# Vitesse de convergence dans le théorème limite central pour des sommes pondérées de champs aléatoires

Davide Giraudo

LMRS, Université de Rouen

Marseille, Séminaire de Probabilités et de Statistique, 9 juin 2017

# Champs aléatoires strictement stationnaires

## Définition

*Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Un champ aléatoire est une collection de variables aléatoires indéxé par $\mathbb{Z}^d$, où $d \in \mathbb{N}^*$.*

Soit $d \in \mathbb{N}^*$. On note $[d] := \{1, \dots, d\}$ et pour $q \in [d]$, $\mathbf{e_q} \in \mathbb{Z}^d$ est le $q$-ième vecteur de la base canonique de $\mathbb{R}^d$.

## Définition

*Un champ aléatoire $(X_\mathbf{n})_{\mathbf{n} \in \mathbb{Z}^d}$ est strictement stationnaire si pour tout entier $k$ et tous $\mathbf{n_1}, \dots, \mathbf{n_k} \in \mathbb{Z}^d$, $q \in [d]$, les vecteurs $(X_{\mathbf{n_i}})_{i=1}^k$ et $\left(X_{\mathbf{n_i}+\mathbf{e_q}}\right)_{i=1}^k$ ont la même loi.*

# Sommes pondérées de champs aléatoires stationnaires

Pour $n \geqslant 1$, soit $b_n := (b_{n,\mathbf{i}})_{\mathbf{i} \in \mathbb{Z}^d} \in \ell^2\left(\mathbb{Z}^d\right)$.

Étant donné un champ aléatoire stationnaire $(X_{\mathbf{j}})_{\mathbf{j} \in \mathbb{Z}^d}$, on s'intéresse au comportement asymptotique de la suite $(S_n)_{n \geqslant 1}$ définie par

$$S_n := \frac{1}{\|b_n\|_{\ell^2}} \sum_{\mathbf{i} \in \mathbb{Z}^d} b_{n,\mathbf{i}} X_{\mathbf{i}}.$$

# Exemple : sommation sur des ensembles

Soit $(\Lambda_n)_{n\geqslant 1}$ une suite de sous-ensembles finis de $\mathbb{Z}^d$ telle que $|\Lambda_n| := \operatorname{Card}(\Lambda_n) \to +\infty$. On pose

$$b_{n,\mathbf{i}} := \begin{cases} 1 & \text{si } \mathbf{i} \in \Lambda_n \\ 0 & \text{sinon.} \end{cases}$$

Ainsi,

$$S_n = \frac{1}{|\Lambda_n|^{1/2}} \sum_{\mathbf{i}\in\Lambda_n} X_{\mathbf{i}}.$$

# Exemple : modèle de régression

On considère le modèle de régression

$$Y_{\mathbf{i}} = g\left(\frac{\mathbf{i}}{n}\right) + X_{\mathbf{i}}, \quad \mathbf{i} \in \Lambda_n := \{1, \ldots, n\}^d,$$

où $g\colon [0,1]^d \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^\infty$ inconnue et $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}$ un champ aléatoire stationnaire. Soient $K$ un noyau défini sur $\mathbb{R}^d$ et $(h_n)_{n\geqslant 1}$ une suite convergent vers 0 et telle que $\lim_{n\to+\infty} nh_n = +\infty$ et $\lim_{n\to+\infty} nh_n^{d+1} = 0$. On pose

$$g_n(\mathbf{x}) = \frac{\sum_{\mathbf{i}\in\Lambda_n} Y_{\mathbf{i}} K\left(\frac{\mathbf{x}-\mathbf{i}/n}{h_n}\right)}{\sum_{\mathbf{i}\in\Lambda_n} K\left(\frac{\mathbf{x}-\mathbf{i}/n}{h_n}\right)}, \quad \mathbf{x} \in [0,1]^d.$$

# Question centrale

Soit $b_n := \left(b_{n,\mathbf{i}}\right)_{\mathbf{i}\in\mathbb{Z}^d} \in \ell^2\left(\mathbb{Z}^d\right)$, $\left\|b_n\right\|_{\ell^2} \neq 0$.

## Question

Donner des conditions sur les poids $\left(b_n\right)_{n\geqslant 1}$, la dépendance et les moments du champ aléatoire stationnaire $\left(X_{\mathbf{i}}\right)_{\mathbf{i}\in\mathbb{Z}^d}$ pour :

- garantir la convergence en loi de la suite $\left(S_n\right)_{n\geqslant 1}$ donnée par

$$S_n := \frac{1}{\left\|b_n\right\|_{\ell^2}} \sum_{\mathbf{i}\in\mathbb{Z}^d} b_{n,\mathbf{i}} X_{\mathbf{i}} \ ;$$

- fournir une vitesse de convergence, *i.e.*, estimer

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left\{S_n \leqslant t\right\} - \mathbb{P}\left\{N \leqslant t\right\}\right|, \quad N \sim \mathcal{N}(0,1).$$

# Champs aléatoires bernoulliens

## Définition

*Le champ aléatoire* $(X_{\mathbf{n}})_{\mathbf{n}\in\mathbb{Z}^d}$ *est dit bernoullien s'il existe un champ aléatoire indépendant identiquement distribué* $(\varepsilon_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}$ *et une fonction* $f : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}$ *telle que* $X_{\mathbf{n}} = f\left((\varepsilon_{\mathbf{n}-\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}\right)$ *pour tout* $\mathbf{n} \in \mathbb{Z}^d$.

## Remarque

*Un champ bernoullien est nécessairement strictement stationnaire.*

# Exemples de champs aléatoires bernoulliens

## Exemples

Soit $(\varepsilon_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}$ un champ i.i.d. centré tel que $\varepsilon_{\mathbf{0}} \in \mathbb{L}^2$.

- **Champ aléatoire linéaire** :

$$X_{\mathbf{n}} = \sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}}\varepsilon_{\mathbf{n}-\mathbf{i}},$$

  où $\sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}}^2 < +\infty$.

- **Champ aléatoire de Volterra** :

$$X_{\mathbf{n}} = \sum_{\mathbf{i},\mathbf{i}'\in\mathbb{Z}^d} a_{\mathbf{i},\mathbf{i}'}\varepsilon_{\mathbf{n}-\mathbf{i}} \cdot \varepsilon_{\mathbf{n}-\mathbf{i}'},$$

  où $\sum_{\mathbf{i},\mathbf{i}'\in\mathbb{Z}^d} a_{\mathbf{i},\mathbf{i}'}^2 < +\infty$ et $a_{\mathbf{i},\mathbf{i}} = 0$ pour tout $\mathbf{i} \in \mathbb{Z}^d$.

# Mesure de dépendance

**Définition (Wei Biao Wu, 2005)**

*Soient* $(X_{\mathbf{n}})_{\mathbf{n}\in\mathbb{Z}^d} := \left(f\left(\varepsilon_{\mathbf{n}-\mathbf{j}}\right)_{\mathbf{j}\in\mathbb{Z}^d}\right)_{\mathbf{n}\in\mathbb{Z}^d}$ *un champ aléatoire bernoullien,* $p \geqslant 1$ *et* $(\varepsilon'_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^d}$ *un champ i.i.d., indépendant du champ i.i.d.* $(\varepsilon_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^d}$ *et ayant la même loi que* $(\varepsilon_{\mathbf{u}})_{\mathbf{u}\in\mathbb{Z}^d}$.

*Le coefficient* $\mathbf{n}\in\mathbb{Z}^d$ *de mesure de dépendance physique est défini par*

$$\delta_{\mathbf{n},p} := \left\| X_{\mathbf{n}} - X^*_{\mathbf{n}} \right\|_p$$

*où* $X^*_{\mathbf{n}} = f\left(\left(\varepsilon^*_{\mathbf{n}-\mathbf{j}}\right)_{\mathbf{j}\in\mathbb{Z}^d}\right)$ *et* $\varepsilon^*_{\mathbf{u}} = \varepsilon_{\mathbf{u}}$ *si* $\mathbf{u} \neq \mathbf{0}$, $\varepsilon^*_{\mathbf{0}} = \varepsilon'_{\mathbf{0}}$.

# Exemples

- Soit $X_{\mathbf{n}} := \sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}}\varepsilon_{\mathbf{n}-\mathbf{i}}$ un champ linéaire. Alors $X_{\mathbf{n}} - X_{\mathbf{n}}^* = a_{\mathbf{n}}\left(\varepsilon_{\mathbf{0}} - \varepsilon_{\mathbf{0}}'\right)$ donc $\delta_{\mathbf{n},p} = |a_{\mathbf{n}}| \left\|\varepsilon_{\mathbf{0}} - \varepsilon_{\mathbf{0}}'\right\|_p$.

# Exemples

- Soit $X_{\mathbf{n}} := \sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}}\varepsilon_{\mathbf{n}-\mathbf{i}}$ un champ linéaire. Alors $X_{\mathbf{n}} - X_{\mathbf{n}}^* = a_{\mathbf{n}}\left(\varepsilon_{\mathbf{0}} - \varepsilon'_{\mathbf{0}}\right)$ donc $\delta_{\mathbf{n},p} = |a_{\mathbf{n}}| \left\|\varepsilon_{\mathbf{0}} - \varepsilon'_{\mathbf{0}}\right\|_p$.
- Soit $X_{\mathbf{n}} = \sum_{\mathbf{i},\mathbf{i}'\in\mathbb{Z}^d} a_{\mathbf{i},\mathbf{i}'}\varepsilon_{\mathbf{n}-\mathbf{i}}\cdot\varepsilon_{\mathbf{n}-\mathbf{i}'}$ un champ aléatoire de Volterra. Alors
$$\delta_{\mathbf{n},p} \leqslant C_p \left\|\varepsilon_{\mathbf{0}} - \varepsilon'_{\mathbf{0}}\right\|_p \left(\left\|\varepsilon_{\mathbf{0}}\right\|_2 + \left\|\varepsilon_{\mathbf{0}}\right\|_p\right)\left(\sum_{\mathbf{l}\in\mathbb{Z}^d} \left|a_{\mathbf{n},\mathbf{l}} + a_{\mathbf{l},\mathbf{n}}\right|^2\right)^{1/2}.$$

# Exemples

- Soit $X_{\mathbf{n}} := \sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}}\varepsilon_{\mathbf{n}-\mathbf{i}}$ un champ linéaire. Alors $X_{\mathbf{n}} - X_{\mathbf{n}}^* = a_{\mathbf{n}}\left(\varepsilon_{\mathbf{0}} - \varepsilon_{\mathbf{0}}'\right)$ donc $\delta_{\mathbf{n},p} = |a_{\mathbf{n}}| \left\|\varepsilon_{\mathbf{0}} - \varepsilon_{\mathbf{0}}'\right\|_p$.
- Soit $X_{\mathbf{n}} = \sum_{\mathbf{i},\mathbf{i}'\in\mathbb{Z}^d} a_{\mathbf{i},\mathbf{i}'}\varepsilon_{\mathbf{n}-\mathbf{i}} \cdot \varepsilon_{\mathbf{n}-\mathbf{i}'}$ un champ aléatoire de Volterra. Alors
$$\delta_{\mathbf{n},p} \leqslant C_p \left\|\varepsilon_{\mathbf{0}} - \varepsilon_{\mathbf{0}}'\right\|_p \left(\left\|\varepsilon_{\mathbf{0}}\right\|_2 + \left\|\varepsilon_{\mathbf{0}}\right\|_p\right) \left(\sum_{\mathbf{l}\in\mathbb{Z}^d} \left|a_{\mathbf{n},\mathbf{l}} + a_{\mathbf{l},\mathbf{n}}\right|^2\right)^{1/2}.$$
- Soit $f_{\mathbf{k}} : \mathbb{R} \to \mathbb{R}$ une fonction telle que la variable aléatoire $Z_{\mathbf{k}} := f_{\mathbf{k}}\left(\varepsilon_{\mathbf{0}}\right)$ soit centrée et appartienne à $\mathbb{L}^p$ et $\sum_{\mathbf{k}\in\mathbb{Z}^d} \left\|Z_{\mathbf{k}}\right\|_2^2 < +\infty$.
On pose
$$X_{\mathbf{n}} := \lim_{N\to+\infty,\mathbb{L}^2} \sum_{-N\mathbf{1}\preccurlyeq\mathbf{j}\preccurlyeq N\mathbf{1}} f_{\mathbf{k}}\left(\varepsilon_{\mathbf{n}-\mathbf{k}}\right).$$
Alors $X_{\mathbf{n}} - X_{\mathbf{n}}^* = f_{\mathbf{n}}\left(\varepsilon_{\mathbf{0}}\right) - f_{\mathbf{n}}\left(\varepsilon_{\mathbf{0}}'\right)$ donc $\delta_{\mathbf{n},2}$ est de l'ordre de $\left\|Z_{\mathbf{n}}\right\|_2$ alors que $\delta_{\mathbf{n},p}$ est de l'ordre de $\left\|Z_{\mathbf{n}}\right\|_p$.

# Théorème limite central

Soit $\tau_{\mathbf{e_q}} : \ell^2 \to \ell^2$ défini par $\tau_{\mathbf{e_q}}\left((a_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}\right) = \left(a_{\mathbf{i}+\mathbf{e_q}}\right)_{\mathbf{i}\in\mathbb{Z}^d}$.

## Théorème (Klicnarova, Volný, Wang (2016))

*Soit $(X_{\mathbf{n}})_{\mathbf{n}\in\mathbb{Z}^d}$ un champ aléatoire bernoullien vérifiant $\sum_{\mathbf{i}\in\mathbb{Z}^d} \delta_{\mathbf{i},2} < \infty$. On suppose que les poids $b_{n,\mathbf{i}}$ vérifient pour tout $q \in \{1, \dots, d\}$,*

$$\lim_{n\to+\infty} \frac{1}{\|b_n\|_{\ell^2}} \left\|\tau_{\mathbf{e_q}}(b_n) - b_n\right\|_{\ell^2} = 0.$$

*Alors la série $\sum_{\mathbf{i}\in\mathbb{Z}^d} |\mathrm{Cov}(X_{\mathbf{0}}, X_{\mathbf{i}})|$ converge et la suite $\left(\sum_{\mathbf{i}\in\mathbb{Z}^d} b_{n,\mathbf{i}} X_{\mathbf{i}} / \|b_n\|_{\ell^2}\right)_{n\geqslant 1}$ converge en loi vers une loi normale centrée ayant pour variance $\sigma^2 = \sum_{\mathbf{i}\in\mathbb{Z}^d} \mathrm{Cov}(X_{\mathbf{0}}, X_{\mathbf{i}})$.*

# Estimation de la vitesse : idée

Soient $Z$ une variable aléatoire et

$$\delta(Z) := \sup_{t\in\mathbb{R}} \left|\mathbb{P}\{Z \leqslant t\} - \mathbb{P}\{\mathcal{N}(0,1) \leqslant t\}\right|.$$

- **Chen, Shao (2004)** Si $I \subset \mathbb{Z}^d$ est fini, $(Y_{\mathbf{i}})_{\mathbf{i}\in I}$ est un champ aléatoire $m$-dépendant centré tel que $\mathbb{E}\left[|Y_{\mathbf{i}}|^p\right] < +\infty$ pour tout $\mathbf{i} \in I$ et un certain $p \in (2,3]$ et $\operatorname{Var}\left(\sum_{\mathbf{i}\in I} Y_{\mathbf{i}}\right) = 1$, alors

$$\delta\left(\sum_{\mathbf{i}\in I} Y_{\mathbf{i}}\right) \leqslant 75\,(10m+1)^{(p-1)d} \sum_{\mathbf{i}\in I} \mathbb{E}\left[|Y_{\mathbf{i}}|^p\right].$$

- Le champ $X_{\mathbf{i}}^{(m)} := \mathbb{E}\left[X_{\mathbf{i}} \mid \sigma\left(\varepsilon_{\mathbf{u}}, \mathbf{i} - m\mathbf{1} \preccurlyeq \mathbf{u} \preccurlyeq \mathbf{i} + m\mathbf{1}\right)\right]$ est $(2m+1)$-dépendant.
- **El Machkouri, Ouchti (2005)** Pour toutes variables aléatoires $Z$ et $Z'$ et tout $p \geqslant 1$,

$$\delta(Z+Z') \leqslant 2\delta(Z) + \|Z'\|_p^{\frac{p}{p+1}}.$$

# Inégalité de moments

### Théorème (El Machkouri, Volný, Wu, 2013)

*Soient* $(a_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d} \in \ell^2\left(\mathbb{Z}^d\right)$ *et* $p \geqslant 2$. *Pour tout champ aléatoire bernoullien* $(X_{\mathbf{n}})_{\mathbf{n}\in\mathbb{Z}^d}$, *l'inégalité suivante a lieu :*

$$\left\|\sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}} X_{\mathbf{i}}\right\|_p \leqslant \left(2p \sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}}^2\right)^{1/2} \cdot \sum_{\mathbf{j}\in\mathbb{Z}^d} \delta_{\mathbf{j},p}.$$

Inégalité de Rosenthal (1972) : pour $p \geqslant 2$, il existe $C = C(p)$ telle que pour toute suite indépendante centrée $(Y_i)_{i\in\mathbb{N}}$ et tout $n \in \mathbb{N}^*$,

$$\mathbb{E}\left[\left|\sum_{i=1}^n Y_i\right|^p\right] \leqslant C \sum_{i=1}^n \mathbb{E}\left[|Y_i|^p\right] + C\left(\sum_{i=1}^n \mathbb{E}\left[Y_i^2\right]\right)^{p/2}.$$

# Inégalité de Rosenthal

## Théorème (G., 2017)

*Soit* $\left\{\varepsilon_{\mathbf{i}}, \mathbf{i} \in \mathbb{Z}^d\right\}$ *un champ i.i.d. Pour toute fonction mesurable* $f : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}$ *telle que* $X_{\mathbf{n}} := f\left(\left(\varepsilon_{\mathbf{n}-\mathbf{i}}\right)_{\mathbf{i}\in\mathbb{Z}^d}\right)$ *admette un moment fini d'ordre* $p \geqslant 2$ *et soit centrée, et toute famille* $\left(a_{\mathbf{i}}\right)_{\mathbf{i}\in\mathbb{Z}^d} \in \ell^2\left(\mathbb{Z}^d\right)$*,*

$$\left\|\sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}} X_{\mathbf{i}}\right\|_p \leqslant \frac{14.5p}{\log p}\left(\sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}}^2\right)^{1/2} \sum_{\mathbf{j}\in\mathbb{N}^d}\left(4\left\|\mathbf{j}\right\|_\infty + 4\right)^{d/2}\delta_{\mathbf{j},2}$$
$$+ \frac{14.5p}{\log p}\left(\sum_{\mathbf{i}\in\mathbb{Z}^d} \left|a_{\mathbf{i}}\right|^p\right)^{1/p} \sum_{\mathbf{j}\in\mathbb{N}^d}\left(4\left\|\mathbf{j}\right\|_\infty + 4\right)^{d(1-1/p)}\delta_{\mathbf{j},p}.$$

# Idée de démonstration

Pour $j \in \mathbb{N}$ et $\mathbf{i} \in \mathbb{Z}^d$, on pose

$$X_{\mathbf{i},j} = \mathbb{E}\left[X_i \mid \sigma\left(\varepsilon_{\mathbf{u}}, \|\mathbf{u}-\mathbf{i}\|_\infty \leqslant j+1\right)\right] - \mathbb{E}\left[X_i \mid \sigma\left(\varepsilon_{\mathbf{u}}, \|\mathbf{u}-\mathbf{i}\|_\infty \leqslant j\right)\right].$$

On fixe $j$ et on contrôle $\left\|\sum_{\mathbf{i}\in\mathbb{Z}^d} a_{\mathbf{i}} X_{\mathbf{i},j}\right\|_p$.

Pour $d=2$ : on coupe $\mathbb{Z}^2$ en blocs $B_{v,w}$, $v,w\in\mathbb{Z}$ de sorte que si $S_{v,w} = \sum_{\mathbf{i}\in B_{v,w}} a_{\mathbf{i}} X_{\mathbf{i},j}$, alors

- la famille $\left(S_{2v,2w}\right)_{v,w\in\mathbb{Z}}$ est indépendante ;
- la famille $\left(S_{2v+1,2w}\right)_{v,w\in\mathbb{Z}}$ est indépendante ;
- la famille $\left(S_{2v,2w+1}\right)_{v,w\in\mathbb{Z}}$ est indépendante ;
- la famille $\left(S_{2v+1,2w+1}\right)_{v,w\in\mathbb{Z}}$ est indépendante.

On applique l'inégalité de Rosenthal à chaque famille, et on contrôle $\left\|X_{\mathbf{i},j}\right\|_p$ par $\sum_{\mathbf{i}:\|\mathbf{i}\|_\infty=j+1} \delta_{\mathbf{i},p}$.

# Résultat général

## Théorème (G., 2017)

*Soit* $p > 2$, $p' := \min\{p, 3\}$ *et* $(X_{\mathbf{j}})_{\mathbf{j}\in\mathbb{Z}^d}$, $b_n$ *comme précédemment. On suppose qu'il existe* $\alpha, \beta > 0$ *tels que les séries suivantes convergent*

$$C_2 := \sum_{\mathbf{i}\in\mathbb{Z}^d} \left(\|\mathbf{i}\|_\infty + 1\right)^{d/2+\alpha} \delta_{\mathbf{i},2} \text{ et } C_p := \sum_{\mathbf{i}\in\mathbb{Z}^d} \left(\|\mathbf{i}\|_\infty + 1\right)^{d(1-1/p)+\beta} \delta_{\mathbf{i},p}.$$

*On suppose que* $\sigma^2 = \sum_{\mathbf{i}\in\mathbb{Z}^d} \operatorname{Cov}(X_{\mathbf{0}}, X_{\mathbf{i}}) > 0$. *Soient*

$$\Delta_n := \sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\sum_{\mathbf{i}\in\mathbb{Z}^d} b_{n,\mathbf{i}} X_{\mathbf{i}}}{\|b_n\|_{\ell^2}} \leqslant t \right\} - \mathbb{P}\left\{ \mathcal{N}\left(0, \sigma^2\right) \leqslant t \right\} \right| \text{ et}$$

$$\varepsilon_n := \sum_{\mathbf{j}\in\mathbb{Z}^d} \mathbb{E}\left[X_{\mathbf{0}} X_{\mathbf{j}}\right] \left( \sum_{\mathbf{i}\in\mathbb{Z}^d} \frac{b_{n,\mathbf{i}} b_{n,\mathbf{i}+\mathbf{j}}}{\|b_n\|_{\ell^2}} - 1 \right).$$

*Soit* $\gamma > 0$. *Il existe* $\kappa$ *et* $n_0 = n_0(\eta)$ *tel que pour tout* $n \geqslant n_0$,

$$\Delta_n \leqslant \kappa \left( \|b_n\|_{\ell^2}^{\gamma(p'-1)d-p'} \|b_n\|_{\ell^{p'}}^{p'} + |\varepsilon_n| + \|b_n\|_{\ell^2}^{-\gamma\alpha\frac{p}{p+1}} + \|b_n\|_{\ell^p}^{\frac{p}{p+1}} \|b_n\|_{\ell^2}^{-\frac{p}{p+1}(\gamma\beta+1)} \right).$$

# Sommes sur des ensembles

## Corollaire

*Soient* $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}$ *un champ aléatoire bernoullien ayant un moment fini d'ordre* $p \geqslant 2$, $p' := \min\{p, 3\}$ *et* $(\Lambda_n)_{n\geqslant 1}$ *une suite de sous-ensembles de* $\mathbb{Z}^d$ *tels que* $|\Lambda_n| \to +\infty$ *et pour tout* $\mathbf{k}\in\mathbb{Z}^d$, $\lim_{n\to+\infty} |\Lambda_n \cap (\Lambda_n - \mathbf{k})| / |\Lambda_n| = 0$. *On suppose qu'il existe* $\alpha, \beta > 0$ *tels que les séries suivantes convergent*

$$C_2 := \sum_{\mathbf{i}\in\mathbb{Z}^d} \left(\|\mathbf{i}\|_\infty + 1\right)^{d/2+\alpha} \delta_{\mathbf{i},2} \text{ et } C_p := \sum_{\mathbf{i}\in\mathbb{Z}^d} \left(\|\mathbf{i}\|_\infty + 1\right)^{d(1-1/p)+\beta} \delta_{\mathbf{i},p}.$$

*et* $\sigma^2 = \sum_{\mathbf{i}\in\mathbb{Z}^d} \operatorname{Cov}(X_{\mathbf{0}}, X_{\mathbf{i}}) > 0$ *Soit* $\gamma > 0$. *Il existe* $n_0 = n_0(\gamma)$ *et* $\kappa$ *telles que pour tout* $n \geqslant n_0$,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\sum_{\mathbf{i}\in\Lambda_n} X_{\mathbf{i}}}{|\Lambda_n|^{1/2}} \leqslant t \right\} - \mathbb{P}\left\{ \mathcal{N}\left(0, \sigma^2\right) \leqslant t \right\} \right| \leqslant \kappa \left( |\Lambda_n|^q + \sum_{\mathbf{j}\in\mathbb{Z}^d} |\mathbb{E}[X_{\mathbf{0}} X_{\mathbf{j}}]| \, a_{n,\mathbf{j}} \right),$$

*où* $a_{n,\mathbf{j}} = \left| \frac{|\Lambda_n \cap (\Lambda_n - \mathbf{j})|}{|\Lambda_n|} - 1 \right|$ *et*

$$q := \max\left\{ \frac{\gamma(p'-1)d - p'}{2} + 1; -\gamma\alpha\frac{p}{2(p+1)}; \frac{1 - p - p\gamma\beta}{p+1} \right\}.$$

# Modèle de régression (1)

On considère le modèle de régression

$$Y_{\mathbf{i}} = g\left(\frac{\mathbf{i}}{n}\right) + X_{\mathbf{i}}, \quad \mathbf{i} \in \Lambda_n := \{1, \ldots, n\}^d,$$

où $g\colon [0,1]^d \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^\infty$ inconnue et $(X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}$ un champ aléatoire stationnaire. Soient $K$ un noyau défini sur $\mathbb{R}^d$ et $(h_n)_{n\geqslant 1}$ une suite convergent vers 0 et telle que $\lim_{n\to+\infty} nh_n = +\infty$ et $\lim_{n\to+\infty} nh_n^{d+1} = 0$. On pose

$$g_n(\mathbf{x}) = \frac{\sum_{\mathbf{i}\in\Lambda_n} Y_{\mathbf{i}} K\left(\frac{\mathbf{x}-\mathbf{i}/n}{h_n}\right)}{\sum_{\mathbf{i}\in\Lambda_n} K\left(\frac{\mathbf{x}-\mathbf{i}/n}{h_n}\right)}, \quad \mathbf{x} \in [0,1]^d.$$

Alors

$$g_n(\mathbf{x}) - \mathbb{E}\left[g_n(\mathbf{x})\right] = \frac{\sum_{\mathbf{i}\in\Lambda_n} X_{\mathbf{i}} K\left(\frac{\mathbf{x}-\mathbf{i}/n}{h_n}\right)}{\sum_{\mathbf{i}\in\Lambda_n} K\left(\frac{\mathbf{x}-\mathbf{i}/n}{h_n}\right)}, \quad \mathbf{x} \in [0,1]^d.$$

# Modèle de régression (2)

Hypothèses sur $K$ :

- $\int_{\mathbb{R}^d} K(\mathbf{u})\,\mathrm{d}\mathbf{u} = 1$,
- $K$ est symétrique, positif, a pour support $[-1,1]^d$.
- Il existe $r$, $c$ et $C$ telles que pour tous $\mathbf{x}, \mathbf{y} \in [-1,1]^d$, $|K(\mathbf{x}) - K(\mathbf{y})| \leqslant r \|\mathbf{x} - \mathbf{y}\|_\infty$ et $c \leqslant K(\mathbf{x}) \leqslant C$.

Soit

$$A_n := (nh_n)^{d/2} \left( \sum_{\mathbf{i} \in \Lambda_n} K^2 \left( \frac{\mathbf{x} - \mathbf{i}/n}{h_n} \right) \right)^{1/2} \|K\|^{-1}_{\mathbb{L}^2(\mathbb{R}^d)} \left( \sum_{\mathbf{i} \in \Lambda_n} K \left( \frac{\mathbf{x} - \mathbf{i}/n}{h_n} \right) \right)^{-1/2} .$$

Sous ces hypothèses, la suite $(A_n)_{n \geqslant 1}$ converge vers $1$.

# Modèle de régression (3)

## Théorème (G., 2017)

*Soient* $p > 2$, $p' := \min\{p, 3\}$ *et* $(X_{\mathbf{j}})_{\mathbf{j}\in\mathbb{Z}^d} = f\left((\varepsilon_{\mathbf{j}-\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}\right)$ *un champ aléatoire bernoullien ayant un moment fini d'ordre* $p$. *On suppose qu'il existe* $\alpha, \beta > 0$ *tels que*

$$C_2 := \sum_{\mathbf{i}\in\mathbb{Z}^d} \left(\|\mathbf{i}\|_\infty + 1\right)^{d/2+\alpha} \delta_{\mathbf{i},2} < \infty \text{ et } C_p := \sum_{\mathbf{i}\in\mathbb{Z}^d} \left(\|\mathbf{i}\|_\infty + 1\right)^{d(1-1/p)+\beta} \delta_{\mathbf{i},p} < \infty.$$

*On suppose que* $\sigma^2 := \sum_{\mathbf{j}\in\mathbb{Z}^d} \operatorname{Cov}(X_{\mathbf{0}}, X_{\mathbf{j}}) > 0$. *Soit* $\Delta_n :=$ $\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left\{(nh_n)^{d/2}\left(g_n(\mathbf{x}) - \mathbb{E}[g_n(\mathbf{x})]\right) \leqslant t\right\} - \mathbb{P}\left\{\mathcal{N}\left(0, \sigma^2 \|K\|_2^2\right) \leqslant t\right\}\right|$,

$$\varepsilon_n := \sum_{\mathbf{j}\in\mathbb{Z}^d} \mathbb{E}[X_{\mathbf{0}}X_{\mathbf{j}}] \left( \sum_{\mathbf{i}\in\Lambda_n\cap(\Lambda_n-\mathbf{j})} \frac{K\left(\frac{\mathbf{x}-\mathbf{i}/n}{h_n}\right) K\left(\frac{\mathbf{x}-(\mathbf{i}-\mathbf{j})/n}{h_n}\right)}{\sum_{\mathbf{k}\in\Lambda_n} K^2\left(\frac{\mathbf{x}-\mathbf{k}/n}{h_n}\right)} - 1 \right).$$

*Soit* $\gamma > 0$. *Il existe* $\kappa > 0$ *et* $n_0 = n_0(\gamma)$ *tels que pour tout* $n \geqslant n_0$,

$$\Delta_n \leqslant \kappa |A_n - 1|^{\frac{p}{p+1}} + |\varepsilon_n| + \kappa (nh_n)^{\frac{d}{2}\left(\gamma(p'-1)d - p' + 2\right)} + (nh_n)^{-\frac{d}{2}\gamma\alpha\frac{p}{p+1}} + (nh_n)^{\frac{2d-p(\gamma\beta+1)}{2(p+1)}}.$$

# Perspectives

- Inégalités de déviation: au lieu de contrôler les moments de $Y := \sum_{\mathbf{i} \in \mathbb{Z}^d} a_{\mathbf{i}} X_{\mathbf{i}}$, on peut chercher une majoration de $\mathbb{P}\{|Y| > t\}$.

- $U$-statistiques : de la forme $U_n = \sum_{\mathbf{i} \in I_n} h(X_{i_1}, \dots, X_{i_r})$, où $I_n = \left\{\mathbf{i} \in \mathbb{N}^d \mid 1 \leqslant i_1 < i_2 < \dots < i_{r-1} < i_r \leqslant n\right\}$, $h \colon \mathbb{R}^r \to \mathbb{R}$ est un noyau symétrique et $(X_k)_{k \geqslant 1}$ est une suite strictement stationnaire.