

Estimations Chapitre 14

Preuve des résultats importants du cours

Estimation ponctuelle

Estimateur

- Preuve :  Moyenne empirique : espérance et variance

On considère une var X d'espérance m et de variance σ^2 et un n -échantillon $(X_1, \dots, X_n)$ de la loi de X . En notant $\bar{X}_n$ la moyenne empirique associé à cet échantillon, on a : $E(\bar{X}_n) = m$, $V(\bar{X}_n) = \frac{\sigma^2}{n}$.

Par linéarité de l'espérance que $E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{m}{n} \sum_{i=1}^n 1 = \frac{nm}{n} = m$.

Puis par indépendance des variables X_i (car $(X_1, \dots, X_n)$ est un n -échantillon de la loi de X) et par propriété de la variance : $V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n^2} \sum_{i=1}^n 1 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

NB : $\bar{X}_n$ la moyenne empirique associé à cet échantillon est un estimateur de l'espérance m de X : car $\bar{X}_n$ est fonction du n -échantillon $(X_1, \dots, X_n)$ de la loi de X et indépendant de m dans sa formulation.

Soit un n -échantillon $(X_1, \dots, X_n)$ d'une loi d'espérance m .

Comme $E(\bar{X}_n) = m$, la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur sans biais de m .

Car le biais est égal à $b_m(\bar{X}_n) = E(\bar{X}_n) - m = 0$.

La moyenne empirique est un estimateur convergent de l'espérance m de X .

Montrons que la moyenne empirique est un estimateur convergent de l'espérance m de X .

C'est à dire montrons que : $\forall \epsilon > 0, \lim_{n \rightarrow +\infty} P(|\bar{X}_n - m| > \epsilon) = 0$.

Pour cela, appliquons l'inégalité de Bienaymé-Tchebychev à $\bar{X}_n$, on peut car $\bar{X}_n$, admet un moment d'ordre 2 :

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

$$\text{Or, } E(\bar{X}_n) = m \text{ et } V(\bar{X}_n) = \frac{\sigma^2}{n}.$$

D'où l'inégalité de B-T donne :

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

On a $|\bar{X}_n - m| > \epsilon \subset |\bar{X}_n - m| \geq \epsilon$ d'où par croissance des probabilités et par positivité des probabilités :

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

De plus $\lim_{n \rightarrow +\infty} \frac{\sigma^2}{n\epsilon^2} = 0$ donc par théorème d'encadrement, on en déduit que $\lim_{n \rightarrow +\infty} P(|\bar{X}_n - m| > \epsilon) = 0$. Donc la moyenne empirique est un estimateur convergent de l'espérance m de X .

Risque quadratique

[HP] Attention ceci est HORS PROGRAMME!

- Preuve : Décomposition biais-variance du risque quadratique

Si un estimateur T_n a un moment d'ordre 2, alors :

$$r_\theta(T_n) = V(T_n) + (b_\theta(T_n))^2.$$

En effet, on a

$$\begin{aligned} r_\theta(T_n) &= E((T_n - \theta)^2) \\ &= E(T_n^2 - 2\theta T_n + \theta^2) \\ &= E(T_n^2) - 2\theta E(T_n) + \theta^2 \text{ par linéarité et propriété de l'espérance} \\ &= E(T_n^2) - E(T_n)^2 + E(T_n)^2 - 2\theta E(T_n) + \theta^2 \\ &= V(T_n) + (E(T_n) - \theta)^2 \\ &= V(T_n) + (b_\theta(T_n))^2 \end{aligned}$$

Estimateur convergent

- Preuve : Condition suffisante de convergence d'un estimateur[HP]

Si pour tout $\theta \in \Theta$, l'estimateur T_n possède un moment d'ordre 2, alors la suite $(T_n)_{n \in \mathbb{N}^*}$ est un estimateur convergent de θ si pour tout $\theta \in \Theta$, $\lim_{n \rightarrow +\infty} r_\theta(T_n) = 0$.

On va appliquer l'inégalité de Markov à $(T_n - \theta)^2$ var positive et ayant une espérance par hypothèses

Alors :

$$\forall a > 0, P((T_n - \theta)^2 \geq a) \leq \frac{E((T_n - \theta)^2)}{a}$$

Avec $\epsilon^2 = a$ et $E((T_n - \theta)^2) = r_\theta(T_n)$, cela donne :

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

Or, $[(T_n - \theta)^2 \geq \epsilon^2] = [|T_n - \theta| \geq \epsilon]$ et une probabilité est toujours positive d'où :

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

Alors si $\lim_{n \rightarrow +\infty} r_\theta(T_n) = 0$, on a par théorème d'encadrement que $\lim_{n \rightarrow +\infty} P(|T_n - \theta| > \epsilon) = 0$. Ce qui signifie que l'estimateur est bien convergent.

1. Si T_n est sans biais, alors la condition du théorème s'écrit : $\lim_{n \rightarrow +\infty} V(T_n) = 0$.
2. Si T_n est asymptotiquement sans biais, alors grâce à la décomposition biais-variance du risque quadratique, la condition du théorème s'écrit aussi : $\lim_{n \rightarrow +\infty} V(T_n) = 0$.
3. De deux estimateurs de θ , on choisit celui qui a le plus petit risque quadratique (ou mieux, celui qui a le risque quadratique qui tend le plus vite vers 0) car ainsi, avec l'inégalité de Markov ou celle de Bienaymé-Tchebychev, on minimise plus vite la probabilité que l'estimateur s'écarte de θ de plus de ϵ .

NB :

On établit la convergence de l'estimateur T_n grâce à l'inégalité de Bienaymé-Tchebychev appliquée à T_n lorsque $E(T_n) = \theta$, ou grâce à l'inégalité de Markov appliquée à $(T_n - \theta)^2$ si $E(T_n) \neq \theta$.