

isque

- EXERCICE 1 -

Dans cet exercice R désigne un réel strictement positif fixé.

On considère une variable aléatoire réelle X de densité f définie par : $f(t) = \begin{cases} 0 & \text{si } t \notin [0; R] \\ \frac{2t}{R^2} & \text{si } t \in [0; R] \end{cases}$.

Soit $n \in \mathbb{N}^*$. On considère $(X_1, X_2, \dots, X_n)$ un n -échantillon de X . On cherche à estimer le réel R .

1. Déterminer la fonction de répartition de X , notée F_X .

Pour tout réel x , on a : $F_X(x) = \int_{-\infty}^x f(t) dt$.

- Si $x < 0$, on a : $F_X(x) = \int_{-\infty}^x 0 dt = 0$.
- Si $x \in [0; R]$, on a : $F_X(x) = \int_{-\infty}^x f(t) dt = \int_0^x \frac{2t}{R^2} dt = \left[\frac{t^2}{R^2} \right]_0^x = \frac{x^2}{R^2}$.
- Si $x > R$, on a : $F_X(x) = \int_{-\infty}^x f(t) dt = 1$ car $X(\Omega) = [0; R]$.

Ainsi, on a :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{R^2} & \text{si } x \in [0; R] \\ 1 & \text{si } x > R \end{cases}$$

2. Montrer que X admet une espérance et une variance et que $E(X) = \frac{2R}{3}$ et $V(X) = \frac{R^2}{18}$.

On peut justifier l'existence de l'espérance et de la variance de X par le fait que son support est borné : $X(\Omega) = [0; R]$.

Sinon, on utilise la rédaction usuelle.

X admet une espérance si et seulement si $\int_{-\infty}^{+\infty} t f(t) dt = \int_0^R \frac{2t^2}{R^2} dt$ converge absolument ce qui est le cas.

X admet un moment d'ordre 2 si et seulement si $\int_{-\infty}^{+\infty} t^2 f(t) dt = \int_0^R \frac{2t^3}{R^2} dt$ converge absolument ce qui est le cas.

On a alors :

$$E(X) = \int_0^R \frac{2t^2}{R^2} dt = \left[\frac{2t^3}{3R^2} \right]_0^R = \frac{2R}{3}$$

et

$$E(X^2) = \int_0^R \frac{2t^3}{R^2} dt = \left[\frac{2t^4}{4R^2} \right]_0^R = \frac{2R^2}{4} = \frac{R^2}{2}$$

$$\text{et donc } V(X) = E(X^2) - E(X)^2 = \frac{R^2}{2} - \left(\frac{2R}{3} \right)^2 = \frac{9R^2 - 8R^2}{18} = \frac{R^2}{18}.$$

3. On note $\overline{X}_n$ la moyenne empirique de X et on considère la variable aléatoire $T_n = \frac{3}{2} \overline{X}_n$.

(a) Montrer que T_n est un estimateur de R .

$T_n = \frac{3 \sum_{i=1}^n X_i}{2n}$ est une fonction de $(X_1, \dots, X_n)$ qui ne dépend pas de R donc c'est un estimateur de R .

(b) Calculer $E(T_n)$ et $V(T_n)$.

On a par linéarité de l'espérance :

$$E(T_n) = \frac{3}{2} E(\overline{X}_n) = \frac{3}{2} \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{3}{2n} n \frac{2R}{3} = R$$

On a :

$$\begin{aligned} V(T_n) &= V\left(\frac{3\overline{X}_n}{2}\right) \\ &= \frac{9}{4} V(\overline{X}_n) \\ &= \frac{9}{4} V\left(\frac{\sum_{i=1}^n X_i}{n}\right) \\ &= \frac{9}{4n^2} \sum_{i=1}^n V(X_i) \quad \text{par indépendance} \\ &= \frac{9}{4n^2} n \frac{R^2}{18} = \frac{R^2}{8n} \end{aligned}$$

(c) Montrer à l'aide de l'inégalité de Bienaymé-Tchebychev que T_n est un estimateur convergent de R .

Comme T_n admet un moment d'ordre 2 avec $E(T_n) = R$ et $V(T_n) = \frac{R^2}{8n}$, on a d'après BT :

$$\forall \epsilon > 0, \quad P(|T_n - E(T_n)| \geq \epsilon) \leq \frac{V(T_n)}{\epsilon^2}.$$

soit

$$\forall \epsilon > 0, \quad 0 \leq P(|T_n - R| \geq \epsilon) \leq \frac{R^2}{8n\epsilon^2}$$

Donc $\lim_{n \rightarrow +\infty} P(|T_n - R| \geq \epsilon) = 0$ d'après le théorème des gendarmes, ce qui montre que T_n est un estimateur convergent de R .

- EXERCICE 2 -

Pour tout entier $m \geq 1$, on considère un échantillon $(Y_1, \dots, Y_m)$ de la loi de Bernoulli de paramètre p .

On rappelle que $\overline{Y}_m = \frac{1}{m} \sum_{i=1}^m Y_i$.

1. A l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que : $P(\overline{Y}_m - \epsilon \leq p \leq \overline{Y}_m + \epsilon) \geq 1 - \frac{1}{4m\epsilon^2}$.

Comme $\overline{Y}_m$ admet un moment d'ordre 2 et que $E(\overline{Y}_m) = p$ et $V(\overline{Y}_m) = \frac{p(1-p)}{n}$ (calcul classique), on a d'après l'inégalité de BT : $\forall \epsilon > 0$,

$$\begin{aligned} P(|\overline{Y}_m - p| \geq \epsilon) &\leq \frac{p(1-p)}{n\epsilon^2} \\ \Leftrightarrow 1 - P(|\overline{Y}_m - p| < \epsilon) &\leq \frac{p(1-p)}{n\epsilon^2} \quad \text{en considérant l'évènement contraire} \\ \Leftrightarrow P(|\overline{Y}_m - p| < \epsilon) &\geq 1 - \frac{p(1-p)}{n\epsilon^2} \\ \Leftrightarrow P(-\epsilon < \overline{Y}_m - p < \epsilon) &\geq 1 - \frac{p(1-p)}{n\epsilon^2} \\ \Leftrightarrow P(\overline{Y}_m - \epsilon < p < \overline{Y}_m + \epsilon) &\geq 1 - \frac{p(1-p)}{n\epsilon^2} \\ \Leftrightarrow P(\overline{Y}_m - \epsilon \leq p \leq \overline{Y}_m + \epsilon) &\geq 1 - \frac{p(1-p)}{n\epsilon^2} \end{aligned}$$

car $[\overline{Y}_m - \epsilon < p < \overline{Y}_m + \epsilon] \subset [\overline{Y}_m - \epsilon \leq p \leq \overline{Y}_m + \epsilon]$ Une étude rapide montre que $\forall p \in [0, 1], \quad p(1-p) \leq 1/4$.

On a alors :

$$\forall \epsilon > 0, \quad P(\overline{Y}_m - \epsilon \leq p \leq \overline{Y}_m + \epsilon) \geq 1 - \frac{1}{4m\epsilon^2}$$

2. En déduire que $\left[\overline{Y}_m - \sqrt{\frac{5}{m}}; \overline{Y}_m + \sqrt{\frac{5}{m}} \right]$ est un intervalle de confiance de p au niveau de confiance 0,95.

Puisqu'on nous donne l'intervalle, il suffit de vérifier qu'il convient en utilisant la question précédente. On

choisit $\epsilon = \sqrt{\frac{5}{m}}$ dans la question précédente, ce qui donne :

$$P\left(\bar{Y}_m - \sqrt{\frac{5}{m}} \leq p \leq \bar{Y}_m + \sqrt{\frac{5}{m}}\right) \geq 1 - \frac{1}{4m\left(\sqrt{\frac{5}{m}}\right)^2}$$

soit

$$P\left(\bar{Y}_m - \sqrt{\frac{5}{m}} \leq p \leq \bar{Y}_m + \sqrt{\frac{5}{m}}\right) \geq 1 - \frac{1}{20} = 0,95$$

ce qui permet de conclure.

- EXERCICE 3 -

Soit a un réel strictement positif et X une variable aléatoire vérifiant $E(X^2) = 2a$ et $V(X^2) = 4a^2$.

On considère n variables aléatoires $(X_1, X_2, \dots, X_n)$ indépendantes et suivant toutes la même loi que X .

On note S_n la variable aléatoire définie par : $S_n = \frac{1}{2n} \sum_{k=1}^n X_k^2$

1. Justifier que S_n est un estimateur de a .

La variable aléatoire S_n est une fonction dépendant des variables aléatoires $(X_1, X_2, \dots, X_n)$ mais indépendante de a donc S_n est un estimateur de a .

2. A l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que S_n est un estimateur convergent de a .

S_n admet une espérance comme somme finie de variables aléatoires en admettant une et on a :

$$\begin{aligned} E(S_n) &= E\left(\frac{1}{2n} \sum_{k=1}^n X_k^2\right) = \frac{1}{2n} \sum_{k=1}^n E(X_k^2) \quad \text{par linéarité de l'espérance} \\ &= \frac{1}{2n} 2na = a \quad \text{car même loi donc } E(X_k^2) = E(X^2) = 2a \end{aligned}$$

S_n admet une variance comme somme finie de variables aléatoires en admettant une et on a :

$$\begin{aligned} V(S_n) &= V\left(\frac{1}{2n} \sum_{k=1}^n X_k^2\right) = \frac{1}{4n^2} \sum_{k=1}^n V(X_k^2) \quad \text{par indépendance} \\ &= \frac{1}{4n^2} 4na^2 = \frac{a^2}{n} \quad \text{car même loi donc } V(X_k^2) = V(X^2) = 4a^2 \end{aligned}$$

On a alors d'après l'inégalité de BT :

$$\forall \epsilon > 0, \quad P(|S_n - E(S_n)| \geq \epsilon) \leq \frac{V(S_n)}{\epsilon^2} \quad \text{soit} \quad \forall \epsilon > 0, \quad 0 \leq P(|S_n - a| \geq \epsilon) \leq \frac{a^2}{n\epsilon^2}$$

On a : $\lim_{n \rightarrow +\infty} \frac{a^2}{n\epsilon^2} = 0$ donc d'après le théorème des gendarmes (avec $P(|S_n - a| \geq \epsilon) \geq 0$), on a : $\forall \epsilon > 0 \quad \lim_{n \rightarrow +\infty} P(|S_n - a| \geq \epsilon) = 0$ ce qui montre que S_n est un estimateur convergent de a .

3. On suppose que a est inférieur ou égal à 1.

(a) **Montrer que :** $\forall \epsilon > 0, P(|S_n - a| \leq \epsilon) \geq 1 - \frac{1}{n\epsilon^2}$.

On a obtenu à la question précédente que :

$$\forall \epsilon > 0, \quad P(|S_n - a| \geq \epsilon) \leq \frac{a^2}{n\epsilon^2}$$

ce qui donne avec l'hypothèse $a \leq 1$:

$$\forall \epsilon > 0, \quad P(|S_n - a| \geq \epsilon) \leq \frac{1}{n\epsilon^2}$$

En considérant l'évènement contraire, on a alors :

$$\forall \epsilon > 0, \quad 1 - P(|S_n - a| < \epsilon) \leq \frac{1}{n\epsilon^2}$$

ce qui donne

$$\forall \epsilon > 0, \quad P(|S_n - a| < \epsilon) \geq 1 - \frac{1}{n\epsilon^2}$$

Enfin, en remarquant que : $(S_n - a < \epsilon) \subset (|S_n - a| < \epsilon)$, on obtient le résultat demandé :

$$\forall \epsilon > 0, \quad P(|S_n - a| \leq \epsilon) \geq P(|S_n - a| < \epsilon) \geq 1 - \frac{13}{n\epsilon^2}$$

(b) **Déterminer les valeurs de n pour laquelle $\left[S_n - \frac{1}{10}; S_n + \frac{1}{10}\right]$ est un intervalle de confiance de a au niveau de confiance au moins égal à 95%.**

On cherche n tel que

$$P\left(S_n - \frac{1}{10} \leq a \leq S_n + \frac{1}{10}\right) \geq 0,95$$

Or, on a

$$\begin{aligned} P(|S_n - a| \leq \epsilon) \geq 1 - \frac{1}{n\epsilon^2} &\iff P(-\epsilon \leq S_n - a \leq \epsilon) \geq 1 - \frac{1}{n\epsilon^2} \\ &\iff P(S_n - \epsilon \leq a \leq S_n + \epsilon) \geq 1 - \frac{1}{n\epsilon^2} \end{aligned}$$

Il s'agit donc de choisir $\epsilon = \frac{1}{10}$ et de chercher n tel que $1 - \frac{1}{n\epsilon^2} \geq 0,95$. On a alors :

$$1 - \frac{1}{n\left(\frac{1}{10}\right)^2} \geq 0,95 \iff \frac{100}{n} \leq 0,05 \iff n \geq \frac{100}{0,05} = 2000$$

Ainsi, pour tout $n \geq 2000$ l'intervalle $\left[S_n - \frac{1}{10}; S_n + \frac{1}{10}\right]$ est un intervalle de confiance de a au niveau de confiance au moins égal à 95%.

- EXERCICE 4 -

Soit a un réel strictement positif et f la fonction définie par : $f : x \mapsto \begin{cases} 0 & \text{si } x < 1 \\ \frac{a}{x^{a+1}} & \text{si } x \geq 1 \end{cases}$.

On admet que f est une densité de probabilité et on note X une variable aléatoire admettant f pour densité. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes, toutes de même loi que X .

1. Déterminer la fonction de répartition F_X de X .

Comme f est nulle sur $]-\infty, 1[$, on a : $X(\Omega) = [1, +\infty[$.

• Pour tout $t < 1$, on a : $F_X(t) = 0$.

• Pour tout $t \geq 1$, on a :

$$F_X(t) = \int_{-\infty}^t \frac{a}{x^{a+1}} dx = \int_1^t \frac{a}{x^{a+1}} dx = \int_1^t ax^{-(a+1)} dx = [-x^{-a}]_1^t = (-t^{-a} + 1^{-a}) = 1 - t^{-a}$$

Ainsi, on a : $F_X(t) = \begin{cases} 0 & \text{si } t < 1 \\ 1 - t^{-a} & \text{si } t \geq 1 \end{cases}$

2. Soit $n \in \mathbb{N}^*$. On pose : $W_n = \ln(X_n)$.

Montrer que la variable aléatoire W_n suit une loi exponentielle de paramètre a .

En déduire l'espérance et la variance de W_n .

Comme $X_n(\Omega) = [1, +\infty[$, on a clairement $W_n(\Omega) = \ln(X_n)(\Omega) = \mathbb{R}_+$.

• Pour tout $t < 0$, on a : $F_{W_n}(t) = 0$.

• Pour tout $t \geq 0$, on a :

$$\begin{aligned} P(W_n \leq t) &= P(\ln(X_n) \leq t) \\ &= P(X_n \leq e^t) \quad \text{car exp est croissante sur } \mathbb{R} \\ &= F_{X_n}(e^t) \\ &= 1 - (e^t)^{-a} \quad \text{car } e^t \geq 1 \\ &= 1 - e^{-at} \end{aligned}$$

Ainsi, on a : $F_{W_n}(t) = \begin{cases} 1 - e^{-at} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$ donc W_n suit la loi exponentielle de paramètre a .

3. On définit, pour tout n de $\mathbb{N}^*$: $M_n = \frac{\ln(X_1) + \dots + \ln(X_n)}{n}$ et $T_n = \sqrt{n}(aM_n - 1)$.

(a) Montrer que la variable aléatoire M_n admet une espérance et une variance et les calculer.

Remarquons que $M_n = \frac{W_1 + \dots + W_n}{n}$ avec les W_i qui suivent la loi exponentielle de paramètre a .

Comme les variables W_i admettent une espérance et une variance, la variable M_n admet également une espérance et une variance et on a :

$$\begin{aligned} E(M_n) &= E\left(\frac{1}{n} \sum_{k=1}^n W_k\right) \\ &= \frac{1}{n} \sum_{k=1}^n E(W_k) \quad \text{par linéarité de l'espérance} \\ &= \frac{1}{n} n \frac{1}{a} = \frac{1}{a} \quad \text{car ce sont des lois exponentielles de paramètre } a \end{aligned}$$

et

$$\begin{aligned} V(M_n) &= V\left(\frac{1}{n} \sum_{k=1}^n W_k\right) \\ &= \frac{1}{n^2} \sum_{k=1}^n V(W_k) \quad \text{par indépendance d'après le lemme des coalitions} \\ &= \frac{1}{n^2} n \frac{1}{a^2} = \frac{1}{na^2} \quad \text{car ce sont des lois exponentielles de paramètre } a \end{aligned}$$

(b) Justifier que la suite de variables aléatoires $(T_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite.

On va appliquer le théorème central limite.

On remarque que $M_n = \frac{W_1 + \dots + W_n}{n}$ est une moyenne de variables aléatoires indépendantes (lemme des coalitions) et de même loi, de même espérance et de même variance non nulle.

Ainsi, d'après le théorème central limite, la variable M_n^* converge en loi vers une variable Z suivant une loi normale centrée réduite.

Or, on a d'après la question précédente :

$$M_n^* = \frac{M_n - E(M_n)}{\sqrt{V(M_n)}} = \frac{M_n - \frac{1}{a}}{\sqrt{\frac{1}{na^2}}} = \frac{aM_n - 1}{\frac{1}{\sqrt{na}}} = \sqrt{n}(aM_n - 1) = T_n$$

Ainsi, $M_n^* = T_n$ converge en loi vers une variable Z suivant une loi normale centrée réduite.

(c) En déduire que l'intervalle $\left[\frac{\sqrt{n}-2}{\sqrt{n}M_n}; \frac{\sqrt{n}+2}{\sqrt{n}M_n}\right]$ est un intervalle de confiance asymptotique pour a au niveau de confiance 95%.

On admettra que $\Phi(2) \geq 0,975$, où Φ désigne la fonction de répartition d'une variable aléatoire suivant la loi normale centrée réduite.

Comme l'intervalle est donné, on va vérifier qu'il fonctionne.

Il s'agit de montrer que : $\lim_{n \rightarrow +\infty} P\left(\frac{\sqrt{n}-2}{\sqrt{n}M_n} \leq a \leq \frac{\sqrt{n}+2}{\sqrt{n}M_n}\right) \geq 0,95$.

On a :

$$\begin{aligned} P\left(\frac{\sqrt{n}-2}{\sqrt{n}M_n} \leq a \leq \frac{\sqrt{n}+2}{\sqrt{n}M_n}\right) &= P(\sqrt{n}-2 \leq a\sqrt{n}M_n \leq \sqrt{n}+2) \quad \text{car } \sqrt{n}M_n \geq 0 \\ &= P(-2 \leq a\sqrt{n}M_n - \sqrt{n} \leq 2) \\ &= P(-2 \leq \sqrt{n}(aM_n - 1) \leq 2) \\ &= P(-2 \leq T_n \leq 2) \\ &\xrightarrow{n \rightarrow +\infty} P(-2 \leq Z \leq 2) \quad \text{d'après le TLC avec } Z \mapsto \mathcal{N}(0,1) \\ &= \Phi(2) - \Phi(-2) \\ &= 2\Phi(2) - 1 \\ &\geq 2 \cdot 0,975 - 1 = 0,95 \quad \text{d'après l'énoncé} \end{aligned}$$

Ainsi, l'intervalle $\left[\frac{\sqrt{n}-2}{\sqrt{n}M_n}; \frac{\sqrt{n}+2}{\sqrt{n}M_n}\right]$ est un intervalle de confiance asymptotique pour a au niveau de confiance 95%.

- EXERCICE 5 -

Soit $(X_1, \dots, X_n)$ un n -échantillon i.i.d de loi parente la loi normale $\mathcal{N}(m, \sigma^2)$ de variance σ^2 connue. i.i.d signifie indépendantes et identiquement distribuées.

On cherche un intervalle de confiance pour l'espérance m et on considère pour cela l'estimateur $T_n = \overline{X_n}$.

1. Calculer l'espérance et la variance de T_n .

On rappelle que $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$. T_n est un estimateur de m car c'est une fonction des X_i indépendante de m .

On a par linéarité de l'espérance :

$$E(T_n) = E(\overline{X_n}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{nm}{n} = m$$

et par indépendance des variables :

$$V(\overline{X_n}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{n\sigma^2}{n} = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow +\infty} 0$$

2. Utilisation de l'inégalité de BT.

(a) Montrer, à l'aide de l'inégalité de BT, que pour tout $t > 0$, on a : $P(T_n - t \leq m \leq T_n + t) \geq 1 - \frac{\sigma^2}{nt^2}$.
 T_n admet un moment d'ordre 2 d'après la question précédente donc d'après l'inégalité de BT, on a :

$$\forall \epsilon > 0, \quad P(|T_n - E(T_n)| \geq \epsilon) \leq \frac{V(T_n)}{\epsilon^2}$$

ce qui donne avec les calculs précédents et en prenant $\epsilon = t$:

$$P(|T_n - m| \geq t) \leq \frac{\sigma^2}{nt^2}$$

On cherche maintenant à isoler m :

$$\begin{aligned} P(|T_n - m| \geq t) &\leq \frac{\sigma^2}{nt^2} \\ \Leftrightarrow 1 - P(|T_n - m| < t) &\leq \frac{\sigma^2}{nt^2} \quad \text{en passant au complémentaire} \\ \Leftrightarrow 1 - P(-t < T_n - m < t) &\leq \frac{\sigma^2}{nt^2} \\ \Leftrightarrow 1 - P(-t < m - T_n < t) &\leq \frac{\sigma^2}{nt^2} \quad \text{en multipliant dans la double inégalité par } -1 \\ \Leftrightarrow 1 - P(T_n - t < m < T_n + t) &\leq \frac{\sigma^2}{nt^2} \\ \Leftrightarrow P(T_n - t \leq m \leq T_n + t) &\geq 1 - \frac{\sigma^2}{nt^2} \end{aligned}$$

- (b) **En déduire que l'intervalle** $I_n(\alpha) = \left[T_n - \frac{\sigma}{\sqrt{n\alpha}}; T_n + \frac{\sigma}{\sqrt{n\alpha}} \right]$ **est un intervalle de confiance de m au niveau de confiance $1 - \alpha$.**
 Il s'agit de trouver t tel que : $P(T_n - t \leq m \leq T_n + t) \geq 1 - \alpha$ ce qui est réalisé d'après la question précédente si $1 - \frac{\sigma^2}{nt^2} = 1 - \alpha$. On obtient alors :

$$1 - \frac{\sigma^2}{nt^2} = 1 - \alpha \iff \frac{\sigma^2}{nt^2} = \alpha \iff \frac{\sigma^2}{n\alpha} = t^2 \iff \frac{\sigma}{\sqrt{n\alpha}} = t$$

Ainsi, pour $t = \frac{\sigma}{\sqrt{n\alpha}}$, on obtient :

$$P\left(T_n - \frac{\sigma}{\sqrt{n\alpha}} \leq m \leq T_n + \frac{\sigma}{\sqrt{n\alpha}}\right) \geq 1 - \alpha$$

ce qui montre que l'intervalle $I_n(\alpha) = \left[T_n - \frac{\sigma}{\sqrt{n\alpha}}; T_n + \frac{\sigma}{\sqrt{n\alpha}} \right]$ est un intervalle de confiance de m au niveau de confiance $1 - \alpha$.

3. Utilisation de la loi normale centrée réduite.

- (a) **Montrer que T_n suit la loi normale $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$.**

Les variables X_i étant indépendantes, on en déduit par stabilité des lois normales que $T_n = \frac{1}{n} \sum_{i=1}^n X_i$

suit également une loi normale. De plus, on a vu que $E(T_n) = m$ et $V(T_n) = \frac{\sigma^2}{n}$ donc T_n suit la loi normale $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$.

- (b) **En déduire que pour tout $t > 0$:** $P\left(\left|\sqrt{\frac{n}{\sigma^2}}(T_n - m)\right| \leq t\right) = 2\Phi(t) - 1$.

Comme T_n suit la loi normale $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$, on a : $T_n^* = \frac{T_n - E(T_n)}{\sqrt{V(T_n)}}$ suit la loi normale $\mathcal{N}(0, 1)$.

Or, $T_n^* = \frac{T_n - m}{\sqrt{\frac{\sigma^2}{n}}} = \sqrt{\frac{n}{\sigma^2}}(T_n - m)$, ce qui donne :

$$\begin{aligned} P\left(\left|\sqrt{\frac{n}{\sigma^2}}(T_n - m)\right| \leq t\right) &= P(|T_n^*| \leq t) \\ &= P(-t \leq T_n^* \leq t) \\ &= \Phi(t) - \Phi(-t) \quad \text{car } T_n \hookrightarrow \mathcal{N}(0, 1) \\ &= 2\Phi(t) - 1 \end{aligned}$$

- (c) **Justifier que pour tout $\alpha \in]0, 1[$, il existe un réel t_α tel que :** $2\Phi(t_\alpha) - 1 = 1 - \alpha$.

On a :

$$2\Phi(t) - 1 = 1 - \alpha \iff 2\Phi(t) = 2 - \alpha \iff \Phi(t) = 1 - \frac{\alpha}{2}$$

Or Φ est une fonction de répartition continue et strictement croissante (car $\Phi' = \varphi > 0$) donc réalise une bijection de $\mathbb{R}$ sur $]0, 1[$.

Comme $\alpha \in]0, 1[$, on a : $1 - \frac{\alpha}{2} \in]1/2, 1[$ donc $1 - \frac{\alpha}{2} \in]0, 1[$ ce qui permet de conclure que l'équation

$\Phi(t) = 1 - \frac{\alpha}{2}$ admet une unique solution $t_\alpha \in \mathbb{R}$ (dans $\mathbb{R}_+$ en fait).

- (d) **En déduire que l'intervalle** $J_n(\alpha) = \left[T_n - \frac{t_\alpha \sigma}{\sqrt{n}}; T_n + \frac{t_\alpha \sigma}{\sqrt{n}} \right]$ **est un intervalle de confiance de m au niveau de confiance $1 - \alpha$.**

Posons $t = t_\alpha$ dans l'équation $P\left(\left|\sqrt{\frac{n}{\sigma^2}}(T_n - m)\right| \leq t\right) = 2\Phi(t) - 1$, ce qui donne :

$$P\left(\left|\sqrt{\frac{n}{\sigma^2}}(T_n - m)\right| \leq t_\alpha\right) = 2\Phi(t_\alpha) - 1$$

$$\iff P\left(\left|\sqrt{\frac{n}{\sigma^2}}(T_n - m)\right| \leq t_\alpha\right) = 1 - \alpha \quad \text{d'après la question précédente}$$

On cherche maintenant à isoler m .

$$P\left(\left|\sqrt{\frac{n}{\sigma^2}}(T_n - m)\right| \leq t_\alpha\right) = 1 - \alpha$$

$$\iff P\left(-t_\alpha \leq \sqrt{\frac{n}{\sigma^2}}(T_n - m) \leq t_\alpha\right) = 1 - \alpha$$

$$\iff P\left(-\sqrt{\frac{\sigma^2}{n}}t_\alpha \leq T_n - m \leq \sqrt{\frac{\sigma^2}{n}}t_\alpha\right) = 1 - \alpha$$

$$\iff P\left(-\frac{\sigma}{\sqrt{n}}t_\alpha \leq m - T_n \leq \frac{\sigma}{\sqrt{n}}t_\alpha\right) = 1 - \alpha \quad \text{en multipliant dans la double inégalité par } -1$$

$$\iff P\left(T_n - \frac{\sigma}{\sqrt{n}}t_\alpha \leq m \leq T_n + \frac{\sigma}{\sqrt{n}}t_\alpha\right) = 1 - \alpha$$

Ainsi, l'intervalle $J_n(\alpha) = \left[T_n - \frac{t_\alpha \sigma}{\sqrt{n}}; T_n + \frac{t_\alpha \sigma}{\sqrt{n}} \right]$ est un intervalle de confiance de m au niveau de confiance $1 - \alpha$.

4. (a) **Déterminer l'étendue des intervalles $I_n(\alpha)$ et $J_n(\alpha)$ (i.e leur longueur).**

• L'étendue de $I_n(\alpha) = \left[T_n - \frac{\sigma}{\sqrt{n\alpha}}; T_n + \frac{\sigma}{\sqrt{n\alpha}} \right]$ notée ℓ_I est :

$$\ell_I = T_n + \frac{\sigma}{\sqrt{n\alpha}} - \left(T_n - \frac{\sigma}{\sqrt{n\alpha}}\right) \text{ soit } \ell_I = 2\frac{\sigma}{\sqrt{n\alpha}}.$$

• L'étendue de $J_n(\alpha) = \left[T_n - \frac{t_\alpha \sigma}{\sqrt{n}}; T_n + \frac{t_\alpha \sigma}{\sqrt{n}} \right]$ notée ℓ_J est :

$$\ell_J = T_n + \frac{t_\alpha \sigma}{\sqrt{n}} - \left(T_n - \frac{t_\alpha \sigma}{\sqrt{n}}\right) \text{ soit } \ell_J = 2\frac{t_\alpha \sigma}{\sqrt{n}}.$$

- (b) **On donne les valeurs numériques de t_α et $\frac{1}{\sqrt{\alpha}}$ suivantes pour différentes valeurs de α .**

α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$1/\sqrt{\alpha}$	10.	7.071	5.773	5.	4.472	4.082	3.779	3.535	3.333	3.162
t_α	2.575	2.326	2.170	2.053	1.959	1.880	1.811	1.750	1.695	1.644

Au vu de ces résultats, déterminer quel est l'intervalle le plus précis entre $I_n(\alpha)$ et $J_n(\alpha)$?

On remarque que : $\ell_I = K \frac{1}{\sqrt{\alpha}}$ et $\ell_J = K t_\alpha$ avec $K = \frac{2\sigma}{\sqrt{n}}$ donc les étendues des intervalles de confiance

dépendent de $\frac{1}{\sqrt{\alpha}}$ et de t_α .

Les résultats indiquent donc que le deuxième intervalle $J_n(\alpha)$ est bien plus précis.

5. **On suppose que le temps de travail hebdomadaire d'un étudiant en classe prépa est une variable aléatoire suivant la loi normale $\mathcal{N}(m, 100)$ de moyenne m inconnue que l'on cherche à estimer.**

Une enquête réalisée auprès de 100 étudiants donne un temps de travail hebdomadaire moyen de 29.32h.

- (a) **Déterminer l'intervalle de confiance $J_n(0.05)$ au niveau de confiance $\alpha = 0.05$ pour la moyenne m .**

Dans cet exemple, on a : $\sigma^2 = 100$ (donc $\sigma = 10$), $n = 100$, $\alpha = 0.05$ et comme estimation de $T_n : T_{100} = 29.32$.

On commence par chercher $t_\alpha = t_{0.05}$ solution de $\Phi(t) = 1 - \frac{0.05}{2} = 0.975$, ce qui donne avec la table de la loi normale $t_{0.05} = 1.96$.

On obtient alors :

$$J_n(0.05) = \left[T_{100} - \frac{t_{0.05}\sigma}{\sqrt{n}}; T_{100} + \frac{t_{0.05}\sigma}{\sqrt{n}} \right] = \left[29.32 - \frac{1.96 \times 10}{10}; 29.32 + \frac{1.96 \times 10}{10} \right] = [27.36; 31.28]$$

- (b) **Quelle valeur de n faut-il choisir pour que la longueur de l'intervalle de confiance au niveau de confiance $\alpha = 0.05$ soit inférieure à 2 ?**

On cherche n tel que $\ell_j = 2 \frac{t_{\alpha} \sigma}{\sqrt{n}} \leq 2$ soit avec les valeurs numériques trouvées à la question précédente :

$$2 \frac{t_{\alpha} \sigma}{\sqrt{n}} \leq 2 \iff 2 \frac{1.96 \times 10}{\sqrt{n}} \leq 2 \iff 19.6 \leq \sqrt{n} \iff 19.6^2 \leq n \iff 384.16 \leq n$$

Ainsi, il faut choisir $n = 385$.

Pour aller plus loin...

- EXERCICE 6 -

Soit $(X_1, \dots, X_n)$ un n -échantillon d'une variable aléatoire X admettant pour espérance m et pour variance σ^2 .

1. **On suppose que m est connu.**

Montrer que $T_n = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ est un estimateur de σ^2 et calculer $E(T_n)$.

T_n est une fonction de $(X_1, \dots, X_n)$ qui ne dépend pas de σ^2 donc c'est un estimateur de σ^2 .

On a par linéarité de l'espérance :

$$\begin{aligned} E(T_n) &= \frac{1}{n} \sum_{i=1}^n E((X_i - m)^2) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i^2 - 2mX_i + m^2) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i^2) - 2mE(X_i) + m^2 \end{aligned}$$

Or, on a : $E(X_i^2) = V(X_i) + E(X_i)^2 = \sigma^2 + m^2$ et $E(X_i) = m$, ce qui donne :

$$E(T_n) = \frac{1}{n} \sum_{i=1}^n \sigma^2 + m^2 - 2m^2 + m^2 = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2$$

2. **On suppose que m n'est pas connu.**

On note $\overline{X}_n$ la moyenne empirique de l'échantillon et $U_n = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2$.

- (a) **Montrer que, pour tout $i \in \llbracket 1; n \rrbracket$: $E\left((X_i - \overline{X}_n)^2\right) = V(X_i - \overline{X}_n)$.**

$$\text{On a : } V(X_i - \overline{X}_n) = E\left((X_i - \overline{X}_n)^2\right) - E(X_i - \overline{X}_n)^2.$$

Or, on a par linéarité de l'espérance : $E(X_i - \overline{X}_n) = E(X_i) - E(\overline{X}_n) = m - m = 0$ un calcul classique (à refaire si besoin) donnant $E(\overline{X}_n) = m$.

$$\text{On a donc : } E\left((X_i - \overline{X}_n)^2\right) = V(X_i - \overline{X}_n).$$

- (b) **Montrer que, pour tout $i \in \llbracket 1; n \rrbracket$: $V(X_i - \overline{X}_n) = \left(1 - \frac{1}{n}\right)^2 V(X_i) + \frac{1}{n^2} \sum_{k \neq i} V(X_k)$.**

$$\text{Comme } \overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k, \text{ on a : } X_i - \overline{X}_n = X_i - \frac{1}{n} \sum_{k=1}^n X_k = X_i - \frac{1}{n} X_i - \frac{1}{n} \sum_{k \neq i} X_k = \left(1 - \frac{1}{n}\right) X_i - \frac{1}{n} \sum_{k \neq i} X_k$$

ce qui donne :

$$V(X_i - \overline{X}_n) = V\left(\left(1 - \frac{1}{n}\right) X_i + (-1) \frac{1}{n} \sum_{k \neq i} X_k\right)$$

car les variables X_i et $\sum_{k=1}^n X_k$ sont indépendantes par le lemme des caolitions. On a alors

$$\begin{aligned} V(X_i - \overline{X}_n) &= V\left(\left(1 - \frac{1}{n}\right) X_i + \frac{1}{n} \sum_{k \neq i} X_k\right) \\ &= \left(1 - \frac{1}{n}\right)^2 V(X_i) + \frac{1}{n^2} \sum_{k \neq i} V(X_k) \end{aligned}$$

par indépendance des X_k .

- (c) **En déduire que :** $V(X_i - \overline{X}_n) = \frac{n-1}{n} \sigma^2$.

Comme $V(X_i) = \sigma^2$ pour tout $i \in \llbracket 1; n \rrbracket$, on a :

$$V(X_i - \overline{X}_n) = \left(1 - \frac{1}{n}\right)^2 V(X_i) + \frac{1}{n^2} \sum_{k \neq i} V(X_k) = \left(\frac{n-1}{n}\right)^2 \sigma^2 + \frac{n-1}{n^2} \sigma^2 = \frac{(n-1)^2 + (n-1)}{n^2} \sigma^2 = \frac{n-1}{n} \sigma^2$$

- (d) **Montrer alors que U_n est un estimateur de σ^2 et calculer $E(U_n)$.**

U_n est une fonction de $(X_1, \dots, X_n)$ qui ne dépend pas de σ^2 donc c'est un estimateur de σ^2 .

On a par linéarité de l'espérance et en utilisant les questions précédentes :

$$\begin{aligned} E(U_n) &= \frac{1}{n} \sum_{i=1}^n E\left((X_i - \overline{X}_n)^2\right) \\ &= \frac{1}{n} \sum_{i=1}^n V(X_i - \overline{X}_n) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{n-1}{n} \sigma^2 \\ &= \frac{n-1}{n} \sigma^2 \end{aligned}$$

- (e) **En déduire un estimateur V_n de σ^2 qui vérifie $E(V_n) = \sigma^2$.**

Posons $V_n = \frac{n}{n-1} U_n$. V_n est une fonction de $(X_1, \dots, X_n)$ qui ne dépend pas de σ^2 donc c'est un estimateur de σ^2 .

On a d'après la question précédente :

$$E(V_n) = \frac{n}{n-1} E(U_n) = \frac{n}{n-1} \frac{n-1}{n} \sigma^2 = \sigma^2$$

- EXERCICE 7 -

Résultat préliminaire.

1. **Soit N un entier supérieur à 2 et Y une variable aléatoire à valeurs dans $\llbracket 1; N \rrbracket$.**

Pour $k \in \llbracket 1; N \rrbracket$, montrer que $P(Y = k) = P(Y \geq k) - P(Y \geq k+1)$.

En déduire que pour toute variable aléatoire Y à valeurs dans $\llbracket 1; N \rrbracket$ on a : $E(Y) = \sum_{k=1}^N P(Y \geq k)$.

- On a :

$$(Y \geq k) = (Y = k) \cup (Y \geq k+1)$$

donc par incompatibilité :

$$P(Y \geq k) = P(Y = k) + P(Y \geq k+1)$$

et donc : $P(Y = k) = P(Y \geq k) - P(Y \geq k+1)$.

- Comme $Y(\Omega) = \llbracket 1; N \rrbracket$, on a :

$$\begin{aligned}
 E(Y) &= \sum_{k=1}^N kP(Y=k) = \sum_{k=1}^N k(P(Y \geq k) - P(Y \geq k+1)) \\
 &= \sum_{k=1}^N kP(Y \geq k) - \sum_{k=1}^N kP(Y \geq k+1) \\
 &= \sum_{k=1}^N kP(Y \geq k) - \sum_{k=2}^{N+1} (k-1)P(Y \geq k) \\
 &= \sum_{k=1}^N kP(Y \geq k) - \sum_{k=2}^N (k-1)P(Y \geq k) \quad \text{car } Y(\Omega) = \llbracket 1; N \rrbracket \\
 &= \underbrace{\sum_{k=1}^N kP(Y \geq k) - \sum_{k=2}^N kP(Y \geq k)}_{=P(Y \geq 1) \text{ sommes télescopiques}} + \sum_{k=2}^N P(Y \geq k) \\
 &= P(Y \geq 1) + \sum_{k=2}^N P(Y \geq k) \\
 &= \sum_{k=1}^N P(Y \geq k)
 \end{aligned}$$

Ainsi, on a bien : $E(Y) = \sum_{k=1}^N P(Y \geq k)$.

Une urne contient N boules numérotées de 1 à N . On sait que $N \geq 2$ mais on ne connaît pas sa valeur exacte et on cherche à l'estimer. Pour cela, on effectue n tirages avec remise et on note, pour tout $k \in \llbracket 1; n \rrbracket$, Z_k le numéro de la boule obtenue au k -ième tirage. Ainsi, les variables $(Z_1, \dots, Z_n)$ sont mutuellement indépendantes.

2. **Pour tout $k \in \llbracket 1; n \rrbracket$, déterminer la loi de Z_k . Donner son espérance et sa variance.**

Les tirages se faisant avec remise et au hasard, Z_k suit une loi uniforme sur $\llbracket 1; N \rrbracket$. $Z_k \hookrightarrow \mathcal{U}(\llbracket 1; N \rrbracket)$ et on a :

$$E(Z_k) = \frac{N+1}{2} \text{ et } V(Z_k) = \frac{N^2-1}{12}.$$

3. **On pose $S_n = \max(Z_1, \dots, Z_n)$.**

- (a) **Calculer pour tout $k \in \llbracket 1; N \rrbracket$, $P(S_n \leq k)$.**

Commençons par remarquer que $P(Z_1 \leq k) = \frac{k}{N}$ par équiprobabilité (le numéro de la boule tirée est inférieur ou égal à k). On a classiquement :

$$\begin{aligned}
 P(S_n \leq k) &= P(\max(Z_1, \dots, Z_n) \leq k) = P((Z_1 \leq k) \cap \dots \cap (Z_n \leq k)) \\
 &= P(Z_1 \leq k) \times \dots \times P(Z_n \leq k) \quad \text{par indépendance} \\
 &= (P(Z_1 \leq k))^n \quad \text{car les variables suivent la même loi} \\
 &= \left(\frac{k}{N}\right)^n
 \end{aligned}$$

$$\text{Ainsi : } P(S_n \leq k) = \left(\frac{k}{N}\right)^n.$$

- (b) **Montrer à l'aide du résultat préliminaire que : $E(S_n) = N - \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n$.**

Comme $S_n(\Omega) = \llbracket 1; N \rrbracket$, on peut utiliser la formule précédente, ce qui donne :

$$E(S_n) = \sum_{k=1}^N P(S_n \geq k) = \sum_{k=1}^N 1 - P(S_n < k) = \sum_{k=1}^N 1 - P(S_n \leq k-1) = N - \sum_{k=0}^{N-1} P(S_n \leq k)$$

Ainsi, d'après le calcul de la question précédente, on a bien : $E(S_n) = N - \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n$.

- (c) **Montrer que pour tout $k \in \llbracket 0; N-1 \rrbracket$, on a : $\int_{k/N}^{(k+1)/N} x^n dx \geq \frac{1}{N} \left(\frac{k}{N}\right)^n$.**

La fonction $x \mapsto x^n$ est croissante sur $[0, 1]$. On a donc :

$$\begin{aligned}
 \frac{k}{N} &\leq x \leq \frac{k+1}{N} \\
 \Rightarrow \left(\frac{k}{N}\right)^n &\leq x^n \leq \left(\frac{k+1}{N}\right)^n \quad \text{car } t \mapsto t^n \text{ est croissante} \\
 \Rightarrow \int_{k/N}^{(k+1)/N} \left(\frac{k}{N}\right)^n dx &\leq \int_{k/N}^{(k+1)/N} x^n dx \leq \int_{k/N}^{(k+1)/N} \left(\frac{k+1}{N}\right)^n dx \quad \text{par croissance de l'intégrale avec } \frac{k+1}{N} > \frac{k}{N} \\
 \Rightarrow \frac{1}{N} \left(\frac{k}{N}\right)^n &\leq \int_{k/N}^{(k+1)/N} x^n dx
 \end{aligned}$$

- (d) **En déduire que : $\frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n \leq \int_0^1 x^n dx$.**

En sommant les inégalités précédentes entre $k=0$ et $k=N-1$, on obtient :

$$\begin{aligned}
 \sum_{k=0}^{N-1} \frac{1}{N} \left(\frac{k}{N}\right)^n &\leq \sum_{k=0}^{N-1} \int_{k/N}^{(k+1)/N} x^n dx \\
 \frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n &\leq \int_0^1 x^n dx \quad \text{d'après la relation de Chasles}
 \end{aligned}$$

- (e) **Montrer que : $E(S_n) \geq N - \frac{N}{n+1}$ puis en déduire que $\lim_{n \rightarrow +\infty} E(S_n) = N$.**

On a vu : $E(S_n) = N - \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n$.

Or, la question précédente donne :

$$\begin{aligned}
 \frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n &\leq \int_0^1 x^n dx \Leftrightarrow \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n \leq N \int_0^1 x^n dx \\
 &\Leftrightarrow \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n \leq N \left[\frac{x^{n+1}}{n+1} \right]_0^1 = \frac{N}{n+1} \\
 &\Leftrightarrow - \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n \geq - \frac{N}{n+1} \\
 &\Leftrightarrow E(S_n) = N - \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n \geq N - \frac{N}{n+1}
 \end{aligned}$$

Comme $S_n(\Omega) = \llbracket 1; N \rrbracket$, on a clairement : $E(S_n) \leq N$, ce qui donne avec la relation précédente :

$$N - \frac{N}{n+1} \leq E(S_n) \leq N$$

Enfin, comme $\lim_{n \rightarrow +\infty} N - \frac{N}{n+1} = \lim_{n \rightarrow +\infty} N = N$, on en déduit par le théorème des gendarmes que :

$$\lim_{n \rightarrow +\infty} E(S_n) = N.$$

On dit que S_n est un estimateur asymptotiquement sans biais de N .