ESTIMATION.

Estimation ponctuelle

Exercice 1.
On a vu que la variance empirique est un estimateur biaisé de la variance.
Soit X une variable aléatoire d'espérance m et de variance σ^2. On pose :

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad S_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2 \qquad T_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2$$

1. Montrer que S_n est un estimateur sans biais de σ^2. Quel est son inconvénient ?
2. Montrer que T_n est un estimateur sans biais de σ^2.(on s'inspirera fortement du calcul vu dans le cours "premier estimateur de la variance" [1])

RÉPONSE:

1. **Outils utilisés** :
   (a) définition de "sans biais"
   (b) linéarité de l'espérance , formule $V(\alpha Y + \beta) = \alpha^2 V(Y)$
   (c) , König–Huygens "à l'envers" $E(Y^2) = V(Y) + (E(Y))^2$

1. on pourrait même ne pas refaire tous ces calculs

Soit $n \in \mathbb{N}^*$

$$\begin{aligned}
E(S_n) &= E\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} E\left((X_k - m)^2\right) && \text{linéarité de l'espérance} \\
&= \frac{1}{n}\sum_{k=1}^{n} \left[\left(E\left(X_k - m\right)\right)^2 + V\left(X_k - m\right)\right] && \text{K–H "à l'envers"} \\
&= \frac{1}{n}\sum_{k=1}^{n} \left[\left(E\left(X_k\right) - m\right)^2 + V\left(X_k - m\right)\right] && \text{linéarité de l'espérance} \\
&= \frac{1}{n}\sum_{k=1}^{n} \left[(m - m)^2 + V\left(X_k\right)\right] && \text{formule } V(\alpha Y + \beta) = \alpha^2 V(Y) \\
&= \frac{1}{n}\sum_{k=1}^{n} \left[\sigma^2\right] \\
&= \sigma^2
\end{aligned}$$

Comme on cherche à estimer σ^2, le biais qui est la différence entre ceu que l'on mesure et ce que l'on veut estimer vaut :

$$b(S_n) = E(S_n) - \sigma^2 = 0$$

l'estimateur S_n est donc non biaisé.

Son inconvénient il faut connaître la valeur exacte de l'espérance pour pouvoir le calculer, or il est courant que l'on cherche à déterminer et la variance et l'espérance d'un échantillon.

2.

Remarque : Ce calcul est presque identique à celui vu pour "premier estimateur de la variance". On a besoin de calculs préalables

$$\begin{aligned}
E(\overline{X_n}) &= E\left(\frac{1}{n-1}\sum_{k=1}^{n}\right) \\
&= \frac{1}{n-1}\sum_{k=1}^{n} E\left(X_k\right) && \text{linéarité de l'espérance} \\
&= \frac{n}{n-1} m
\end{aligned}$$

et, d'une autre écriture de T_n

$$\begin{aligned}
T_n &= \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 \\
&= \frac{1}{n-1}\sum_{k=1}^{n}(X_k^2 + \overline{X}_n^2 - 2X_k\overline{X}_n) \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 + \sum_{k=1}^{n}\overline{X}_n^2 - 2\sum_{k=1}^{n} X_k\overline{X}_n\right) \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 + n\overline{X}_n^2 - 2\overline{X}_n\sum_{k=1}^{n} X_k\right) && \text{constantes par rapport à l'indice } k \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 + n\overline{X}_n^2 - 2n\overline{X}_n\overline{X}_n\right) && \text{définition de } \overline{X}_n \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\overline{X}_n^2\right) \\
&= \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2\right) - \frac{n}{n-1}\overline{X}_n^2
\end{aligned}$$

En utilisant la formule de König–Huygens "à l'envers" sur $\overline{X_n}$

$$E(\overline{X_n}^2) = V(\overline{X_n}) + E(\overline{X_n})^2 = \frac{\sigma^2}{n} + m^2$$

On a de même

$$E(X_k^2) = V(X_k) + E(X_k)^2 = \sigma^2 + m^2$$

On peut donc calculer l'espérance de T_n

$$\begin{aligned}
E(T_n) &= E\left(\frac{1}{n-1}\sum_{k=1}^{n} X_k^2 - \frac{n}{n-1}\overline{X}_n^2\right) \\
&= \frac{1}{n-1}\sum_{k=1}^{n} E\left(X_k^2\right) - \frac{n}{n-1}E\left(\overline{X}_n^2\right) && \text{linéarité} \\
&= \frac{1}{n-1}\sum_{k=1}^{n}\left(\sigma^2 + m^2\right) - \frac{n}{n-1}\left(\frac{\sigma^2}{n} + m^2\right) \\
&= \frac{n}{n-1}\left(\sigma^2 + m^2\right) - \frac{n}{n-1}\left(\frac{\sigma^2}{n} + m^2\right) \\
&= \frac{n-1}{n-1}\sigma^2 \\
&= \sigma^2
\end{aligned}$$

L'espérance de l'estimateur est égale à ce que l'on veut estimer, l'estimateur T_n est sans biais.

Remarque : L'inconvénient précédent a disparu, le calcul de T_n ne dépend que de l'échantillon et non pljus de la connaissance préalables de m

$*$

Exercice 2.
D'après Romain Boillaud On dispose de $n > 2$ observations indépendantes X_1, X_2,...X_n de même loi de Poisson de paramètre $\lambda > 0$ inconnu. On souhaite estimer le paramètre $\theta = \mathrm{e}^{-\lambda}$.

Pour tout $k \in \mathbb{N}^*$, on pose

$$Y_k = \begin{cases} 1 & \text{si } X_k = 0 \\ 0 & \text{sinon} \end{cases}$$

On introduit également :

$$\overline{Y}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k \qquad S_n = \sum_{k=1}^{n} X_k$$

1. (a) Montrer que Y_k suit une loi de Bernouilli de paramètre $\mathrm{e}^{-\lambda}$.
   (b) Calculer $E(Y_k)$ puis $E(\overline{Y}_n)$. En déduire que $\overline{Y}_n$ est un estimateur sans biais de θ.
   (c) Calculer la variance de $\overline{Y}_n$ et en déduire que cette estimateur est convergent.
2. Expliquer pourquoi S_n suit une loi de Poisson et trouver son paramètre
3. On pose pour $i \in \mathbb{N}$
$$\varphi(i) = \mathbb{P}_{S_n=i}(X_1 = 0)$$
Montrer que pour $i \in \mathbb{N}$
$$\varphi(i) = \left(\frac{n-1}{n}\right)^i$$
4. On pose $T_n = \varphi(S_n)$
   (a) À l'aide du théorème de transfert, montrer que $E(T_n) = \mathrm{e}^{-\lambda}$.
   (b) Montrer que T_n est un estimateur sans biais de θ.
   (c) Montrer $V(T_n) = \mathrm{e}^{-2\lambda}\left(\mathrm{e}^{\lambda/n} - 1\right)$ et en déduire que T_n est un estimateur convergent.
5. Comparer les risques quadratiques de T_n et $\overline{Y}_n$.

RÉPONSE:

1. (a) Une variable aléatoire qui ne prend que deux valeurs 0 et 1 est une variable de Bernoulli, pour connaître son paramètre il faut calculer $P(Y_k = 1) = P(X_k = 0) = \dfrac{\lambda^0}{0!}\mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda}$
$$Y_n \hookrightarrow \mathcal{B}(\mathrm{e}^{-\lambda})$$

(b) D'après le cours

$$E(Y_n) = \mathrm{e}^{-\lambda}$$

Calcul classique

$$\begin{aligned} E(\overline{Y_n}) &= E\left(\frac{1}{n}\sum_{k=1}^{n} Y_k\right) \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left(Y_k\right) && \text{linéarité de l'espérance} \\ &= \frac{1}{n}\sum_{k=1}^{n} \mathrm{e}^{-\lambda} \\ &= \frac{n}{n}\mathrm{e}^{-\lambda} \\ & \mathrm{e}^{-\lambda} \end{aligned}$$

$$E(\overline{Y_n}) = \theta$$

L'espérance de l'estimateur est égale à ce que l'on veut estimer, c'est la définition de "sans biais"

(c) **outils et techniques pour le calcul de la variance**

- propriétés de la loi de Bernoulli
- $V(X+Y) = V(X) + V(Y)$ **en cas d'indépendance**
- $V(\alpha X + \beta) = \alpha^2 V(X)$

On a

$$\begin{aligned} V(\overline{Y_n}) &= V\left(\frac{1}{n}\sum_{k=1}^{n} Y_k\right) \\ &= \frac{1}{n^2} V\left(\sum_{k=1}^{n} Y_k\right) && \text{formule } V(\alpha X + \beta) = \alpha^2 V(X) \\ &= \frac{1}{n^2}\sum_{k=1}^{n} V\left(Y_k\right) && \text{les } (Y_k) \text{ sont indépendants} \\ &= \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{e}^{-\lambda}\left(1 - \mathrm{e}^{-\lambda}\right) && \text{loi de Bernoulli} \\ &= \frac{1}{n^2}\mathrm{e}^{-\lambda}\left(1 - \mathrm{e}^{-\lambda}\right)\sum_{k=1}^{n} 1 && \text{constante / k} \\ &= \frac{\mathrm{e}^{-\lambda}\left(1 - \mathrm{e}^{-\lambda}\right)}{n} \end{aligned}$$

outils pour "estimateur convergent

i. Techniques proches de celles vue en cours
ii. Décomposition du risque quadratique en biais variance.

iii. Caractérisation d'un estimateur convergent avec la limite du risque quadratique

D'après le cours

$$r(\overline{T_n}) = b(\overline{Y_n})^2 + V(\overline{Y_n}) = 0^2 + V(\overline{Y_n}) = \frac{\mathrm{e}^{-\lambda}\left(1-\mathrm{e}^{-\lambda}\right)}{n}$$

Or

$$\lim_{n\to+\infty} \frac{\mathrm{e}^{-\lambda}\left(1-\mathrm{e}^{-\lambda}\right)}{n} = 0$$

Donc

$$\lim_{n\to+\infty} r(\overline{Y_n}) = 0$$

Donc l'estimateur $\overline{Y_n}$ est convergent

2. On sait que la somme de variables aléatoires suivant des lois de Poisson indépendantes suit une loi de Poisson, plus précisément

$$S_n \hookrightarrow \mathcal{P}\left(n\lambda\right)$$

3. Soit $i \in \mathbb{N}$

$$\begin{aligned}
\varphi(i) &= \mathbb{P}_{S_n=i}(X_1=0) \\
&= \frac{\mathbb{P}([S_n=i]\cap[X_1=0])}{\mathbb{P}(S_n=i)} && \text{définition d'une proba conditionnelle} \\
&= \frac{\mathbb{P}([S_{n-1}=i]\cap[X_1=0])}{\mathbb{P}(S_n=i)} \\
&= \frac{\mathbb{P}(S_{n-1}=i)\mathbb{P}(X_1=0)}{\mathbb{P}(S_n=i)} && \text{indépendance par lemme des coalitions} \\
&= \frac{\dfrac{((n-1)\lambda)^i}{i!}\mathrm{e}^{-\lambda(n-1)}\mathbb{P}(X_1=0)}{\mathbb{P}(S_n=i)} && \text{car } S_{n-1} \hookrightarrow \mathcal{P}((n-1)\lambda) \\
&= \frac{\dfrac{((n-1)\lambda)^i}{i!}\mathrm{e}^{-\lambda(n-1)}\mathbb{P}(X_1=0)}{\dfrac{(n\lambda)^i}{i!}\mathrm{e}^{-n\lambda}} && \text{car } S_n \hookrightarrow \mathcal{P}(n\lambda) \\
&= \frac{\dfrac{((n-1)\lambda)^i}{i!}\mathrm{e}^{-\lambda(n-1)}\mathrm{e}^{-\lambda}}{\dfrac{(n\lambda)^i}{i!}\mathrm{e}^{-n\lambda}} && \text{car } X_1 \hookrightarrow \mathcal{P}(\lambda) \\
&= \frac{(n-1)^i\mathrm{e}^{-n\lambda}}{n^i\mathrm{e}^{-n\lambda}} \\
&= \left(\frac{n-1}{n}\right)^i
\end{aligned}$$

4. On pose $T_n = \varphi(S_n)$

(a) Le théorème de transfert, et comme les sommes sont finies donc absolument convergentes

$$
\begin{aligned}
E(T_n) &= E(\varphi(S_n)) \\
&= \sum_{i=0}^{+\infty} \varphi(i)\mathbb{P}(S_n = i) \\
&= \sum_{i=0}^{n} \mathbb{P}_{S_n=i}(X_1 = 0)\mathbb{P}(S_n = i) && \text{définition} \\
&= \mathbb{P}(X_1 = 0) && \text{la CVA est assurée par probabilités totales avec le SCE } ([S_n = i])_i \\
&= \mathrm{e}^{-\lambda}\frac{\lambda^0}{0!} && \text{car } X_1 \hookrightarrow \\
&= \mathrm{e}^{-\lambda}
\end{aligned}
$$

(b) On vient de montrer que l'espérance de l'estimateur T_n est égale à $\mathrm{e}^{-\lambda} = \theta$ qui est le paramètre que l'on veut estimer.

T_n est un estimateur sans biais de θ.

Sous réserve de convergence (la convergence absolue se confond avec la convergence car les séries sont à termes positifs)

$$
\begin{aligned}
E(T_n^2) &= E((\varphi(S_n))^2) \\
&= \sum_{i=0}^{+\infty} \varphi(i)^2 \mathbb{P}(S_n = i) && \text{théorème de transfert} \\
&= \sum_{i=0}^{+\infty} \varphi(i)^2 \mathbb{P}(S_n = i) && \text{théorème de transfert} \\
&= \sum_{i=0}^{+\infty} \left(\left(\frac{n-1}{n} \right)^i \right)^2 \mathbb{P}(S_n = i) && \text{question précédente} \\
&= \sum_{i=0}^{+\infty} \left(\left(\frac{n-1}{n} \right)^i \right)^2 \mathrm{e}^{-n\lambda} \frac{(n\lambda)^i}{i!} && \text{loi de } S_n \\
&= \sum_{i=0}^{+\infty} \left(\left(\frac{n-1}{n} \right)^2 \right)^i \mathrm{e}^{-n\lambda} \frac{(n\lambda)^i}{i!} \\
&= \mathrm{e}^{-n\lambda} \sum_{i=0}^{+\infty} \frac{\left(\left(\frac{n-1}{n} \right)^2 (n\lambda) \right)^i}{i!} \\
&= \exp(-n\lambda) \cdot \exp\left(\left(\frac{n-1}{n} \right)^2 (n\lambda) \right) && \text{série exponentielle qui assure la convergence} \\
&= \exp(-n\lambda) \cdot \exp\left(\left(\frac{n-1}{n} \right)^2 (n\lambda) \right) \\
&= \mathrm{e}^{\frac{\lambda}{n} - 2\lambda} && \text{calculs}
\end{aligned}
$$

La formule de König–Huygens assure l'existence d'une variance et

$$
\begin{aligned}
V(T_n) &= E(T_n^2) - E(T_n)^2 \\
&= \mathrm{e}^{\frac{\lambda}{n} - 2\lambda} - \left(\mathrm{e}^{-\lambda}\right)^2 \\
&= \mathrm{e}^{\frac{\lambda}{n} - 2\lambda} - \left(\mathrm{e}^{-2\lambda}\right) \\
&= \mathrm{e}^{\frac{\lambda}{n} - 2\lambda} - \mathrm{e}^{-2\lambda} \\
&= \mathrm{e}^{-2\lambda} \left(\mathrm{e}^{\frac{\lambda}{n}} - 1 \right)
\end{aligned}
$$

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

D'après la décomposition risque quadratique en Biais/variance

$$r(T_n) = (b(T_n))^2 + V(T_n) = 0^2 + \mathrm{e}^{-2\lambda}\left(\mathrm{e}^{\dfrac{\lambda}{n}} - 1\right)$$

Or, à $\lambda > 0$ fixé

$$\lim_{n\to+\infty} \frac{\lambda}{n} = 0$$

par continuité de l'exponentielle

$$\lim_{n\to+\infty} \exp\left(\frac{\lambda}{n}\right) = \mathrm{e}^0 = 1$$

et donc

$$\lim_{n\to+\infty} r(T_n) = 0$$

D'après le résultat du cours

(c) $(T_n)_{n\in\mathbb{N}^*}$ est un estimateur convergent de $\theta = \mathrm{e}^{-\lambda}$

5. On étudie $R(T_n) - R(\overline{Y_n})$ (longue étude des variations d'une fonction ...) et on trouve

$$r(\overline{Y_n}) \leqslant r(T_n)$$

*

Estimation par intervalle de confiance

Exercice 3 (Bernouilli en utilisant l'inégalité de Bienaymé Tchebichev)**.**
On suppose que $X \hookrightarrow \mathcal{B}(p)$ et que $(X_1, X_2, \dots, X_n)$ est un échantillon. On pose

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

1. Donner l'espérance et la variance de $\overline{X}_n$. (cours)
2. Montrer que $\overline{X_n}$ est un estimateur sans biais et convergent vers p. (cours).
3. Montrer que $\mathbb{P}(|\overline{X}_n - p| \leqslant \varepsilon) \geqslant 1 - \dfrac{p(1-p)}{n\varepsilon^2}$
4. Soit $\alpha \in]0;1[$ Trouver α tel que $1 - \dfrac{p(1-p)}{n\varepsilon^2} = 1 - \alpha$
5. Montrer que $\mathbb{P}\left(|\overline{X}_n - p| \leqslant \sqrt{\dfrac{p(1-p)}{n\alpha}}\right) \geqslant 1 - \alpha$.
6. Montrer que pour tout $p \in]0;1[$ $p(1-p) \leqslant 1/4$. En déduire que
7. Montrer que $\mathbb{P}\left(|\overline{X}_n - p| \leqslant \dfrac{1}{2\sqrt{n\alpha}}\right) \geqslant 1 - \alpha$.
8. En déduire un intervalle de confiance de niveau de confiance $1 - \alpha$ de p.

RÉPONSE:

1. C'est des calculs très classiques

$$\begin{aligned}
E(\overline{X_n}) &= E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} E\left(X_k\right) && \text{linéarité} \\
&= \frac{1}{n}(np) = p
\end{aligned}$$

$$\begin{aligned}
V(\overline{X_n}) &= V\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n^2}V\left(\sum_{k=1}^{n} X_k\right) && \text{formule } V(aX+b) = a^2V(X) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} V\left(X_k\right) && \text{indépendance des variables aléatoire} \\
&= \frac{npq}{n^2} = \frac{pq}{n}
\end{aligned}$$

Remarque : Dans le cadre de cet exercice nous savons que $X_n \hookrightarrow \mathcal{B}(n,p)$ on pourrait aller un peu plus. La méthode proposée est celle classique.

2. **outils utilisées**
   - Définition d'estimateur sans biais.
   - décomposition du risque quadratique en biais/variance.
   - Caractérisation d'un estimateur par l'étude du risque quadratique.

   On a vu que $E(\overline{X_n}) = p$ donc l'espérance de l'estimateur est égale à la quantité que l'on veut estimer.

   $\overline{X_n}$ est un estimateur sans biais de p

   D'après la formule du cours

$$r(\overline{X_n}) = b(\overline{X_n})^2 + V(\overline{X_n}) = 0 + \frac{pq}{n}$$

   Donc

$$\lim_{n\to+\infty} r_n(\overline{X_n}) = 0$$

   $\overline{X_n}$ est un estimateur convergent p

3. **outils utilisées**

- Bienaymée Tchebichef
- Evénement contraire

Soit $\varepsilon > 0$, comme $\overline{X_n}$ admet une variance alors en utilisant l'inégalité de Bienaymée Tchebichev

$$\mathbb{P}(\left|\overline{X_n} - E(\overline{X_n})\right|) > \varepsilon \leqslant \frac{V(\overline{X_n})}{\varepsilon^2}$$

donc

$$\mathbb{P}(\left|\overline{X_n} - p\right| > \varepsilon) \leqslant \frac{p(1-p)}{n\varepsilon^2}$$

donc

$$1 - \mathbb{P}(\left|\overline{X_n} - p\right| > \varepsilon) \geqslant 1 - \frac{p(1-p)}{n\varepsilon^2}$$

en passant à l'événement contraire

$$\mathbb{P}(\left|\overline{X_n} - p\right| \leqslant \varepsilon) \geqslant 1 - \frac{p(1-p)}{n\varepsilon^2}$$

4.

$$\begin{aligned} 1 - \frac{p(1-p)}{n\varepsilon^2} = 1 - \alpha \Leftrightarrow \frac{p(1-p)}{n\varepsilon^2} &= \alpha \\ \Leftrightarrow \varepsilon &= \sqrt{\frac{p(1-p)}{n\alpha}} \end{aligned}$$

5. On savait que

$$\mathbb{P}(\left|\overline{X_n} - p\right| \leqslant \varepsilon) \geqslant 1 - \frac{p(1-p)}{n\varepsilon^2}$$

et en utilisant la question précédente

$$\mathbb{P}\left(|\overline{X}_n - p| \leqslant \sqrt{\frac{p(1-p)}{n\alpha}}\right) \geqslant 1 - \alpha$$

6. On peut étudier les variations de $x \mapsto (1-x)x$ sur l'intervalle $[0;\,1]$ et en déduire la majoration demandée.

Méthode plus rapide

$$\begin{aligned} (1-p)p \leqslant \frac{1}{4} &\Leftrightarrow 4(1-p)p \leqslant 1 && \text{car } 4 > 0 \\ &\Leftrightarrow 0 \leqslant 1 - 4p + 4p^2 \\ &\Leftrightarrow 0 \leqslant (1-2p)^2 \end{aligned}$$

La dernière inégalité est vraie, donc par remontée d'équivalences :

$$\forall p \in [0;\,1] \qquad p(1-p) \leqslant \frac{1}{4}$$

7. D'après la question précédente

$$p(1-p) \leqslant \frac{1}{4}$$

donc comme les quantités considérées sont positives

$$\sqrt{\frac{p(1-p)}{n\alpha}} \leqslant \sqrt{\frac{1}{4n\alpha}}$$

On a donc

$$\left[|\overline{X_n} - p| \leqslant \sqrt{\frac{p(1-p)}{n\alpha}}\right] \subset \left[|\overline{X_n} - p| \leqslant \sqrt{\frac{1}{4n\alpha}}\right]$$

Par croissance d'une probabilité

$$\mathbb{P}\left(|\overline{X_n} - p| \leqslant \sqrt{\frac{p(1-p)}{n\alpha}}\right) \leqslant \mathbb{P}\left(|\overline{X_n} - p| \leqslant \sqrt{\frac{1}{4n\alpha}}\right)$$

et comme

$$1-\alpha \leqslant \mathbb{P}\left(|\overline{X_n} - p| \leqslant \sqrt{\frac{p(1-p)}{n\alpha}}\right)$$

On obtient

$$1-\alpha \mathbb{P}\left(|\overline{X_n} - p| \leqslant \sqrt{\frac{1}{4n\alpha}}\right)$$

8. **outils utilisés**
   - Définition d'un intervalle de confiance
   - "enlever des valeurs absolues" $|Y-b| \leqslant a \Leftrightarrow b-a \leqslant Y \leqslant b+a$

   L'inégalité précédente peut s'écrire

$$1-\alpha \leqslant \mathbb{P}\left(\overline{X_n} - \sqrt{\frac{1}{4n\alpha}} \leqslant p \leqslant \overline{X_n} + \sqrt{\frac{1}{4n\alpha}}\right)$$

   La valeur que l'on veut estimer est compris dans l'intervalle de confiance $\left[\overline{X_n} - \sqrt{\frac{1}{4n\alpha}}; \overline{X_n} + \sqrt{\frac{1}{4n\alpha}}\right]$ non pas de fa con sur mais avec un probabilité plus grande que $1-\alpha$.

*

Exercice 4 (Bernouilli : en utilisant le théorème limite-centrale)**.**
On reprend les mêmes notations que dans l'exercice 3. On note Φ la fonction de répartition de la loi normale centrée réduite.

1. Pourquoi Φ est elle une fonction bijective de $]-\infty; +\infty[$ dans $]0; 1[$.
2. Exprimer $\overline{X_n}^*$ la variable réduite centrée associée à $\overline{X_n}$.
3. En citant u théorème du cours expliquer pourquoi

$$\lim_{n\to+\infty} \mathbb{P}\left(a \leqslant \sqrt{n}\frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \leqslant b\right) = \Phi(b) - \Phi(a)$$

4. Soit $\alpha \in]0;\, 1[$, on pose $t_\alpha = \Phi^{-1}(1 - {}^{\alpha}/_{2})$. Montrer que

$$\lim_{n\to+\infty} \mathbb{P}\left(\overline{X_n} - \frac{t_\alpha p(1-p)}{\sqrt{n}} \leqslant p \leqslant \overline{X_n} + \frac{t_\alpha p(1-p)}{\sqrt{n}}\right) = 1 - \alpha$$

5. en déduire que $\left[\overline{X_n} - \frac{t_\alpha}{2\sqrt{n}};\, \overline{X_n} + \frac{t_\alpha}{2\sqrt{n}}\right]$ est un intervalle de confiance asymptotique au niveau $1 - \alpha$

RÉPONSE:

1. la fonction Φ est continue comme fonction de répartition d'une variable à densité ? De plus elle dérivable sur $\mathbb{R}$ et pour $t \in \mathbb{R}$

$$\Phi'(t) = \phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-{}^{t^2}/_{2}\right)$$

Donc ϕ est strictement positive et donc Φ est strictement croissante.
De plus

$$\lim_{x\to-\infty} \Phi(x) = 0 \qquad \lim_{x\to+\infty} \Phi(x) = 1$$

On peut donc appliquer le théorème de la bijection monotone pour affirmer qur Φ réalise une bijection de $]-\infty;\, +\infty[$ vers $]\lim_{x\to-\infty} \Phi(x);\, \lim_{x\to+\infty} \Phi(x)[\, =]0;\, 1[$.

2. **Outils utilisés**
   - $Y^* = \dfrac{Y - E(Y)}{\sigma_Y}$ la formule pour centrée et réduire une variable
   - Le calcul (classique)de $E(\overline{X_n})$ et $V(\overline{X_n})$

$S_n = X_1 + X_2 + \cdots + X_n$ est la somme de n variables de Bernoulli indépendante de même paramêtre p donc

$$S_n \hookrightarrow \mathcal{B}(n, p)$$

et on ait

$$\overline{X_n} = \frac{S_n}{n}$$

donc

$$E(\overline{X_n}) = \frac{1}{n} \times np = p \qquad \sigma_{\overline{X_n}} = \sqrt{V(\overline{X_n})} = \sqrt{\frac{1}{n^2} V(S_n)} = \sqrt{\frac{p(1-p)}{n}}$$

Donc

$$\begin{aligned}
\overline{X_n}^* &= \frac{\overline{X_n} - E(\overline{X_n})}{\sigma_{\overline{X_n}}} \\
&= \frac{\overline{X_n} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \\
&= \sqrt{n} \times \frac{\overline{X_n} - p}{\sqrt{p(1-p)}}
\end{aligned}$$

$$\overline{X_n}^* = \sqrt{n} \times \frac{\overline{X_n} - p}{\sqrt{p(1-p)}}$$ est la variable réduite centrée associée à $\overline{X_n}$.

3. Les variables aléatoires étant toute de même loi indépendantes et admettant une variance, nous pouvons appliquer le théorème central de la limite, pour a et b deux réels tels que $a \leqslant b$

$$\lim_{n\to+\infty} \mathbb{P}\left(a \leqslant \overline{X_n}^* \leqslant b\right) = \Phi(b) - \Phi(a)$$

$$\lim_{n\to+\infty} \mathbb{P}\left(a \leqslant \sqrt{n}\frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \leqslant b\right) = \Phi(b) - \Phi(a)$$

4. **outils utilisés**
   - manipulation d'inégalité
   - propriétés de Φ.

   Par définition $t_\alpha = \Phi^{-1}(1 - \alpha/2)$, donc

$$\Phi(t_\alpha) = 1 - \alpha/2$$

De plus on sait que pour tout réel x

$$\Phi(-x) = 1 - \Phi(x)$$

donc

$$\Phi(-t_\alpha) = 1 - (1 - \alpha/2) = \alpha/2$$

de plus

$$\begin{aligned}
-t_\alpha \leqslant \sqrt{n}\frac{\overline{X_n} - p}{\sqrt{p(1-p)}} \leqslant t_\alpha &\Leftrightarrow -t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \overline{X_n} - p \leqslant t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} && \text{positivité}\\
&\Leftrightarrow -t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant p - \overline{X_n} \leqslant t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} && \text{en } \times -1\\
&\Leftrightarrow \overline{X_n} - t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant p \leqslant t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} + \overline{X_n}
\end{aligned}$$

La question précédente donne

$$\lim_{n\to+\infty} \mathbb{P}\left(-t_\alpha \leqslant \sqrt{n}\frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \leqslant t_\alpha\right) = \Phi(t_\alpha) - \Phi(-t_\alpha)$$

et donc

$$\lim_{n\to+\infty} \mathbb{P}\left(\overline{X}_n - t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant p \leqslant t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} + \overline{X}_n\right) = 1 - \alpha/2 - \alpha/2$$

$$\lim_{n\to+\infty} \mathbb{P}\left(\overline{X_n} - \frac{t_\alpha p(1-p)}{\sqrt{n}} \leqslant p \leqslant \overline{X_n} + \frac{t_\alpha p(1-p)}{\sqrt{n}}\right) = 1 - \alpha$$

5. C'est la définition du cours

$*$

Exercice 5.
On veut estimer la masse m d'un certain objet. Pour cela, on effectue des pesées successives et l'on note m la moyenne obtenue. On admet que la variable aléatoire renvoyant le résultat d'une pesée de l'objet étudié suit la loi normale $\mathcal{N}(m, \sigma^2)$ avec $\sigma = 0.1$. On effectue une suite de pesées et on note X_i le résultat le la i-ième pesée.

1. . On note $\overline{X_n} = \dfrac{1}{n} \sum_{k=1}^{n} X_k$. Quelle est la loi suivie par $\overline{X_n}$?
2. Calculer l'espérance et la variance de $\overline{X_n}$.
3. On note $\overline{X_n}^* = \sqrt{n}\dfrac{\overline{X_n} - m}{\sigma}$. Quelle est la loi suivie par $\overline{X_n}^*$?
4. On effectue 10 pesées et on obtient une masse moyenne de 45,5 grammes.
5. Trouver un réel x tel que
$$\mathbb{P}(|\overline{X}_{10} - m| \leqslant \frac{\sigma}{\sqrt{10}}x) \geqslant 0.9$$
On donne $\Phi(1,65) \approx 0,95$
6. Montrer que $\left[45,5 - \dfrac{\sigma}{\sqrt{10}}x;\ 45,5 + \dfrac{\sigma}{\sqrt{10}}x\right]$ est un intervalle de confiance au niveau de confiance 90% pour la masse m.
7. Déterminer quelle valeur de n il faut choisir pour que l'intervalle de confiance de niveau de confiance 90% soit de longueur 0.05.