V - Estimation

I - Définitions

Définition 1 - Échantillon

Soient n un entier naturel non nul et X une variable aléatoire. Un *n-échantillon* de X est un *n*-uplet (X_1, \ldots, X_n) tel que les variables aléatoires X_1, \ldots, X_n soient mutuellement indépendantes et de même loi que X.

Définition 2 - Estimateur

Soient X une variable aléatoire suivant une loi de probabilité dépendant d'un paramètre θ variant dans un intervalle I de $\mathbb R$ et (X_1,\ldots,X_n) un *n*-échantillon de X. Un estimateur de θ est une variable aléatoire $\varphi(X_1,\ldots,X_n)$ où φ est une application de \mathbb{R}^n dans \mathbb{R} .

Exemple 1 - Estimateurs

- Soit (X_1,\ldots,X_n) un n-échantillon d'une variable aléatoire X. Les quantités suivantes sont des estimateurs
 - de $\mathbf{E}[X]$:

 * $T_1 = \frac{1}{n} \sum_{k=1}^{n} X_k$.

 * $T_3 = \frac{X_1 + X_2}{2}$.

 * $T_4 = \frac{X_1 X_2}{2}$.

- $\star T_2 = \frac{1}{n-1} \sum_{k=1}^n X_k.$
- Si $X \hookrightarrow \mathcal{B}(p)$, on peut chercher à estimer p (qui est égal à son espérance), sa variance $\sigma^2 = p(1-p)...$
- Si $X \hookrightarrow \mathscr{P}(\lambda)$, on peut chercher à estimer λ (qui est égal à son espérance), la quantité $\mathbf{P}([X=0]) = e^{-\lambda}, \dots$
- Si $X \hookrightarrow \mathcal{U}(\llbracket 1, n \rrbracket)$, on peut chercher à estimer n: avec les résultats d'un dé, on peut estimer son nombre de faces.

Définition 3 - Biais

Soit T un estimateur de θ . Le biais de T est le réel $b_{\theta}(T) = \mathbf{E}[T] - \theta$. Si $b_{\theta}(T) = 0$, l'estimateur T est un estimateur sans biais de θ .

Exemple 2 - Calculs de biais

Reprenons l'exemple précédent en notant $m = \mathbf{E}[X]$.

• Soit $T_1 = \frac{1}{n} \sum_{k=1}^{n} X_k$. D'après la linéarité de l'espérance,

$$\mathbf{E}[T_1] - m = \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}[X_k] - m = \frac{nm}{n} - m = 0.$$

Ainsi, T_1 est un estimateur sans biais de m.

• Soit $T_2 = \frac{1}{n-1} \sum_{k=1}^{n} X_k$. D'après la linéarité de l'espérance,

$$\mathbf{E}[T_2] - m = \frac{1}{n-1} \sum_{k=1}^{n} \mathbf{E}[X_k] - m = \frac{nm}{n-1} - m = \frac{m}{n(n-1)}.$$

Ainsi, T_2 est un estimateur de m de biais $\frac{m}{n(n-1)}$.

• Soit $T_3 = \frac{X_1 + X_2}{2}$. D'après la linéarité de l'espérance.

$$\mathbf{E}[T_3] - m = \frac{\mathbf{E}[X_1] + \mathbf{E}[X_2]}{2} - m = 0.$$

Ainsi, T_3 est un estimateur sans biais de m.

• Soit $T_4 = \frac{X_1 X_2}{2}$. Les variables aléatoires étant indépen-

$$\mathbf{E}[T_4] - m = \frac{\mathbf{E}[X_1]\mathbf{E}[X_2]}{2} - m = \frac{m^2}{2} - m = \frac{m(m-2)}{2}.$$

Ainsi, T_4 est un estimateur de m de biais $\frac{m(m-2)}{2}$.

Chapitre V - Estimation D 2

Définition 4 - Risque quadratique

Soit T un estimateur de θ . Le risque quadratique de T est défini par

$$R_{\theta}(T) = \mathbf{E}\left[(T-\theta)^2\right].$$

En particulier,

$$R_{\theta}(T) = b_{\theta}(T)^2 + \mathbf{V}(T)$$

Si T est un estimateur sans biais de θ , alors $R_{\theta}(T) = \mathbf{V}(T)$.

Exemple 3 - Calculs de risques quadratiques

Reprenons les exemples précédents, pour les estimateurs sans biais. On suppose que X admet une variance.

 \bullet Comme T_1 est un estimateur sans biais de m, alors les variables aléatoires étant indépendantes,

$$R_m(T_1) = \mathbf{V}(T_1) = \mathbf{V}\left(\frac{1}{n}\sum_{k=1}^n X_k\right) = \frac{1}{n^2}\sum_{k=1}^n \mathbf{V}(X_k)$$
$$= \frac{\mathbf{V}(X)}{n}.$$

• Comme T_3 est sans biais et X_1 et X_2 sont indépendantes,

$$R_m(T_3) = \mathbf{V}(T_3) = \frac{\mathbf{V}(X_1) + \mathbf{V}(X_2)}{4} = \frac{\mathbf{V}(X)}{2}.$$

Définition 5 - Meilleur estimateur

L'estimateur T_1 est un meilleur estimateur que T_2 si

$$\forall \theta \in I, R_{\theta}(T_1) \leqslant R_{\theta}(T_2).$$

Exemple 4 - Comparaison d'estimateurs

En reprenant l'exemple précédent, T_1 est un meilleur estimateur de $\mathbf{E}[X]$ que T_3 dès que $n \ge 2$.

II - Estimation d'une proportion

Théorème 1 - Loi faible des grands nombres

Soit X une variable aléatoire admettant une moyenne m et une variance σ^2 . Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi que X. Alors,

$$\forall \ \varepsilon > 0, \ \lim_{n \to +\infty} \mathbf{P} \left(\left| \frac{1}{n} \sum_{k=1}^{n} X_k - m \right| \geqslant \varepsilon \right) = 0.$$

Définition 6 - Estimateur convergent

Soit T_n un estimateur de θ . L'estimateur T_n est convergent si pour tout $\varepsilon > 0$, $\lim_{n \to +\infty} \mathbf{P}(|T_n - \theta| \ge \varepsilon) = 0$.

Théorème 2 - Estimation ponctuelle d'une proportion

Soit X une variable aléatoire qui suit une loi de Bernoulli de paramètre p. Soit (X_1, \ldots, X_n) un n-échantillon de X et $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Alors, \overline{X}_n est un estimateur sans biais et convergent de p.

Exemple 5 - Sondage

On souhaite connaître la proportion de français favorables à une réforme donnée. On modélise la réponse d'un individu en considérant une variable aléatoire X suivant une loi de Bernoulli de paramètre p. On interroge n français choisis indépendamment dans la population. On note X_i la réponse donnée par le i^e individu interrogé : 1 si l'individu est favorable et 0 sinon. On suppose que X_i suit la même loi que X. Pour estimer p, on va donc utiliser la quantité $\frac{X_1+\dots+X_n}{n}$.