

Bootstrap pour les séries temporelles

Notations

On va considérer des $(X_t)_{t \in \mathbb{Z}}$ (faiblement) stationnaires.

- ▶ On définit $\mu = \mathbb{E}(X_t)$
- ▶ la fonction d'autocovariance $\gamma_X(h) = \text{Cov}(X_s, X_{s+h})$
- ▶ la fonction d'autocorrelation

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

ne dépendant pas de s

Estimation

A partir des observations X_1, \dots, X_n (de la série stationnaire X), on peut calculer

- ▶ la **moyenne** $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$ *estimation de μ .*
- ▶ la **fonction d'autocovariance** empirique

$$\hat{\gamma}_X(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X})(X_t - \bar{X}) \quad \forall 0 < h < n$$

- ▶ la **fonction d'autocorrélation** empirique

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)} \quad \leftarrow$$

Théorème

on a un TCL

Sous des conditions générales, si X est un bruit blanc, alors pour n assez grand, l'ACF empirique, $\hat{\rho}_X(h)$, for $h = 1, 2, \dots, H$, où H est fixé mais arbitraire, est approximativement de loi gaussienne de moyenne nulle et d'écart-type

$$\sigma_{\hat{\rho}_X(h)} = \frac{1}{\sqrt{n}} \quad \leftarrow$$

Modèle ARMA(p, q)

Un processus ARMA(p, q) X est stationnaire et défini via

$$\Phi(B)X_t = \Theta(B)\omega_t$$

écriture identifiable / la + simple

où $\omega \sim WN(0, \sigma^2)$, Φ est un polynôme d'ordre p , Θ un polynôme d'ordre q et Φ et Θ n'ont pas de zéro commun. *de l'ARMA*

- ▶ Ce processus ARMA(p, q) est causal ssi

$$\Phi(z) = 0 \Leftrightarrow |z| > 1.$$

- ▶ Il est inversible ssi les racines $\Theta(z)$ sont hors du cercle unité.

Representations causale et inversible

Considérons un processus ARMA causal, inversible défini par $\Phi(B)X_t = \Theta(B)\omega_t$. Il peut être réécrit comme

- ▶ un MA(∞) (décomposition de Wold):

$$X_t = \frac{\Theta(B)}{\Phi(B)}\omega_t = \psi(B)\omega_t = \sum_{k \geq 0} \psi_k \omega_{t-k}$$

- ▶ ou un AR(∞)

$$\omega_t = \frac{\Phi(B)}{\Theta(B)}X_t = \pi(B)X_t = \sum_{k \geq 0} \pi_k X_{t-k} \quad \pi_0 = 1$$

Remarque : on écrit un processus AR(p) comme $X_t = \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t$

$$\omega_t = \sum_{k=0}^p \pi_k X_{t-k}$$
$$\Rightarrow \omega_t = \pi_0 X_t + \sum_{k=1}^p \pi_k X_{t-k}$$

remarque
centrale
pour le bootstrap

$$\phi_0^* = -\frac{\pi_j}{\pi_0}$$

hypothèse très forte (pas forcément nécessaire)

mais on considère le cas le + simple

Equation de Yule-Walker pour un AR(p)

On considère les AR(p) (sans perte de généralité pour la suite)

La fonction d'autocovariance et les paramètres du modèle vérifient

$$\Gamma_p \phi^* = \gamma_p \quad \text{et} \quad \sigma^{2,*} = \gamma(0) - (\phi^*)^T \gamma_p$$

↑
coeff de l'AR(p)

où


$$\Gamma_p = (\gamma(k-j))_{1 \leq j, k \leq p}, \quad \phi^* = (\phi_1^*, \dots, \phi_p^*)^T, \quad \text{et} \quad \gamma_p = (\gamma(1), \dots, \gamma(p))^T.$$

En notant $\hat{\gamma}_p = (\hat{\gamma}(1), \dots, \hat{\gamma}(p))^T$ les autocovariances empiriques, on obtient

$$\begin{bmatrix} \hat{\phi} \\ \hat{\sigma}^2 \end{bmatrix} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p \quad \text{and} \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}^T \hat{\gamma}_p.$$

Si WNW (0,0,2) alors la loi de $\hat{\phi}$ est inconnue

Si le bruit est gaussien, on peut définir les estimateurs au maximum de vraisemblance.

On sait estimer la fonction d'auto covariance donc avec Yule Walker on sait estimer les coefficients de l'AR(p).  on est à p fixé, ou p est inconnu → à estimer

Loi asymptotique des estimateurs

Sous des conditions faibles sur ω , et si le $AR(p)$ est causal, l'estimateur de Yule-Walker vérifie

$$\sqrt{n}(\hat{\phi} - \phi^*) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^{2,*} \Gamma_p^{-1})$$
$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^{2,*}.$$

TCL à la
vitesse \sqrt{n} pour
 $\hat{\phi} \rightarrow$ ou est

Les estimateurs au maximum de vraisemblance dans ce modèle ont la même loi asymptotique.

"raffermis" sur
le bootstrap.

Prédiction linéaire

$$\text{AR}(p): X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \omega_t$$

Meilleur prédicteur linéaire pour un AR(p)

Si X est un processus $\text{AR}(p)$ causal alors le meilleur prédicteur linéaire de X_t basé sur $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ est

$$X_t^{(p)} = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}$$

L'erreur de prédiction est alors donnée par

$$e_t^{(p)} = \underbrace{X_t^{(p)}} - \underbrace{X_t} = \omega_t$$

Autoregressive-sieve bootstrap in Kreiss 1992

$(X_t) : \text{ARMA}(p, q)$
 $p, q, \Phi, \Theta, \sigma^2$ inconnus

On approxime la représentation $\text{AR}(\infty)$ par une représentation $\text{AR}(p_n)$ (avec $p_n \rightarrow \infty$ quand $n \rightarrow \infty$), puis on ré-échantillonne suivant la procédure suivante

→ **Bootstrap paramétrique**

1. Estimer l'ordre \hat{p}_n et les coefficients associés $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{\hat{p}_n,n}$ (par la méthode de Yule-Walker ou le maximum de vraisemblance). hypothèse très très forte
2. Définir les résidus

valeur prédite dans le modèle AR estimé au point 1.

$$\tilde{e}_t^{(\hat{p}_n)} = \hat{X}_t^{(\hat{p}_n)} - X_t = \sum_{j=1}^{\hat{p}_n} \hat{\phi}_{j,n} X_{t-j} - X_t \text{ pour } t = \hat{p}_n + 1, \dots, n$$

↑
observation

les centrer en définissant

$$\hat{e}_t^{(\hat{p}_n)} = \tilde{e}_t^{(\hat{p}_n)} - \frac{1}{n - \hat{p}_n} \sum_{t=\hat{p}_n+1}^n \tilde{e}_t^{(\hat{p}_n)}$$

3. Pour $b = 1, \dots, B$, reconstruire

$$X_{b, \hat{p}_n+1}^{S, \star}, \dots, X_{b, n}^{S, \star}$$

← obs d'une AR (\hat{p}_n) .

plus compliqué
 de la même manière suit un \hat{p}_n (très proche du bootstrap en resampling)
 comme un $\text{AR}(\hat{p}_n)$ avec les coefficients estimés $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{\hat{p}_n,n}$ et un tirage aléatoire avec remise dans les résidus $(\hat{e}_t^{(\hat{p}_n)})_{t=\hat{p}_n+1, \dots, n}$ → bootstrap

Bootstrap paramétrique

Cas 1 X_1, \dots, X_n iid de loi $\mathcal{E}(\lambda)$

→ on peut faire un bootstrap non-paramétrique comme décrit au début de cours.

→ autre solution à appliquer si on est sûr du modèle (\mathcal{E}).

① on estime λ par $\hat{\lambda} = (X_1, \dots, X_n)$

② ré-échantillonnage: on simule b tirant la loi $\mathcal{E}(\hat{\lambda})$.

$X_b^* = (X_{b,1}^*, \dots, X_{b,n}^*) \quad b = 1, \dots, B.$

③ on a $\hat{\lambda}_b^*$ à partir de X_b^* . --
le reste est comme au début de cours.

Cas 2 $(Y_1, X_1) \dots (Y_n, X_n)$ iid.

avec $Y_i = \beta_0 + X_i \beta + \varepsilon_i$ avec $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$
(modèle linéaire gaussien).

$\Leftrightarrow Y_i \sim \mathcal{N}(\beta_0 + X_i \beta, \sigma^2)$

① on estime β_0, β, σ^2 à partir de $(Y_1, X_1) \dots (Y_n, X_n)$
→ $\hat{\beta}_0, \hat{\beta}, \hat{\sigma}^2$.

② ré-échantillonnage.

pour un X_i on simule $Y_{b,i}^*$ suivant
la $\mathcal{N}(\hat{\beta}_0 + X_i \hat{\beta}, \hat{\sigma}^2)$ // conditionnellement
à X_1, \dots, X_n
ou à X .

$$\left(\hat{\phi}_{t, t}^{(p_n)} \right)_{t=1, \dots, n}$$

résidus estimés et centrés.

$\hat{\phi}_{1, n} - \hat{\phi}_{p_n, n}$ & calculés à partir des X_1, \dots, X_n observations de départ.

Bootstrap dans l'AR (\hat{p}_n) estimé: ou ajout de la lia

$$\hat{X}_{b, t}^* = \sum_{j=1}^{\hat{p}_n} \hat{\phi}_{j, n} \hat{X}_{b, t-j}^* + \hat{\epsilon}_{b, t}^*$$

tiré aléatoirement et avec remise dans $(\hat{\epsilon}_t^{(p_n)})$

$t = p+1, \dots, n$

$$\Delta \quad t-j < p+1$$

$$X_{b, t-j}^* = X_{t-j} \quad \text{pour } t-j \geq p+1$$

$$\hat{X}_{b, t-j}^* = X_{t-j}$$

$X_1, \dots, X_{\hat{p}_n}, X_{\hat{p}_n+1}, \dots, X_n$

on peut calculer le prédicteur linéaire

$$\hat{X}_t = \sum_{j=1}^{\hat{p}_n} \hat{\phi}_{j, n} X_{t-j} \quad t \geq \hat{p}_n + 1$$

Block bootstrap, in Kunsch 1989; Politis and Romano 1994

l : déterministe \nearrow l : aléatoire.

0 bo: X_{1-1}, X_n

1. On commence par "circulariser" la série en posant

$$X_{n+1} = X_1, X_{n+2} = X_2, \dots$$

bootstrap
non-paramétrique

2. On choisit un entier $1 < l < n$, qui correspond au nombre de blocs, et on pose $N = \lfloor n/l \rfloor$, c'est la taille des blocs.

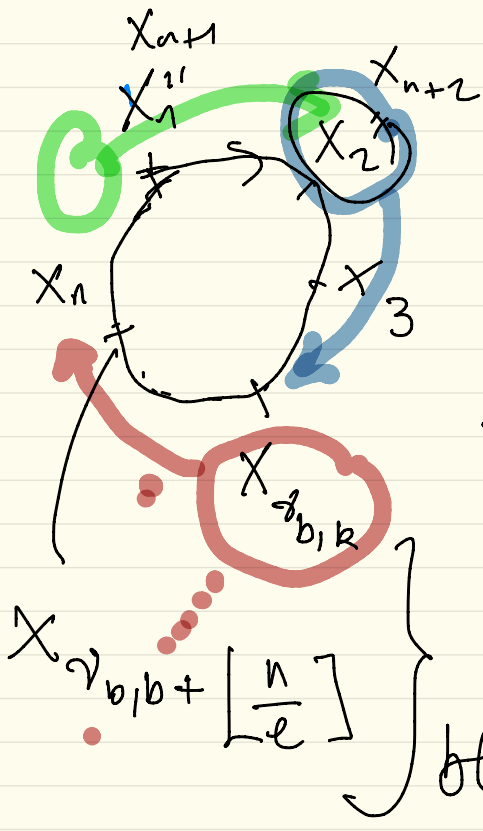
3. Pour $b = 1, \dots, B$

3.1 on tire aléatoirement et avec remise N entiers $\nu_{b,k}$ ($k = 1, \dots, N$) entre 1 et n

3.2 on crée les blocs $X_{\nu_{b,k}}, X_{\nu_{b,k}+1}, \dots, X_{\nu_{b,k}+l-1}$ de longueur l

3.3 on les concatène pour créer la série bootstrapée

$$(X_{b,1}^{B,*}, \dots, X_{b,n}^{B,*}) = (X_{\nu_{b,1}}, X_{\nu_{b,1}+1}, \dots, X_{\nu_{b,1}+l-1}, \dots, X_{\nu_{b,N}}, X_{\nu_{b,N}+1}, \dots, X_{\nu_{b,N}+l-1}).$$



on obtient
une série infinie.

l : longueur de
bloc.

$\gamma_{b, k} \leftarrow$ on tire
aléatoire avec
uniforme dans
 $[1, n]$.

bloc de taille $\lfloor \frac{n}{l} \rfloor$

bloc 1 **bloc 2** **bloc 3**

$$\left(X_{v_{1,1}, 1} \dots X_{v_{1,1} + \lfloor \frac{n}{l} \rfloor}, X_{v_{2,1}, 1} \dots X_{v_{2,1} + \lfloor \frac{n}{l} \rfloor}, \dots \right)$$

$$= (X_{1,1}^*, \dots, X_{1,n}^*)$$

il faut concaténer N blocs.

le choix de l va être crucial :

	$l=1$ ($N=n$)	taille
	\vdots	$\lfloor \frac{n}{l} \rfloor = n$
	$l=n$ ($N=1$)	\vdots
nombre de blocs.		$\lfloor \frac{n}{l} \rfloor = 1$

$l=1$: on observe dans tous les échantillons bootstrap la série de départ

par ex. $(x_4, x_5, \dots, x_n, x_1, x_2, x_3)$

$l=n$: cela correspond au bootstrap iid.

Si on imagine qu'on veut estimer $\gamma(2)$.

$\rightarrow l=n$ on a aussi la dépendance temporelle \rightarrow ça ne marche pas.

$\rightarrow l=1$ on a aussi la dépendance entre x_n et x_1 .

Autres applications

En général, pour faire du bootstrap à partir d'observation non i.i.d., il convient de faire très attention à la construction des échantillons bootstrappés (on l'a vu en régression et en séries temporelles). C'est le cas pour les processus markoviens, pour les données censurées et plus généralement pour les processus de comptage.

Une référence possible (il y en a beaucoup d'autres) Davison and Hinkley 1997

References I



Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*. Vol. 1. Cambridge university press, 1997.



Bradley Efron. *The jackknife, the bootstrap and other resampling plans*. SIAM, 1982.



Bradley Efron. “Bootstrap methods: another look at the jackknife”. In: *Breakthroughs in Statistics*. Springer, 1992, pp. 569–593.



Peter Hall. *The bootstrap and Edgeworth expansion*. Springer Science & Business Media, 2013.



Jens-Peter Kreiss. “Bootstrap procedures for AR (∞)—processes”. In: *Bootstrapping and Related Techniques*. Springer, 1992, pp. 107–113.



Hans R Kunsch. “The jackknife and the bootstrap for general stationary observations”. In: *The annals of Statistics* (1989), pp. 1217–1241.

References II



Dimitris N Politis and Joseph P Romano. "The stationary bootstrap".
In: *Journal of the American Statistical association* 89.428 (1994),
pp. 1303–1313.



Dimitris N Politis, Joseph P Romano, and Michael Wolf.
Subsampling Springer series in statistics. 1999.



Maurice H Quenouille. "Notes on bias in estimation". In: *Biometrika*
43.3/4 (1956), pp. 353–360.



John W Tukey. "Bias and confidence in not-quite large samples". In:
Annals of Mathematical Statistics. Vol. 29. 2. INST
MATHEMATICAL STATISTICS IMS BUSINESS OFFICE-SUITE 7,
3401 INVESTMENT BLVD, HAYWARD, CA 94545. 1958,
pp. 614–614.