



HAL
open science

Prévision statistique paramétrique par séparation temporelle.

Olivier Faugeras

► **To cite this version:**

Olivier Faugeras. Prévision statistique paramétrique par séparation temporelle.. Annales de l'ISUP, 2008, LII (3), pp.43-60. hal-03633609

HAL Id: hal-03633609

<https://hal.science/hal-03633609v1>

Submitted on 7 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pub. Inst. Stat. Univ. Paris
LII, fasc. 3, 2008, 43 à 60

PRÉVISION STATISTIQUE PARAMÉTRIQUE PAR SÉPARATION TEMPORELLE.

OLIVIER FAUGERAS

L.S.T.A., Université Paris 6,

175 rue du Chevaleret, Boîte 158, 75013 Paris, France

Résumé Soit $\mathbb{X} = \{X_t, t \in \mathbb{Z}\}$ un processus réel faiblement stationnaire de carré intégrable, de loi indexée par un paramètre θ , observé sur $0 \leq t \leq T$. On cherche à prédire la variable aléatoire X_{T+1} non observée par une statistique \hat{X}_{T+1} qui soit $\sigma(X_t, 0 \leq t \leq T)$ mesurable avec comme critère l'erreur quadratique de prévision $E_\theta(\hat{X}_{T+1} - X_{T+1})^2$. Il est bien connu que l'espérance conditionnelle $X_{T+1}^* := E_\theta(X_{T+1} | X_0^T) := r_\theta(X_0^T)$ minimise cette erreur. Néanmoins ce prédicteur probabiliste n'est pas accessible au statisticien car le paramètre θ est inconnu et doit être estimé par une statistique $\hat{\theta}_T$. Le comportement du prédicteur statistique plug-in $r_{\hat{\theta}_T}(X_0^T)$ est alors difficile à étudier. On se propose de construire un prédicteur statistique lorsque le prédicteur probabiliste r_θ dépend approximativement des k_T dernières valeurs $(X_{T-i}, i = 1, \dots, k_T)$ avec $k_T \rightarrow \infty$. En estimant θ par $\hat{\theta}_{\phi(T)}$ sur l'intervalle $[0, \phi(T)]$, on construit alors le prédicteur statistique $r_{\hat{\theta}_{\phi(T)}}(X_{T-k_T}^T)$ et on étudie sa consistance et sa loi limite asymptotique sous des hypothèses de régularité, de mélangeance et des conditions entre k_T et $\phi(T)$.

Abstract Let $\mathbb{X} = \{X_t, t \in \mathbb{Z}\}$ be a real-valued weakly stationary square integrable time series, with law indexed by a parameter θ , observed on a time interval $0 \leq t \leq T$. We are interested in forecasting the unobserved random variable X_{T+1} by a function \hat{X}_{T+1} of the observations $(X_i, i = 0, \dots, T)$, with the quadratic error criteria $E_\theta(\hat{X}_{T+1} - X_{T+1})^2$. It is well known that the conditional expectation $X_{T+1}^* := E_\theta(X_{T+1} | X_0^T) := r_\theta(X_0^T)$ is a solution to this minimization problem. Nonetheless, this probabilistic forecaster is not a genuine statistical one, since it depends on the unknown value of the parameter θ , which has to be estimated by an estimator $\hat{\theta}_T$. The plug-in statistical forecaster induced $r_{\hat{\theta}_T}(X_0^T)$ is then a difficult object to study. In this paper, we propose to deal with the case where the probabilistic forecaster depends approximatively only on the last k_T values of the time series $(X_{T-i}, i = 1, \dots, k_T)$. By estimating θ by $\hat{\theta}_{\phi(T)}$ on the interval $[0, \phi(T)]$, we build a statistical predictor $r_{\hat{\theta}_{\phi(T)}}(X_{T-k_T}^T)$ and show its consistency and derive its limit in distribution under regularity, mixing, and assumptions on k_T and $\phi(T)$.

1. Introduction.

1.1. *Motivation.* Soit $\mathbb{X} = \{X_t, t \in \mathbb{Z}\}$ un processus à valeurs réelles faiblement stationnaire de carré intégrable, défini sur $(\Omega, \mathcal{A}, \mathbb{P})$, de loi \mathbb{P} indexée par un paramètre θ à valeurs dans \mathbb{R}^d , observé sur $0 \leq t \leq T$. On cherche à prédire la variable aléatoire X_{T+1} non observée par une statistique \hat{X}_{T+1} qui soit $\sigma(X_t, 0 \leq t \leq T)$ mesurable, de carré intégrable et qui minimise l'erreur quadratique $E_\theta(\hat{X}_{T+1} - X_{T+1})^2$. En notant X_a^b la σ -algèbre engendrée par $(X_t, a \leq t \leq b)$ et $r_\theta(X_0^T)$ l'espérance conditionnelle $E_\theta(X_{T+1}|X_0^T)$, rappelons alors le lemme évident suivant :

LEMME 1.1. *L'erreur de prévision se décompose en un terme probabiliste et un terme d'approximation statistique :*

$$E_\theta(X_{T+1} - \hat{X}_{T+1}(X_0^T))^2 = E_\theta(X_{T+1} - r_\theta(X_0^T))^2 + E_\theta(r_\theta(X_0^T) - \hat{X}_{T+1}(X_0^T))^2$$

Le premier terme s'appelle erreur probabiliste et ne dépend que du processus et le second terme s'appelle erreur statistique de prévision et résulte du choix de \hat{X}_{T+1} par le statisticien. L'erreur de prévision est donc minimisée pour le choix de $\hat{X}_{T+1}(X_0^T) = E_\theta(X_{T+1}|X_0^T) := r_\theta(X_0^T)$.

Néanmoins le choix de ce prédicteur, que l'on qualifiera de probabiliste, n'est pas satisfaisant d'un point de vue statistique car le paramètre θ étant inconnu, l'espérance conditionnelle n'est pas accessible au statisticien. On est donc naturellement amené à construire un estimateur \hat{r}_T de cette espérance conditionnelle $r_\theta(\cdot)$ basé sur l'échantillon $(X_i, i = 0, \dots, T)$ pour obtenir le prédicteur statistique $\hat{r}_T(X_0^T)$. Dans un cadre paramétrique où l'on suppose la forme de la fonction de régression r_θ connue, cela se traduit par estimer le paramètre θ par $\hat{\theta}_T$ et construire le prédicteur statistique plug-in $r_{\hat{\theta}_T}(X_0^T)$.

Cependant, le fait que les variables (dépendantes) (X_0, \dots, X_T) servent à la fois dans le problème (statistique) d'estimation de θ et comme valeurs d'entrée dans le calcul (probabiliste) de la fonction de régression, rend l'étude de l'erreur de prévision statistique malaisé. Une manière usuelle de procéder dans la littérature est d'introduire une hypothèse supplémentaire sur la structure de dépendance du processus (voir par exemple Caires et Ferreira [2005] pour une discussion), typiquement markovien d'ordre k , afin de simplifier la fonction de régression $r_\theta(X_0^T)$ en $r_\theta(X_{T-k+1}^T)$, ce qui revient à considérer le problème de la prévision à « passé » fini. Dans le cadre simplifié d'un processus ARMA ayant une structure linéaire, la méthode de Box-Jenkins ou du filtre de Kalman (voir par exemple Box et Jenkins [1994] ou Brockwell et Davis [1996]) permet de traiter ce problème.

1.2. *Séparation temporelle.* On se propose ici de ne pas faire cette hypothèse mais de séparer les problèmes probabiliste et statistique de façon temporelle. On se place dans le cadre où la fonction de régression $r_\theta(\cdot)$ dépend approximativement des k_T dernières valeurs $(X_{T-i}, i = 1, \dots, k_T)$ avec $k_T \leq T, k_T \rightarrow \infty$.

$$X_{T+1}^* := E_\theta [X_{T+1} | X_0^T] := \sum_{i=0}^{k_T} r_i(X_{T-i}, \theta) + \eta_{k_T}(X, \theta)$$

où chaque fonction r_i représente la contribution (additive) de la X_{T-i} valeur, et où $\eta_{k_T}(X, \theta)$ est une fonction de carré intégrable asymptotiquement négligeable dans un sens à préciser. Ce modèle additif est une extension d'un cas particulier étudié par Bosq [2005]. Pour plus de détails concernant les modèles additifs, on pourra se référer à, par exemple, Härdle et al. [2004].

Dans le cas étudié par Bosq [2005] chapitre 2, le prédicteur probabiliste a pour structure

$$X_{T+h}^* := E_\theta(X_{T+h} | X_{-\infty}^T) = r_{T,h}(Y_T, \theta)$$

où Y_T est une variable $\sigma(X_{\pi(T)}^T)$ mesurable, telle que $0 < \pi(T) \leq T$ et $\lim_{T \rightarrow \infty} \pi(T)/T = 1$, i.e. qui représente le proche passé. Le prédicteur statistique est alors construit à partir d'un estimateur $\hat{\theta}_{\psi(T)}$ du paramètre,

$$\hat{X}_{T+h} = r_{T+h}(\hat{\theta}_{\psi(T)}, Y_T)$$

avec $0 < \psi(T) < T, \pi(T) - \psi(T) \rightarrow \infty$ et $\psi(T)/T \rightarrow 1$. La consistance, la vitesse de convergence et la loi limite du prédicteur statistique sont alors obtenues.

Dans cet article, on suppose qu'on dispose d'un estimateur consistant $\hat{\theta}_T$ de θ . En estimant θ par $\hat{\theta}_{\varphi(T)}$ sur l'intervalle $[0, \varphi(T)]$ avec $\varphi(T) \rightarrow \infty$, on construit alors le prédicteur statistique

$$\hat{X}_{t+1} = r_{\hat{\theta}_{\varphi(T)}}(X_{T-k_T}^T)$$

Cette étude a pour but de montrer que si le processus est mélangeant, alors on peut séparer le problème d'estimation de θ sur $[0, \varphi(T)]$, du problème probabiliste sur la « mémoire » du processus entre $[T - k_T, T]$.

Plus précisément, on montrera dans la section 2 la consistance du prédicteur, i.e. la convergence vers 0 de l'erreur statistique de prévision, avant de montrer un exemple d'application inspiré par la décomposition de Wold dans la section 3, pour finir par l'étude de la loi asymptotique du prédicteur statistique dans la section 4.

1.3. *Discussion sur le modèle.* On a dit que le processus est approximativement k_T markovien, ce qui revient à considérer que k_T est imposé par le processus. On peut aussi considérer le modèle ci-dessus comme un modèle additif non-linéaire généralisé au sens où le processus vérifie

$$X_{T+1} = \sum_{i=0}^{+\infty} r_i(X_{T-i}; \theta) + \varepsilon_{T+1}$$

où l'innovation (ε) est telle que $E[\varepsilon_{T+1}|X_{-\infty}^T] = 0$ et où la convergence de la série est à comprendre au sens de la convergence en moyenne quadratique. Une condition pour la convergence de cette série est donnée dans le corollaire 3.1 de Rio [2000] (le premier corollaire 3.1 p.51), rappelé dans le lemme ci-dessous.

LEMME 1.2. *Soit $(Y_i)_{i \in \mathbb{N}}$ une suite de variables réelles centrées de variance finie. Soit Q_i est la fonction de quantile de $|Y_i|$ i.e. l'inverse généralisé continu à droite de la fonction $H_{Y_i}(t) = P(|Y_i| > t)$, et $\alpha(y) = \alpha[y]$ où $[y]$ désigne la partie entière de y et $\alpha(k)$ le coefficient de mélange fort de Rosenblatt [1956] (voir ci-dessous). Alors la série $\sum_{i=1}^{\infty} Y_i$ converge p.s. si la condition suivante est réalisée :*

$$\sum_{i=1}^{\infty} \int_0^1 \alpha^{-1}(u) Q_i^2(u) du < +\infty$$

On notera que ce lemme généralise le théorème des deux séries de Kolmogorov qui traite du cas i.i.d. et requière la convergence des moments d'ordre 1 et 2.

La convergence de la série $\sum_{i=0}^{+\infty} r_i(X_{T-i}; \theta)$ entraîne à son tour que

$$\sum_{i=k}^{+\infty} r_i(X_{T-i}; \theta) \xrightarrow{p.s.} 0$$

pour $k \rightarrow +\infty$. En posant $\eta_{k_T} = \sum_{i=k_T+1}^{+\infty} r_i(X_{T-i}; \theta)$, l'écriture

$$X_{T+1}^* := E_{\theta} [X_{T+1} | X_0^T] := \sum_{i=0}^{k_T} r_i(X_{T-i}, \theta) + \eta_{k_T}(X, \theta)$$

avec l'introduction de $\eta_{k_T}(X, \theta)$ peut se comprendre comme un choix du statisticien de prendre un $k_T \rightarrow +\infty$ de façon à rendre la contribution du passé lointain dans la prévision négligeable, i.e. à avoir $\eta_{k_T}(X, \theta) \rightarrow 0$ pour $T \rightarrow +\infty$.

2. Consistance du prédicteur statistique. On rappelle la notion de α -mélangeance (cf. Rosenblatt [1956]) :

DÉFINITION 2.1. Soit (Ω, \mathcal{A}, P) un espace probabilisé et \mathcal{B}, \mathcal{C} deux sous-tribus de \mathcal{A} . On définit le coefficient de α -mélange entre les deux tribus \mathcal{B}, \mathcal{C} par

$$\alpha(\mathcal{B}, \mathcal{C}) = \sup_{\substack{B \in \mathcal{B} \\ C \in \mathcal{C}}} |P(B \cap C) - P(B)P(C)|$$

et le coefficient de α -mélange d'ordre k pour le processus $\mathbb{X} = \{X_t, t \in \mathbb{N}\}$ défini sur l'espace probabilisé (Ω, \mathcal{A}, P) par

$$\alpha(k) = \sup_{t \in \mathbb{N}} \alpha(\sigma(X_s, s \leq t), \sigma(X_s, s \geq t+k))$$

On rappelle en outre l'inégalité de Davydov (cf. Bosq [1998], p. 21) : Notons $\sigma(X)$ la σ -algèbre des événements engendrés par la variable X et $\|X\|_q = \{\mathbb{E}(X^q)\}^{1/q}$ pour $1 \leq q \leq \infty$.

LEMME 2.2. Soient $X \in L^q(\mathbb{P})$ et $Y \in L^r(\mathbb{P})$, si $q > 1$, $r > 1$ et $\frac{1}{r} + \frac{1}{q} = 1 - \frac{1}{p}$, alors

$$|Cov(X, Y)| \leq 2p(2\alpha(\sigma(X), \sigma(Y)))^{\frac{1}{p}} \|X\|_q \|Y\|_r.$$

On se place donc dans le cadre suivant :

- Le processus \mathbb{X} est du second ordre, faiblement stationnaire, α mélangeant.
- On suppose que le prédicteur probabiliste s'écrit :

$$X_{T+1}^* := E_\theta [X_{T+1} | X_0^T] = \sum_{i=0}^{k_T} r_i(X_{T-i}, \theta) + \eta_{k_T}(X, \theta)$$

On effectue alors les hypothèses suivantes :

Hypothèse \mathbf{H}_0 sur le processus \mathbb{X} .

- (i) $\lim_{T \rightarrow \infty} E_\theta(\eta_{k_T}^2(X, \theta)) = 0$;
- (ii) pour tout $i \in \mathbb{N}$, $\|r_i(X_{T-i}, \theta_1) - r_i(X_{T-i}, \theta_2)\| \leq H_i(X_{T-i}) \|\theta_1 - \theta_2\|$, $\forall \theta_1, \theta_2$;
- (iii) il existe $r > 1$ tel que $\sup_{i \in \mathbb{N}} E_\theta H_i^{2r}(X_{T-i}) < \infty$.

Hypothèse \mathbf{H}_1 sur l'estimateur $\hat{\theta}_T$. On suppose qu'on dispose d'un estimateur consistant $\hat{\theta}_T$ de θ à la vitesse (paramétrique) T .

- (i) $\limsup_{T \rightarrow \infty} T \cdot E_\theta(\hat{\theta}_T - \theta)^2 < \infty$;
 (ii) il existe $q > 1$ tel que $\limsup_{T \rightarrow \infty} T^q E(\hat{\theta}_T - \theta)^{2q} < \infty$.

Hypothèse \mathbf{H}_2 sur les coefficients.

- (i) $\frac{k_T^2}{\varphi(T)} \xrightarrow{T \rightarrow \infty} 0$;
 (ii) $(T - k_T - \varphi(T)) \xrightarrow{T \rightarrow \infty} \infty$.

On est alors en mesure de formuler le théorème suivant :

THÉORÈME 2.3. *Si les hypothèses $\mathbf{H}_0, \mathbf{H}_1, \mathbf{H}_2$ sont vérifiées, alors*

$$\limsup_{T \rightarrow \infty} E_\theta(\hat{X}_{T+1} - X_{T+1}^*)^2 = 0$$

DÉMONSTRATION.

$$\begin{aligned} & E_\theta(X_{T+1}^* - \hat{X}_{T+1})^2 \\ &= E_\theta \left(\sum_{i=0}^{k_T} (r_i(X_{T-i}, \theta) - r_i(X_{T-i}, \hat{\theta}_{\varphi(T)})) + \eta_{k_T}(X, \theta) \right)^2 \\ &\leq 2E_\theta(\eta_{k_T}^2(X, \theta)) + 2E_\theta \left(\sum_{i=0}^{k_T} (r_i(X_{T-i}, \theta) - r_i(X_{T-i}, \hat{\theta}_{\varphi(T)})) \right)^2 \\ &\leq 2E_\theta(\eta_{k_T}^2(X, \theta)) + 2(k_T + 1) \sum_{i=0}^{k_T} E_\theta \left(r_i(X_{T-i}, \theta) - r_i(X_{T-i}, \hat{\theta}_{\varphi(T)}) \right)^2 \\ &\leq 2E_\theta(\eta_{k_T}^2(X, \theta)) + 2(k_T + 1) \sum_{i=0}^{k_T} E_\theta \left(H_i(X_{T-i}) \left\| \hat{\theta}_{\varphi(T)} - \theta \right\| \right)^2 \end{aligned}$$

par application de l'hypothèse \mathbf{H}_0 (ii), d'où

$$\begin{aligned} E_\theta(X_{T+1}^* - \hat{X}_{T+1})^2 &\leq 2E_\theta(\eta_{k_T}^2(X, \theta)) \\ &\quad + 2(k_T + 1) \sum_{i=0}^{k_T} E_\theta H_i^2(X_{T-i}) E_\theta \left\| \hat{\theta}_{\varphi(T)} - \theta \right\|^2 \\ &\quad + 2(k_T + 1) \sum_{i=0}^{k_T} \delta_{i,T} \\ &\leq I_1 + I_2 + I_3 \end{aligned}$$

où on a appliqué l'inégalité du lemme 2.2 avec $X = H_i^2(X_{T-i})$ et $Y = \|\hat{\theta}_{\varphi(T)} - \theta\|^2$, et où on a posé

$$\begin{aligned} I_1 &= 2E_\theta(\eta_{k_T}^2(X, \theta)) \\ I_2 &= 2(k_T + 1) \sum_{i=0}^{k_T} E_\theta H_i^2(X_{T-i}) E_\theta \|\hat{\theta}_{\varphi(T)} - \theta\|^2 \\ I_3 &= 2(k_T + 1) \sum_{i=0}^{k_T} \delta_{i,T} \end{aligned}$$

avec

$$\delta_{i,T} = 2p(2\alpha(T-i-\varphi(T)))^{1/p} \left(E_\theta H_i^{2r}(X_{T-i})\right)^{1/r} \left(E_\theta \|\hat{\theta}_{\varphi(T)} - \theta\|^{2q}\right)^{1/q}$$

Par l'hypothèse \mathbf{H}_0 (i), $\lim_{T \rightarrow \infty} I_1 = 0$.

Par l'hypothèse \mathbf{H}_1 (i), $\limsup_{T \rightarrow \infty} \varphi(T) \cdot E_\theta (\hat{\theta}_{\varphi(T)} - \theta)^2 < \infty$ Donc,

$$\limsup_{T \rightarrow \infty} I_2 \leq \limsup_{T \rightarrow \infty} \frac{k_T}{\varphi(T)} \cdot \sum_{i=0}^{k_T} E_\theta H_i^2(X_{T-i})$$

Or par \mathbf{H}_1 (iii), il existe $r > 1$ tel que $\sup_{i \in \mathbb{N}} E_\theta H_i^{2r}(X_{T-i}) < \infty$, donc

$$\limsup_{T \rightarrow \infty} I_2 \leq \limsup_{T \rightarrow \infty} \frac{k_T^2}{\varphi(T)}.$$

$$I_3 = 4pk_T \left(E_\theta \|\hat{\theta}_{\varphi(T)} - \theta\|^{2q}\right)^{\frac{1}{q}} \sum_{i=0}^{k_T} (2\alpha(T-i-\varphi(T)))^{1/p} \left(E_\theta H_i^{2r}(X_{T-i})\right)^{\frac{1}{r}}$$

Par les hypothèses \mathbf{H}_0 (iii) et \mathbf{H}_1 (ii), on a

$$\begin{aligned} \sup_{i \in \mathbb{N}} \left(E_\theta H_i^{2r}(X_{T-i})\right)^{1/r} &< \infty \\ \limsup_{T \rightarrow \infty} \varphi(T)^q E_\theta (\hat{\theta}_{\varphi(T)} - \theta)^{2q} &< \infty \end{aligned}$$

Donc,

$$\limsup_{T \rightarrow \infty} I_3 \leq 4p \cdot \limsup_{T \rightarrow \infty} \frac{k_T}{\varphi(T)} \sum_{i=0}^{k_T} (2\alpha(T-i-\varphi(T)))^{1/p}$$

Comme X est α -mélangeant, $\alpha(k)$ est décroissant lorsque k croît, donc les coefficients de la somme ci dessus sont majorés par $\alpha(T - k_T - \varphi(T))$, donc

$$\limsup_{T \rightarrow \infty} I_3 \leq \limsup_{T \rightarrow \infty} \frac{k_T^2}{\varphi(T)} \alpha^{1/p}(T - k_T - \varphi(T))$$

Les conditions \mathbf{H}_2 (i), et \mathbf{H}_2 (ii) sur les coefficients assurent alors que

$$\limsup_{T \rightarrow \infty} I_3 = \limsup_{T \rightarrow \infty} I_2 = 0.$$

□

Dans le cadre d'une prévision à « mémoire fixe », i.e. où $k_T = k$, les hypothèses se simplifient et on peut formuler le corollaire suivant :

COROLLAIRE 2.4. *On suppose que le prédicteur probabiliste s'écrit :*

$$X_{T+1}^* := E_\theta [X_{T+1} | X_0^T] = \sum_{i=0}^k r_i(X_{T-i}, \theta) + \eta_T(X, \theta)$$

On garde les hypothèses $\mathbf{H}_0, \mathbf{H}_1, \mathbf{H}_2$ que l'on modifie de la façon suivante :

- pour l'hypothèse \mathbf{H}_0 (i) sur le processus \mathbb{X} par $\lim_{T \rightarrow \infty} E_\theta(\eta_T^2(X, \theta)) = 0$;
- pour les hypothèses \mathbf{H}_2 sur les coefficients par
 - \mathbf{H}_2 (i) $\varphi(T) \rightarrow \infty$;
 - \mathbf{H}_2 (ii) $T - \varphi(T) \xrightarrow{T \rightarrow \infty} \infty$.

Alors, sous ces nouvelles hypothèses, $\limsup_{T \rightarrow \infty} E_\theta(\hat{X}_{T+1} - X_{T+1}^*)^2 = 0$.

3. Exemple d'application. Dans cette section on explicite les hypothèses du théorème 2.3 en basant notre discussion sur la décomposition de Wold d'un processus (cf. par exemple Brockwell & Davis [1996]) : pour un processus linéaire à temps discret, faiblement stationnaire, centré, purement non déterministe et inversible, sa décomposition peut s'écrire sous la forme suivante :

$$X_T = e_T + \sum_{i=1}^{k_T} \varphi_i(\theta) X_{T-i} + \sum_{i > k_T} \varphi_i(\theta) X_{T-i}$$

avec $\sum_{i=1}^{\infty} \varphi_i^2(\theta) < \infty$. On a alors

$$\begin{aligned} X_{T+1}^* &:= E \left[X_{T+1} \middle| X_{-\infty}^T \right] \\ &= \sum_{i=1}^{k_{T+1}} \varphi_i(\theta) X_{T+1-i} + \sum_{i>k_{T+1}} \varphi_i(\theta) X_{T+1-i} + E \left[e_{T+1} \middle| X_{-\infty}^T \right] \\ &:= \sum_{i=0}^{k_T} r_i(X_{T-i}, \theta) + \eta_{k_T}(X, \theta) \end{aligned}$$

avec

$$r_i(X_{T-i}, \theta) := \varphi_{i+1}(\theta) X_{T-i}$$

$$\eta_{k_T}(X, \theta) := \sum_{i>k_T} \varphi_{i+1}(\theta) X_{T-i} + E[e_{T+1} | X_{-\infty}^T]$$

Si pour tout i , φ_i est dérivable et $\varphi'_i(\cdot)$ est borné, on peut écrire pour tous θ_1 et θ_2 ,

$$\begin{aligned} \|r_i(X_{T-i}, \theta_1) - r_i(X_{T-i}, \theta_2)\| &= \|(\varphi_{i+1}(\theta_1) - \varphi_{i+1}(\theta_2)) X_{T-i}\| \\ &\leq \|\varphi'_{i+1}(\cdot)\|_{\infty} \|\theta_1 - \theta_2\| \|X_{T-i}\| \end{aligned}$$

et les hypothèses \mathbf{H}_0 (ii), (iii), sont vérifiées avec $H_i = id$. \mathbf{H}_0 (iv) est vérifiée si X admet un moment d'ordre $2r$, pour un $r > 1$. Si $E[e_{T+1} | X_{-\infty}^T] = 0$,

$$E\eta_{k_T}^2(X, \theta) = \sum_{i>k_T} \sum_{j>k_T} \varphi_{i+1}(\theta) \varphi_{j+1}(\theta) \text{cov}(X_{T-i}, X_{T-j})$$

en utilisant l'inégalité du lemme 2.2,

$$E\eta_{k_T}^2(X, \theta) \leq 2^{(2p+1)/p} p \|X_0\|_q \|X_0\|_r \sum_{i,j>k_T} \varphi_{i+1}(\theta) \varphi_{j+1}(\theta) \alpha^{1/p} (|i-j+1|)$$

et la condition,

$$\sum_{i,j} \varphi_{i+1}(\theta) \varphi_{j+1}(\theta) \alpha^{1/p} (|i-j+1|) < \infty$$

assure la validité de l'hypothèse \mathbf{H}_0 (i).

On a donc montré la proposition suivante :

PROPOSITION 3.1. *Si \mathbb{X} vérifie les conditions*

1. $\forall i, \varphi_i$ est dérivable et $\|\varphi'_i(\cdot)\|_\infty < \infty$;
2. il existe un $r > 1$ tel que (X_t) admet un moment d'ordre $2r$;
3. $E[e_{T+1} | X_{-\infty}^T] = 0$;
4. \mathbb{X} est α -mélangeant vérifie $\sum_{i,j} \varphi_{i+1}(\theta) \varphi_{j+1}(\theta) \alpha^{1/p} (|i-j|) < \infty$.

Alors \mathbb{X} vérifie les conditions du théorème 2.3.

Dans le cas encore plus particulier d'un processus auto-régressif d'ordre p , le corollaire 2.4 s'applique avec les conditions suivantes :

COROLLAIRE 3.2. Si \mathbb{X} vérifie les conditions

1. $\forall i, \varphi_i$ est dérivable et $\|\varphi'_i(\cdot)\|_\infty < \infty$;
2. il existe un $r > 1$ tel que (X_t) admet un moment d'ordre $2r$;
3. $E[e_{T+1} | X_{-\infty}^T] = 0$;
4. \mathbb{X} est α -mélangeant.

Alors \mathbb{X} vérifie les conditions du corollaire 2.4.

4. Loi limite du prédicteur statistique.

4.1. *Un lemme d'indépendance asymptotique.* On note $\overset{d}{\rightsquigarrow}$ la convergence en loi. On établit tout d'abord une conséquence de la mélangeance sur la convergence en loi des vecteurs aléatoires par le lemme suivant :

LEMME 4.1. Soit (X'_n) et (X''_n) deux suites de variables aléatoires réelles de lois respectives P'_n et P''_n définies sur l'espace probabilisé (Ω, \mathcal{A}, P) . On suppose que (X'_n) et (X''_n) sont asymptotiquement deux à deux mélangeants au sens où il existe une suite de coefficients $\alpha(n)$ avec $\alpha(n) \xrightarrow{n \rightarrow \infty} 0$ tels que, pour tous boréliens A et B de \mathcal{R} ,

$$|P(X'_n \in A, X''_n \in B) - P(X'_n \in A)P(X''_n \in B)| \leq \alpha(n)$$

Alors, si

1. $X'_n \overset{d}{\rightsquigarrow} X'$ de loi P' ;
 2. $X''_n \overset{d}{\rightsquigarrow} X''$ de loi P'' ;
- $(X'_n, X''_n) \overset{d}{\rightsquigarrow} (X', X'')$, et la loi de (X', X'') est $P' \otimes P''$.

DÉMONSTRATION. D'après Billingsley [1968] Théorème 2.1, $P_n \rightarrow P \Leftrightarrow \lim_n P_n(C) = P(C)$ pour tout C , ensemble de continuité de P , i.e. pour tout ensemble C tel que $P(\partial C) = 0$ où ∂C représente la frontière de C .

$X'_n \xrightarrow{d} X'$ donc $\forall A$ ensemble de continuité de la loi de X' , $P(X'_n \in A) \rightarrow P(X' \in A)$. $X''_n \xrightarrow{d} X''$ donc $\forall B$ ensemble de continuité de la loi de X'' , $P(X''_n \in B) \rightarrow P(X'' \in B)$. Donc quel que soit $\varepsilon > 0$, il existe un rang n_0 tel que pour tout $n > n_0$, on ait

$$|P(X'_n \in A)P(X''_n \in B) - P(X' \in A)P(X'' \in B)| \leq \varepsilon$$

Par ailleurs, comme X'_n et X''_n sont deux à deux α -mélangeants, on a

$$|P(X'_n \in A, X''_n \in B) - P(X'_n \in A)P(X''_n \in B)| \leq \alpha(n)$$

Donc, pour tout $\varepsilon > 0$,

$$\begin{aligned} & |P(X'_n \in A, X''_n \in B) - P(X' \in A)P(X'' \in B)| \\ & \leq |P(X'_n \in A, X''_n \in B) - P(X'_n \in A)P(X''_n \in B)| \\ & + |P(X'_n \in A)P(X''_n \in B) - P(X' \in A)P(X'' \in B)| \\ & \leq \alpha(n) + \varepsilon \\ & \leq 2\varepsilon \text{ pour } n \text{ assez grand} \end{aligned}$$

D'après Billingsley [1968] Théorème 3.1, $(X'_n, X''_n) \xrightarrow{d} (X', X'')$ si et seulement si pour tout A , ensemble de continuité de la loi de X' , pour tout B ensemble de continuité de la loi de X'' , on a $P(X'_n \in A, X''_n \in B) \rightarrow P(X' \in A, X'' \in B)$ quand $n \rightarrow \infty$.

Donc $(X'_n, X''_n) \xrightarrow{d} (X', X'')$ et la loi limite de (X', X'') est $P' \otimes P''$. \square

4.2. *Convergence en loi.* On notera par la suite ∂_θ l'opération de dérivation par rapport à θ . On reprend le cadre de la section 2, en introduisant des hypothèses supplémentaires. Pour plus de clarté, nous écrivons toutes les hypothèses nécessaires :

Hypothèses sur le processus \mathbb{X}

- (i) \mathbb{X} est un processus du second ordre, faiblement stationnaire, α mélangeant.
- (ii) On suppose que le prédicteur probabiliste s'écrit :

$$X_{T+1}^* := E_\theta \left[X_{T+1} \mid X_0^T \right] = \sum_{i=0}^{k_T} r_i(X_{T-i}, \theta) + \eta_{k_T}(X, \theta)$$

Hypothèse \mathbf{H}'_0 .

- (i) $\theta \rightarrow r_i(X_{T-i}, \theta)$ est deux fois différentiable en θ ;
 (ii) la norme infinie de $\partial_\theta^2 r_i(X_{T-i}, \cdot)$ est borné en probabilité uniformément en i , i.e. $\partial_\theta^2 r_i(X_{T-i}, \cdot)$ est uniformément tendue :

$$\sup_i \left\| \partial_\theta^2 r_i(X_{T-i}, \cdot) \right\|_\infty = O_P(1);$$

(iii) $\eta_{k_T}(X, \theta) = o_P \left(\sqrt{\frac{1}{\varphi(T)}} \right);$

- (iv) $\sum_{i=0}^{+\infty} \partial_\theta r_i(X_{T-i}; \theta)$ existe et converge presque sûrement vers un vecteur V lorsque $T \rightarrow +\infty$.

Hypothèse \mathbf{H}'_1 sur l'estimateur $\hat{\theta}_T$.

(i) $\sqrt{T}(\hat{\theta}_T - \theta) \stackrel{d}{\sim} N(0, \sigma^2(\theta)).$

Hypothèse \mathbf{H}'_2 sur les coefficients .

- (i) $k_T = o(\sqrt{\varphi(T)});$
 (ii) $(T - k_T - \varphi(T)) \xrightarrow{T \rightarrow \infty} \infty.$

On a alors le théorème suivant :

THÉORÈME 4.2. Si les hypothèses $\mathbf{H}'_0, \mathbf{H}'_1, \mathbf{H}'_2$ sont vérifiées, alors

$$\sqrt{\varphi(T)}(\hat{X}_{T+1} - X_{T+1}^*) \stackrel{d}{\sim} \langle U, V \rangle$$

où U et V sont deux variables indépendantes, U de loi $N(0, \sigma^2(\theta))$ et V la loi limite de $\sum_{i=0}^{+\infty} \partial_\theta r_i(X_{T-i}; \theta)$.

DÉMONSTRATION. Par l'hypothèse \mathbf{H}'_0 (i), il existe une suite de variables $(\xi_{T,i})$ qui soient $\sigma(X_0^{T-i})$ mesurable et incluses dans un voisinage de θ de sorte que

$$\begin{aligned} r_i(X_{T-i}; \hat{\theta}_{\phi(T)}) - r_i(X_{T-i}; \theta) &= (\hat{\theta}_{\phi(T)} - \theta)^t \cdot \partial_\theta r_i(X_{T-i}; \theta) \\ &+ (\hat{\theta}_{\phi(T)} - \theta)^t \frac{\partial_\theta^2 r_i(X_{T-i}; \xi_{T,i})}{2} (\hat{\theta}_{\phi(T)} - \theta) \end{aligned}$$

où $(\hat{\theta}_{\phi(T)} - \theta)^t$ désigne la transposée du vecteur $(\hat{\theta}_{\phi(T)} - \theta)$ et $\partial_\theta r_i(X_{T-i}; \theta)$

le gradient par rapport à θ de $r_i(X_{T-i}; \theta)$. Donc,

$$\begin{aligned}\hat{X}_{T+1} - X_{T+1}^* &= \sum_{i=0}^{k_T} (r_i(X_{T-i}, \hat{\theta}_{\varphi(T)}) - r_i(X_{T-i}, \theta)) - \eta_{k_T}(X, \theta) \\ &= \sum_{i=0}^{k_T} (\hat{\theta}_{\varphi(T)} - \theta)^t \cdot \partial_{\theta} r_i(X_{T-i}, \theta) \\ &\quad + \sum_{i=0}^{k_T} (\hat{\theta}_{\varphi(T)} - \theta)^t \frac{\partial_{\theta}^2 r_i(X_{T-i}; \xi_{T,i})}{2} (\hat{\theta}_{\varphi(T)} - \theta) \\ &\quad - \eta_{k_T}(X, \theta)\end{aligned}$$

Donc,

$$\begin{aligned}\sqrt{\varphi(T)}(\hat{X}_{T+1} - X_{T+1}^*) &= \sqrt{\varphi(T)}(\hat{\theta}_{\varphi(T)} - \theta)^t \sum_{i=0}^{k_T} (\partial_{\theta} r_i(X_{T-i}, \theta)) \\ &\quad + \sqrt{\varphi(T)} \frac{(\hat{\theta}_{\varphi(T)} - \theta)^t}{2} \sum_{i=0}^{k_T} \partial_{\theta}^2 r_i(X_{T-i}; \xi_{T,i}) (\hat{\theta}_{\varphi(T)} - \theta) \\ &\quad - \sqrt{\varphi(T)} \eta_{k_T}(X, \theta) \\ &:= I_1 + I_2 - I_3\end{aligned}$$

Etude de I_2 .

Par l'hypothèse \mathbf{H}'_1 (i), $\sqrt{\varphi(T)}(\hat{\theta}_{\varphi(T)} - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$. Donc $\|\hat{\theta}_{\varphi(T)} - \theta\|^2 = O_P(1/\varphi(T))$. En rajoutant la condition \mathbf{H}'_0 (ii), $\sup_i \partial_{\theta}^2 r_i(X_{T-i}, \xi_{T,i}) = O_P(1)$, on obtient

$$I_2 = \sqrt{\varphi(T)} O_P\left(\frac{1}{\varphi(T)}\right) \sum_{i=0}^{k_T} O_P(1) = O_P\left(\frac{k_T}{\sqrt{\varphi(T)}}\right)$$

Etude de I_3 .

$I_3 = \sqrt{\varphi(T)} \eta_{k_T}(X, \theta)$. Si on suppose, par l'hypothèse \mathbf{H}'_0 (iii), que

$$\eta_{k_T}(X, \theta) = o_P\left(\sqrt{\frac{1}{\varphi(T)}}\right)$$

alors $I_3 = o_P(1)$.

Etude de I_1 .

On pose $I_1 = \langle U_T, V_T \rangle$ avec

$$U_T = \sqrt{\varphi(T)}(\hat{\theta}_{\varphi(T)} - \theta)$$

$$V_T = \sum_{i=0}^{k_T} \partial_{\theta} r_i(X_{T-i}, \theta)$$

U_T est $\sigma(X_0^{\varphi(T)})$ mesurable et V_T est $\sigma(X_{T-k_T}^T)$ mesurable donc, U_T et V_T ont pour coefficient de α -mélange, $\alpha(T - k_T - \varphi(T))$, et sont donc deux à deux mélangeants pour $T - k_T - \varphi(T) \rightarrow \infty$ (hypothèse \mathbf{H}'_2 (ii)). Par le lemme 4.1, il suffit donc d'étudier la convergence en loi de U_T et V_T séparément pour en déduire la convergence de (U_T, V_T) et par continuité celle de $U_T V_T$.

Or, par hypothèse, $U_T = \sqrt{\varphi(T)}(\hat{\theta}_{\varphi(T)} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta)) := U$. Pour V_T , la condition \mathbf{H}'_0 (iv), assure l'existence d'un vecteur

$$V := \sum_{i=0}^{+\infty} \partial_{\theta} r_i(X_{T-i}; \theta)$$

tel que

$$\sum_{i=0}^T \partial_{\theta} r_i(X_{T-i}; \theta) \xrightarrow{\mathcal{L}^2} V$$

Donc pour la sous suite k_T , $V_T \rightarrow V$ en moyenne quadratique. Donc $V_T \xrightarrow{d} V$ en loi.

En utilisant le lemme 4.1 précité, on obtient alors

$$\langle U_T, V_T \rangle \xrightarrow{d} \langle U, V \rangle$$

On a donc $I_1 + I_2 - I_3 = \langle U_T, V_T \rangle + o_P\left(\frac{k_T}{\sqrt{\varphi(T)}}\right) + o_P(1)$. Si $k_T = o(\sqrt{\varphi(T)})$ (hypothèse \mathbf{H}'_2 (i)), $I_1 + I_2 - I_3 = \langle U_T, V_T \rangle + o_P(1)$. Par le lemme de Slutsky, on en déduit que

$$\sqrt{\varphi(T)}(\hat{X}_{T+1} - X_{T+1}^*) \xrightarrow{d} \langle U, V \rangle$$

avec U de loi $\mathcal{N}(0, \sigma^2(\theta))$. □

4.3. *Discussion.* Les conditions sur les coefficients k_T et $\varphi(T)$ donnent une indication du compromis à effectuer entre l'approximation probabiliste et l'estimation statistique. En effet, pour

$$\eta_{k_T} = \sum_{i=k_T+1}^{+\infty} \partial_{\theta} r_i(X_{T-i}; \theta)$$

On a par Minkowski,

$$\|\eta_{k_T}\|_2 \leq \sum_{i=k_T}^{+\infty} \|\partial_\theta r_i(X_{T-i}; \theta)\|_2$$

Si par exemple

$$\|\partial_\theta r_i(X_{T-i}; \theta)\|_2 \sim \frac{1}{i^a}$$

pour un $a > 1$, alors

$$\|\eta_{k_T}\|_2 \sim \sum_{i=k_T}^{+\infty} \frac{1}{i^a} \sim \frac{k_T^{1-a}}{1-a}$$

et la condition \mathbf{H}'_0 (iii) est vérifiée pour

$$\frac{\sqrt{\varphi(T)}}{k_T^{a-1}} = o(1)$$

On doit effectuer un compromis pour rendre compatibles cette dernière condition et la condition \mathbf{H}'_2 (i) : $k_T = o(\sqrt{\varphi(T)})$.

Remarque : on aurait pu généraliser ce résultat à d'autres vitesses d'estimation que \sqrt{T} mais on a préféré ne pas présenter un cas plus général pour des raisons de lisibilité.

4.4. *Cas de la mémoire fixe.* Lorsque $k_T = k$ est fixe, le théorème précédent donne le corollaire suivant :

COROLLAIRE 4.3. *On suppose que le prédicteur probabiliste s'écrit :*

$$X_{T+1}^* := E_\theta [X_{T+1} | X_0^T] = \sum_{i=0}^k r_i(X_{T-i}, \theta) + \eta_T(X, \theta)$$

sous les hypothèses modifiées suivantes

Hypothèse \mathbf{H}'_0 .

(i) $\theta \rightarrow r_i(X_{T-i}, \theta)$ est deux fois différentiable en θ ;

(ii) la norme infinie de $\partial_\theta^2 r_i(X_{T-i}, \cdot)$ est borné en probabilité uniformément en i , i.e. $\partial_\theta^2 r_i(X_{T-i}, \cdot)$ est uniformément tendue :

$$\sup_i \left\| \partial_\theta^2 r_i(X_{T-i}, \cdot) \right\|_\infty = O_P(1);$$

(iii) $\eta_T(X, \theta) = o_P\left(\sqrt{\frac{1}{\varphi(T)}}\right)$;

(iv) $\sum_{i=0}^k \partial_\theta r_i(X_{T-i}; \theta)$ converge en loi vers un vecteur V lorsque $T \rightarrow +\infty$.

Hypothèse \mathbf{H}'_1 sur l'estimateur $\hat{\theta}_T$.

$$(i) \sqrt{T}(\hat{\theta}_T - \theta) \overset{d}{\rightsquigarrow} N(0, \sigma^2(\theta)).$$

Hypothèse \mathbf{H}'_2 sur les coefficients .

$$(i) (T - \varphi(T)) \xrightarrow{T \rightarrow \infty} \infty.$$

$$\sqrt{\varphi(T)}(\hat{X}_{T+1} - X_{T+1}^*) \overset{d}{\rightsquigarrow} \langle U, V \rangle$$

où U et V sont deux variables indépendantes.

4.5. Application à la prévision d'un $\mathbf{AR}(l)$. Un exemple important dans les applications est celui d'un processus linéaire autorégressif d'ordre l ,

$$X_t = \sum_{i=1}^l \theta_i X_{t-i} + \varepsilon_t$$

où (ε_t) est une suite i.i.d centrée. On pose $P(z) = z^l - \sum_{i=1}^l \theta_i z^{l-i}$. D'après Doukhan [1994], théorème 6 et corollaire 3 de la section 2.4.1.2, si le polynôme P a toutes ses racines à l'intérieur du disque unité ouvert et si la distribution marginale commune des (ε_t) est dominée par la mesure de Lebesgue, le processus (X_t) est stationnaire et géométriquement ergodique donc géométriquement mélangeant. Il existe donc $a > 0$ et $b > 0$ tel que le coefficient de α -mélange vérifie $\alpha(u) \leq a \exp(-bu)$, pour $u \geq 0$.

D'après Box et Jenkins [1994], la stationnarité du processus entraîne que les coefficients θ satisfont les équations de Yule-Walker, à savoir $\gamma(k) = \theta_1 \gamma(k-1) + \dots + \theta_l \gamma(k-l)$ pour $k > 0$, où on a noté $\gamma(k) = \text{Cov}(X_0, X_k)$ les autocovariances. On considère alors l'estimateur sans biais des autocovariances $\hat{\gamma}_T = (\hat{\gamma}_T(1), \dots, \hat{\gamma}_T(l))$ défini par

$$\hat{\gamma}_T(i) = \frac{1}{T-i} \sum_{t=1}^{T-i} X_t X_{t+i}$$

pour $i = 1, \dots, l$ et $T > i$. En inversant l'équation de Yule-Walker, on obtient un estimateur de θ :

$$\hat{\theta}_T = \hat{\Gamma}_{l-1}^{-1} \hat{\gamma}_T$$

où $\hat{\gamma}_T$ est le vecteur $(\hat{\gamma}_T^1, \dots, \hat{\gamma}_T^l)$ et $\hat{\Gamma}_{l-1}$ la matrice

$$\begin{bmatrix} 1 & \hat{\gamma}_T(1) & \cdots & \hat{\gamma}_T(l-1) \\ \hat{\gamma}_T(1) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\gamma}_T(1) \\ \hat{\gamma}_T(l-1) & \cdots & \hat{\gamma}_T(1) & 1 \end{bmatrix}$$

On a alors, cf. Gouriéroux et Monfort [1990] propriété 9.7 et 9.8, Si X est un processus gaussien, stationnaire admettant une représentation moyenne mobile infinie, tel que $\sum_{k=-\infty}^{+\infty} |k\gamma(k)| < +\infty$, alors

$$\sqrt{T}(\hat{\gamma}_T - \gamma) \overset{d}{\rightsquigarrow} N(0, \Sigma)$$

où N est un vecteur Gaussien l dimensionnel de matrice de covariance $\Sigma = [\Sigma_{kl}] = [\sum_{j=-\infty}^{+\infty} \gamma(j)(\gamma(j+k+l) + \gamma(j+k-l))]$.

On peut aussi estimer θ par la méthode du maximum de vraisemblance, où la variance σ^2 de ε_t est supposée inconnue. On a alors, sous les conditions de la propriété 9.51 de Gouriéroux et Monfort [1990],

$$\sqrt{T} \begin{pmatrix} \hat{\theta}_T - \theta \\ \hat{\sigma}_T^2 - \sigma^2 \end{pmatrix} \overset{d}{\rightsquigarrow} N(0, J(\theta, \sigma^2)^{-1})$$

où $J(\theta, \sigma^2)$ est la matrice d'information de Fisher. La matrice de variance covariance asymptotique de $\sqrt{T}(\hat{\theta}_T - \theta)$ vaut $\sigma^2/\gamma(0)\Gamma_{l-1}^{-1}$, où Γ_{l-1} est l'analogue sans estimateurs de la matrice $\hat{\Gamma}_{l-1}$ décrite précédemment.

Le prédicteur probabiliste s'écrit alors

$$X_{T+1}^* := E_\theta[X_{T+1}|X_0^T] = \sum_{i=0}^{l-1} \theta_{i+1} X_{T-i}$$

qui est de la forme voulue avec $r_i(X_{T-i}, \theta) := \theta_{i+1} X_{T-i}$ et $\eta_T(X, \theta) = 0$. On utilise l'une ou l'autre des méthodes d'estimation de θ pour construire le prédicteur statistique

$$\hat{X}_{T+1} = \sum_{i=0}^{l-1} \hat{\theta}_{\varphi(T)}(i+1) X_{T-i}$$

Les conditions \mathbf{H}'_0 (i)-(iii) d'application du corollaire 4.3 avec $k = l-1$ sont trivialement satisfaites. $\partial_{\theta_j} r_i(X_{T-i}, \theta) = X_{T-i}$ pour $j = i+1$, et 0 sinon. On a donc

$$\sum_{i=0}^{l-1} \partial_{\theta} r_i(X_{T-i}; \theta) = \sum_{i=0}^{l-1} \partial_{\theta}(\theta_{i+1} X_{T-i}) = \begin{bmatrix} X_T \\ X_{T-1} \\ \vdots \\ X_{T-l+1} \end{bmatrix} := V_T$$

Par ergodicité du processus, il existe un vecteur V tel que $V_T \overset{d}{\rightsquigarrow} V$, et la condition \mathbf{H}'_0 (iv) est vérifiée. En choisissant par exemple $\varphi(T) = T - \log T$, le corollaire 4.3 entraîne que

$$\sqrt{\varphi(T)}(\hat{X}_{T+1} - X_{T+1}^*) \overset{d}{\rightsquigarrow} \langle U, V \rangle$$

où U est la loi limite de $\sqrt{T}(\hat{\theta}_T - \theta)$.

Bibliographie

- [1] Billingsley, P. (1968) Convergence of Probability measures. John Wiley and Sons, Inc., New York-London-Sydney.
- [2] Bosq, D. (1998) Nonparametric statistics for stochastic processes. *Lecture Notes in Statistics 110, Second Edition*, Estimation and prediction, Springer, New York.
- [3] Bosq, D. (2005) Inférence et prévision en grandes dimensions. Collection « Économie et Statistiques Avancées ». Economica, Paris.
- [4] Box, G., Jenkins, G., Reinsel, G. (1994) Time series analysis : Forecasting and control. *Third edition*, Prentice Hall Inc. Englewood Cliffs, NJ.
- [5] Brockwell, P.J., Davis, R.A. (1996) Time Series : Theory and Methods. Springer-Verlag, New York.
- [6] Caires, S.; Ferreira, J. A. (2005) On the non-parametric prediction of conditionally stationary sequences. *Stat. Inference Stoch. Process.* 8 , no. 2, 151-184.
- [7] Doukhan, P. (1994) Mixing. Properties and examples. *Lecture Notes in Statistics*, 85. Springer-Verlag, New York, 1994.
- [8] Gouriéroux, C.; A. Monfort. (1990) Séries Temporelles et modèles dynamiques. Collection « Économie et Statistiques Avancées ». Economica, Paris.
- [9] Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). Nonparametric and Semiparametric Models. Springer.
- [10] Rio, E. [2000] Théorie asymptotique des processus aléatoires faiblement dépendants. *Mathématiques & Applications (Berlin)*, 31. Springer-Verlag, Berlin, 2000.
- [11] Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* 42, 43-47.

E-MAIL: olivier.faugeras@gmail.com