

Notations et préliminaires.

Le modèle est $X_i \sim \mathcal{U}[0, \theta]$, $\theta > 0$ inconnu, de densité

$$f_\theta(x) = \frac{1}{\theta} \mathbf{1}_{[0, \theta]}(x),$$

d'où

$$\mathbb{E}[X] = \frac{\theta}{2}, \quad \mathbb{V}[X] = \frac{\theta^2}{12}.$$

On note $\hat{\theta}_i$ les estimateurs demandés, $M_n = \max(X_1, \dots, X_n)$, et

$$R(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{V}[\hat{\theta}] + b(\hat{\theta})^2$$

l'erreur quadratique moyenne, avec $b(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$.

1°. Méthode des moments. On égalise moment théorique et moment empirique d'ordre 1 :

$$\mathbb{E}[X] = \frac{\theta}{2} = \bar{X} \implies \hat{\theta}_1 = 2\bar{X}.$$

Biais : $\mathbb{E}[\hat{\theta}_1] = 2\mathbb{E}[\bar{X}] = 2 \cdot \frac{\theta}{2} = \theta$, l'estimateur est sans biais. Erreur quadratique : comme $b = 0$,

$$R(\hat{\theta}_1) = \mathbb{V}[2\bar{X}] = 4 \frac{\mathbb{V}[X]}{n} = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

Convergence : $R(\hat{\theta}_1) \rightarrow 0$, donc $\hat{\theta}_1 \rightarrow \theta$ dans L^2 et en probabilité; la loi forte des grands nombres donne même la convergence presque sûre. Le théorème central limite fournit la vitesse :

$$\sqrt{n}(\hat{\theta}_1 - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\theta^2}{3}\right).$$

Défaut : $\hat{\theta}_1$ peut être inférieur à M_n , donc prendre une valeur incompatible avec les observations.

2°. Maximum de vraisemblance. La vraisemblance de l'échantillon $(x_1, \dots, x_n)$, tous positifs, s'écrit

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbf{1}_{[0, \theta]}(x_i) = \frac{1}{\theta^n} \mathbf{1}_{\{\theta \geq \max_i x_i\}}.$$

On ne peut pas dériver : L est nulle pour $\theta < \max_i x_i$ puis strictement décroissante en θ . Elle est donc maximale au plus petit θ admissible :

$$\hat{\theta}_2 = M_n = \max(X_1, \dots, X_n).$$

Loi de M_n : pour $x \in [0, \theta]$,

$$F_n(x) = \mathbb{P}[\forall i, X_i \leq x] = \left(\frac{x}{\theta}\right)^n, \quad f_n(x) = \frac{nx^{n-1}}{\theta^n}.$$

D'où, pour $k \geq 1$,

$$\mathbb{E}[M_n^k] = \int_0^\theta x^k \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{n+k} \theta^k.$$

Biais : $\mathbb{E}[\hat{\theta}_2] = \frac{n}{n+1} \theta$, donc

$$b(\hat{\theta}_2) = -\frac{\theta}{n+1} < 0.$$

L'estimateur sous-estime systématiquement θ (c'est inévitable puisque $M_n \leq \theta$ p.s.), mais il est asymptotiquement sans biais.

Variance et erreur quadratique :

$$\mathbb{V}[M_n] = \frac{n}{n+2} \theta^2 - \frac{n^2}{(n+1)^2} \theta^2 = \frac{n \theta^2}{(n+1)^2 (n+2)},$$

$$R(\hat{\theta}_2) = \frac{n \theta^2}{(n+1)^2 (n+2)} + \frac{\theta^2}{(n+1)^2} = \frac{2 \theta^2}{(n+1)(n+2)}.$$

Convergence : $R(\hat{\theta}_2) = O(1/n^2) \rightarrow 0$, donc convergence L^2 et en probabilité. Comme M_n est croissante et que

$$\mathbb{P}[|M_n - \theta| > \varepsilon] = \left(1 - \frac{\varepsilon}{\theta}\right)^n$$

est le terme général d'une série convergente, le lemme de Borel-Cantelli donne la convergence presque sûre. La vitesse est ici en n (et non $\sqrt{n}$), avec une loi limite non gaussienne :

$$n(\theta - M_n) \xrightarrow{\mathcal{L}} \mathcal{E}(1/\theta),$$

car $\mathbb{P}[n(\theta - M_n) > t] = (1 - t/(n\theta))^n \rightarrow e^{-t/\theta}$.
Correction du biais : il suffit de multiplier par $\frac{n+1}{n}$,

$$\hat{\theta}_3 = \frac{n+1}{n} M_n, \quad \mathbb{E}[\hat{\theta}_3] = \theta,$$

$$R(\hat{\theta}_3) = \mathbb{V}[\hat{\theta}_3] = \left(\frac{n+1}{n}\right)^2 \frac{n \theta^2}{(n+1)^2 (n+2)},$$

soit $R(\hat{\theta}_3) = \frac{\theta^2}{n(n+2)}$. On a bien $R(\hat{\theta}_3) < R(\hat{\theta}_2)$ dès que $n \geq 2$.

3°. Estimateur plug-in. Principe de substitution : on écrit θ comme une fonctionnelle de la loi, puis on remplace la loi par la loi empirique. Ici la médiane théorique vaut $\theta/2$, soit $\theta = 2F_\theta^{-1}(1/2)$, d'où

$$\hat{\theta}_4 = 2F_n^{-1}(1/2) = 2X_{(m+1)},$$

où $X_{(1)} \leq \dots \leq X_{(n)}$ est l'échantillon ordonné : avec $n = 2m+1$ impair, la médiane empirique est exactement $X_{(m+1)}$.

Loi de $X_{(m+1)}$: en posant $U_i = X_i/\theta \sim \mathcal{U}[0, 1]$, on a $X_{(m+1)} = \theta U_{(m+1)}$ avec

$$U_{(m+1)} \sim \text{Beta}(m+1, n-m) = \text{Beta}(m+1, m+1),$$

de densité $\frac{(2m+1)!}{(m!)^2} u^m (1-u)^m$ sur $[0, 1]$.

Biais : cette densité est symétrique par rapport à $1/2$, donc $\mathbb{E}[U_{(m+1)}] = 1/2$ et

$$\mathbb{E}[\hat{\theta}_4] = 2\theta \cdot \frac{1}{2} = \theta :$$

$\hat{\theta}_4$ est sans biais.

Erreur quadratique : pour $\text{Beta}(a, b)$ la variance vaut $\frac{ab}{(a+b)^2(a+b+1)}$, ici avec $a = b = m + 1$ et $a + b = n + 1$:

$$\mathbb{V}[\mathbf{U}_{(m+1)}] = \frac{(m+1)^2}{(n+1)^2(n+2)} = \frac{1}{4(n+2)},$$

$$\mathbf{R}(\hat{\theta}_4) = \mathbb{V}[2\theta \mathbf{U}_{(m+1)}] = \frac{4\theta^2}{4(n+2)} = \frac{\theta^2}{n+2}.$$

Convergence : $\mathbf{R}(\hat{\theta}_4) \rightarrow 0$, donc convergence L^2 et en probabilité (et p.s. par le résultat classique sur la médiane empirique). La vitesse est en $\sqrt{n}$:

$$\sqrt{n}(\hat{\theta}_4 - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta^2),$$

conformément à la variance asymptotique $1/(4nf_0(\theta/2)^2)$ de la médiane empirique.

4°. Comparaison. Les quatre risques quadratiques sont

$$\mathbf{R}(\hat{\theta}_1) = \frac{\theta^2}{3n}, \quad \mathbf{R}(\hat{\theta}_2) = \frac{2\theta^2}{(n+1)(n+2)},$$

$$\mathbf{R}(\hat{\theta}_3) = \frac{\theta^2}{n(n+2)}, \quad \mathbf{R}(\hat{\theta}_4) = \frac{\theta^2}{n+2}.$$

Pour $n \geq 2$:

$$\mathbf{R}(\hat{\theta}_3) < \mathbf{R}(\hat{\theta}_2) \ll \mathbf{R}(\hat{\theta}_1) < \mathbf{R}(\hat{\theta}_4).$$

Deux familles se dégagent : les estimateurs fondés sur une moyenne ou une médiane ont un risque en θ^2/n , ceux fondés sur le maximum un risque en θ^2/n^2 . Le meilleur est donc

$$\hat{\theta}_3 = \frac{n+1}{n} M_n.$$

Remarques :

- entre $\hat{\theta}_1$ et $\hat{\theta}_4$, la moyenne bat la médiane d'un facteur 3 asymptotiquement : pour la loi uniforme, l'information utile est dans les extrêmes, pas au centre;
- la vitesse en $1/n^2$ ne contredit pas la borne de Cramér-Rao : le support $[0, \theta]$ dépend de θ , le modèle n'est pas régulier et la borne ne s'applique pas;
- M_n est une statistique exhaustive et complète; $\hat{\theta}_3$ étant sans biais et fonction de M_n , le théorème de Lehmann-Scheffé assure qu'il est optimal parmi les estimateurs sans biais. Si l'on accepte un léger biais, on fait encore mieux : en minimisant $\mathbb{E}[(cM_n - \theta)^2]$ en c on trouve $c = \frac{n+2}{n+1}$ et

$$\mathbf{R}\left(\frac{n+2}{n+1} M_n\right) = \frac{\theta^2}{(n+1)^2} < \frac{\theta^2}{n(n+2)}.$$

5°. Approche bayésienne. La loi *a priori*

$\pi(\theta) = \theta^{-2} \mathbf{1}_{[1, +\infty[}(\theta)$ est bien une densité (loi de Pareto de paramètres 1 et 1), car $\int_1^{+\infty} \theta^{-2} d\theta = 1$. Par la formule de Bayes, et en notant $a = \max(1, M_n)$:

$$\pi(\theta | x) \propto L(\theta) \pi(\theta) = \frac{1}{\theta^n} \mathbf{1}_{\{\theta \geq M_n\}} \cdot \frac{1}{\theta^2} \mathbf{1}_{\{\theta > 1\}},$$

soit

$$\pi(\theta | x) \propto \theta^{-(n+2)} \mathbf{1}_{\{\theta > a\}}.$$

La constante de normalisation s'obtient par

$$\int_a^{+\infty} \theta^{-(n+2)} d\theta = \frac{a^{-(n+1)}}{n+1},$$

d'où la loi *a posteriori*

$$\pi(\theta | x) = (n+1) \frac{a^{n+1}}{\theta^{n+2}} \mathbf{1}_{\{\theta > a\}},$$

c'est-à-dire une loi de Pareto de paramètre d'échelle $a = \max(1, M_n)$ et d'indice $n+1$ (la famille de Pareto est donc conjuguée pour ce modèle).

Sous coût quadratique, l'estimateur de Bayes est la moyenne *a posteriori* :

$$\hat{\theta}_5 = \mathbb{E}[\theta | x] = (n+1) a^{n+1} \int_a^{+\infty} \theta^{-(n+1)} d\theta,$$

soit

$$\hat{\theta}_5 = \frac{n+1}{n} \max(1, M_n).$$

La loi *a priori* π a pour densité $\pi(\theta) = \frac{1}{\theta^2} \mathbf{1}_{[1, +\infty[}(\theta)$.

C'est la loi d'une variable aléatoire T dont les réalisations sont les valeurs de θ . La vraisemblance de l'échantillon est

$$L(x|\theta) = \frac{1}{\theta^n} \mathbf{1}_{[0, \theta]^n}(x) \quad (1)$$

où $x = (x_1, \dots, x_n)$ est une observation de l'échantillon. La densité de la loi jointe (X, T) peut alors s'écrire

$$f_{(X, T)}(x, \theta) = L(x|\theta) \pi(\theta) \quad (2)$$

$$= \frac{1}{\theta^{n+2}} \mathbf{1}_{[0, \theta]^n}(x) \mathbf{1}_{[1, +\infty[}(\theta) \quad (3)$$

D'après le théorème de Bayes, la loi marginale de X a pour densité

$$f_X(x) = \int_1^{+\infty} \frac{1}{\theta^{n+2}} \mathbf{1}_{[0, \theta]^n}(x) d\theta \quad (4)$$

l'indicatrice sous l'intégrale vaut 1 si tous les x_i sont compris entre 0 et θ et sinon elle vaut 0. Autrement dit, elle vaut 1 si, et seulement si,

$$0 \leq x_{(1)} \leq x_{(n)} \leq \theta \quad (5)$$

On va donc intégrer soit entre 1 et l'infini si 1 est plus grand que θ , soit entre θ et l'infini. Notons $\sigma = \max(1, \theta)$. On a alors

$$f_X(x) = \int_{\sigma}^{+\infty} \frac{d\theta}{\theta^{n+2}} = \frac{1}{n+1} \frac{1}{\sigma^{n+1}} \quad (6)$$

On en déduit, d'après la formule de Bayes, la vraisemblance *a posteriori* de θ sachant l'observation x (en notant T la variable aléatoire dont une réalisation est θ) :

$$L(T = \theta | X = x) = \frac{n+1}{\theta^{n+2}} \sigma^{n+1} \mathbf{1}_{[1, +\infty[}(\theta) \mathbf{1}_{[0, \theta]^n}(x) \quad (7)$$

L'estimateur bayésien de la moyenne est alors l'espérance de cette loi *a posteriori* et vaut

$$\mathbb{E}[T|X = x] = (n+1)\sigma^{n+1} \int_{\sigma}^{+\infty} \frac{d\theta}{\theta^{n+1}} = \frac{n+1}{n} \sigma \mathbb{1}_{[x_{(1)} \geq 0]} \quad (8)$$

On en déduit que

$$\mathbb{E}[T|X] = \frac{n+1}{n} \max(1, X_{(n)}) \quad (9)$$

Interprétation : dès que $M_n \geq 1$, ce qui se produit avec probabilité tendant vers 1 (et toujours si $\theta > 1$ et n grand), on retrouve exactement $\hat{\theta}_3 = ((n+1)/n)M_n$.
L'a priori n'intervient que pour imposer la contrainte $\theta > 1$ lorsque les données sont trop peu informatives.
 Sous coût absolu on prendrait la médiane *a posteriori*, $a 2^{1/(n+1)}$, et le maximum *a posteriori* vaudrait simplement a .