Lecture 5: Rates of multivariate normal approximation for statistics of geometric structures

Spring school at Darmstadt, 25 February-March 1, 2019

Probabilistic Analysis of Geometric Structures

· **Lecture 1**: Probabilistic analysis of Euclidean optimization problems

· **Lecture 2**: Central limit theorems for statistics of geometric structures

· **Lecture 3**: Limit theory for statistics of geometric structures via stabilizing score functions

· **Lecture 4**: Statistics of random polytopes

· **Lecture 5**: Rates of multivariate normal approximation for statistics of geometric structures

Lecture 5: Rates of multivariate normal approximation for statistics of geometric structures

Spring school at Darmstadt, 25 February-March 1, 2019

Introduction

Questions pertaining to geometric structures on random input $\mathcal{X} \subset \mathbb{R}^d$ often involve analyzing sums of spatially correlated terms

$$\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X}),$$

where the $\mathbb{R}$-valued score function ξ, defined on pairs $(x, \mathcal{X})$, represents the interaction of x with respect to $\mathcal{X}$.

Introduction

Questions pertaining to geometric structures on random input $\mathcal{X} \subset \mathbb{R}^d$ often involve analyzing sums of spatially correlated terms

$$\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X}),$$

where the $\mathbb{R}$-valued score function ξ, defined on pairs $(x, \mathcal{X})$, represents the interaction of x with respect to $\mathcal{X}$.

The sums describe some global feature of the random structure in terms of local contributions $\xi(x, \mathcal{X}),\ x \in \mathcal{X}$.

Introduction

Questions pertaining to geometric structures on random input $\mathcal{X} \subset \mathbb{R}^d$ often involve analyzing sums of spatially correlated terms

$$\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X}),$$

where the $\mathbb{R}$-valued score function ξ, defined on pairs $(x, \mathcal{X})$, represents the interaction of x with respect to $\mathcal{X}$.

The sums describe some global feature of the random structure in terms of local contributions $\xi(x, \mathcal{X}),\ x \in \mathcal{X}$.

When $\mathcal{X}$ is a Poisson point process $\mathcal{P}_s$ of intensity s on a fixed subset of $\mathbb{R}^d$, then much is known concerning central limit theorems for $\sum_{x \in \mathcal{P}_s} \xi(x, \mathcal{P}_s)$ as the intensity s tends to ∞.

Total edge length of nearest neighbor graph

Total edge length of graphs. $\mathcal{X} \subset \mathbb{R}^d$ finite. Given $x \in \mathcal{X}$, let x_{NN} be the nearest neighbor (NN) of x.

· Undirected nearest neighbor graph on $\mathcal{X}$: include an edge $\{x, y\}$ if $y = x_{NN}$ and/or $x = y_{NN}$.

· For $x \in \mathcal{X}$, put

$$\xi(x, \mathcal{X}) := \begin{cases} \frac{1}{2}||x - x_{NN}|| & \text{if} \quad x, x_{NN} \quad \text{are mutual NN} \\ ||x - x_{NN}|| & \text{otherwise.} \end{cases}$$

Total edge length of nearest neighbor graph

Total edge length of graphs. $\mathcal{X} \subset \mathbb{R}^d$ finite. Given $x \in \mathcal{X}$, let x_{NN} be the nearest neighbor (NN) of x.

· Undirected nearest neighbor graph on $\mathcal{X}$: include an edge $\{x, y\}$ if $y = x_{NN}$ and/or $x = y_{NN}$.

· For $x \in \mathcal{X}$, put

$$\xi(x, \mathcal{X}) := \begin{cases} \frac{1}{2}||x - x_{NN}|| & \text{if } \ x, x_{NN} \ \text{ are mutual NN} \\ ||x - x_{NN}|| & \text{otherwise.} \end{cases}$$

· Total edge length of NN graph on $\mathcal{X}$: $\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X})$.

Set-up

$W \subset \mathbb{R}^d$, $d \geq 1$, a fixed measurable set.

$g : W \to \mathbb{R}^+$, assume g is Lipschitz.

$\mathcal{P}_{sg}$, a Poisson point process on W with intensity sg.

$\mathbf{N}$: the set of simple σ-finite counting measures on $\mathbb{R}^d$.

$(\xi_s^{(i)})_{s \geq 1}$, $i \in \{1, ..., m\}$, measurable maps ('scores') from $W \times \mathbf{N} \to \mathbb{R}$.

Set-up

$W \subset \mathbb{R}^d$, $d \geq 1$, a fixed measurable set.

$g : W \to \mathbb{R}^+$, assume g is Lipschitz.

$\mathcal{P}_{sg}$, a Poisson point process on W with intensity sg.

$\mathbf{N}$: the set of simple σ-finite counting measures on $\mathbb{R}^d$.

$(\xi_s^{(i)})_{s \geq 1},\ i \in \{1, ..., m\}$, measurable maps ('scores') from $W \times \mathbf{N} \to \mathbb{R}$.

$H_s^{(i)} := H_s^{(i)}(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg}), \quad A_i \subset W.$

Goal. Find rates of multivariate normal convergence for the vector

$$\left(\frac{H_s^{(1)} - \mathbb{E}\, H_s^{(1)}}{\sqrt{\mathrm{Var} H_s^{(1)}}}, ..., \frac{H_s^{(m)} - \mathbb{E}\, H_s^{(m)}}{\sqrt{\mathrm{Var} H_s^{(m)}}} \right)$$

as intensity $s \to \infty$.

Three assumptions on scores $(\xi_s^{(i)})_{s\geq 1}$

Recall:

$$H_s^{(i)} := H_s^{(i)}(\mathcal{P}_{sg}) := \sum_{x\in\mathcal{P}_{sg}\cap A_i} \xi_s^{(i)}(x,\mathcal{P}_{sg}), \quad A_i \subset W.$$

1. Assume for all $i \in \{1, ..., m\}$ that there is a translation invariant measurable map

$$\xi^{(i)} : W \times \mathbf{N} \to \mathbb{R}$$

such that $\xi_s^{(i)}$ equals $\xi^{(i)}$ on an $s^{1/d}$-dilation of the underlying arguments:

$$\xi_s^{(i)}(x,\mathcal{M}) = \xi^{(i)}(s^{1/d}x, s^{1/d}\mathcal{M}),\ x \in W, \mathcal{M} \in \mathbf{N},\ s \geq 1.$$

Three assumptions on scores $(\xi_s^{(i)})_{s\geq 1}$

Recall:

$$H_s^{(i)} := H_s^{(i)}(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg}), \quad A_i \subset W.$$

1. Assume for all $i \in \{1, ..., m\}$ that there is a translation invariant measurable map

$$\xi^{(i)} : W \times \mathbf{N} \to \mathbb{R}$$

such that $\xi_s^{(i)}$ equals $\xi^{(i)}$ on an $s^{1/d}$-dilation of the underlying arguments:

$$\xi_s^{(i)}(x, \mathcal{M}) = \xi^{(i)}(s^{1/d}x, s^{1/d}\mathcal{M}), \; x \in W, \mathcal{M} \in \mathbf{N}, \; s \geq 1.$$

So

$$\xi_s^{(i)}(x, \mathcal{P}_{sg}) = \xi^{(i)}(s^{1/d}x, s^{1/d}\mathcal{P}_{sg}) = \xi^{(i)}(\mathbf{0}, s^{1/d}(\mathcal{P}_{sg} - x)).$$

Three assumptions on scores $(\xi_s^{(i)})_{s\geq 1}$

2. Stabilization. For $s \geq 1$ we say that $R_s : W \times \mathbf{N} \to \mathbb{R}^+$ is a radius of stabilization for $\xi_s^{(i)}$, $i \in \{1, ..., m\}$, if for all $x \in W$, $\mathcal{M} \in \mathbf{N}$, $s \geq 1$, we have

$$\xi_s^{(i)}(x, \mathcal{M} \cap B^d(x, R_s(x, \mathcal{M}))) = \xi_s^{(i)}(x, \mathcal{M}),$$

Loosely speaking, this says the score is determined by data at distance $R_s(x, \mathcal{M})$ from x, i.e, scores are determined by the 'local data'.

Three assumptions on scores $(\xi_s^{(i)})_{s\geq 1}$

2. Stabilization. For $s \geq 1$ we say that $R_s : W \times \mathbf{N} \to \mathbb{R}^+$ is a radius of stabilization for $\xi_s^{(i)}$, $i \in \{1, ..., m\}$, if for all $x \in W$, $\mathcal{M} \in \mathbf{N}$, $s \geq 1$, we have

$$\xi_s^{(i)}(x, \mathcal{M} \cap B^d(x, R_s(x, \mathcal{M}))) = \xi_s^{(i)}(x, \mathcal{M}),$$

Loosely speaking, this says the score is determined by data at distance $R_s(x, \mathcal{M})$ from x, i.e, scores are determined by the 'local data'.

Stabilization implies the following 'continuity condition' as $s \to \infty$:

$$\xi_s^{(i)}(x, \mathcal{P}_{sg}) = \xi^{(i)}(s^{1/d}x, s^{1/d}\mathcal{P}_{sg}) = \xi^{(i)}(\mathbf{0}, s^{1/d}(\mathcal{P}_{sg} - x)) \xrightarrow{\mathcal{D}} \xi^{(i)}(\mathbf{0}, \mathcal{P}_{g(x)}).$$

Three assumptions on scores $(\xi_s^{(i)})_{s\geq 1}$

Exponential Stabilization. We say that $\xi_s^{(i)}$, $i \in \{1, ..., m\}$, are exponentially stabilizing wrt $\mathcal{P}_{sg}$ if there are radii of stabilization $(R_s)_{s\geq 1}$ and constants C_{stab} and $c_{stab} \in (0, \infty)$ such that

$$\mathbb{P}(R_s(x, \mathcal{P}_{sg}) \geq r) \leq C_{stab} \exp(-c_{stab} s r^d), \ r \geq 0, x \in W, s \geq 1,$$

$$\mathbb{P}(R_s(x, \mathcal{P}_{sg(x)}) \geq r) \leq C_{stab} \exp(-c_{stab} s r^d), \ r \geq 0, x \in W, s \geq 1.$$

Three assumptions on scores $(\xi_s^{(i)})_{s\geq 1}$

Exponential Stabilization. We say that $\xi_s^{(i)}$, $i \in \{1, ..., m\}$, are exponentially stabilizing wrt $\mathcal{P}_{sg}$ if there are radii of stabilization $(R_s)_{s\geq 1}$ and constants C_{stab} and $c_{stab} \in (0, \infty)$ such that

$$\mathbb{P}(R_s(x, \mathcal{P}_{sg}) \geq r) \leq C_{stab} \exp(-c_{stab} s r^d), \ r \geq 0, x \in W, s \geq 1,$$

$$\mathbb{P}(R_s(x, \mathcal{P}_{sg(x)}) \geq r) \leq C_{stab} \exp(-c_{stab} s r^d), \ r \geq 0, x \in W, s \geq 1.$$

This says that scores $\xi_s^{(i)}$, $i \in \{1, ..., m\}$, have spatial dependencies which decay exponentially fast.

Idea: sums of exponentially stabilizing scores should behave like sums of i.i.d. random variables.

Three assumptions on scores $(\xi_s^{(i)})_{s\geq 1}$

3. p-**Moment Condition.** We say that $\xi_s^{(i)}$, $i \in \{1, ..., m\}$, satisfy a p-moment condition, $p \geq 1$, if there is $C_p \in (0, \infty)$ such that for all $i \in \{1, ..., m\}$, we have

$$\sup_{s\in[1,\infty)} \sup_{x\in W} \mathbb{E}\,|\xi_s^{(i)}(x, \mathcal{P}_{sg})|^p \leq C_p,$$

$$\sup_{s\in[1,\infty)} \sup_{x\in W} \mathbb{E}\,|\xi_s^{(i)}(x, \mathcal{P}_{sg(x)})|^p \leq C_p.$$

Four distances between m-dimensional vectors

We define four distances between distributions of two m-dimensional random vectors.

(i) $\mathcal{H}_m^{(2)}$: set of all C^2-functions $h : \mathbb{R}^m \to \mathbb{R}$ such that

$$|h(x) - h(y)| \leq ||x - y||, \; x, y \in \mathbb{R}^m,$$

$$\sup_{x \in \mathbb{R}^m} ||\text{Hess } h(x)||_{\text{op}} \leq 1.$$

Given two m-dimensional random vectors Y, Z we put

$$d_2(Y, Z) := \sup_{h \in \mathcal{H}_m^{(2)}} |\mathbb{E}\, h(Y) - \mathbb{E}\, h(Z)|$$

if $\mathbb{E}\,||Y||, \mathbb{E}\,||Z|| < \infty$.

Four distances between m-dimensional vectors

(ii) $\mathcal{H}_m^{(3)}$: set of C^3-functions $h : \mathbb{R}^m \to \mathbb{R}$ such that absolute values of the second and third partial derivatives are bounded by 1.

Given m-dimensional random vectors Y, Z we put

$$d_3(Y, Z) := \sup_{h \in \mathcal{H}_m^{(3)}} |\mathbb{E}\, h(Y) - \mathbb{E}\, h(Z)|$$

if $\mathbb{E}\,||Y||^2, \mathbb{E}\,||Z||^2 < \infty$.

Four distances between m-dimensional vectors

(ii) $\mathcal{H}_m^{(3)}$: set of C^3-functions $h : \mathbb{R}^m \to \mathbb{R}$ such that absolute values of the second and third partial derivatives are bounded by 1.

Given m-dimensional random vectors Y, Z we put

$$d_3(Y, Z) := \sup_{h \in \mathcal{H}_m^{(3)}} |\mathbb{E}\, h(Y) - \mathbb{E}\, h(Z)|$$

if $\mathbb{E}\,||Y||^2, \mathbb{E}\,||Z||^2 < \infty$.

(iii)

$$d_{convex}(Y, Z) := \sup_{h \in \mathcal{I}} |\mathbb{E}\, h(Y) - \mathbb{E}\, h(Z)|,$$

where $\mathcal{I}$ is the set of indicators of convex sets in $\mathbb{R}^m$.

Four distances between m-dimensional vectors

(iv) For all $\ell \in \mathbb{N}$ we introduce the distance

$$d_{\mathbb{H}_\ell}(Y, Z) := \sup_{h \in \mathbb{H}_\ell} |\mathbb{E}\, h(Y) - \mathbb{E}\, h(Z)|,$$

where $\mathbb{H}_\ell$ is the set of indicator functions of intersections of ℓ closed half-spaces in $\mathbb{R}^m$.

Four distances between m-dimensional vectors

(iv) For all $\ell \in \mathbb{N}$ we introduce the distance

$$d_{\mathbb{H}_\ell}(Y, Z) := \sup_{h \in \mathbb{H}_\ell} |\mathbb{E}\, h(Y) - \mathbb{E}\, h(Z)|,$$

where $\mathbb{H}_\ell$ is the set of indicator functions of intersections of ℓ closed half-spaces in $\mathbb{R}^m$.

For $m = \ell = 1$, $d_{\mathbb{H}_1}$ is the univariate Kolmogorov distance d_K. Thus we may consider $d_{\mathbb{H}_\ell}$ to be a multi-dimensional generalization of d_K.

Results: Covariance asymptotics, univariate CLT

Recall for $A_i \subset W, i \in \{1, ..., m\}$,

$$H_s^{(i)} := H_s^{(i)}(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg}), \ \ s \geq 1.$$

Centered version: $\bar{H}_s^{(i)} := H_s^{(i)} - \mathbb{E}\, H_s^{(i)}$.

Results: Covariance asymptotics, univariate CLT

Recall for $A_i \subset W, i \in \{1, ..., m\}$,

$$H_s^{(i)} := H_s^{(i)}(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg}), \;\; s \geq 1.$$

Centered version: $\bar{H}_s^{(i)} := H_s^{(i)} - \mathbb{E}\, H_s^{(i)}$.

Theorem. Assume for all $i \in \{1, ..., m\}$ that the scores $(\xi_s^{(i)})_{s \geq 1}$
(i) are exponentially stabilizing, and
(ii) satisfy the p-moment condition for some $p > 4$.

Then for all $i, j \in \{1, ..., m\}$ we have as $s \to \infty$

$$\frac{\mathrm{Cov}(H_s^{(i)}, H_s^{(j)})}{s} \to \sigma_{ij}, \qquad \frac{\bar{H}_s^{(i)}}{\sqrt{s}} \xrightarrow{\mathcal{D}} N(0, \sigma_{ii}).$$

Def. $N_\Sigma :=$ multivariate normal with covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq m}$.

Results: Main Theorem

Assume $\Sigma = (\sigma_{ij})_{1\le i,j\le m}$ is positive definite.

Thm (Schulte + Y.) Assume $\forall i \in \{1, ..., m\}$ that the scores $(\xi_s^{(i)})_{s\ge 1}$
(i) are exponentially stabilizing, and
(ii) satisfy the p-moment condition, for some $p > 6$.

Then there is a constant $C \in (0, \infty)$ such that

$$\tilde{d}(s^{-1/2}(\bar{H}_s^{(1)}, ..., \bar{H}_s^{(m)}), N_\Sigma) \le Cs^{-1/d},\ s \ge 1,\ (*)$$

for $\tilde{d} \in \{d_2, d_3, d_{\mathbb{H}_\ell}\}$. If

$$\max_{i\le m} |H_s^{(i)}(\mathcal{P}_{sg} \cup \{y\}) - H_s^{(i)}(\mathcal{P}_{sg})| \le \tilde{C},\ y \in \mathbb{R}^d,$$

then (*) also holds for $\tilde{d} = d_{convex}$.

Results: Variance asymptotics

Mecke formula relates the expectation of a function summed over a point process to an integral involving the mean measure of the point process:

$$\begin{aligned}
&\mathrm{Var} \sum_{x \in \mathcal{P}_{sg} \cap W} \xi_s(x, \mathcal{P}_{sg}) \\
&= s \int_W \mathbb{E}\,(\xi_s(x, \mathcal{P}_{sg}))^2 g(x) dx \\
&\quad + s^2 \int_W \int_W [\mathbb{E}\,\xi_s(x, \mathcal{P}_{sg} \cup \{y\}) \xi_s(y, \mathcal{P}_{sg} \cup \{x\}) \\
&\qquad - \mathbb{E}\,\xi_s(x, \mathcal{P}_{sg}) \mathbb{E}\,\xi_s(y, \mathcal{P}_{sg})] g(x) g(y) dy dx.
\end{aligned}$$

Divide by s.

Results: Variance asymptotics

$$s^{-1}\mathrm{Var} \sum_{x \in \mathcal{P}_{sg} \cap W} \xi_s(x, \mathcal{P}_{sg}) = \int_W \mathbb{E}\,(\xi_s(x, \mathcal{P}_{sg}))^2 g(x) dx$$

$$+ s \int_W \int_W [\mathbb{E}\,\xi_s(x, \mathcal{P}_{sg} \cup \{y\}) \xi_s(y, \mathcal{P}_{sg} \cup \{x\})$$

$$- \mathbb{E}\,\xi_s(x, \mathcal{P}_{sg}) \mathbb{E}\,\xi_s(y, \mathcal{P}_{sg})] g(y) g(x) dy dx.$$

· Recall: $\xi_s(x, \mathcal{P}_{sg} \cup \{y\}) = \xi(s^{1/d} x, s^{1/d}(\mathcal{P}_{sg} \cup \{y\}))$.

Results: Variance asymptotics

$$s^{-1}\mathrm{Var}\sum_{x\in\mathcal{P}_{sg}\cap W}\xi_s(x,\mathcal{P}_{sg}) = \int_W \mathbb{E}\,(\xi_s(x,\mathcal{P}_{sg}))^2 g(x)dx$$
$$+\, s\int_W\int_W [\mathbb{E}\,\xi_s(x,\mathcal{P}_{sg}\cup\{y\})\xi_s(y,\mathcal{P}_{sg}\cup\{x\})$$
$$-\,\mathbb{E}\,\xi_s(x,\mathcal{P}_{sg})\mathbb{E}\,\xi_s(y,\mathcal{P}_{sg})]g(y)g(x)dydx.$$

- Recall: $\xi_s(x,\mathcal{P}_{sg}\cup\{y\}) = \xi(s^{1/d}x, s^{1/d}(\mathcal{P}_{sg}\cup\{y\}))$.
- Recall: as $s\to\infty$

$$\xi_s(x,\mathcal{P}_{sg}) = \xi(s^{1/d}x, s^{1/d}\mathcal{P}_{sg}) = \xi(\mathbf{0}, s^{1/d}(\mathcal{P}_{sg}-x)) \xrightarrow{\mathcal{D}} \xi(\mathbf{0},\mathcal{P}_{g(x)}).$$

Results: Variance asymptotics

$$s^{-1}\mathrm{Var}\sum_{x\in\mathcal{P}_{sg}\cap W}\xi_s(x,\mathcal{P}_{sg}) = \int_W \mathbb{E}\,(\xi_s(x,\mathcal{P}_{sg}))^2 g(x)dx$$
$$+ s\int_W\int_W [\mathbb{E}\,\xi_s(x,\mathcal{P}_{sg}\cup\{y\})\xi_s(y,\mathcal{P}_{sg}\cup\{x\})$$
$$- \mathbb{E}\,\xi_s(x,\mathcal{P}_{sg})\mathbb{E}\,\xi_s(y,\mathcal{P}_{sg})]g(y)g(x)dydx.$$

· Recall: $\xi_s(x,\mathcal{P}_{sg}\cup\{y\}) = \xi(s^{1/d}x, s^{1/d}(\mathcal{P}_{sg}\cup\{y\}))$.
· Recall: as $s\to\infty$

$$\xi_s(x,\mathcal{P}_{sg}) = \xi(s^{1/d}x, s^{1/d}\mathcal{P}_{sg}) = \xi(\mathbf{0}, s^{1/d}(\mathcal{P}_{sg}-x)) \xrightarrow{\mathcal{D}} \xi(\mathbf{0},\mathcal{P}_{g(x)}).$$

· Translate $s^{1/d}x$ to the origin, translate $s^{1/d}(\mathcal{P}_{sg}\cup\{y\})$ by x to get $s^{1/d}((\mathcal{P}_{sg}-\{x\})\cup(\{y\}-\{x\}))$, put $y = x + s^{-1/d}z$, $dz = sdy$.

· For each $x\in W$, inside integral converges to $\int_{\mathbb{R}^d}[\mathbb{E}\,\xi(\mathbf{0},\mathcal{P}_{g(x)}\cup\{z\})\xi(z,\mathcal{P}_{g(x)}\cup\{\mathbf{0}\}) - \mathbb{E}\,\xi(\mathbf{0},\mathcal{P}_{g(x)})\mathbb{E}\,\xi(z,\mathcal{P}_{g(x)})]g(x)dz$.

Results: Variance asymptotics

Conclusion: as $s \to \infty$ we get

$$
\begin{aligned}
s^{-1}\mathrm{Var} \sum_{x \in \mathcal{P}_{sg} \cap W} \xi_s(x, \mathcal{P}_{sg}) &= \int_W \mathbb{E}\,(\xi_s(x, \mathcal{P}_{sg}))^2 g(x)dx \\
&\quad + s \int_W \int_W [\mathbb{E}\,\xi_s(x, \mathcal{P}_{sg} \cup \{y\})\xi_s(y, \mathcal{P}_{sg} \cup \{x\}) \\
&\qquad - \mathbb{E}\,\xi_s(x, \mathcal{P}_{sg})\mathbb{E}\,\xi_s(y, \mathcal{P}_{sg})]g(y)g(x)dydx \\
&\to \sigma^2,
\end{aligned}
$$

where σ^2 is sum of $\int_W \mathbb{E}\,(\xi(x, \mathcal{P}_{g(x)}))^2 g(x)dx$ and

$$
\begin{aligned}
\int_W \int_{\mathbb{R}^d} [\mathbb{E}\,\xi(\mathbf{0}, \mathcal{P}_{g(x)} \cup \{z\})\xi(z, \mathcal{P}_{g(x)} \cup \{\mathbf{0}\}) \\
- \mathbb{E}\,\xi(\mathbf{0}, \mathcal{P}_{g(x)})\mathbb{E}\,\xi(z, \mathcal{P}_{g(x)})]g^2(x)dzdx.
\end{aligned}
$$

Three remarks

(i) A main ingredient to the proof:

$$\left|\sigma_{ij} - \frac{\mathsf{Cov}(H_s^{(i)}, H_s^{(j)})}{s}\right| \le C s^{-1/d}, \quad s \ge 1.$$

Three remarks

(i) A main ingredient to the proof:

$$\left| \sigma_{ij} - \frac{\mathsf{Cov}(H_s^{(i)}, H_s^{(j)})}{s} \right| \leq C s^{-1/d}, \quad s \geq 1.$$

(ii) If we replace N_Σ by $N_{\Sigma(s)}$, where $\Sigma(s)$ is the covariance matrix of

$$s^{-1/2}(\bar{H}_s^{(1)}, ..., \bar{H}_s^{(m)}),$$

then the rates of multivariate normal convergence improve to

$$\tilde{d}(s^{-1/2}(\bar{H}_s^{(1)}, ..., \bar{H}_s^{(m)}), N_{\Sigma(s)}) \leq C s^{-1/2}, \;\; s \geq 1,$$

for $\tilde{d} \in \{d_2, d_3, d_{\mathbb{H}_\ell}, d_{convex}\}$. Rates are not improvable in general.

Three remarks, cont'd

(iii) Comparison with literature

(a) Rinott and Rotar (1996): obtain rates of multivariate convergence for sums of locally dependent r.v.; rates involve extra logarithmic factors.

(b) Penrose and Wade (2008): consider the special case $\xi_s^{(1)} = ... = \xi_s^{(m)}$ and all sets $A_i, i \in \{1, ..., m\}$, are disjoint. They obtain rate of normal convergence $O(s^{-1/(2d+\epsilon)}),\ \epsilon > 0$.

(c) Peccati and Zheng (2010): bounds involve O-U operator

Three applications

(i) **Multivariate statistics of kNN graph.** Let $k \in \mathbb{N}$ and $\mathcal{X} \subset \mathbb{R}^d$ a finite point set. For $x, y \in \mathcal{X}$, we put an undirected edge between x and y if x is one of the k nearest neighbors of y and/or y is a k nearest neighbor of x. Put

$$H^{(k)}(\mathcal{X}) := \text{sum of lengths of edges in kNN on } \mathcal{X}.$$

(i) **Multivariate statistics of kNN graph.** Let $k \in \mathbb{N}$ and $\mathcal{X} \subset \mathbb{R}^d$ a finite point set. For $x, y \in \mathcal{X}$, we put an undirected edge between x and y if x is one of the k nearest neighbors of y and/or y is a k nearest neighbor of x. Put

$$H^{(k)}(\mathcal{X}) := \text{sum of lengths of edges in kNN on } \mathcal{X}.$$

Theorem. Let $\mathcal{P}_{sg}$ be a Poisson point process on $[0,1]^d$ with intensity sg, $g \in \text{Lip}([0,1]^d)$, g bounded away from 0 and ∞. Then for all $k_i \in \mathbb{N}$, $1 \leq i \leq m$, we have

$$\tilde{d}(s^{-1/2}(\bar{H}^{(k_1)}(s^{1/d}\mathcal{P}_{sg}), ..., \bar{H}^{(k_m)}(s^{1/d}\mathcal{P}_{sg})), N_\Sigma) \leq Cs^{-1/d}, \;\; s \geq 1,$$

for $\tilde{d} \in \{d_2, d_3, d_{\mathbb{H}_\ell}\}$.

Three applications

(ii) **Multivariate statistics of random geometric graph.** Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite point set. Put $H_s^{(i)}(\mathcal{X})$ to be the number of components of random geometric graph $G(s^{1/d}\mathcal{X}, s^{1/d}r)$ of size i.

Three applications

(ii) **Multivariate statistics of random geometric graph.** Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite point set. Put $H_s^{(i)}(\mathcal{X})$ to be the number of components of random geometric graph $G(s^{1/d}\mathcal{X}, s^{1/d}r)$ of size i.

Theorem. Let $\mathcal{P}_{sg}$ be a Poisson point process on $[0,1]^d$ with intensity sg, $g \in \text{Lip}([0,1]^d)$, g bounded away from 0 and ∞. When $r = \rho s^{-1/d}$ we have for all $i_j \in \mathbb{N}$, $1 \leq j \leq m$

$$\tilde{d}(s^{-1/2}(\bar{H}_s^{(i_1)}(\mathcal{P}_{sg}), ..., \bar{H}_s^{(i_m)}(\mathcal{P}_{sg})), N_\Sigma) \leq Cs^{-1/d}, \ s \geq 1,$$

for $\tilde{d} \in \{d_2, d_3, d_{\mathbb{H}_\ell}, d_{convex}\}$.

Three Applications

(iii) **Multivariate statistics for equality of distributions.** Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite point set. Consider the undirected nearest neighbors graph $NNG(\mathcal{X})$ on $\mathcal{X}$. Color the nodes of $\mathcal{X}$ with color i with probability $\pi_i, 1 \leq i \leq m$.

Let $H^{(i)}(\mathcal{X})$ be the number of edges in $NNG(\mathcal{X})$ which join nodes of color $i, 1 \leq i \leq m$.

Three Applications

(iii) **Multivariate statistics for equality of distributions.** Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite point set. Consider the undirected nearest neighbors graph $NNG(\mathcal{X})$ on $\mathcal{X}$. Color the nodes of $\mathcal{X}$ with color i with probability $\pi_i, 1 \le i \le m$.

Let $H^{(i)}(\mathcal{X})$ be the number of edges in $NNG(\mathcal{X})$ which join nodes of color $i, 1 \le i \le m$.

Theorem. Let $\mathcal{P}_{sg}$ be a Poisson point process on $[0,1]^d$ with intensity sg, $g \in \text{Lip}([0,1]^d)$, g bounded away from 0 and ∞. We have

$$\tilde{d}(s^{-1/2}(\bar{H}^{(1)}(s^{1/d}\mathcal{P}_{sg}), ..., \bar{H}^{(m)}(s^{1/d}\mathcal{P}_{sg})), N_\Sigma) \le Cs^{-1/d}, \ s \ge 1,$$

for $\tilde{d} \in \{d_2, d_3, d_{\mathbb{H}_\ell}, d_{convex}\}$.

This vector features in tests for equality of distributions.

Proof idea

Theorem (Schulte + Y.) Let $F = (F_1, \dots, F_m)$, $m \in \mathbb{N}$, be a vector of Poisson functionals $F_1, \dots, F_m$ with $\mathbb{E}\, F_i = 0$, $i \in \{1, \dots, m\}$, and assume there is $p \in (4, \infty)$ such that

$$\mathbb{E}\, |D_x F_i|^p < \infty, \quad \lambda\text{-a.e. } x \in \mathbb{X},$$

and

$$\mathbb{E}\, |D^2_{x_1,x_2} F_i|^p < \infty, \quad \lambda\text{-a.e. } x_1, x_2 \in \mathbb{X}$$

for all $i \in \{1, \dots, m\}$.

Proof idea

Theorem (Schulte + Y.) Let $F = (F_1, \dots, F_m)$, $m \in \mathbb{N}$, be a vector of Poisson functionals $F_1, \dots, F_m$ with $\mathbb{E}\, F_i = 0$, $i \in \{1, \dots, m\}$, and assume there is $p \in (4, \infty)$ such that

$$\mathbb{E}\, |D_x F_i|^p < \infty, \quad \lambda\text{-a.e. } x \in \mathbb{X},$$

and

$$\mathbb{E}\, |D^2_{x_1,x_2} F_i|^p < \infty, \quad \lambda\text{-a.e. } x_1, x_2 \in \mathbb{X}$$

for all $i \in \{1, \dots, m\}$.
(a) For positive definite $\Sigma = (\sigma_{ij})_{i,j \in \{1,\dots,m\}} \in \mathbb{R}^{m \times m}$,

$$d_3(F, N_\Sigma) \leq \frac{m}{2} \sum_{i,j=1}^{m} |\sigma_{ij} - \operatorname{Cov}(F_i, F_j)| + \frac{3m^{3/2}}{2} \Gamma_1(p) + \frac{m^2}{4} \Gamma_2(p).$$

Proof idea

(a) For positive definite $\Sigma = (\sigma_{ij})_{i,j\in\{1,\dots,m\}} \in \mathbb{R}^{m\times m}$,

$$d_3(F, N_\Sigma) \leq \frac{m}{2} \sum_{i,j=1}^{m} |\sigma_{ij} - \mathrm{Cov}(F_i, F_j)| + \frac{3m^{3/2}}{2}\Gamma_1(p) + \frac{m^2}{4}\Gamma_2(p).$$

$\Gamma_1(p)$ and $\Gamma_2(p)$ are integrals of moments of first and second order difference operators.

Proof idea

Theorem (Schulte + Y.) (b) For positive definite $\Sigma \in \mathbb{R}^{m\times m}$,

$$d_2(F, N_\Sigma) \le \|\Sigma^{-1}\|_{op}\|\Sigma\|_{op}^{1/2}\sum_{i,j=1}^{m}|\sigma_{ij} - \mathrm{Cov}(F_i, F_j)| + 3\|\Sigma^{-1}\|_{op}\|\Sigma\|_{op}\sqrt{m}\Gamma_1(p) + \frac{\sqrt{2\pi}}{8}\|\Sigma^{-1}\|_{op}^{3/2}\|\Sigma\|_{op} m^2\Gamma_2(p).$$

Proof idea

Theorem (Schulte + Y.) (c) Let $F = (F_1, \dots, F_m)$, $m \in \mathbb{N}$, be a vector of Poisson functionals $F_1, \dots, F_m$ with $\mathbb{E}\, F_i = 0$, $i \in \{1, \dots, m\}$, and assume there is $p \in (6, \infty)$ such that

$$\mathbb{E}\,|D_x F_i|^p < \infty, \quad \lambda\text{-a.e. } x \in \mathbb{X},$$

and

$$\mathbb{E}\,|D^2_{x_1,x_2} F_i|^p < \infty, \quad \lambda\text{-a.e. } x_1, x_2 \in \mathbb{X}$$

for all $i \in \{1, \dots, m\}$. If $\Sigma \in \mathbb{R}^{m \times m}$ is positive definite then for any $\ell \in \mathbb{N}$ there exists a constant $C^{(1)} \in (0, \infty)$ also depending on m, l and Σ such that

$$d_{\mathbb{H}_\ell}(F, N_\Sigma) \leq C^{(1)} \max \left\{ \sum_{i,j \in \{1,\dots,m\}} |\sigma_{ij} - \mathrm{Cov}(F_i, F_j)|, \Gamma_1(p), \Gamma_3(p), \sqrt{\Gamma_4(p)} \right\}.$$

Proof idea

Theorem (Schulte + Y.) (d) Assume the conditions of part (c). Let $\Sigma \in \mathbb{R}^{m\times m}$ be positive definite and assume that there is a constant $\varrho \in (0,\infty)$ such that, for $i \in \{1,\dots,m\}$ and λ-a.e. $x \in \mathbb{X}$, $|D_x F_i| \leq \varrho$ $\mathbb{P}$-a.s. Then there exists a constant $C^{(2)} \in (0,\infty)$ depending on m and Σ such that

$$
\begin{aligned}
& d_{convex}(F, N_\Sigma) \\
& \leq C^{(2)} \max\Big\{ \sum_{i,j\in\{1,\dots,m\}} |\sigma_{ij} - \mathrm{Cov}(F_i,F_j)|, \Gamma_1(p), \Gamma_3(p), \varrho^3\lambda(A), \\
& \sqrt{\varrho^4\lambda(A)}, \lambda(A)^{-1} \int_{\mathbb{X}\setminus A} \mathbb{P}(D_x F \neq 0)\,\lambda(dx) \Big\}
\end{aligned}
$$

for any $A \in \mathcal{F}$ with $0 < \lambda(A) < \infty$.

Extensions of Main Result

$W \subset \mathbb{R}^d$, $d \geq 1$, a fixed measurable set.

$\mathcal{P}_{sg}$, a Poisson point process on W with intensity sg, $g : W \to \mathbb{R}^+$ is Lipschitz.

$H_s^{(i)} := H_s^{(i)}(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg}), \quad A_i \subset W.$

We have found rates of multivariate normal convergence for the vector

$$\left(\frac{H_s^{(1)} - \mathbb{E}\, H_s^{(1)}}{\sqrt{\mathrm{Var} H_s^{(1)}}}, ..., \frac{H_s^{(m)} - \mathbb{E}\, H_s^{(m)}}{\sqrt{\mathrm{Var} H_s^{(m)}}} \right), \quad \text{as intensity } s \to \infty.$$

Extensions:

(i) points in $\mathcal{P}_{sg}$ may carry independent marks

(ii) rates of multivariate normal convergence for random measures

$$\mu_s^{(i)}(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg}) \delta_x, \quad A_i \subset W.$$

THANK YOU