Exercice | **PC et CC** | **ESCP 1999** | ★★ | **Un premier exemple à savoir par coeur**

Soit $a \in]0,1[$ et X une variable aléatoire, définie sur un espace probabilisé $(\Omega, \mathcal{B}, P)$, à valeurs dans $[0,1]$.
On suppose que la loi conditionnelle de X conditionnée par $[X \leqslant a]$ est la loi uniforme sur $[0,a]$ et que sa loi conditionnelle conditionnée par $[X > a]$ est la loi uniforme sur $[a,1]$.

On suppose de plus que $P[X \leqslant a] = \dfrac{1}{2}$.

Q1. Montrer que X est une variable aléatoire à densité et en trouver une densité.

Déterminer $E(X)$. Pour gagner du temps on admet que $V(X) = \dfrac{1}{48}\left(4\,a^2 - 4\,a + 5\right)$.

On cherche maintenant à estimer le paramètre inconnu a.

Q2. Soit $(X_1, \ldots, X_n)$ n variables aléatoires indépendantes de même loi que X (*je préfère soit $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires inépendantes de même loi que X*).

On pose $M_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$. Déterminer l'espérance $E(M_n)$ et la variance $V(M_n)$.
En déduire un estimateur sans biais T_n de a.
La suite d'estimateurs $(T_n)_{n \geqslant 1}$ est-elle convergente ?

Q3. Montrer que $V(T_n) \leqslant \dfrac{5}{12n} \cdot$ $n \in \mathbb{N}^*$, $\alpha \in]0,1[$ et ε est un réel strictement positif.

Donnés en utilisant l'inégalité de Bienaymé-Tchebychev, une condition suffisante sur ε pour que $[T_n - \varepsilon, T_n + \varepsilon]$ soit un intervalle de confiance au risque α pour a (c'est-à-dire : $P(a \in [T_n - \varepsilon, T_n + \varepsilon]) \geqslant 1 - \alpha$ ou $P(a \notin [T_n - \varepsilon, T_n + \varepsilon]) < \alpha$).

Quel intervalle de confiance au risque α retenez vous ?

On retiendra que Bienaymé-Tchebychev peut permettre de construire des intervalles de confiance... malheureusement souvent médiocres

Q4. Quelle est la limite en loi de la suite $\left(\dfrac{T_n - a}{\sqrt{V(T_n)}}\right)_{n \geqslant 1}$? (*faire propre, simple et juste*)

On suppose que n est suffisamment grand pour identifier la loi de $\dfrac{T_n - a}{\sqrt{V(T_n)}}$ à cette loi limite.

Trouver ε pour que $P\left(\left|\dfrac{T_n - a}{\sqrt{V(T_n)}}\right| \leqslant \varepsilon\right) \geqslant 0.95$ (rappel $0.975 = \Phi(1.96)$).

En déduire un intervalle de confiance pour a à un risque inférieur à 5%

En s'inspirant de ce qui précède donner un intervalle de confiance pour a à un risque inférieur à α (rappelons que Φ est une bijection de $\mathbb{R}$ sur $]0,1[$).

Cette démarche est à bien comprendre et à savoir faire par coeur.

On retiendra que le théorème de la limite centrée peut permettre de construire des intervalles de confiance... malheureusement, le plus souvent, à une approximation près

Q5. Au niveau de risque 5%, comparer les longueurs des deux intervalles de confiance trouvés dans les questions 3 et 4.

EX .. Q1 X prend ses valeurs dans $[0,1]$. $\forall x\in]-\infty,0[$, $F_X(x)=0$. $\forall x\in[1,+\infty[$, $F_X(x)=1$.

Notons que $\forall x\in\mathbb{R}$, $P_{\{X\leq a\}}(X\leq x)=\begin{cases}0 \text{ si } x\in]-\infty,0[\\ \frac{x}{a} \text{ si } x\in[0,a[\\ 1 \text{ si } x\in[1,+\infty[\end{cases}$.

$\forall x\in\mathbb{R}$, $P_{\{X>a\}}(X\leq x)=\begin{cases}0 \text{ si } x\in]-\infty,a[\\ \frac{x-a}{1-a} \text{ si } x\in[a,1[\\ 1 \text{ si } x\in[1,+\infty[\end{cases}$

Soit $x\in[0,1[$. $(\{X\leq a\},\{X>a\})$ est un système complet d'événements. La formule des probabilités totales donne :

$F_X(x)=P(X\leq a)P_{\{X\leq a\}}(X\leq x)+P(X>a)P_{\{X>a\}}(X\leq x)=\frac{1}{2}\left[P_{\{X\leq a\}}(X\leq x)+P_{\{X>a\}}(X\leq x)\right]$.

1er cas.. $x\in[0,a[$. $F_X(x)=\frac{1}{2}\left[\frac{x}{a}+0\right]=\frac{x}{2a}$.

2ème cas.. $x\in[a,1[$. $F_X(x)=\frac{1}{2}\left[1+\frac{x-a}{1-a}\right]$

Finalement $\underline{\forall x\in\mathbb{R},\ F_X(x)=\begin{cases}0 \text{ si } x\in]-\infty,0[\\ \frac{x}{2a} \text{ si } x\in[0,a[\\ \frac{1}{2}\left[1+\frac{x-a}{1-a}\right] \text{ si } x\in[a,1[\\ 1 \text{ si } x\in[1,+\infty[\end{cases}}$

$x\mapsto 0$, $x\mapsto 1$, $x\mapsto\frac{x}{2a}$, $x\mapsto\frac{1}{2}\left[1+\frac{x-a}{1-a}\right]$ sont de classe C^1 sur $\mathbb{R}$ donc F_X est de classe C^1 sur $]-\infty,0[$, $[0,a[$, $[a,1[$, $[1,+\infty[$. Alors F_X est ~~est~~ de classe C^1 sur $\mathbb{R}-\{0,a,1\}$, continue à droite en 0, a et 1.

$\lim_{x\mapsto 0^-}F_X(x)=0=\frac{0}{2a}=F_X(0)$, $\lim_{x\mapsto a^-}F_X(x)=\frac{1}{2}=\frac{1}{2}\left[1+\frac{a-a}{1-a}\right]=F_X(a)$, $\lim_{x\mapsto 1^-}F_X(x)=1=F_X(1)$

Ainsi F_X est continue à gauche en 0, a et 1.

Finalement F_X est continue sur $\mathbb{R}$ et au moins de classe C^1 sur $\mathbb{R}-\{0,a,1\}$.

X est une variable aléatoire à densité.

$\forall x\in]-\infty,0[\cup]1,+\infty[$, $F'_X(x)=0$, $\forall x\in]0,a[$, $F'_X(x)=\frac{1}{2a}$ et $\forall x\in]a,+\infty[$, $F'_X(x)=\frac{1}{2(1-a)}$.

$\underline{\text{la fonction } f_X \text{ définie par } \forall x\in\mathbb{R},\ f_X(x)=\begin{cases}\frac{1}{2a} \text{ si } x\in[0,a]\\ \frac{1}{2(1-a)} \text{ si } x\in\\ 0 \text{ sinon}\end{cases}}$ $\underline{\text{est une densité de } X \text{ car}}$ →

elle est positive sur son domaine et coïncide avec F_X' sur $\mathbb{R}$ privé d'un ensemble fini de point. Soit $k \in \mathbb{N}$.

$\int_{-\infty}^{0} t^k f_X(t)\,dt$ existe et vaut 0 ; $\int_0^a t^k f_X(t)\,dt$ existe et vaut $\frac{1}{2a}\,\frac{a^{k+1}}{k+1}$;

$\int_a^1 t^k f_X(t)\,dt$ existe et vaut $\frac{1}{2(1-a)}\times\frac{1-a^{k+1}}{k+1}$ et $\int_1^{+\infty} t^k f_X(t)\,dt$ existe et vaut 0.

Ainsi $\int_{-\infty}^{+\infty} t^k f_X(t)\,dt$ existe et vaut $\frac{1}{2(k+1)}\left[a^k + \frac{1-a^{k+1}}{1-a}\right]$.

Alors $E(X^k)$ existe et vaut $\frac{1}{2(k+1)}\left[a^k + \frac{1-a^{k+1}}{1-a}\right]$.

Ainsi $E(X)$ existe et vaut $\frac{1}{4}\left[a + \frac{1-a^2}{1-a}\right] = \frac{2a+1}{4}$. $E(X) = \frac{2a+1}{4}$.

$E(X^2)$ existe et vaut $\frac{1}{6}\left[a^2 + \frac{1-a^3}{1-a}\right] = \frac{1}{6}\left[a^2+1+a+a^2\right] = \frac{2a^2+a+1}{6}$.

Alors $V(X)$ existe et vaut $\frac{2a^2+a+1}{6} - \frac{(2a+1)^2}{16} = \frac{1}{48}\left[16a^2+8a+8-12a^2-12a-3\right]$

$V(X) = \frac{1}{48}(4a^2-4a+5)$.

Q2 $E(M_n) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{2a+1}{4}$ et $V(M_n) \overset{\text{indépendance}}{=} \frac{1}{n^2}\sum_{k=1}^{n} V(X_k) = \frac{1}{48n}(4a^2-4a+5)$.

$E(M_n) = \frac{2a+1}{4}$ et $V(M_n) = \frac{1}{48n}(4a^2-4a+5)$.

Posons $T_n = 2M_n - \frac{1}{2}$. $E(T_n) = 2\left(\frac{2a+1}{4}\right) - \frac{1}{2} = a$. $E(T_n) = a$ et $V(T_n) = 4V(M_n) = \frac{4a^2-4a+5}{12n}$

T_n est un estimateur sans biais de a.

$\forall \varepsilon \in \mathbb{R}_+^*$, $0 \leq P(|T_n - a| \geq \varepsilon) = P(|T_n - E(T_n)| \geq \varepsilon) \leq \frac{V(T_n)}{\varepsilon^2} = \frac{4a^2-4a+5}{12n\varepsilon^2}$.

Par encadrement : $\forall \varepsilon \in \mathbb{R}_+^*$, $\lim_{n\to+\infty} P(|T_n - a| \geq \varepsilon) = 0$.

T_n est un estimateur sans biais et convergent de a.

(Q3) Soit $\varepsilon \in \mathbb{R}_+^*$

$P(a \in [T_n-\varepsilon, T_n+\varepsilon]) = P(T_n - \varepsilon \leq a \leq T_n+\varepsilon) = P(-\varepsilon \leq T_n - a \leq \varepsilon) = P(|T_n - a| \leq \varepsilon) = P(|T_n - E(T_n)| \leq \varepsilon)$

Alors :

$P(a \in [T_n-\varepsilon, T_n+\varepsilon]) \geq 1-\alpha \Leftrightarrow P(|T_n - E(T_n)| \leq \varepsilon) \geq 1-\alpha \Leftrightarrow P(|T_n - E(T_n)| > \varepsilon) \leq \alpha.$

Or $P(|T_n - E(T_n)| > \varepsilon) \leq P(|T_n - E(T_n)| \geq \varepsilon) \underset{\text{B.T.}}{\leq} \dfrac{V(T_n)}{\varepsilon^2}.$

Notons que : $V(T_n) = \dfrac{1}{12n}(4a^2-4a+5) = \dfrac{1}{12n}[4\underbrace{a(a-1)}_{\leq 0}+5] \leq \dfrac{5}{12n}$. $\quad V(T_n) \leq \dfrac{5}{12n}$.

Alors $P(|T_n - E(T_n)| > \varepsilon) \leq \dfrac{5}{12n\varepsilon^2}.$

Ainsi pour avoir $P(a \in [T_n-\varepsilon, T_n+\varepsilon]) \geq 1-\alpha$ il suffit d'avoir $\dfrac{5}{12n\varepsilon^2} \leq \alpha$, c'est à dire $\varepsilon \geq \sqrt{\dfrac{5}{12n\alpha}}$.

$[T_n-\varepsilon, T_n+\varepsilon]$ est un intervalle de confiance de au risque α dès que $\varepsilon \geq \sqrt{\dfrac{5}{12n\alpha}}$.

(Q4) $\dfrac{T_n - a}{\sqrt{V(T_n)}} = \dfrac{2M_n - \frac{1}{2} - a}{\sqrt{4V(M_n)}} = \dfrac{2\left(M_n - \frac{2a+1}{4}\right)}{2\sqrt{V(M_n)}} = \dfrac{M_n - E(M_n)}{\sqrt{V(M_n)}}.$

- Les variables aléatoires X_n
  - → sont indépendantes
  - → ont même loi
  - → ont une espérance et une variance non nulle.

Alors la suite de terme général $\dfrac{M_n - E(M_n)}{\sqrt{V(M_n)}}$ converge en loi vers une variable aléatoire qui suit la loi normale centrée réduite.

Alors $\dfrac{T_n - a}{\sqrt{V(T_n)}}$ converge en loi vers une variable aléatoire qui suit la loi normale centrée réduite.

Dans la suite nous supposerons que $\dfrac{T_n - a}{\sqrt{V(T_n)}} \hookrightarrow \mathcal{N}(0,1).$

Traitons le cas général. Remplaçons alors 0,95 par $1-\alpha$ ou 0,05 par α.

Soit $\varepsilon \in \mathbb{R}_+^*$. Nous notons ϕ la fonction de répartition d'une var qui suit la loi normale centrée réduite.

Notons que
- $P\left(\left|\frac{T_n-a}{\sqrt{V(T_n)}}\right| \leq \varepsilon\right) = 2\phi(\varepsilon)-1$

- $P\left(\left|\frac{T_n-a}{\sqrt{V(T_n)}}\right| \leq \varepsilon\right) = P\left(-\varepsilon \leq \frac{T_n-a}{\sqrt{V(T_n)}} \leq \varepsilon\right) = P\left(a \in \left[T_n-\varepsilon\sqrt{V(T_n)}, T_n+\varepsilon\sqrt{V(T_n)}\right]\right)$.

Alors $\left[T_n-\varepsilon\sqrt{V(T_n)}, T_n+\varepsilon\sqrt{V(T_n)}\right]$ est un intervalle de confiance de a à la confiance α si et seulement si $2\phi(\varepsilon)-1 \geq 1-\alpha$ ou $\phi(\varepsilon) \geq 1-\frac{\alpha}{2}$.

$\alpha \in]0,1[$. $1-\frac{\alpha}{2} \in]\frac{1}{2},1[$. ϕ est une bijection de $\mathbb{R}$ dans $]0,1[$.

Ainsi $\exists !\, t_\alpha \in \mathbb{R}$, $\phi(t_\alpha)=1-\frac{\alpha}{2}$. Notons que $t_\alpha \in \mathbb{R}_+^*$ car $1-\frac{\alpha}{2} > \frac{1}{2}$.

Si t_α est l'unique élément de $\mathbb{R}$ et même de $\mathbb{R}_+^*$ tel que $\phi(t_\alpha)=1-\frac{\alpha}{2}$ alors

$\left[T_n-t_\alpha\sqrt{V(T_n)}, T_n+t_\alpha\sqrt{V(T_n)}\right]$ est un intervalle de confiance de a au risque α.

Malheureusement $V(T_n)$ dépend de a ... mais est majoré par $\frac{5}{12n}$.

Remarquons que $P\left(a \in \left[T_n-t_\alpha\sqrt{V(T_n)}, T_n+t_\alpha\sqrt{V(T_n)}\right]\right) \leq P\left(a \in \left[T_n-t_\alpha\sqrt{\frac{5}{12n}}, T_n+t_\alpha\sqrt{\frac{5}{12n}}\right]\right)$

"par croissance"

Si t_α est l'unique élément de $\mathbb{R}$ et même de $\mathbb{R}_+^*$ tel que $\phi(t_\alpha)=1-\frac{\alpha}{2}$ alors

$\left[T_n-t_\alpha\sqrt{\frac{5}{12n}}, T_n+t_\alpha\sqrt{\frac{5}{12n}}\right]$ est un intervalle de confiance de a au risque α.

Remarque... Si $\alpha=0,05$, $1-\frac{\alpha}{2}=0,975$ et $t_\alpha = 1,96$.

(Q5) Dans Q3 l'intervalle de confiance a pour longueur : $2\times\sqrt{\frac{5}{12n\times 0,05}} = 2\times\sqrt{\frac{100}{12n}}$.

— Q4 — : $2t_{0,05}\sqrt{\frac{5}{12n}} = 2\times 1,96\sqrt{\frac{5}{12n}}$

$$\frac{2\times\sqrt{\frac{100}{12n}}}{2\times 1,96\sqrt{\frac{5}{12n}}} = \frac{10}{1,96\times\sqrt{5}} = \frac{2\sqrt{5}}{1,96} \simeq 2,28\ !$$

Le second intervalle de confiance est bien meilleur que le premier.

Exercice Un sondage effectué auprès de 100 ménages a donné les résultats suivants : 40% de ménages satisfaits du produits P et 60% de mécontents. Donner un intervalle de confiance au risque de 5% de la proportion de satisfaits.

$f_{obs} = 0,4$. $n = 100$. $\alpha = 0,05$. $t_\alpha = 1,96$ (on veut que $2\phi(t_\alpha) - 1 \geq 1 - \alpha = 0,95$ donc on veut que $\phi(t_\alpha) \geq \frac{0,95+1}{2} = 0,975$ ... d'où la valeur de t_α).

L'intervalle de confiance demandé est : $\left[0,4 - \frac{1,96}{2\sqrt{100}} ; 0,4 + \frac{1,96}{2\sqrt{10}}\right] = [0,302 ; 0,498]$

Exercice Dans un scrutin le dépouillement des n premiers bulletins donne 60% de votes favorables au candidat A.

Q1. Déterminer n pour que l'on puisse affirmer avec moins de 5% de risque d'erreurs que A obtiendra entre 58% et 62% de voix.

Q2. On suppose que A obtient effectivement 60% de voix. Trouver la probabilité pour que les partisans de A soient en minorité dans un échantillon donné de 100 électeurs ?

(Q1) Il nous faut trouver $[0,58 ; 0,62]$ comme intervalle de confiance au risque de 5% de la proportion p de bulletins favorables à A.

Cet intervalle de confiance est à priori $\left[f_{obs} - \frac{t_\alpha}{2\sqrt{n}}, f_{obs} + \frac{t_\alpha}{2\sqrt{n}}\right]$. avec $\alpha = 0,05$

Nous savons depuis longtemps que $t_\alpha = 1,96$.

Cet intervalle est donc $\left[0,6 - \frac{1,96}{2\sqrt{n}}, 0,6 + \frac{1,96}{2\sqrt{n}}\right]$.

Finalement tout intervalle de ce type avec $\frac{1,96}{2\sqrt{n}} \leq 0,02$ conviendra.

Nous choisirons alors n tel que : $\left(\frac{1,96}{2 \times 0,02}\right)^2 \leq n$. Or $\left(\frac{1,96}{2 \times 0,02}\right)^2 = 2401$

Nous choisirons donc $\underline{\underline{n = 2401}}$.

(Q2) Notons X le nombre de partisans de A dans un échantillon donné de 100 électeurs

X suit une loi binomiale de paramètres 100 et 0,6

Les paramètres permettent d'approximer X par une var Y suivant une loi normale de paramètres $60, \sqrt{100 \times 0,6 \times 0,4} = \sqrt{24}$.

$P(X < 50) \simeq P(Y < 50) = P\left(\frac{Y-60}{\sqrt{24}} \leq -\frac{10}{\sqrt{24}}\right) = \phi\left(-\frac{10}{\sqrt{24}}\right) = 1 - \phi\left(\frac{10}{\sqrt{24}}\right)$.

$\frac{10}{\sqrt{24}} \simeq 2,04$. $1 - \phi(2,04) \simeq 0,0217$. La probabilité pour que les partisans de A soient en minorité est sensiblement : $\underline{\underline{0,0217}}$.

Exercice Afin d'étudier la proportion p de consommateurs satisfaits par le produit A on a interrogé 100 consommateurs. 56 d'entre eux ont déclaré être satisfaits par ce produit. Donner un intervalle de confiance à 95% de p.

Pour nous $\alpha = 0,05$. Cherchons t_α tel que $2\phi(t_\alpha) - 1 \geqslant 1-\alpha$; c'est à dire tel que $\phi(t_\alpha) = \frac{2-\alpha}{2} = \frac{2-0,05}{2} = 0,975 \simeq \phi(1,96)$. Nous retiendrons $t_\alpha = 1,96$.

Un intervalle de confiance à 95% de confiance de p est $\left[f_{obs} - \frac{t_\alpha}{2\sqrt{100}}, f_{obs} + \frac{t_\alpha}{2\sqrt{100}}\right]$

$f_{obs} = 0,56$. $f_{obs} - \frac{t_\alpha}{2\sqrt{100}} \simeq 0,462$ et $f_{obs} + \frac{t_\alpha}{2\sqrt{100}} = 0,658$

Nous retiendrons comme intervalle de confiance : $[0,462 ; 0,658]$.

Exercice Une usine fabrique des câbles. On suppose que la charge maximale supportée par un câble exprimée en tonnes est une variable aléatoire suivant une loi normale de paramètres m et $0,5$.

Une étude portant sur 50 câbles a donné une moyenne des charges maximales supportées égale à 12,2 tonnes.

Q1. Déterminer l'intervalle de confiance à 99% de la charge maximale moyenne de tous les câbles fabriqués par l'usine.

Q2. Quelle doit être la taille minimale de l'échantillon étudié pour que la longueur de l'intervalle de confiance à 99% soit inférieure ou égale à $0,2$.

Q1.. $m_{obs} = 12,2$. $n = 50$. $\alpha = 0,01$.

cherchons t_α tel que $2\phi(t_\alpha) - 1 \geqslant 1-\alpha$; c'est à dire tel que : $\phi(t_\alpha) \geqslant \frac{2-\alpha}{2} = 0,995$

$\phi(2,57) \leqslant 0,995 \leqslant \phi(2,58)$. Nous prendrons donc $t_\alpha = 2,58$

L'intervalle de confiance de m à 99% de confiance est : $\left[m_{obs} - \frac{t_\alpha}{\sqrt{50}} 0,5 ; m_{obs} + \frac{t_\alpha}{\sqrt{50}} 0,5\right]$

$m_{obs} - \frac{t_\alpha}{\sqrt{50}} \times 0,5 \simeq 12,2 - \frac{2,58}{\sqrt{50}} \times 0,5 \simeq 12,0175$

$m_{obs} + \frac{t_\alpha}{\sqrt{50}} \times 0,5 \simeq 12,2 + \frac{2,58}{\sqrt{50}} \times 0,5 \simeq 12,3824$

Nous retiendrons comme intervalle de confiance $[12,01 ; 12,39]$

Q2.. La longueur de l'intervalle de confiance à 99% est : $2 \times \frac{t_\alpha}{\sqrt{n}} \times 0,5 = \frac{2,58}{\sqrt{n}}$

$\frac{2,58}{\sqrt{n}} \leqslant 0,2 \Leftrightarrow n \geqslant \left(\frac{2,58}{0,2}\right)^2 \simeq 166,41$

Nous retiendrons $n = 167$ comme taille minimale de l'échantillon pour que la longueur de l'intervalle de confiance à 99% soit inférieure à $0,2$.

Exercice Pour estimer le résultat d'une élection à deux candidats, on effectue un sondage sur 1000 personnes et 540 déclarent voter pour A. Au seuil de confiance de 95%, peut-on estimer que A sera élu ?

L'intervalle de confiance de la proportion p des partisans de A au seuil de confiance de 95% est $\left[f_{obs} - \frac{1,96}{2\sqrt{1000}}, f_{obs} + \frac{1,96}{2\sqrt{1000}}\right]$

$f_{obs} - \frac{1,96}{2\sqrt{1000}} = 0,54 - \frac{1,96}{2\sqrt{1000}} = 0,5090 > 0,5$. <u>Au seuil de confiance de 95% on peut estimer que A sera élu.</u>

Exercice Si l'écart type de la durée de vie* d'un modèle de lampe électrique est estimé à 100 heures, quelle doit être la taille de l'échantillon à prélever pour que l'erreur sur l'estimation de la durée de vie moyenne n'excède pas 20 heures et ce avec une probabilité de 95% (resp. 99%). (*durée de vie qui suit une loi normale)

L'intervalle de confiance à la confiance de 95% de la durée de vie moyenne d'une lampe est $\left[m_{obs} - \frac{t_\alpha \times 100}{\sqrt{n}}, m_{obs} + \frac{t_\alpha \times 100}{\sqrt{n}}\right]$ avec $t_\alpha = 1,96$ puisque $\alpha = 0,05$

n est solution de : $\frac{t_\alpha \times 100}{\sqrt{n}} \leq 20$, c'est-à-dire de : $n \geq \left(\frac{1,96 \times 100}{20}\right)^2 \simeq 96,04$

Nous obtenons $n = 97$.

Pour la confiance à 99% nous obtenons <u>$n = 167$.</u>

($t_\alpha = 2,58$ car $\alpha = 0,01$ et $\left(\frac{2,58 \times 100}{20}\right)^2 \simeq 166,41$

Exercice Une machine fabrique des rondelles en séries. Leur diamètre d est une variable aléatoire gaussienne dont l'écart type est de 1 millimètre. m est la moyenne.
On prélève au hasard un échantillon de 9 rondelles. Les mesures des diamètres en mm sont les suivantes : 20,1 19,9 20,0 19,8 19,7 20,2 20,1 23,1 22,8.
Donner un intervalle de confiance au risque de $0,05$ (resp. $0,01$) de m.

$m_{obs} = \frac{619}{30}$. Si $\alpha = 0,05$, $t_\alpha = 1,96$. $n = 9$.

L'intervalle de confiance au risque de 0,05 est : $\left[m_{obs} - \frac{t_\alpha \times 1}{\sqrt{9}}, m_{obs} + \frac{t_\alpha \times 1}{\sqrt{9}}\right]$

Nous obtenons : <u>[19,98 ; 21,29]</u>

Pour $\alpha = 0,01$ on remplace t_α par 2,58 et on trouve : <u>[19,77 ; 21,50].</u>

Exercice 1 **PC et CC** Un sondage consiste à proposer l'affirmation "A" à certaines personnes d'une population donnée. Le sujet abordé étant délicat, le stratagème suivant est mis en place afin de mettre en confiance les personnes sondées pour qu'elles ne mentent pas ...

L'enquêteur dispose d'un paquet de 20 cartes, numérotées de 1 à 20, qu'il remet à la personne sondée. Celle-ci tire une carte au hasard et ne la montre pas à l'enquêteur. La règle est alors la suivante :
• si la carte porte le numéro 1, la personne sondée répond "vrai" si elle est d'accord avec l'affirmation "A" et 'faux" sinon.
• si la carte porte un autre numéro, la personne sondée répond "vrai" si elle n'est pas d'accord avec l'affirmation "A" et 'faux" sinon.
Le but de l'enquête est d'évaluer la proportion p de gens de cette population qui sont réellement d'accord avec l'affirmation "A".

Q1. On interroge une personne selon ce procédé et on considère l'événement suivant, noté V : la personne répond "vrai". On note $\theta = P(V)$.

En utilisant la formule des probabilités totales, exprimer θ en fonction de p, puis en déduire p en fonction de θ.

Q2. Certaines considérations théoriques laissent penser que $p = \dfrac{17}{18}\cdot$

a) Vérifier que $\theta = \dfrac{1}{10}\cdot$

b) Calculer la probabilité pour qu'une personne ayant répondu "vrai" soit d'accord avec l'affirmation "A".

On revient au cas général où l'on ne connaît ni p, ni θ.

Q3. On considère un échantillon aléatoire, de taille n, extrait de la population considérée et on note S_n, le nombre de réponses "vrai" obtenues. On suppose n assez grand pour pouvoir considérer que cet échantillonnage est assimilable à un tirage avec remise.

a) Donner la loi de S_n ainsi que son espérance et sa variance.

b) Montrer que $\dfrac{S_n}{n}$, est un estimateur sans biais et convergent de θ.

Q4. Dans cette question, on suppose que l'on a réalisé un échantillon de 100 personnes et on constate que 23 personnes ont répondu "vrai".

a) Donner une estimation ponctuelle de θ et de p.

b) Donner un intervalle de confiance à 95% de θ puis de p.

On rappelle que, si Φ désigne la fonction de répartition d'une variable X suivant la loi normale $\mathcal{N}(0, 1)$, alors $\Phi(1,96) = 0,975$.

EXERCICE

1) Notons T_1 l'événement «la personne tire la carte numéro 1 ».

$(T_1, \overline{T_1})$ est un système complet d'événements. La formule des probabilités totales donne alors :

$$\theta = P(V) = P(V/T_1)\ P(T_1) + P(V/\overline{T_1})\ P(\overline{T_1}).$$

Sachant que la personne a tiré la carte numéro 1, elle répond "vrai" si et seulement si elle est d'accord avec l'affirmation «A». La proportion de gens de cette population qui sont réellement d'accord avec l'affirmation «A» étant p : $P(V/T_1) = p$.

Sachant que la personne n'a pas tiré la carte numéro 1, elle répond "vrai" si et seulement si elle n'est pas d'accord avec l'affirmation «A» ; alors : $P(V/\overline{T_1}) = 1 - p$.

De plus : $P(T_1) = \dfrac{1}{20}$ et $P(\overline{T_1}) = \dfrac{19}{20}\cdot$ Finalement : $\theta = P(V) = \dfrac{1}{20}\,p + \dfrac{19}{20}\,(1-p)$.

Ainsi : $\boxed{\theta = \dfrac{19 - 18\,p}{20}}$. Ceci donne encore $18\,p = 19 - 20\,\theta$ et donc : $\boxed{p = \dfrac{19 - 20\,\theta}{18}}$.

2) a. $p = \dfrac{17}{18}\cdot$ Alors $\theta = \dfrac{19 - 18p}{20} = \dfrac{19 - 17}{20} = \dfrac{1}{10}\cdot$ $\boxed{\theta = \dfrac{1}{10}}$.

b. Notons D_A l'événement «une personne est d'accord avec l'affirmation «A»». On cherche : $P(D_A/V)$.

$P(D_A/V) = \dfrac{P(V/D_A)\ P(D_A)}{P(V)} = \dfrac{P(V/D_A)\ p}{\theta}\cdot$

Sachant que la personne est d'accord avec «A», elle répond vrai si et seulement si elle tire la carte numéro 1.

Alors : $P(V/D_A) = \dfrac{1}{20}$ et : $P(D_A/V) = \dfrac{(1/20)\ p}{\theta} = \dfrac{p}{20\,\theta} = \dfrac{17/18}{20\ (1/10)} = \dfrac{17}{36}\cdot$

$\boxed{\text{La probabilité pour qu'une personne ayant répondu "vrai" soit d'accord avec «A» est : } \dfrac{17}{36}}$.

3) a. Si l'échantillonnage est assimilable à un tirage avec remise, on peut alors affirmer que :

$\boxed{S_n \text{ suit une loi binômiale de paramètres } n \text{ et } \theta}$.

Alors $\boxed{\forall k \in [\![0, n]\!],\ P(S_n = k) = \binom{n}{k}\ \theta^k\ (1-\theta)^{n-k},\ E(S_n) = n\theta \text{ et } V(S_n) = n\theta(1-\theta)}$.

b. $E\left(\dfrac{S_n}{n}\right) = \dfrac{1}{n}\ E(S_n) = \dfrac{1}{n}\ n\theta = \theta$. Alors $\boxed{\dfrac{S_n}{n} \text{ est un estimateur sans biais de } \theta}$.

$V\left(\dfrac{S_n}{n}\right) = \dfrac{1}{n^2}\ V(S_n) = \dfrac{1}{n^2}\ n\theta(1-\theta) = \dfrac{1}{n}\ \theta(1-\theta)$. Ainsi : $\lim\limits_{n\to+\infty} V\left(\dfrac{S_n}{n}\right) = 0$.

Donc $\boxed{\dfrac{S_n}{n} \text{ est un estimateur sans biais et convergent de } \theta}$.

4) a. 23 personnes ont répondu "vrai" sur un échantillon de 100 personnes.

Alors $\boxed{\text{une estimation ponctuelle de } \theta \text{ est} : 0,23}$.

Comme $p = \dfrac{19 - 20\,\theta}{18}$ et $\dfrac{19 - 20(23/100)}{18} = \dfrac{19 \times 5 - 23}{90} = \dfrac{4}{5} = 0,8$:

$\boxed{\text{une estimation ponctuelle de } p \text{ est } 0,8}$.

b. D'après le cours, si $t_{0,05}$ est un réel vérifiant $2\Phi(t_{0,05}) - 1 = 0,95$ alors $\left[\dfrac{23}{100} - \dfrac{t_{0,05}}{2\sqrt{100}}, \dfrac{23}{100} + \dfrac{t_{0,05}}{2\sqrt{100}}\right]$ est un intervalle de confiance à 95% de θ.

$2\Phi(t_{0,05}) - 1 = 0,95 \iff \Phi(t_{0,05}) = \dfrac{0,95+1}{2} = 0,975 = \Phi(1,96)$. Nous poserons alors : $t_{0,05} = 1,96$.

Alors : $\dfrac{23}{100} - \dfrac{t_{0,05}}{2\sqrt{100}} = 0,23 - 0,098 = 0,132$ et $\dfrac{23}{100} + \dfrac{t_{0,05}}{2\sqrt{100}} = 0,23 + 0,098 = 0,328$.

Ainsi $\boxed{[0,132; 0,328] \text{ est un intervalle de confiance à } 95\% \text{ de } \theta}$.

$p = \dfrac{19 - 20\,\theta}{18}$, donc $\theta \in [0,132; 0,328] \iff p \in \left[\dfrac{19 - 20 \times 0,328}{18}, \dfrac{19 - 20 \times 0,132}{18}\right]$.

$0,6911$ est une valeur approchée par défaut de $\dfrac{19 - 20 \times 0,328}{18}$.

$0,9089$ est une valeur approchée par exès de $\dfrac{19 - 20 \times 0,132}{18}$.

Ainsi $\boxed{[0,6911; 0,9089] \text{ est un intervalle de confiance à } 95\% \text{ de } p}$.
