Bootstrap

Clément Dell'Aiera, Guillaume Rateau

Résumé

Table des matières

1 Bootstrap et valeurs extrêmes

1.1 Théorèmes des valeurs extrêmes

Théorème 1. Si , pour une fonction de répartition F, il existe des suites réelles $a_n > 0$ et b_n telles que, en tout point de continuité de F :

$$\lim_{n\infty} F^n(a_n x + b_n) = G_\alpha(x),$$

la loi limite G est de la forme :

$$G_\alpha(x) = \exp(-(1+\alpha x)^{-\frac{1}{\alpha}}) \quad \text{pour } 1 + \alpha x > 0$$

Le paramètre α donne de l'information sur la queue de la distribution : c'est un paramètre d'intérêt servant à prédire les évenements rares. Toutefois, les méthodes de Bootstrap naives ne fonctionnent plus dans le cas des valeurs extrêmes, à cause d'un potentiel biais.

Remarque : $G_0(x) = \exp\{e^{-x}\}$.

1. Si $\alpha > 0$, on dit que F est dans le domaine d'attraction de la loi de Fréchet, domaine des lois à queues épaisses (décroissance en loi de puissance).
2. Si $\alpha = 0$, on dit que F est dans le domaine d'attraction de la loi de Gumbel (queues légères).
3. Si $\alpha > 0$, on dit que F est dans le domaine d'attraction de la loi de Weibull (queues finies).

Domaine	**Gumbel**	**Fréchet**	**Weibull**
Lois	Normale Exponentielle Fréchet Lognormale Gamma Weibull	Cauchy Pareto Student	Uniforme Beta

FIGURE 1 – Quelques exemples de lois appartenant à différents domaines.

2 Estimation de la vitesse de convergence par régression

Dans toute cette partie, l'échantillon considéré est stationnaire et est foretement mélangeant :

$$\alpha(k) = \sup |P(A \cup B) - P(A)P(B)| \to 0 \quad \text{quand} \quad k \to \infty.$$

On se donne un échantillon $X_1, .., X_n, ..$ et une statistique d'intérêt T_n. Soit Y_i le sous-échantillon $(X_i, X_{i+1} ..., X_{i+b_n-1})$ pour $i \in [|1; q|]$ et $q = n - b_n + 1$. Soit $T_{b_n,i}$ la statistique obtenue à partir de l'échantillon Y_i avec le taux d'échantillonage b_n. On définit :

$$K_{b_n}(x|X^n, \tau) = \frac{1}{q} \sum_{i=1}^{q} 1_{\{\tau_n T_{b_n,i} \leq x\}}.$$

Théorème 2. On suppose que : $||K_{b_n}(.|X^n, \tau) - K(.|P)||_\infty = o_P(1)$.
Si K est continue et atteint ses bornes sur un compact, sur lequel elle est strictement croissante, alors, lorsque n tend vers l'infini :

$$\tau_{b_n} K_{b_n}^{-1}(t|X^n) = K^{-1}(x, P) + o_P(1)$$

On se place dans les hypothèse

$$\lim_{n \to +\infty} b_n = +\infty \quad \text{et} \quad \lim_{n \to +\infty} \frac{b_n}{n} = 0.$$

Supposons que $\tau_n = n^{-\gamma}$. En passant au log dans la conclusion du théorème précédent, on obtient :

$$\log |K_{b_n}^{-1}(x|X^n, \tau)| = \log |K^{-1}(x, P)| + \gamma \log b_n + o_P(1).$$

En faisant alors varier les taux de sous-échantillonage $b_{n,i}$, on peut alors estimer γ par moindres carrés ordinaires :

$$\hat{\gamma} = \frac{\sum_{i=1}^{I} (y_i - \overline{y})(\log b_{i,n} - \overline{\log})}{\sum_{i=1}^{I} (\log b_{i,n} - \overline{\log})^2},$$

où $y_i = \log |K_{b_{n,i}}^{-1}(t|X^n)|$.

Une deuxième méthode est inspirée par la remarque suivante : il suffit d'observer des différences de quantiles. En effet, si l'on prend l'équation du théorème 2 pour deux quantiles d'ordre $0 < t_1 < t_2$, et que l'on soutrait l'une à l'autre, après avoir passé au log, on obtient :

$$\log |K_{b_n}^{-1}(t_1|X^n) - K_{b_n}^{-1}(t_2|X^n)| = \log |K^{-1}(t_1|P) - K^{-1}(t_2|P)| + \gamma \log b_n + o_P(1).$$

Avec plusieurs quantiles d'ordre $0 < t_1 < .... < t_J < 1$, on obtient un système :

$$y_{ij} = \log |K_{b_{n,i}}^{-1}(t_j|X^n)| = a_j + \gamma \log b_{n,i} + u_{ij},$$

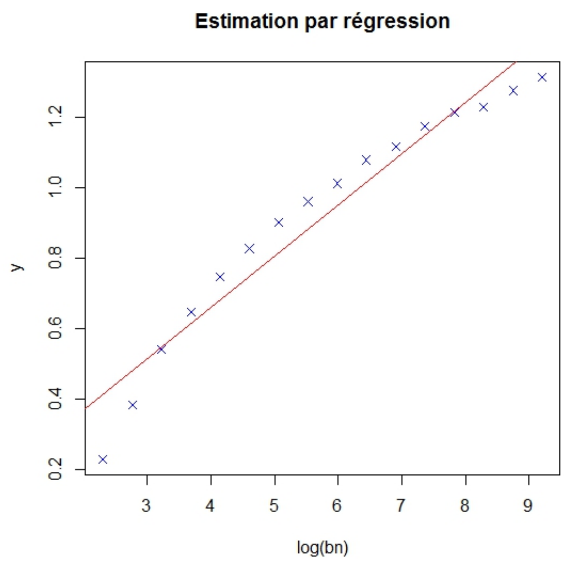

FIGURE 2 – Régression des log-quantiles obtenus par bootstrap sur $\log b_n$ dans le cas de la loi normale. Ici $I = 16$ et $N = 10^5$, on obtient $\gamma_I = 0.1494989$.

où $a_j = \log |K^{-1}(t_j|P)|$ et $u_{ij} = o_P(1)$.

Un estimateur de type ANOVA est

$$\gamma_{IJ} = \frac{\sum_{i=1}^{I}(y_{i,.} - \overline{y})(\log b_{i,n} - \overline{\log})}{\sum_{i=1}^{I}(\log b_{i,n} - \overline{\log})^2},$$

Théorème 3. Soit X une suite stationnaire et fortement mélangeante, dont la statistique d'intérêt $\tau_n T_n$ possède une distribution asymptotique, lorsque $\tau_n = n^{-\gamma}$, $\gamma > 0$ inconnu. On suppose de plus que K est strictement croissante sur un intervalle borné et continue. On choisit des point t_j dans $(0, 1)$ et différents taux de sous-échnatilonage n^{β_i}, $1 > \beta_1 > ... > \beta_I > 0$. Alors :

$$\gamma_{IJ} = \gamma + o_P((\log n)^{-1})$$

Voici des estimateurs de γ obtenus par la méthode donnée par le théorème 3. Ici $I = 16$ et $J = 10$, les échnantillons sont de taille $N = 100000$, et l'on a pris $\beta_i = 10^i$ pour i allant de 1 à 4.

Lois	Normale	Pareto	Uniforme
γ_{IJ}	0.003967868	0.003950971	0.004037952

FIGURE 3 – Simulations.

Références

[1] Dimitris N. Politis Patrice Bertail, Christian Haefke. A subsampling approach to estimating the distribution of diverging statistics with applications to assessing financial market risks. 2001.

[2] Joseph P. Romano Patrice Bertail, Dimitris N. Politis. On subsampling estimatorswith unknown rate of convergence. 1995.