

IFT-3245

Démonstration 7, 19 novembre 2012.

1. Exemple d'échec du bootstrap.

On considère un échantillon X_1, \dots, X_n issu d'une loi uniforme $[0, \theta]$, où θ est inconnu. On voudrait estimer θ .

Quel est l'estimateur de maximum de vraisemblance ?

Sous θ , la densité de X_i est $1/\theta$, pour tout i . Dès lors, la vraisemblance s'écrit

$$\prod_{i=1}^n \frac{1}{\theta},$$

avec la condition $\theta \geq \max\{X_1, \dots, X_n\}$. Ceci conduit à l'estimateur de maximum de vraisemblance

$$\max_{\theta \geq \max\{X_1, \dots, X_n\}} \frac{1}{\theta^n}.$$

Cette fonction est décroissante en θ , aussi le maximum est atteint avec le plus petit θ réalisable. En d'autres termes, la contrainte devient active :

$$\hat{\theta}_n = \max\{X_1, \dots, X_n\}.$$

En triant les observations, on peut aussi écrire

$$\hat{\theta}_n = X_{(n)}.$$

Déterminons la distribution de cet estimateur. Fonction de répartition :

$$\begin{aligned} F(y) &= P[X_{(n)} < y] \\ &= P[X_1 < y \cap X_2 < y \cap \dots \cap X_n < y] \\ &= \prod_{i=1}^n P[X_i < y] \\ &= \left(\frac{y}{\theta}\right)^n \end{aligned}$$

si $y \in [0, \theta]$. $F(y) = 0$ si $y \leq 0$ et 1 si $y \geq \theta$. On obtient la densité en dérivant $F(y)$ par rapport à y (sur $[0, \theta]$) :

$$f(y) = \frac{ny^{n-1}}{\theta^n}$$

L'estimateur est biaisé. En effet :

$$\begin{aligned} E_\theta[\max\{X_1, \dots, X_n\}] &= \int_0^\theta \frac{ny^{n-1}}{\theta^n} dy \\ &= \left[\frac{n}{\theta^n} \frac{y^{n+1}}{n+1} \right]_0^\theta \\ &= \frac{n}{n+1} \theta. \end{aligned}$$

La probabilité pour la procédure bootstrap de renvoyer le maximum des observations est

$$p = 1 - \left(1 - \frac{1}{n}\right)^n$$

où

$$\left(1 - \frac{1}{n}\right)^n$$

correspond à la probabilité de ne jamais tirer $X_{(n)}$.

Soit $T_n = n(\theta - X_{(n)})$, $T_n^* = n(X_{(n)} - X_{(n)}^*)$. Supposons $\theta = 1$. Alors, pour $t \geq 0$,

$$\begin{aligned} P_{\hat{F}_n}[T_n^* \leq t] &\geq P_{\hat{F}_n}[T_n^* = 0] \\ &= P_{\hat{F}_n}[X_{(n)}^* = X_{(n)}] \\ &= 1 - P_{\hat{F}_n}[X_{(n)}^* < X_{(n)}] \\ &= 1 - \left(\frac{n-1}{n}\right)^n \\ &\rightarrow 1 - e^{-1}, \text{ comme } n \rightarrow \infty. \end{aligned}$$

Prenons par exemple $t = 0.0001$. On a

$$\lim_n P_F[T_n \leq t] = 1 - e^{-0.0001} \approx 0.$$

Aussi $P_{F_n}[T_n^* \leq t]$ ne converge pas vers $P_F[T_n \leq t]$.