RESUME DU COURS DE MATHEMATIQUES

ESTIMATION

ESTIMATION

ECHANTILLONS

DEFINITION−1 : *Soit $\mathcal{L}$ une loi de probabilité définie sur un espace probabilisé $(\Omega, \mathcal{T}, P)$. On appelle échantillon aléatoire de taille n (ou $n-$échantillon) de la loi $\mathcal{L}$ toute suite $\varepsilon_n = (X_1, \ldots, X_n)$ définie sur (Ω) et à valeurs dans $\mathbb{R}^n$ où $X_1, \ldots, X_n$ sont n variables aléatoires , mutuellement indépendantes définies sur l'espace probabilisé $(\Omega, \mathcal{T}, P)$ et qui suivent la même loi de probabilité $\mathcal{L}$.*

Pour tout $\omega \in \Omega$ tel que $\{\omega\} \in \mathcal{T}$ la réalisation de ε_n, c'est-à-dire $(X_1(\omega), \ldots, X_n(\omega))$, notée $(x_1, \ldots, x_n)$, s'appellera un échantillon observé.

DEFINITION−2 : *On appelle statistique sur un échantillon $\varepsilon_n = (X_1, \ldots, X_n)$ d'une loi de probabilité $\mathcal{L}$ toute variable aléatoire $Y = \varphi(X_1, \ldots, X_n)$ sur $(\Omega, \mathcal{T}, P)$ où φ est une application de $\mathbb{R}^n$ dans $\mathbb{R}$*

DEFINITION−3 : *Soit $\varepsilon_n = (X_1, \ldots, X_n)$ un $n-$échantillon de la loi $\mathcal{L}$. La statistique $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ s'appelle moyenne empirique de la loi $\mathcal{L}$ associée à l'échantillon ε_n*

THEOREME : Soit $\overline{X_n}$ la moyenne empirique associée à un $n-$échantillon d'une loi $\mathcal{L}$ d'espérance m et de variance σ^2, alors

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

THEOREME : La moyenne empirique $\overline{X_n}$ associée à un $n-$échantillon d'une loi $\mathcal{L}$ d'espérance m converge en proba-

bilité (donc en loi) vers une variable certaine égale à m.

THEOREME : Soit $\overline{X_n}$ la moyenne empirique associée à un $n-$échantillon d'une loi $\mathcal{L}$ d'espérance m et de variance σ^2, alors la variable centrée réduite associée a $\overline{X_n}$ converge en loi vers une variable Y qui suit la loi normale $\mathcal{N}(0,1)$.

$$\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}} \text{ converge en loi vers } Y$$

ESTIMATEURS

Position du problème

On suppose que les résultats observés sur l'échantillon $(x_1, \ldots, x_n)$ sont les réalisations d'une variable X, dont la loi est inconnue. On veut, à partir de l'échantillon observé, rechercher quelle loi théorique on peut prendre pour X. Dans certains cas, on ne connaît rien a priori sur la loi de X. Mais, très souvent, on est dans la situation où la loi cherchée appartient à une certaine famille de lois (classique ou pas) dépendant d'un paramètre θ, que nous noterons $\mathcal{L}_\theta$, et on suppose que l'on connaît l'ensemble Θ dans lequel varie le paramètre θ. Par exemple, pour une famille de lois de Bernoulli, le paramètre p varie dans $[0, 1]$, pour une famille de lois de Poisson, comme pour une famille de lois exponentielles le paramètre λ varie dans $\mathbb{R}^{+*}$.

La loi que l'on cherche à évaluer est alors déterminée par son paramètre θ_0 et on suppose raisonnablement qu'à deux valeurs distinctes du paramètre correspondent deux lois différentes.

En résumé : on veut estimer la valeur θ_0 du paramètre d'une loi à partir d'un échantillon aléatoire d'une variable X qui suit cette loi inconnue, valeur θ_0 qui permet de déterminer sans ambiguïté la loi des variables $X_1, \ldots, X_n$ qui constituent l'échantillon, qui sont mutuellement indépendantes et qui suivent la même loi que X. Dans la pratique, on est souvent amené à estimer non pas directement θ_0 mais une fonction de θ_0

DEFINITION : *Soit $g\ :\ \Theta \longrightarrow \mathbb{R}$ et $\varepsilon_n = (X_1, \ldots, X_n)$ un $n-$échantillon d'une loi μ_{θ_0}. On appelle estimateur de $g(\theta_0)$ toute statistique (toute fonction de $X_1, \ldots, X_n$) qui prend ses valeurs dans $g(\Theta)$, ensemble des valeurs possibles de $g(\theta_0)$*

On notera $T_n = \varphi(X_1, \ldots, X_n)$ un estimateur de $g(\theta_0)$

Toutes les variables X_i $(1 \leq i \leq n)$ suivent la même loi μ_θ, donc T_n dépend de θ. Les valeurs observées de l'échantillon, grâce auxquelles on cherchera à évaluer $g(\theta_0)$ sont des réalisations de T_n

DEFINITION : *Soit $T_n = \varphi(X_1, \ldots, X_n)$ un estimateur de $g(\theta_0)$. Une estimation de $g(\theta_0)$ est une réalisation de T_n, donc la valeur $\varphi(x_1, \ldots, x_n)$ où $(x_1, \ldots, x_n)$ est une réalisation de l'échantillon $(X_1, \ldots, X_n)$.*

En pratique on désignera par la même lettre θ la variable et la vraie valeur θ_0 que l'on cherche à estimer.

Exemples :

1) on dispose d'un dé dont ne sait pas s'il est équilibré. On le lance n fois de manière indépendante et on choisit une fois pour toutes $i \in [\![1, 6]\!]$ (on cherche donc à estimer la probabilité d'apparition du numéro i)

On considère la variable de Bernoulli X_k qui prend la valeur 1 si au k ème lancer le numéro i est apparu et qui prend la valeur 0 sinon. Le paramètre de X_k est $p = P(\text{numéro } i \text{ apparaisse })$. Ici $\theta = p$, la famille de loi est la famille de loi de Bernoulli, de paramètre p inconnu et $\Theta = [0, 1]$. On pourra