

Je vous propose ici quelques méthodes classiques de recherche d'estimateurs, ainsi que de détermination de d'intervalles de confiance ne reposant pas tous sur le théorème limite central.

Dans ce problème, on souhaite estimer le paramètre θ (où $\theta \in]0; +\infty[$) d'une loi uniforme sur l'intervalle $[0; \theta]$.

n désigne un entier naturel non nul, et $(X_1, \dots, X_n)$ est un n -échantillon d'une variable X de loi uniforme sur $[0; \theta]$.

Lorsque que T_n est un estimateur de θ , on note $r_\theta(T_n)$ son risque quadratique.

α désigne un réel de $]0; 1[$, appelé *risque*, dont la *confiance* associée $1 - \alpha$ et on note t_α l'unique réel strictement positif tel que $\mathbb{P}(|Z| \leq t_\alpha) = 1 - \alpha$ pour toute variable Z de loi normale centrée réduite. Ainsi, en notant Φ la fonction de répartition de la loi normale centrée réduite, α et t_α sont liés par la relation

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

A - Construction d'estimateurs ponctuels

1. Méthode des moments

Dans cette méthode, on exploite un lien entre un moment - ici l'espérance -, et le paramètre θ .

(a) Rappeler $\mathbb{E}(X)$ et justifier que la moyenne empirique

$$\overline{X}_n \stackrel{\text{déf.}}{=} \frac{1}{n} \sum_{k=1}^n X_k$$

est un estimateur sans biais et convergent de $\frac{\theta}{2}$.

(b) En déduire que

$$M_n \stackrel{\text{déf.}}{=} 2\overline{X}_n$$

est un estimateur convergent de θ .

(c) Préciser le risque quadratique de M_n .

(d) On considère un estimateur T_n combinaison linéaire des X_i , c'est-à-dire qu'il existe n réels $\alpha_1, \dots, \alpha_n$ tels que

$$T_n = \sum_{k=1}^n \alpha_k X_k.$$

i. Donner une condition nécessaire et suffisante pour que T_n soit un estimateur sans biais de θ .

ii. On suppose que T_n est un estimateur sans biais de θ .

Montrer que le risque quadratique de T_n est minimum si, et seulement si, $T_n = M_n$. On pourra utiliser l'inégalité de Cauchy-Schwarz.

Ainsi, parmi tous les estimateurs sans biais combinaisons linéaires des X_i , M_n est le plus efficace.

2. Méthode du maximum de vraisemblance

Soit $(x_1, \dots, x_n)$ une réalisation de l'échantillon $(X_1, \dots, X_n)$. On cherche la valeur de θ rendant cette réalisation la probable.

Soit f_θ la densité de X définie par

$$f_\theta : x \mapsto \begin{cases} \frac{1}{\theta} & \text{si } x \in [0; \theta], \\ 0 & \text{sinon.} \end{cases}$$

(a) Montrer que $f_\theta(x_1) \times f_\theta(x_2) \times \dots \times f_\theta(x_n)$ est maximum lorsqu'on a $\theta = \max(x_1, x_2, \dots, x_n)$.

On pose alors

$$S_n \stackrel{\text{déf.}}{=} \max(X_1, X_2, \dots, X_n).$$

(b) Déterminer une densité de S_n et montrer que S_n est un estimateur asymptotiquement sans biais de θ .

(c) Déterminer le risque quadratique de S_n et justifier que S_n est un estimateur convergent de θ .

(d) Justifier que

$$R_n \stackrel{\text{déf.}}{=} \frac{n+1}{n} S_n$$

est un estimateur sans biais et convergent de θ .

(e) R_n est-il plus efficace (i.e. de risque quadratique inférieur) que S_n ?

3. Méthode des statistiques d'ordre

Dans cette méthode, on exploite un lien entre un quantile - ici la médiane -, et le paramètre θ .

(a) Justifier qu'il existe un unique réel m , que l'on déterminera, vérifiant

$$\mathbb{P}(X \leq m) = \mathbb{P}(X \geq m).$$

Dans toute la suite de cette question 3., on suppose que n est un entier impair et on note k l'unique entier tel que

$$n = 2k + 1.$$

On ordonne les variables $X_1, X_2, \dots, X_n$ et on note $Y_1, Y_2, \dots, Y_n$ les variables obtenues.

Autrement dit, soit σ une bijection de $\llbracket 1; n \rrbracket$ sur lui-même telle que

$$\begin{aligned} \forall i \in \llbracket 1; n \rrbracket, \quad Y_i &= X_{\sigma(i)}, \\ Y_1 &\leq Y_2 \leq \dots \leq Y_n. \end{aligned}$$

Soit enfin

$$\mu_n \stackrel{\text{déf.}}{=} Y_{k+1}$$

la *médiane empirique* de l'échantillon $(X_i)_{1 \leq i \leq n}$. On notera qu'il y a ainsi k variables X_i inférieures à μ_n , et k supérieures à μ_n .

(b) Dans cette question, on détermine une densité f_{μ_n} de μ_n . Soit $x \in [0; \theta]$.

i. Soit, pour tout i de $\llbracket 1; n \rrbracket$, χ_i la variable indicatrice de l'événement $[X_i \leq x]$ et soit

$$\Sigma_n = \sum_{i=1}^n \chi_i.$$

Justifier que Σ_n suit la loi binomiale $\mathcal{B}\left(n, \frac{x}{\theta}\right)$.

ii. Comparer les événements $[\mu_n \leq x]$ et $[\Sigma_n \geq k+1]$.

iii. En déduire que la fonction de répartition F_{μ_n} vérifie

$$F_{\mu_n}(x) = \frac{1}{\theta^n} \sum_{i=k+1}^n \binom{n}{i} x^i (\theta - x)^{n-i}.$$

iv. En déduire qu'on définit une densité de μ_n en posant

$$f_{\mu_n}(x) \stackrel{\text{déf.}}{=} \begin{cases} \frac{n}{\theta^n} \binom{n-1}{\frac{n-1}{2}} x^{\frac{n-1}{2}} (\theta - x)^{\frac{n-1}{2}} & \text{si } x \in [0; \theta], \\ 0 & \text{sinon.} \end{cases}$$

(c) Dans cette question, on détermine les moments de μ_n .

Pour p et q entiers naturels, on pose

$$I_{p,q} \stackrel{\text{déf.}}{=} \int_0^1 u^p (1-u)^q du.$$

i. Montrer que, pour $p \in \mathbb{N}$ et $q \in \mathbb{N}^*$, $I_{p,q} = \frac{q}{p+1} I_{p+1,q-1}$.

ii. En déduire, pour tout p et q de $\mathbb{N}$, $I_{p,q} = \frac{p!q!}{(p+q+1)!}$.

iii. En déduire que $\mathbb{E}(\mu_n) = \frac{\theta}{2}$ et $\mathbb{V}(\mu_n) = \frac{\theta^2}{4(n+2)}$.

(d) On pose $\nu_n \stackrel{\text{déf.}}{=} 2\mu_n$.

Justifier que ν_n est un estimateur sans biais et convergent de θ .

4. Comparaison de ces quatre estimateurs

Comparer les risques quadratiques des estimateurs M_n , S_n , R_n et ν_n , puis les classer du plus efficace au moins efficace.

B - Construction d'intervalles de confiance**5. Intervalles de confiance par M_n**

- (a) Justifier que $\sqrt{3n} \frac{M_n - \theta}{\theta}$ converge en loi vers la loi normale centrée réduite.
- (b) En déduire que $\sqrt{3n} \frac{M_n - \theta}{M_n}$ converge en loi vers la loi normale centrée réduite.
- (c) Montrer que $\left[M_n - \frac{M_n t_\alpha}{\sqrt{3n}}; M_n + \frac{M_n t_\alpha}{\sqrt{3n}} \right]$ est un intervalle de confiance asymptotique de θ de niveau de confiance $1 - \alpha$.

6. Intervalles de confiance par S_n et R_n

- (a) Déterminer un réel λ_n , dépendant de α et de n , tel que $[S_n; \lambda_n S_n]$ soit un intervalle de confiance de θ au niveau de confiance $1 - \alpha$.
- (b) En déduire un intervalle de confiance de θ de niveau de confiance $1 - \alpha$ construit à l'aide de R_n .

7. Intervalles de confiance par les statistiques d'ordre

- (a) On suppose $n = 2k$ avec $k \in \mathbb{N}^*$.
On note D_n (respectivement U_n) le nombre de variables X_i telles que $X_i < \frac{\theta}{2}$ (respectivement telles que $X_i > \frac{\theta}{2}$). Quelle est la loi suivie par D_n et U_n ?
- (b) Justifier que $\mathbb{P}(D_n \geq k) = \mathbb{P}(D_n \leq k)$, puis que $\mathbb{P}(D_n \geq k) > \frac{1}{2}$.
- (c) Soit s_α le plus petit entier tel que $\mathbb{P}(D_n \geq s_\alpha) \leq \frac{\alpha}{2}$.
Justifier l'existence de s_α , et montrer que $s_\alpha > \frac{n}{2}$.
- (d) Justifier que $\mathbb{P}\left(\frac{\theta}{2} > Y_{s_\alpha}\right) \leq \frac{\alpha}{2}$.
- (e) Montrer que $[2Y_{s_\alpha}; 2Y_{n-s_\alpha+1}]$ est un intervalle de confiance de niveau de confiance au moins $1 - \alpha$ de θ .
- (f) Par quelle loi peut-on approcher la loi commune de D_n et U_n lorsque n devient grand ?
En déduire le calcul approché de s_α suivant :

$$s_\alpha \simeq \frac{n + t_\alpha \sqrt{n}}{2}.$$

C - Simulation et calcul des intervalles avec SCILAB**8. Simulation**

Compléter l'algorithme suivant afin de générer un n -échantillon de X et de calculer les quatre estimateurs sur cet échantillon.

```
t=input('theta ?'); n=input('n ?');
X=grand(1,...,"unf",0,...);
Mn=.....; Sn=.....; Rn=.....; nun=.....;
```

9. Calcul des intervalles de confiance

On rappelle qu'en Scilab `cdfnor("X",0,1,y,1-y)` permet de calculer $\Phi^{-1}(y)$.

Compléter le script précédant pour qu'il demande α puis calcule, à l'aide des variables **Mn**, **Sn** et **nun**, les intervalles de confiance construits dans la partie précédente.