

Une méthode de rééchantillonnage statistique pour tester la cyclostationnarité à l'ordre deux

Sofiane MAIZ¹, Mohamed EL BADAOU¹, Jacek LEŚKOW², Christine SERVIERE³

¹Université de Lyon, Université de Saint Etienne, Jean Monnet, F-42000 Saint Etienne LASPI, F-42334, IUT de Roanne
20 Avenue de Paris, 42334 Roanne Cedex, France

²Cracow University of Technology, Institute of Mathematics 24 Warszawska, 31-155 Krakow, Poland

³Université de Grenoble, Institut polytechnique de Grenoble CNRS:UMR5216, GIPSA-lab, Grenoble, France
sofiane.maiz@univ-st-etienne.fr, mohamed.elbadaoui@univ-st-etienne.fr,
leskow@wsb-nlu.edu.pl, christine.serviere@gipsa-lab.grenoble-inp.fr,

Résumé – Ce papier présente une méthode de rééchantillonnage statistique, appelée *Subsampling*, pour l'estimation des coefficients de Fourier de la fonction d'autocovariance des processus cyclostationnaires d'ordre 2 (CS2). L'exploitation de la convergence de la distribution de la version du subsampling à la distribution asymptotique de l'estimateur initial a permis de simplifier considérablement l'inférence statistique pour les processus CS2. Cela a conduit à la construction de test statistique avec des intervalles de confiance valides pour l'estimateur étudié. L'applicabilité de cette méthode sur un modèle de simulation ainsi que sur des données biomécaniques a également été prouvée. De plus, l'estimation de la CS2 basée sur le *Subsampling* a permis de réaliser une description complète et innovante des différents états de fatigue d'un coureur.

Abstract – This paper presents a statistical resampling method for the estimation of the Fourier coefficients of the autocovariance function in second order cyclostationary processes (CS2). This method is referred to as *Subsampling*. The convergence of the distribution of the subsampling version to the asymptotic distribution of the initial estimator allowed the construction of valid confidence intervals, together with a statistical test for the presence of CS2. The applicability of this procedure to CS2 simulated and real data is proved. Moreover, the subsampling-based second order CS analysis allowed to make a full and an innovative description and characterization of different fatigue states of a runner.

1 Introduction

Les séries temporelles et les processus stochastiques modélisant les phénomènes périodiques et quasi-périodiques sont bien connus et ont été bien étudiés dans les dernières décennies. Les signaux et séries temporelles générés par de tels modèles reflètent des périodicités dans leurs moments d'ordre un et d'ordres supérieurs. Une classe de ces modèles est celle des modèles cyclostationnaires (CS) [1, 2]. Beaucoup de recherches se sont penchées sur la manière d'estimer l'autocovariance périodique des processus CS, et à quel genre de résultat asymptotique on pourrait obtenir. Des recherches récentes [2, 3, 4, 8] ont montré que ces estimateurs sont asymptotiquement normal. Cependant, un problème majeur apparaît avec la matrice de covariance de la distribution asymptotique. En effet, cette covariance dépend des moments d'ordre quatre des données étudiées, et est très difficile à calculer [4]. Par conséquent, l'inférence pour les tests d'hypothèses ainsi que la détermination des intervalles de confiance sont très difficiles à construire.

De nombreux travaux de recherche liés aux procédures de bootstrap et de rééchantillonnage statistique ont montré la consistance ainsi que la possible application de ces méthodes pour de tels modèles non-stationnaires, en particulier, pour les modèles cyclostationnaires [4, 3, 8]. Dans ce papier, nous nous focali-

sons sur la consistance d'une méthode de sous-échantillonnage statistique ou *subsampling*, pour l'estimateur des paramètres $|a(\lambda, \tau)|$ d'une série temporelle $\{X(t) : t \in \mathbb{Z}\}$. Les nombres complexes $a(\lambda, \tau)$ représentent les coefficients de Fourier de la fonction d'autocovariance $B(t, \tau) = \sum a(\lambda, \tau)e^{i2\pi\lambda t}$. L'objectif de ce travail est donc de construire des intervalles de confiance valides pour l'estimateur étudié, afin de tester si le processus considéré est CS2 ou pas.

La section 2 de ce papier contient les définitions basiques et nécessaires, ainsi que la formulation du problème. Le résultat fondamental de la consistance du subsampling pour l'estimateur $|a(\lambda, \tau)|$, de même que l'algorithme de la méthode sont présentés dans la section 3. La section 4 traite de l'application et de la validation de la procédure de subsampling sur des données de simulation. L'applicabilité de cette méthode sur des signaux réels issus de la biomécanique à des fins de détection et de caractérisation de la fatigue d'un coureur professionnel est démontrée dans la section 5. La section 6 conclut le papier.

2 Formulation du problème

Une série temporelle $\{X(t) : t \in \mathbb{Z}\}$ est dite CS2 si ses moments d'ordre un et deux sont périodiques. Plus exactement,

si sa moyenne $\mu_X(t) = E(X(t))$ et sa fonction d'autocovariance $B_X(t, \tau) = \text{cov}(X(t), X(t + \tau))$ sont périodiques dans le temps, de période T : $\mu_X(t) = \mu_X(t + T)$ et $B_X(t, \tau) = B_X(t + T, \tau)$.

Dans ce papier, l'attention est portée sur les propriétés CS à l'ordre 2, i.e. sur la fonction d'autocovariance $B_X(t, \tau)$. Pour cela, il est supposé que $\mu_X(t)$ est nulle. De plus, la représentation de Fourier de la fonction d'autocovariance est utilisée:

$$B_X(t, \tau) = \sum_{\lambda \in \Lambda_T} a(\lambda, \tau) e^{i2\pi\lambda t} \quad (1)$$

Avec

$$a(\lambda, \tau) = \frac{1}{T} \sum_{t=1}^T B_X(t, \tau) e^{-i2\pi\lambda t}, \quad (2)$$

Et l'ensemble $\Lambda_T = \{\lambda : a(\lambda, \tau) \neq 0\}$ est supposé fini.

Cette représentation est très utile pour résoudre le problème d'estimation de $B_X(t, \tau)$ via l'estimation de ses coefficients de Fourier complexes $a(\lambda, \tau)$. En pratique, la période T de l'autocovariance $B_X(t, \tau)$ n'est pas connue. Par conséquent, les éléments de l'ensemble Λ_T ne sont pas facilement obtenus. C'est pourquoi, nous montrons dans ce papier l'intérêt du subsampling pour l'estimation et l'identification des fréquences cycliques $\lambda \in \Lambda_T$ du processus étudié.

Soit l'échantillon disponible $\{X(1), X(2), \dots, X(n)\}$ de la série temporelle $\{X(t) : t \in \mathbb{Z}\}$. Alors, l'estimateur naturel du paramètre $a(\lambda, \tau)$ est défini comme suit:

$$\hat{a}_n(\lambda, \tau) = \frac{1}{n - |\tau|} \sum_{1 - \min\{\tau, 0\}}^{n - \max\{\tau, 0\}} X(t + \tau) X(t) e^{-i2\pi\lambda t} \quad (3)$$

$\hat{a}_n(\lambda, \tau)$ est appelé l'estimateur de l'autocovariance (auto-corrélation) cyclique. Il combine l'information sur le retard, par son argument τ et sur la fréquence cyclique, par son argument λ . Il est très important de rappeler que l'analyse statistique de cet estimateur est fondamentale pour l'identification du comportement CS2 du processus sous étude. Cela signifie que si pour toutes les fréquences λ non nulles, le paramètre $\hat{a}_n(\lambda, \tau)$ n'est pas significativement différent de zéro, alors on pourrait suggérer que $X(t)$ n'est pas CS2. Cependant, si le paramètre $\hat{a}_n(\lambda, \tau)$ est significativement différent de zéro pour des fréquences λ non nulles, alors $X(t)$ est CS2.

L'idée principale du subsampling a été introduite dans [5] et est brièvement décrite dans ce qui suit. Soit une série temporelle $\{X(t) : t \in \mathbb{Z}\}$ gouvernée par une loi de probabilité quelconque P appartenant à une certaine classe de mesures de probabilité \mathcal{P} . L'objectif ici est de construire des intervalles de confiance pour le paramètre $\theta(P) \in \{\theta(P) : P \in \mathcal{P}\}$.

Soit $\hat{\theta}_n = \hat{\theta}_n(X(1), X(2), \dots, X(n))$ l'estimateur de $\theta(P)$, basé sur l'échantillon $\{X(1), X(2), \dots, X(n)\}$ de la série temporelle $\{X(t) : t \in \mathbb{Z}\}$. Pour toute séquence b tendant à l'infini avec n , telle que $b < n$, et pour tout $t \in \{1, \dots, n - b + 1\}$, on définit:

$$\hat{\theta}_{n,b,t} = \hat{\theta}_b(X(t), \dots, X(t + b - 1)) \quad (4)$$

l'estimateur du paramètre $\theta(P)$ basé sur le sous-échantillon $\{X(t), X(t + 1), \dots, X(t + b - 1)\}$. Le paramètre t peut être interprété comme étant le point de départ pour le re-calcule de $\hat{\theta}_n$ sur une longueur de bloc b . Ensuite, il devient simple de dériver les quantiles à partir de la distribution empirique des répliques de l'estimateur issues de la procédure du subsampling, afin de remplacer les quantiles de la distribution initiale.

3 Estimation des coefficients de Fourier via le subsampling

L'estimateur $\hat{a}_n(\lambda, \tau)$ pour chaque entier τ et pour chaque réel $\lambda \in \Lambda_T$ est basé sur l'échantillon $\{X(1), \dots, X(n)\}$ provenant de la série temporelle $\{X(t) : t \in \mathbb{Z}\}$. Cet estimateur complexe est traité dans deux dimensions, avec des coordonnées égales à sa partie imaginaire et à sa partie réelle.

Soit b une longueur de bloc telle que $b < n$, et soit $t \in \{1, \dots, n - b + 1\}$. Alors, on définit l'estimateur

$$\hat{a}_{n,b,t}(\lambda, \tau) = \frac{1}{b} \sum_{j=t}^{t+b} X(j) X(j + \tau) e^{-i2\pi\lambda j} \quad (5)$$

L'égalité (5) donne l'estimateur de $a(\lambda, \tau)$ basé sur les sous-échantillons $(X(t), \dots, X(t + b - 1))$, avec t pouvant être interprété comme le point de départ du nouveau calcul de $\hat{a}_n(\lambda, \tau)$ sur un bloc de longueur b . Il est à noter que cet estimateur est différent de celui proposé dans [4]. Ce nouvel estimateur, contrairement à celui dans [4], prend en compte les indices temps de chaque sous-échantillon, afin de préserver les structures temporelles de chaque sous-échantillon utilisé lors de la procédure du subsampling. En outre, la distribution empirique du subsampling est donnée par:

$$L_{n,b}(x) = \frac{1}{n - b + 1} \sum_{t=1}^{n-b+1} \mathbf{1}\{\sqrt{b}(|\hat{a}_{n,b,t}(\lambda, \tau)| - |\hat{a}_n(\lambda, \tau)|) \leq x\} \quad (6)$$

Avec $\mathbf{1}\{B\}$ la fonction indicatrice de l'événement B . Sous ces notations, le subsampling pour l'estimateur $|\hat{a}_n(\lambda, \tau)|$ du paramètre $|a(\lambda, \tau)|$ est consistant.

$$\left(|\hat{a}_n(\lambda, \tau)| - \frac{c_{n,b}(1 - \frac{\alpha}{2})}{\sqrt{n}}, |\hat{a}_n(\lambda, \tau)| - \frac{c_{n,b}(\frac{\alpha}{2})}{\sqrt{n}} \right), \quad (7)$$

Tel que $\alpha \in (0, 1)$ est le niveau de signification, et $c_{n,b}(1 - \frac{\alpha}{2})$, $c_{n,b}(\frac{\alpha}{2})$ les quantiles dérivés à partir de la procédure de subsampling. Ainsi, il devient assez simple de construire des tests de signification pour les fréquences cycliques. Cela est fait en testant si les coefficients de Fourier $a(\lambda, \tau)$ de la fonction d'autocovariance $B(t, \tau)$ sont nuls en dehors d'un ensemble fini de fréquences cycliques ou pas.

3.1 Algorithme du subsampling

Dans cette partie, l'algorithme du subsampling pour l'estimation des coefficients de Fourier, ainsi que pour le test des fréquences cycliques est présenté. La procédure de l'estimation est basée sur $\{X(1), X(2), \dots, X(n)\}$ de la série temporelle $\{X(t) : t \in \mathbb{Z}\}$

- a- Estimer $\hat{a}_n(\lambda, \tau)$ basé sur tous les échantillons de longueur n . L'estimation de $\hat{a}_n(\lambda, \tau)$ se fait selon (3)
- b- Choisir une taille de bloc (entier positif) $b(< n)$, et un niveau de confiance $\alpha \in (0, 1)$
- c- Pour $t = 1, 2, \dots, n - b + 1$,
Utiliser (5) pour estimer $\hat{a}_{n,b,t}(\lambda, \tau)$.
- d- Définir

$$V_{re_t}^{n,b,t} = \Re[\sqrt{b}(\hat{a}_{n,b,t}(\lambda, \tau) - \hat{a}_n(\lambda, \tau))]$$
, et

$$V_{im_t}^{n,b,t} = \Im[\sqrt{b}(\hat{a}_{n,b,t}(\lambda, \tau) - \hat{a}_n(\lambda, \tau))]$$

Avec $V_{re_t}^{n,b,t}$ et $V_{im_t}^{n,b,t}$ respectivement les parties réelle et imaginaire de la statistique définie ci-dessus, de taille $n - b + 1$ correspondant aux $n - b + 1$ sous-échantillons.
- e- Ordonner les valeurs de $V_{re_t}^{n,b,t}$ et $V_{im_t}^{n,b,t}$ afin d'obtenir les quantiles $c_{n,b}(1 - \frac{\alpha}{2})$ et $c_{n,b}(\frac{\alpha}{2})$.
- f- Définir les intervalles de confiance pour la partie imaginaire et la partie réelle

$$\left(\hat{a}_n(\lambda, \tau) - \frac{c_{n,b}(1 - \frac{\alpha}{2})}{\sqrt{n}}, \hat{a}_n(\lambda, \tau) - \frac{c_{n,b}(\frac{\alpha}{2})}{\sqrt{n}} \right)$$

Donc, à partir de l'algorithme de la procédure du subsampling décrit ci-dessus, il devient simple de dériver et construire un test de significativité des fréquences cycliques. En effet, si pour chaque λ et τ , zéro appartient à l'intervalle de confiance, alors $a(\lambda, \tau)$ est non significatif. Par conséquent, la fréquence λ testée est dite *non significative*. Cependant, si pour un λ donné et un τ donné, zéro n'appartient pas à l'intervalle de confiance, alors on pourra dire que la fréquence cyclique λ est significative.

4 Simulation

Afin d'illustrer la manière dont le subsampling et la construction de tests sont réalisés en pratique, l'attention est focalisée sur un modèle CS théorique. Soit $\{U_i(t), t \in \mathbb{R}\}$, $i = 1, 2$, deux processus aléatoires stationnaires, indépendants, de moyennes nulles et avec des fonctions de covariance $r_i(\cdot)$, respectivement pour $i = 1, 2$. Le modèle étudié sera la somme de deux modulations d'amplitudes comme suit:

$$X(t) = U_1(t)\cos(2\pi f_1 t) + U_2(t)\cos(2\pi f_2 t), \quad (8)$$

où $n = 8192$ est la longueur des données, f_1 et f_2 sont respectivement égales à $0.1Hz$ et $0.11Hz$.

L'ensemble des fréquences cycliques obtenu théoriquement sera $\Lambda = \{-2f_2, -2f_1, 0, 2f_1, 2f_2\}$, et:

$$\begin{aligned} a(0, \tau) &= \frac{1}{2}r_1(\tau)\cos(2\pi f_1 \tau) + \frac{1}{2}r_2(\tau)\cos(2\pi f_2 \tau) \\ a(2f_1, \tau) &= \frac{1}{4}r_1(\tau)e^{i2\pi f_1 \tau} \quad ; \quad a(2f_2, \tau) = \frac{1}{4}r_2(\tau)e^{i2\pi f_2 \tau} \\ \overline{a(2f_1, \tau)} &= a(-2f_1, \tau) \quad ; \quad \overline{a(2f_2, \tau)} = a(-2f_2, \tau) \end{aligned}$$

Et avec comme test:

$$\begin{cases} H_0 & : a(\lambda, \tau) = 0 \\ H_1 & : a(\lambda, \tau) \neq 0. \end{cases}$$

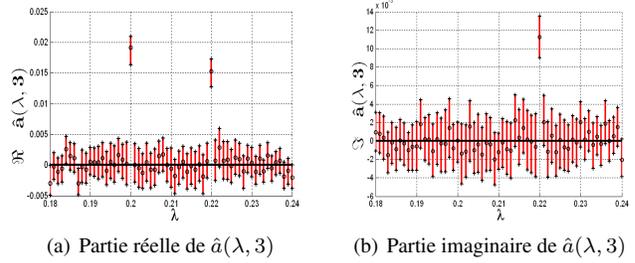


FIG. 1: Les valeurs de l'estimateur et les intervalles de confiance via subsampling pour $\alpha = 0.05$

Afin de tester l'hypothèse H_0 contre l'hypothèse H_1 , la procédure du subsampling a été appliquée à $X(t)$, en utilisant une longueur de bloc $b = \sqrt{n}$ et $\alpha = 0.05$. Les résultats obtenus pour les parties réelle et imaginaire de l'estimateur sont représentés dans la figure(1). Ces résultats montrent clairement que le subsampling appliqué au modèle de simulation $X(t)$ a réussi à identifier les vraies fréquences cycliques théoriques $\{2f_0; 2f_1\}$. Cela signifie que l'hypothèse H_0 à tester est rejetée, donc $X(t)$ est CS2.

5 Application sur des signaux de course à pied et caractérisation de la fatigue

La locomotion humaine, en particulier, la course à pied, se traduit par des séquences de gestes cycliques et répétés. La variabilité de ces séquences pourrait révéler des défaillances motrices. L'objectif de cette partie est d'analyser les foulées d'un coureur professionnel, à partir de ses signaux de réaction au sol, afin de détecter et caractériser la fatigue au cours du temps.

Pour ce faire, des signaux de réaction au sol d'un coureur professionnel ont été acquis sur un tapis de course expérimental, munis de capteurs accélérométriques [6]. Le participant a couru 24 heures, et trois signaux ont été acquis à différents temps; à $T = 0h$, à $T = 12h$ et à $T = 24h$. La fréquence d'échantillonnage des signaux était de 1000 Hz, une durée d'acquisition de 20 secondes, et la vitesse du tapis fixée à 12km/h.

Dans un premier temps, les signaux acquis ont été synchronisés sur le cycle de la foulée afin de rendre la période constante [7]. Ensuite, la moyenne synchrone, qui représente la composante déterministe et périodique du signal a été estimée, puis

retranchée du signal, dans le but de ne garder que la composante du second ordre (CS2).

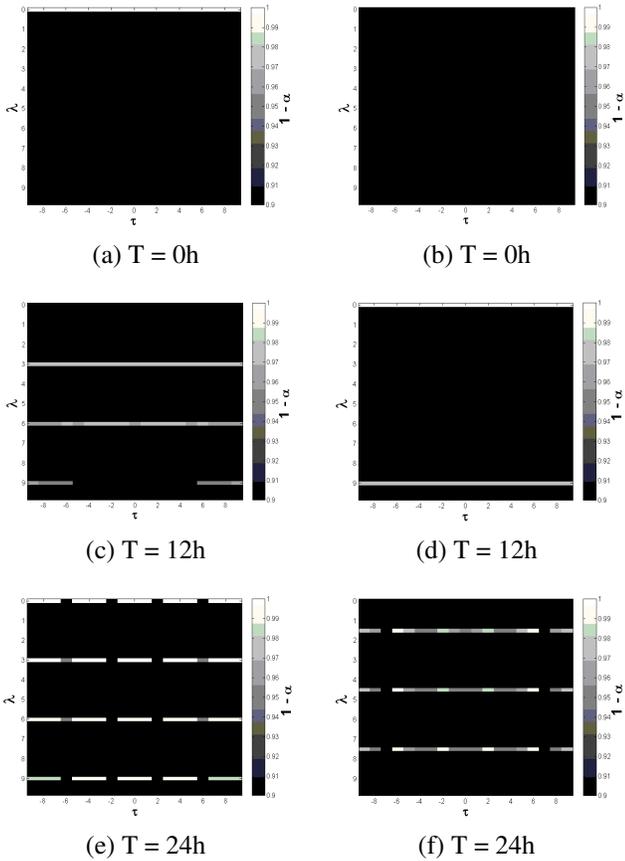


FIG. 2: (a),(c) et (e). Signatures fréquentielles des parties réelles de $\hat{a}(\lambda, \tau)$ pour $T=0$; 12 et 24h, respectivement. (b),(d) et (f). Signatures fréquentielles des parties imaginaires de $\hat{a}(\lambda, \tau)$ respectivement pour $T=0$; 12 et 24h, en fonction de λ , τ and α .

Après l'estimation des fréquences cycliques et leur test via subsampling, en utilisant une longueur de bloc $b = \sqrt{n}$, des cartographies représentant les différentes signatures fréquentielles correspondant à chaque état de fatigue ont été construites et présentées dans la figure (2). Ces signatures fréquentielles associent pour chaque τ et λ de $\hat{a}(\lambda, \tau)$ la plus haute probabilité de signification. Ces cartographies permettent de donner une description complète du comportement fréquentiel au second ordre des signaux traités.

Les signatures fréquentielles obtenues pour les trois différents temps ($T = 0h$, $T = 12h$ et $T = 24h$) ont permis de mettre en évidence des différences significatives sur le contenu fréquentiel du second ordre. Les résultats obtenus suggèrent que lorsque la fatigue apparaît, des fréquences cycliques significatives correspondant à la fréquence du pas du coureur et de ses harmoniques apparaissent. Ceci est dû aux fluctuations des pas causées par l'apparition de la fatigue. De plus, le niveau de signification de ces fréquences cycliques augmente avec le ni-

veau de fatigue du coureur. En outre, de nouvelles fréquences cycliques très significatives correspondant à la fréquence de la foulée du coureur et de ses harmoniques apparaissent avec le temps. Cela traduit des fluctuations de foulées, en plus des fluctuations de pas, quand l'état de fatigue du coureur est très élevé.

6 Conclusion

Ce papier a démontré que la procédure du subsampling a considérablement simplifié l'inférence pour les processus cyclostationnaires à l'ordre 2. Cela a permis la construction d'intervalles de confiance valides pour l'estimateur de $a(\lambda, \tau)$, ainsi que des tests pour la présence de la CS2. Les cartographies de signatures fréquentielles dérivées de cette méthode pourraient être très utiles dans différents domaines d'application, comme la mécanique, l'acoustique et la surveillance des machines tournantes, afin d'identifier et de caractériser beaucoup de défauts, qui ont pour caractéristique, l'apparition sur les propriétés statistiques du second ordre des signaux. Cependant, le problème majeur du subsampling semble être le choix de la longueur du bloc utilisé. Pour cela, des travaux sont en ce moment effectués pour déterminer la longueur optimale de bloc utilisé dans la procédure du subsampling.

Références

- [1] Gardner, W. A., an introduction to cyclostationary signals. Chapter 1 in Cyclostationarity in Communications and Signal Processing. *IEEE Press*, New York, pp. 1-90, 1994.
- [2] Hurd, H., Periodically Correlated Random Sequences - Spectral Theory and Practice. John Wiley & Sons, Inc., Hoboken, New Jersey, 2007.
- [3] Dehay, D. and Leśkow, J., Functional limit theory for the spectral covariance estimator. *Journal of Applied Probability* 33. 1077-92, 1996.
- [4] Łenart, Ł., Leśkow, J. and Synowiecki, R., Subsampling in estimation of autocovariance for PC time series, *Journal of Time Series Analysis* 29 (6). 995-1018, 2008.
- [5] Politis, D., Romano, J. and Wolf, M., Subsampling, *Springer Series in Statistics*, New-York, 1999.
- [6] Sabri, K., El Badaoui, M., Guillet, F., Belli, A., Millet, G., Morin, J.B., Cyclostationary modelling of ground reaction force signals, *Signal Processing* 90, 1146-1152, 2010.
- [7] Bonnardot, F., El Badaoui, M., Randall, R.B., Danière, J., Guillet, F. Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation), *MSSP* 19, 2005.
- [8] Łenart, Ł., Non-parametric frequency identification and estimation in mean function for almost periodically correlated time series, *Journal of Multivariate Analysis* 115, 252-269, 2013.