Inégalités de déviation pour des U-statistiques

Davide Giraudo
IRMA

Journées MAS 2022
Rouen, 30 août 2022

Plan

Introduction aux U-statistiques

Inégalités de déviation

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 $h\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$, la U-statistique de noyau h, définie 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., fût introduite par **Hoeffding** (1948).

On peut prendre comme estimateur U_n/C_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) $Y_j := \sum_{i=1}^{j-1} h\left(X_i, X_j\right)$.

Exemple 1 : $h(x,y) = x + y$

On suppose que $(X_i)_{i\geqslant 1}$ est i.i.d., centrée, $\mathbb{E}\left[X_1^2\right] = 1$ et $h(x,y) = x + y$. Alors

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

donc $\left(U_n/n^{3/2}\right)_{n\geqslant 2}$ converge en loi vers une loi normale centrée réduite et $U_n/n^{1+1/p} \to 0$ presque sûrement pour tout $1 \leqslant p < 2$.

Exemple 2 : $h(x,y) = x \cdot y$

On suppose que $(X_i)_{i \geqslant 1}$ est i.i.d., centrée, $\mathbb{E}\left[X_1^2\right] = 1$ et $h(x,y) = x \cdot y$.
Alors

$$\begin{aligned} U_n &= \sum_{1 \leqslant i < j \leqslant n} h(X_i, X_j) \\ &= \sum_{1 \leqslant i < j \leqslant n} X_i X_j \\ &= \frac{1}{2}\left(\left(\sum_{k=1}^{n} X_k\right)^2 - \sum_{k=1}^{n} X_k^2\right) \end{aligned}$$

donc $(2U_n/n)_{n \geqslant 2}$ converge en loi vers $N^2 - 1$, où N est de loi normale centrée réduite et $U_n/n^{2/p} \to 0$ presque sûrement pour tout $1 \leqslant p < 2$.

Le noyau h influe donc sur la normalisation et la loi limite.

Autres 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 / C_n^2 fournit un estimateur de la variance.

3. Estimateur de Grassberger-Procaccia : pour $t > 0$ fixé, $h\left(x, y\right) = \mathbf{1}\left\{|x - y| \leqslant t\right\}$.

Propriété de martingale ?

Soit $D_j := \sum_{i=1}^{j-1} h(X_i, X_j)$ avec $(X_i)_{i \geqslant 0}$ i.i.d.. Alors $U_n = \sum_{j=2}^{n} D_j$. On cherche à savoir si $(D_j)_{j \geqslant 2}$ est une suite d'accroissements d'une martingale pour la filtration $(\mathcal{F}_j)_{j \geqslant 1}$, où $\mathcal{F}_j = \sigma(X_k, 1 \leqslant k \leqslant j)$.
En utilisant l'identité

$$\mathbb{E}[Y \mid \mathcal{F} \vee \mathcal{G}] = \mathbb{E}[Y \mid \mathcal{F}]$$

valable si $\mathcal{G}$ est indépendante de $\sigma(Y) \vee \mathcal{F}$, on obtient

$$\begin{aligned}
\mathbb{E}[D_j \mid \mathcal{F}_{j-1}] &= \sum_{i=1}^{j-1} \mathbb{E}[h(X_i, X_j) \mid \sigma(X_k, 1 \leqslant k \leqslant j-1)] \\
&= \sum_{i=1}^{j-1} \mathbb{E}[h(X_i, X_j) \mid \sigma(X_i) \vee \sigma(X_k, 1 \leqslant k \leqslant j-1, k \neq i)] \\
&= \sum_{i=1}^{j-1} \mathbb{E}[h(X_i, X_j) \mid \sigma(X_i)] \\
&= \sum_{i=1}^{j-1} h_1(X_i), \text{ avec } h_1(x) = \mathbb{E}[h(x, X_2)]
\end{aligned}$$

donc $\mathbb{E}[D_j \mid \mathcal{F}_{j-1}] = 0$ si et seulement si $\mathbb{E}[h(X_1, X_2) \mid X_1] = 0$.

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

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

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

$$h_1(x) = \mathbb{E}[h(x, X_2)] - \theta, \quad h_2(y) = \mathbb{E}[h(X_1, y)] - \theta,$$

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

Alors

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

et

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

Cas symétrique

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

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

La symétrie entraîne que $h_1 = h_2$ donc

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

Le terme $(n-1)\sum_{i=1}^{n} h_1(X_i)$ 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 $\mathbb{E}\left[h(X_1, X_2) \mid X_1\right] = 0$ presque sûrement.

Dans la suite, on supposera h symétrique.

Traitement du terme dégénéré

On note $U_n(h_3) = \sum_{1 \leqslant i < j \leqslant n} h_3(X_i, X_j)$ la partie dégénérée.

On utilise la propriété de martingale pour la sommation en i et j et l'inégalité de Burkholder.

- Moment d'ordre 2 : si $(i,j) \neq (k,\ell)$, alors $\mathbb{E}[h_3(X_i,X_j)\, h(X_k,X_\ell)] = 0$ (on conditionne par rapport à $X_1,\ldots,X_{\ell-1}$ si $\ell > j$, ou par rapport à $X_{i+1},\ldots,X_n$ si $i < k$ et on traite les autres cas en échangeant les rôles de (i,j) et (k,ℓ)). Par conséquent,

$$\mathbb{E}\left[\max_{2 \leqslant n \leqslant N} U_n(h_3)^2\right] \leqslant 4 \sum_{1 \leqslant i < j \leqslant N} \mathbb{E}\left[h_3^2(X_i,X_j)\right] \leqslant KN^2 \mathbb{E}\left[h_3^2(X_1,X_2)\right].$$

- Moments d'ordre $1 < p < 2$:

$$\mathbb{E}\left[\max_{2 \leqslant n \leqslant N} |U_n(h_3)|^p\right] \leqslant K_p N^2 \mathbb{E}\left[|h_3(X_1,X_2)|^p\right].$$

- Moments d'ordre $p > 2$:

$$\mathbb{E}\left[\max_{2 \leqslant n \leqslant N} |U_n(h_3)|^p\right] \leqslant K_p N^p \mathbb{E}\left[|h_3(X_1,X_2)|^p\right].$$

Loi des grands nombres

Proposition (**Giné, Zinn** (1991))

Soient $(X_i)_{i \geqslant 1}$ *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* $\mathbb{E}\left[\left|h\left(X_1, X_2\right)\right|^p\right] < \infty$*.*

- *Si h est dégénérée pour* $(X_i)_{i \geqslant 1}$ *(*$\mathbb{E}\left[h\left(X_1, X_2\right) \mid X_1\right] = 0$ *p.s.), alors*

$$\frac{1}{n^{2/p}} \left| \sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right) \right| \to 0 \ p.s.$$

- *Si on suppose simplement que* $\mathbb{E}\left[h\left(X_1, X_2\right)\right] = 0$*, alors*

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

Contrôle de la fonction maximale

Proposition (G. (2021))

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

- *si* h *est dégénérée pour* $(X_i)_{i\geqslant 1}$ *(*$\mathbb{E}\left[h\left(X_1,X_2\right)\mid X_1\right]=0$ *p.s.), alors*

$$t^p\mathbb{P}\left\{\sup_{n\geqslant 1}\frac{1}{n^{2/p}}\left|\sum_{1\leqslant i<j\leqslant n}h\left(X_i,X_j\right)\right|>t\right\}\leqslant\kappa_p\mathbb{E}\left[\left|h\left(X_1,X_2\right)\right|^p\right].$$

- *Si on suppose simplement que* $\mathbb{E}\left[h\left(X_1,X_2\right)\right]=0$*, alors pour tout* $t>0$*,*

$$t^p\mathbb{P}\left\{\sup_{n\geqslant 1}\frac{1}{n^{1+1/p}}\left|\sum_{1\leqslant i<j\leqslant n}h\left(X_i,X_j\right)\right|>t\right\}\leqslant\kappa_p\mathbb{E}\left[\left|h\left(X_1,X_2\right)\right|^p\right].$$

Loi des logarithmes itérés bornée

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

Proposition (**Arcones, Giné** (1995))

Soient $(X_i)_{i \geqslant 1}$ 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 $\mathbb{E}\left[h(X_1, X_2)^2\right] < \infty$.

- *Si h est dégénérée pour $(X_i)_{i \geqslant 1}$ ($\mathbb{E}[h(X_1, X_2) \mid X_1] = 0$ p.s.), alors*

$$\left\| \sup_{n \geqslant 1} \frac{1}{n \,\mathrm{LL}(n)} |U_n| \right\|_p \leqslant C_p \|h(X_1, X_2)\|_2$$

- *Si on suppose simplement que $\mathbb{E}[h(X_1, X_2)] = 0$, alors*

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

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

Théorème limite central fonctionnel

Soit

$$\sigma_n(t) := \left(U_{[nt]} - (nt - [nt])\left(U_{[nt]+1} - U_{[nt]}\right)\right), t \in [0,1], n \geqslant 2,$$

où $[\cdot]$ désigne la partie entière. Autrement dit, $\sigma_n(k/n) = U_k$ et $t \mapsto \sigma_n(t)$ est continue.

Mandelbaum et Taqqu (1984) :

- Si $\mathbb{E}\left[\mathbb{E}\left[h(X_1, X_2) \mid X_1\right]^2\right] = \sigma^2 > 0$, alors

$$n^{-3/2}\sigma_n(t) \to \sigma W \text{ en loi dans } C[0,1],$$

  où W est un mouvement brownien standard.
- Si $\mathbb{E}\left[\mathbb{E}\left[h(X_1, X_2) \mid X_1\right]^2\right] = 0$, alors il existe une suite de réels $(a_i)_{i \geqslant 1}$ et une suite de mouvements browniens standard indépendants $\left(B^{(i)}\right)_{i \geqslant 1}$ tels que

$$n^{-1}\sigma_n(t) \to \sum_{i=1}^{\infty} a_i \left(\left(B_t^{(i)}\right)^2 - t\right) \text{ en loi dans } C[0,1].$$

Généralisations des U-statistiques

1. On peut considérer des U-statistiques d'ordre supérieur : pour $h\colon \mathbb{R}^k \to \mathbb{R}$,
$$U_n = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} h\left(X_{i_1}, \ldots, X_{i_k}\right)$$

2. On peut également remplacer h par une fonction dépendant de $(i_1, \ldots, i_k)$:
$$U_n = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} h_{i_1, \ldots, i_k}\left(X_{i_1}, \ldots, X_{i_k}\right)$$

3. Les variables aléatoires X_i peuvent prendre leur valeurs dans un espace measurable $(S, \mathcal{S})$.

Lien avec les champs aléatoires

On peut intepréter une U-statistique comme la somme d'un champ aléatoire non-stationnaire $\left(h\left(X_{i_1}, \ldots, X_{i_k}\right)\right)_{i_1,\ldots,i_k \in \mathbb{Z}}$ sur un certain sous-ensemble de $\mathbb{Z}^k$.

de la Pena, S. J. Montgomery-Smith (1995) : l'inégalité suivante a lieu

$$\mathbb{P}\left(\left|\sum_{1\leqslant i_1<\cdots<i_k\leqslant n} h_{i_1,\ldots,i_k}\left(X_{i_1}, \ldots, X_{i_k}\right)\right| > t\right) \leqslant C_k \mathbb{P}\left(C_k\left|\sum_{1\leqslant i_1<\cdots<i_k\leqslant n} h_{i_1,\ldots,i_k}\left(X^{(1)}_{i_1}, \ldots, X^{(k)}_{i_k}\right)\right| > t\right),$$

où $\left(X^{(1)}_{i_1}\right)_{i_1\in\mathbb{Z}}, \ldots, \left(X^{(k)}_{i_k}\right)_{i_k\in\mathbb{Z}}$ sont des copies indépendantes de $\left(X_i\right)_{i\in\mathbb{Z}}$ et C_k ne dépend que de k.

Données dépendantes, bref état de l'art

- Loi des grands nombres
  - **Arcones** (1998), **Dehling, Sharipov** (2009) (données β-mélangeantes)
  - **G.** (2021) : $X_i = f\left(\left(\varepsilon_{i-u}\right)_{u\in\mathbb{Z}}\right)$ (décalage bernoullien), où $\left(\varepsilon_u\right)_{u\in\mathbb{Z}}$ est i.i.d.
- Théorème limite central :
  - **Dehling, Wendler** (2010), α et β-mélange
  - **Arcones** (1994) : β-mélange
  - **Hsing, Wu** (2004), **G.** (2021) cas de décalage bernoullien.
- Loi des logarithmes itérés
  - **G.** (2021), décalage bernoullien,
  - **Dehling, Wendler** (2010), α et β-mélange

Plan

Introduction aux U-statistiques

Inégalités de déviation

Motivation : loi des grands nombres

Soit $U_n = \frac{1}{C_n^2} \sum_{1 \leqslant i < j \leqslant n} h(X_i, X_j)$ avec $(X_i)_{i \geqslant 1}$ i.i.d. On suppose que $\mathbb{E}[|h(X_1, X_2)|]$ et $\mathbb{E}[h(X_1, X_2)] = 0$. Alors $U_n / C_n^2 \to 0$ presque sûrement. On cherche à quantifier la convergence, par exemple en établissant la convergence de

$$\sum_{N \geqslant 2} N^{\beta} \mathbb{P}\left(\max_{2 \leqslant n \leqslant N} |U_n| > N^{\alpha} \varepsilon \right)$$

pour certains α et β.

Une inégalité permettant de majorer $\mathbb{P}(\max_{2 \leqslant n \leqslant N} |U_n| > t)$ à l'aide de la queue de $|h(X_1, X_2)|$ fournit une condition d'intégrabilité sur $h(X_1, X_2)$, α et β garantissant $\sum_{N \geqslant 2} N^{\beta} \mathbb{P}(\max_{2 \leqslant n \leqslant N} |U_n| > N^{\alpha} \varepsilon) < \infty$.

Une inégalité sur les moments combinée à l'inégalité de Markov ne serait pas suffisante dans ce contexte.

Motivation : tension pour les théorèmes limites fonctionnels

Soit $\rho\colon [0,1] \to \mathbb{R}_+$ de la forme $\rho(t) = t^{1/2} (\log(C/t))^{\beta}$.

$$\mathcal{H}_\rho^o := \left\{ x\colon [0,1] \to \mathbb{R} \mid \lim_{\delta \to 0} \sup_{0 < t-s < \delta \leqslant 1} \frac{|x(t) - x(s)|}{\rho(t-s)} = 0 \right\}.$$

On cherche à étudier la convergence de

$$\sigma_n(t) := \left(U_{[nt]} - (nt - [nt]) \left(U_{[nt]+1} - U_{[nt]} \right) \right), t \in [0,1], n \geqslant 2,$$

adéquatement normalisé, dans $\mathcal{H}_\rho^o$. Pour établir la tension il faut contrôler la queue de $|U_{n_2} - U_{n_1}|$.

Cas où h est borné, inégalité exponentielle

Théorème (Arcones (1995))

Soit $U_n = \sum_{1 \leqslant i \leqslant j \leqslant n} h(X_i, X_j)$ *une U-statistique dégénérée, avec* $H = \sup_{x,y \in \mathbb{R}} |h(x,y)| < \infty$*. Alors l'inégalité suivante a lieu pour tout* $t > 0$ *:*

$$\mathbb{P}\left(\frac{1}{n}|U_n| > t\right) \leqslant 2 \exp\left(-\frac{C}{H} t\right),$$

où C *est une constante numérique.*

Notons que l'exposant en t diffère de celui de l'inégalité d'Azuma-Hoeffding (t^2), mais est optimal car $(|U_n| / n)$ peut converger en loi vers $\left|N^2 - 1\right| / 2$ ($X_i = \pm 1$ avec probabilité $1/2$ et $h(x,y) = xy \mathbf{1}_{|x| \leqslant 1} \mathbf{1}_{|y| \leqslant 1}$.

Autres inégalités avec h borné

- Cas i.i.d. : **Houdré et Reynaud-Bouret** (2002), avec des noyaux pouvant dépendre de (i, j) et des constantes explicites.

- Suite α-mélangeantes :
  - **Shen, Han et Witten** (2020) sous une condition sur la transformée de Fourier du noyau
  - **F. Han** (2018)

- Fonctions de chaînes de Markov : **Duchemin, de Castro et Lacour** (2021), avec des noyaux pouvant dépendre de (i, j).

Une inégalité exponentielle

Théorème (G. (2021))

Soit $U_n = \sum_{1 \leqslant i \leqslant j \leqslant n} h(X_i, X_j)$ *une U-statistique dégénérée. Alors l'inégalité suivante a lieu pour tous* $t, y > 0$ *:*

$$\mathbb{P}\left(\frac{1}{N} \max_{2 \leqslant n \leqslant N} |U_n| > t\right) \leqslant A \exp\left(-B\frac{t}{y}\right) + C \int_1^{\infty} \mathbb{P}\left(|h(X_1, X_2)| > yu\right) u \log u \, \mathrm{d}u,$$

où A*,* B *et* C *sont des constantes numériques.*

Application à la vitesse de convergence dans la loi des grands nombres

Théorème (G. (2021))

Soit $U_n = \sum_{1 \leqslant i \leqslant j \leqslant n} h(X_i, X_j)$ *une U-statistique dégénérée. Soit* $\alpha \in]1, 2[$. *On suppose qu'il existe* $\gamma > 0$ *tel que*

$$\forall R > 0, \quad \mathbb{E}\left[\exp\left(R \left|h\left(X_1, X_2\right)\right|^{\gamma}\right)\right] < +\infty.$$

Alors

$$\forall \varepsilon > 0, \quad \sum_{N \geqslant 1} \exp\left(2^{N(\alpha-1)\frac{\gamma}{\gamma+1}}\right) \mathbb{P}\left\{\max_{2 \leqslant n \leqslant 2^N} |U_n| > \varepsilon 2^{N\alpha}\right\} < +\infty.$$

Application au TLCF dans les espaces hölderiens, cas non dégénéré

On rappelle que $\sigma_n\left(\frac{k}{n}\right) = U_k$ et σ_n est affine sur $[k/n, (k+1)/n]$, $1 \leqslant k \leqslant n-1$.

Soit $\rho_\beta(t) = t^{1/2}(\log(C/t))^\beta$.

Théorème (G. (2021))

Soit $(X_i)_{i \geqslant 1}$ une suite i.i.d. et h telle que $\mathbb{E}[h(X_1, X_2)] = 0$. Soit $Y = |h(X_1, X_2)|$.
On suppose que $\sigma^2 = \mathbb{E}\left[\mathbb{E}[h(X_1, X_2) \mid X_1]^2\right] \neq 0$, $\beta > 1/2$ et

$$\forall A > 0, \mathbb{E}\left[\exp\left(AY^{\frac{1}{\beta - 1/2}}\right)\right] < +\infty$$

Alors $\frac{1}{n^{3/2}}\sigma_n(\cdot) \to \sigma^2 W$ en loi dans $\mathcal{H}^o_{\rho_\beta}$, où W est un mouvement brownien standard.

Application au TLCF dans les espaces hölderiens, cas dégénéré

On rappelle que $\sigma_n\left(\frac{k}{n}\right) = U_k$ et σ_n est affine sur $[k/n, (k+1)/n]$, $1 \leqslant k \leqslant n-1$.

Soit $\rho_\beta(t) = t^{1/2}(\log(C/t))^\beta$.

Théorème (G. (2021))

Soit $(X_i)_{i\geqslant 1}$ une suite i.i.d. et h telle que $\mathbb{E}[h(X_1, X_2)] = 0$. Soit $Y = |h(X_1, X_2)|$.
On suppose que $\mathbb{E}\left[\mathbb{E}[h(X_1, X_2) \mid X_1]^2\right] = 0$, $\beta > 1$ et

$$\forall A > 0, \mathbb{E}\left[\exp\left(AY^{\frac{1}{\beta-1}}\right)\right] < +\infty$$

Alors $\frac{1}{n}\sigma_n(t) \to \sum_{i=1}^{\infty} a_i \left(\left(B_t^{(i)}\right)^2 - t\right)$ en loi dans $\mathcal{H}^o_{\rho_\beta}$, où $\left(B^{(i)}\right)$ est une suite de mouvements browniens standard indépendants et $\sum_{i\geqslant 1} a_i^2 < \infty$.

Perspectives

1. Inégalités de déviation pour des U-statistiques de la forme $\sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} h_{i_1,\ldots,i_k}(X_{i_1},\ldots,X_{i_k})$ où les fonctions $h_{i_1,\ldots,i_k}$ ne sont pas nécessairement bornées.

2. Cas de U-statistiques à valeurs dans un espace de Banach.

3. Donnés dépendantes (α-mélange, décalage de Bernoulli).

4. Applications en statistique : détection de changements multiples de paramètres.