \exp\left(-\frac{X_i - \theta}{\theta} \right) > 0.$$

   On obtient donc que la vraisemblance est croissante sur $]0, X_{(1)}]$. On obtient donc un unique maximum qui est $\theta = X_{(1)}$ et on obtient ainsi l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{MV} = X_{(1)}$.

7. On considère maintenant $X_{(1)} = \min_{i=1,\ldots,n}(X_i)$. Donner la fonction de répartition de X. En déduire celle de $X_{(1)}$.

   Pour tout $t \geq \theta$, on peut soit calculer $\mathbb{P}[X \leq t]$ comme des bourrins, soit passer par le fait que pour tout $x \geq 0$, $\mathbb{P}[Y \leq x] = 1 - \exp\left(-\frac{x}{\theta}\right)$. On a alors pour tout $t \geq \theta$

$$\mathbb{P}[X \leq t] = \mathbb{P}[Y + \theta \leq t] = \mathbb{P}[Y \leq t - \theta] = 1 - \exp\left(-\frac{t - \theta}{\theta} \right).$$

   De plus, pour tout $t \geq \theta$, on a

$$\mathbb{P}\left[X_{(1)} \geq t \right] = \mathbb{P}\left[\forall i, X_{(i)} \geq t \right]$$

   Par indépendance des X_i, on obtient

$$\begin{aligned} \mathbb{P}\left[X_{(1)} \geq t \right] &= \prod_{i=1}^{n} \mathbb{P}\left[X_i \geq t \right] \\ &= \left(\exp\left(-\frac{t - \theta}{\theta} \right) \right)^n \\ &= \exp\left(-n \frac{t - \theta}{\theta} \right). \end{aligned}$$

Et donc, en particulier,

$$F_{X_{(1)}}(t) = 1 - \exp\left(-n\frac{t-\theta}{\theta}\right).$$

8. Donner la loi de $n\left(X_{(1)} - \theta\right)$.

   On a pour tout $t \geq 0$,

$$\mathbb{P}\left[n\left(X_{(1)} - \theta\right) \leq t\right) = \mathbb{P}\left[X_{(1)} \leq \theta + \frac{t}{n}\right] = 1 - \exp\left(-\frac{t}{\theta}\right),$$

   et donc la variable aléatoire $n\left(X_{(1)} - \theta\right)$ suit une loi exponentielle de paramètre θ^{-1}.

9. En déduire la convergence en probabilité de $X_{(1)}$.

   On a pour tout $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left[\left|X_{(1)} - \theta\right| \geq \epsilon\right] &= \mathbb{P}\left[X_{(1)} - \theta \geq \epsilon\right] \\
&= \mathbb{P}\left[n\left(X_{(1)} - \theta\right) \geq n\epsilon\right] \\
&= \exp\left(-n\frac{\epsilon}{\theta}\right) \\
&\rightarrow_{n\rightarrow\infty} 0.
\end{aligned}$$

10. Montrer que l'estimateur $X_{(1)}$ est biaisé mais asymptotiquement sans biais.

    On peut facilement vérifier que $X_{(1)} - \theta$ qui suit une loi exponentielle de paramètre $\frac{\theta}{n}$, et on obtient donc $\mathbb{E}\left[\left(X_{(1)} - \theta\right)\right] = \frac{\theta}{n}$, et on obtient en particulier

$$\mathbb{E}\left[X_{(1)}\right] = \frac{n+1}{n}\theta \xrightarrow[n\rightarrow\infty]{} 0.$$

11. En déduire la convergence en moyenne quadratique de $X_{(1)}$.

    On a, comme $X_{(1)} - \theta$ suit une loi exponentielle de paramètre $\frac{n}{\theta}$,

$$\begin{aligned}
\mathbb{E}\left[\left(X_{(1)} - \theta\right)^2\right] &= \mathbb{V}\left[X_{(1)} - \theta\right] + \left(\mathbb{E}\left[X_{(1)} - \theta\right]\right)^2 \\
&= \frac{\theta^2}{n^2} + \frac{\theta^2}{n^2} \\
&= 2\frac{\theta^2}{n^2}.
\end{aligned}$$

12. Donner un nouvel intervalle de confiance pour le paramètre θ.

    Comme $\theta \leq X_{(1)}$, on va chercher à construire un intervalle de confiance de la forme

$$IC_{1-\alpha} = \left[f\left(x_1, ..., x_n\right); x_{(1)}\right]$$

et donc en particulier, on va chercher $c_\alpha > 0$ tel que

$$\mathbb{P}\left[X_{(1)} - c_\alpha \leq \theta \leq X_{(1)}\right] = \mathbb{P}\left[X_{(1)} - c_\alpha \leq \theta\right] = 1 - \alpha.$$

Comme on connait la loi de $n\left(X_{(1)} - \theta\right)$, on va le faire apparaitre. On a alors

$$\begin{aligned}
1 - \alpha = \mathbb{P}\left[X_{(1)} - c_\alpha \leq \theta\right] &= \mathbb{P}\left[X_{(1)} - \theta - c_\alpha \leq 0\right] \\
&= \mathbb{P}\left[X_{(1)} - \theta \leq c_\alpha\right] \\
&= \mathbb{P}\left[n\left(X_{(1)} - \theta\right) \leq nc_\alpha\right] \\
&= F_Y\left(nc_\alpha\right)
\end{aligned}$$

où F_Y est la fonction de répartition de Y, i.e de la loi exponentielle de paramètre θ^{-1}. On obtient alors

$$c_\alpha = \frac{F_Y^{-1}(1-\alpha)}{n}$$

Il reste donc à calculer $F_Y^{-1}(1-\alpha)$. Pour tout $\alpha \in (0,1)$,

$$\begin{aligned}
F_Y(x) = 1 - \alpha &\Leftrightarrow 1 - \alpha = 1 - \exp\left(-\frac{x}{\theta}\right) \\
&\Leftrightarrow \alpha = \exp\left(-\frac{x}{\theta}\right) \\
&\Leftrightarrow \ln(\alpha) = -\frac{x}{\theta} \\
&\Leftrightarrow x = -\theta \ln(\alpha).
\end{aligned}$$

On obtient donc $c_\alpha = -\frac{\theta \ln(\alpha)}{n}$, et en particulier

$$\mathbb{P}\left[X_{(1)} - \frac{\theta \ln(\alpha)}{n} \leq \theta\right] = 1 - \alpha.$$

De plus, on a

$$\begin{aligned}
1 - \alpha &= \mathbb{P}\left[X_{(1)} + \frac{\theta \ln(\alpha)}{n} \leq \theta \leq X_{(1)}\right] \\
&= \mathbb{P}\left[X_{(1)} \leq \theta\left(1 - \frac{\ln(\alpha)}{n}\right)\right] \\
&= \mathbb{P}\left[X_{(1)}\left(1 - \frac{\ln(\alpha)}{n}\right)^{-1} \leq \theta\right]
\end{aligned}$$

On obtient donc l'intervalle de confiance

$$IC_{1-\alpha}(\theta) = \left[x_{(1)}\left(1 - \frac{\ln(\alpha)}{n}\right)^{-1} ; x_{(1)}\right].$$

13. Commenter.

    Pour construire des intervalles de confiances, on prendra le dernier car non seulement le deuxième estimateur est plus rapide que le premier, et on obtient donc normalement des intervalles plus précis, mais l'intervalle de confiance obtenue est non asymptotique.

Exercice 4 : Soit X une variable aléatoire de densité $f_\theta(x) = \frac{2}{\theta^2} x \mathbf{1}_{[0,\theta]}(x)$, avec $\theta > 0$. On considère dans ce qui suit des réalisations $x_1, \ldots, x_n$ des variables aléatoires indépendantes $X_1, \ldots, X_n$ de même loi que X.

1. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.

   On a
   $$\mathbb{E}[X] = \int_0^\theta \frac{2x^2}{\theta} dx = \frac{2}{3\theta^2} \left[x^3\right]_0^\theta = \frac{2}{3}\theta.$$
   De la même façon
   $$\mathbb{E}\left[X^2\right] = \frac{2}{\theta^2} \int_0^\theta x^3 dx = \frac{1}{2\theta} \left[0, x^4\right]_0^\theta = \frac{1}{2}\theta^2.$$
   et donc $\mathbb{V}[X] = \frac{1}{18}\theta^2$.
2. Par la méthode des moments, en déduire un estimateur $\hat{\theta}_n$ de θ.

   On obtient donc, comme $\mathbb{E}[X] = (2/3)\theta$ l'estimateur $\hat{\theta}_n = \frac{3}{2}\overline{X}_n$.
3. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?

   Comme $\overline{X}_n$ converge presque sûrement vers $(2/3)\theta$ (LFGN), par le théorème de continuité, on obtient la forte consistance de $\hat{\theta}_n$. De plus, on a le TLC
   $$\sqrt{n}\left(\overline{X}_n - \frac{2}{3}\theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} Z \sim \mathcal{N}\left(0, \frac{1}{18}\theta^2\right).$$
   Par le théorème de continuité, on obtient donc
   $$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \frac{3}{2}Z \sim \mathcal{N}\left(0, \frac{1}{8}\theta^2\right).$$
4. Soit $\alpha \in (0,1)$, construire un intervalle de confiance asymptotique de niveau $1-\alpha$ du paramètre θ.

   1ere version : Le TLC précédent peut s'écrire
   $$2\sqrt{2n}\frac{\hat{\theta}_n - \theta}{\theta} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$
   et comme on ne connaît pas θ, on va en particulier s'intéresser à la variable aléatoire
   $$2\sqrt{2n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} = \frac{\theta}{\hat{\theta}_n} 2\sqrt{2n}\frac{\hat{\theta}_n - \theta}{\theta}.$$
   A l'aide du théorème de continuité, on a $\frac{\theta}{\hat{\theta}_n}$ qui converge en probabilité vers 1, et donc, en

appliquant le Théorème de Slutsky, on obtient

$$2\sqrt{2n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Et donc, si on note $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite, on a

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2} \leq 2\sqrt{2n}\frac{\hat{\theta}_n - \theta}{\hat{\theta}_n} \leq q_{1-\alpha/2}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}} \leq \hat{\theta}_n - \theta \leq q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[-\hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}} \leq -\theta \leq -\hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\hat{\theta}_n + q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}} \geq \theta \geq \hat{\theta}_n - q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}}\right]
\end{aligned}$$

et on obtient ainsi l'intervalle de confiance asymptotique

$$IC_{1-\alpha}(\theta) = \left[\theta_n - q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}}; \theta_n + q_{1-\alpha/2}\frac{\hat{\theta}_n}{2\sqrt{2n}}\right]$$

2eme version : Le TLC de la question précédente peut s'écrire

$$2\sqrt{2n}\left(\frac{\hat{\theta}_n}{\theta} - 1\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

On obtient donc

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty} \mathbb{P}\left(-q_{1-\alpha/2} \leq 2\sqrt{2n}\left(\frac{\hat{\theta}_n}{\theta} - 1\right) \leq q_{1-\alpha/2}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[-\frac{q_{1-\alpha/2}}{2\sqrt{2n}} \leq \frac{\hat{\theta}_n}{\theta} - 1 \leq \frac{q_{1-\alpha/2}}{2\sqrt{2n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[1 - \frac{q_{1-\alpha/2}}{2\sqrt{2n}} \leq \frac{\hat{\theta}_n}{\theta} \leq 1 + \frac{q_{1-\alpha/2}}{2\sqrt{2n}}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\hat{\theta}_n} - \frac{q_{1-\alpha/2}}{2\sqrt{2n}\hat{\theta}_n} \leq \frac{1}{\theta} \leq \frac{1}{\hat{\theta}_n} - \frac{q_{1-\alpha/2}}{2\sqrt{2n}\hat{\theta}_n}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\hat{\theta}_n\left(1 - \frac{q_{1-\alpha/2}}{2\sqrt{2n}}\right)^{-1} \geq \theta \geq \hat{\theta}_n\left(1 + \frac{q_{1-\alpha/2}}{2\sqrt{2n}}\right)^{-1}\right].
\end{aligned}$$

On obtient donc l'intervalle de confiance asymptotique

$$IC_{1-\alpha}(\theta) = \left[\theta_n \left(1 + \frac{q_{1-\alpha/2}}{2\sqrt{2n}}\right)^{-1} ; \theta_n \left(1 - \frac{q_{1-\alpha/2}}{2\sqrt{2n}}\right)^{-1}\right]$$

5. Donner l'estimateur du maximum de vraisemblance.

   On a la vraisemblance qui est donnée par

$$L_n(\theta) = \prod_{i=1}^{n} \frac{2}{\theta^2} X_i \mathbf{1}_{[0,\theta]}(X_i) = \prod_{i=1}^{n} \frac{2}{\theta^2} X_i \mathbf{1}_{[x_i,+\infty[}(\theta).$$

   La vraisemblance est donc nulle si il existe i tel que $\theta < X_i$ et elle est donc non nulle si pour tout i, $\theta \leq X_i$, i.e si $\theta \leq X_{(n)} = \max_i X_i$. De plus, la vraisemblance est strictement décroissante sur sur $[X_{(n)}, +\infty[$ et le maximum est donc atteint en $X_{(n)}$. On obtient donc l'EMV $\hat{\theta}_n^{MV} = \max_i X_{(i)} = X_{(n)}$.

6. On considère maintenant $X_{(n)} = \max_{i=1,\ldots,n} X_i$. Donner la fonction de répartition de X.

   Pour tout $t \in [0, \theta]$, on a

$$F_X(t) = \int_0^t \frac{2}{\theta^2} x dx = \frac{1}{\theta^2}\left[x^2\right]_0^t = \frac{t^2}{\theta^2}$$

7. En déduire la fonction de répartition de $X_{(n)}$.

   Pour tout $t \in [0, \theta]$, on a

$$\mathbb{P}\left[X_{(n)} \leq t\right] = \mathbb{P}\left[\forall i, X_i \leq t\right]$$

   Par indépendance des X_i, on obtient donc

$$\mathbb{P}\left[X_{(n)} \leq t\right] = \prod_{i=1}^{n} \mathbb{P}\left[X_i \leq t\right] = \left(\frac{t}{\theta}\right)^{2n}.$$

8. En déduire la densité de $X_{(n)}$.

   On obtient donc pour tout x,

$$f_{X_{(n)}}(x) = \frac{2n}{\theta}\left(\frac{t}{\theta}\right)^{2n-1} \mathbf{1}_{[0,\theta]}(x).$$

9. L'estimateur $X_{(n)}$ est-il sans biais ?

   On a,

$$\mathbb{E}\left[X_{(n)}\right] = \int_0^\theta \frac{2n}{\theta} x \left(\frac{x}{\theta}\right)^{2n-1} dx = \int_0^\theta 2n \left(\frac{x}{\theta}\right)^{2n} dx = \frac{2n}{2n+1}\theta$$

   L'estimateur est donc biaisé mais asymptotiquement sans biais.

10. Calculer le risque quadratique $\mathbb{E}\left[\left(X_{(n)}-\theta\right)^2\right]$.

Pour calculer le risque quadratique, on va utiliser la décomposition suivante,

$$\mathbb{E}\left[\left(X_{(n)}-\theta\right)^2\right] = \mathbb{E}\left[X_{(n)}^2\right] - 2\theta\mathbb{E}\left[X_{(n)}\right] + \theta^2$$

et on va donc devoir calculer le moment d'ordre 2 de $X_{(n)}$. On a,

$$\mathbb{E}\left[X_{(n)}^2\right] = \int_0^\theta \frac{2n}{\theta}x^2\left(\frac{x}{\theta}\right)^{2n-1}dx = \theta\int_0^\theta 2n\left(\frac{x}{\theta}\right)^{2n+1}dx = \frac{2n}{2n+2}\theta^2.$$

On obtient donc

$$\begin{aligned}\mathbb{E}\left[\left(X_{(n)}-\theta\right)^2\right] &= \frac{n}{n+1}\theta^2 - 2\frac{2n}{2n+1}\theta^2 + \theta^2 \\ &= \frac{2n^2+n-4n^2-4n+2n^2+3n+1}{(n+1)(2n+1)}\theta^2 \\ &= \frac{1}{(n+1)(2n+1)}\theta^2\end{aligned}$$

11. Donner la convergence en loi de $n\left(\theta - X_{(n)}\right)$.

Pour tout $t \geq 0$, on a

$$\begin{aligned}\mathbb{P}\left[n\left(\theta - X_{(n)}\right) \leq t\right] &= \mathbb{P}\left[\theta - X_{(n)} \leq \frac{t}{n}\right] \\ &= \mathbb{P}\left[X_{(n)} \geq \theta - \frac{t}{n}\right] \\ &= 1 - \left(1 - \frac{t}{n\theta}\right)^{2n} \\ &= 1 - \exp\left(2n\ln\left(1 - \frac{t}{n\theta}\right)\right) \\ &\sim_{n\to\infty} 1 - \exp\left(-2n\frac{t}{n\theta}\right) \\ &= 1 - \exp\left(-\frac{2t}{\theta}\right)\end{aligned}$$

et $n\left(\theta - X_{(n)}\right)$ converge donc vers une loi exponentielle de paramètre $2\theta^{-1}$.

12. Donner un nouvel intervalle de confiance au niveau $1-\alpha$ de θ.

On pourrait utiliser la convergence en loi de la question précédente pour construire un intervalle de confiance asymptotique. Cependant, comme on connait la loi de $X_{(n)}$, on va essayer de construire un intervalle non asymptotique. Comme $\theta \geq X_{(n)}$, on va chercher un intervalle de la forme

$$\left[x_{(n)}; x_{(n)} + c_\alpha\right]$$

et plus précisément, on va chercher c_α tel que

$$\mathbb{P}\left[X_{(n)} + c_\alpha \geq \theta\right] = 1 - \alpha$$

On a

$$\begin{aligned}
\mathbb{P}\left[X_{(n)} + c_\alpha \geq \theta\right] = 1 - \alpha &\Leftrightarrow \mathbb{P}\left[\theta - X_{(n)} \leq c_\alpha\right] = 1 - \alpha \\
&\Leftrightarrow \mathbb{P}\left[n\left(\theta - X_{(n)}\right) \leq nc_\alpha\right] = 1 - \alpha \\
&\Leftrightarrow 1 - \left(1 - \frac{c_\alpha}{\theta}\right)^{2n} = 1 - \alpha \\
&\Leftrightarrow \left(1 - \frac{c_\alpha}{\theta}\right)^{2n} = \alpha \\
&\Leftrightarrow \ln\left(1 - \frac{c_\alpha}{\theta}\right) = \frac{\ln(\alpha)}{2n} \\
&\Leftrightarrow \frac{c_\alpha}{\theta} = 1 - \exp\left(\frac{\ln(\alpha)}{2n}\right) \\
&\Leftrightarrow c_\alpha = \theta\left(1 - \exp\left(\frac{\ln(\alpha)}{2n}\right)\right)
\end{aligned}$$

A noter que $1 - \exp\left(\frac{\ln(\alpha)}{2n}\right) \sim \frac{\ln(\alpha)}{2n}$. On obtient donc

$$\begin{aligned}
1 - \alpha &= \mathbb{P}\left[X_{(n)} + \theta\left(1 - \exp\left(\frac{\ln(\alpha)}{2n}\right)\right) \geq \theta\right] \\
&= \mathbb{P}\left[X_{(n)} \geq \theta \exp\left(\frac{\ln(\alpha)}{2n}\right)\right] \\
&= \mathbb{P}\left[X_{(n)} \exp\left(-\frac{\ln(\alpha)}{2n}\right) \geq \theta\right]
\end{aligned}$$

On obtient donc l'intervalle de confiance non asymptotique suivant :

$$IC_{1-\alpha}(\theta) = \left[x_{(n)}; x_{(n)} \exp\left(-\frac{\ln(\alpha)}{2n}\right)\right].$$

13. Quel intervalle de confiance choisiriez vous ?

    Les deux premiers intervalles sont asymptotique, et on ne peut aucunement assurer à quel point la loi de la variable aléatoire utilisée est proche de celle d'une loi normale. Comme le dernier intervalle est non-asymptotique, il est plus "fiable" mais il est aussi plus précis asymptotiquement.

Exercice 5 : Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes suivant une loi de Poisson de paramètre $\theta > 0$.

1. Donner $\mathbb{E}\left[X_1\right]$ et $\mathbb{V}\left[X_1\right]$ et en déduire un estimateur $\hat{\theta}_n$ de θ.

On a $\mathbb{E}[X_1] = \mathbb{V}[X_1] = \theta$. On obtient donc l'estimateur $\hat{\theta}_n = \overline{X}_n$ qui fortmenent consistant (LFGN).

2. Donner la normalité asymptotique de l'estimateur $\hat{\theta}_n$. On a

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\theta).$$

3. Par la méthode du plug-in, donner un intervalle de confiance asymptotique à 95% pour θ. D'après la question précédente, on a

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sqrt{\theta}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Comme on ne connait pas $\sqrt{\theta}$, on va s'intéresser à la variable

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sqrt{\hat{\theta}_n}} = \frac{\sqrt{\theta}}{\sqrt{\hat{\theta}_n}}\frac{\hat{\theta}_n - \theta}{\sqrt{\theta}}.$$

Comme l'estimateur $\hat{\theta}_n$ est fortement consistant, par le théorème de continuité, on a

$$\frac{\sqrt{\theta}}{\sqrt{\hat{\theta}_n}} \xrightarrow[n\to\infty]{p.s} 1$$

et on obtient en particulier la convergence en probabilité. En appliquant le théorème de Slutsky, on obtient

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sqrt{\hat{\theta}_n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

De plus, si on note $q_{1-\alpha}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite, on obtient

$$\begin{aligned}
1-\alpha &= \lim_{n\to\infty}\mathbb{P}\left[-q_{1-\alpha/2} \leq \sqrt{n}\frac{\hat{\theta}_n - \theta}{\sqrt{\hat{\theta}_n}} \leq q_{1-\alpha/2}\right] \\
&= \lim_{n\to\infty}\mathbb{P}\left[-q_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_n}}{\sqrt{n}} \leq \hat{\theta}_n - \theta \leq q_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_n}}{\sqrt{n}}\right] \\
&= \lim_{n\to\infty}\mathbb{P}\left[\hat{\theta}_n - q_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_n}}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n + q_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_n}}{\sqrt{n}}\right].
\end{aligned}$$

4. Trouver une fonction $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ telle que

$$\sqrt{n}\left(g\left(\hat{\theta}_n\right)\right) - g(\theta)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

On a, par la delta-méthode, pour toute fonction g dérivable en θ,

$$\sqrt{n}\left(g\left(\hat{\theta}_n\right)-g(\theta)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0,\left(g'(\theta)\right)^2\theta\right)$$

et on cherche donc g tel que $g'(\theta)=\frac{1}{\sqrt{\theta}}$. On choisit donc $g : x \longmapsto 2\sqrt{x}$ qui est bien dérivable en $\theta>0$. On obtient alors

$$\sqrt{n}\left(2\sqrt{\hat{\theta}_n}-2\sqrt{\theta}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

5. En déduire un nouvel intervalle de confiance pour θ.

   Soit $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite. On a

$$\begin{aligned}1-\alpha &= \lim_{n\to\infty}\mathbb{P}\left[-q_{1-\alpha/2}\leq 2\sqrt{n}\left(\sqrt{\hat{\theta}_n}-\sqrt{\theta}\right)\leq q_{1-\alpha/2}\right]\\ &= \lim_{n\to\infty}\mathbb{P}\left[\sqrt{\hat{\theta}_n}-\frac{q_{1-\alpha/2}}{2\sqrt{n}}\leq\sqrt{\theta}\leq\sqrt{\hat{\theta}_n}+\frac{q_{1-\alpha/2}}{2\sqrt{n}}\right].\end{aligned}$$

   Comme pour n suffisamment grand on a $\sqrt{\hat{\theta}_n}-\frac{q_{1-\alpha/2}}{2\sqrt{n}}\geq 0$ presque sûrement, on obtient

$$\begin{aligned}1-\alpha &= \lim_{n\to\infty}\mathbb{P}\left[\left(\sqrt{\hat{\theta}_n}-\frac{q_{1-\alpha/2}}{2\sqrt{n}}\right)^2\leq\theta\leq\left(\sqrt{\hat{\theta}_n}+\frac{q_{1-\alpha/2}}{2\sqrt{n}}\right)^2\right]\\ &= \lim_{n\to\infty}\mathbb{P}\left[\hat{\theta}_n-q_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_n}}{\sqrt{n}}+\frac{q_{1-\alpha/2}^2}{4n}\leq\theta\leq\hat{\theta}_n+q_{1-\alpha/2}\frac{\sqrt{\hat{\theta}_n}}{\sqrt{n}}+\frac{q_{1-\alpha/2}^2}{4n}\right]\end{aligned}$$

6. Vérifier que les deux résultats sont équivalents, i.e si on note $[a_n,b_n]$ et $[a_n',b_n']$ les deux intervalles obtenus, on a

$$a_n\sim_{n\to\infty}a_n' \qquad \text{et} \qquad b_n\sim_{n\to\infty}b_n'.$$

   On a

$$\frac{a_n'}{a_n}=1+\frac{q_{1-\alpha/2}^2}{4n\left(\hat{\theta}_n-q_{1-\alpha/2}\sqrt{\hat{\theta}_n}/\sqrt{n}\right)}\xrightarrow[n\to\infty]{}1.$$

Exercice 6 : Loi de Laplace translatée Soit Y une variable aléatoire suivant une loi de Laplace, i.e de densité f_Y définie pour tout $x\in\mathbb{R}$ par

$$f_Y=\frac{1}{2}\exp\left(-|x|\right).$$

Soit $\theta>0$ et X une variable aléatoire suivant une loi de Laplace translatée, i.e de densité f_X définie pour tout $x\in\mathbb{R}$ par

$$f_X=\frac{1}{2}\exp\left(-|\theta-x|\right).$$

Dans ce qui suit on considère des variables aléatoires indépendantes $X_1, \ldots, X_n$ de même loi que X.

1. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.

   On a $\mathbb{E}[Y] = 0$ et $\mathbb{V}[Y] = 1$ et donc $\mathbb{E}[X] = \theta$ et $\mathbb{V}[X] = 1$. S'ils ne connaissent pas l'espérance et la variance d'une loi de Laplace, qu'ils fassent des IPPs, ça les changera du ski...

2. Par la méthode des moments, donner un estimateur de θ. Est-il consistant ? Asymptotiquement normal ? Que pouvez vous en conclure ?

   On a l'estimateur $\overline{X}_n$ qui est consistant par la LGN et asymptotiquement efficace...

3. Construire un intervalle de confiance asymptotique de niveau 0.90 pour θ. On a

$$IC_{1-\alpha}(\theta) = \left[\overline{x}_n \pm 1.645 \frac{1}{\sqrt{n}} \right].$$

4. Que vaut l'estimateur du maximum de vraisemblance ?

   On a

$$L_n(\theta) = \prod_{i=1}^{n} \frac{1}{2} \exp\left(-|X_i - \theta|\right) = \frac{1}{2^n} \exp\left(-\sum_{i=1}^{n} |X_i - \theta|\right)$$

   et donc

$$l_n(\theta) = -n \ln(2) - \sum_{i=1}^{n} |X_i - \theta|$$

   et

$$l_n'(\theta) = \sum_{i=1}^{n} \frac{X_i - \theta}{|X_i - \theta|}$$

   On tombe sur un cas où on ne sait pas calculer l'estimateur du maximum de vraisemblance. En réalité, il s'agit de prendre la médiane de $X_1, \ldots, X_n$ que l'on ne sait pas calculer explicitement. Par contre, on peut l'approcher via l'algorithme de Weiszfeld (que l'on peut voir comme un algo de point fixe ou comme un algo de gradient déterministe)

$$m_{t+1} = \frac{\sum_{i=1}^{n} \frac{x_i}{|x_i - m_t|}}{\sum_{i=1}^{n} \frac{1}{|x_i - m_t|}}$$

   où via un algorithme de gradient stochastique... En fait, un minimiseur est la médiane empirique mais ils verront ça l'année prochaine.

5. On note $\hat{m}_n$ l'estimateur du maximum de vraisemblance, on suppose qu'il existe et est consistant. Que pouvez-vous en conclure ?

   Il est asymptotiquement efficace (dernier théorème du cours), et donc

$$\sqrt{n}\left(\hat{m}_n - \theta\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Exercice 7 : Débiaiser ou ne pas débiaiser On considère une variable aléatoire $X \sim \mathcal{U}([0,\theta])$ avec $\theta > 0$et on considdère l'estimateur $\hat{\theta}_n = X_{(n)}$.

1. Est-il biaisé ? Calculer l'erreur quadratique moyenne.

   On a pour tout $x \in [0,\theta]$,
   $$\mathbb{P}\left[X_{(n)} \leq x\right] = \left(\frac{x}{\theta}\right)^n$$
   et on a donc
   $$f_{X_{(n)}}(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}\mathbf{1}_{[0,\theta]}(x)$$
   On obtient donc
   $$\mathbb{E}\left[X_{(n)}\right] = \int_0^\theta x\frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}dx = \frac{n}{n+1}\theta$$
   et l'estimateur est donc biaisé. De plus, on a
   $$\mathbb{E}\left[X_{(n)}^2\right] = \int_0^\theta x^2\frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}dx = n\theta\int_0^\theta \left(\frac{x}{\theta}\right)^{n+1}dx = \frac{n}{n+2}\theta^2$$
   On obtient donc
   $$\begin{aligned}\mathbb{V}\left[\hat{\theta}_n\right] &= \mathbb{E}\left[\hat{\theta}_n^2\right] - \left(\mathbb{E}\left[\hat{\theta}_n\right]\right)^2 \\ &= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 \\ &= \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\theta^2 \\ &= \frac{n}{(n+2)(n+1)^2}\theta^2.\end{aligned}$$
   On a alors,
   $$\mathrm{EQM}\left(\hat{\theta}_n,\theta\right) = \mathbb{V}\left[\hat{\theta}_n\right] + B\left(\hat{\theta}_n,\theta\right)^2 = \frac{n}{(n+2)(n+1)^2}\theta^2 + \frac{n^2}{(n+1)^2}\theta^2 = \frac{2\theta^2}{(n+1)(n+2)}$$

2. Proposer un estimateur non biaisé et calculer son erreur quadratique moyenne. On considère maintenant l'estimateur non biaisé $\tilde{\theta}_n$ défini par
   $$\tilde{\theta}_n = \frac{n+1}{n}X_{(n)}.$$
   On a
   $$\mathbb{V}\left[\tilde{\theta}_n\right] = \frac{(n+1)^2}{n^2}\mathbb{V}\left[X_{(n)}\right] = \frac{\theta^2}{n(n+2)} = \mathrm{EQM}\left(\tilde{\theta}_n,\theta\right),$$
   et on obtient donc un meilleur estimateur.

3. On considère l'estimateur $\hat{\theta}_{\alpha,n} = \alpha X_{(n)}$. Calculer l'erreur quadratique moyenne.

   on obtient
   $$\mathbb{E}\left[\alpha X_{(n)}\right] = \alpha\frac{n}{n+1}\theta \qquad \text{et} \qquad B\left(\alpha X_{(n)},\theta\right) = \frac{\alpha n - n - 1}{n+1}\theta$$

et

$$\begin{aligned}\mathrm{EQM}\left(\alpha X_{(n)},\theta\right) &= \alpha^2\mathbb{E}\left[X_{(n)}^2\right] - 2\alpha\theta\mathbb{E}\left[X_{(n)}\right] + \theta^2\\ &= \theta^2\left(\frac{n}{n+2}\alpha^2 - \frac{2n}{n+1}\alpha + 1\right)\end{aligned}$$

4. Choisir α afin de minimiser l'erreur quadratique moyenne.
   $\alpha = \frac{n+2}{n+1}$
5. Conclure.
   On choisit $X_{(\alpha)}$ pour α optimal même si celui ci est biaisé !

Exercice 8 : J'existe ? On considère une variable aléatoire $X \sim \mathcal{B}\left(p, \lambda^{-1}\right)$ avec $\lambda > 1$ et $\hat{\theta}(X)$ un estimateur de λ.

1. Calculer $\mathbb{E}\left[\hat{\theta}(X)\right]$.
   On a
   $$\mathbb{E}\left[\hat{\theta}(X)\right] = \sum_{k=0}^{n} \lambda^{-k}\left(1-\frac{1}{\lambda}\right)^{n-k}\hat{\theta}(k)$$
2. Sachant qu'un polynôme non nul de degré p admet au plus p racines, conclure.
   On suppose par l'absurde que l'estimateur est non biaisé et on a alors
   $$\lambda = \sum_{k=0}^{n} \lambda^{-k}\left(1-\frac{1}{\lambda}\right)^{n-k}\hat{\theta}(k)$$
   ce que l'on peut réécrire, pour tout $\lambda > 1$ comme
   $$\lambda^{n+1} - \sum_{k=0}^{n}(\lambda-1)^{n-k}\hat{\theta}(k) = 0$$
   et comme $\hat{\theta}(k)$ est déterministe et ne dépend pas de λ, c'est gagné.

Exercice 9 : Estimation de deux paramètres. Soit $Y \sim \mathcal{E}(\lambda)$ avec $\lambda > 0$ inconnu. Soit $X = Y + \theta$ avec $\theta > 0$ inconnu. On admettra que X a pour densité $f_{\theta,\lambda}$ définie pour tout x par

$$f_{\theta,\lambda}(x) = \lambda\exp\left(-\lambda(x-\theta)\right)\mathbf{1}_{[\theta,+\infty[}(x).$$

Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d de même loi que X.

1. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[Y]$.
   On a $\mathbb{E}[X] = \theta + \lambda^{-1}$ et $\mathbb{V}[X] = \mathbb{V}[Y] = \lambda^{-2}$.
2. Déterminer l'estimateur du maximum de vraisemblance de θ.
   On a
   $$L_n(\theta) = \prod_{i=1}^{n}\lambda\exp\left(-\lambda\left(X_i-\theta\right)\right)\mathbf{1}_{[\theta,+\infty[}\left(X_i\right) = \lambda^n\exp\left(\lambda\theta - \lambda n\overline{X}_n\right)\mathbf{1}_{[0,X_{(1)}]}(\theta)$$

et la fonction $\theta \longmapsto \lambda^n \exp\left(\lambda\theta - \lambda n \overline{X}_n\right)$ est clairement croissante. On obtient donc

$$\hat{\theta}_n = X_{(1)}.$$

3. Calculer la fonction de répartition de $X_{(1)}$.
   Pour tout $t \geq \theta$,

$$\begin{aligned}
\mathbb{P}\left[X_{(1)} \leq t\right] &= 1 - \mathbb{P}\left[X_{(1)} > t\right] \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}\left[X_i > t\right] \\
&= 1 - \left(\mathbb{P}\left[X_i > t\right]\right)^n \\
&= 1 - \left(\mathbb{P}\left[Y_i > t - \theta\right]\right)^n \\
&= 1 - \exp\left(-n\lambda\left(t - \theta\right)\right).
\end{aligned}$$

4. En déduire l'erreur quadratique moyenne de $X_{(1)}$.
   1ere version : version bourrin On a pour tout $x \geq \theta$,

$$f_{X_{(1)}}(x) = n\lambda \exp\left(-n\lambda(x - \theta)\right)$$

et on obtient donc

$$\begin{aligned}
\mathbb{E}\left[X_{(1)}\right] &= \int_{\theta}^{+\infty} x n\lambda \exp\left(-n\lambda(x-\theta)\right) dx \\
&= \left[-x \exp\left(-n\lambda(x-\theta)\right)\right]_{\theta}^{+\infty} + \int_{\theta}^{+\infty} \exp\left(-n\lambda(x-\theta)\right) dx \\
&= \theta + \left[-\frac{1}{n\lambda} \exp\left(-n\lambda(x-\theta)\right)\right]_{\theta}^{+\infty} \\
&= \frac{1}{n\lambda} + \theta
\end{aligned}$$

et l'estimateur est donc biaisé. De la même façon, on a

$$\begin{aligned}
\mathbb{E}\left[X_{(1)}^2\right] &= \int_{\theta}^{+\infty} x^2 n\lambda \exp\left(-n\lambda(x-\theta)\right) dx \\
&= \left[-x^2 \exp\left(-n\lambda(x-\theta)\right)\right]_{\theta}^{+\infty} + 2\int_{\theta}^{\infty} x \exp\left(-n\lambda(x-\theta)\right) dx \\
&= \theta^2 + \frac{2}{n\lambda} \mathbb{E}\left[X_{(1)}\right] \\
&= \theta^2 + \frac{2}{n\lambda}\left(\frac{1}{n\lambda} + \theta\right)
\end{aligned}$$

et on obtient donc

$$\mathbb{V}\left[X_{(1)}\right] = \frac{1}{n^2\lambda^2}.$$

2eme version : un peu plus subtile On peut remarquer $X_{(1)} = \theta + Z$ où $Z \sim \mathbb{E}(n\lambda)$ et on obtient directement l'espérance et la variance.

Conclusion : On a par la décomposition biais variance

$$\text{EQM}\left(X_{(1)}, \theta\right) = \frac{2}{n^2\lambda^2}$$

5. Déduire de la question 5
$$\sqrt{n}\left(X_{(1)} - \theta\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$
Que pouvez vous en déduire ?
Via question 5 : On a pour tout $\epsilon > 0$
$$\mathbb{P}\left[\sqrt{n}\left(X_{(1)} - \theta\right) \geq \epsilon\right] = \mathbb{P}\left[X_{(1)} \geq \theta + \frac{\epsilon}{\sqrt{n}}\right] = \exp(-\sqrt{n}\lambda) \xrightarrow[n\to\infty]{} 0.$$
Via question 6 : On a pour tout $\epsilon > 0$, via Markov
$$\mathbb{P}\left[\sqrt{n}\left(X_{(1)} - \theta\right) \geq \epsilon\right] \leq n\frac{\text{EQM}\left(X_{(1)}, \theta\right)}{\epsilon^2} \leq \frac{2}{n\lambda^2\epsilon^2} \xrightarrow[n\to+\infty]{} 0.$$
On en déduit que $X_{(1)}$ converge en probabilité vers θ.
6. En déduire un estimateur de λ.
On a l'estimateur
$$\hat{\lambda}_n = \left(\overline{X}_n - \hat{\theta}_n\right)^{-1}.$$
Remarque : L'estimateur est presque sûrement bien définie car $\overline{X}_n = \hat{\theta}_n \Rightarrow \forall i, j, X_i = X_j$, ce qui est un évènement de probabilité nulle.
7. Montrer que l'estimateur de λ est consistant. Pour s'aider, on admettra que si (A_n) et (B_n) sont deux suites de variables aléatoires convergeant en probabilités vers a et b, et $g : I \times J \longrightarrow \mathbb{R}$ est une fonction continue en (a, b), alors
$$g\left(A_n, B_n\right) \xrightarrow[n\to+\infty]{\mathbb{P}} g(a, b).$$
avec I, J des intervalles ouvert de $\mathbb{R}$.
On a (LGN) $\overline{X}_n$ qui converge en probabilité vers $\theta + \lambda^{-1}$ et $\hat{\theta}_n$ converge en probabilité vers θ. De plus, la fonction $g : (x, y) \longmapsto (x - y)^{-1}$ est continue sur $\left\{(x, y) \in \mathbb{R}^2, x \neq y\right\}$ et donc continue en $\left(\theta + \lambda^{-1}, \theta\right)$. On a donc la consistance de $\hat{\lambda}_n = g\left(\overline{X}_n, \hat{\theta}_n\right)$.
8. Montrer que
$$\sqrt{n}\left(\overline{X}_n - \hat{\theta}_n - \frac{1}{\lambda}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right).$$
On a la décomposition

$$\sqrt{n}\left(\overline{X}_n - \hat{\theta}_n - \frac{1}{\lambda}\right) = \sqrt{n}\left(\overline{X}_n - \theta - \frac{1}{\lambda}\right) - \sqrt{n}\left(X_{(1)} - \theta\right).$$

On a vu que le second terme converge en probabilité vers 0 et pour le premier, on a, comme $X = Y + \theta$,

$$\sqrt{n}\left(\overline{X}_n - \theta - \frac{1}{\lambda}\right) = \sqrt{n}\left(\overline{Y}_n - \frac{1}{\lambda}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right),$$

et on conclut grâce à Slutsky.

9. En déduire la normalité asymptotique de l'estimateur de λ.

   On applique la delta méthode avec $g : x \longmapsto x^{-1}$, et on obtient

$$\sqrt{n}\left(g\left(\overline{X}_n - \hat{\theta}_n\right) - g\left(\frac{1}{\lambda}\right)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \left(g'\left(\frac{1}{\lambda}\right)\right)^2 \frac{1}{\lambda^2}\right)$$

   i.e

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \lambda^2\right).$$

10. Soit $\alpha \in (0,1)$, en déduire un intervalle de confiance asymptotique de λ.

    On a

$$\sqrt{n}\frac{\hat{\lambda}_n - \lambda}{\lambda} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

Version 1 : Par Slutsky (détailler comme d'habitude)

$$\sqrt{n}\frac{\hat{\lambda}_n - \lambda}{\hat{\lambda}_n} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

et on a donc,

$$\lim_{n\to\infty} \mathbb{P}\left[-q_{1-\alpha/2} \leq \sqrt{n}\frac{\hat{\lambda}_n - \lambda}{\hat{\lambda}_n} \leq q_{1-\alpha/2}\right] = 1-\alpha$$

avec $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite, et on obtient donc

$$IC_{1-\alpha}(\lambda) = \left[\lambda_n \pm q_{1-\alpha/2}\frac{\lambda_n}{\sqrt{n}}\right].$$

Version 2 : Soit $q_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi normale centrée réduite, on a

$$\lim_{n\to+\infty} \mathbb{P}\left[-q_{1-\alpha/2} \leq \sqrt{n}\left(\frac{\hat{\lambda}_n}{\lambda} - 1\right) \leq q_{1-\alpha/2}\right] = 1-\alpha$$

et on obtient donc l'intervalle de confiance

$$IC_{1-\alpha}(\lambda) = \left[\frac{\lambda_n}{1 + \frac{q_{1-\alpha/2}}{\sqrt{n}}}; \frac{\lambda_n}{1 - \frac{q_{1-\alpha/2}}{\sqrt{n}}}\right]$$

Remarque : Qu'il n'oublie pas de dire que $q_{1-\alpha/2}$ est le quantile... je ne suis pas devin, je ne suis pas censé deviner.

11. Montrer que

$$n\left(X_{(1)}-\theta\right)\sim\mathcal{E}(\lambda).$$

Pour tout $t\geq 0$, on a

$$\mathbb{P}\left[n\left(X_{(1)}-\theta\right)\leq t\right]=\mathbb{P}\left[X_{(1)}\leq\theta+\frac{t}{n}\right]=1-\exp\left(-\lambda t\right).$$

ce qui et bien la fonction de répartition de $\mathcal{E}(\lambda)$.

12. Pour tout $\alpha\in(0,1)$, donner le quantile $q_{\lambda,1-\alpha}$ d'ordre $1-\alpha$ de la loi exponentielle de paramètre λ.

On a

$$\begin{aligned}F_Y(t)=1-\alpha &\Leftrightarrow 1-\exp\left(-\lambda t\right)=1-\alpha\\ &\Leftrightarrow \exp\left(-\lambda t\right)=\alpha\\ &\Leftrightarrow -\lambda t=\ln(\alpha)\\ &\Leftrightarrow t=\frac{-\ln(\alpha)}{\lambda}.\end{aligned}$$

13. Donner la convergence de

$$n\left(X_{(1)}-\theta\right)-\left(q_{\hat{\lambda}_n,1-\alpha}-q_{\lambda,1-\alpha}\right).$$

avec $q_{\hat{\lambda}_n}=\frac{-\ln(\alpha)}{\hat{\lambda}_n}$.

On a $n\left(X_{(1)}-\theta\right)\sim\mathcal{E}(\lambda)$ et par le théorème de continuité,

$$q_{\hat{\lambda}_n,1-\alpha}-q_{\lambda,1-\alpha}\xrightarrow[n\to+\infty]{\mathbb{P}}0$$

et on obtient donc, par Slutsky

$$n\left(X_{(1)}-\theta\right)-\left(q_{\hat{\lambda}_n,1-\alpha}-q_{\lambda,1-\alpha}\right)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{E}(\lambda).$$

14. Déterminer un intervalle de confiance asymptotique de niveau $1-\alpha$ pour θ.

On a d'après la question précédente,

$$\mathbb{P}\left[n\left(X_{(1)}-\theta\right)-\left(q_{\hat{\lambda}_n,1-\alpha}-q_{\lambda,1-\alpha}\right)\leq q_{\lambda,1-\alpha}\right]\xrightarrow[n\to\infty]{}1-\alpha$$

i.e

$$\mathbb{P}\left[n\left(X_{(1)}-\theta\right)\leq q_{\hat{\lambda}_n,1-\alpha}\right]\xrightarrow[n\to+\infty]{}1-\alpha$$

i.e

$$\mathbb{P}\left[\theta \geq X_{(1)} - \frac{q_{\hat{\lambda}_n,1-\alpha}}{n}\right] \xrightarrow[n\to+\infty]{} 1-\alpha$$

et comme $X_{(1)} \geq \theta$, on obtient

$$\mathbb{P}\left[X_{(1)} - \frac{q_{\hat{\lambda}_n,1-\alpha}}{n} \leq \theta \leq X_{(1)}\right] = \mathbb{P}\left[\theta \geq X_{(1)} - \frac{q_{\hat{\lambda}_n,1-\alpha}}{n}\right] \xrightarrow[n\to+\infty]{} 1-\alpha$$

et on obtient donc l'intervalle de confiance asymptotique

$$IC_{1-\alpha}(\theta) = \left[x_{(1)} - \frac{q_{\lambda_n,1-\alpha}}{n}; x_{(1)}\right] = \left[x_{(1)} + \frac{\ln(\alpha)}{n\lambda_n}; x_{(1)}\right]$$

Exercice 10 : On considère une urne dans laquelle il y a m_1 boules rouges et m_2 boules noires. Le nombre de boules de chaque couleur est inconnu, et le nombre de boules total de boules est bien trop conséquent pour que l'on s'amuse à les compter. L'objectif est donc de proposer différentes stratégies pour estimer la proportion $p = \frac{m_1}{m_1+m_2}$ de boules rouges.

1. On propose d'effectuer N tirages avec remise et on note $X_i = 1$ si le i-ème tirage est une boule rouge et $X_i = 0$ sinon.

   (a) Proposer un estimateur de p et donner son erreur quadratique moyenne.

   On prend l'estimateur $\overline{X}_n$ et on a directement

   $$\text{EQM}\left(\overline{X}_n, p\right) = \frac{p(1-p)}{n}.$$

   (b) Donner sa normalité asymptotique et en déduire un intervalle de confiance asymptotique de niveau $1-\alpha$ de p.

   On a

   $$\sqrt{N}\left(\overline{X}_N - p\right) \xrightarrow[N\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, p(1-p)\right)$$

   et on a donc (Slutsky)

   $$\sqrt{N}\frac{\overline{X}_N - p}{\sqrt{p(1-p)}} \xrightarrow[N\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,1\right)$$

   et on obtient donc l'intervalle de confiance

   $$IC_{1-\alpha}(p) = \left[\overline{x}_N \pm q_{1-\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{N}}\right]$$

2. On propose d'effectuer N tirages sans remise et on note Y le nombre de boules rouges tirées ($N << m_1 + m_2$).

   (a) Quelle est la loi de Y. On admettra que $\mathbb{E}\left[Y\right] = Np$ et $\mathbb{V}[Y] = Np(1-p)\frac{m_1+m_2-N}{m_1+m_2-1}$.

On a pour tout $k \in \mathbb{N}$,

$$\mathbb{P}[Y = k] = \frac{\binom{m_1}{k}\binom{m_2}{N-k}}{\binom{m_1+m_2}{N}}$$

En effet, c'est bien le nombre de façon de choisir k boules rouges parmis m_1 et $N-k$ boules noirs parmis m_2 sur le nombre de façon de choisir N boules parmis $m_1 + m_2$. En réalité, Y suit une loi géométrique (pas à connaître).

(b) Proposer un nouvel estimateur de p et donner son erreur quadratique moyenne. Quel estimateur choisiriez-vous ?

On va considérer l'estimateur $\hat{p}_N = \frac{Y}{N}$. C'est un estimateur sans biais, et on a

$$\mathrm{EQM}(\hat{p}_N, p) = \frac{1}{N^2}\mathbb{V}[Y] = \frac{p(1-p)}{N}\frac{m_1 + m_2 - N}{m_1 + m_2 - 1}$$

et l'erreur est donc plus petite pour tout p. On va donc choisir cet estimateur.

(c) Donner un intervalle de confiance de niveau au moins $1 - \alpha$ de p.

Malheureusement, on n' pas accès aux quantiles de la loi hyper géométrique. Par contre, l'inégalité de Bienaymé Tchebychef nous donne

$$\mathbb{P}[|\hat{p}_N - p| \geq \epsilon] \leq \frac{\mathbb{V}[\hat{p}_N]}{\epsilon^2} \leq \frac{1}{4N\epsilon^2}\frac{m_1 + m_2 - N}{m_1 + m_2 - 1} = \alpha \Leftrightarrow \epsilon = \frac{1}{2\sqrt{N\alpha}}\sqrt{\frac{m_1 + m_2 - N}{m_1 + m_2 - 1}}$$

On obtient ainsi l'intervalle de confiance

$$IC_{1-\alpha}(p) = \left[\hat{p}_N \pm \frac{1}{2\sqrt{N\alpha}}\sqrt{\frac{m_1 + m_2 - N}{m_1 + m_2 - 1}}\right]$$

3. On propose d'effectuer la même expérience n fois, mais en n'effectuant que $K = \frac{N}{n}$ (on suppose K entier) tirages sans remise et on note Y_i le nombre de boules rouges tirées à la i-ème expérience.

(a) Proposer un nouvel estimateur de p et donner son erreur quadratique moyenne. Quel estimateur choisiriez-vous ?

On considère $\overline{Y}_n = \frac{1}{N}\sum_{i=1}^{n} Y_i$. C'est un estimateur sans biais et on a

$$\mathrm{EQM}\left(\overline{Y}_n, p\right) = \mathbb{V}\left[\overline{Y}_n\right] = \frac{n}{N^2}\mathbb{V}[Y_1] = \frac{nKp(1-p)}{N^2}\frac{m_1 + m_2 - K}{m_1 + m_2 - 1} = \frac{p(1-p)}{N}\frac{m_1 + m_2 - K}{m_1 + m_2 - 1}$$

On obtient donc une meilleur erreur que pour le premier mais une moins bonne que pour le deuxième.

(b) Donner sa normalité asymptotique et en déduire un nouvelle intervalle de confiance.

On a

$$\sqrt{n}\left(\overline{Y}_n - p\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{p(1-p)}{N}\frac{m_1 + m_2 - K}{m_1 + m_2 - 1}\right)$$

et on obtient donc l'intervalle de confiance

$$IC_{1-\alpha}\left[\overline{y}_n \pm q_{1-\alpha/2}\sqrt{\frac{\overline{Y}_n(1-\overline{Y}_n)}{nK}\frac{m_1+m_2-K}{m_1+m_2-1}}\right] = \left[\overline{y}_n \pm q_{1-\alpha/2}\sqrt{\frac{\overline{Y}_n(1-\overline{Y}_n)}{N}\frac{m_1+m_2-K}{m_1+m_2-1}}\right]$$

4. On a $m_1 = 2500, m_2 = 7500, N = 1000, K = 10$ et on obtient des intervalles de confiance (dans l'ordre) de taille $0.054, 0.14, 0.053$. Comment interpréter ces résultats ?

   Le fait que le deuxième soit ("bizarement") plus grand vient du côté "grossier" de la majoration via Markov. On a $\frac{1}{\sqrt{\alpha}} = 4.5$ alors que $q_{1-\alpha} = 1.96$. De plus, le fait que le troisième ne soit pas boucoup plus précis vient du fait que $\frac{m_1+m_2-K}{m_1+m_2-1}$ est proche de 1.