Théorèmes limites pour les U-statistiques de données bernoulliennes

Davide Giraudo
Euler International Mathematical Institute (Saint-Pétersbourg)

Séminaire Approx, EDP et Modèles aléatoires
Calais, 7 janvier 2021

Définition des U-statistiques

Afin d'estimer $\mathbb{E}\left[h\left(X, Y\right)\right]$ via une moyenne empirique, où X et Y sont i.i.d. et $h\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, on définit la U-statistique de noyau h par

$$U_n := \sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right), \quad n \geqslant 2,$$

où $\left(X_j\right)_{j \geqslant 1}$ est i.i.d. et de même loi que X. On peut prendre comme estimateur $U_n / \binom{n}{2}$.

Objectif général : comprendre le comportement asymptotique de U_n.

Notons que pour chaque $i < j$, $h\left(X_i, X_j\right)$ a la même loi que $h\left(X_1, X_2\right)$. La U-statistique U_n peut-être vue comme les sommes partielles des variables aléatoires (non-indépendantes) $Y_j := \sum_{i=1}^{j-1} h\left(X_i, X_j\right)$.

Exemples de noyaux

On rappelle que

$$U_n := \sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right), \quad n \geqslant 2.$$

1. Estimateur de différence de moyenne de Gini : $h\left(x, y\right) = |x - y|$.
2. Estimateur de variance : $h\left(x, y\right) := \left(x - y\right)^2 / 2$. Alors $U_n / \binom{n}{2}$ donne un estimateur de la variance.
3. Estimateur de Grassberger-Procaccia : pour $t > 0$ fixé, $h\left(x, y\right) = \mathbf{1}\left\{|x - y| \leqslant t\right\}$.

Outil : décomposition d'Hoeffding

Soit $h\colon \mathbb{R}^2 \to \mathbb{R}$ une fonction mesurable et $(X_i)_{i\geqslant 1}$ une suite i.i.d. Soit

$$U_n = \sum_{1\leqslant i<j\leqslant n} h\left(X_i, X_j\right).$$

On définit $\theta := \mathbb{E}\left[h\left(X_1, X_2\right)\right]$,

$$h_1\left(x\right) = \mathbb{E}\left[h\left(x, X_1\right)\right] - \theta, \quad h_2\left(y\right) = \mathbb{E}\left[h\left(X_1, y\right)\right] - \theta,$$

$$h_3\left(x, y\right) = h\left(x, y\right) - h_1\left(x\right) - h_2\left(y\right) - \theta.$$

Alors

$$U_n = \binom{n}{2}\theta + \sum_{i=1}^{n-1}\left(n-i\right) h_1\left(X_i\right) + \sum_{j=2}^{n}\left(j-1\right) h_2\left(X_j\right) + \sum_{1\leqslant i<j\leqslant n} h_3\left(X_i, X_j\right)$$

et

$$\mathbb{E}\left[h_3\left(X_i, X_j\right) \mid X_1, \dots, X_{j-1}\right] = \mathbb{E}\left[h_3\left(X_i, X_j\right) \mid X_{i+1}, \dots, X_n\right] = 0.$$

On remarque que $\mathbb{E}\left[h_3\left(X_i, x\right)\right] = 0$ pour tout x dans le support de la loi de X_1.

Cas symétrique

On suppose h symétrique, *i.e.*, $h(x,y) = h(y,x)$ pour tous $x, y \in \mathbb{R}$. On a obtenu

$$U_n = \binom{n}{2}\theta + \sum_{i=1}^{n-1}(n-i)\,h_1(X_i) + \sum_{j=2}^{n}(j-1)\,h_2(X_j) + \sum_{1\leqslant i<j\leqslant n} h_3(X_i, X_j)\,.$$

La symétrie entraîne que $h_1 = h_2$ donc

$$U_n = \binom{n}{2}\theta + n\sum_{i=1}^{n} h_1(X_i) + \sum_{1\leqslant i<j\leqslant n} h_3(X_i, X_j)\,.$$

Le terme $n\sum_{i=1}^{n} h_1(X_i)$ est appelé partie linéaire ; le terme $\sum_{1\leqslant i<j\leqslant n} h_3(X_i, X_j)$ partie dégénérée.

On dit que le noyau h est dégénéré si $\mathbb{E}[h(X_1, X_2) \mid X_1] = 0$ presque sûrement (ou de manière équivalente, $\mathbb{E}[h(X_1, x)] = 0$ pour tout x dans le support de la loi de X_0).

Dans la suite, on supposera h symétrique.

Loi des grands nombres

Proposition (Giné, Zinn (1991))

Soient $(X_i)_{i\in\mathbb{Z}}$ *une suite i.i.d. et* $h\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$ *une fonction mesurable. Soit* $1\leqslant p<2$. *On suppose que* $\mathbb{E}\left[\left|h\left(\varepsilon_0,\varepsilon_1\right)\right|^p\right]<\infty$ *et que pour tout* $x\in\operatorname{supp}\left(\mathbb{P}_{X_1}\right)$, $\mathbb{E}\left[h\left(X_0,x\right)\right]=0$. *Alors :*

1. *la suite* $\left(n^{-2/p}\left|\sum_{1\leqslant i<j\leqslant n}h\left(X_i,X_j\right)\right|\right)_{n\geqslant 1}$ *converge vers* 0 *presque sûrement ;*
2. *pour tout* $t>0$,

$$t^p\mathbb{P}\left\{\sup_{n\geqslant 1}\frac{1}{n^{2/p}}\left|\sum_{1\leqslant i<j\leqslant n}h\left(X_i,X_j\right)\right|>t\right\}\leqslant\kappa_p\mathbb{E}\left[\left|h\left(\varepsilon_0,\varepsilon_1\right)\right|^p\right].$$

Sans l'hypothèse $\mathbb{E}\left[h\left(X_0,x\right)\right]=0$, la suite $\left(n^{-1-1/p}\left|\sum_{1\leqslant i<j\leqslant n}\left(h\left(X_i,X_j\right)-\mathbb{E}\left[h\left(X_1,X_2\right)\right]\right)\right|\right)_{n\geqslant 1}$ converge vers 0 presque sûrement

Loi des logarithmes itérés bornée

Si $h(X_1, X_2) \in \mathbb{L}^2$, alors la variable aléatoire

$$\sup_{n \geqslant 1} \frac{1}{n^{3/2}\sqrt{\mathrm{LL}(n)}} \left| U_n - \mathbb{E}[U_n] \right|$$

est presque sûrement finie, où $L \colon \mathbb{R}_+ \to \mathbb{R}_+$ est définie par $L(x) = \max\{\ln x, 1\}$ et $\mathrm{LL}(x) := L \circ L(x)$.

Plus précisément, pour tout $1 < p < 2$, l'inégalité suivante a lieu (Arcones, Giné, 1995) :

$$\left\| \sup_{n \geqslant 1} \frac{1}{n^{3/2}\sqrt{\mathrm{LL}(n)}} \left| U_n - \mathbb{E}[U_n] \right| \right\|_p \leqslant C_p \left\| h(X_1, X_2) \right\|_2$$

Théorème limite central

Si $h(X_1, X_2) \in \mathbb{L}^2$, alors la convergence en loi suivante a lieu

$$\frac{1}{n^{3/2}}\left(U_n - \mathbb{E}\left[U_n\right]\right) \to \sqrt{\mathbb{E}\left[\mathbb{E}\left[h(X_1, X_2) \mid X_1\right]^2\right]}N,$$

où N est de loi normale centrée réduite.

Il est possible que $\mathbb{E}\left[h(X_1, X_2) \mid X_1\right] = 0$ presque sûrement (par exemple si $h(x, y) = xy$ et X_1 est centrée). Si $\mathbb{E}\left[h(X_1, X_2) \mid X_1\right] = 0$, alors la convergence en loi suivante a lieu

$$\frac{1}{n}U_n \to \sum_{k \geqslant 1} \lambda_k \left(N_k^2 - 1\right),$$

où $(N_k)_{k \geqslant 1}$ est une suite i.i.d. de loi normale centrée réduite et $\sum_{k \geqslant 1} \lambda_k^2$ converge.

Suites bernoulliennes

Definition

On dit qu'une suite de variables aléatoires $(X_i)_{i\in\mathbb{Z}}$ est bernoullienne s'il existe une suite i.i.d. $(\varepsilon_u)_{u\in\mathbb{Z}}$ et une fonction mesurable $f\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ telles que $X_j := f\left((\varepsilon_{j-i})_{i\in\mathbb{Z}}\right)$ presque sûrement.

Examples

1. Fonction de processus linéaire : $X_i = g\left(\sum_{k\in\mathbb{Z}} a_k\varepsilon_{i-k}\right)$, où $g\colon \mathbb{R}\to\mathbb{R}$ est hölderienne et $\sum_{k\in\mathbb{Z}} a_k^2 < \infty$.
2. Processus de Volterra : $X_i = \sum_{k,k'\in\mathbb{Z},k\neq k'} a_{k,k'}\varepsilon_{i-k}\varepsilon_{i-k'}$, où $\sum_{k,k'\in\mathbb{Z},k\neq k'} a_{k,k'}^2 < \infty$.

But : fournir des conditions sur la dépendance de $(X_i)_{i\in\mathbb{Z}}$ et le noyau h pour obtenir les théorèmes limites précédents pour $U_n := \sum_{1\leqslant i<j\leqslant n} h(X_i, X_j)$.

Mesure de dépendence pour les suites

Soit $(X_j)_{j\in\mathbb{Z}}$ une suite bernoullienne de la forme $X_j := f\left((\varepsilon_{j-i})_{i\in\mathbb{Z}}\right)$ où $(\varepsilon_u)_{u\in\mathbb{Z}}$ est i.i.d. et $f\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ est mesurable. Soit

$$V_{j,\ell} := (\varepsilon_u)_{u=j-\ell}^{j+\ell}.$$

La variable aléatoire $\mathbb{E}\left[X_j \mid V_{j,\ell}\right]$ peut s'exprimer comme une fonction de $V_{j,\ell}$ et par stationarité, cette fonction est indépendante de j. Par conséquent, on peut écrire

$$\mathbb{E}\left[X_j \mid V_{j,\ell}\right] = f_\ell\left(V_{j,\ell}\right).$$

Lorsque $X_0 \in \mathbb{L}^p$, on définit

$$\delta_{\ell,p} := \left\|f_\ell\left(V_{j,\ell}\right) - f_{\ell-1}\left(V_{j,\ell-1}\right)\right\|_p.$$

Par le théorème de convergence des martingales, $\delta_{\ell,p} \to 0$ lorsque $\ell \to \infty$.

Example

Pour $X_i = g\left(\sum_{k\in\mathbb{Z}} a_k \varepsilon_{i-k}\right)$, où $g\colon \mathbb{R} \to \mathbb{R}$ est γ-hölderienne,

$$\delta_{\ell,p} \leqslant C_{p,\gamma}\left(|a_\ell|^\gamma + |a_{-\ell}|^\gamma\right)\|X_0\|_{p\gamma}^\gamma.$$

Mesure de dépendence pour les U-statistiques

Soit $(X_j)_{j\in\mathbb{Z}}$ une suite bernoullienne de la forme $X_j := f\left((\varepsilon_{j-i})_{i\in\mathbb{Z}}\right)$ où $(\varepsilon_u)_{u\in\mathbb{Z}}$ est i.i.d. et $f\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ est mesurable. Soit

$$V_{j,\ell} := (\varepsilon_u)_{u=j-\ell}^{j+\ell}.$$

Soit f_ℓ telle que

$$\mathbb{E}\left[X_j \mid V_{j,\ell}\right] = f_\ell\left(V_{j,\ell}\right).$$

Pour $p \geqslant 1$ et $\ell \geqslant 1$, on définit le coefficient

$$\theta_{\ell,p} := \sup_{j\geqslant 0} \left\| h\left(f_\ell\left(V_{0,\ell}\right), f_\ell\left(V_{j,\ell}\right)\right) - h\left(f_{\ell-1}\left(V_{0,\ell-1}\right), f_{\ell-1}\left(V_{j,\ell-1}\right)\right)\right\|_p$$

et pour $\ell = 0$, $\theta_{0,p} = \sup_{j\geqslant 0} \left\| h\left(f_0\left(V_{j,\ell}\right), f_0\left(V_{0,\ell}\right)\right)\right\|_p$.

Si $\sum_{\ell\geqslant 1} \theta_{\ell,p} < \infty$, alors

$$\sum_{1\leqslant i<j\leqslant n} h\left(X_i, X_j\right) = \sum_{1\leqslant i<j\leqslant n} h\left(f_0\left(V_{i,0}\right), f_0\left(V_{j,0}\right)\right)$$
$$+ \sum_{\ell\geqslant 1} \sum_{1\leqslant i<j\leqslant n} h\left(f_\ell\left(V_{i,\ell}\right), f_\ell\left(V_{j,\ell}\right)\right) - h\left(f_{\ell-1}\left(V_{i,\ell-1}\right), f_{\ell-1}\left(V_{j,\ell-1}\right)\right),$$

où la convergence a lieu dans $\mathbb{L}^p$.

Exemple : noyau hölderien

On suppose que le noyau $h\colon \mathbb{R}^2 \to \mathbb{R}$ vérifie pour tous $x, x', y, y' \in \mathbb{R}$,

$$|h(x,y) - h(x',y')| \leqslant C\,|x - x'|^{\alpha} + C\,|y - y'|^{\alpha},$$

où C est indépendante de x, x', y, y'. On rappelle que

$$\theta_{\ell,p} := \sup_{j \geqslant 0} \left\| h\left(f_\ell\left(V_{0,\ell}\right), f_\ell\left(V_{j,\ell}\right)\right) - h\left(f_{\ell-1}\left(V_{0,\ell-1}\right), f_{\ell-1}\left(V_{j,\ell-1}\right)\right)\right\|_p,$$

où

$$V_{j,\ell} := \left(\varepsilon_u\right)_{u=j-\ell}^{j+\ell}.$$

et f_ℓ est telle que

$$\mathbb{E}\left[X_j \mid V_{j,\ell}\right] = f_\ell\left(V_{j,\ell}\right).$$

Alors

$$\theta_{\ell,p} \leqslant 2C\left\|\mathbb{E}\left[X_0 \mid V_{0,\ell}\right] - \mathbb{E}\left[X_0 \mid V_{0,\ell-1}\right]\right\|_{p\alpha}^{\alpha} = 2C\delta_{\ell,p\alpha}^{\alpha}.$$

Exemple : estimateur de variance

Soit $h\left(x,y\right) := \left(x-y\right)^2/2$. Grâce à l'inégalité

$$\left\|Y^2-Z^2\right\|_p^p = \mathbb{E}\left[\left|Y-Z\right|^p\left|Y+Z\right|^p\right] \leqslant \mathbb{E}\left[\left|Y-Z\right|^{2p}\right]^{1/2}\mathbb{E}\left[\left|Y+Z\right|^{2p}\right]^{1/2},$$

on déduit que

$$\theta_{\ell,p} \leqslant 2\left\|X_0\right\|_{2p}\left\|\mathbb{E}\left[X_0 \mid V_{0,\ell}\right]-\mathbb{E}\left[X_0 \mid V_{0,\ell-1}\right]\right\|_{2p} = 2\left\|X_0\right\|_{2p}\delta_{\ell,2p}.$$

Exemple : estimateur de Grassberger-Procaccia

Soit $t \in \mathbb{R}$ fixé et $h\colon (x,y) \in \mathbb{R}^2 \mapsto \mathbf{1}\left\{|x-y| \leqslant t\right\}$.

On suppose que pour tous ℓ et j, la variable aléatoire $\mathbb{E}\left[X_0 \mid \sigma\left(\varepsilon_u, -\ell \leqslant u \leqslant \ell\right)\right] - \mathbb{E}\left[X_j \mid \sigma\left(\varepsilon_u, j-\ell \leqslant u \leqslant j+\ell\right)\right]$ a une densité $g_{\ell,j}$ dont la norme $\mathbb{L}^r\left(\mathbb{R}\right)$ est bornée indépendemment de ℓ et j pour un certain $r > 1$.

Alors l'inégalité suivante a lieu

$$\theta_{\ell,p} \leqslant C \left\|\mathbb{E}\left[X_0 \mid V_{0,\ell}\right] - \mathbb{E}\left[X_0 \mid V_{0,\ell-1}\right]\right\|_q^{\frac{q}{p}\frac{r}{q(r-1)+r}} .$$

Utilisation de la mesure de dépendence

On rappelle que

$$\theta_{\ell,p} := \sup_{j \geqslant 0} \left\| h\left(f_\ell\left(V_{0,\ell}\right), f_\ell\left(V_{j,\ell}\right)\right) - h\left(f_{\ell-1}\left(V_{0,\ell-1}\right), f_{\ell-1}\left(V_{j,\ell-1}\right)\right)\right\|_p,$$

où

$$V_{j,\ell} := \left(\varepsilon_u\right)_{u=j-\ell}^{j+\ell}.$$

etf_ℓ est telle que

$$\mathbb{E}\left[X_j \mid V_{j,\ell}\right] = f_\ell\left(V_{j,\ell}\right).$$

De plus, si $\sum_{\ell \geqslant 1} \theta_{\ell,p} < \infty$, alors

$$\begin{aligned}\sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right) &= \sum_{1 \leqslant i < j \leqslant n} h\left(f_0\left(V_{i,0}\right), f_0\left(V_{j,0}\right)\right) \\ &+ \sum_{\ell \geqslant 1} \sum_{1 \leqslant i < j \leqslant n} h\left(f_\ell\left(V_{i,\ell}\right), f_\ell\left(V_{j,\ell}\right)\right) - h\left(f_{\ell-1}\left(V_{i,\ell-1}\right), f_{\ell-1}\left(V_{j,\ell-1}\right)\right).\end{aligned}$$

Par conséquent, $\sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right)$ peut s'exprimer comme une série de U-statistiques mettant en jeu des suites stationnaires qui sont une fonction de $2\ell + 1$ variables aléatoires indépendantes.

U-statistiques de données $2\ell+1$-dépendentes

Soit $(X_i)_{i\in\mathbb{Z}}$ une suite de variables aléatoires telle que $X_i = f(\varepsilon_{i-\ell},\dots,\varepsilon_{i+\ell})$, où $(\varepsilon_k)_{k\in\mathbb{Z}}$ est i.i.d. et $f\colon \mathbb{R}^{2\ell+1}\to\mathbb{R}$ est mesurable.

Le traitement des sommes partielles de ces suites est en général aisé. En effet,

$$\sum_{i=1}^{n} X_i = \sum_{r=0}^{4\ell+1} \sum_{\ell:1\leqslant(4\ell+2)k+r\leqslant n} X_{(4\ell+2)k+r}$$

et pour chaque $r\in\{0,4\ell+1\}$, la suite $\left(X_{(4\ell+2)k+r}\right)_{k\in\mathbb{Z}}$ est i.i.d..

Pour les U-statistiques, la situation est plus complexe ; par exemple si $X_i = f(\varepsilon_{i-1},\varepsilon_i)$, on peut écrire

$$\begin{aligned}\sum_{1\leqslant i<j\leqslant 2n} h(X_i,X_j) = &\sum_{1\leqslant i'<j'\leqslant n} h(X_{2i'},X_{2j'}) + \sum_{1\leqslant i'<j'\leqslant n} h(X_{2i'-1},X_{2j'-1})\\ &+ \sum_{1\leqslant i'<j'\leqslant n} h(X_{2i'},X_{2j'-1}) + \sum_{1\leqslant i'<j'\leqslant n} h(X_{2i'-1},X_{2j'})\end{aligned}$$

Les deux premiers termes peuvent s'exprimer comme une U-statistique dont les données sont des vecteurs i.i.d., mais ce n'est pas le cas des deux derniers.

Décomposition de Hoeffding

Proposition (G., 2021+)

Soient $(X_i)_{i\in\mathbb{Z}}$ une suite bernoullienne, $h\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$ une fonction mesurable. On suppose que $\sum_{\ell\geqslant 1}\theta_{\ell,p}<\infty$. Alors

$$\sum_{1\leqslant i<j\leqslant n}\left(h\left(X_i,X_j\right)-\mathbb{E}\left[h\left(X_i,X_j\right)\right]\right)=n\sum_{i=1}^{n}Y_i + \sum_{\ell\geqslant 0}\sum_{1\leqslant a,b\leqslant 4\ell+2}\sum_{1\leqslant i<j\leqslant[n/2\ell]}h^{(\ell)}_{a,b}\left(\varepsilon^{a,b}_i,\varepsilon^{a,b}_j\right)+R_n,$$

où la suite $(Y_i)_{i\geqslant 1}$ est bernoulienne et centrée, et pour chaque $\ell\geqslant 0$ et $1\leqslant a,b\leqslant 4\ell+2$, $\left(\varepsilon^{a,b}_i\right)_{i\in\mathbb{Z}}$ est une suite indépendante de vecteurs de dimension $4\ell+2$ et la U-statistique associée à $h^{(\ell)}_{a,b}$ est dégénérée.

Le terme R_n joue un rôle négligeable dans le cadre des théorèmes limites.

Loi des grands nombres

Théorème (G., 2021+)

Soit $(\varepsilon_u)_{u\in\mathbb{Z}}$ *une suite i.i.d.,* $f\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ *et* $h\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$ *des fonctions mesurables et* $X_i = f\left(\left(\varepsilon_{i-k}\right)_{k\in\mathbb{Z}}\right)$*. Soit* $p\in[1,2)$*. Alors*

$$\sup_{t>0} t\mathbb{P}\left\{\sup_{n\geqslant 1}\frac{1}{n^{1+1/p}}\left|\sum_{1\leqslant i<j\leqslant n}\left(h\left(X_i,X_j\right)-\mathbb{E}\left[h\left(X_i,X_j\right)\right]\right)\right|>t\right\}^{1/p} \leqslant c_p\left(\theta_{0,p}+\sum_{\ell\geqslant 1}\ell^{1-1/p}\theta_{\ell,p}\right).$$

Si $\sum_{\ell\geqslant 1}\ell^{1-1/p}\theta_{\ell,p}<\infty$*, alors*

$$\frac{1}{n^{1+1/p}}\left|\sum_{1\leqslant i<j\leqslant n}\left(h\left(X_i,X_j\right)-\mathbb{E}\left[h\left(X_i,X_j\right)\right]\right)\right|\to 0 \text{ p. s..}$$

Loi des logarithmes itérés

Théorème (G., 2021+)

Soit $(\varepsilon_u)_{u\in\mathbb{Z}}$ une suite i.i.d., $f\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ et $h\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$ des fonctions mesurables et $X_i = f\left(\left(\varepsilon_{i-k}\right)_{k\in\mathbb{Z}}\right)$. Soit $p \in [1,2)$. L'inégalité suivante a lieu :

$$\left\| \sup_{n\geqslant 1} \frac{1}{n^{3/2}\sqrt{\mathrm{LL}\,(n)}} \left| \sum_{1\leqslant i<j\leqslant n} \left(h\left(X_i, X_j\right) - \mathbb{E}\left[h\left(X_i, X_j\right)\right]\right) \right| \right\|_p \leqslant c_p \left(\theta_{0,2} + \sum_{\ell\geqslant 1} \ell^{1/2} \theta_{\ell,2} \right),$$

où c_p ne dépend que de p.

Théorème limite central

Théorème (G., 2021+)

Soient $(\varepsilon_u)_{u\in\mathbb{Z}}$ *une suite i.i.d.,* $f\colon \mathbb{R}^{\mathbb{Z}}\to\mathbb{R}$ *et* $h\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$ *des fonctions mesurables et*

$$X_j := f\left(\left(\varepsilon_{j-k}\right)_{k\in\mathbb{Z}}\right), \quad U_n := \sum_{1\leqslant i<j\leqslant n} h\left(X_i, X_j\right).$$

Supposons que

$$\sum_{\ell\geqslant 0} \ell^{1/2}\theta_{\ell,2} < +\infty; \quad \sum_{\ell\geqslant 0} \ell^{2}\theta_{\ell,1} < +\infty$$

et

$$\sum_{k\in\mathbb{Z}} \left|\operatorname{Cov}\left(Y_0, Y_k\right)\right| < +\infty,$$

où Y_k *est donnée par* $Y_k = \mathbb{E}\left[h\left(X_k, X_0'\right) \mid \varepsilon_u, u\in\mathbb{Z}\right]$ *et* X_0' *a la même loi que* X_0 *et est indépendante de* $\sigma\left(\varepsilon_u, u\in\mathbb{Z}\right)$*. Alors*

$$\frac{1}{n^{3/2}}\left(U_n - \mathbb{E}\left[U_n\right]\right) \to \sqrt{\sum_{k\in\mathbb{Z}} \operatorname{Cov}\left(Y_0, Y_k\right)} N, \quad N\sim\mathcal{N}(0,1).$$

Perspectives

1. Traitement des U-statistiques d'ordre supérieur : $U_n = \sum_{1 \leqslant i_1 < \cdots < i_r \leqslant n} h\left(X_{i_1}, \ldots, X_{i_r}\right)$.

2. Utilisation de méthodes d'approximation par martingales pour obtenir des théorèmes limites.

3. Obtention de vitesse de convergence dans le théorème limite central : contrôle de
$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{U_n - \mathbb{E}\left[U_n\right]}{\sigma n^{3/2}} \leqslant x \right\} - \mathbb{P}\left\{N \leqslant x\right\} \right|.$$

4. Théorèmes limites fonctionnels : $W_n\left(t\right) = \sum_{1 \leqslant i < j \leqslant [nt]} h\left(X_i, X_j\right)$.