Travaux dirigés

Estimation

ECS2 – Lycée La Bruyère, Versailles

Année 2017/2018

Exercice 1

Question 1

C'est du cours !

- Tout d'abord, $\overline{X}_n$ est fonction de $X_1, \dots, X_n$ donc c'est un estimateur.
- Puis
$$\mathbb{E}(\overline{X}_n) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k) = m$$
donc $\overline{X}_n$ est un estimateur sans biais de m.

- De même, par indépendance de $X_1, \dots, X_n$,
$$\mathbb{V}(\overline{X}_n) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{V}(X_k) = \frac{\sigma^2}{n}.$$
Pour $\varepsilon > 0$ donné, l'inégalité de Bienaymé-Tchebychev donne :
$$\mathbb{P}\big(|\overline{X}_n - \mathbb{E}(\overline{X}_n)| \geqslant \varepsilon\big) \leqslant \frac{\mathbb{V}(\overline{X}_n)}{\varepsilon^2}$$
donc (on retrouve la LGN) :
$$\mathbb{P}\big(|\overline{X}_n - m| \geqslant \varepsilon\big) \leqslant \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$
Il en résulte par encadrement que
$$\lim_{n\to\infty}\mathbb{P}\big(|\overline{X}_n - m| \geqslant \varepsilon\big) = 0$$
et $\overline{X}_n$ est donc un estimateur convergent de m.

Exercice 1

Question 2

Il vient :
$$\mathbb{E}(T_n) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k - m)^2 = \frac{1}{n}\sum_{k=1}^{n}\mathbb{V}(X_k) = \sigma^2$$
et T_n est donc un estimateur sans biais de σ^2.

Exercice 1

Question 3.a

Pour $n \in \mathbb{N}$, V_n est bien un estimateur car fonction du n-échantillon $(X_1, \dots, X_n)$.
Par ailleurs,
$$\begin{aligned}V_n &= \frac{1}{n}\sum_{k=1}^{n}(X_k^2 - 2X_k\overline{X}_n + \overline{X}_n^2) = \frac{1}{n}\Big(\sum_{k=1}^{n}X_k^2 - 2\overline{X}_n\sum_{k=1}^{n}X_k + n\overline{X}_n^2\Big)\\ &= \frac{1}{n}\Big(\sum_{k=1}^{n}X_k^2 - n\overline{X}_n^2\Big) = \frac{1}{n}\sum_{k=1}^{n}X_k^2 - \overline{X}_n^2.\end{aligned}$$
Par suite,
$$\begin{aligned}\mathbb{E}(V_n) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k^2) - \mathbb{E}(\overline{X}_n^2)\\ &= \frac{1}{n}\sum_{k=1}^{n}\big(\mathbb{V}(X_k) + \mathbb{E}(X_k)^2\big) - \big(\mathbb{V}(\overline{X}_n) + \mathbb{E}(\overline{X}_n)^2\big)\\ &= \frac{1}{n}\sum_{k=1}^{n}(\sigma^2 + m^2) - \Big(\frac{\sigma^2}{n} + m^2\Big) = \Big(1 - \frac{1}{n}\Big)\sigma^2.\end{aligned}$$

D'où le biais de V_n :
$$b(V_n) = \mathbb{E}(V_n) - \sigma^2 = -\frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$$
et V_n est donc un estimateur de σ^2 asymptotiquement sans biais.

Exercice 1

Question 3.b

D'après la question précédente,
$$\widetilde{V}_n = \frac{1}{1-\frac{1}{n}}V_n = \frac{n}{n-1}V_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2$$
est un estimateur sans biais de σ^2. En effet,
$$\mathbb{E}(\widetilde{V}_n) = \frac{n}{n-1}\mathbb{E}(V_n) = \sigma^2.$$

Exercice 2

Question 1

Par linéarité de l'espérance,
$$\mathbb{E}(\overline{X}_n) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k) = \frac{1}{n}\sum_{k=1}^{n}\frac{\theta}{2} = \frac{\theta}{2}$$
et $\overline{X}_n$ est donc un estimateur sans biais de $\frac{\theta}{2}$.
Par suite, $T_n = 2\overline{X}_n$ est un estimateur sans biais de θ :
$$\mathbb{E}(T_n) = \theta.$$

Exercice 2
Question 2.a

Les X_i ont pour fonction de répartition commune :

$$F : x \in \mathbb{R} \longmapsto \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{\theta} & \text{si } 0 \leqslant x < \theta \\ 1 & \text{si } x \geqslant \theta \end{cases}.$$

Puis, pour $x \in \mathbb{R}$,

$$\begin{aligned} F_{M_n}(x) &= \mathbb{P}(M_n \leqslant x) = \mathbb{P}\big(\max(X_1, \dots, X_n) \leqslant x\big) \\ &= \mathbb{P}(X_1 \leqslant x, \dots, X_n \leqslant x) = \mathbb{P}(X_1 \leqslant x) \cdots \mathbb{P}(X_n \leqslant x) = F(x)^n \end{aligned}$$

par indépendance de $X_1, \dots, X_n$.

Ainsi M_n admet pour fonction de répartition

$$F_{M_n} : x \in \mathbb{R} \longmapsto \begin{cases} 0 & \text{si } x < 0 \\ \left(\frac{x}{\theta}\right)^n & \text{si } 0 \leqslant x < \theta \\ 1 & \text{si } x \geqslant \theta \end{cases},$$

continue sur $\mathbb{R}$ (même en 0 et θ) et de classe $\mathscr{C}^1$ sur $\mathbb{R} \setminus \{0, \theta\}$. La variable M_n est donc à densité donnée par

$$f_{M_n} = F'_{M_n} : x \in \mathbb{R} \longmapsto \begin{cases} \frac{n}{\theta^n} x^{n-1} & \text{si } 0 \leqslant x \leqslant \theta \\ 0 & \text{sinon} \end{cases}.$$

Exercice 2
Question 2.b

La variable M_n étant presque sûrement bornée, elle admet une espérance

$$\mathbb{E}(M_n) = \int_{-\infty}^{+\infty} t f_{M_n}(t)\,\mathrm{d}t = \frac{n}{\theta^n} \int_0^\theta t^n\,\mathrm{d}t = \frac{n}{n+1}\theta.$$

Par suite, $U_n = \frac{n+1}{n} M_n$ est un estimateur sans biais de θ :

$$\mathbb{E}(U_n) = \frac{n+1}{n}\,\mathbb{E}(M_n) = \theta.$$

Exercice 2
Question 3

L'estimateur T_n de θ étant non biaisé, son risque quadratique est donné par

$$r(T_n) = \mathbb{V}(2\overline{X}_n) = \mathbb{V}\Big(\frac{2}{n} \sum_{k=1}^n X_k\Big) = \frac{4}{n^2} \sum_{k=1}^n \mathbb{V}(X_k) = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

par indépendance de $X_1, \dots, X_n$. Il converge vers 0 lorsque $n \to \infty$, ce qui assure que T_n est un estimateur convergent de θ.

On peut raisonner de même pour U_n :

$$r(U_n) = \mathbb{V}\Big(\frac{n+1}{n} M_n\Big) = \Big(\frac{n+1}{n}\Big)^2 \big(\mathbb{E}(M_n^2) - \mathbb{E}(M_n)^2\big)$$

où

$$\mathbb{E}(M_n^2) = \int_{-\infty}^{+\infty} t^2 f_{M_n}(t)\,\mathrm{d}t = \frac{n}{\theta^n} \int_0^\theta t^{n+1}\,\mathrm{d}t = \frac{n}{n+2}\theta^2$$

si bien que

$$r(U_n) = \frac{1}{n(n+2)}\theta^2 \xrightarrow[n\to\infty]{} 0,$$

d'où l'on déduit que U_n est un estimateur convergent de θ.

Exercice 2
Question 4

Le meilleur estimateur est celui qui possède le risque quadratique le plus faible. Or un équivalent de chacun d'eux montre clairement que $r(U_n) = o\big(r(T_n)\big)$ lorsque $n \to \infty$. L'estimateur U_n est donc meilleur estimateur de θ que T_n pour n assez grand.

Exercice 3
Question 1.a

La fonction $f = f_{a,b}$ est continue sur $\mathbb{R} \setminus \{a\}$, positive sur $\mathbb{R}$ avec :

$$\forall x \geqslant a, \quad \int_a^x f(t)\,\mathrm{d}t = 1 - \exp\Big(-\frac{x-a}{b}\Big) \xrightarrow[x\to+\infty]{} 1.$$

La fonction f étant nulle sur $]-\infty, a[$, on en déduit la convergence de l'intégrale

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1.$$

Toutes les conditions sont donc réunies pour faire de f une densité de probabilité.

Exercice 3
Question 1.b

La variable $Y = \frac{X-a}{b}$ a pour densité

$$x \longmapsto b f_{a,b}(a + bx) = \begin{cases} 0 & \text{si } x < 0 \\ \mathrm{e}^{-x} & \text{si } x \geqslant 0 \end{cases}.$$

Elle suit donc la loi exponentielle $\mathcal{E}(1)$.
À ce titre, elle admet espérance et variance données par $\mathbb{E}(Y) = \mathbb{V}(Y) = 1$, d'où l'on déduit celles de X :

$$\mathbb{E}(X) = \mathbb{E}(a + bY) = a + b\,\mathbb{E}(Y) = a + b$$

et

$$\mathbb{V}(X) = \mathbb{V}(a + bY) = b^2\,\mathbb{V}(Y) = b^2.$$

Exercice 3

Question 2.a

La variable X de loi $\mathcal{E}(a,b)$ admet pour fonction de répartition

$$F : x \in \mathbb{R} \longmapsto \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{si } x < a \\ 1 - \exp(-\frac{x-a}{b}) & \text{si } x \geqslant a \end{cases}.$$

Classiquement, l'indépendance de $X_1, \dots, X_n$ donne :

$$\forall x \in \mathbb{R}, \quad \mathbb{P}(T_n > x) = \mathbb{P}(X_1 > x, \dots, X_n > x) = \mathbb{P}(X > x)^n,$$

d'où l'on déduit la fonction de répartition de la variable T_n :

$$x \longmapsto 1 - (1 - F(x))^n = \begin{cases} 0 & \text{si } x < a \\ 1 - \exp(-n\frac{x-a}{b}) & \text{si } x \geqslant a \end{cases}.$$

Comme la fonction de répartition caractérise la loi, la variable T_n suit donc la loi $\mathcal{E}(a, \frac{b}{n})$.

Exercice 3

Question 2.b

La variable aléatoire T_n est un estimateur de a comme fonction de $X_1, \dots, X_n$. D'après **1.b.** et **a.**,

$$\mathbb{E}(T_n) = a + \frac{b}{n} \xrightarrow[n\to\infty]{} a$$

d'où :

$$r(T_n) = \mathbb{V}(T_n) + b(T_n)^2 = 2\frac{b^2}{n^2} \xrightarrow[n\to\infty]{} 0.$$

Il en ressort que T_n est un estimateur asymptotiquement sans biais et convergent de a.

Remarque. La convergence peut s'obtenir grâce à l'inégalité de Markov appliquée à la variable positive $T_n - a$: pour $\varepsilon > 0$,

$$\mathbb{P}(|T_n - a| \geqslant \varepsilon) = \mathbb{P}(T_n - a \geqslant \varepsilon) \leqslant \frac{\mathbb{E}(T_n - a)}{\varepsilon} = \frac{b}{n\varepsilon} \xrightarrow[n\to\infty]{} 0$$

ce qui permet de conclure par encadrement.

Exercice 3

Question 3.a

Il vient :

$$\forall n \in \mathbb{N}^*, \quad \mathbb{E}(U_n) = \mathbb{E}(X) - \mathbb{E}(T_n) = (a+b) - \left(a + \frac{b}{n}\right) = \left(1 - \frac{1}{n}\right)b.$$

Exercice 3

Question 3.b

D'après l'inégalité de Cauchy-Schwarz,

$$\operatorname{cov}(S_n, T_n)^2 \leqslant \mathbb{V}(S_n)\,\mathbb{V}(T_n) = nb^2\frac{b^2}{n^2} = \frac{b^4}{n} \xrightarrow[n\to\infty]{} 0$$

d'où l'on déduit par encadrement que $\operatorname{cov}(S_n, T_n)$ converge vers 0 lorsque $n \to \infty$.

Exercice 3

Question 3.c

Il résulte de **a.** que $\mathbb{E}(U_n) \to b$ lorsque $n \to \infty$. Ainsi U_n est un estimateur asymptotiquement sans biais de b.
Par ailleurs,

$$\begin{aligned} \mathbb{V}(U_n) = \mathbb{V}\left(\frac{S_n}{n} - T_n\right) &= \mathbb{V}\left(\frac{S_n}{n}\right) - 2\operatorname{cov}\left(\frac{S_n}{n}, T_n\right) + \mathbb{V}(T_n) \\ &= \frac{b^2}{n} - \frac{2}{n}\operatorname{cov}(S_n, T_n) + \frac{b^2}{n^2} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

Dans ces conditions,

$$r(U_n) = \mathbb{V}(U_n) + b(U_n)^2 \xrightarrow[n\to\infty]{} 0,$$

ce qui assure que U_n est un estimateur convergent de b.

Exercice 5

Question 1.a

Pour $k \in [\![1, n]\!]$, la variable aléatoire Y_k ne prend que les valeurs 0 et 1 donc suit une loi de Bernoulli de paramètre

$$\mathbb{P}(Y_k = 1) = \mathbb{P}(X_k = 0) = \mathbf{e}^{-\lambda} = \theta.$$

Dès lors,

$$\mathbb{E}(\overline{Y}_n) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(Y_k) = \theta$$

et $\overline{Y}_n$ est un estimateur sans biais de θ.

Exercice 5

Question 1.b

Les variables Y_k sont indépendantes, étant respectivement fonctions des X_k, elles-mêmes indépendantes. Dès lors, on a classiquement :

$$\mathbb{V}(\overline{Y}_n) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{V}(Y_k) = \frac{\theta(1-\theta)}{n}.$$

L'estimateur $\overline{Y}_n$ étant sans biais, son risque quadratique $r(\overline{Y}_n) = \mathbb{V}(\overline{Y}_n)$ converge donc vers 0 lorsque $n \to \infty$. Dans ces conditions, $\overline{Y}_n$ est un estimateur convergent de θ.

Exercice 5

Question 2

Pour $j \in \mathbb{N}$, on a par indépendance de X_1 et $X_2 + \cdots + X_n$ (lemme des coalitions) :

$$\begin{aligned} \varphi(j) = \mathbb{P}_{[S_n = j]}(X_1 = 0) &= \frac{\mathbb{P}(X_1 = 0, X_1 + \cdots + X_n = j)}{\mathbb{P}(X_1 + \cdots + X_n = j)} \\ &= \frac{\mathbb{P}(X_1 = 0, X_2 + \cdots + X_n = j)}{\mathbb{P}(X_1 + \cdots + X_n = j)} \\ &= \frac{\mathbb{P}(X_1 = 0)\,\mathbb{P}(X_2 + \cdots + X_n = j)}{\mathbb{P}(X_1 + \cdots + X_n = j)}. \end{aligned}$$

Comme $X_1, \dots, X_n$ sont mutuellement indépendantes et suivent des lois de Poisson de paramètre λ, les variables $S_n = X_1 + \cdots + X_n$ et $X_2 + \cdots + X_n$ suivent aussi des lois de Poisson, de paramètres respectifs $n\lambda$ et $(n-1)\lambda$, ce qui permet d'achever le calcul :

$$\varphi(j) = \frac{\mathbf{e}^{-\lambda}\mathbf{e}^{-(n-1)\lambda}\frac{((n-1)\lambda)^j}{j!}}{\mathbf{e}^{-n\lambda}\frac{(n\lambda)^j}{j!}} = \left(\frac{n-1}{n}\right)^j.$$

Exercice 5

Question 3.a

La variable $T_n = \varphi(S_n) = \left(\frac{n-1}{n}\right)^{S_n}$ est indépendante du paramètre inconnu λ et ne dépend que du n-échantillon $(X_1, \dots, X_n)$. C'est donc un estimateur de θ.
Par ailleurs, le théorème de transfert assure que :

$$\mathbb{E}(T_n) = \sum_{k=0}^{\infty} \varphi(k)\,\mathbb{P}(S_n = k) = \sum_{k=0}^{\infty} \left(\frac{n-1}{n}\right)^k \mathrm{e}^{-n\lambda} \frac{(n\lambda)^k}{k!}$$
$$= \mathrm{e}^{-n\lambda} \sum_{k=0}^{\infty} \frac{\big((n-1)\lambda\big)^k}{k!} = \mathrm{e}^{-n\lambda}\mathrm{e}^{(n-1)\lambda} = \mathrm{e}^{-\lambda} = \theta$$

puisque la série ci-dessus converge absolument et T_n est donc un estimateur sans biais de θ.

Remarque. On peut également écrire, en utilisant la formule des probabilités totales au SCE associé à la variable X_n :

$$\mathbb{E}(T_n) = \sum_{k=0}^{\infty} \varphi(k)\,\mathbb{P}(S_n = k) = \sum_{k=0}^{\infty} \mathbb{P}_{[S_n=k]}(X_1 = 0)\,\mathbb{P}(S_n = k) = \mathbb{P}(X_1 = 0) = \theta.$$

Exercice 5

Question 3.b

Par le même raisonnement,

$$\mathbb{E}(T_n^2) = \sum_{k=0}^{\infty} \varphi(k)^2\,\mathbb{P}(S_n = k) = \sum_{k=0}^{\infty} \left(\frac{n-1}{n}\right)^{2k} \mathrm{e}^{-n\lambda} \frac{(n\lambda)^k}{k!}$$
$$= \mathrm{e}^{-n\lambda} \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{(n-1)^2}{n}\lambda\right)^k = \mathrm{e}^{-n\lambda}\mathrm{e}^{\frac{(n-1)^2}{n}\lambda} = \mathrm{e}^{(-2+\frac{1}{n})\lambda}.$$

Par suite,

$$\mathbb{V}(T_n) = \mathbb{E}(T_n^2) - \mathbb{E}(T_n)^2 = \mathrm{e}^{-2\lambda}\left(\mathrm{e}^{\frac{\lambda}{n}} - 1\right).$$

L'estimateur T_n étant non biaisé, on en déduit son risque quadratique :

$$r(T_n) = \mathbb{V}(T_n) = \mathrm{e}^{-2\lambda}\left(\mathrm{e}^{\frac{\lambda}{n}} - 1\right) \xrightarrow[n\to\infty]{} 0,$$

d'où il ressort que T_n un estimateur convergent de θ.

Exercice 5

Question 4

Les estimateurs $\overline{Y}_n$ et T_n étant sans biais, leurs risques quadratiques sont donnés par leurs variances :

$$r(\overline{Y}_n) = \mathbb{V}(\overline{Y}_n) = \frac{\theta(1-\theta)}{n}$$

et

$$r(T_n) = \mathbb{V}(T_n) = \theta^2(\theta^{-1/n} - 1).$$

Une rapide étude de variations montre que la fonction $f : t \longmapsto n\mathrm{e}^{t/n} - \mathrm{e}^t - n + 1$ est négative sur $\mathbb{R}_+$, d'où l'on déduit que

$$r(T_n) - r(\overline{Y}_n) = \frac{\theta^2}{n}\left(n\theta^{-1/n} - n - \theta^{-1} + 1\right) = \frac{\theta^2}{n} f(\lambda) \leqslant 0,$$

et T_n est donc meilleur estimateur de θ que $\overline{Y}_n$.

Exercice 6

Question 1

La fonction f_a est continue sur $\mathbb{R} \setminus \{1\}$, positive sur $\mathbb{R}$, avec :

$$\int_a^x f_a(t)\,\mathrm{d}t = 1 - \frac{1}{x^a} \xrightarrow[x\to+\infty]{} 1$$

d'où, sachant que f_a est nulle sur $]-\infty, 1[$, la convergence de l'intégrale

$$\int_{-\infty}^{+\infty} f_a(t)\,\mathrm{d}t = 1.$$

La fonction f_a est donc une densité de probabilité.

Exercice 6

Question 2.a

Pour tout $(x_1, \dots, x_n, a) \in \mathbb{R}^n \times \mathbb{R}_+^*$,

$$L(x_1, \dots, x_n, a) = \begin{cases} 0 & \text{si } \exists i \in [\![1, n]\!],\, x_i < 1 \\ \frac{a^n}{(x_1 \cdots x_n)^{a+1}} & \text{si } \forall i \in [\![1, n]\!],\, x_i \geqslant 1 \end{cases}.$$

Exercice 6

Question 2.b

Pour $(x_1, \dots, x_n) \in [1, +\infty[^n$ donné, on a :

$$\forall a \in \mathbb{R}_+^*, \quad h(a) = L(x_1, \dots, x_n, a) = \frac{a^n}{(x_1 \cdots x_n)^{a+1}}$$

et

$$g(a) = \ln h(a) = n \ln a - (a+1) \sum_{k=1}^{n} \ln x_k.$$

On détermine sans difficulté les variations de g :

a	0		$\widehat{a}$		$+\infty$
g	$-\infty$	$\nearrow$		$\searrow$	$-\infty$

qui admet donc un maximum en

$$\widehat{a} = \frac{n}{\sum_{k=1}^{n} \ln x_k}.$$

Il en va de même de la fonction $h = \exp \circ g$ puisque la fonction exp est croissante.

Exercice 6

Question 3.a

La fonction de répartition de X_k est donnée par :

$$x \longmapsto \mathbb{P}(X_k \leqslant x) = \int_{-\infty}^{x} f_a(t)\,\mathrm{d}t = \begin{cases} 0 & \text{si } x < 1 \\ 1 - \frac{1}{x^a} & \text{si } x \geqslant 1 \end{cases}$$

d'où l'on déduit celle de Y_k :

$$\mathbb{P}(Y_k \leqslant y) = \mathbb{P}(X_k \leqslant \mathrm{e}^y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 - \mathrm{e}^{-ay} & \text{si } y \geqslant 0 \end{cases}.$$

Comme celle-ci caractérise la loi de Y_k, on en déduit que Y_k suit la loi exponentielle $\mathcal{E}(a)$.

Exercice 6
Question 3.b

Les variables $aY_1, \dots, aY_n$ sont indépendantes (par indépendance de $X_1, \dots, X_n$) et suivent la loi $\mathcal{E}(1) = \gamma(1)$ d'après la question **a.**. Dans ces conditions, la variable $aS_n = aY_1 + \cdots + aY_n$ suit la loi $\gamma(n)$ i.e. admet pour densité

$$f_{aS_n} : y \in \mathbb{R} \longmapsto \begin{cases} 0 & \text{si } y < 0 \\ \frac{y^{n-1}}{(n-1)!}\mathbf{e}^{-y} & \text{si } y \geqslant 0 \end{cases}.$$

On en déduit par transformation affine la densité suivante pour $S_n = \frac{aS_n}{a}$:

$$f_{S_n} : x \in \mathbb{R} \longmapsto af_{aS_n}(ax) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{a^n}{(n-1)!}x^{n-1}\mathbf{e}^{-ax} & \text{si } x \geqslant 0 \end{cases}$$

Exercice 6
Question 4

Par définition,

$$T_n = \frac{n}{\sum_{k=1}^{n} \ln X_k} = \frac{n}{S_n}.$$

Le changement de variable affine $u = at$ donne :

$$\int_{-\infty}^{+\infty} \frac{n}{t} f_{S_n}(t)\,\mathrm{d}t = \frac{na^n}{(n-1)!}\int_0^{+\infty} t^{n-2}\mathbf{e}^{-at}\,\mathrm{d}t = \frac{na}{(n-1)!}\int_0^{+\infty} u^{n-2}\mathbf{e}^{-u}\,\mathrm{d}u.$$

On en déduit, par référence aux intégrales Γ $(n-1>0)$, la convergence absolue de l'intégrale ci-dessous et donc, d'après le théorème de transfert, l'existence de

$$\mathbb{E}(T_n) = \int_{-\infty}^{+\infty} \frac{n}{t} f_{S_n}(t)\,\mathrm{d}t = \frac{na}{(n-1)!}\Gamma(n-1) = \frac{n}{n-1}a.$$

On montre de la même façon que T_n admet un moment d'ordre 2 égal à

$$\mathbb{E}(T_n^2) = \frac{n^2a^2}{(n-1)!}\Gamma(n-2) = \frac{n^2a^2}{(n-1)(n-2)}$$

et donc une variance

$$\mathbb{V}(T_n) = \mathbb{E}(T_n^2) - \mathbb{E}(T_n)^2 = \frac{n^2}{(n-1)^2(n-2)}a^2.$$

Exercice 6
Question 5

De la question **4.**, on déduit que

$$b(T_n) = \frac{a}{n-1} \xrightarrow[n\to\infty]{} 0,$$

ce qui montre que T_n est un estimateur asymptotiquement sans biais de a.
De plus,

$$r(T_n) = \mathbb{V}(T_n) + b(T_n)^2 \xrightarrow[n\to\infty]{} 0$$

d'où l'on déduit que T_n est un estimateur convergent de a.

Exercice 7
Question 1

Le calcul a déjà été fait dans l'exercice 1 :

$$\begin{aligned} S_n^2 &= \frac{1}{n}\sum_{k=1}^{n}(X_k^2 - 2X_k\overline{X}_n + \overline{X}_n^2) = \frac{1}{n}\Big(\sum_{k=1}^{n} X_k^2 - 2\overline{X}_n\sum_{k=1}^{n} X_k + n\overline{X}_n^2\Big) \\ &= \frac{1}{n}\Big(\sum_{k=1}^{n} X_k^2 - n\overline{X}_n^2\Big) = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \overline{X}_n^2. \end{aligned}$$

Exercice 7
Question 2

Déjà vu dans l'exercice 1 : d'après **1.**,

$$\begin{aligned} \mathbb{E}(S_n^2) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k^2) - \mathbb{E}(\overline{X}_n^2) \\ &= \frac{1}{n}\sum_{k=1}^{n}\big(\mathbb{V}(X_k) + \mathbb{E}(X_k)^2\big) - \big(\mathbb{V}(\overline{X}_n) + \mathbb{E}(\overline{X}_n)^2\big) \\ &= \frac{1}{n}\sum_{k=1}^{n}(\sigma^2 + m^2) - \Big(\frac{\sigma^2}{n} + m^2\Big) = \Big(1 - \frac{1}{n}\Big)\sigma^2. \end{aligned}$$

On observe en particulier que $\mathbb{E}(S_n^2)$ converge vers σ^2, si bien que S_n^2 est un estimateur asymptotiquement sans biais de σ^2.

Exercice 7
Question 3.a

Les variables

$$T_n = \frac{1}{n}\sum_{k=1}^{n} X_k^2, \quad n \geqslant 1$$

sont des moyennes empiriques pour la suite $(X_n^2)_{n\geqslant 1}$. Les variables X_n^2 étant indépendantes (car les X_n le sont), de même loi et admettant un moment d'ordre 2 (car X admet un moment d'ordre 4), la loi des grands nombres assure que (T_n) converge en probabilité vers $\mathbb{E}(X^2) = \sigma^2 + \mu^2$.

Exercice 7
Question 3.b

D'après la loi des grands nombres appliquée cette fois à la suite (X_n) de variables aléatoires indépendantes et de même loi, admettant une espérance et une variance communes, la suite $(\overline{X}_n)_n$ converge en probabilité vers $\mathbb{E}(X) = \mu$. Par continuité de la fonction $x \longmapsto -x^2$, on en déduit que la suite $(-\overline{X}_n^2)_n$ converge en probabilité vers $-\mathbb{E}(X)^2 = -\mu^2$.
Par théorème opératoire sur la convergence en probabilité établi en TD, on en déduit que $S_n^2 = T_n - \overline{X}_n^2$ converge en probabilité vers $(\sigma^2 + \mu^2) - \mu^2 = \sigma^2$ lorsque $n \to \infty$.

Exercice 7

Question 3.c

La convergence en probabilité de $S_n \geqslant 0$ vers $\sigma^2 \geqslant 0$ lorsque $n \to \infty$ établie en **b.** implique, par continuité de $x \longmapsto \sqrt{x}$ sur $\mathbb{R}_+$, celle de $\sqrt{S_n^2} = S_n$ vers $\sqrt{\sigma^2} = \sigma$.
Ainsi S_n est un estimateur convergent de σ.

Exercice 8

Question 1

Par hypothèse, les résultats des n pesées forment un n-échantillon $(X_1, \dots, X_n)$ i.i.d. de loi $\mathcal{N}(m, \sigma^2)$, où m est le paramètre inconnu et $\sigma = 0{,}1$.
Dans ces conditions, la moyenne empirique

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

suit également une loi normale, d'espérance m et de variance σ^2/n. En notant $z_{1-\alpha/2}$ le quantile de la loi normale centrée réduite à l'ordre $1 - \frac{\alpha}{2}$, on a donc :

$$\mathbb{P}\Big(\big|\overline{X}_n - m\big| \leqslant z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = \mathbb{P}\big(\big|\overline{X}_n^*\big| \leqslant z_{1-\alpha/2}\big) \geqslant 1 - \alpha.$$

En d'autres termes,

$$\Big[\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\Big]$$

est un intervalle de confiance pour m au niveau de confiance $1 - \alpha$.

L'énoncé donne $n = 10$ et $\overline{x}_n = \overline{X}_n(\omega) = 72{,}40$. Pour avoir un intervalle au niveau de confiance 90%, on prend $\alpha = 0{,}1$; on lit dans la table de la loi normale centrée réduite le quantile correspondant $z_{1-\alpha/2} \simeq 1{,}65$. On est ainsi conduit à l'intervalle de confiance $[72{,}347; 72{,}453]$.

Exercice 8

Question 2

La longueur de l'intervalle de confiance obtenu précédemment est $2\frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}$. On cherche donc n tel que

$$2z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant 0{,}05 \quad\Longleftrightarrow\quad n \geqslant \Big(2z_{1-\alpha/2}\frac{\sigma}{0{,}05}\Big)^2$$

ce qui donne, pour $\alpha = 0{,}1$, $n \geqslant 44$.

Exercice 8

Question 3

Comme n est trop petit pour légitimer une approximation gaussienne, on n'a d'autre recours que l'inégalité de Bienaymé-Tchebychev. Pour $\varepsilon > 0$,

$$\mathbb{P}\big(\big|\overline{X}_n - m\big| < \varepsilon\big) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Pour avoir un niveau de confiance $1 - \alpha$, on choisit donc ε tel que $\frac{\sigma^2}{n\varepsilon^2} \leqslant \alpha$, c'est-à-dire $\varepsilon \geqslant \frac{\sigma}{\sqrt{n\alpha}}$.
On est alors conduit à l'intervalle de confiance

$$\Big[\overline{X}_n - \frac{\sigma}{\sqrt{n\alpha}}, \overline{X}_n + \frac{\sigma}{\sqrt{n\alpha}}\Big]$$

pour m au niveau de confiance $1 - \alpha$.
Pour $\alpha = 0{,}1$, on obtient $[72{,}3; 72{,}5]$.

Exercice 10

Question 1.a

Pour $k \in [\![1, n]\!]$, la variable X_k ne prend que les valeurs 0 et 1 donc suit une loi de Bernoulli. On en calcule le paramètre $q = \mathbb{P}(X_k = 1)$ grâce à la formule des probabilités totales. En notant Y_k la variable aléatoire donnant le résultat de l'expérience de Bernoulli réalisée secrètement par le k-ième individu, et en utilisant le système complet qui lui est associé, on obtient :

$$\begin{aligned} q = \mathbb{P}(X_k = 1) &= \mathbb{P}(Y_k = 0)\,\mathbb{P}_{[Y_k=0]}(X_k = 1) + \mathbb{P}(Y_k = 1)\,\mathbb{P}_{[Y_k=1]}(X_k = 1) \\ &= (1-\alpha)(1-p) + \alpha p = (2\alpha - 1)p + 1 - \alpha. \end{aligned}$$

Remarque. Le calcul de $\mathbb{P}_{[Y_k=1]}(X_k = 1)$ n'est pas tout à fait immédiat : en introduisant la variable

$$Z_k = \begin{cases} 1 & \text{si la } k\text{-ième personne interrogée est favorable au projet} \\ 0 & \text{sinon} \end{cases},$$

indépendante de Y_k, il vient :

$$\mathbb{P}_{[Y_k=1]}(X_k = 1) = \mathbb{P}_{[Y_k=1]}(Z_k = 1) = \mathbb{P}(Z_k = 1) = p.$$

Exercice 10

Question 1.b

Pour tout $n \in \mathbb{N}^*$, $(X_1, \dots, X_n)$ est un n-échantillon de loi de Bernoulli de paramètre q, admettant même espérance et variance. La moyenne empirique

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

est donc, par théorème (à redémontrer), un estimateur sans biais et convergent de $\mathbb{E}(X_1) = q$.

Exercice 10

Question 1.c

La formule de la question **a.** permet d'exprimer p en fonction de q :

$$p = \frac{q - 1 + \alpha}{2\alpha - 1}$$

ce qui conduit à définir, à partir de l'estimateur S_n de q obtenu en **b.**, l'estimateur

$$T_n = \frac{S_n - 1 + \alpha}{2\alpha - 1}$$

pour p.

Par linéarité de l'espérance,

$$\mathbb{E}(T_n) = \frac{\mathbb{E}(S_n) - 1 + \alpha}{2\alpha - 1} = \frac{q - 1 + \alpha}{2\alpha - 1} = p$$

et T_n est donc un estimateur sans biais de p.

Par ailleurs, $T_n = f(S_n)$ où $f : x \longmapsto \frac{x-1+\alpha}{2\alpha-1}$ est continue sur $\mathbb{R}$. Puisque S_n est un estimateur convergent de q, on en déduit que T_n est un estimateur convergent de $f(q) = p$.

Exercice 10

Question 1.d

À n fixé, le risque quadratique

$$r(T_n) = \mathbb{V}(T_n) = \frac{1}{(2\alpha-1)^2}\mathbb{V}(S_n) = \frac{q(1-q)}{(2\alpha-1)^2 n} = \frac{p(1-p)}{n} + \frac{\alpha(1-\alpha)}{(2\alpha-1)^2 n}$$

tend vers $+\infty$ lorsque $\alpha \to \frac{1}{2}$. Au contraire, il est minimum lorsque $\alpha \to 0$ ou $\alpha \to 1$, mais une valeur de α trop proche de 0 ou 1 fera perdre son intérêt à la procédure (puisque le résultat de l'expérience secrète sera presque toujours le même).
On a donc intérêt à choisir α suffisamment éloigné de $\frac{1}{2}$ pour avoir un risque quadratique modéré, mais pas trop proche de 0 ou 1 pour préserver la confidentialité du sondage.

Exercice 10

Question 2.a

Méthode 1 : Bienaymé-Tchebychev
D'après l'inégalité de Bienaymé-Tchebychev appliquée à l'estimateur T_n sans biais de p, il vient pour $\varepsilon > 0$:

$$\mathbb{P}\big(|T_n - p| > \varepsilon\big) \leqslant \mathbb{P}\big(|T_n - p| \geqslant \varepsilon\big) \leqslant \frac{\mathbb{V}(T_n)}{\varepsilon^2} = \frac{q(1-q)}{(2\alpha-1)^2 n}\frac{1}{\varepsilon^2}$$

si bien qu'on aura

$$\mathbb{P}\big(p \in [T_n - \varepsilon, T_n + \varepsilon]\big) \geqslant 1 - \beta \quad \text{i.e.} \quad \mathbb{P}\big(|T_n - p| > \varepsilon\big) \leqslant \beta$$

dès que

$$\frac{q(1-q)}{(2\alpha-1)^2 n}\frac{1}{\varepsilon^2} \leqslant \beta \quad \text{i.e.} \quad \varepsilon \geqslant \frac{1}{|2\alpha-1|}\sqrt{\frac{q(1-q)}{n\beta}}$$

et en particulier lorsque $\varepsilon \geqslant \frac{1}{2|2\alpha-1|\sqrt{n\beta}}$ puisque $q(1-q) \leqslant \frac{1}{4}$.

On obtient ainsi l'intervalle de confiance

$$I_n = \left[T_n - \frac{1}{2\,|2\alpha-1|\,\sqrt{n\beta}},\, T_n + \frac{1}{2\,|2\alpha-1|\,\sqrt{n\beta}}\right]$$

pour p au niveau de confiance $1 - \beta$.

Méthode 2 : théorème limite central
D'après le théorème limite central appliqué à la suite (X_n) de variables indépendantes et identiquement distribuées admettant une variance, la suite (S_n^*) converge en loi vers une variable Z de loi normale centrée réduite. Or, comme T_n est fonction affine de S_n,

$$S_n^* = T_n^* = \frac{\sqrt{n}}{\sigma}(T_n - p) \quad \text{où} \quad \sigma = \frac{\sqrt{q(1-q)}}{|2\alpha-1|}.$$

En utilisant l'inégalité $q(1-q) \leqslant \frac{1}{4}$, il vient $\sigma \leqslant \frac{1}{2|2\alpha-1|}$ si bien qu'en notant $z_{1-\beta/2}$ le quantile d'ordre $1 - \frac{\beta}{2}$ de la loi normale centrée réduite,

$$\mathbb{P}\left(|T_n - p| \leqslant \frac{z_{1-\beta/2}}{2\,|2\alpha-1|\,\sqrt{n}}\right) \geqslant \mathbb{P}\left(|T_n - p| \leqslant z_{1-\beta/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \beta_n$$

en notant

$$\beta_n = 1 - \mathbb{P}\big(|T_n^*| \leqslant z_{1-\beta/2}\big) \xrightarrow[n\to\infty]{} 1 - \mathbb{P}(|Z| \leqslant z_{1-\beta/2}) = \beta.$$

On obtient ainsi l'intervalle de confiance asymptotique

$$J_n = \left[T_n - \frac{z_{1-\beta/2}}{2\,|2\alpha-1|\,\sqrt{n}},\, T_n + \frac{z_{1-\beta/2}}{2\,|2\alpha-1|\,\sqrt{n}}\right]$$

pour p au niveau de confiance $1 - \beta$.

Méthode 3 : TLC & Slutsky
On reprend la méthode précédente mais, plutôt que de majorer

$$\sigma = \frac{\sqrt{q(1-q)}}{|2\alpha-1|},$$

on l'estime par

$$\widehat{\sigma}_n = \frac{\sqrt{S_n(1-S_n)}}{|2\alpha-1|}.$$

Puisque S_n converge en probabilité vers q (à valeurs dans $]0,1[$) lorsque $n \to \infty$,

$$\frac{\sigma}{\widehat{\sigma}_n} = \sigma\frac{|2\alpha-1|}{\sqrt{S_n(1-S_n)}} \xrightarrow[n\to\infty]{\mathcal{P}} \sigma\frac{|2\alpha-1|}{\sqrt{q(1-q)}} = 1$$

car la fonction $x \longmapsto \sigma\frac{|2\alpha-1|}{\sqrt{x(1-x)}}$ est continue sur $]0,1[$.

Le lemme de Slutsky assure alors que

$$Z_n = \frac{\sigma}{\widehat{\sigma}_n}\cdot\frac{\sqrt{n}}{\sigma}(T_n - p) = |2\alpha-1|\,\sqrt{n}\frac{T_n - p}{\sqrt{S_n(1-S_n)}}$$

converge en loi vers la loi normale centrée réduite lorsque $n \to \infty$.
Par conséquent,

$$\mathbb{P}(|Z_n| \leqslant z_{1-\beta/2}) = \mathbb{P}\left(|T_n - p| \leqslant z_{1-\beta/2}\frac{\sqrt{S_n(1-S_n)}}{|2\alpha-1|\,\sqrt{n}}\right) \xrightarrow[n\to\infty]{} 1 - \beta$$

d'où il ressort que

$$K_n = \left[T_n - z_{1-\beta/2}\frac{\sqrt{S_n(1-S_n)}}{|2\alpha-1|\,\sqrt{n}},\, T_n + z_{1-\beta/2}\frac{\sqrt{S_n(1-S_n)}}{|2\alpha-1|\,\sqrt{n}}\right]$$

est un intervalle de confiance asymptotique pour p au niveau de confiance $1 - \beta$.

L'énoncé donne $\alpha = \frac{1}{6}$, $n = 1000$, $s_n = S_n(\omega) = 0{,}425$ d'où $t_n = T_n(\omega) = 0{,}6125$.
Pour avoir un intervalle de confiance au niveau de confiance 95%, on choisit $\beta = 0{,}05$ et on lit dans la table de la loi normale $z_{1-\beta/2} \simeq 1{,}96$.
On obtient alors les intervalles de confiance :

$$I_{1000} \simeq [0{,}506, 0{,}719], \quad J_{1000} \simeq [0{,}566; 0{,}659]$$

et

$$K_{1000} \simeq [0{,}566; 0{,}659].$$

Remarque. L'intervalle K_{1000} est très légèrement plus précis que l'intervalle J_{1000}, lui-même largement plus précis que l'intevalle I_{1000}.

Exercice 10

Question 2.b

```
n=10^3;   // nombre de simulations
N=100;    // nombre d'IC
alpha=1/6;
p=rand();   // choix aléatoire de p
Z=grand(N,n,'bin',1,p);   // avis des personnes interrogées
Y=grand(N,n,'bin',1,alpha);   // résultats des expériences de Bernoulli
X=Y.*Z+(1-Y).*(1-Z);   // réponses obtenues
S=sum(X,'c')/n;
T=(S-1+alpha)/(2*alpha-1);
r=1.96*sqrt(S.*(1-S))/abs(2*alpha-1)/sqrt(n);   // rayon des IC
plot2d(1:N,p*ones(1,N));
for k=1:N
  if (abs(p-T(k))<=r(k)) then opt='b-+';
  else opt='r-+';   // choix de la couleur selon que l'IC contient p
  end
  plot([k,k],[T(k)-r(k),T(k)+r(k)],opt);   // représentation de l'IC
end
```


Exercice 11

Question 1.a

Pour $i \in [\![1, n]\!]$ donné, la variable D_i donne le temps d'attente du premier succès (obtenir un poisson marqué) lors d'une répétition d'expériences de Bernoulli (le « tirage » d'un poisson à partir du $i+1$-ième) indépendantes et de même paramètre de succès $p = \frac{m}{N}$. Elle suit donc la loi géométrique $\mathcal{G}(\frac{m}{M})$. Son espérance et sa variance sont données par :

$$\mathbb{E}(D_i) = \frac{1}{p} = \frac{N}{m} \quad \text{et} \quad \mathbb{V}(D_i) = \frac{1-p}{p^2} = \frac{N(N-m)}{m^2}.$$

Exercice 11

Question 1.b

On a bien sûr $X_n = D_1 + \cdots + D_n$ d'où :

$$\mathbb{E}(X_n) = \sum_{i=1}^{n} \mathbb{E}(D_i) = n\frac{N}{m}$$

et, par indépendace des variables $D_1, \ldots, D_n$,

$$\mathbb{V}(X_n) = \sum_{i=1}^{n} \mathbb{V}(D_i) = n\frac{N(N-m)}{m^2}.$$

Exercice 11

Question 1.c

La variable $T_n = \frac{m}{n}X_n$ est fonction de $X_1, \ldots, X_n$; c'est donc un estimateur de N, sans biais puisque

$$\mathbb{E}(T_n) = \frac{m}{n}\mathbb{E}(X_n) = N$$

d'après **b.**.

Exercice 11

Question 2

D'après le théorème limite central appliqué à la suite $(D_i)_{i\geqslant 1}$ de variables aléatoires indépendantes et identiquement distribuées, admettant un moment d'ordre 2, la variable centrée réduite associée à

$$\frac{1}{n}\sum_{i=1}^{n} D_i = \frac{X_n}{n} = \frac{T_n}{m}$$

converge en loi vers la loi normale centrée réduite lorsque $n \to \infty$. En notant $z_{1-\alpha/2}$ le quantile de niveau $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite, on a donc :

$$\mathbb{P}(|T_n^*| \leqslant z_{1-\alpha/2}) = \mathbb{P}(|T_n - N| \leqslant z_{1-\alpha/2}\sigma(T_n)) \xrightarrow[n\to\infty]{} 1 - \alpha.$$

Sachant que $\sigma(T_n) \leqslant 100$, on en déduit que

$$\left[T_n - 100z_{1-\alpha/2}, T_n + 100z_{1-\alpha/2}\right]$$

est un intervalle de confiance asymptotique pour N au niveau de confiance $1 - \alpha$.
Pour $\alpha = 0{,}9$, on trouve $z_{1-\alpha/2} \simeq 1{,}64$.

Exercice 11

Question 3

Avec les données de l'énoncé, $m = 200$, $n = 50$ et $X_n = 450$, on obtient l'intervalle de confiance $[1636, 1964]$ pour N au niveau de confiance 90%.

Exercice 11

Question 4.a

La variable Y_n compte le nombre de succès (pêcher un poisson marqué) lors d'une succession de n épreuves de Bernoulli indépendantes de même paramètre de succès $p = \frac{m}{N}$. Elle suit donc la loi binomiale $\mathcal{B}(n, \frac{m}{N})$. Elle admet donc espérance $\mathbb{E}(Y_n) = \frac{nm}{N}$ si bien que $U_n = \frac{1}{nm}Y_n$ est un estimateur sans biais de $\frac{1}{N}$:

$$\mathbb{E}(U_n) = \frac{1}{nm}\mathbb{E}(Y_n) = \frac{1}{N}.$$

Exercice 11

Question 4.b

L'événement $[Y_n = 0]$ a une probabilité non nulle : $\frac{nm}{Y_n}$ n'est donc pas une variable aléatoire bien définie.

Exercice 11

Question 5.a

D'après le théorème de transfert, la variable finie V_n admet pour espérance :

$$\begin{aligned}\mathbb{E}(V_n) &= \sum_{k=0}^{n} \frac{m(n+1)}{k+1}\,\mathbb{P}(Y_n = k)\\ &= \sum_{k=0}^{n} \frac{m(n+1)}{k+1}\binom{n}{k}\left(\frac{m}{N}\right)^k\left(1-\frac{m}{N}\right)^{n-k}\\ &= N\sum_{k=0}^{n}\binom{n+1}{k+1}\left(\frac{m}{N}\right)^{k+1}\left(1-\frac{m}{N}\right)^{n-k}\\ &= N\sum_{k=1}^{n+1}\binom{n+1}{k}\left(\frac{m}{N}\right)^{k}\left(1-\frac{m}{N}\right)^{n+1-k}\\ &= N\left(1-\left(1-\frac{m}{N}\right)^{n+1}\right).\end{aligned}$$

Exercice 11

Question 5.b

On observe sur l'expression précédente que

$$\mathbb{E}(U_n) = N\left(1-\left(1-\frac{m}{N}\right)^{n+1}\right) \xrightarrow[n\to\infty]{} N,$$

ce qui montre que U_n est un estimateur asymptotiquement sans biais de N.

Exercice 12

D'après le théorème limite central appliqué à la suite (X_n) de variables indépendantes et identiquement distribuées admettant une espérance λ et une variance λ communes,

$$T_n = \sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} = \overline{X}_n^*$$

converge en loi vers la loi normale centrée réduite lorsque $n \to \infty$.
Par ailleurs, la loi des grands nombres donne la convergence en probabilité de la suite $(\overline{X}_n)$ vers $\mathbb{E}(X_1) = \lambda$, les variables étant à valeurs dans $[0, +\infty[$. Par continuité de la fonction $x \longmapsto \sqrt{\lambda/x}$ en λ, on en déduit que

$$\sqrt{\frac{\lambda}{\overline{X}_n}} \xrightarrow[n\to\infty]{\mathcal{P}} \sqrt{\frac{\lambda}{\lambda}} = 1.$$

Dans ces conditions, le lemme de Slutsky assure que

$$Z_n = \sqrt{\frac{\lambda}{\overline{X}_n}}\,T_n = \sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\overline{X}_n}}$$

converge en loi vers la loi normale centrée réduite.

Par conséquent, si $z_{1-\alpha/2}$ est le quantile de la loi normale centrée réduite à l'ordre $1-\frac{\alpha}{2}$, on a

$$\mathbb{P}(|Z_n| \leqslant z_{1-\alpha/2}) = \mathbb{P}\left(|\overline{X}_n - \lambda| \leqslant z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

Par suite,

$$\left[\overline{X}_n - z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}, \overline{X}_n + z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}\right]$$

est donc un intervalle de confiance asymptotique pour λ au niveau de confiance $1-\alpha$.

Exercice 13

D'après la loi des grands nombres appliquée à la suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, admettant même espérance $\mathbb{E}(X_n) = \frac{1}{p}$ et même variance $\mathbb{V}(X_n) = \frac{q}{p^2}$, la suite de terme général

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad n \geqslant 1,$$

converge en probabilité vers $\frac{1}{p}$. Ces variables sont à valeurs dans $]0, +\infty[$ où la fonction $x \longmapsto \frac{1}{x}$ est continue. Par conséquent, la suite $(Y_n) = \left(\frac{1}{\overline{X}_n}\right)$ converge en probabilité vers p et, par un argument similaire,

$$\sqrt{\frac{q}{1-Y_n}} \xrightarrow[n\to\infty]{\mathcal{P}} \sqrt{\frac{q}{1-p}} = 1.$$

D'un autre côté, le théorème limite central appliqué à la suite (X_n) de variables aléatoires indépendantes et identiquement distribuées admettant un moment d'ordre 2 assure que

$$\overline{X}_n^* = \frac{\overline{X}_n - \frac{1}{p}}{\sqrt{\frac{q}{np^2}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Après avoir vérifié sans peine que

$$T_n = \sqrt{n}\frac{p - Y_n}{Y_n\sqrt{1-Y_n}} = \sqrt{\frac{q}{1-Y_n}}\overline{X}_n^*,$$

on déduit du lemme de Slutzky que la suite (T_n) converge en loi vers la loi $\mathcal{N}(0,1)$.

Par conséquent, si $z_{1-\alpha/2}$ est le quantile de la loi normale centrée réduite à l'ordre $1-\frac{\alpha}{2}$, on a

$$\mathbb{P}(|T_n| \leqslant z_{1-\alpha/2}) = \mathbb{P}\left(|Y_n - p| \leqslant z_{1-\alpha/2}\frac{Y_n\sqrt{1-Y_n}}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

Par suite,

$$\left[Y_n - z_{1-\alpha/2}\frac{Y_n\sqrt{1-Y_n}}{\sqrt{n}}, Y_n + z_{1-\alpha/2}\frac{Y_n\sqrt{1-Y_n}}{\sqrt{n}}\right]$$

est donc un intervalle de confiance asymptotique pour p au niveau de confiance $1-\alpha$.