TCL pour U-statistiques géométriques dans le chaos poissonien

GT Probas, MAP 5
Université Paris Descartes

Raphaël Lachièze-Rey
En collaboration avec Giovanni Peccati, Université de Luxembourg

15 novembre 2011

Définitions

Dans cet exposé :

- μ : mesure loc. finie sur X loc. compact (sans atomes).
- η : Processus ponctuel de Poisson d'intensité μ.
- η_λ : Processus ponctuel de Poisson d'intensité $\lambda\mu, \lambda > 0$.
- $h(x_1, \dots, x_k)$: Fonction symétrique $X^k \mapsto \mathbb{R}$.

[Reitzner, Schulte 2011] : **U-statistique de noyau** h :

$$\boxed{F = F(\eta) = \sum_{x_1,\dots,x_k \in \eta,\, x_i \neq x_j} h(x_1, \dots, x_k).}$$

Comportement asymptotique de $F_\lambda := F(\eta_\lambda)$.
Question :

$$\tilde{F}_\lambda = \frac{F_\lambda - \mathbb{E}F_\lambda}{\sqrt{\mathrm{Var}(F_\lambda)}} \Rightarrow \mathcal{N}(0,1)$$

quand $\lambda \to \infty$?

Exemples 1 : Graphes aléatoires (k=2)

Règle : On connecte x et y s'ils satisfont une certaine condition géométrique déterministe :

- Unit disk graph : $d(x,y) \leq \alpha_\lambda$ pour une métrique d et $\alpha_\lambda > 0$.

notation : $x \sim y$.

$$F_\lambda = \sum_{x \neq y \in \eta_\lambda} \varphi(x,y) 1_{x \sim y}$$

- **Exemple** : Longueur totale des arêtes.

Hypergaphes aléatoires : k=2,3,...

- Complexe de Vietoris-Rips : On connecte $(x_1,\dots,x_k)$ si $diam(x_1,\dots,x_k) \leq \alpha_\lambda$.
- **Exemple** : Aire totale des triangles ayant des sommets dans η_λ.

Exemple 2 : Particules en interaction

On suppose que les points de η_λ interagissent k par k :

$$\lambda = \sum_{x_1,\ldots,x_k \in \eta_\lambda, x_i \neq x_j} U(x_1, \ldots, x_k)$$

- Hypothèses sur la décroissance de U à l' ∞.

Intersection de sous-espaces

[Thaele, Schulte 2011]

- W : corps convexe.
- $\mathcal{H}$: $\{H_1, H_2, \ldots\}$ Processus de Poisson ponctuel d'intensité $\lambda\mu$,$\lambda > 0$, dans l'espace des plans affines de $\mathbb{R}^2$.

F : $\#$ Nombres d'intersection de plans dans $W \Rightarrow$ U-stat. de noyau :

$$h(H_1, H_2) = \mathbf{1}_{H_1 \cap H_2 \cap W \neq \emptyset}.$$

Généralisation :

- $F_{k,j}$: j-ème volume intrinsèque d'un processus de sous-espaces de dimension k dans $\mathbb{R}^d$.
- Thaele et Schulte ont montré la normalité asymptotique de $F_{k,j}$ lorsque $\lambda \to \infty$ (en utilisant la décomposition de Wiener-Ito).

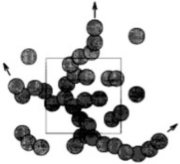
Processus marqués

Modèle germe-grain :

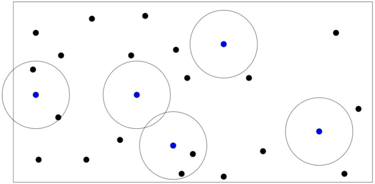
- $\mathcal{K}$: compacts de $\mathbb{R}^d$.
- η à valeurs dans $\mathbb{R}^d \times \mathcal{K}$

Pour $k \geq 1$:

$$F = \#\{\text{ chaines de longueur } k\}$$



Télécommunications :



bleus : serveurs
noir : clients.
$F = \#$ clients desservis (comptés avec leur multiplicité).

Intégrale de Wiener-Ito

- Soit $f(x_1,\ldots,x_k)$: noyau sur X^k.
- **Intégrale de Wiener-Ito d'ordre k par rapport à η** :

$$I_k(f)=\int_{X^k} f(x_1,\ldots,x_k)d(\eta-\mu)^{\otimes k}(x_1,\ldots,x_k).$$

- **Ecriture intégrale d'une U-statistique** :

$$U_k(h)=\int_{X^k} h(x_1,\ldots,x_k)d\eta^{\otimes k}(x_1,\ldots,x_k).$$

Une U-statistique admet une écriture comme somme finie d'intégrales de Wiener-Ito (formule du binome).

$$U_k(h)=\sum_{j=0}^{k} I_j(h_j)$$

$$h_j(x_1,\ldots,x_j)=\binom{k}{j}\int_{X^{k-j}} h(x_1,\ldots,x_k)d\mu^{\otimes k-j}(x_{j+1},\ldots,x_k).$$

Chaos poissonien

Théorème

Toute variable aléatoire $F = F(\eta)$ telle que $\mathbb{E}(F^2) < \infty$ admet dans $L^2(\mathbb{P})$ une unique décomposition orthogonale

$$F = \sum_{j \geq 0} I_j(f_j),$$

où $f_j \in L^2(X^j, \mathbb{R})$.

Terme constant : $I_0(f_0) = \mathbb{E}(F)$.
Remarque : $\mathrm{Var}(I_j(f_j)) = \|f_j\|^2_{L^2(\mu^j)}$. On pose $\tilde{f}_j = \frac{f_j}{\|f_j\|_{L^2}}$

- Isoler les noyaux f_j de variance prédominante,
- Montrer que chaque $I_j(\tilde{f}_j)$ suit un TCL.

Méthode de Stein

Lemme

Une variable aléatoire N a la loi normale ssi pour toute fonction h dérivable à support compact,

$$\mathbb{E}(h(N) - Nh'(N)) = 0.$$

Méthode de Stein :
Calculer la distance entre la loi d'une variable F et d'une gaussienne N en terme de $|\mathbb{E}(h(F) - Fh'(F))|$, pour h lisse.

Distance de Wasserstein

$$\boxed{d_W(F,N) = \sup_{h \in Lip_1} |\mathbb{E}h(F) - h(N)|}$$

$d_W(F,N) \to 0$ implique $F \Rightarrow \mathcal{N}(0,1)$.
En particulier,

$$d_K(F,N) = \sup_{t \in \mathbb{R}} |\mathbb{P}(F \leq t) - \mathbb{P}(N \leq t)| \leq \sqrt{d_W(F,N)}$$

Reitzner, Schulte (à venir) : Bornes en termes de d_K.

Méthode de Stein (suite)

Pour chaque $h \in Lip_1$, l'équation de Stein

$$g'(y) - yg(y) = h(y) - \mathbb{E}(h(N))$$

admet une solution : g_h, qui satisfait

$$|g_h'(y)| \leq 1, \quad |g_h''(y)| \leq 2, \; y \in \mathbb{R}.$$

On en déduit que

$$d_W(F, N) = \sup_{h \in Lip_1} |\mathbb{E}(h(F)) - \mathbb{E}(h(N))|$$

$$\boxed{d_W(F, N) \leq \sup_{|g'| \leq 1, |g''| \leq 2} |\mathbb{E}(g'(F) - Fg(F))|.}$$

C'est la **Borne de Stein** (existe pour d'autres lois que la loi normale).
Pour estimer cette borne, on utilise le **calcul de Malliavin**.

Calcul de Malliavin

Idee : Exprimer $g'(F)$ et $Fg(F)$ en termes d'operateurs linéaires sur la serie

$$F = \sum_{q \geq 1} I_q(f_q).$$

Opérateurs linéaires

- $D_z F = \sum_{q \geq 1} q I_{q-1}(f_q(\cdot, z))$,
- $LF = -\sum_{q \geq 1} q I_q(f)$
- $L^{-1}F = -\sum_{q \geq 1} \frac{1}{q} I_q(f)$.

Peccati, Sole, Utzet, Taqqu 2010 :

- Bornes génériques en termes des opérateurs de Malliavin

$$d_W(F, N) \leq \mathbb{E}|1 - \langle DF, -DL^{-1}F \rangle| + \int_X \mathbb{E}[D_z F^2 | D_z L^{-1} F] \mu(dz).$$

Reitzner, Schulte 2011

η_λ : Processus de Poisson d'intensité $\lambda\mu, \lambda > 0$.

Théorème

Soit une U-statistique $U_k(h_\lambda)$ *telle que* $U_k(|h_\lambda|) \in L^2(\mathbb{P})$.

- $h_\lambda = h$ *ne dépend pas de* λ *(U-stat. géométrique) et* $h_1 \neq 0$ *le 1er noyau n'est pas identiquement nul.*
- *OU* $h_\lambda(x_1, \ldots, x_k) = \mathbf{1}_{\mathrm{diam}(x_1,\ldots,x_k) \geq \lambda^{1/d}}$ *(+Bornes plus générales pour "U-statistiques locales")*

Alors $d_W(\tilde{F}, N) \leq C\lambda^{-1/2}$.

Applications (Reitzner, Schulte 2011)

- Intersections d'hyperplans dans un convexe.
- **Problème de Sylvester**. η_λ : PPP dans un convexe compact W. Avec
$$h(x_1, \ldots, x_k) = \mathbf{1}_{\{x_1,\ldots,x_k \text{ sont en position convexe}\}},$$
$F_\lambda = U_k(h)$ est asymptotiquement normal, donc $\lambda^{-k} F_\lambda$ est un estimateur asymptotiquement gaussien de
$$p_k = \mathbb{P}(\ k \text{ variables uniformes dans } W \text{ sont en position convexe})$$
- Longueur totale du **Unit disk graph** (2 points sont reliés s'ils sont à distance $\leq \lambda^{-1/d}$ dans $[0,1]^d$).

Contractions

Formule de multiplication

$$I_q(f)I_k(g) = \sum_{r=0}^{q\wedge k}\sum_{l=0}^{r} \varkappa_{q,k,r,l} I_{q+k-r-l}(f \star_r^l g)$$

- $\varkappa_{q,k,r,l}$: Terme combinatoire.
- $f \star_r^l g$: Symmétrisée de la contraction de f et g.

$$\begin{aligned} f\star_r^l g(x_1,\dots,x_{k-l},t_1,\dots,t_{k-l},y_1,\dots,y_{r-l}) \\ = \int f(x_1,\dots,x_{q-l},y_1,\dots,y_{r-l},z_1,\dots,z_l) \\ g(t_1,\dots,t_{k-l},y_1,\dots,y_{r-l},z_1,\dots,z_l)d\mu^{\otimes l}(z_1,\dots,z_l) \end{aligned}$$

On met r variables en commun, parmi lesquelles on en intègre l.

Bornes en termes de contractions

Théorème (Peccati, Solé, Utzet, Taqqu 2010)

Soit $F = \sum_{i=1}^{k} I_i(f_i)$, $f_i \in L^2(X^i)$.
Alors

$$d_W(F, N) \leq C[\max_{1} \|f_i \star_r^l f_j\|_{L^2} + \max_{1 \leq i \leq k} \|f_i\|_{L^4}] + C' \frac{\mathrm{Var}^2(F) - 1}{\mathrm{Var}(F) \vee 1}.$$

où $\max_{1}$ *porte sur*

$$i \leq j \leq k, 0 \leq l \leq r \leq i, (i, j) \neq (1, 1), r \geq 1, l < j.$$

Exemple pour $k = q = 2$

Soit

- $f : X^2 \mapsto \mathbb{R}_+$

telle que $\|f\|_{L^2} \neq 0$.

$$f \star_1^1 f(x,y) = \int f(x,z) f(y,z) \mu(dz)$$

$$\|\tilde{f} \star_1^1 \tilde{f}\|_{L^2}^2 = \frac{\int f(x,z) f(y,z) f(x,t) f(y,t) \mu^{\otimes 4}(dx,dy,dz,dt)}{\int f(x,y) f(x,y) f(z,t) f(z,t) \mu^{\otimes 4}(dx,dy,dz,dt)}$$

Si f est "petite loin de la diagonale" (ex : $f(x,y) = \mathbf{1}_{\|x-y\| \leq \lambda^{-1/d}}$),

- **intégrale numérateur** : on ne compte que les termes où (x,y,z,t) sont "proches"
- **Intégrale dénominateur** : on compte les termes où (x,y) sont proches, et (z,t) sont proches. ($\Rightarrow$ plus de termes !)

Dernier ingrédient : TCL multi-dimensionnel

Théorème (Peccati, Zheng 2010)

Soit $F = \sum_{i=1}^{k} I_i(f_i)$*,* $f_i \in L^2(X^i)$*.*
Si

$$\max_{1} \|f_i \star_r^l f_j\|_{L^2} + \max_{1 \leq i \leq k} \|f_i\|_{L^4} \to 0,$$

alors pour φ *3 fois dérivable tq* $\|\varphi''\|_\infty, \|\varphi'''\|_\infty < \infty$

$$\mathbb{E}(\varphi(I_1(f_1), \dots, I_k(f_k))) - \mathbb{E}(\varphi(N_1, \dots, N_k)) \to 0$$

où $N_i \sim \mathcal{N}(0, 1)$*.*

"4th moment bound" pour les U-stat. à noyau positif

On suppose que $h_i \geq 0, 1 \leq i \leq k$.

Théorème

Si

$$B = [\max_1 \|f_i \star_r^l f_j\|_{L^2} + \max_{1 \leq i \leq k} \|f_i\|_{L^4}]$$

$$cB \leq \sqrt{\mathbb{E}(F^4 - 3)} \leq CB$$

- Traditionellement, les auteurs vérifiaient $\mathbb{E}(F_\lambda^k) \to \mathbb{E}(N^k)$, pour $k \geq 1$.
- La condition sur les contractions est **nécessaire** pour la normalité asymptotique si $\{F_\lambda^4;\ \lambda > 0\}$ est uniformément intégrable.

U-stat. géométriques

$$F = \sum_{\eta_k^{\neq}} h(x_1, \ldots, x_k) = \sum_{i=1}^{k} I_i(h_i).$$

où h ne dépend pas de λ.

Théorème

- *Si $h_1 \neq 0$, $\tilde{F} \Rightarrow N$ à la vitesse $\lambda^{-1/2}$ (pour d_W). (R,S, '11)*
- *Si $h_1 = 0$ et $h_2 \neq 0$, $\tilde{F}$ n'est pas aymptotiquement normale.*
- *Si $h_1 = 0, h_2 = 0, \ldots, h_{i-1} = 0, h_i \neq 0, h_i \geq 0$, $\tilde{F}$ ne suit pas un TCL.*

Graphes aléatoires géométriques

On suppose

$$F_\lambda = \sum_{x \neq y \in \eta_\lambda} \mathbf{1}_{\alpha_\lambda(x-y)\in H} = \mathbb{E}(F_\lambda) + I_1(h_1) + I_2(h_2).$$

pour un certain $H \subseteq \mathbb{R}^d$ (Unit disk graph :$H = B(0,1)$) et un facteur d'échelle $\alpha_\lambda > 0$. On pose

$$\psi(\lambda) = \frac{\ell(H \cap Q_\lambda)}{\ell(Q_\lambda)}$$

où $Q_\lambda = [-\lambda^{1/d}, \lambda^{1/d}]^d$.

Conditions analytiques sur H

On pose $F = \mathbb{E}(F) + F_1 + F_2$.

Théorème

On fait "l'hypothèse de régularité" que

$$0 < \inf_{\lambda > 0} \frac{\psi(c\lambda)}{\psi(\lambda)} \leq \sup_{\lambda > 0} \frac{\psi(c\lambda)}{\psi(\lambda)} < \infty,$$

pour tout $c > 0$. Alors

$$\begin{aligned} \mathrm{Var}(\tilde{F}_1) &\sim \lambda^3 \psi(\lambda)^2 \\ \mathrm{Var}(\tilde{F}_2) &\sim \lambda^2 \psi(\lambda) \\ d_W(\tilde{F}_2, N) &\leq C\sqrt{\psi(\lambda)} \end{aligned}$$

U-stats locales généralisées

$$F = \sum_{x_1,\ldots,x_k \in \eta_\lambda^{\neq}} h(\lambda^{1/d}(x_1,\ldots,x_k))$$

pour $h \geq 0$ stationnaire sur $(\mathbb{R}^d)^k$.

Théorème

Si il existe $\varepsilon > 0$ tel que

$$|h(0,x_2,\ldots,x_k)| \leq C \prod_{i=2}^{k} 1 \wedge \|x_i\|^{-d-\varepsilon},$$

Alors $\tilde{F}$ suit un TCL à la vitesse $\lambda^{-1/2}$.

Remarques

- Ne rentre pas dans le cadre de la stabilisation (la contribution d'un x dépend de ce qui se passe loin de x)
- Possibilité de considérer des processus stationnaires marqués.
- Autres remises à l'échelle possibles