Bootstrap et rééchantillonnage

Atelier SFdS - Partie 2

Magalie Fromont et Myriam Vimond

CREST (Ensai) - IRMAR (Université Européenne de Bretagne)

Novembre 2012

Plan

Consistance du bootstrap

Définitions

$R_n(\mathbb{X}, P)$ une racine dont la loi est notée μ_n
$R_n^* = R_n(\mathbb{X}^*, P_n)$ sa réplication bootstrap, dont la loi conditionnelle sachant $\mathbb{X}$ est notée μ_n^*.

Étant donnée une distance ρ sur un espace de lois de probabilités contenant μ_n et μ_n^*, le bootstrap de $R_n(\mathbb{X}, P)$ est dit :

– **faiblement consistant** si $\rho(\mu_n, \mu_n^*) \xrightarrow{(\mathbb{P})} 0$.

– **fortement consistant** si $\rho(\mu_n, \mu_n^*) \xrightarrow{p.s.} 0$.

Distance de Kolmogorov sur $\mathcal{P}(\mathbb{R})$ l'ensemble des mesures de probabilité sur $\mathbb{R}$:

$k(\mu, \mu') = \sup_{x \in \mathbb{R}} |F_\mu(x) - F_{\mu'}(x)|$, où $F_\mu(x)$, $F_{\mu'}$ sont les f.d.r. associées à μ et μ'.

Distance de Mallows-Wasserstein :

$(B, \|.\|)$ espace de Banach séparable,
$\mathcal{P}(B)$ ensemble des mesures de probabilité sur B,
$\Gamma_2(B) = \left\{\mu \in \mathcal{P}(B) / \int \|x\|^2 d\mu(x) < \infty\right\}$,

$d_2(\mu, \mu') = \left(\inf_{(R,R')/R\sim\mu, R'\sim\mu'} \mathbb{E}\left[\|R - R'\|^2\right]\right)^{1/2}$ définit une distance sur $\Gamma_2(B)$.

Propriété de la distance de Mallows-Wasserstein

Soit $\mu_n,\ \mu \in \Gamma_2(B)$,
$d_2(\mu_n, \mu) \to_{n\to\infty} 0 \Leftrightarrow \mu_n \rightsquigarrow \mu$ et $\int \|x\|^2 d\mu_n(x) \to_{n\to\infty} \int \|x\|^2 d\mu(x)$.

✔ Distance populaire en statistique, adaptée à la plupart des problèmes qui se posent avec le bootstrap (estimation de loi, d'espérance, de variance).

Consistance du bootstrap

Un exemple : le bootstrap de la moyenne

X_i à valeurs réelles, $\theta_P = \int x dP(x)$,
$R_n(\mathbb{X}, P) = \sqrt{n}\left(\overline{\mathbb{X}} - \theta(P)\right)$ de loi μ_n,
$R_{n*} = \sqrt{n}\left(\overline{\mathbb{X}^*} - \overline{\mathbb{X}}\right)$ de loi conditionnelle sachant $\mathbb{X}$ μ_n^*.

Consistance du bootstrap de la moyenne

Si les X_i sont i.i.d. de variance $\sigma^2 < \infty$,

- $\mu_n \rightsquigarrow \mathcal{N}(0, \sigma^2)$ (TCL).
- $\mu_n^* \rightsquigarrow \mathcal{N}(0, \sigma^2)$ p.s.
- $k(\mu_n, \mu_n^*) \overset{p.s.}{\to}_{n\to\infty} 0$ et $d_2(\mu_n, \mu_n^*) \overset{p.s.}{\to}_{n\to\infty} 0$.

$\hookrightarrow$ Singh (1981) pour la distance k (+ vitesses sous conditions de moments supplémentaires),
$\hookrightarrow$ Bickel et Freedman (1981) pour la distance d_2.

Question : Que se passe-t-il si $\mathbb{E}[X_i^2] = \infty$?

Réciproque : Giné, Zinn (1989)

S'il existe une suite de variables aléatoires $T_n(\mathbb{X})$, une suite $a_n \nearrow \infty$, et une mesure de probabilité aléatoire μ^* non dégénérée de probabilité > 0, telles que :

$$\mu_n^* = \mathcal{L}\left(\frac{\sum_{i=1}^n X_i^*}{an} - T_n(\mathbb{X}) \middle| \mathbb{X} \right) \rightsquigarrow \mu^* \text{ p.s.,}$$

alors $a_n \simeq \sqrt{n}$, $\mathbb{E}[X_i^2] < \infty$, et $\mu_n^* \rightsquigarrow \mathcal{N}(0, \mathbb{V}(X_i))$.

✔ Résultats également pour la consistance faible.

Application 1 : bootstrap de la moyenne "studentisée"

$\tilde{R}_n(\mathbb{X}, P) = \sqrt{n}\left(\overline{\mathbb{X}} - \theta(P)\right)/S(\mathbb{X})$, où $S^2(\mathbb{X})$ estimateur plug-in ou empirique de la variance des X_i, de loi $\tilde{\mu}_n$,
$\tilde{R}^*_n = \sqrt{n}\left(\overline{\mathbb{X}^*} - \overline{\mathbb{X}}\right)/S(\mathbb{X}^*)$ de loi conditionnelle sachant $\mathbb{X}$ $\tilde{\mu}^*_n$.

Consistance du bootstrap de la moyenne "studentisée"

Si les X_i sont i.i.d. de variance $\sigma^2 < \infty$,

- $\forall \varepsilon > 0$, $P\left(|S(\mathbb{X}^*) - \sigma| > \varepsilon \middle| \mathbb{X}\right) \xrightarrow{p.s.} 0$.
- $k(\tilde{\mu}_n, \tilde{\mu}^*_n) \xrightarrow{p.s.} 0$ et $d_2(\tilde{\mu}_n, \tilde{\mu}^*_n) \xrightarrow{p.s.} 0$.

✔ Généralisation au cas multivarié, et au m out of n bootstrap.

Application 2 : méthode Delta

$\tilde{R}_n(\mathbb{X}, P) = \sqrt{n}\left(g(\overline{\mathbb{X}}) - g(\theta(P))\right)$ de loi $\tilde{\mu}_n$,
$\tilde{R}_{n^*} = \sqrt{n}\left(g(\overline{\mathbb{X}^*}) - g(\overline{\mathbb{X}})\right)$ de loi conditionnelle sachant $\mathbb{X}$ $\tilde{\mu}_n^*$.

Méthode Delta pour le bootstrap de la moyenne

Si les X_i sont i.i.d. de variance $\sigma^2 < \infty$, g fonction continûment différentiable en $\theta(P)$, avec $g'(\theta(P)) \neq 0$, alors

$$k(\tilde{\mu}_n, \tilde{\mu}_n^*) \xrightarrow{p.s.} 0.$$

✔ Généralisation au cas multivarié, et au m out of n bootstrap.

Question philosophique...

✘ Puisque le bootstrap de la moyenne est fortement consistant ssi le TCL est vérifié, quid de son intérêt ?

✔ Dans le cas de la moyenne comme dans celui de certaines autres statistiques (mais pas toutes, attention !), sous certaines conditions de moments, μ_n^* approche plus précisément μ_n que la loi asymptotique.

Illustration

$\mathbb{X}$ 30-échantillon de la loi $\mathcal{E}(1)$.
Bootstrap de la moyenne "studentisée".

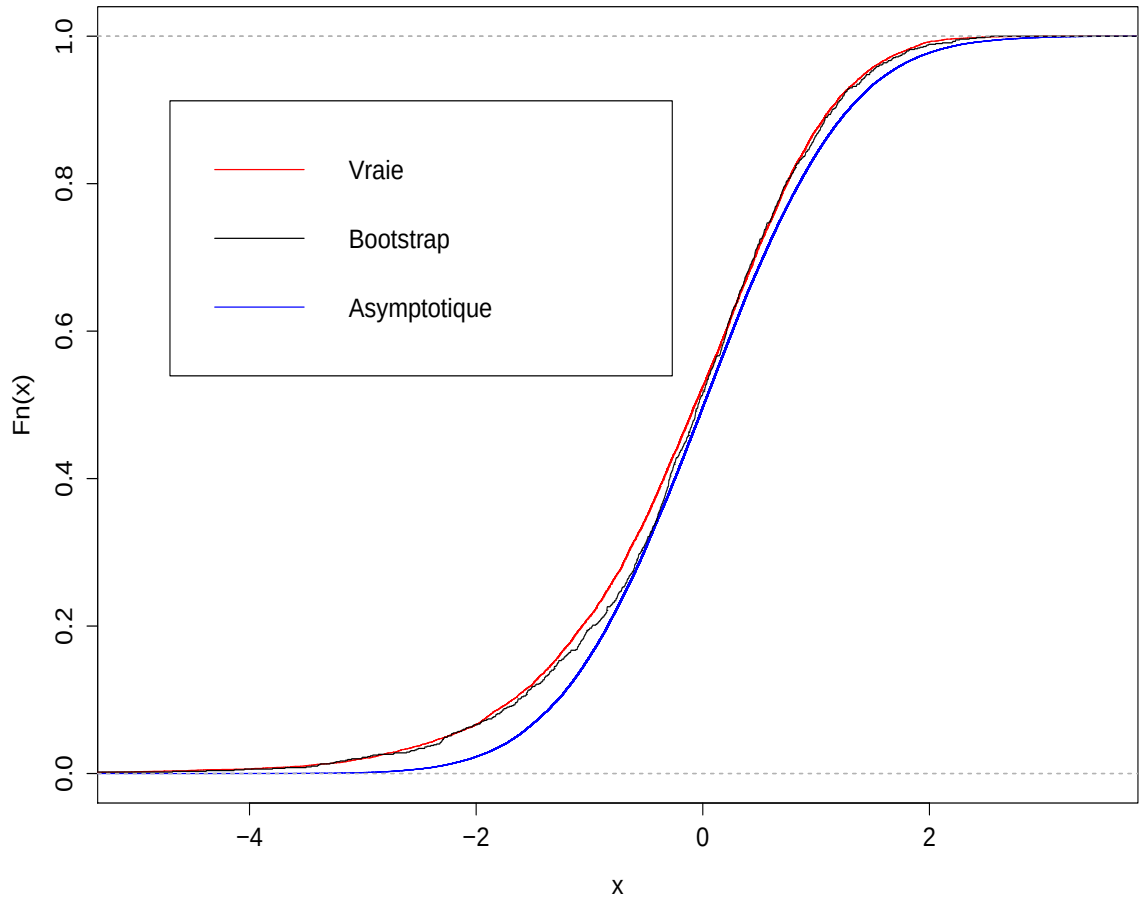


FIGURE: F.d.r. réelle, empirique bootstrap, asymptotique

Validité au second ordre

Précision à l'ordre k d'un IC

Soit $\mathcal{I}_n$ un intervalle de confiance pour θ de niveau $1-\alpha$.

$\mathcal{I}_n$ est **consistant** si $\mathbb{P}(\theta \in \mathcal{I}_n) \to 1-\alpha$.

$\mathcal{I}_n$ est **précis à l'ordre k** si $\mathbb{P}(\theta \in \mathcal{I}_n) = 1-\alpha+\mathcal{O}\left(1/n^{k/2}\right)$.

Objectif : étude de la précision de l'IC table bootstrap à l'aide des développements d'Edgeworth

Une illustration

Soit $\hat{\theta}$ un estimateur $\sqrt{n}$-consistant,

$$Z_n = \sqrt{n}\left(\hat{\theta} - \theta\right)/\hat{\sigma}_n \overset{\mathbb{P}}{\rightsquigarrow} \mathcal{N}(0,1),$$

où $\hat{\sigma}_n^2$ est un estimateur consistant de la variance σ^2.

Z_n admet **un développement d'Edgeworth** à l'ordre 2 si

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}(Z_n \leq x) - \phi(x) - \frac{1}{\sqrt{n}} p_1(x)\phi'(x) - \frac{1}{n} p_2(x)\phi'(x) \right| = O\left(1/n^{3/2}\right)$$

où p_1 (*fonction paire*), p_2 (*fonction impaire*) polynômes de degré 3 et 6

$$Z_n = \sqrt{n}\left(\hat{\theta} - \theta\right)/\hat{\sigma}_n \overset{\mathbb{P}}{\rightsquigarrow} \mathcal{N}(0,1)$$

Soit $0 < \alpha < 1$ et $\mathbf{z}_\alpha$ le quantile d'ordre α de $\mathcal{N}(0,1)$,

- $\mathcal{I}_n^\infty = \left]-\infty, \hat{\theta} - z_\alpha \hat{\sigma}/\sqrt{n}\right]$ est un $\mathrm{IC}_{1-\alpha}(\theta)$ précis à l'ordre 1

$$\mathbb{P}\left(\theta \in \mathcal{I}_n^\infty\right) = 1 - \mathbb{P}(Z_n \leq z_\alpha) = 1 - \alpha + \mathcal{O}\left(1/\sqrt{n}\right).$$

- $\mathcal{J}_n^\infty = \left[\hat{\theta} \pm z_\alpha \hat{\sigma}/\sqrt{n}\right]$ est un $\mathrm{IC}_{1-2\alpha}(\theta)$ précis à l'ordre 2

$$\mathbb{P}\left(\theta \in \mathcal{J}_n^\infty\right) = \mathbb{P}(z_\alpha \leq Z_n \leq z_{1-\alpha}) = 1 - 2\alpha + \mathcal{O}\left(1/n\right).$$

ici $\mathcal{N}(0,1)$ symétrique, sinon la précision est à l'ordre 1

$$Z_n^* = \sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right)/\hat{\sigma}_n^* \overset{\mathbb{P}^*}{\rightsquigarrow} \mathcal{N}(0,1)$$

Soit $0 < \alpha < 1$ et $\mathbf{z}^*_{\alpha,\mathbf{n}}$ le quantile d'ordre α de $Z_n^*|\mathbb{X}^*$.

Quelle est la précision des intervalles

- $\mathcal{I}_n^* = \left]-\infty, \hat{\theta} - z^*_{\alpha,n}\hat{\sigma}/\sqrt{n}\right]$?
- $\mathcal{J}_n^* = \left[\hat{\theta} \pm z^*_{\alpha,n}\hat{\sigma}/\sqrt{n}\right]$?

Validité au second ordre

Développement d'Edgeworth - Bootstrap de la moyenne

Étude d'un cas simple

Soit $X_1, \ldots, X_n$ un n-échantillon de P de moyenne θ (inconnue) et de variance σ^2 (connue).

Un estimateur sans biais de θ : $\hat{\theta} = \overline{X}$.

$$Z_n = \sqrt{n}\left(\hat{\theta} - \theta\right)/\sigma \text{ converge en loi vers } \mathcal{N}(0,1).$$

Validité au second ordre

Développement d'Edgeworth - Construction

Développement de la fonction caractéristique

Soit $\chi(t)$ la fonction caractéristique de $Y_1 = (X_1 - \theta)/\sigma$.
Soit $\chi_n(t)$ la fonction caractéristique de Z_n

$$\chi_n(t) = \left\{\chi(t/\sqrt{n})\right\}^n \longrightarrow e^{-t^2/2}.$$

Par un développement en série entière de $\log \chi(t)$ en fonction des cumulants κ_ℓ de Y :

$$\chi_n(t) = e^{-t^2/2} \times \exp\left(\sum_{\ell \geq 3} \frac{\kappa_\ell}{(\sqrt{n})^{\ell-2}} \frac{(it)^\ell}{\ell!}\right)$$
$$= e^{-t^2/2} + \frac{1}{\sqrt{n}} r_1(it) e^{-t^2/2} + \frac{1}{n} r_2(it) e^{-t^2/2} + \ldots,$$

où $\quad r_1(t) = \frac{1}{6}\kappa_3 t^3, \qquad r_2(t) = \frac{1}{24}\kappa_4 t^4 + \frac{1}{72}\kappa_3^2 t^6, \ldots$
$\kappa_3 = \mathbb{E}(Y - \mathbb{E}(Y))^3, \quad \kappa_4 = \mathbb{E}(Y - \mathbb{E}(Y))^4 - 3\mathbb{V}(Y)^2 \ldots$

Identification via les polynômes d'Hermitte

Soit $H_n(x)$ la f.d.r de Z_n.

Soit He_ℓ le polynôme d'Hermitte d'ordre ℓ.
Par linéarité de la transformée de Fourier,

$$\chi_n(t) = e^{-t^2/2} + \frac{1}{\sqrt{n}} r_1(it) e^{-t^2/2} + \frac{1}{n} r_2(it) e^{-t^2/2} + \ldots,$$

$$\Rightarrow \quad H_n(x) = \phi(x) + \frac{1}{\sqrt{n}} R_1(x) + \frac{1}{n} R_2(x) + \ldots$$

où $R_1(x) = p_1(x)\phi'(x), \quad R_2(x) = p_2(x)\phi'(x), \ldots$

- $p_1(x) = -\frac{1}{6}\kappa_3(x^2 - 1)$, correction de l'asymétrie
- $p_2(x) = -\frac{1}{24}\kappa_4 \mathrm{He}_3(x) + \frac{1}{72}\kappa_3^2 \mathrm{He}_5(x)$, correction de la curtose

Validité au second ordre

Précision au second ordre des quantiles

Développement de Cornish Fisher

Soit $\mathbf{z_{\alpha,n}}$ le quantile d'ordre α de la loi de Z_n.

Par identification à partir du développement d'Edgeworth, on obtient un développement similaire sur les quantiles :

$$z_{\alpha,n} = z_\alpha + \frac{1}{\sqrt{n}} q_1(z_\alpha) + \frac{1}{n} q_2(z_\alpha) + \ldots,$$

où les q_j sont des polynômes fonctions des cumulants κ_ℓ.

Conditions d'existence des développements d'Edgeworth et de Cornish Fisher

Hypothèses à vérifier sur P :

- P admet des moments finis jusqu'à un certain ordre,

  Pour un développement à l'ordre 1 : $\int |x|^3 dP < \infty$
  Pour un développement à l'ordre 2 : $\int |x|^4 dP < \infty$

- Le support de P n'est pas inclus dans un réseau discret.

  $\Rightarrow$ **n'est pas vérifié par** P_n.

$\Rightarrow$ pas de développement d'Edgeworth de Z_n par *plug-in*

Mais, en utilisant **la convergence uniforme** des fonctions caractéristiques, on en obtient une version *plug-in* !

Version bootstrap des développements

Soit $H_n^*(x) = \mathbb{P}(Z_n^* \leq x | \mathbb{X})$ la f.d.r. de $Z_n^* | \mathbb{X}$.
Soit $z_{\alpha,n}^*$ le quantile d'ordre α de H_n^*

Si P admet un moment d'ordre 8, alors H_n^* admet :

✔ un développement d'Edgeworth à l'ordre 2,

$$\sup_{x \in \mathbb{R}} \left| H_n^*(x) - \phi(x) - \frac{1}{\sqrt{n}} \hat{p}_1(x)\phi'(x) - \frac{1}{n} \hat{p}_2(x)\phi'(x) \right| = O_{\mathbb{P}}\left(1/n^{3/2}\right),$$

où $\hat{p}_1$ et $\hat{p}_2$ sont des estimateurs *plug-in* de p_1 et p_2.

✔ un développement de Cornish-Fisher,

$$z_{\alpha,n}^* = z_\alpha + \frac{1}{\sqrt{n}} \hat{q}_1(z_\alpha) + \frac{1}{n} \hat{q}_2(z_\alpha) + O_{\mathbb{P}}\left(1/n^{3/2}\right),$$

où $\hat{q}_1$ et $\hat{q}_2$ sont des estimateurs *plug-in* de q_1 et q_2.

D'après les développement de Cornish-Fisher,

$$z_{\alpha,n} - z_\alpha = O\left(1/\sqrt{n}\right)$$

$$z^*_{\alpha,n} - z_{\alpha,n} = \frac{1}{\sqrt{n}}\left(\hat{q}_1(z_\alpha) - q_1(z_\alpha)\right) + O_{\mathbb{P}}\left(1/n\right)$$

Par le TCL, $\hat{q}_1(z_\alpha) - q_1(z_\alpha) = O_{\mathbb{P}}\left(1/\sqrt{n}\right)$,

$$z^*_{\alpha,n} - z_{\alpha,n} = O_{\mathbb{P}}\left(1/n\right)$$

Le quantile bootstrap $z^*_{\alpha,n}$ est plus proche en probabilité du quantile réel $z_{\alpha,n}$.

Validité au second ordre

Probabilités de recouvrement

Pour un $\mathrm{IC}_{1-\alpha}(\theta)$ unilatéral,

- $\mathcal{I}_n = \left]-\infty, \hat{\theta} - z_{\alpha,n}\sigma/\sqrt{n}\right]$ IC exact

$$\mathbb{P}\left(\theta \in \mathcal{I}_n\right) = 1 - \alpha.$$

- $\mathcal{I}_n^{\infty} = \left]-\infty, \hat{\theta} - z_{\alpha}\sigma/\sqrt{n}\right]$ IC asymptotique

$$\mathbb{P}\left(\theta \in \mathcal{I}_n^{\infty}\right) = 1 - \alpha + \mathcal{O}\left(1/\sqrt{n}\right).$$

- $\mathcal{I}_n^{*} = \left]-\infty, \hat{\theta} - z_{\alpha,n}^{*}\sigma/\sqrt{n}\right]$ IC bootstrap

$$\mathbb{P}\left(\theta \in \mathcal{I}_n^{*}\right) = 1 - \alpha + \mathcal{O}\left(1/n\right).$$

Bootstrap du processus empirique

Pour aller plus loin

Bandes de confiance bootstrap pour la fonction de répartition associée à P, bootstrap des M-estimateurs (régression) ?

✔ Solutions "au cas par cas" (c.f. Bickel et Freedman (1981), Singh (1981)), mais pas pour toutes les racines (c.f. statistiques de test de Kolmogorov-Smirnov généralisées...)
✔ Le bootstrap du processus empirique permet de répondre à ces questions en même temps...
✘ ... mais il requiert un investissement plus important sur le plan théorique au départ.

Définition de la convergence faible

Il sera fait abstraction ici des problèmes de mesurabilité pour les processus empiriques : on admettra l'existence d'une définition de la convergence faible pour les processus empiriques (c.f. Van der Vaart, Wellner (1996)).

Bootstrap du processus empirique

Consistance

On considère la distance d_{BL1} de Lipschitz bornée :
$d_{BL1}(\mu, \mu') = \sup_{\{f \ / \ \|f\|_\infty \leq 1,\ |f(x)-f(y)| \leq |x-y| \ \forall x,y\}} \left| \int f d\mu - \int f d\mu' \right|$.

$\mathcal{F}$ classe de fonctions mesurables d'enveloppe finie,
$Z_n(f) = \sqrt{n}\,(P_n - P)\,(f)$ et $Z_n^*(f) = \sqrt{n}\,(P_n^* - P_n)\,(f)$.
Soit $\mathcal{F}_\delta = \left\{ f - g \ / \ f, g \in \mathcal{F}, \left(P((f-g) - P(f-g))^2 \right)^{1/2} < \delta \right\}$.

Consistance du bootstrap du processus empirique

Si $\mathcal{F}$ est P-Donsker et $\mathcal{F}_\delta$ mesurable $\forall \delta > 0$,

$d_{BL1}(Z_n^* | \mathbb{X}, Z) \overset{(\tilde{\mathbb{P}})}{\to} 0$, où Z est un pont Brownien.

Si $\tilde{\mathbb{P}}\left(\sup_{f \in \mathcal{F}} |f - P(f)|^2 \right) < \infty$, convergence $\tilde{\mathbb{P}}$-p.s.

✔ Généralisation au m out of n bootstrap, et au bootstrap à poids (c.f. Van der Vaart et Wellner (1996))

Plan

Queues de distributions épaisses

Bootstrap de la moyenne

Soit $\mathbb{X}$ un n-échantillon de P d'espérance $\mu = \int x dP$.

Si P est dans le domaine d'attraction de la loi normale, $\sigma^2 = \int (x-\mu)^2 dP < \infty$, alors :

- par le TCL, $Z_n = \sqrt{n}\left(\overline{\mathbb{X}} - \mu\right)/\sigma \rightsquigarrow \mathcal{N}(0,1)$
- pour H_n^* la f.d.r de $Z_n^*|\mathbb{X}$,

$$\sup_x \left| H_n^*(x) - \phi(x) \right| \overset{p.s.}{\to} 0,$$

et H_n^* est une meilleure approximation de la f.d.r de Z_n que la f.d.r de $\mathcal{N}(0,1)$, ϕ.

Que se passe-t-il lorsque P n'est plus dans le domaine d'attraction de la loi normale, i.e. $\int x^2 dP = \infty$?

Queues de distributions épaisses

Lois stables d'ordre α

Soit $\mathbb{X}$ un n-échantillon de P telle que :

$$1 - F(x) \sim_{+\infty} x^{-\alpha} L(x), \qquad F(-x) \sim_{+\infty} c x^{-\alpha} L(x)$$

pour un $\alpha \in]1, 2[$ et $L : \mathbb{R}_+^* \to \mathbb{R}_+^*$ à *variation lente* au voisinage de l'infini.

Comme $1 < \alpha < 2$, $\int |x| dP < \infty$ et $\int x^2 dP = \infty$

P est **dans le domaine d'attraction de** Q_α s'il existe des constantes $a_n > 0$ et b_n telles que

$$\left(\sum_{i=1}^{n} X_i - b_n \right) / a_n \rightsquigarrow Q_\alpha.$$

Les lois stables généralisent le rôle de la loi normale dans le TCL.

Convergence de la moyenne empirique

Si P est dans le domaine d'attraction d'une loi stable centrée Q d'ordre $\alpha \in]1,2]$, alors

$$R_n = n^{1-1/\alpha}\left(\overline{\mathbb{X}} - \mu\right) \rightsquigarrow Q.$$

La loi Q est décrite par 3 paramètres α, $|\beta| < 1$, $\sigma > 0$,

$$\int e^{itx} dQ_\alpha(x) = \begin{cases} \exp\left(-\sigma^\alpha |t|^\alpha \left(1 - i\beta \mathrm{sgn}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right)\right) & \text{si } \alpha \neq 1 \\ \exp\left(-\sigma|t|\left(1 - i\beta\frac{\pi}{2}\mathrm{sgn}(t)\log|t|\right)\right) & \text{si } \alpha = 1 \end{cases}$$

Q dépend de (α, β, σ) inconnus. Les estimer ?
Densité de Q généralement inconnue. Quantiles de Q ?

Queues de distributions épaisses

Pourquoi pas le bootstrap d'Efron ?

Domaine d'attraction de la loi normale (Athreya, 1987)

Loi de pareto
$(\tau = 3)$

$$F(x) = \left(1 - \frac{1}{x^{\tau}}\right) \mathbb{1}_{x>1}$$

La racine :

$$R_n = \sqrt{n}\left(\overline{X} - \mu\right)/\sigma.$$

où $n = 100$.

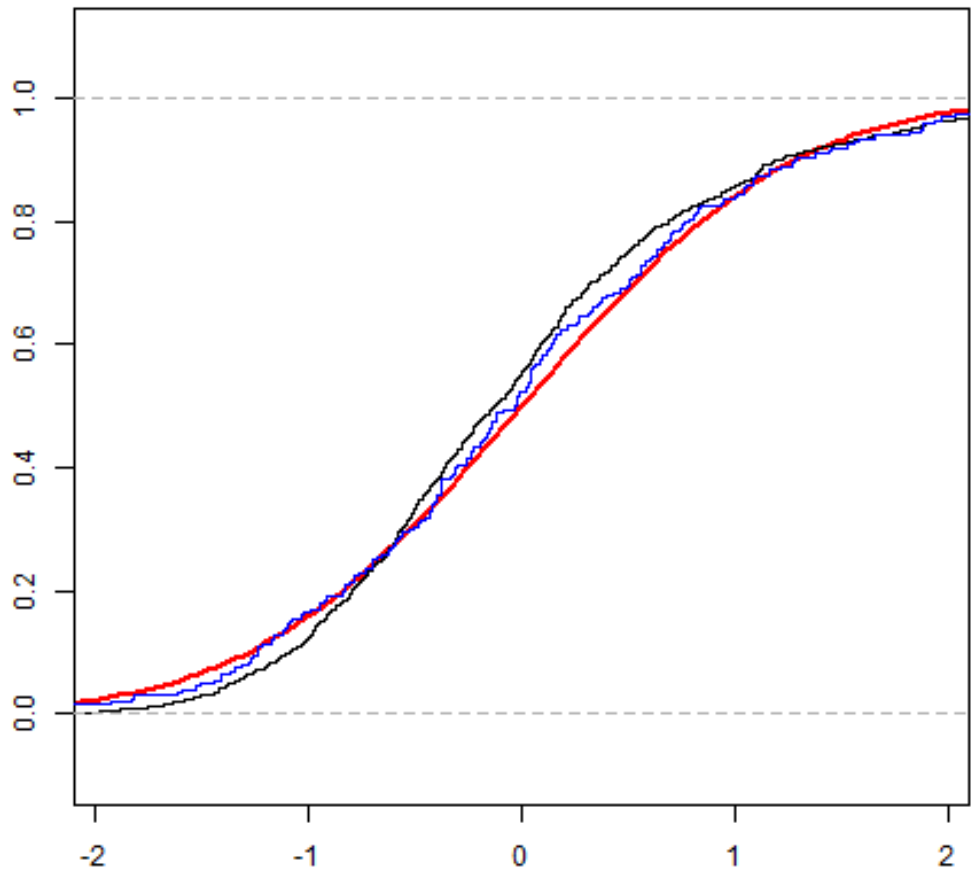


FIGURE: F.d.r. de la loi asymptotique, de la racine R_n, de la racine bootstrap R_n^*

En dehors du domaine d'attraction de la loi normale (Athreya, 1987)

Loi de pareto
($\tau = 1.5$)

$$F(x) = \left(1 - \frac{1}{x^\tau}\right) \mathbb{1}_{x>1}$$

La racine :

$$R_n = n\left(\overline{X} - \mu\right)/X_{(n)}.$$

où $n = 100$.

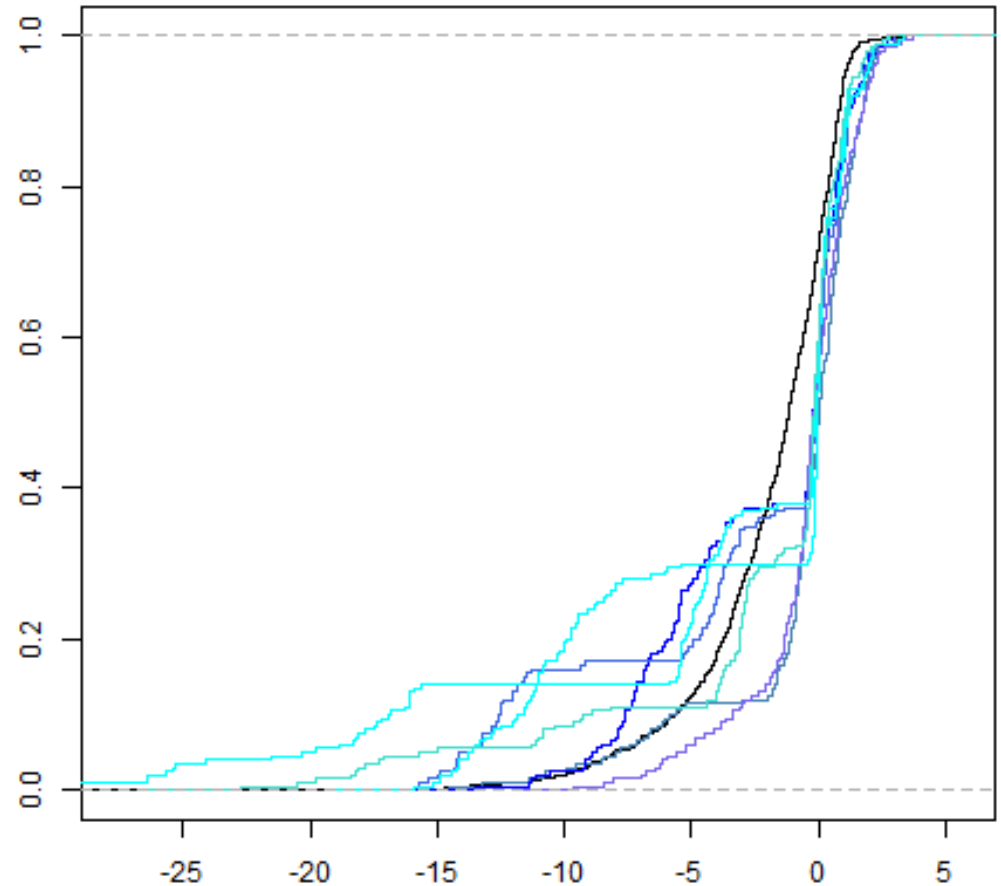


FIGURE: F.d.r. de la racine R_n et de la racine bootstrap R_n^* pour différents échantillons $\mathbb{X}$

L'explication (KRISHNA ATHREYA (1987))

Soit H la f.d.r. associée à la loi stable Q
Soit $\hat{\alpha}_n$ un estimateur consistant de α
Soit H_n^* la f.d.r. associée à $R_n^* = n^{1-1/\hat{\alpha}_n}\left(\overline{\mathbb{X}^*} - \overline{\mathbb{X}}\right)|\mathbb{X}$

Si le bootstrap de la moyenne convergeait, alors

$$\forall x \in \mathbb{R}, \quad H_n^*(x) \to H(x) \quad \text{p.s.}$$

Or, il s'avère que $H_n^*(x)$ converge vers une v.a. $H(x, Z)$.

Queues de distributions épaisses

Remèdes au bootstrap d'Efron

Avec un estimateur de α

- Le $m|n$ bootstrap ou le $\binom{n}{m}$ bootstrap

Soit $\hat{\alpha}_n$ un estimateur $\log(n)$-consistant de α
Soit H la f.d.r. associée à la loi stable Q
Soit H_n la f.d.r. associée à R_n
Soit $H^*_{m,n}$ la f.d.r associée à $R^*_{m,n} = n^{1-1/\hat{\alpha}_n}\left(\overline{\mathbb{X}^*_m} - \overline{\mathbb{X}}\right)|\mathbb{X}$

Politis-Romano-Wolf (1999) pour le $\binom{n}{m}$ bootstrap

Sous les hypothèses, $m \to \infty$, $m = o(n)$, et H_n est continue,

$$\sup_{x\in\mathbb{R}} \left|H^*_{m,n}(x) - H(x)\right| = o_{\mathbb{P}}(1)$$

$$\mathbb{P}\left(R_n \leq q^*_{m,n}(\delta)\right) \to \delta,$$

où $q^*_{m,n}(\delta)$ est le quantile d'ordre δ de $H^*_{m,n}$.

La différence entre avec ou sans remise est négligeable si $m^2/n = o(1)$.

Loi de pareto ($\tau = 1.5$) :

$$F(x) = \left(1 - \frac{1}{x^\tau}\right) \mathbb{1}_{x>1}$$

La racine :

$$R_n = n\left(\overline{\mathbb{X}} - \mu\right)/X_{(n)}.$$

où $n = 100$,
$m = 25$

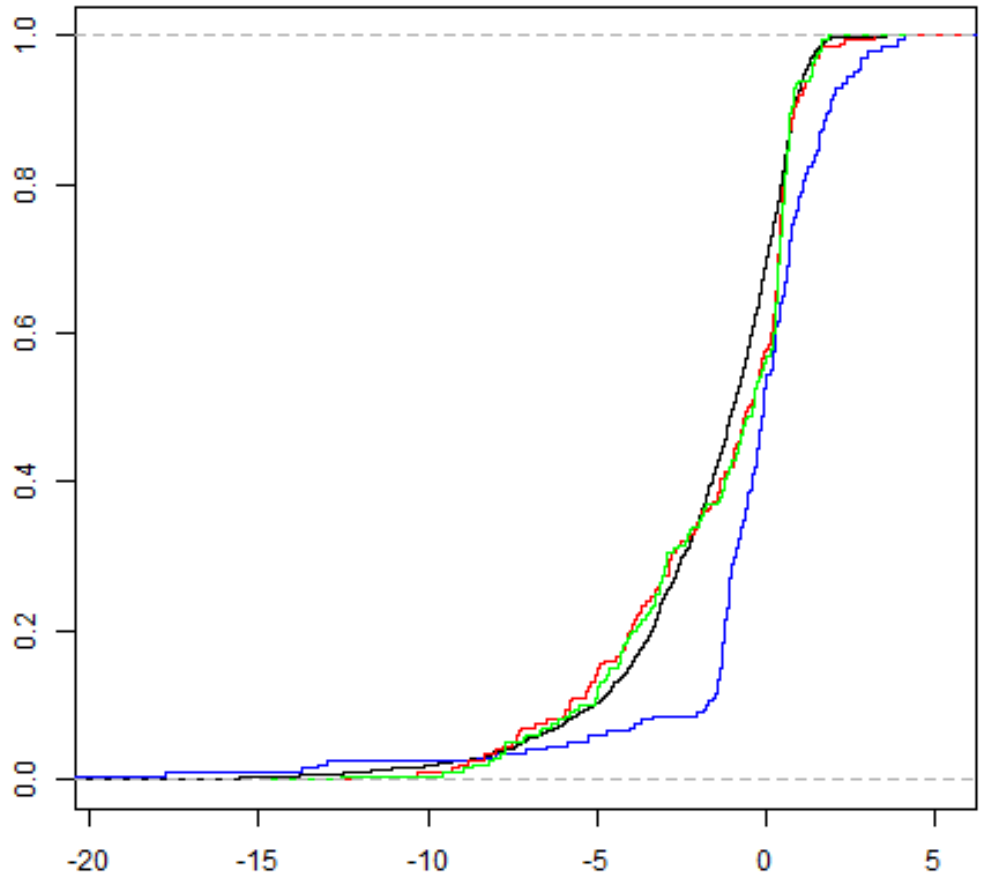


FIGURE: F.d.r. de la racine R_n et des racines bootstrap (Efron, m out of n avec remise et m out of n sans remise)

La différence entre avec ou sans remise est négligeable si $m^2/n = o(1)$.

Loi de pareto ($\tau = 1.5$) :

$$F(x) = \left(1 - \frac{1}{x^\tau}\right) \mathbb{1}_{x>1}$$

La racine :

$$R_n = n\left(\overline{\mathbb{X}} - \mu\right)/X_{(n)}.$$

où $n = 400$,
$m = 30$

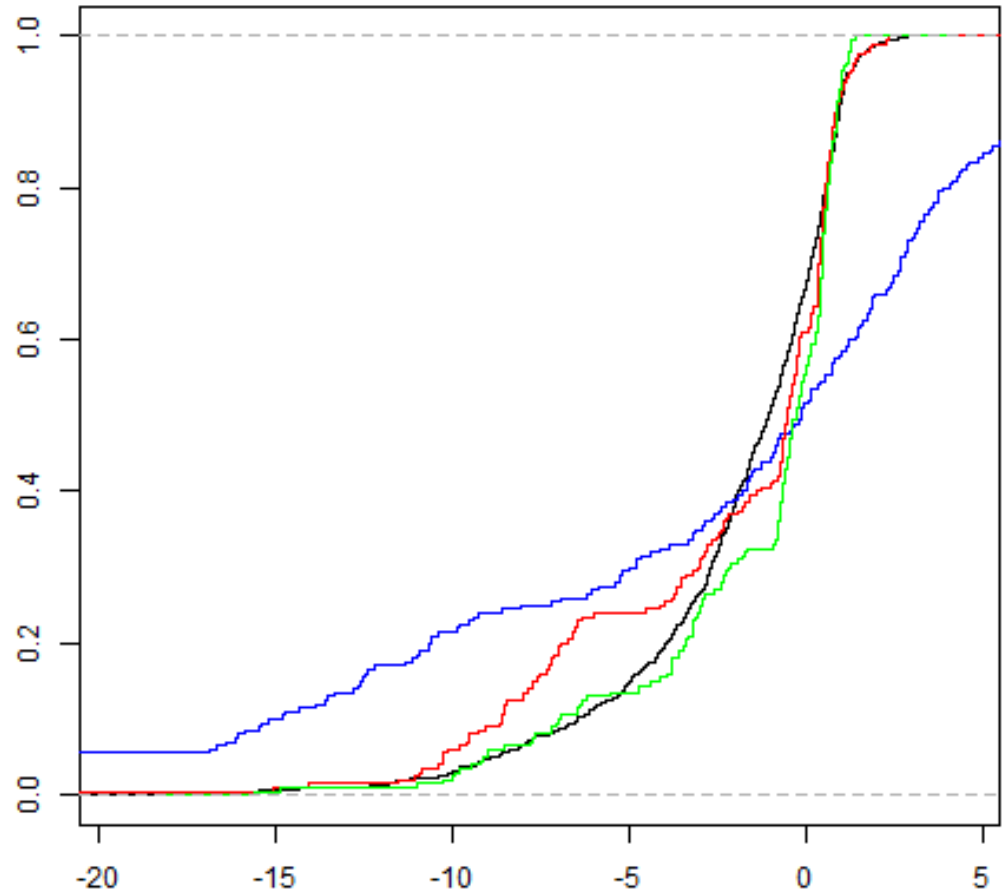


FIGURE: F.d.r. de la racine R_n et des racines bootstrap (Efron, m out of n avec remise et m out of n sans remise)

Avec une racine auto-normalisée

- Le $m|n$ bootstrap ou le $\binom{n}{m}$ bootstrap

Soit $\hat{\sigma}_n^2$ l'estimateur sans biais de la variance de P

$$S_n = \sqrt{n}\left(\bar{X} - \mu\right)/\hat{\sigma}_n \quad \text{converge en loi}$$

car S_n se décompose comme le quotient de termes convergeant vers des lois stables,

$$S_n = \frac{X_1 + \ldots + X_n - n\mu}{n^{1/\alpha}} \sqrt{\frac{n^{2/\alpha}}{(X_1 - \mu)^2 + \ldots + (X_n - \mu)^2}}$$

Politis-Romano-Wolf (1999) pour $\binom{n}{m}$ bootstrap

Idem avec les même hypothèses sur S_n

Loi de pareto
($\tau = 1.5$) :

$$F(x) = \left(1 - \frac{1}{x^\tau}\right) \mathbb{1}_{x>1}$$

La racine :

$$R_n = \sqrt{n}\left(\overline{\mathbb{X}} - \mu\right)/\hat{\sigma}_n.$$

où $n = 100$,
$m = 25$

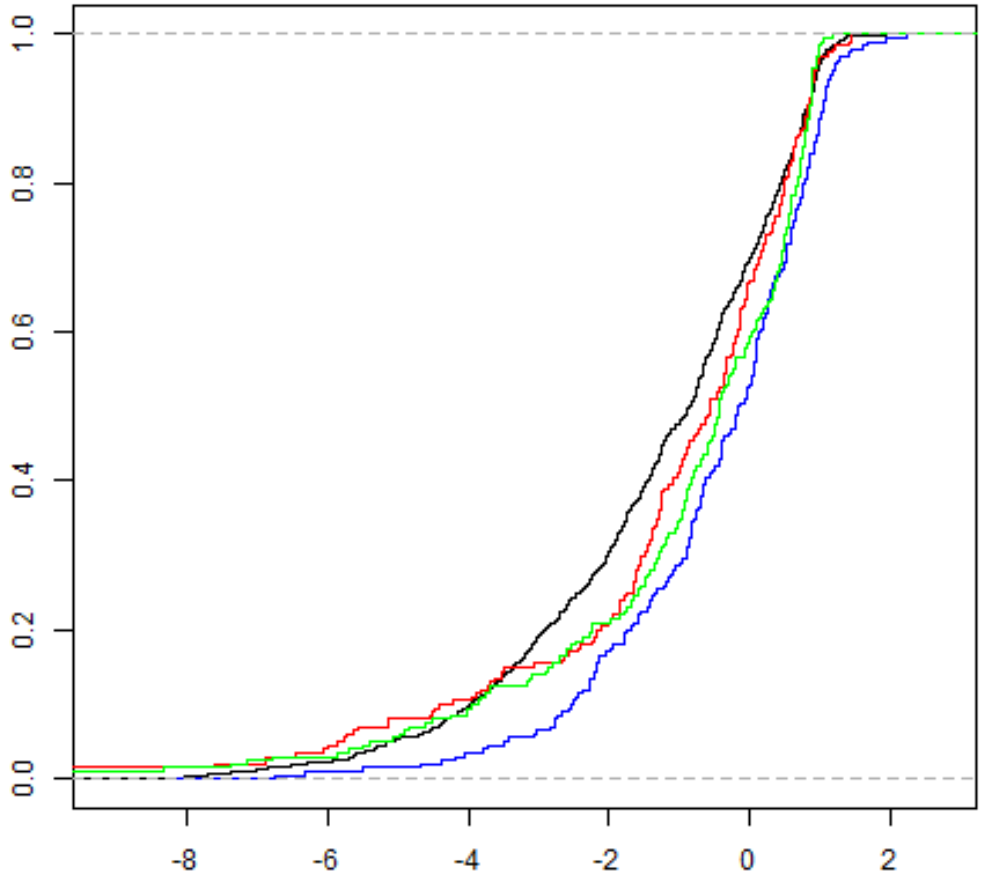


FIGURE: F.d.r. de la racine R_n et des racines bootstrap (Efron, m out of n avec remise et m out of n sans remise)

Loi de pareto
($\tau = 1.5$)

$$F(x) = \left(1 - \frac{1}{x^\tau}\right) \mathbb{1}_{x>1}$$

La racine :

$$R_n = \sqrt{n}\left(\overline{\mathbb{X}} - \mu\right)/\hat{\sigma}_n.$$

où $n = 400$,
$m = 30$

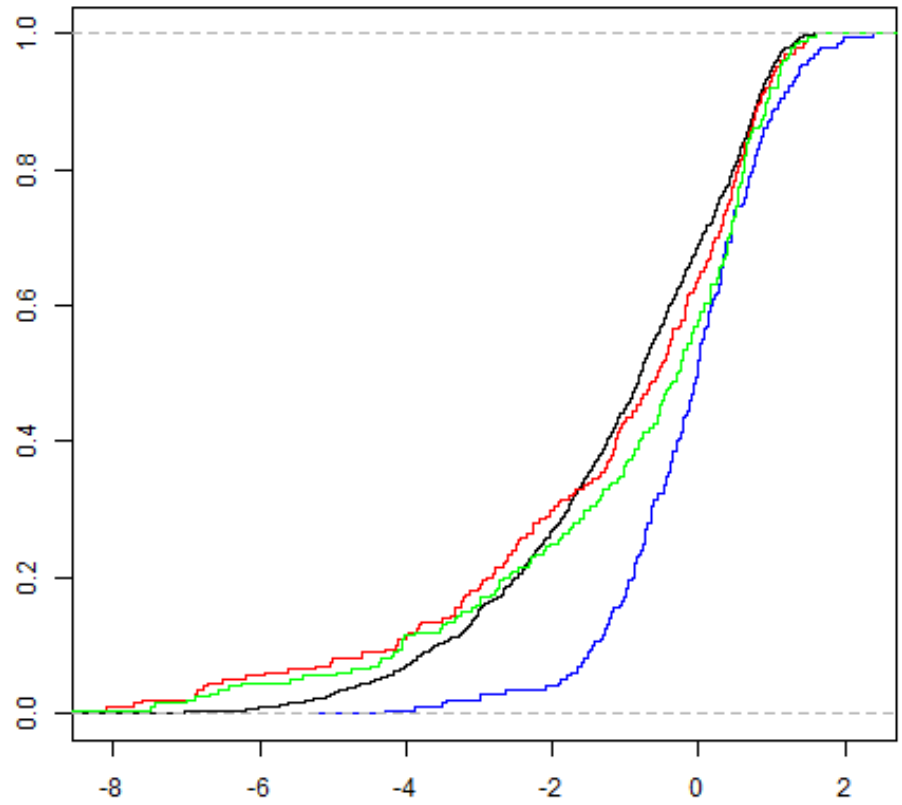


FIGURE: F.d.r. de la racine R_n et des racines bootstrap (Efron, m out of n avec remise et m out of n sans remise)

Bootstrap paramétrique

Adriana Cornea et Russell Davidson (2011) ont étudié le test de :

$$(H_0)\ \mu = 0 \text{ v.s. } (H_1)\ \mu \neq 0,$$

en considérant la racine auto-normalisée comme statistique de test, et en proposant une méthode de bootstrap paramétrique pour évaluer les p-valeurs.

- Plus $\alpha \to 1^+$, plus la convergence du bootstrap est lente,
- Quel que soit α, la convergence du bootsrap paramétrique semble plus rapide.
- Pour $n = 100$, le bootstrap paramétrique donne de meilleurs résultats.
- Pour $n = 1000$, les méthodes semblent équivalentes.

Plan

Valeurs extrêmes

Exemple d'échec pour le maximum

- $\mathbb{X}$ un n-échantillon de $\mathcal{U}[0, \theta]$.
- $\hat{\theta} = X_{n:n}$ un estimateur n-consistant de θ

$$R_n = n(\theta - \hat{\theta}) \overset{(\mathbb{P})}{\rightsquigarrow} \mathcal{E}(1/\theta)$$

La réplication bootstrap de R_n ne converge pas vers $\mathcal{E}(1)$

$$\mathbb{P}^* \left(R_n^* = 0\right) = \mathbb{P}^* \left(\hat{\theta}_n^* = \hat{\theta}_n\right) \to 1 - e^{-1}.$$

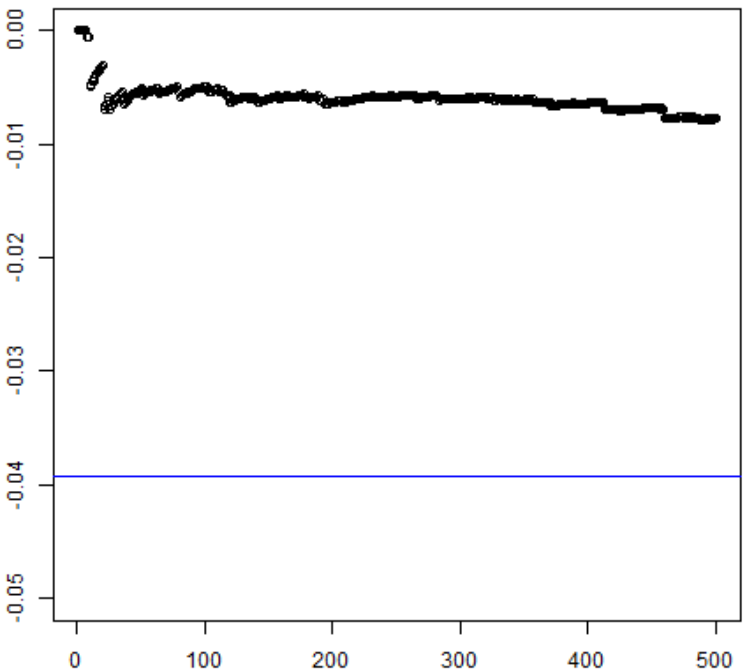
Un remède pour cet exemple est le bootstrap paramétrique :

$$\mathbb{X}^* \text{ est un } n - \text{échantillonn de } \mathcal{U}[0, \hat{\theta}].$$

Évaluation du biais de $\hat{\theta} = X_{n:n}$ (avec $n = 50$, $\theta = 2$) :

$$\mathbb{E}\left(\hat{\theta} - \theta\right) = -\theta/(n-1)$$

Bootstrap d'Efron



Bootstrap paramétrique

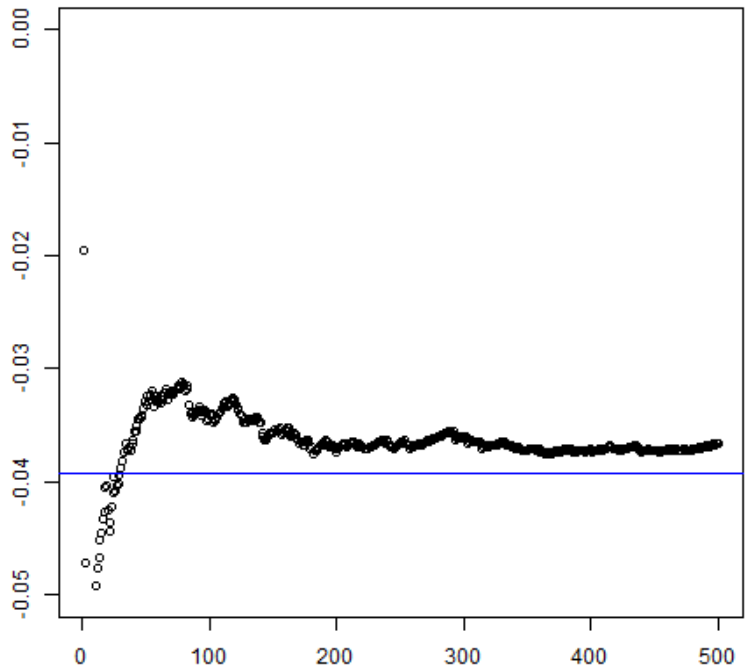


FIGURE: Comparaison des évaluations bootstrap du biais en fonction du nombre d'itérations B avec la valeur du biais théorique

Valeurs extrêmes

Rappels

Soit $\mathbb{X}$ un n-échantillon de P.
On note $X_{1:n} \leq \ldots \leq X_{n:n}$ les statistiques d'odre associées.

Soit G une f.d.r.

P est **dans le domaine d'attraction de** G pour le minimum s'il existe des constantes $a_n > 0$ et b_n telles que pour tout point de continuité x de G,

$$\mathbb{P}\left(\frac{1}{a_n}(X_{1:n} - b_n) \leq x\right) \to G(x).$$

$$R_n = \frac{1}{a_n}(X_{1:n} - b_n) \rightsquigarrow G.$$

G appartient à l'une de ces trois familles, et les coefficients a_n et b_n peuvent être déterminés en fonction de F selon la nature de G :

1. $G(x) = 1 - \exp(-e^x)$
   $a_n = \gamma_n - F^{-1}(1/(en))$, $b_n = \gamma_n$,
2. $G(x) = \mathbb{1}_{x>0} + (1 - \exp(-(-x)^{-\alpha}))\,\mathbb{1}_{x<0}$, avec $\alpha > 0$
   $a_n = \gamma_n$, $b_n = 0$,
3. $G(x) = (1 - \exp(-(x)^{\alpha}))\,\mathbb{1}_{x>0}$, avec $\alpha > 0$
   $a_n = \gamma_n - \theta_F$, $b_n = \theta_F$,

où $\gamma_n = F^{-1}(1/n)$, $\theta_F = \inf\{x : \ F(x) > 0\}$, et F la f.d.r. de P

Valeurs extrêmes

Échec du bootsrap d'Efron

Soit $\mathbb{X}^*$ un n-échantillon de P_n,

$$R_n^* = \frac{1}{a_n}(X_{1:n}^* - b_n)$$

Krishna Athreya et Jun-ichiro Fukuchi (1994) ont établi que la f.d.r. H_n^* associée à $R_n^*|\mathbb{X}$ converge vers un processus stochastique $H(x, Z)$.

Valeurs extrêmes

Remèdes au bootstrap d'Efron

Le $m|n$ bootstrap ou le $\binom{n}{m}$ bootstrap

Soit $\hat{a}_m$, $\hat{b}_m$ des estimateurs *plug-in* de a_m et b_m, de façon que

$$\hat{a}_m / a_m \xrightarrow{\mathbb{P}} 1, \qquad (\hat{b}_m - b_m)/\hat{a}_m \xrightarrow{\mathbb{P}} 0.$$

Soit $R^*_{m|n} = \left(X^*_{1:m} - \hat{b}_m\right)/\hat{a}_m$ la réplication de R_n à partir d'un échantillon $m|n$ bootstrap.
Soit $H^*_{m|n}$ la f.d.r associée à $R^*_{m,n}|\mathbb{X}$

Athreya-Fukuchi (1997) pour le $m|n$ bootstrap

Si $m \to \infty$ et $m = o(n)$, alors

$$\sup_{x \in \mathbb{R}} \left| H^*_{m|n}(x) - G(x) \right| = o_{\mathbb{P}}(1)$$

Exemple

Loi uniforme
($\theta = 2$)

La racine :

$R_n = n\,(\theta - X_{n:n})\,/\theta.$

où $n = 50$,
$m = 15$.

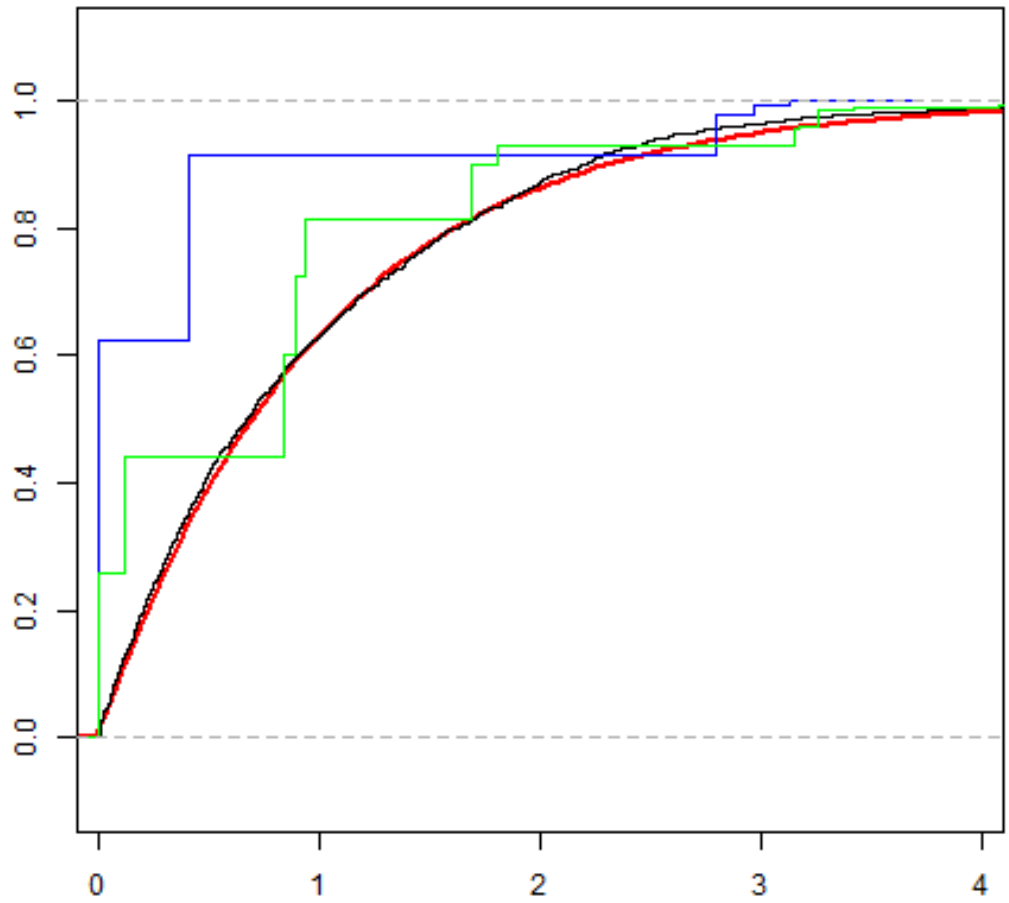


FIGURE: F.d.r. de la loi asymptotique, de la racine R_n, des racines bootstrap R^*_n, $R^*_{m|n}$

Exemple

Loi uniforme
$(\theta = 2)$

La racine :

$R_n = n\,(\theta - X_{n:n})\,/\theta.$

où $n = 400$,
$m = 20$.

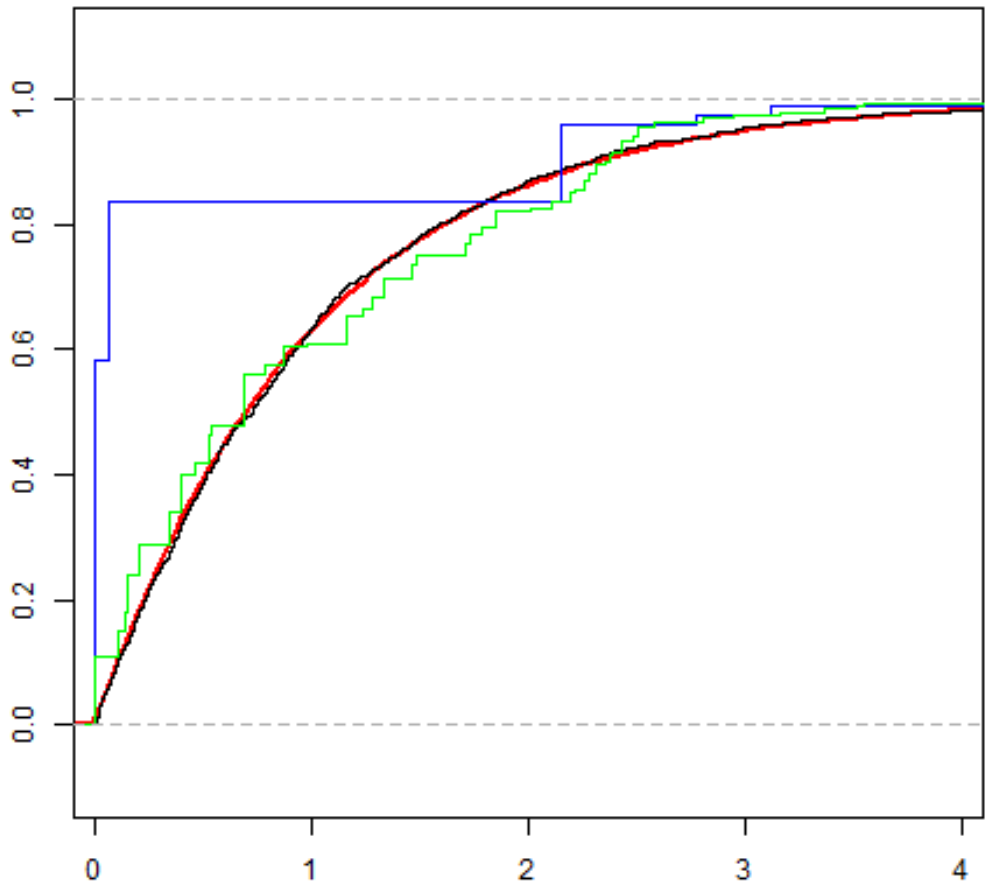


FIGURE: F.d.r. de la loi asymptotique, de la racine R_n, des racines bootstrap R_n^*, $R_{m|n}^*$

Le bootstrap lisse

Soit $\tilde{F}_n$ un estimateur à noyau de F,

$$\tilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

K est à support borné, $h_m/a_m = o_{\mathbb{P}}(1)$

Soit $\hat{a}_m$, $\hat{b}_m$ des estimateurs *plug-in* de a_m et b_m.
Soit $\mathbb{X}^*_m$ un m-échantillon de $\tilde{F}_n$, et $\tilde{R}^*_{m|n}$ la réplication bootstrap associée.
Soit $\tilde{H}^*_{m|n}$ la f.d.r. associée à $\tilde{R}^*_{m,n}|\mathbb{X}$.

Athreya-Fukuchi (1997) pour le $m|n$ bootstrap

Si $m \to \infty$ et $mF_n(a_m x + b_m) \to c(x)$ pour un sous-ensemble dense de $x \in \mathbb{R}$, alors

$$\sup_{x\in\mathbb{R}} \left|\tilde{H}^*_{m|n}(x) - G(x)\right| = o_{\mathbb{P}}(1)$$

Plan

U-statistiques

Soit $\mathbb{X} = (X_1, \ldots, X_n)$ un n-échantillon de P.
Soit $\theta = \mathbb{E}h(X_1, \ldots, X_r)$ où h fonction symétrique et $r \leq n$.

Un estimateur sans biais de θ est donné par :

$$U_n = \frac{1}{\binom{n}{r}} \sum_{1 \leq i_1 < \ldots < i_r \leq n} h(X_{i_1}, \ldots, X_{i_r}).$$

U_n est une **U-statistique** d'ordre r et de noyau h.

U-statistiques

La décomposition de Hoeffding

$$U_n = \theta + \sum_{c=1}^{r} \binom{r}{c} U_n^{(c)}$$

- $U_n^{(c)}$ est une U-statistique d'ordre c et de noyau $h^{(c)}$,

$$h^{(c)}(x_1, \ldots, x_c) = \int \ldots \int h(u_1, \ldots, u_r) \prod_{i=1}^{c} (\delta_{x_i} - P)(u_i) \prod_{j=c+1}^{r} dP(u_j).$$

- les v.a. $U_n^{(c)}$ sont décorrélées et $U_n^{(c)} = O_{\mathbb{P}}\left(1/n^{c/2}\right)$.
- U_n est **P-dégénérée à l'ordre $c-1$** si

$$\begin{cases} \mathbb{E}\left[h(X_1, \ldots, X_r) | X_1 \ldots X_{c-1}\right] = \mathbb{E}h(X_1, \ldots, X_r) & \text{p.s.} \\ \mathbb{E}\left[h(X_1, \ldots, X_r) | X_1 \ldots X_c\right] \neq \mathbb{E}h(X_1, \ldots, X_r) & \text{p.s.} \end{cases}$$

$$h^{(1)} \equiv 0, \ldots \; h^{(c-1)} \equiv 0, \text{ et } \quad h^{(c)} \not\equiv 0.$$

U-statistiques

Convergence d'une U-statistique selon l'ordre de dégénérescence

Si U_n est P-non-dégénérée, $\mathbb{E}h\left(X_1, \ldots X_r\right)^2 < \infty$,
U_n est un estimateur $\sqrt{n}$-consistant et asymptotiquement normal de θ, et

$$\begin{aligned} U_n - \theta &= rU_n^{(1)} + O_{\mathbb{P}}\left(1/n\right) \\ &= \frac{r}{n}\sum_{i=1}^{n} h^{(1)}(X_i) + O_{\mathbb{P}}\left(1/n\right) \end{aligned}$$

Si U_n est P-dégénérée à l'ordre $c-1$,

$$U_n - \theta = \binom{r}{c} U_n^{(c)} + O_{\mathbb{P}}\left(1/n^{(c+1)/2}\right),$$

où $\sqrt{\binom{n}{c}\binom{r}{c}}U_n^{(c)}$ converge en loi vers un chaos Gaussien $K_{P,h,c}$.

U-statistiques

Pourquoi pas le bootstrap d'Efron ?

✘ Parce qu'il n'est pas consistant dans le cas dégénéré

Exemple

- $\mathbb{X}$ n-échantillon de P d'espérance μ, de variance σ^2
- $\theta = \mu^2$, $h(x_1, x_2) = x_1 x_2$
- Décomposition de Hoeffding :

$$h^{(1)}(x_1) = (x_1 - \mu)\mu, \quad h^{(2)}(x_1, x_2) = (x_1 - \mu)(x_2 - \mu).$$

Si $\mu \neq 0$, U_n est P-non-dégénérée,

$$\sqrt{n}(U_n - \theta) \overset{\mathbb{P}}{\rightsquigarrow} \mathcal{N}(0, 4\theta\sigma^2), \qquad \sqrt{n}(U_n^* - \bar{X}^2) \overset{\mathbb{P}^*}{\rightsquigarrow} \mathcal{N}(0, 4\theta\sigma^2).$$

Si $\mu = 0$, U_n est P-dégénérée à l'ordre 1 et

$$n(U_n - \theta) \rightsquigarrow \sigma^2(Z^2 - 1), \qquad n(U_n^* - \bar{X}^2) \rightsquigarrow \sigma^2(Z^2 - 1 + YZ),$$

où Z, Y i.i.d. $\sim \mathcal{N}(0, 1)$.

Illustration : cas non dégénéré

Loi normale
$(\mu = 2, \sigma = \sqrt{2})$

La racine :

$R_n = \sqrt{n}\,(U - \theta) \,/\, \sqrt{\theta\sigma^2}$,

avec $n = 200$.

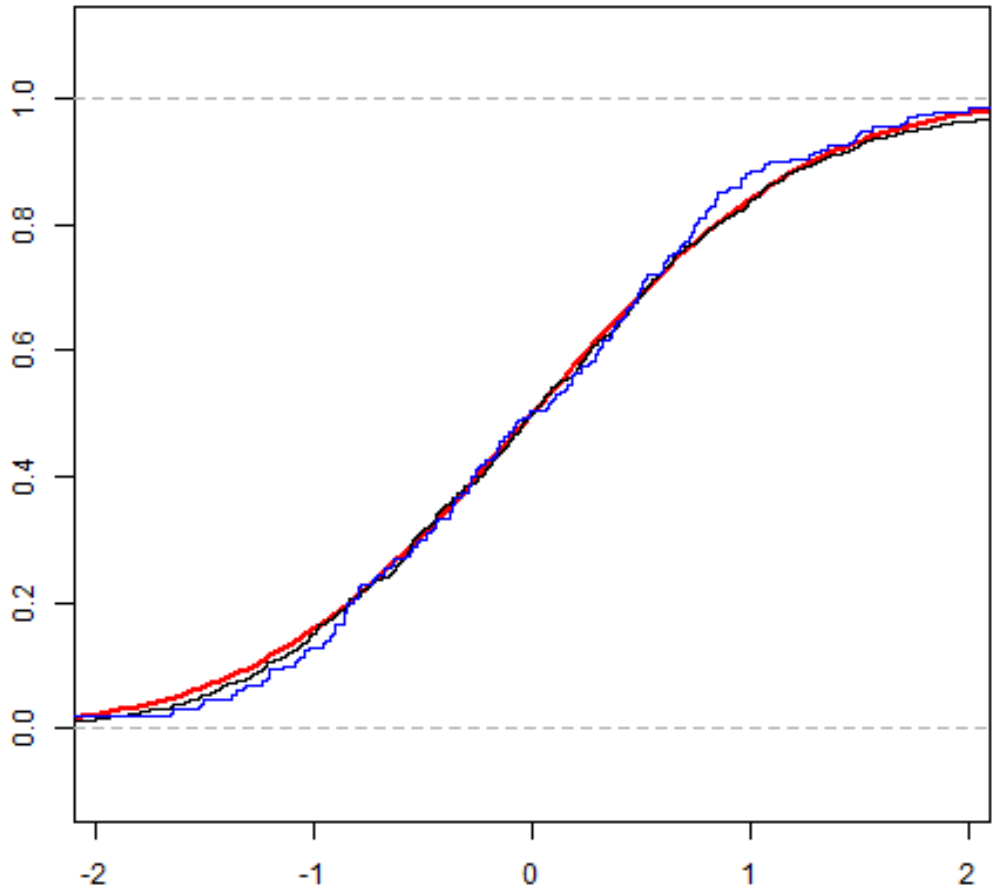


FIGURE: F.d.r. de la loi asymptotique, de la racine R_n, et de la racine bootstrap R_n^*

Illustration : cas dégénéré

Loi normale
$(\mu = 0, \sigma = \sqrt{2})$

La racine :

$R_n = n\,(U - \theta)\,/\sigma^2,$

avec $n = 200$.

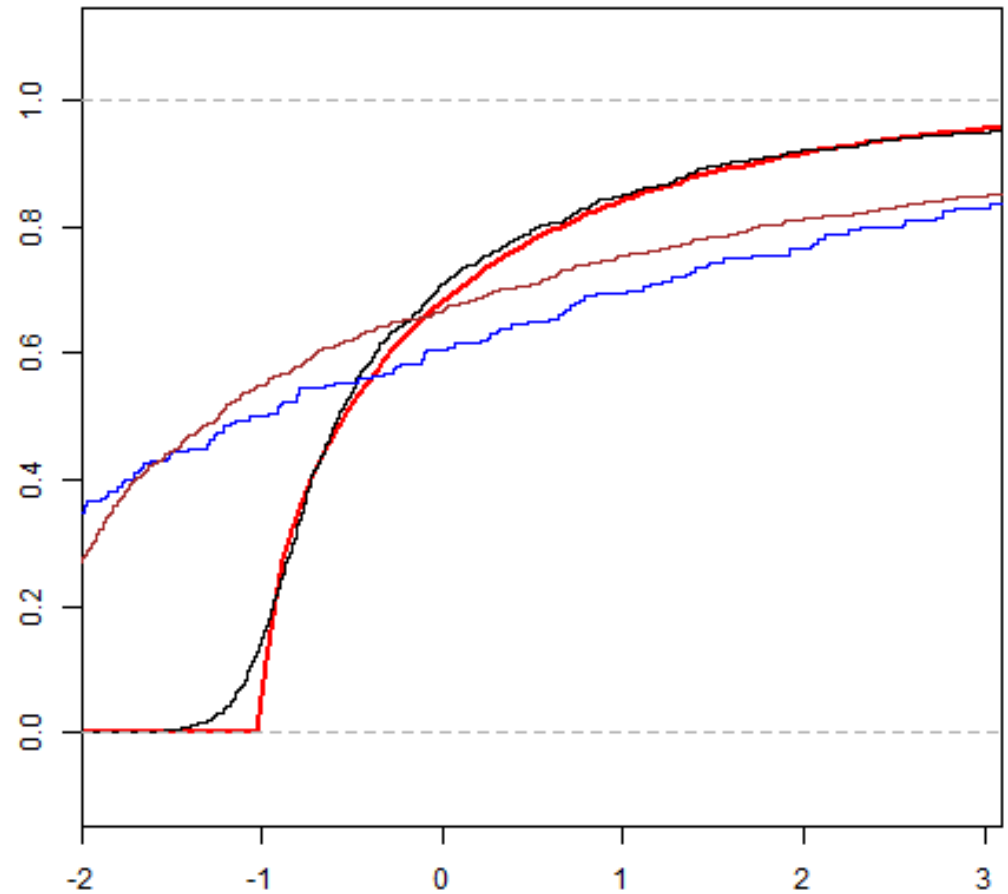


FIGURE: F.d.r. de la loi asymptotique, de la racine R_n, de la racine bootstrap R_n^*, de la loi limite de R_n^*

Bootstrap d'Efron pour $H_n^{(c)}$

Si U_n est P-dégénérée à l'ordre $c-1$,

$$n^{c/2}(U_n-\theta)\rightsquigarrow K_{P,h,c}, \qquad \binom{n}{c}^{1/2}\binom{r}{c}U_n^{(c)}\rightsquigarrow K_{P,h,c}.$$

$U_n^{*(c)}(\mathbb{X}_{n|n}^*)$ est la U-statistique de noyau

$$h^{*(c)}(x_1,\ldots,x_c)=\int\ldots\int h(u_1,\ldots,u_r)\prod_{i=1}^{c}(\delta_{x_i}-P_n)(u_i)\prod_{j=c+1}^{r}dP_n(u_j).$$

Arcones-Giné (1992)

Si $\mathbb{E}\left[h\left(X_{i_1},\ldots X_{i_r}\right)\right]^{d/r}<\infty$, pour $d=\#\{i_1,\ldots i_r\}$, alors

$$\binom{n}{c}^{1/2}\binom{r}{c}U_n^{*(c)}|\mathbb{X}\rightsquigarrow K_{P,h,c} \text{ en probabilité.}$$

Le $m|n$ bootstrap ou le $\binom{n}{m}$ bootstrap

Si U_n est P-dégénérée à l'ordre $c-1$,

$$n^{c/2}(U_n - \theta) \overset{\mathbb{P}}{\rightsquigarrow} K_{P,h,c}.$$

Soit $U^*_{m|n} = U_m\left(\mathbb{X}^*_{m|n}\right)$ la réplication bootstrap de U_n

Bretagnolle (1992)

Si $m = o(n)$ et $\mathbb{E}\left[h\left(X_{i_1}, \ldots X_{i_r}\right)\right]^{d/r} < \infty$, pour $d = \#\{i_1, \ldots i_r\}$, alors

$$\binom{n}{c}^{1/2}\left(U^*_{m|n} - U_n\right)|\mathbb{X} \rightsquigarrow K_{P,h,c} \text{ en probabilité.}$$

U-statistiques

Exemple : test d'indépendance de Hoeffding

Soit $F(y,z)$ la f.d.r de $X = (Y,Z)$.
On note F_Y et F_Z les f.d.r. resp. de Y et Z.

$$\Delta(F) = \int\int \{F(y,z) - F_Y(y)F_Z(z)\}^2 \, dF(y,z)$$

Sous l'hypothèse que F, F_Y, F_Z sont continues,

$$Y \perp\!\!\!\perp Z \quad \Leftrightarrow \quad \Delta(F) = 0$$

Un estimateur sans biais de $\Delta(F)$ est une U-stat D_n ($r = 5$) t. q.

Sous (H_0) $Y \perp\!\!\!\perp Z$, $\quad D_n = O_{\mathbb{P}}(1/n)$ (1-dégénérée)

Sous (H_1) $Y \not\perp\!\!\!\perp Z$, $\quad \sqrt{n}(D_n - \Delta) \rightsquigarrow \mathcal{N}(0, 25\delta_1^2)$ (non-dégénérée)

Test d'hypothèses :

$$(H_0)\ Y \perp\!\!\!\perp Z \qquad \text{v.s.} \qquad (H_1)\ Y \not\perp\!\!\!\perp Z$$

Problème : évaluer $q_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi limite de nD_n sous (H_0) (i.e. de $K_{P,h,2}$).

- Sous (H_0), $nD_n = \binom{5}{2} nD_n^{(2)} + o_{\mathbb{P}}(1)$
- Sous (H_1), $D_n = \Delta + \binom{5}{1} D_n^{(1)} + \binom{5}{2} D_n^{(2)} + o_{\mathbb{P}}(1/n)$

Soit $q^*_{1-\alpha}$ le quantile de la loi limite de $\left|\binom{5}{2} D_n^{*(2)}\right|$.
Le test rejetant (H_0) si $nD_n > q^*_{1-\alpha}$ est de niveau asymptotique α.

Plan

Régression linéaire multiple

Modèle de régression linéaire multiple

$$Y_i = \beta_0 x_{i,0} + \beta_1 x_{i,1} + \ldots + \beta_{p-1} x_{i,p-1} + \varepsilon_i, \quad \text{pour} \quad i = 1 \ldots n,$$

- $p \leq n$.
- Y_i variable aléatoire observée (variable à expliquer).
- $x_{i,0}, x_{i,1}, \ldots, x_{i,p-1}$ valeurs réelles (variables explicatives), souvent $x_{i,0} = 1$ pour tout i.
- $\beta_0, \beta_1, \ldots, \beta_{p-1}$ paramètres réels inconnus (cœfficients de régression).
- ε_i variables aléatoires, non observées (erreurs ou bruits), vérifiant les conditions $(C_1) - (C_3)$

  (C_1) $\mathbb{E}[\varepsilon_i] = 0$ (centrage).

  (C_2) $Cov(\varepsilon_i, \varepsilon_j) = 0$ (non corrélation).

  (C_3) $\mathbb{V}(\varepsilon_i) = \sigma^2$ (inconnue) (homoscédasticité).

Expression vectorielle

$$Y = \mathbb{X}\beta + \varepsilon,$$

avec

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \mathbb{X} = \begin{pmatrix} x_{1,0} & \dots & x_{1,p-1} \\ \vdots & & \vdots \\ x_{n,0} & \dots & x_{n,p-1} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix} \text{ et } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Sous $(C_1) - (C_3)$, on a alors :

- $\mathbb{E}[\varepsilon] = 0$ et $\mathbb{E}[Y] = \mathbb{X}\beta$,
- $\mathbb{V}(\varepsilon) = \mathbb{V}(Y) = \sigma^2 I_n$.

Hypothèse : $\text{rang}(\mathbb{X}) = p \Leftrightarrow \mathbb{X}'\mathbb{X}$ définie positive.

On note $c_{j+1} = \left[(\mathbb{X}'\mathbb{X})^{-1}\right]_{j+1,j+1}$.

Estimateur des moindres carrés ordinaires

- $\hat{\beta} \in \operatorname{argmin}_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p-1} \beta_j x_{i,j} \right)^2 = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \|Y - \mathbb{X}\beta\|^2$.

  $\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y$, $\mathbb{E}[\hat{\beta}] = \beta$ et $\mathbb{V}(\hat{\beta}) = \sigma^2(\mathbb{X}'\mathbb{X})^{-1}$.
- Vecteur des valeurs ajustées : $\hat{Y} = H_{\mathbb{X}}Y$, avec $H_{\mathbb{X}} = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$.
- Matrice chapeau : $H_{\mathbb{X}}$, dont les cœfficients sont notés $h_{i,j}$. $h_{i,i} = x_i(\mathbb{X}'\mathbb{X})^{-1}x_i'$, où $x_i = (x_{i,0}, \ldots, x_{i,p-1})$.
- Vecteur des résidus : $\hat{\varepsilon} = Y - \hat{Y}$, $\mathbb{E}[\hat{\varepsilon}] = 0$, $\mathbb{V}(\hat{\varepsilon}) = (1 - h_{i,i})\sigma^2$.
- Estimateur sans biais de la variance : $\widehat{\sigma^2} = \hat{\varepsilon}'\hat{\varepsilon}/(n-p)$.
- Prédiction : soit $x_{n+1} = (x_{n+1,0}, \ldots, x_{n+1,p-1})$, on souhaite prédire une nouvelle observation d'une variable $Y_{n+1} = \beta_0 x_{n+1,0} + \ldots + \beta_{p-1}x_{n+1,p-1} + \varepsilon_{n+1} = x_{n+1}\beta + \varepsilon_{n+1}$, avec $\mathbb{E}[\varepsilon_{n+1}] = 0$, $\mathbb{V}(\varepsilon_{n+1}) = \sigma^2$ et $Cov(\varepsilon_{n+1}, \varepsilon_i) = 0$. $\hat{Y}^p_{n+1} = x_{n+1}\hat{\beta}$.

Régression linéaire multiple

Modèle paramétrique gaussien

(C_4) ε est un vecteur gaussien, de sorte que :
$(C_1) - (C_4) \Leftrightarrow \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$.

M matrice réelle de taille $q \times p$ de rang q $(q \leq p)$

Proposition

- $M\hat{\beta} \sim \mathcal{N}(M\beta, \sigma^2 \left[M(\mathbb{X}'\mathbb{X})^{-1}M'\right])$.
- $\frac{1}{\sigma^2}\left[M(\hat{\beta}-\beta)\right]' \left[M(\mathbb{X}'\mathbb{X})^{-1}M'\right]^{-1} \left[M(\hat{\beta}-\beta)\right] \sim \chi^2(q)$.
- $(n-p)\widehat{\sigma^2}/\sigma^2 \sim \chi^2(n-p)$.
- Les estimateurs $\hat{\beta}$ et $\widehat{\sigma^2}$ sont indépendants.

$\hookrightarrow$ Racines pivotales pour la construction d'intervalles de confiance, statistiques de test.

Intervalles de confiance

Soit $\alpha \in]0,1[$. On note $t_{n-p}(u)$ le u-quantile de la loi $\mathcal{T}(n-p)$.
Un IC de niveau de confiance $(1-\alpha)$ pour β_j est donné par :

$$\hat{I}_j = \left[\hat{\beta}_j - t_{n-p}(1-\alpha/2)\sqrt{\widehat{\sigma^2} c_{j+1}}; \hat{\beta}_j + t_{n-p}(1-\alpha/2)\sqrt{\widehat{\sigma^2} c_{j+1}}\right].$$

Tests d'hypothèses sur les cœfficients de régression

$(H_0)\ \beta_j = 0$ versus $(H_1)\ \beta_j \neq 0$.

Soit $\alpha \in]0,1[$, $T(Y) = \hat{\beta}_j / \sqrt{\widehat{\sigma^2} c_{j+1}}$.
Le test rejetant (H_0) si $|T(Y)| \geq t_{n-p}(1-\alpha/2)$ est de niveau α.

Intervalle de prédiction

Un intervalle de prédiction pour Y_{n+1} de niveau de confiance $(1-\alpha)$ est donné par

$$\widehat{I}^p_{n+1} = \Big[\hat{Y}^p_{n+1} - t_{n-p}(1-\alpha/2)\sqrt{\widehat{\sigma^2}(1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})};$$
$$\hat{Y}^p_{n+1} + t_{n-p}(1-\alpha/2)\sqrt{\widehat{\sigma^2}(1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}\Big].$$

Régression linéaire multiple

Théorèmes limites

Les ε_i sont supposés i.i.d. de loi inconnue P, vérifiant $(C_1) - (C_3)$.

Pour chaque n, la dépendance de Y, $\mathbb{X}$, $\hat{\beta}$, $\hat{\varepsilon}$, $\widehat{\sigma^2}$, c_{j+1} est précisée par $^{(n)}$.

Proposition

Si $\mathbb{X}^{(n)\prime}\mathbb{X}^{(n)}/n \to_{n\to+\infty} V$ définie positive, alors

- $\sqrt{n}\left(\hat{\beta}^{(n)} - \beta\right) \rightsquigarrow \mathcal{N}(0, \sigma^2 V^{-1})$.
- $\sqrt{\mathbb{X}^{(n)\prime}\mathbb{X}^{(n)}/\widehat{\sigma^2}^{(n)}}\left(\hat{\beta}^{(n)} - \beta\right) \rightsquigarrow \mathcal{N}(0, I_p)$.

Intervalles de confiance asymptotiques

On note $z(u)$ le u-quantile de la loi $\mathcal{N}(0,1)$. Un intervalle de confiance asymptotique de niveau de confiance asymptotique $(1-\alpha)$ pour β_j est donné par :

$$\hat{I}_j^{(n)} = \left[\hat{\beta}_j^{(n)} - z(1-\alpha/2)\sqrt{\widehat{\sigma^2}^{(n)} c_{j+1}^{(n)}}; \hat{\beta}_j + z(1-\alpha/2)\sqrt{\widehat{\sigma^2} c_{j+1}^{(n)}}\right].$$

$\hookrightarrow$ Même idée pour les tests d'hypothèses et les intervalles de prédiction.

Estimation bootstrap des résidus

Les ε_i sont supposés i.i.d. de loi inconnue P, vérifiant (C_1) – (C_3). On considère une racine $R_n(\mathcal{Y}, \beta)$.
Soit $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \overline{\hat{\varepsilon}}$ ($\tilde{\varepsilon}_i = \hat{\varepsilon}_i$ si le modèle est avec constante).

Le bootstrap des résidus se résume selon le schéma suivant.

β	$\leftrightarrow$	$\hat{\beta}$
P	$\leftrightarrow$	$\tilde{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{\tilde{\varepsilon}_i}$
ε_i i.i.d. de loi P	$\leftrightarrow$	ε_i^* i.i.d. de loi $\tilde{P}_n$ conditionnellement à $\mathcal{Y}$, échantillon bootstrap associé au vecteur des résidus recentrés.
$\mathcal{Y}_i = x_i\beta + \varepsilon_i$	$\leftrightarrow$	$\mathcal{Y}_i^* = x_i\hat{\beta} + \varepsilon_i^*$.
$\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathcal{Y}$	$\leftrightarrow$	$\beta^* = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathcal{Y}^*$.
$R_n(\mathcal{Y}, \beta)$	$\leftrightarrow$	$R_n^* = R_n(\mathcal{Y}^*, \hat{\beta})$

Estimation bootstrap de la loi de $R_n(Y, \beta)$

La loi de $R_n(Y, \beta)$ est estimée par la loi conditionnelle de $R_n^* = R_n(Y^*, \hat{\beta})$ sachant Y.

Rééchantillonnage des résidus

Approximation de Monte Carlo et rééchantillonnage

Conditionnellement à Y, une réalisation de ε_i^* peut être simulée en tirant au hasard avec remise n valeurs dans $\{\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n\}$.

La loi conditionnelle de R_n^* sachant $Y = y$ peut donc être approchée par une méthode de Monte Carlo.

Intervalles de confiance bootstrap

Racines considérées :

- $R_n(Y, \beta) = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma^2} c_{j+1}}} \leftrightarrow R_n^* = R_n(Y^*, \hat{\beta}) = \frac{\beta_j^* - \hat{\beta}_j}{\sqrt{\sigma^{*2} c_{j+1}}}$, où $\sigma^{*2} = (Y^* - \mathbb{X}\beta^*)'(Y^* - \mathbb{X}\beta^*)/(n-p)$.
- $\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y \leftrightarrow \beta^* = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y^*$.

Intervalles de confiance bootstrap

Soit $\alpha \in]0,1[$. On note :
$t_u*(Y)$ le u-quantile de la loi de $R_n^*|Y$,
$q_u^*(Y)$ le u-quantile de la loi de $\beta^*|Y$.

- Intervalle de confiance bootstrap-t pour β_j :
$$I_j^* = \left[\hat{\beta}_j - t_{1-\alpha/2}^*(Y)\sqrt{\widehat{\sigma^2} c_{j+1}}; \hat{\beta}_j - t_{\alpha/2}^*(Y)\sqrt{\widehat{\sigma^2} c_{j+1}}\right].$$
- Intervalle de confiance bootstrap percentile pour β_j :
$$I_j^* = \left[q_{\alpha/2}^*(Y), q_{1-\alpha/2}^*(Y)\right].$$

Intervalles de confiance bootstrap approchés

- $\varepsilon^{*1}, \ldots, \varepsilon^{*B}$ B échantillons bootstrap de $\tilde{\varepsilon}$.
- $Y^{*b} = \mathbb{X}\hat{\beta} + \varepsilon^{*b}$, β^{*b} estimateur des MCO calculé sur Y^{*b}.
- $R_n^{*b} = R_n(Y^{*b}, \hat{\beta})$.
- $(R_n^{*(1)}, \ldots, R_n^{*(B)})$ stat. d'ordre associée à $(R_n^{*b})_b$.
- $(\beta^{*(1)}, \ldots, \beta^{*(B)})$ stat. d'ordre associée à $(\beta^{*b})_b$.

$t^*_{\alpha/2}(Y)$ et $t^*_{1-\alpha/2}(Y)$ sont approchés par $R_n^{*(\lceil B\alpha/2 \rceil)}$ et $R_n^{*(\lceil B-B\alpha/2 \rceil)}$.
$q^*_{\alpha/2}(Y)$ et $q^*_{1-\alpha/2}(Y)$ sont approchés par $\beta^{*(\lceil B\alpha/2 \rceil)}$ et $\beta^{*(\lceil B-B\alpha/2 \rceil)}$.

- Intervalle de confiance bootstrap-t approché pour β_j :
$$\left[\hat{\beta}_j - R_n^{*(\lceil B-B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2} c_{j+1}}; \hat{\beta}_j - R_n^{*(\lceil B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2} c_{j+1}}\right].$$
- Intervalle de confiance bootstrap percentile approché :
$$\left[\beta^{*(\lceil B\alpha/2 \rceil)}; \beta^{*(\lceil B-B\alpha/2 \rceil)}\right].$$

Algorithme IC bootstap-t pour les cœfficients de régression

Variable

B : *entier* 999, 9999...

Début

Calculer $\hat{\beta}$ estimateur des MCO sur $\mathcal{Y}$

Calculer $\tilde{\varepsilon}$ vecteur des résidus recentrés

Calculer $\widehat{\sigma^2}$ estimateur de la variance sur $\mathcal{Y}$

Pour b **variant de** 1 **à** B

Calculer ε^{*b} échantillon bootstrap de $\tilde{\varepsilon}$

Calculer $\mathcal{Y}^{*b} = \mathbb{X}\hat{\beta} + \varepsilon^{*b}$

Calculer β^{*b} estimateur des MCO sur $\mathcal{Y}^{*b}$

Calculer σ^{2*b} estimateur de la variance sur $\mathcal{Y}^{*b}$

Calculer $R_n^{*b} = R_n(\mathcal{Y}^{*b}, \hat{\beta})$

FinPour

Calculer $(R_n^{*(1)}, \ldots, R_n^{*(B)})$ statistique d'ordre de $(R_n^{*b})_b$

Retourner $\hat{\beta}_j - R_n^{*(\lceil B - B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2} c_{j+1}}, \hat{\beta}_j - R_n^{*(\lceil B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2} c_{j+1}}$

Fin

Régression linéaire multiple

Bootstrap et rééchantillonnage des résidus

Algorithme IC bootstap percentile pour les cœfficients de régression

Variable

B : *entier* 999, 9999...

Début

Calculer $\hat{\beta}$ estimateur des MCO sur $\mathcal{Y}$

Calculer $\tilde{\varepsilon}$ vecteur des résidus recentrés

Pour b **variant de** 1 **à** B

Calculer ε^{*b} échantillon bootstrap de $\hat{\varepsilon}$

Calculer $\mathcal{Y}^{*b} = \mathbb{X}\hat{\varepsilon} + \varepsilon^{*b}$

Calculer β^{*b} estimateur des MCO sur $\mathcal{Y}^{*b}$

FinPour

Calculer $(\beta^{*(1)}, \ldots, \beta^{*(B)})$ statistique d'ordre de $(\beta^{*b})_b$

Retourner $\beta^{*(\lceil B\alpha/2 \rceil)}$, $\beta^{*(\lceil B-B\alpha/2 \rceil)}$

Fin

Tests d'hypothèses bootstrap

$(H_0)\ \beta_j = 0$ v.s. $(H_1)\ \beta_j \neq 0$

- Statistique de test : $T(Y) = \dfrac{\hat{\beta}_j}{\sqrt{\widehat{\sigma^2} c_{j+1}}}$.
- Racine : $R_n(Y, \beta) = \dfrac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma^2} c_{j+1}}} \leftrightarrow R_n^* = R_n(Y^*, \hat{\beta}) = \dfrac{\beta_j^* - \hat{\beta}_j}{\sqrt{\sigma^{*2} c_{j+1}}}$.

Tests d'hypothèses bootstrap

Soit $\alpha \in]0, 1[$.
Soit $t_u^*(\mathbb{X})$ le $(1-\alpha)$ quantile de la loi conditionnelle de R_n^* sachant Y. Un test bootstrap est donné par :

$$\Phi_\alpha * (\mathbb{X}) = \begin{cases} 1 & \text{si } T(Y) < t^*_{\alpha/2}(Y) \text{ ou } T(Y) > t^*_{1-\alpha/2}(Y) \\ 0 & \text{sinon.} \end{cases}$$

✘ Si la loi de la racine $R_n(Y, \beta)$ semble très éloignée de la loi normale, on changera de statistique de test (typiquement $T(Y) = \hat{\beta}$), et de racine.

Tests d'hypothèses bootstrap approchés

- $\varepsilon^{*1}, \ldots, \varepsilon^{*B}$ B échantillons bootstrap de $\tilde{\varepsilon}$.
- $Y^{*b} = \mathbb{X}\hat{\beta} + \varepsilon^{*b}$, β^{*b} estimateur des MCO calculé sur Y^{*b}.
- $R_n^{*b} = R_n(Y^{*b}, \hat{\beta})$.
- $(R_n^{*(1)}, \ldots, R_n^{*(B)})$ stat. d'ordre associée à $(R_n^{*b})_b$.

$t^*_{\alpha/2}(Y)$ et $t^*_{1-\alpha/2}(Y)$ sont approchés par $R_n^{*(\lceil B\alpha/2 \rceil)}$ et $R_n^{*(\lceil B - B\alpha/2 \rceil)}$.

- Premier test : on rejette (H_0) si $T(Y) < R_n^{*(\lceil B\alpha/2 \rceil)}$ ou $T(Y) > R_n^{*(\lceil B - B\alpha/2 \rceil)}$.
- Deuxième test : étant donnée une observation y de Y, on rejette (H_0) si $\frac{\#\{b,\ R_n^{*b} \leq T(y)\}+1}{B+1}$ ou $\frac{\#\{b,\ R_n^{*b} \geq T(y)\}+1}{B+1}$ est $\leq \alpha/2$.

Algorithme Tests bootstap pour les cœfficients de régression

Variable

B : *entier* $999, 9999$

Début

Calculer $\hat{\beta}$ estimateur des MCO sur Y

Calculer $\tilde{\varepsilon}$ vecteur des résidus recentrés

Calculer $\widehat{\sigma^2}$ estimateur de la variance sur Y

Pour b **variant de** 1 **à** B

Calculer ε^{*b} échantillon bootstrap de $\hat{\varepsilon}$

Calculer $Y^{*b} = \mathbb{X}\hat{\varepsilon} + \varepsilon^{*b}$

Calculer β^{*b} estimateur des MCO sur Y^{*b}

Calculer σ^{2*b} estimateur de la variance sur Y^{*b}

Calculer $R_n^{*b} = R_n(Y^{*b}, \hat{\beta})$

FinPour

Retourner $\frac{\#\{b,\ R_n^{*b} \leq T(y)\}+1}{B+1}$ et $\frac{\#\{b,\ R_n^{*b} \geq T(y)\}+1}{B+1}$

Fin

Intervalles de prédiction bootstrap-t

Racine considérée :

$$R_n(Y,\beta) = \frac{x_{n+1}\hat{\beta} - Y_{n+1}}{\sqrt{\widehat{\sigma^2}(1+x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}} \leftrightarrow R_n^* = \frac{x_{n+1}(\beta^* - \hat{\beta}) - \varepsilon^*_{n+1}}{\sqrt{\sigma^{2^*}(1+x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}},$$

où $\varepsilon^*_{n+1} \sim \tilde{P}_n$.

Intervalle de prédiction

Soit $\alpha \in]0,1[$, $t_u*(Y)$ le u-quantile de la loi de $R_n^*|Y$.

Intervalle de prédiction bootstrap-t pour Y_{n+1} :

$$I^*_{n+1} = \Big[\hat{Y}^p_{n+1} - t^*_{1-\alpha/2}(Y)\sqrt{\widehat{\sigma^2}(1+x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})};$$
$$\hat{Y}^p_{n+1} - t^*_{\alpha/2}(Y)\sqrt{\widehat{\sigma^2}(1+x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}\Big].$$

Intervalles de prédiction bootstrap approchés

- $\varepsilon^{*1}, \ldots, \varepsilon^{*B}$ échantillons bootstrap de taille $(n+1)$ de $\tilde{\varepsilon}$.
- $Y^{*b} = \mathbb{X}\hat{\beta} + \varepsilon^{*b}$, et β^{*b} estimateur des MCO calculé sur Y^{*b}.
- $R_n^{*b} = \dfrac{x_{n+1}(\beta^* - \hat{\beta}) - \varepsilon_{n+1}^{*b}}{\sqrt{\sigma^{2*b}(1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}}$.
- $(R_n^{*(1)}, \ldots, R_n^{*(B)})$ stat. d'ordre associée à $(R_n^{*b})_b$.

$t^*_{\alpha/2}(Y)$ et $t^*_{1-\alpha/2}(Y)$ sont approchés par $R_n^{*(\lceil B\alpha/2 \rceil)}$ et $R_n^{*(\lceil B - B\alpha/2 \rceil)}$.

Intervalle de prédiction bootstrap-t approché :

$$\Big[\hat{Y}^p_{n+1} - R_n^{*(\lceil B - B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2}(1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})};$$
$$\hat{Y}^p_{n+1} - R_n^{*(\lceil B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2}(1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}\Big].$$

Algorithme Intervalle de prédiction bootstap-t

Variable

B : *entier* 999, 9999...

Début

Calculer $\hat{\beta}$ estimateur des MCO sur $\mathcal{Y}$

Calculer $\widehat{\tilde{\varepsilon}}$ vecteur des résidus recentrés

Calculer $\widehat{\sigma^2}$ estimateur de la variance sur $\mathcal{Y}$

Pour b **variant de** 1 **à** B

Calculer ε^{*b} échantillon bootstrap de taille $(n+1)$

Calculer $\mathcal{Y}^{*b} = \mathbb{X}\hat{\beta} + \varepsilon^{*b}$

Calculer β^{*b} estimateur des MCO sur $\mathcal{Y}^{*b}$

Calculer σ^{2*b} estimateur de la variance sur $\mathcal{Y}^{*b}$

Calculer $R_n^{*b} = R_n(\mathcal{Y}^{*b}, \hat{\beta})$

FinPour

Calculer $(R_n^{*(1)}, \ldots, R_n^{*(B)})$ statistique d'ordre de $(R_n^{*b})_b$

Retourner $\hat{Y}^p_{n+1} - R_n^{*(\lceil B - B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2}(1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}$

Retourner $\hat{Y}^p_{n+1} - R_n^{*(\lceil B\alpha/2 \rceil)} \sqrt{\widehat{\sigma^2}(1 + x_{n+1}(\mathbb{X}'\mathbb{X})^{-1}x'_{n+1})}$

Fin

Bootstrap des résidus : avantages

✔ Approche stable et efficace si les hypothèses du modèle sont vérifiées.

Bootstrap des résidus : inconvénients

✘ Les variables explicatives doivent pouvoir être considérées comme déterministes.
✘ Approche très sensible aux écarts au modèle.

Deux alternatives possibles

✔ Bootstrap par paires qui prend en compte le caractère aléatoire des variables explicatives, et qui ne se base pas sur la structure linéaire du modèle.

✔ Bootstrap sauvage qui sera moins sensible à une éventuelle hétéroscédasticité.

Régression linéaire multiple

Bootstrap et rééchantillonnage des paires

Modèle considéré : les x_i sont maintenant considérées comme des variables aléatoires X_i.

$\hookrightarrow$ Résultats précédents valables en conditionnant par $X_i = x_i$.

Estimation bootstrap par paires

$$\beta \quad \leftrightarrow \quad \hat{\beta}$$

$$Z = (Z_1, \ldots, Z_n), \quad Z_i = (X_i, Y_i) \quad \leftrightarrow \quad Z^* = (Z_1^*, \ldots, Z_n^*),\ Z_i^* = (X_i^*, Y_i^*),$$

n-échantillon bootstrap associé à Z

$$\mathbb{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \ Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \leftrightarrow \quad \mathbb{X}^* = \begin{pmatrix} X_1^* \\ \vdots \\ X_n^* \end{pmatrix}, \ Y^* = \begin{pmatrix} Y_1^* \\ \vdots \\ Y_n^* \end{pmatrix}$$

$$\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y \quad \leftrightarrow \quad \beta^* = (\mathbb{X}^{*\prime}\mathbb{X}^*)^{-1}\mathbb{X}^{*\prime}Y^*$$

$$R_n(Z, \beta) = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma^2} c_{j+1}}} \quad \leftrightarrow \quad R_n^* = R_n(Z^*, \hat{\beta}) = \frac{\beta_j^* - \hat{\beta}_j}{\sqrt{\sigma^{*2} c_{j+1}^*}}, \text{ où}$$

$$c_{j+1}^* = \left[(\mathbb{X}^{*\prime}\mathbb{X}^*)^{-1}\right]_{j+1,j+1}$$

Estimation bootstrap de la loi de $R_n(Z, \beta)$

La loi de $R_n(Z, \beta)$ est estimée par la loi conditionnelle de $R_n^* = R_n(Z^*, \hat{\beta})$ sachant Z.

Rééchantillonnage des paires

Approximation de Monte Carlo et rééchantillonnage

Conditionnellement à $Z = z$, une réalisation de l'échantillon bootstrap Z^* peut être simulée en tirant au hasard avec remise n valeurs dans $\{z_1, \ldots, z_n\}$.

La loi conditionnelle de R_n^* sachant $Y = y$ peut donc être approchée par une méthode de Monte Carlo.

$\hookrightarrow$ Intervalles de confiance bootstrap-t, bootstrap percentiles, tests d'hypothèses, intervalles de prédiction bootstrap peuvent être construits en adaptant les définitions et algorithmes précédents au bootstrap par paires.

Régression linéaire multiple

Bootstrap sauvage

Modèle hétéroscédastique : conditonnellement à $X_i = x_i$, les ε_i sont indépendantes, centrées, mais de variance σ_i^2.

Estimation bootstrap sauvage

$$
\begin{array}{lcl}
\beta & \leftrightarrow & \hat{\beta} \\
\varepsilon_i & \leftrightarrow & \varepsilon_i^* = \tilde{\varepsilon}_i W_i, \text{ où } W_1, \ldots, W_n \text{ i.i.d.,} \\
 & & \mathbb{E}[W_i] = 0, \mathbb{E}[W_i^2] = 1, \mathbb{E}[W_i^3] = 1. \\
Y_i = X_i\beta + \varepsilon_i & \leftrightarrow & Y_i^* = X_i\hat{\beta} + \varepsilon_i^*. \\
\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y & \leftrightarrow & \beta^* = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y^*. \\
R_n(\mathbb{X}, Y, \beta) & \leftrightarrow & R_n^* = R_n(\mathbb{X}, Y^*, \hat{\beta}).
\end{array}
$$

Estimation bootstrap sauvage de la loi de $R_n(\mathbb{X}, \Upsilon, \beta)$

La loi de $R_n(\mathbb{X}, \Upsilon, \beta)$ est estimée par la loi conditionnelle de $R_n^* = R_n(\mathbb{X}, \Upsilon^*, \hat{\beta})$ sachant $\mathbb{X}, \Upsilon$.

Approximation de Monte Carlo

Approximation de Monte Carlo

Les W_i peuvent être simulées par une méthode classique (inversion de la fonction de répartition, acceptation-rejet...).

La loi conditionnelle de R_n^* sachant $\mathbb{X} = x$ et $\Upsilon = y$ peut donc être approchée par une méthode de Monte Carlo.

$\hookrightarrow$ Intervalles de confiance, tests d'hypothèses, intervalles de prédiction bootstrap peuvent être construits en adaptant les définitions et algorithmes précédents.
$\hookrightarrow$ Méthode généralisable à des modèles de régression non paramétrique.

Régression linéaire multiple

Éléments de bibliographie

Ouvrage

- Fox, J. and Weisberg, S. (2011). An R Companion to Applied Regression + Appendix Bootstrapping Regression Models in R (2012).

Articles

- Freedman, D. A. (1981). Bootstrapping regression models. Ann. Statist.
- Härdle, W., and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. Ann. Statist.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensionnal linear models. Ann. Statist.
- Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Statist.

Plan

Estimation d'une erreur de prédiction

Leave-one-out bootstrap

$Z = (Z_1, \ldots, Z_n)$ un n-échantillon d'une loi P,
$Z_i = (X_i, Y_i)$, où les Y_i sont à valeurs dans $\mathbb{R}$ (régression réelle) ou dans $\{-1, 1\}$ (discrimination).

(X, Y) une nouvelle v.a. de loi P supposée indépendante de Z.

ϕ_Z une règle de prédiction de Y à partir de X, construite sur Z.

- $l(y, y') = (y - y')^2$ (perte quadratique) dans le cas de la régression.
- $l(y, y') = \mathbb{1}_{y \neq y'}$ dans le cas de la discrimination.

Erreur de prédiction moyenne : $\mathcal{E}[\phi] = \mathbb{E}_{Z \sim P^{\otimes n}} \mathbb{E}_{(X,Y) \sim P} \left[l\left(Y, \phi_Z(X)\right) \right]$.

Erreur apparente : $\mathcal{E}_n[\phi] = \frac{1}{n} \sum_{i=1}^{n} l\left(Y_i, \phi_Z(X_i)\right)$.

✘ Sous-estimation de l'erreur de prédiction moyenne.

Estimateur Leave-One-Out

$\hat{\mathcal{E}}[\phi]^{LOO} = \frac{1}{n}\sum_{i=1}^{n} l\left(Y_i, \phi_{Z_{(i)}}(X_i)\right)$, où :
$Z_{(i)}$ est l'échantillon Z privé de Z_i,
$\phi_{Z_{(i)}}$ est la règle de prédiction construite sur $Z_{(i)}$.
$\hookrightarrow$ Critère du PRESS en régression.

Estimateur Leave-One-Out bootstrap

$\mathcal{E}^*[\phi] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[l\left(Y_i, \phi_{Z^*_{(i)}}(X_i)\right)\middle| Z\right]$, où :
$Z^*_{(i)}$ est un échantillon bootstrap de taille n associé à $Z_{(i)}$,
$\phi_{Z^*_{(i)}}$ est la règle de prédiction construite sur $Z^*_{(i)}$.

$\hookrightarrow \mathcal{E}[\phi] = \mathbb{E}_{(X,Y)\sim P}\mathbb{E}_{Z\sim P^{\otimes n}}\left[l\left(Y, \phi_Z(X)\right)\right]$.
Plug-in pour $(X, Y) \sim P$: $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{Z\sim P^{\otimes n}}\left[l\left(Y_i, \phi_Z(X_i)\right)\right]$, puis pour chaque i, $Z \sim P^{\otimes n}$ remplacé par $\hat{P}_{(i)} = \frac{1}{n}\sum_{j\neq i}\delta_{Z_j}$.

✘ Sur-estimation de l'erreur de prédiction moyenne.

Approximation de Monte Carlo

Un échantillon bootstrap Z^* associé à Z ne contenant pas Z_i est aussi un échantillon bootstrap $Z^*_{(i)}$ de taille n associé à $Z_{(i)}$.

$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[l\left(Y_i,\phi_{Z^*_{(i)}}(X_i)\right)\middle|Z=z\right]$ peut donc être approchée par une méthode de Monte Carlo.

Algorithme Estimation LOO bootstrap

Variable

B : *entier assez grand*

Début

Pour b **variant de** 1 **à** B

Générer z^{*b} échantillon bootstrap associé à z

Calculer $I_i^b = 1$ si $z_i \notin z^{*b}$, 0 sinon.

FinPour

Retourner $\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{b=1}^{B} I_i^b l\left(Y_i,\phi_{z^{*b}}(X_i)\right)}{\sum_{b=1}^{B} I_i^b}$

Fin

Estimation d'une erreur de prédiction

Leave-one-out bootstrap corrigé

Le problème

Pour tout $j \neq i$, $P(Z_j \in Z^*_{(i)}) = 1 - \left(1 - \frac{1}{n}\right)^n \to_{n\to\infty} 1 - e^{-1} \simeq 0.632$, donc le support de l'échantillon bootstrap $Z^*_{(i)}$ se compose d'environ $0.632(n-1)$ éléments de $Z_{(i)}$ si n est grand.

Une solution : Leave-One-Out bootstrap 0.632 (Efron 1983)

$$\mathcal{E}^{*0.632}[\phi] = 0.632\mathcal{E}^*[\phi] + 0.368\mathcal{E}_n[\phi].$$

✘ Si $\mathcal{E}_n[\phi]$ est très petite, voire nulle (1-ppv en discrimination par ex.), correction trop forte.

Une autre solution : Leave-One-Out bootstrap $0.632+$ (Efron et Tibshirani 1997)

Soit $\hat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[l\left(Y_i, \phi_Z(X_j)\right)\right]$.

Taux de sur-apprentissage relatif :

$$\hat{R} = \begin{cases} \frac{\mathcal{E}^*[\phi]-\mathcal{E}_n[\phi]}{\hat{\gamma}-\mathcal{E}_n[\phi]} & \text{si } \mathcal{E}^*[\phi], \hat{\gamma} > \mathcal{E}_n[\phi] \\ 0 & \text{sinon.} \end{cases}$$

$$\mathcal{E}^{*0.632+}[\phi] = \mathcal{E}^{*0.632}[\phi] + \left(min(\mathcal{E}^*[\phi], \hat{\gamma}) - \mathcal{E}_n[\phi]\right)\frac{0.368 \times 0.632\hat{R}}{1 - 0.368\hat{R}}.$$

Estimation d'une erreur de prédiction

Éléments de bibliographie

Articles

- Efron, B. (1983). Estimating the Error Rate of a Prediction Rule : Improvement on Cross-Validation. JASA.
- Efron, B., and Tibshirani, R. (1997). Improvements on Cross-Validation : The .632+ Bootstrap Method. JASA.

Plan

Séries temporelles

Données q dépendantes : Block Bootstraps

Soit $\mathbb{X} = \{X_t,\ t = \ldots, -2, -1, 0, 1, 2, \ldots\}$ une série temporelle :

- Fortement stationnaire i.e. $\forall k, l$, $\{X_1, \ldots, X_k\} =^{(\mathcal{L})} \{X_{1+l}, \ldots, X_{k+l}\}$.
- q-dépendante i.e. pour tout t, $\{\ldots, X_{t-1}, X_t\}$ et $\{X_{t+q+1}, X_{t+q+2}, \ldots\}$ sont indépendantes, pour un q fixé.

Inconsistance du bootstrap "naïf" (Singh(1981)

Soit $\mathbb{X}_n*$ un n-échantillon bootstrap associé à $\mathbb{X}_n = (X_1, \ldots, X_n)$, $\mathbb{E}[X_i] = \mu$, $\mathbb{V}(X_i) = \sigma^2$.

- $\sqrt{n}\left(\overline{\mathbb{X}_n^*} - \overline{\mathbb{X}_n}\right) \rightsquigarrow \mathcal{N}(0, \sigma^2)$ p.s.
- $\sqrt{n}\left(\overline{\mathbb{X}_n} - \mu\right) \rightsquigarrow \mathcal{N}\left(0, \sigma^2 + \sum_{i=1}^{q-1} \mathrm{cov}(X_1, X_{1+i})\right)$.

✘ Inconsistance du bootstrap "naïf" si $\sum_{i=1}^{q-1} \mathrm{cov}(X_1, X_{1+i}) \neq 0$.
↪ Les éléments de l'échantillon bootstrap, cond. à $\mathbb{X}_n$, sont indépendants...

✔ Idée : bootstrapper en gardant au maximum la structure de l'échantillon d'origine.

Block bootstrap

Soit $k_n \to \infty$ et $l_n \to \infty$, $n = k_n l_n$.

Pour $j = 1 \dots k_n$,
$\mathcal{B}_j = \{X_{(j-1)l_n+1}, X_{(j-1)l_n+2}, \dots, X_{jl_n}\}$ (bloc de taille l_n)

Estimation BB de la loi d'une racine

- Soit $\mathcal{B}_1^* := (X_1^*, \dots, X_{l_n}^*), \dots, \mathcal{B}_B^* := (X_{(B-1)l_n+1}^*, \dots, X_{Bl_n}^*)$ B blocs tirés au hasard avec remise dans $\{\mathcal{B}_1, \dots, \mathcal{B}_{k_n}\}$.
- Soit $\mathbb{X}^* = (X_1^*, \dots, X_{l_n}^*, \dots, X_{(B-1)l_n+1}^*, \dots, X_{Bl_n}^*)$, composé des $m = Bl_n$ élements de $\mathcal{B}_1^*, \dots, \mathcal{B}_B^*$ mis "bout à bout".
- Soit P_m^* la mesure empirique associée à $\mathbb{X}^*$.

La loi de $R_n(\mathbb{X}_n, P_n)$ est estimée par la loi de $R_n(\mathbb{X}^*, P_m^*)|\mathbb{X}$.

✘ Peu de blocs disponibles.
✘ Problème du choix de l_n

Moving block bootstrap

Soit $l_n \to \infty$ avec $l_n/n \to 0$.

Pour $j = 1 \dots n - l_n + 1$, $\mathcal{B}_j = \{X_j, \dots, X_{j+l_n-1}\}$ (bloc de taille l_n).

Estimation MBB de la loi d'une racine

- Soit $\mathcal{B}_1^* := (X_1^*, \dots, X_{l_n}^*), \dots, \mathcal{B}_B^* := (X_{(B-1)l_n+1}^*, \dots, X_{Bl_n}^*)$ B blocs tirés au hasard avec remise dans $\{\mathcal{B}_1, \dots, \mathcal{B}_{n-l_n+1}\}$.
- Soit $\mathbb{X}^* = (X_1^*, \dots, X_{l_n}^*, \dots, X_{(B-1)l_n+1}^*, \dots, X_{Bl_n}^*)$, composé des $m = Bl_n$ élements de $\mathcal{B}_1^*, \dots, \mathcal{B}_B^*$ mis "bout à bout".
- Soit P_m^* la mesure empirique associée à $\mathbb{X}^*$.

La loi de $R_n(\mathbb{X}_n, P_n)$ est estimée par la loi de $R_n(\mathbb{X}^*, P_m^*)|\mathbb{X}$.

✔ Plus de blocs disponibles !
✘ Mais toujours le problème du choix de l_n, et des problèmes de bord...
↪ Circular Block Bootstrap, Stationary Block Bootstrap.
↪ Choix de l_n par sous-échantillonnage, ou par plug-in.

Séries temporelles AR(p) stationnaires

$X_t = \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} + \varepsilon_t$, où

- $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ est une suite de v.a. i.i.d. de loi P centrée,
- $\beta = (\beta_1, \ldots, \beta_p)'$ vecteur de paramètres inconnus, tels que $\beta(z) := 1 - \sum_{j=1}^{p} \beta_j z^j \neq 0$ pour tout $z \in \mathbb{C}$, $|z| < 1$.

On dispose d'une observation x_n de $\mathbb{X}_n = (X_1, \ldots, X_n)$.

Estimateur des moindres carrés ordinaires de β :
$\hat{\beta} = (V_n' V_n)^{-1} V_n' (X_{p+1}, \ldots, X_n)'$, où V_n est une matrice $(n-p) \times p$ dont la $i$ème ligne est $(X_{i+p-1}, \ldots, X_i)$.

Résidus : $\hat{\varepsilon}_t = X_t - \hat{\beta}_1 X_{t-1} + \ldots - \hat{\beta}_p X_{t-p}$, $t = p+1, \ldots, n$.

Résidus recentrés : $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \overline{\hat{\varepsilon}}$, où $\overline{\hat{\varepsilon}} = \frac{1}{n-p} \sum_{t=p+1}^{n} \hat{\varepsilon}_t$.

Le bootstrap des résidus se résume selon le schéma suivant.

β	$\leftrightarrow$	$\hat{\beta}$
P	$\leftrightarrow$	$\tilde{P}_n = \frac{1}{n-p}\sum_{t=p+1}^{n} \delta_{\tilde{\varepsilon}_t}$
ε_t i.i.d. de loi P	$\leftrightarrow$	ε_t^* i.i.d. de loi $\tilde{P}_n$ conditionnellement à $\mathbb{X}$, échantillon bootstrap associé au vecteur des résidus recentrés.
$\mathbb{X}$	$\leftrightarrow$	$\mathbb{X}^*$ solution stationnaire de $X_t^* = \hat{\beta}_1 X_{t-1}^* + \ldots + \beta_p X_{t-p}^* + \varepsilon_t^*$.
$\hat{\beta}$	$\leftrightarrow$	$\beta^* = (V_n^{*\prime} V_n^*)^{-1} V_n^{*\prime} (X_{p+1}^*, \ldots, X_n^*)'$.
$R_n(\mathbb{X}_n, \beta)$	$\leftrightarrow$	$R_n^* = R_n((X_1^*, \ldots, X_n^*), \hat{\beta})$

Estimation bootstrap des résidus de la loi d'une racine

La loi d'une racine $R_n(\mathbb{X}_n, P_n)$ est estimée par la loi conditionnelle de $R_n^* = R_n((X_1^*, \ldots, X_n^*), \hat{\beta})$ sachant $\mathbb{X}$.

En pratique, on pose $(x_{1-p}^* = x_{1-p}, \ldots, x_0^* = x_0)$ et on génère $(x_1^*, \ldots, x_n^*)$ de façon récursive.

$\hookrightarrow$ Algorithmes pour l'estimation de biais, variance, MSE, pour la construction d'IC, de tests d'hypothèses, d'intervalles de prédiction.

Séries temporelles ARMA (p, q) stationnaires

$X_t = \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \ldots + \alpha_q \varepsilon_{t-q}$, où

- $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ est une suite de v.a. i.i.d. de loi P centrée,
- $\beta = (\beta_1, \ldots, \beta_p)'$ et $\alpha = (\alpha_1, \ldots, \alpha_q)'$ paramètres inconnus,
  $\beta(z) := 1 - \sum_{j=1}^{p} \beta_j z^j \neq 0$ pour tout $z \in \mathbb{C}$, $|z| < 1$,
  $\alpha(z) := 1 - \sum_{j=1}^{q} \alpha_j z^j \neq 0$ pour tout $z \in \mathbb{C}$, $|z| < 1$,
  $\beta_p \neq 0$, $\alpha_q \neq 0$, $\alpha(z)$ et $\beta(z)$ n'ont pas de racine commune.

Il existe $\rho > 1$ tel que pour tout z, $|z| \leq \rho$:
$[\beta(z)]^{-1} = \sum_{j=0}^{\infty} b_j z^j$, $[\alpha(z)]^{-1} = \sum_{j=0}^{\infty} a_j z^j$,
$[\beta(z)]^{-1}\alpha(z) = sum_{j=0}^{\infty} c_j z^j = \left(\sum_{j=0}^{\infty} d_j z^j\right)$.

$\hookrightarrow \mathbb{X}$ peut être vue comme une série AR(∞) :

- $X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$,
- $\varepsilon_t = \sum_{j=0}^{\infty} d_j X_{t-j}$,
- $\varepsilon_t = \sum_{j=0}^{\infty} a_j \left(X_{t-j} - \beta_1 X_{t-j-1} - \ldots - \beta_p X_{t-j-p}\right)$.

On dispose d'une observation x_{n+p} de $\mathbb{X}_{n+p} = (X_{1-p}, \ldots, X_n)$.

Estimateurs de β et α :
$\hat{\beta}$ et $\hat{\alpha}$ tels que $\sum_{j=1}^{p} |\hat{\beta}_j - \beta_j| + \sum_{j=1}^{q} |\hat{\alpha}_j - \alpha_j| \to 0$ en proba.

Il existe $1 < \rho_0 < \rho$ t.q. si $\tilde{\alpha}(z) := 1 - \sum_{j=1}^{q} \hat{\alpha}_j z^j$,
$[\hat{\alpha}(z)]^{-1} = \sum_{j=0}^{\infty} \hat{a}_j z^j$, $|z| \leq \rho_0$ pour de grandes valeurs de n, avec grande proba.

Résidus : $\hat{\varepsilon}_t = \sum_{j=1}^{t} \hat{a}_{j-1} \left(X_{t+1-j} - \sum_{k=1}^{p} \hat{\beta}_k X_{t+1-j-k} \right)$, $t = 1, \ldots, n$.

Résidus recentrés : $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \overline{\hat{\varepsilon}}$, où $\overline{\hat{\varepsilon}} = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_t$.

Le bootstrap des résidus se résume selon le schéma suivant.

β, α	$\leftrightarrow$	$\hat{\beta}, \hat{\alpha}$
P	$\leftrightarrow$	$\tilde{P}_n = \frac{1}{n}\sum_{t=1}^{n} \delta_{\tilde{\varepsilon}_t}$
ε_t i.i.d. de loi P	$\leftrightarrow$	ε_t^* i.i.d. de loi $\tilde{P}_n$ conditionnellement à $\mathbb{X}$, échantillon bootstrap associé au vecteur des résidus recentrés.
$\mathbb{X}$	$\leftrightarrow$	$\mathbb{X}^*$ solution stationnaire de $X_t^* = \sum_{j=1}^{p} \hat{\beta}_j X_{t-1}^* + \sum_{j=1}^{q} \hat{\alpha}_j \varepsilon_{t-j}^* + \varepsilon_t^*$.
$\hat{\beta}, \hat{\alpha}$	$\leftrightarrow$	β^*, α^* estimateurs calculés sur $\mathbb{X}^*$.
$R_n(\mathbb{X}_{n+p}, \beta, \alpha)$	$\leftrightarrow$	$R_n^* = R_n((X_{1-p}^*, \ldots, X_n^*), \hat{\beta}, \hat{\alpha})$

Estimation bootstrap des résidus de la loi d'une racine

La loi d'une racine $R_n(\mathbb{X}_{n+p}, \beta, \alpha)$ est estimée par la loi conditionnelle de R_n^* sachant $\mathbb{X}$.

En pratique, on pose $x_t^* = \varepsilon_t^* = 0$ pour tout $t \leq -\max(p, q)$ et on génère les suivants de façon récursive.

$\hookrightarrow$ Algorithmes pour l'estimation de biais, variance, MSE, pour la construction d'IC, de tests d'hypothèses, d'intervalles de prédiction.

- Lahiri, S. N. (2003) Bootstrap for dependent data.

Articles

- Allen, M. and Datta, S. (1999). A note on bootstrapping M-estimators in ARMA models.
- Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions.
- Bose, A. (1990). Bootstrap in moving average models.
- Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series.
- Künsch, H. R. (1989). The Jackknife and the bootstrap for general stationary observations.
- Kreiss, J. P., and Franke, J. (1992). Bootstrapping stationary autoregressive moving-average models.
- Liu, R. Y., and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence.