Probab. Theory Relat. Fields **111**, 585–608 (1998)

Probability Theory and Related Fields

Processus empiriques absolument réguliers et entropie universelle

Emmanuel Rio

URA no. 0743 CNRS, Université de Paris-Sud, Bât. 425, Mathématique, F-91405 Orsay Cedex France.
e-mail: emmanuel.rio@math.u-psud.fr

Reçu: 21 février 1997 / Version revisée: 24 novembre 1997

English title: Absolutely regular empirical processes and universal entropy

Abstract. We extend Pollard's uniform central limit theorem for empirical processes indexed by families of functions $\mathcal{F}$ to β-mixing empirical processes. Under some suitable universal entropy conditions on $\mathcal{F}$, we prove a functional central limit theorem for stationary and β-mixing sequences under the minimal sommability condition on the β-mixing coefficients $\sum_n \beta_n < \infty$.

Mathematics Subject Classification (1991): 60F05

Résumé. Nous étendons le théorème limite central fonctionnel de Pollard pour les processus empiriques indexés par des familles de fonctions $\mathcal{F}$ d'entropie universelle intégrable aux suites d'observations absolument régulières et stationnaires. Sous des conditions d'entropie universelles sur $\mathcal{F}$ raisonnables, nous montrons un théorème limite central uniforme pour les suites de variables aléatoires absolument régulières et stationnaires sous la condition minimale de sommabilité des coefficients de β-mélange $\sum_n \beta_n < \infty$.

1. Introduction

Soit $(X_i)_{i\in\mathbb{Z}}$ une suite de variables aléatoires à valeurs dans $\mathcal{X}$ espace polonais muni de sa tribu borélienne $\mathcal{B}(\mathcal{X})$, strictement stationnaire. Soit P la loi commune des observations, $P_n = n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n})$ la probabilité empirique, et $Z_n = \sqrt{n}(P_n - P)$ la mesure empirique centrée et normalisée. Soit $\mathscr{F}$ une classe de fonctions mesurables de $\mathscr{X}$ dans $\mathbb{R}$, de carré intégrable sous P.

1.1. Observations indépendantes

Supposons les variables X_i indépendantes. La théorie des théorèmes limites centraux (TLC) uniformes pour le processus empirique normalisé $\{Z_n(f) : f \in \mathscr{F}\}$ a commencé par l'étude de la convergence uniforme du processus empirique sur certaines classes de parties. Le premier résultat est celui de Donsker (1952) pour la classe $\mathscr{F} = \{\mathbb{I}_{]-\infty,t]} : t \in \mathbb{R}\}$. Dans son article fondateur, Dudley (1978) a donné une définition rigoureuse de la notion de TLC uniforme en termes de convergence faible du processus empirique Z_n dans l'espace de Banach (pas toujours séparable) $\mathscr{L}_\infty(\mathscr{F})$. Il a aussi donné deux conditions suffisantes pour le TLC uniforme dans le cas des classes de parties. La première est une condition portant sur l'intégrabilité de l'entropie avec crochets dans $\mathscr{L}_1(P)$. La seconde est la propriété combinatoire de Vapnik et Chervonenkis, vérifiée par un certain nombre de classes de parties usuelles, telles que la classe des intervalles, la classe des polygones à nombre de faces borné, la classe des boules euclidiennes, etc... Dans ce cas le TLC uniforme est vérifié pour toute probabilité P. Depuis la théorie des TLC uniformes pour les processus empiriques a été développée dans la direction des classes de fonctions. Dudley (1984) a étendu son résultat sur les classes de parties d'entropie avec crochets dans $\mathscr{L}_1(P)$ intégrable aux classes de fonctions bornées. Puis, en tronquant les fonctions de façon adaptative, Ossiander (1987) a étendu ce résultat aux classes de fonctions d'entropie avec crochets dans $\mathscr{L}_2(P)$ intégrable. Nous renvoyons à Andersen et al. (1988) pour des résultats plus récents et une bibliographie détaillée. Une autre approche possible est de considérer des classes de fonctions d'entropie universelle intégrable. Pour ces classes, Kolchinskii (1981) et Pollard (1982) ont obtenu un TLC uniforme sous des conditions sur l'entropie universelle L^1 et sur l'entropie universelle L^2, étendant ainsi le TLC uniforme de Dudley pour les classes de parties de Vapnik-Chervonenkis aux classes de fonctions.

1.2. Observations faiblement dépendantes

Il existe de nombreuses notions de dépendance faible [voir Doukhan (1994)]. Nous allons ici considérer ici celles qui s'expriment en termes de coefficients de mélange entre des tribus engendrées par le passé et le futur de la suite $(X_i)_{i\in\mathbb{Z}}$. Nous renvoyons à Doukhan et al. (1995) pour une brève historique des résultats de convergence des processus empiriques connus sous diverses conditions de mélange. Nous nous intéressons ici aux suites dites absolument régulières ou encore β-mélangeantes, pour lesquelles les coefficients de mélange sont définis en termes de convergence en variation de la loi du futur après l'instant n conditionnellement au passé avant l'instant 0 vers sa limite. Cette notion de dépendance est bien adaptée à l'étude des processus empiriques, car elle permet l'utilisation de techniques de couplage, qui sont l'analogue non markovien des techniques de renouvellement pour les chaînes de Markov irréductibles, apériodiques et positivement récurrentes.

Les premières extensions des TLC fonctionnels généraux concernent les chaînes de Markov. Levental (1988) a étendu le TLC uniforme de Pollard (1982) aux chaînes de Markov stationnaires, ayant une unique loi invariante et dont les temps de renouvellement ont un moment d'ordre $2+\delta$ fini. Le fait remarquable est que la condition sur la fonction d'entropie universelle est la même que dans le cas indépendant. Ces chaînes sont souvent absolument régulières et la vitesse de décroissance de leurs coefficients de régularité est liée aux propriétés d'intégrabilité de leurs temps de renouvellement [voir Bolthausen (1980, 1982), Tuominen et Tweedie (1994)]: dans certains cas l'existence d'un moment d'ordre $2+\delta$ des temps de renouvellement est équivalente à la condition $\sum n^\delta \beta_n < \infty$, et les coefficients de mélange fort sont alors du même ordre de grandeur ceux de β-mélange.

Les suites β-mélangeantes comprennent cependant d'autres exemples d'intérêt que les chaînes de Markov. Par exemple les processus linéaires sont β-mélangeants sous certaines conditions [Gorodetskii (1982), Pham et Tran (1985) ou Doukhan (1994)]. Récemment, Arcones et Yu (1994) ont obtenu un TLC fonctionnel pour les suites β-mélangeantes générales, mais en se restreignant à des familles de fonctions un peu moins générales, appelées classes sous-graphes de Vapnik-Chervonenkis. Les conditions de mélange imposées sont toujours un peu plus fortes que celles impliquant le TLC usuel pour les sommes partielles normalisées. Le second TLC fonctionnel général est l'extension de Doukhan et al. (1995) du théorème d'Ossiander (1987) aux suites β-mélangeantes. Là encore, les conditions de mélange demandées ne sont pas minimales (voir Doukhan et al. (1995), pages 403–405) sauf pour les classes plongées dans $\mathscr{L}_\infty(P)$ et d'entropie dans $\mathscr{L}_\infty(P)$ intégrable.

Dans cet article, nous allons donner une extension des théorèmes de Kolchinskii et de Pollard aux suites d'observations β-mélangeantes. Nous réussissons ici partiellement: la condition de β-mélange requise est identique à celle du TLC usuel pour les sommes normalisées mais la condition sur l'entropie universelle L^2 de la classe $\mathscr{F}$ est un peu plus forte que dans le cas indépendant.

Pour obtenir ce résultat, nous nous servons d'une inégalité de variance de Viennet (1997), établie à partir de l'inégalité de covariance de Delyon (1990), qui montre l'existence d'une mesure positive et bornée Q pour laquelle l'écart-type de $Z_n(f)$ est majoré par la norme de f dans $\mathscr{L}_2(Q)$. Cette mesure positive Q joue le rôle que P jouait dans le cas indépendant, et sert donc de mesure de référence dans le chaînage restreint. La seconde innovation technique réside dans l'utilisation d'un théorème de couplage maximal de Goldstein (1979), qui nous permet de remplacer la suite de variables initiales successivement par des suites de variables de plus en plus proches d'une suite indépendante au fur et à mesure que nous considérons des réseaux de maille de plus en plus petite.

Dans la section 2, nous donnons le résultat principal, ainsi que quelques applications. La section 3 est consacrée au théorème de couplage maximal et à ses conséquences. Enfin nous montrons la tension du processus empirique dans la section 4.

2. Définitions et résultats

Dans la suite, l'espace probabilisé sous-jacent $(\Omega, \mathscr{T}, \mathbb{P})$ est supposé contenir une variable aléatoire U de loi uniforme sur $[0,1]$, indépendante de la suite $(X_i)_{i\in\mathbb{Z}}$. On suppose de plus l'existence d'une fonction positive et mesurable F de $\mathscr{X}$ dans $\mathbb{R}$ telle que $|f| \leq F$ pour tout élément f de $\mathscr{F}$.

Définition 2.1. *Pour toute mesure positive Q et toute fonction numérique f intégrable sous Q, on pose $E_Q(f) = \int f\,dQ$. Pour tout réel $r \geq 1$, on note $\mathscr{L}_r(Q)$ l'espace des fonctions numériques f telles que $|f|^r$ soit intégrable sous Q et on le munit de la norme usuelle $\|\,.\,\|_{r,Q}$, définie par $\|f\|_{r,Q}^r = \int |f^r| dQ$. On note $\mathscr{L}_\infty(Q)$ l'espace des fonctions essentiellement bornées sous Q.*

Définition 2.2. *Soit d un écart sur $\mathscr{F}$ et soit $\delta > 0$. Une partie $\mathscr{G}$ de $\mathscr{F}$ est dite δ-écartée pour d si, pour tout couple (f,g) d'éléments distincts de $\mathscr{G}$, on a $d(f,g) > \delta$. Une partie δ-écartée est dite maximale si il n'existe pas de partie δ-écartée la contenant strictement.*

Définition 2.3. *Soit* $r \geq 1$ *et* δ *dans* $]0,1]$. *Pour* Q *loi sur* $\mathscr{X}$ *telle que* F *soit dans* $\mathscr{L}_r(Q)$, *on note* $N_r^F(\delta, \mathscr{F}, Q)$ *la borne supérieure des cardinaux des parties finies* $\mathscr{G}$ *de* $\mathscr{F}$ *qui sont* $\delta\|F\|_{r,Q}$*-écartées pour l'écart associé à* $\|\cdot\|_{r,Q}$. *On définit la fonction d'entropie* $H_r^F(\delta, \mathscr{F}, Q)$ *comme le logarithme de* $N_r^F(\delta, \mathscr{F}, Q) \vee 2$. *Soit* $\mathscr{A}(\mathscr{X})$ *l'ensemble des lois de probabilités de support fini sur* $\mathscr{X}$. *On pose*

$$H_r^F(\delta, \mathscr{F}) = \sup\{H_r^F(\delta, \mathscr{F}, Q) : Q \in \mathscr{A}(\mathscr{X})\} \ .$$

Cette fonction est appelée fonction de (r)-entropie universelle de $(\mathscr{F}, F)$.

Comme pour les suites indépendantes, la classe $\mathscr{F}$ doit satisfaire à une condition de mesurabilité. Nous supposerons donc que la classe $\mathscr{F}$ est l'image admissible d'un espace localement compact à base dénombrable $\mathscr{Y}$ muni de sa tribu borélienne $\mathscr{B}(\mathscr{Y})$.

Définition 2.4. Soit $\mathscr{F}$ *est dite image admissible si il existe un espace localement compact à base dénombrable* $\mathscr{Y}$ *muni de sa tribu borélienne* $\mathscr{B}(\mathscr{Y})$ *et une application surjective* T *de* $\mathscr{Y}$ *sur* $\mathscr{F}$ *telle que l'application de* $(\mathscr{X} \times \mathscr{Y}, \mathscr{B}(\mathscr{X}) \otimes \mathscr{B}(\mathscr{Y}))$ *dans* $\mathbb{R}$ *muni de sa tribu borélienne qui à* (x, y) *associe* $T(y)(x)$ *soit mesurable.*

Rappelons la notion de dépendance faible utilisée ici.

Définition 2.5. *Soient* $\mathscr{A}$ *et* $\mathscr{B}$ *deux sous-tribus de* $(\Omega, \mathscr{T}, \mathbb{P})$. *Le coefficient de* β*-mélange (ou d'absolue régularité) de Volkonskii et Rozanov* (1959) *est défini par*

$$\beta(\mathscr{A}, \mathscr{B}) = \frac{1}{2} \sup\left\{\sum_{i \in I} \sum_{j \in J} \left|\mathbb{P}(A_i \cap B_j) - \mathbb{P}(\mathbf{A_i})\mathbb{P}(B_j)\right|\right\} \ ,$$

le maximum étant pris sur toutes les partitions finies $(A_i)_{i \in I}$ et $(B_j)_{j \in J}$ de Ω *composées respectivement d'éléments de* $\mathscr{A}$ *et de* $\mathscr{B}$.

Définition 2.6. *Pour* k *et* l *dans* $\mathbb{Z}$, *on pose* $\mathscr{F}_k = \sigma(X_i : i \leq k)$ *et* $\mathscr{G}_l = \sigma(X_i : i \geq l)$. *Les coefficients* $(\beta_n)_{n \geq 0}$ *de* β*-mélange de la suite stationnaire* $(X_i)_{i \in \mathbb{Z}}$ *sont définis par* $\beta_0 = 1$ *et* $\beta_n = \beta(\mathscr{F}_0, \mathscr{G}_n)$ *pour* $n > 0$.

Rappelons que la variance d'un processus empirique associé à une suite d'observations β-mélangeantes peut différer fortement de la variance d'un processus empirique associé à des observations indépendantes. En particulier la distance usuelle sur $\mathscr{L}_2(P)$ ne majore pas la distance induite par la covariance du processus empirique. Cependant les inégalités de variance de Viennet (1997), résumées dans la proposition ci-dessous, montrent l'existence d'une mesure positive Q

de masse totale finie telle que la distance induite par la covariance du processus empirique soit majorée par la distance usuelle sur $\mathscr{L}_2(Q)$.

Proposition 1. *Pour la suite $(\beta_n)_{n\geq 0}$ des coefficients de β-mélange de la suite $(X_i)_{i\in\mathbb{Z}}$ définie ci-dessus, il existe une suite $(b_i)_{i>0}$ de fonctions mesurables de $\mathscr{X}$ dans $[0,1]$, telle que $\mathbb{E}_P(b_i) \leq \beta_i$ et, pour tout $i > 0$ et toute fonction f dans $\mathscr{L}_2(P)$,*

$$(a) \qquad |\mathrm{Cov}(f(X_0), f(X_i))| \leq 2\int_{\mathscr{X}} b_i f^2\, dP \ .$$

Supposons que $\sum \beta_n$ converge et posons $B = 1 + 4\sum_{i>0} b_i$. Alors B est dans $\mathscr{L}_1(P)$ et $Q = BP$ est une mesure positive de masse finie telle que, pour toute f dans $\mathscr{L}_2(Q)$,

$$(b) \quad \mathrm{Var}\, Z_n(f) \leq \mathrm{Var}\, f(X_0) + 2\sum_{i>0} |\mathrm{Cov}(f(X_0), f(X_i))| \leq \int_{\mathscr{X}} f^2 dQ \ .$$

Soit $\Gamma(f,f) = \mathrm{Var}\, f(X_0) + 2\sum_{i>0}\mathrm{Cov}(f(X_0), f(X_i))$. *Alors, pour f dans $\mathscr{L}_2(Q)$,*

$$(c) \qquad \lim_{n\to\infty} \mathrm{Var}\, Z_n(f) = \Gamma(f,f) \text{ et } \Gamma(f,f) \leq \|f\|_{2,Q}^2 \ .$$

Bien que cette proposition soit montrée dans Viennet (1997), il est important pour nous d'avoir une forme explicite des variables b_i à l'aide du théorème de couplage maximal. Nous construirons donc les variables b_i dans la section 3. Les résultats ci-dessous seront établis uniquement pour les variables b_i définies dans la section 3 par (3.5).

Sous la contrainte de β-mélange $\sum_n \beta_n < \infty$ le TLC d'Ibragimov et Linnik (1971) entraîne la convergence marginale dans $\mathscr{L}_\infty(P)$ (voir Hall et Heyde (1980) pour une preuve complète de ce TLC). Soit $\mathscr{F}$ une classe de fonctions incluse dans $\mathscr{L}_2(Q)$. En partant de la proposition 1 et en appliquant la méthode de troncature d'Ibragimov et Linnik (1971), il est facile de montrer la convergence des marginales de dimension finie de $\{Z_n(f) : f \in \mathscr{F}\}$ vers celles d'un vecteur gaussien de fonction de covariance Γ. La question naturelle est donc de savoir si il est possible d'étendre le TLC uniforme de Pollard (1982) pour les suites indépendantes aux classes de fonctions d'entropie universelle L^2 intégrable ayant une fonction enveloppe F dans $\mathscr{L}_2(Q)$. Cependant un tel résultat semble assez difficile d'accès. Comme Doukhan et al. (1995), nous demanderons à F d'être dans le sous-espace de $\mathscr{L}_2(Q)$ défini ci-dessous.

Définition 2.7. *Supposons que $\sum_n \beta_n < \infty$. Soit $\beta^{-1}(u) = \inf\{k \in \mathbb{N} : \beta_k \leq u\}$ la fonction inverse de la fonction $\beta : x \to \beta_{[x]}$. On* *note* Q_f *la fonction de quantile* de $|f(X_0)|$, *qui est l'inverse càdlàg de la fonction* $t \to \mathbb{P}(|f(X_0)| > t)$. *Soit* $\mathscr{L}_{2,\beta}(P)$ *l'espace des fonctions numériques* f *telles que*

$$\|f\|_{2,\beta}^2 = \int_0^1 \beta^{-1}(u) Q_f^2(u) du < \infty \ . \tag{2.1}$$

Notons $\|f\|_{2,\beta}$ le réel positif défini par (2.1). Alors $(\mathscr{L}_{2,\beta}(P), \| \, . \, \|_{2,\beta})$ est un espace vectoriel normé contenant $\mathscr{L}_\infty(P)$ (voir Doukhan et al. (1995) pour plus de détails). Si Q est la mesure définie dans la proposition 1, alors (voir Viennet (1997) lemme 2.2.2)

$$\|f\|_{2,Q} \leq 2\|f\|_{2,\beta} \text{ et par conséquent } \mathscr{L}_{2,\beta}(P) \subset \mathscr{L}_2(Q) \ . \tag{2.2}$$

Donnons maintenant un TLC uniforme sous une condition d'entropie universelle L^2 pour des classes de fonctions non bornées. Ce théorème est la conséquence d'une majoration des fluctuations donnée dans la section 4. Cette majoration permet aussi d'obtenir un TLC uniforme sous la condition d'entropie universelle de Pollard, quitte à renforcer les hypothèses sur les coefficients de mélange et la fonction enveloppe.

Théorème 1. *Soit* $(X_i)_{i\in\mathbb{Z}}$ *une suite stationnaire de variables aléatoires de loi* P, *satisfaisant la condition de mélange* $\sum_{n>0} \beta_n < \infty$. *Soit* $\mathscr{F}$ *une classe de fonctions numériques image admissible. Supposons qu'il existe une fonction* F *dans* $\mathscr{L}_{2,\beta}(P)$ *pour laquelle* $|f| \leq F$ *pour toute* f *dans* $\mathscr{F}$ *et que*

$$\int_0^1 \sqrt{H_2^F(x, \mathscr{F}) \log(1/x)}\, dx < \infty \ .$$

Alors il existe une suite $(Z^{(n)})_{n>0}$ *de processus gaussiens indexés par* $\mathscr{F}$, *de fonction de covariance* Γ *et à trajectoires p.s. uniformément continues pour la métrique induite sur* $\mathscr{F}$ par $\mathscr{L}_2(Q)$, *telle que*

$$\sup_{f\in\mathscr{F}} \left| Z_n(f) - Z^{(n)}(f) \right| \xrightarrow{\mathbb{P}} 0 \textit{ quand } n \to \infty \ .$$

Nous allons maintenant comparer ce théorème avec les résultats existants.

Supposons que $\mathscr{F} = \{\mathbb{I}_S : S \in \mathscr{S}\}$ pour une certaine classe de parties $\mathscr{S}$ et que la fonction $H_2^1(\, . \,, \mathscr{F})$ est finie au voisinage de 0. Alors $\mathscr{S}$ est une classe de parties possédant la propriété combinatoire de Vapnik et Chervonenkis. Un résultat de Dudley (1978) assure alors que $H_2^1(x, \mathscr{F}) = \mathscr{O}(\log(1/x))$ quand $x \to 0$. Le théorème 1 s'applique donc dans ce cas, et améliore des résultats antérieurs de Arcones et Yu (1994) et de Doukhan et al. (1995) pour les classes de parties.

Prenons maintenant l'exemple des classes de sous-graphes Vapnik-Chervonenkis, dont la fonction enveloppe F est dans $\mathscr{L}_{2,\beta}(P)$. Pour ces classes $H_2^F(x,\mathscr{F}) = \mathscr{O}(\log(1/x))$ quand $x \to 0$. Le théorème 1 donne donc le TLC uniforme sous des conditions de mélange plus faibles que celles proposées par Arcones et Yu (1994): si F est dans $\mathscr{L}_r(P)$, alors la condition de mélange s'écrit $\sum_n n^{2/(r-2)}\beta_n < \infty$ et, si les coefficients de β-mélange décroissent géométriquement, alors F est dans $\mathscr{L}_{2,\beta}(P)$ dès que $E_P(F^2 \log^+ F) < \infty$.

Soit maintenant une classe de fonctions $\mathscr{F}$ plongée dans $\mathscr{L}_\infty(P)$, dont la fonction d'entropie dans $\mathscr{L}_\infty(P)$ satisfait la condition intégrale

$$\int_0^1 \sqrt{H_\infty(x,\mathscr{F})\log(1/x)}\,dx < \infty \ .$$

Alors $\mathscr{F}$ satisfait TLC uniforme. Cependant le théorème 1 dans Doukhan et al. (1995) donne le même résultat sous une condition d'entropie plus faible. En fait la sous-optimalité du théorème 1 pour l'entropie dans $\mathscr{L}_\infty(P)$ provient d'un choix inadéquat des paramètres du chaînage et non de la technique de couplage maximal.

3. Couplage maximal

Le TLC fonctionnel résulte de la convergence des marginales de dimension finie, qui est une conséquence du TLC de Doukhan et al. (1994) et de la tension du processus empirique sous la métrique induite sur $\mathscr{F}$ par $\mathscr{L}_2(Q)$ (voir Pollard (1990), théorème 10.2).

Pour établir la tension, nous allons utiliser le théorème de couplage maximal (Goldstein (1979), théorème 3.3) ci-dessous.

Théorème 2. *Soit $(\xi_i)_{i\in\mathbb{Z}}$ une suite de variables aléatoires à valeurs dans un espace polonais $\mathscr{X}$. Supposons que $(\Omega,\mathscr{T},\mathbb{P})$ contient une variable U de loi uniforme sur $[0,1]$, indépendante de $(\xi_i)_{i\in\mathbb{Z}}$. Alors on peut construire une suite $(\xi_i^*)_{i\in\mathbb{Z}}$ de même loi que la suite $(\xi_i)_{i\in\mathbb{Z}}$, indépendante de $\mathscr{A}_0 = \sigma(\xi_i : i \le 0)$, mesurable pour la tribu engendrée par U* et *$(\xi_i)_{i\in\mathbb{Z}}$, et couplée avec la suite initiale ainsi. Soit $\mathscr{B}_n = \sigma(\xi_i : i \ge n)$. Pour tout entier $n > 0$,*

$$\mathbb{P}(\xi_k = \xi_k^* \ \textit{pour tout} \ k \ge n) = 1 - \beta(\mathscr{A}_0,\mathscr{B}_n) \ .$$

3.1. Couplage maximal et processus empiriques

Les critères de tension récents pour les processus empiriques β-mélangeants sont fondés sur une forme affaiblie du théorème 2: la suite $(X_i)_{i\in\mathbb{Z}}$ est remplacée par une suite de variables ayant des propriétés d'indépendance convenables. On majore alors l'erreur due à la substitution ainsi que les fluctuations du processus empirique associé à la nouvelle suite. Comme les auteurs précédents, nous allons commencer par cette substitution en un coup.

Soit K un entier naturel à débattre et $q_0 = 2^K$. En appliquant récursivement le théorème 2, il est facile de constuire une suite $(X_i^0)_{i>0}$ de variables aléatoires ayant les propriétés suivantes. Si $q = q_0$, alors

1. Pour tout i positif, $U_i^0 = (X_{iq+1}^0, \ldots, X_{iq+q}^0)$ a même loi que $U_i = (X_{iq+1}, \ldots, X_{iq+q})$.
2. Les blocs $(U_{2i}^0)_{i\geq 0}$ sont indépendants ainsi que les blocs $(U_{2i+1}^0)_{i\geq 0}$.
3. De plus $\mathbb{P}(U_i \neq U_i^0) \leq \beta_q$ pour tout $i \geq 0$.

L'erreur de substitution se majore alors grâce au lemme suivant dans le cas des classes de fonctions bornées (voir Doukhan et al. (1995) pages 407–408 pour la preuve).

Lemme 1. *Posons* $S_n^0(f) = f(X_1^0) + \cdots + f(X_n^0)$ et $Z_n^0(f) = n^{-1/2}(S_n^0(f) - nE_P(f))$. *Soit* $\mathscr{F}_M$ *une classe de fonctions de* $\mathscr{X}$ *dans* $[-M, M]$*, image admissible. Alors*

$$\mathbb{E}\left(\sup_{f\in\mathscr{F}_M}\left|Z_n(f) - Z_n^0(f)\right|\right) \leq 2\sqrt{n}\,M\beta_q \ .$$

Pour un choix convenable de q (voir section 4), le majorant du lemme 1 tendra vers 0 quand n tend vers l'infini, et nous serons donc ramenés au module de continuité de Z_n^0.

La différence essentielle avec le cas indépendant réside dans les contrôles exponentiels pour les variables $|Z_n^0(f) - Z_n^0(g)|$ pour f et g bornées par M et proches dans $\mathscr{L}_2(Q)$. Pour les variables indépendantes, l'inégalité de Bernstein (voir Pollard (1984) pages 191–194) s'applique avec M comme borne commune des variables et donc l'oscillation du processus Z_n sur des réseaux de petite maille est convenablement majorée (voir Pollard (1984) pour une description détaillée de la méthode de chaînage restreint utilisée dans le cas indépendant). Par contre, pour le processus Z_n^0 obtenu dans le cas β-mélangeant, l'inégalité de Bernstein s'applique avec q_0M comme borne commune des variables. Ce nombre peut être très proche de $\sqrt{n}$. Une application sans discernement de la méthode de chaînage restreint donne alors de trop grandes erreurs sur les réseaux de petite maille. Pour remédier à cet inconvénient, nous allons utiliser le théorème 2 pour remplacer progressivement le processus empirique Z_n^0 par des processus Z_n^k associés à des suites de blocs indépendants de largeur $q_k = q_0 2^{-k}$ au fur et à mesure que nous prendrons des réseaux de maille plus en plus petite. Nous pourrons alors utiliser l'inégalité de Bernstein avec Mq_k comme borne commune des variables à la k-ième étape du chaînage et obtenir ainsi la tension sous des conditions moins restrictives.

Nous devons donc construire de proche en proche des suites $(X_i^k)_{i>0}$ de variables possédant des propriétés analogues aux propriétés 1, 2 et 3 avec $q = q_0 2^{-k}$. Pour ce faire, nous allons nous placer dans un bloc $U_i^0 = (X_{iq+1}^0, \dots, X_{iq+q}^0)$ et construire les variables $(X_{iq+1}^k, \dots, X_{iq+q}^k)$ par induction sur k à partir du bloc initial U_i^0 et d'un stock d'innovations indépendantes de loi uniforme sur $[0, 1]$. Le procédé de construction sera identique à l'intérieur de chaque bloc. Il nous suffit donc de détailler la construction effectuée dans le premier bloc.

Soit $q = q_0$. Prenons $i = 0$ et construisons la suite $(X_1^1, \dots, X_q^1)$ à partir de U_0^0 et de variables auxiliaires. Soit $(u_i)_{i>0}$ une suite de variables indépendantes de loi uniforme sur $[0, 1]$, indépendante des suites de v.a. ci-dessus. D'après le lemme de Skorohod (1976), il existe une suite $(X_i^{0*})_{i\in\mathbb{Z}}$ ayant même loi que $(X_i)_{i\in\mathbb{Z}}$, telle que les variables de cette suite sont des fonctions mesurables de $(u_0, X_1^0, \dots, X_q^0)$ et $(X_1^0, \dots, X_1^q) = (X_1^{0*}, \dots, X_q^{0*})$ p.s. En appliquant le théorème 2 de couplage maximal nous allons remplacer la seconde moitié de la suite $(X_1^0, \dots, X_q^0)$ par une suite de variables ne dépendant pas des variables définies antérieurement. Appliquons le théorème 2 aux variables $\xi_i = X_{i-q_1}^{0*}$. Nous obtenons ainsi une suite $(\xi_i^*)_{i\in\mathbb{Z}}$ de variables aléatoires indépendante de la suite X^{0*}. La suite $(X_i^{1*})_{i\in\mathbb{Z}}$ est alors définie par $X_i^{1*} = \xi_{i+q_1}^*$ pour tout $i \in \mathbb{Z}$. Pour finir, on pose $X_i^1 = X_i^0$ pour i dans $[1, q_1]$ et $X_i^1 = X_i^{1*}$ pour $i \in [q_1 + 1, 2q_1]$.

Comme les suites finies $(X_i^1)_{i\in[1,q_1]}$ et $(X_{i+q_1}^1)_{i\in[1,q_1]}$ sont indépendantes et ont même loi que $(X_i^0)_{i\in[1,q_1]}$, nous pouvons reprendre ce procédé pour la suite $(X_i^1)_i$ sur les deux sous-intervalles de longueur $q_1 = q_0/2$ à l'aide de nouvelles variables auxiliaires indépendantes de toutes les variables ci-dessus. En procédant par induction, on montre aisément la proposition suivante.

Proposition 2. *Soit K entier naturel et $q_0 = 2^K$. On peut construire une famille $(X_i^k)_{i>o}$ de suites de variables aléatoires par induction sur k dans $[0, K]$ ayant les propriétés suivantes:*

(i) *Soit $q = q_0$. Posons $W_i = (X^k_{iq+1}, \dots, X^k_{iq+q})_{k\in[0,K]}$. Alors les blocs $(W_i)_{i\geq 0}$ sont équidistribués. De plus les blocs $(W_{2i})_{i\geq 0}$ sont indépendants ainsi que les $(W_{2i+1})_{i\geq 0}$.*

(ii) *Posons $W^k_i = (X^{k-1}_{2iq_k+1}, \dots, X^{k-1}_{(2i+2)q_k}, X^k_{2iq_k+1}, \dots, X^k_{(2i+2)q_k})$. Alors les $(W^k_i)_{i\geq 0}$ sont équidistribués. De plus, si $(X^*_i)_{i\in\mathbb{Z}}$ est la suite construite à partir de $(X_i)_{i\in\mathbb{Z}}$ à l'aide du théorème 2, ces blocs ont même loi que $(X_{1-q_k}, \dots, X_{q_k}, X_{1-q_k}, \dots, X_0, X^*_1, \dots X^*_{q_k})$.*

(iii) *Les blocs $W^k_1, \dots, W^k_{2^{k-1}}$ sont indépendants.*

Cette proposition est à la base du calcul de tension de la section 4.

3.2. Couplage maximal et inégalités de variance

Montrons (a) de la proposition 1 en appliquant le théorème 2 de couplage maximal à la suite $(X_i)_{i\in\mathbb{Z}}$. Soit $(X^*_i)_{i\in\mathbb{Z}}$ la suite ainsi construite. Pour tout $i > 0$, les variables X^*_i et X_0 sont indépendantes. Par conséquent

$$\operatorname{Cov}(f(X_0), f(X_i)) = \mathbb{E}(f(X_0)(f(X_i) - f(X^*_i))\mathbb{I}_{X_i\neq X^*_i}) \ . \tag{3.1}$$

Donc

$$|\operatorname{Cov}(f(X_0), f(X_i))| \leq \mathbb{E}(\ (|f(X_0)f(X_i)| + |f(X_0)f(X^*_i)|)\mathbb{I}_{X_i\neq X^*_i}) \ ,$$

puis, en notant que $2|f(x)f(y)| \leq f^2(x) + f^2(y)$,

$$\begin{aligned} 2|\operatorname{Cov}(f(X_0), f(X_i))| \leq \mathbb{E}(2f^2(X_0)&\mathbb{I}_{X_i\neq X^*_i} \\ &+ f^2(X_i)\mathbb{I}_{X_i\neq X^*_i} + f^2(X^*_i)\mathbb{I}_{X_i\neq X^*_i}) \ . \end{aligned} \tag{3.2}$$

Soient b_{0i}, b_{i0} et b^*_{i0} les fonctions mesurables de $\mathscr{X}$ dans $[0,1]$ définies par

$$\begin{aligned} b_{0i}(X_0) = \mathbb{E}(\mathbb{I}_{X_i\neq X^*_i} | X_0);\ b_{i0}(X_i) = \mathbb{E}(\mathbb{I}_{X_i\neq X^*_i} | X_i); \\ b^*_{i0}(X^*_i) = \mathbb{E}(\mathbb{I}_{X_i\neq X^*_i} | X^*_i) \ . \end{aligned} \tag{3.3}$$

Ces fonctions ont une moyenne sous P inférieure ou égale à β_i, d'après (b) du théorème 2. De plus on montre facilement en partant de (3.2) que

$$2|\operatorname{Cov}(f(X_0), f(X_i))| \leq E_P(f^2(2b_{0i} + b_{i0} + b^*_{i0})) \ . \tag{3.4}$$

Par conséquent la variable b_i définie par

$$b_i = (2b_{0i} + b_{i0} + b^*_{i0})/4 \tag{3.5}$$

satisfait (a) de la proposition 1. Dans la suite, les variables b_i sont définies par (3.5).

4. Tension des processus empiriques β-mélangeants

Pour montrer le théorème 1, nous devons montrer l'équicontinuité asymptotique de Z_n pour la métrique induite par $\mathscr{L}_2(Q)$. Quitte à diviser les fonctions par une constante, on peut supposer que $\|F\|_{2,\beta} = 1/2$, auquel cas $\|F\|_{2,Q} \le 1$.

Notation 4.1. On pose $d(f,g) = \|f-g\|_{2,Q}$ et $H(x) = H_2^F(x,\mathscr{F})$.

La fonction $H(x)$ ainsi définie est un majorant de la fonction d'entropie de $(\mathscr{F},d)$. Pour montrer la tension, nous allons remplacer H par une fonction $\mathbb{H}$ plus régulière à l'aide du lemme suivant (voir Doukhan et al. (1995), page 410).

Lemme 2. *Soit H une fonction décroissante de $[0,1]$ dans $\mathbb{R}^+$. Alors il existe une fonction décroissante $\mathbb{H}$ majorant H, telle que $x \to x^4\mathbb{H}(x)$ soit croissante et, pour tout v positif,*

$$\varphi(v) = \int_0^v \sqrt{\mathbb{H}(x)}\,dx \le 4\int_0^v \sqrt{H(x)}\,dx \ .$$

Nous allons définir des fonctions tronquées ainsi.

Notation 4.2. Soit $M \ge 1$, $f_M = f\mathbb{I}_{F\le M}$ et $\mathscr{F}_M = \{f_M : f \in \mathscr{F}\}$.

Afin de mettre en oeuvre la technique de chaînage restreint, nous allons définir des projections sur les réseaux finis de maille δ et de cardinal au plus $H(\delta)$ comme suit.

Définition 4.1. *Soit δ dans $]0,1[$ et $\mathscr{F}_\delta$ une partie de $\mathscr{F}$ δ-écartée maximale pour d. Soit Π_δ une application de $\mathscr{F}$ dans $\mathscr{F}_\delta$ telle que $d(f,\Pi_\delta f) \le \delta$ pour toute f dans $\mathscr{F}$, ayant la propriété de mesurabilité suivante: l'application de $(\mathscr{X}\times\mathscr{Y},\mathscr{B}(\mathscr{X})\otimes\mathscr{B}(\mathscr{Y}))$ dans $\mathbb{R}$ muni de sa tribu borélienne qui à (x,y) associe $\Pi_\delta T(y)(x)$ est mesurable (une telle selection mesurable existe: voir Dudley (1984), page 120). On note $\Pi_\delta f_M = (\Pi_\delta f)_M$*

Le théorème 1 s'obtient à partir de l'évaluation suivante du module de continuité du processus empirique tronqué par des arguments identiques à ceux donnés dans Doukhan et al. (1995) pour passer du cas borné au cas non borné. Nous nous contenterons donc de montrer l'analogue suivant du théorème 2 de Doukhan et al. (1995).

Proposition 3. *Soit $L(x) = \log(x \vee e)$. Sous les hypothèses du théorème 1 et sous l'hypothèse supplémentaire $\|F\|_{2,\beta} = 1/2$, il existe une constante universelle C_0, telle que pour tout entier q dans $[1,n]$ puissance de 2 et tout δ dans $]0,1[$,*

$$\mathbb{E}\left(\sup_{f\in\mathscr{F}}|Z_n(f_M-\Pi_\delta f_M)|\right)\le C_0\left(1+\frac{qM\mathbb{H}(\delta)}{\delta\sqrt{n}}\right)\int_0^\delta\sqrt{H(x)L(\delta/x)}\,dx$$
$$+4\sqrt{n}M\beta_q\ .$$

Preuve. Pour montrer ce résultat, nous allons nous servir d'un chaînage restreint.

Définition 4.2. *Soit* $K>0$ *entier et* $q=q_0=2^K$. *On pose* $S_n^k(f)=f(X_1^k)+\cdots+f(X_n^k)$ *et* $Z_n^k(f)=n^{-1/2}(S_n^k(f)-nP(f))$, *les variables* X_i^k *étant celles de la proposition 2. Soit* $\delta_k=\delta 2^{-k/3}$. *On pose* $\Pi_K=\Pi_{\delta_K}$ *et on définit récursivement les projections* $(\Pi_k)_{k\in[0,K[}$ *par* $\Pi_k=\Pi_{\delta_k}\circ\Pi_{k+1}$.

Un calcul élémentaire montre que

$$\begin{aligned}Z_n(f_M-\Pi_0 f_M)&=(Z_n-Z_n^0)(f_M-\Pi_0 f_M)+Z_n^K(f_M-\Pi_K f_M)\\&\quad+\sum_{l=1}^K Z_n^{l-1}(\Pi_l f_M-\Pi_{l-1}f_M)\\&\quad+\sum_{k=1}^K(Z_n^{k-1}-Z_n^k)(f_M-\Pi_k f_M)\ .\end{aligned}\tag{4.2}$$

Donc, en appliquant le lemme 1 (rappelons que $q_0=2^K$),

$$\mathbb{E}\left(\sup_{f\in\mathscr{F}}|Z_n(f_M-\Pi_\delta f_M)|\right)\le 4\sqrt{n}M\beta_{q_0}+\mathbb{E}_1+\mathbb{E}_2+\mathbb{E}_3\ ,\tag{4.3}$$

où

$$\mathbb{E}_1=\sum_{l=1}^K\mathbb{E}\left(\sup_{f\in\mathscr{F}}\left|Z_n^{l-1}(\Pi_l f_M-\Pi_{l-1}f_M)\right|\right),$$
$$\mathbb{E}_2=\mathbb{E}\left(\sup_{f\in\mathscr{F}}\left|Z_n^K(f_M-\Pi_K f_M)\right|\right)\ \text{et}$$
$$\mathbb{E}_3=\sum_{k=1}^K\mathbb{E}\left(\sup_{f\in\mathscr{F}}\left|\left(Z_n^{k-1}-Z_n^k\right)(f_M-\Pi_k f_M)\right|\right)\ .$$

Notons que, puisque $\mathscr{F}$ est image admissible, il résulte de la définition 2.4, de la propriété de mesurabilité de la définition 4.1 et du théorème T32, page 17, dans Dellacherie (1972) que les v.a. ci-dessus sont effectivement mesurables: on peut donc considérer leurs espérances.

4.1. Majoration de $\mathbb{E}_1$

Majorons

$$\mathbb{E}_{1,l} = \mathbb{E}\left(\sup_{f\in\mathscr{F}} |Z_n^l(\Pi_{l+1}f_M - \Pi_l f_M)| \right)$$

pour l dans $[0, K-1]$ fixé. D'après la proposition 2, Les blocs $(X^l_{iq_l+1}, \dots, X^l_{(i+1)q_l})_{i\in[1,2^l]}$ sont indépendants et chaque bloc a pour loi jointe la loi jointe de q_l variables consécutives de la suite $(X_i)_{i\in\mathbb{Z}}$. En appliquant (i) de la proposition 2, nous pouvons ainsi décomposer la mesure empirique Z_n^l en une somme de deux mesures empiriques $Z^l_{n,1}$ et $Z^L_{n,2}$ obtenues à partir de blocs indépendants de variables de largeur q_l, chaque bloc ayant la loi de $(X_1, \dots, X_{q_l})$ (sauf le dernier).

Posons $g = \Pi_{l+1}f_M - \Pi_l f_M$ pour l'instant. La somme $S_n^l(g)$ se décompose en somme de deux sommes de variables aléatoires indépendantes de même loi que $g(X_1) + \cdots + g(X_{q_l})$, sauf la variable relative au dernier bloc, qui a la loi de $g(X_1) + \cdots + g(X_{n-q_l[n/q_l]})$. En tout cas, ces variables sont bornées par $2Mq_l$ et de variance majorée par la largeur du bloc multipliée par $\|g\|^2_{2,Q}$, d'après la proposition 1. Or, par définition des projections Π_k,

$$\|g\|^2_{2,Q} = E_Q(\mathbb{I}_{F\le M}(\Pi_{l+1}f - \Pi_l f)^2) \le \delta_l^2 \quad \text{pour tout } f \in \mathscr{F} \ .$$

Donc, en appliquant deux fois le corollaire A.1 de l'annexe et en notant que la famille $\{\Pi_{l+1}f_M - \Pi_l f_M : f \in \mathscr{F}\}$ est de cardinal inférieur à $\exp(\mathbb{H}(\delta_{l+1}))$,

$$\mathbb{E}_{1,l} \le 4\delta_l\sqrt{\mathbb{H}(\delta_{l+1})} + 4n^{-1/2}q_l M\mathbb{H}(\delta_{l+1}) \ . \tag{4.4}$$

Le premier terme de (4.4) est une contribution infinitésimale à l'intégrale d'entropie. Comparons le second terme au premier terme:

$$\frac{q_l\mathbb{H}(\delta_{l+1})}{\delta_l\sqrt{n\mathbb{H}(\delta_{l+1})}} = \frac{q_l}{\delta_l\delta^2_{l+1}} \, \frac{\delta^2_{l+1}\sqrt{\mathbb{H}(\delta_{l+1})}}{\sqrt{n}} \ .$$

Comme $\delta_k = \delta_0 2^{-k/3}$ et $q_k = q_0 2^{-k}$, le premier rapport du terme de droite est constant. Comme la fonction $x \to x^2\sqrt{\mathbb{H}(x)}$ est croissante le second rapport est décroissant en l. Donc le rapport du terme de droite sur le terme de gauche dans (4.4) décroit quand l croit. Il en résulte que

$$\mathbb{E}_1 \le 4\Big(1 + n^{-1/2} q_0 M \delta_0^{-1} \sqrt{\mathbb{H}(\delta_1)}\Big)$$

$$\Big(\delta_0 \sqrt{\mathbb{H}(\delta_1)} + \cdots + \delta_{K-1} \sqrt{\mathbb{H}(\delta_1)}\Big) \ . \tag{4.5}$$

Comme les réels δ_k sont en progression géométrique de raison $2^{-1/3}$ et $x \to x^2\sqrt{\mathbb{H}(x)}$ croit, $\sqrt{\mathbb{H}(\delta_1)} \le 2^{2/3}\sqrt{\mathbb{H}(\delta_0)}$. Puisque $\delta_0 = \delta$, il est aisé d'en déduire que

$$\mathbb{E}_1 \le 25\Big(1 + 2n^{-1/2} q M \delta^{-1} \sqrt{\mathbb{H}(\delta)}\Big)\varphi(\delta) \ . \tag{4.6}$$

4.2. Majoration de $\mathbb{E}_3$

Notation 4.3. Soit $S_{n,1}^k(f)$ la variable obtenue en sommant les variables $f(X_i^k)$ sur l'ensemble des indices i dans $[1, n]$ tels que $[(i-1)/q_0]$ soit pair et $S_{n,2}^k(f) = S_n^k(f) - S_{n,1}^k(f)$.

Fixons k dans $[1, K]$ et majorons le k-ième terme de la somme constituant $\mathbb{E}_3$. Pour majorer ce terme, il suffit de majorer les deux quantités suivantes séparément:

$$\mathbb{E}_{3,k}^s = n^{-1/2}\mathbb{E}(\sup_{f \in \mathscr{F}}(|(S_{n,s}^{k-1} - S_{n,s}^k)(f_M - \Pi_k f_M)|) \text{ avec } s = 1 \text{ ou } s = 2 \ . \tag{4.7}$$

Posons $I_j = [(2j+1)q_k, (2j+2)q_k \wedge n] \cap \mathbb{N}$. Soit

$$V_j^k(h) = \sum_{i \in I_j} (h(X_i^k) - h(X_i^{k-1})) \ . \tag{4.8}$$

Alors $(S_{n,s}^{k-1} - S_{n,s}^k)(h) = \sum_j V_j^k(h)$, la somme étant prise sur les indices j tels que $[j2^{-k}]$ soit de la même parité que $s-1$. Fixons s. La proposition 2 assure que les variables $\{V_j^k(f_M - \Pi_k f_M) : f \in \mathscr{F}\}$ sont indépendantes dans leur ensemble quand j décrit l'ensemble d'indices décrit ci-dessus. Puisque la classe $\mathscr{F}$ est image admissible, les inégalités de symétrisation s'appliquent (voir Dudley (1984) pour la mesurabilité des v.a. considérées) et donc, si $(\varepsilon_j)_j$ est une suite de signes symétriques indépendants, indépendante de la tribu engendrée par toutes les variables définies auparavant,

$$\mathbb{E}_{3,k}^s \le 2n^{-1/2}\mathbb{E}\left(\sup_{f \in \mathscr{F}} \left| \sum \varepsilon_j V_j^k(f_M - \Pi_k f_M)\right|\right) = 2w_k(\delta_k) \ . \tag{4.9}$$

Il suffit donc de majorer $w_k(\delta_k)$. Pour cela, nous allons appliquer la méthode de chaînage conditionnellement à la tribu engendrée par les variables (X_i^k, X_i^{k-1}).

Donnons-nous des suites $(x_i^k)_{i>0}$ et $(x_i^{k-1})_{i>0}$ d'éléments de $\mathscr{X}$ et posons

$$v_j^k(h) = \sum_{i\in I_j} (h(x_i^k) - h(x_i^{k-1})) \ .$$

Soit $\mathscr{G}$ une famille de fonctions, pouvant dépendre de $(x_1^k, \dots, x_n^k, x_1^{k-1}, \dots, x_n^{k-1}))$, de cardinal $\exp(H)$. En appliquant le lemme A.1. avec $L(t) = \frac{1}{2} t^2 \sup\{\sum (v_j^k(g))^2 : g \in \mathscr{G}\}$, on obtient

$$\mathbb{E}\left(\sup_{g\in\mathscr{G}} \left|\sum \varepsilon_j v_j^k(g)\right|\right) \le 2\sqrt{H} \sup_{g\in\mathscr{G}} \sqrt{\sum (v_j^k(g))^2} \ . \tag{4.10}$$

Notons que la somme constituant $v_j^k(h)$ peut être restreinte à l'ensemble B_j des entiers i de I_j tels que $x_i^k \neq x_i^{k-1}$. Si B_j est de cardinal n_j, alors, par l'inégalité de Cauchy-Schwarz,

$$\sum_j (v_j^k(g))^2 \le \sum_j 2n_j \sum_{i\in B_j} (g^2(x_i^k) + g^2(x_i^{k-1})) \ . \tag{4.11}$$

La mesure de référence Q_n^k choisie pour le chaînage sera donc

$$Q_n^k = n^{-1} \sum_j 2n_j \sum_{i\in B_j} (\delta_{x_i^k} + \delta_{x_i^{k-1}}) \ . \tag{4.12}$$

Soit $\mathscr{G}_k$ une partie de $\mathscr{F}$, $\delta_k \|F\|_{2,Q_n^k}$-écartée maximale pour $\|\,.\,\|_{2,Q_n^k}$ (noter que $\mathscr{G}_k$ est de cardinal $\exp(\mathbb{H}(\delta_k))$ au plus). Soit Π'_k une application de $\mathscr{F}$ dans $\mathscr{G}_k$ telle que

$$\|f - \Pi'_k f\|_{2,Q_n^k} \le \delta_k \|F\|_{2,Q_n^k} \text{ pour toute } f \in \mathscr{F} \ . \tag{4.13}$$

D'après l'inégalité triangulaire,

$$\begin{aligned} &\mathbb{E}\left(\sup_{f\in\mathscr{F}} \left|\sum \varepsilon_j v_j^k(f_M - \Pi_k f_M)\right|\right) \\ &\le \mathbb{E}\left(\sup_{f\in\mathscr{F}} \left|\sum \varepsilon_j v_j^k((f - \Pi'_k f)\mathbb{I}_{F\le M})\right|\right) \\ &\quad + \mathbb{E}\left(\sup_{f\in\mathscr{F}} \left|\sum \varepsilon_j v_j^k(\Pi'_k f_M - \Pi_k f_M)\right|\right) \ . \end{aligned} \tag{4.14}$$

En appliquant la méthode de chaînage usuelle (voir Pollard (1990), théorème 3.5 page 12) au premier terme, il est facile de montrer que

$$(4.15)\quad \mathbb{E}\left(\sup_{f\in\mathscr{F}}\left|\sum \varepsilon_j v_j^k((f-\Pi'_k f)\mathbb{I}_{F\le M})\right|\right) \le 9\sqrt{n}\|F\|_{2,Q_n^k}\varphi(\delta_k/2)\ .$$

Pour le second terme, on note que la famille $\{\Pi'_k f_M - \Pi_k f_M : f\in\mathscr{F}\}$ est finie et de cardinal $\exp(2\mathbb{H}(\delta_k))$ au plus. Par conséquent, d'après le lemme A.1 appliqué comme précédemment,

$$\mathbb{E}\left(\sup_{f\in\mathscr{F}}\left|\sum \varepsilon_j v_j^k(\Pi'_k f_M - \Pi_k f_M)\right|\right)$$
$$(4.16)\qquad \le 4\sqrt{\mathbb{H}(\delta_k)\sup_{f\in\mathscr{F}}\sum(v_j^k(\Pi'_k f_M - \Pi_k f_M))^2}\ .$$

Mais, par l'inégalité triangulaire suivie de (4.11) et (4.13)

$$\sup_{f\in\mathscr{F}}\sqrt{\sum(v_j^k(\Pi'_k f_M - \Pi_k f_M))^2}$$
$$(4.17)\qquad \le \delta_k\sqrt{n}\|F\|_{2,Q_n^k} + \sup_{f\in\mathscr{F}}\sqrt{\sum(v_j^k(f_M - \Pi_k f_M))^2}\ .$$

Donc

$$\mathbb{E}\left(\sup_{f\in\mathscr{F}}\left|\sum \varepsilon_j v_j^k(f_M - \Pi_k f_M)\right|\right)$$
$$\le 9\sqrt{n}\|F\|_{2,Q_n^k}\varphi(\delta_k)$$
$$(4.18)\qquad + 4\sqrt{\mathbb{H}(\delta_k)}\sup_{f\in\mathscr{F}}\sqrt{\sum(v_j^k(f_M - \Pi_k f_M))^2}\ .$$

Des arguments identiques à ceux après la décomposition (4.3) montrent que l'application

$$\left(x_1^k,\dots,x_n^k,x_1^{k-1},\dots,x_n^{k-1}\right) \longrightarrow \sup_{f\in\mathscr{F}}\sum\left(v_j^k(f_M - \Pi_k f_M)\right)^2$$

est mesurable pour la tribu complétée pour la loi de $(X_1^k,\dots,X_n^k,X_1^{k-1},\dots,X_n^{k-1})$ de la tribu borélienne. Donc, en intégrant (4.18) sous cette loi,

$$w_k(\delta_k) \le 9\mathbb{E}(\|F\|_{2,Q_n^k})\varphi(\delta_k)$$

$$(4.19)\qquad + 4\sqrt{\mathbb{H}(\delta_k)\mathbb{E}\left(\sup_{f\in\mathcal{F}} n^{-1}\sum (V_j^k(f_M - \Pi_k f_M))^2\right)}\ ,$$

où Q_n^k désigne maintenant la mesure aléatoire obtenue en prenant $x_i^k = X_i^k$ et $x_i^{k-1} = X_i^{k-1}$ dans (4.12). Pour contrôler la quantité sous le radical, on applique encore l'inégalité de symétrisation (voir Pollard (1990), théorème 2.2) suivie de la majoration $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, ce qui donne:

$$\sqrt{\mathbb{E}\left(\sup_{f\in\mathcal{F}} \sum (V_j^k(f_M - \Pi_k f_M))^2\right)}$$

$$\le \sqrt{\sup_{f\in\mathcal{F}} \mathbb{E}\left(\sum (V_j^k(f_M - \Pi_k f_M))^2\right)}$$

$$(4.20)\qquad + \sqrt{2\mathbb{E}\left(\sup_{f\in\mathcal{F}} \left|\sum \varepsilon_j (V_j^k(f_M - \Pi_k f_M))^2\right|\right)}\ .$$

Pour achever la majoration, nous allons faire apparaître une inéquation du second degré dans laquelle intervient $w_k(\delta_k)$ à l'aide d'un théorème de comparaison. Comme

$$|V_j^k(f_M - \Pi_k f_M)| \le q_k \|f_M - \Pi_k f_M\|_\infty \le 2q_k M \text{ p.s.}$$

et comme la fonction $x \to x^2$ est lipschitzienne de rapport $4Mq_k$ sur $[-2Mq_k, 2Mq_k]$, le théorème 4.12 de comparaison dans Ledoux et Talagrand (1991) (appliqué conditionnellement aux variables X_i^k et X_i^{k-1}) assure que

$$(4.21)\qquad \mathbb{E}\left(\sup_{f\in\mathcal{F}} \left|\sum \varepsilon_j (V_j^k(f_M - \Pi_k f_M))^2\right|\right) \le 8q_k M\sqrt{n}\, w_k(\delta_k)\ .$$

Majorons l'autre terme de (4.20). D'après (4.11), $\sum (V_j^k(g))^2 \le nQ_n^k(g^2)$. Donc

$$(4.22)\quad \mathbb{E}\left(\sum (V_j^k(g))^2\right) \le 2q_k \sum_j \sum_{i\in I_j} \mathbb{E}\left((g^2(X_i^k) + g^2(X_i^{k-1}))\mathbb{I}_{X_i^k \ne X_i^{k-1}}\right)\ ,$$

car $n_j \le q_k$. Notons $|I_j|$ le cardinal de I_j. D'après (ii) de la proposition 2 et (3.3),

$$\begin{aligned}
&\sum_{i\in I_j} \mathbb{E}\Big((g^2(X_i^k)+g^2(X_i^{k-1}))\mathbb{I}_{X_i^k\neq X_i^{k-1}}\Big)\\
&\quad =\sum_{i=1}^{|I_j|}\mathbb{E}\big((g^2(X_i^*)+g^2(X_i))\big)\mathbb{I}_{X_i\neq X_i^*}\big)\\
&\quad \leq \sum_{i>0}\mathbb{E}(b_{i0}^*(X_i^*)g^2(X_i^*)+b_{i0}(X_i)g^2(X_i)\,)\\
&\quad \leq 4\sum_{i>0}E_P(b_ig^2)\leq E_Q(g^2)\ , \qquad (4.23)
\end{aligned}$$

si les fonctions b_i sont définies par (3.5). Donc, pour la mesure positive Q du théorème 1,

$$\mathbb{E}\left(\sum_j (V_j^k(g))^2\right)\leq \sum_j 2q_kE_Q(g^2)\leq nE_Q(g^2) \qquad (4.24)$$

(rappelons que $q_0\leq n$ et que la sommation porte sur les entiers j tels que $[j2^{-k}]$ ait même parité que $s-1$). De plus, si $g=f_M-\Pi_kf_M$ alors

$$\mathbb{E}_Q(g^2)\leq \mathbb{E}_Q((f-\Pi_kf)^2\mathbb{I}_{F\leq M})\leq \mathbb{E}_Q((f-\Pi_kf)^2)\leq \delta_k^2\ . \qquad (4.25)$$

En mettant bout à bout (4.19), (4.20), (4.21), (4.24) et (4.25), nous obtenons l'inéquation du second degré en $\sqrt{w_k(\delta_k)}$ suivante:

$$w_k(\delta_k)\leq \psi_k+16(n^{-1/2}q_kM\mathbb{H}(\delta_k))^{1/2}\sqrt{w_k(\delta_k)}\ , \qquad (4.26)$$

où

$$\psi_k=9\mathbb{E}(\|F\|_{2,Q_n^k})\varphi(\delta_k)+4\delta_k\sqrt{\mathbb{H}(\delta_k)}\ .$$

L'étude du signe du trinôme montre alors que

$$\sqrt{w_k(\delta_k)}\leq 8\sqrt{n^{-1/2}q_kM\mathbb{H}(\delta_k)}+\sqrt{\psi_k+64n^{-1/2}q_kM\mathbb{H}(\delta_k)}\ . \qquad (4.27)$$

En élevant au carré, nous en déduisons que

$$\begin{aligned}
w_k(\delta_k)&\leq 2^8n^{-1/2}q_kM\mathbb{H}(\delta_k)+2\psi_k\\
&\leq 2^7(n^{-1/2}q_{k-1}M\mathbb{H}(\delta_k)+\delta_{k-1}\sqrt{\mathbb{H}(\delta_k)}\,)+18\mathbb{E}(\|F\|_{2,Q_n^k})\varphi(\delta_k)\ . \qquad (4.28)
\end{aligned}$$

Les coefficients de mélange interviennent dans (4.28) à travers $\mathbb{E}(\|F\|_{2,Q_n^k})$. Pour majorer cette quantité, on note d'abord que $\mathbb{E}(\|F\|_{2,Q_n^k})\leq (\mathbb{E}(\|F\|^2_{2,Q_n^k})\,)^{1/2}$, puis on applique (ii) de la proposition 2. Ainsi

$$\left(\mathbb{E}(\|F\|_{2,Q_n^k})\right)^2 \leq \frac{2}{q_k}\sum_{i=1}^{q_k}\sum_{j=1}^{q_k}\mathbb{E}\left(\mathbb{I}_{X_i\neq X_i^*}\mathbb{I}_{X_j\neq X_j^*}(F^2(X_i)+F^2(X_i^*))\right)$$

$$\leq \frac{4}{q_k}\sum_{i=1}^{q_k}\sum_{j=1}^{q_k}\int_0^{\beta_i\wedge\beta_j} Q_F^2(u)du \ , \tag{4.29}$$

puisque $\mathbb{P}(X_i \neq X_i^*, X_j \neq X_j^*) \leq \beta_i \wedge \beta_j$. Donc

$$\left(\mathbb{E}(\|F\|_{2,Q_n^k})\right)^2 \leq \frac{4}{q_k}\int_0^1 (\beta^{-1}(u)\wedge q_k)^2 Q_F^2(u)du \ . \tag{4.30}$$

En sommant en k, il est facile d'en déduire que

$$\sum_{k=1}^{K}\left(\mathbb{E}(\|F\|_{2,Q_n^k})\right)^2 \leq 16\int_0^1 \beta^{-1}(u)Q_F^2(u)du \ . \tag{4.31}$$

Le premier terme de (4.28) est, à un facteur 2^5 près, le majorant obtenu pour $\mathbb{E}_{1,k-1}$ dans (4.4). La somme en k est donc majorée comme $\mathbb{E}_1$ dans (4.6), à un facteur 2^5 près. Pour majorer la quantité obtenue en sommant en k le second terme, on applique l'inégalité de Cauchy-Schwarz suivie de (4.31). Ainsi

$$\sum_{k=1}^{K} w_k(\delta_k) \leq 800\left(1+2n^{-1/2}qM\delta^{-1}\sqrt{\mathbb{H}(\delta)}\right)\varphi(\delta)$$

$$+\,72\|F\|_{2,\beta}\left(\sum_{k=1}^{K}\varphi^2(\delta_k)\right)^{1/2} \ . \tag{4.32}$$

Enfin, d'après (4.9), $\mathbb{E}_3$ est inférieur ou égal à $4w_1(\delta_1)+\cdots+4w_K(\delta_K)$.

4.3. Fin de la preuve de la proposition 3

La majoration de $\mathbb{E}_2$ est identique à celle du module de continuité du processus empirique dans le cas indépendant et est donc laissée au lecteur. Comme $\|F\|_{2,\beta} = 1/2$, Nous déduisons donc de (4.6) et (4.32) que

$$\mathbb{E}_1+\mathbb{E}_2+\mathbb{E}_3 \leq \ 6425\left(1+2n^{-1/2}qM\delta^{-1}\sqrt{\mathbb{H}(\delta)}\right)\varphi(\delta)$$

$$+\,288\left(\sum_{k=1}^{K}\varphi^2(\delta_k)\right)^{1/2} \ . \tag{4.33}$$

Pour conclure, il nous reste à montrer que

$$\left(\sum_{k=1}^{K}\varphi^2(\delta_k)\right)^{1/2}\leq 9\int_0^{\delta}\sqrt{H(x)L(\delta/x)}\,dx\ . \tag{4.34}$$

En appliquant le théorème de Fubini et ensuite le lemme 2,

$$\sum_{k=1}^{K}\varphi^2(\delta_k)\leq 16\int_0^{\delta}\int_0^{\delta}\sqrt{H(x)H(y)}\sum_{k=1}^{K}\mathbb{I}_{\delta_k>x\vee y}\ dx\,dy\ .$$

Or

$$\sum_{k=1}^{K}\mathbb{I}_{\delta_k>x\vee y}\leq\frac{3}{\log 2}L\left(\frac{\delta}{x\vee y}\right)\leq\frac{3}{\log 2}\sqrt{L(\delta/x)L(\delta/y)}$$

puisque $\delta_k=\delta 2^{-k/3}$, et une seconde application du théorème de Fubini donne (4.34). Par conséquent, la proposition 3 est établie pour $C_0=60000$.

4.4. Un TLC uniforme sous la condition d'entropie universelle de Pollard

Revenons sur la majoration de $\mathbb{E}_3$. L'inégalité (4.32) provient de (4.28) et (4.30) via l'inégalité de Cauchy-Schwarz. Supposons seulement $\varphi(1)<\infty$. Alors l'inégalité de Cauchy-Schwarz est inefficace, et nous pouvons seulement montrer à partir de (4.28) et (4.30) que

$$\mathbb{E}_3\leq 3200\,\varphi(\delta)\left(\frac{2qM\sqrt{\mathbb{H}(\delta)}}{\delta\sqrt{n}}+\sum_{l=0}^{K}\left(2^{-l}\int_0^1\left(\beta^{-1}(u)\wedge 2^l\right)^2 Q_F^2(u)du\right)^{1/2}+1\right)\ . \tag{4.35}$$

Cette majoration conduit à la proposition suivante.

Proposition 4. *Sous la contrainte de β-mélange*

$$\sum_{n>0}Q_F(\beta_n)(\beta_n/n)^{1/2}<\infty\ , \tag{a}$$

il existe une constante C dépendant uniquement des coefficients de β-mélange et de F, telle que pour tout entier q dans $[1,n]$ *puissance de* 2 *et tout δ dans* $]0,1[$,

$$\mathbb{E}\left(\sup_{f\in\mathscr{F}}|Z_n(f_M-\Pi_\delta f_M)|\right)\leq C(1+n^{-1/2}qM\delta^{-1}\mathbb{H}(\delta)\,)\varphi(\delta)+4\sqrt{n}\,M\beta_q\ .$$

Preuve. La proposition 4 résulte de (4.3), (4.6) et (4.35), en notant que

$$\sum_{l=0}^{\infty} 2^{-l/2}\left(\int_0^1 (\beta^{-1}(u) \wedge 2^l)^2 Q_F^2(u)du\right)^{1/2} < \infty$$

si la condition (a) de la proposition 4 est réalisée. ■

Sous la condition de β-mélange (a), cette proposition conduit au TLC uniforme suivant.

Théorème 3. *Soit* $(X_i)_{i\in\mathbb{Z}}$ *une suite stationnaire de variables aléatoires de loi P, satisfaisant la condition de mélange* $\sum_{n>0}(\beta_n/n)^{1/2} < \infty$. *Soit* $\mathscr{F}$ *une classe de fonctions numériques image admissible, de fonction enveloppe F vérifiant (a) de la proposition 4 et de fonction d'entropie universelle H telle que* $\int_0^1 \sqrt{H(x)}\,dx < \infty$. *Alors les conclusions du théorème 1 sont valides.*

Remarque. Si $\mathscr{F}$ est une classe de fonctions numériques uniformément bornées, la condition (a) de la proposition 4 sera réalisée dès que $\sum_{n>0}(\beta_n/n)^{1/2} < \infty$. En particulier la condition de β-mélange $\beta_n = \mathcal{O}(n^{-1}(\log n)^{-2-\varepsilon})$ pour un certain $\varepsilon > 0$ est suffisante pour le TLC uniforme. Si la fonction enveloppe F est telle que $P(F > x) = \mathcal{O}(x^{-r})$, alors (a) sera réalisée dès que $\beta_n = \mathcal{O}(n^{-r/(r-2)}(\log n)^{-2-\varepsilon})$ pour un certain $\varepsilon > 0$.

Annexe: une version d'un lemme de Pisier

Dans cette annexe[1], nous donnons un majorant du maximum d'un nombre fini de variables aléatoires réelles ayant une transformée de Laplace uniformément majorée.

Lemme A.1. *Considérons N variables aléatoires réelles* $Z_1, \ldots, Z_N$. *Supposons qu'il existe une fonction convexe L finie et croissante sur un voisinage à droite de* 0, *telle que* $\log \mathbb{E}(\exp(tZ_i)) \le L(t)$ *pour tout t positif et tout i dans* $[1, N]$. *Soit* h_L *sa transformée de Young. Alors* $\mathbb{E}(\max\{Z_1, \ldots, Z_N\}) \le h_L^{-1}(\log N)$.

Preuve. D'après l'inégalité de Jensen, pour tout t positif,

$$\begin{aligned} t\mathbb{E}(\max\{Z_1, \ldots, Z_N\})) - L(t) &\le \log \mathbb{E}(\exp(t\max\{Z_1, \ldots, Z_N\})) - L(t) \\ &\le \log\left(\sum_{i=1}^{N} \mathbb{E}(\exp(tZ_i))\right) - L(t) \le \log N \ . \end{aligned}$$

[1]Je remercie P. Massart de m'avoir suggéré la forme présente du lemme A.1.

Donc, en maximisant en t, $h_L(\mathbb{E}(\max\{Z_1, \dots, Z_N\})) \leq \log N$, ce qui achève la preuve. ■

Corollaire A.1. *Soient $Y_1, \dots, Y_n$ des v.a. indépendantes, à valeurs dans $\mathscr{X}$. Soit $\mathscr{G}$ une famille finie de fonctions mesurables de $\mathscr{X}$ dans $[-C, C]$, de cardinal* $\exp(H)$. *Posons* $T_n(g) = g(Y_1) + \cdots + g(Y_n)$ et $V = \sup\{\mathbb{E}(g^2(Y_1) + \cdots + g^2(Y_n)) : g \in \mathscr{G}\}$ *Alors*

$$\mathbb{E}(\sup\{T_n(g) - \mathbb{E}(T_n(g)) : g \in \mathscr{G}\}) \leq \sqrt{2VH} + (CH/3) \ .$$

Preuve. Le lemme A.1 s'applique avec $L(t) = Vt^2/(2 - 2Ct/3)$. Si $\lambda = h_L^{-1}(z)$, alors la droite d'équation $y = \lambda x - z$ est tangente à l'hyperbole $y = L(x)$, et donc l'équation $2(\lambda x - z)(1 - Cx/3) = Vx^2$ a une racine double en x. Donc $\Delta = (\lambda - Cz/3)^2 - 2zV = 0$, et par conséquent $\lambda \leq \sqrt{2Vz} + Cz/3$, ce qui achève la preuve.

Références

Andersen, N.T., Giné, E., Ossiander, M., Zinn, J.: The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. Probab. Th. Rel. Fields **77,** 271–305 (1988)

Arcones, M.A. and Yu, B.: Central limit theorems for empirical and U-processes of stationary mixing sequences. J. Theoret. Prob. **7,** 47–71 (1994)

Bolthausen, E.: The Berry-Esseen theorem for functionals of discrete Markov chains. Z. Wahrsch. verw. Gebiete **54,** 59–73 (1980)

Bolthausen, E.: The Berry-Esseen theorem for strongly mixing Harris recurrent Markov chains. Z. Wahrsch. verw. Gebiete **60,** 283–289 (1982)

Dellacherie, C.: Capacités et processus stochastiques. Ergebnisse der mathematik und ihrer grenzgebiete. Springer, Berlin (1972)

Delyon, B.: Publication interne IRISA Rennes (1990)

Donsker, M.: Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov's theorems. Ann. Math. Stat. **23,** 277–281 (1952)

Doukhan, P.: Mixing. Properties and Examples. Lecture Notes in Statistics **85**. Berlin. Springer 1994

Doukhan, P., Massart, P., Rio, E.: The functional central limit theorem for strongly mixing processes. Annales inst. H. Poincaré Probab. Statist. **30,** 63–82 (1994)

Doukhan, P., Massart, P., Rio, E.: Invariance principles for absolutely regular processes. Ann. Inst. H. Poincaré Probab. Statist. **31,** 393–427 (1995)

Dudley, R.M.: Central limit theorems for empirical measures. Ann. Probab. **6**. 899–929 (1978)

Dudley, R.M.: A course on empirical processes. Ecole d'été de probabilités de Saint-Flour XII-1982. Lectures Notes in Math. **1097,** 1–142. Berlin: Springer (1984)

Goldstein, S.: Maximal coupling. Z. Wahrsch. verw. Gebiete **46,** 193–204 (1979)

Gorodetskii, V.V.: On the strong mixing property for linear sequences. Theory Probab. Appl. **22,** 411–413 (1982)

Hall, P., Heyde, C.C.: Martingale limit theory and its applications. Academic Press 1980

Ibragimov, I.A., Linnik, Y.V.: Independent and stationary sequences of random variables. Amsterdam: Wolters-Noordhoff 1971

Kolchinskii, V.I.: On the central limit theorem for empirical measures (In Russian). Teor. vero. i. mat. stat. (Kiev) **24,** 63–75 (1981)

Ledoux, M., Talagrand, M.: Probability in Banach spaces. Isoperimetry and Processes. Berlin: Springer 1991

Levental, E.: Uniform limit theorems for Harris recurrent Markov chains. Probab. Th. Rel. Fields **80,** 101–118 (1988)

Ossiander, M.: A central limit theorem under metric entropy with L_2-bracketing. Ann. Probab. **15,** 897–919 (1987)

Pham, T.D., Tran, L.T.: Some mixing properties of time series models. Stochastic Process. Appl. **19,** 297–303 (1985)

Pollard, D.: A central limit theorem for empirical processes. J. Aust. Math. Soc. Ser. A **33,** 235–248 (1982)

Pollard, D.: Convergence of stochastic processes. Berlin: Springer 1984

Pollard, D.: Empirical processes : theory and applications. NSF-CBMS Regional Conference Series in Probability and Statistics. Hayward-Alexandria: IMS-ASA 1990

Skorohod, A.V.: On a representation of random variables. Theory Probab. Appl. **21,** 628–632 (1976)

Tuominen, P., Tweedie, R.L.: Subgeometric rates of convergence of f-ergodic Markov chains. Adv. Appl. Prob. **26,** 775–798 (1994)

Viennet, G.: Inequalities for absolutely regular sequences: application to density estimation. Prob. Th. Rel. Fields **107,** 467–492 (1997)

Volkonskii, V.A., Rozanov, Y.A.: Some limit theorems for random functions I. Theory Probab. Appl. **4,** 178–197 (1959)