

- EXERCICE 1 -

On reprend l'exemple et on note $(X_1, \dots, X_n)$ un n -échantillon de X .

On considère $T = X_1$ et $\overline{X}_n$ la moyenne empirique de $(X_1, \dots, X_n)$.

1. Vérifier que T et $\overline{X}_n$ sont deux estimateurs de p .

La variable $T = X_1$ est bien une fonction des variables $(X_1, \dots, X_n)$ et ne dépendant pas de p . C'est donc un estimateur de p .

La variable $\overline{X}_n$ est bien une fonction des variables $(X_1, \dots, X_n)$ et ne dépendant pas de p . C'est donc un estimateur de p .

2. On interroge 100 personnes. La première répond "non" et sur le total, 27 répondent "oui".

Donner les estimations correspondants aux estimateurs T et $\overline{X}_{100}$.

La valeur prise par $T = X_1$ est 0 par hypothèse donc l'estimation associée au sondage de l'exemple est $t = x_1 = 0$.

La valeur prise par $\overline{X}_n$ est 0,27 par hypothèse donc l'estimation associée au sondage de l'exemple est $\overline{x}_n = 0,27$.

- EXERCICE 2 -

On reprend l'exercice 1 et on note $(X_1, \dots, X_n)$ un n -échantillon de X .

1. Calculer le biais et le risque quadratique de l'estimateur $T_n = X_1$ puis de l'estimateur $\overline{X}_n$.

- On a, puisque $X_1 \hookrightarrow \mathcal{B}(p)$:

$$E(T_n) = E(X_1) = p$$

donc l'estimateur T_n est sans biais (car $E(T_n) - p = 0$).

Comme l'estimateur T_n est sans biais, son risque quadratique est égal à sa variance soit :

$$r(T_n) = V(T_n) = V(X_1) = pq$$

- On a par linéarité de l'espérance :

$$E(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{np}{n} = p$$

donc l'estimateur $\overline{X}_n$ est sans biais.

Comme l'estimateur $\overline{X}_n$ est sans biais, son risque quadratique est égal à sa variance soit par indépendance des variables :

$$r(\overline{X}_n) = V(\overline{X}_n) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{npq}{n} = \frac{pq}{n}$$

2. Quel estimateur de p choisir?

Les deux estimateurs sont sans biais mais $\overline{X}_n$ a un risque quadratique plus faible (qui tend vers 0 lorsque n tend vers $+\infty$). On a donc intérêt à choisir $\overline{X}_n$ comme estimateur de p .

- EXERCICE 3 -

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

Montrer que $\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$ est un estimateur sans biais et convergent de m .

Il s'agit de montrer que $\lim_{n \rightarrow +\infty} P(|\overline{X}_n - m| \geq \epsilon) = 0$.

On peut tout simplement remarquer qu'il s'agit de la loi faible des grands nombres ou refaire la démonstration avec l'IBT.

$\overline{X}_n$ admet une espérance et une variance comme somme finie de variables aléatoires en admettant et on a :

- $E(\overline{X}_n) = E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} nm = m$ par linéarité de l'espérance et car les X_i ont même espérance m .
- $V(\overline{X}_n) = V\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n V(X_k) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$ par indépendance et car les X_i ont même variance σ^2 .

On a d'après l'inégalité de BT :

$$\forall \epsilon > 0, \quad 0 \leq P(|\overline{X}_n - m| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Or, $\lim_{n \rightarrow +\infty} \frac{\sigma^2}{n\epsilon^2} = 0$ donc d'après le théorème des gendarmes, on a : $\forall \epsilon > 0, \quad \lim_{n \rightarrow +\infty} P(|\overline{X}_n - m| \geq \epsilon) = 0$.

- EXERCICE 4 -

Soit T_n un estimateur de θ admettant un moment d'ordre 2.

1. Montrer, en utilisant l'inégalité de Markov, que : $\forall \epsilon > 0, \quad P(|T_n - \theta| \geq \epsilon) \leq \frac{E((T_n - \theta)^2)}{\epsilon^2} = \frac{r_\theta(T_n)}{\epsilon^2}$

L'inégalité de Markov s'écrit pour Y admettant une espérance : $\forall \epsilon > 0, \quad P(|Y| \geq \epsilon) \leq \frac{E(Y)}{\epsilon^2}$.

En remarquant que $|T_n - \theta| \geq \epsilon \Leftrightarrow |T_n - \theta|^2 \geq \epsilon^2$ puis en choisissant $Y = |T_n - \theta|^2$ dans l'inégalité de Markov (Y admet une espérance car T_n admet un moment d'ordre 2), on obtient :

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

2. Démontrer alors le théorème 2.

On a vu :

$$0 \leq P(|T_n - \theta| \geq \epsilon) \leq \frac{r_\theta(T_n)}{\epsilon^2}$$

Si $\lim_{n \rightarrow +\infty} r_\theta(T_n) = 0$, alors d'après le théorème des gendarmes, on en déduit que $\lim_{n \rightarrow +\infty} P(|T_n - \theta| \geq \epsilon) = 0$ ce qui prouve T_n est un estimateur convergent de θ (par définition).