Rates of Multivariate Normal Approximation for Statistics in Geometric Probability

Joseph Yukich (joint with Matthias Schulte, Bern)

Lehigh University

Charles Stein’s Influence on Probability
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 $\xi$, 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 $\xi(x, \mathcal{X}), x \in \mathcal{X}$. 

·
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 $\xi$, 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 $W$ of $\mathbb{R}^d$, then much is known concerning central limit theorems for

$$\sum_{x \in \mathcal{P}_s \cap W} \xi(x, \mathcal{P}_s)$$

as the intensity $s$ tends to $\infty$. 
Set-up

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

\[ g : W \to \mathbb{R}^+; \ g \text{ is Lipschitz.} \]

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

\[ \mathbb{N}: \text{ the set of simple } \sigma\text{-finite counting measures on } \mathbb{R}^d. \]

\[ (\xi_{s}^{(1)})_{s \geq 1}, \ldots, (\xi_{s}^{(m)})_{s \geq 1}, \text{ measurable maps (‘scores’) from } W \times \mathbb{N} \to \mathbb{R}. \]

\[ H_{s}^{(i)} := H_{s}^{(i)}(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_{i}} \xi_{s}^{(i)}(x, \mathcal{P}_{sg}), \ A_{i} \subset W. \]
Set-up

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

$g : W \rightarrow \mathbb{R}^+; \ g \text{ is Lipschitz.}$

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

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

$(\xi_s^{(1)})_{s \geq 1}, \ldots, (\xi_s^{(m)})_{s \geq 1}$, measurable maps (‘scores’) from $W \times \mathbb{N} \rightarrow \mathbb{R}$.

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

**Goal.** Find rates of multivariate normal convergence for the $m$-vector

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

as intensity $s \rightarrow \infty$. 
Assumptions on scores \((\xi_s^{(i)})_{s\geq 1}\)

\begin{align*}
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.
\end{align*}

1. Assume for all \(i \in \{1, \ldots, m\}\) that \(\xi_s^{(i)}\) is the score \(\xi^{(i)}\) at \(x\) evaluated on an \(s\)-dilation of the underlying point set:

\[
\xi_s^{(i)}(x, \mathcal{M}) = \xi^{(i)}(x, x + s^{1/d}(\mathcal{M} - x)), \quad x \in W, \mathcal{M} \in \mathbb{N}, \quad s \geq 1.
\]
Assumptions on scores \((\xi_s^{(i)})_{s \geq 1}\)

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

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

Loosely speaking, this says the scores \(\xi_s^{(i)}, i \in \{1, \ldots, m\}\) are determined by data at distance \(R_s(x, \mathcal{M})\) from \(x\).
Assumptions on scores \((\xi_s^{(i)})_{s \geq 1}\)

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

\[
P(R_s(x, P_{sg}) \geq r) \leq C_{stab} \exp(-c_{stab}sr^d), \quad r \geq 0, x \in W, s \geq 1,
\]
Exponential Stabilization. We say that \((\xi_s^{(1)})_{s \geq 1}, \ldots, (\xi_s^{(m)})_{s \geq 1}\), are exponentially stabilizing wrt \(P_{sg}\) if there are radii of stabilization \((R_s)_{s \geq 1}\) and constants \(C_{stab}\) and \(c_{stab} \in (0, \infty)\) such that

\[
P(R_s(x, P_{sg}) \geq r) \leq C_{stab} \exp(-c_{stab}sr^d), \quad r \geq 0, x \in W, s \geq 1,
\]

This says that scores \((\xi_s^{(1)})_{s \geq 1}, \ldots, (\xi_s^{(m)})_{s \geq 1}\) have spatial dependencies which decay exponentially fast.

Idea: Sums of exponentially stabilizing scores should behave like sums of i.i.d. random variables.
Assumptions on scores \((\xi_s^{(i)})_{s \geq 1}\)

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

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

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

Put \( H_s := H_s(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg}} \xi_s(x, \mathcal{P}_{sg}) \).
\( \mathcal{P}_{sg} \), a Poisson point process on \( W \) with intensity \( sg \).

Put \( H_s := H_s(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg}} \xi_s(x, \mathcal{P}_{sg}) \).

**Theorem (Lachiéze-Rey, Schulte + Y. (2019))** Assume \((\xi_s), s \geq 1\), are exponentially stabilizing and satisfy the \( p \)-moment condition for some \( p \in (4, \infty) \). If \( \text{Var} \, H_s = \Omega(s) \), then

\[
d_K \left( \frac{H_s - \mathbb{E} \, H_s}{\sqrt{\text{Var} \, H_s}}, N(0,1) \right) \leq \frac{c}{\sqrt{s}}, \quad s \geq 1.
\]
Univariate CLT

\( P_{sg}, \) a Poisson point process on \( W \) with intensity \( sg. \)

Put \( H_s := H_s(P_{sg}) := \sum_{x \in P_{sg}} \xi_s(x, P_{sg}). \)

**Theorem (Lachiéze-Rey, Schulte + Y. (2019))** Assume \( (\xi_s), s \geq 1, \) are exponentially stabilizing and satisfy the \( p \)-moment condition for some \( p \in (4, \infty). \) If \( \text{Var} H_s = \Omega(s), \) then

\[
d_K\left( \frac{H_s - \mathbb{E} H_s}{\sqrt{\text{Var} H_s}}, N(0, 1) \right) \leq \frac{c}{\sqrt{s}}, \quad s \geq 1.
\]

Question: what are good bounds for

\[
d_K\left( \frac{H_s - \mathbb{E} H_s}{\sqrt{s}}, N(0, 1) \right)\?
We define four distances between distributions of two $m$-dimensional random vectors.

(i) $\mathcal{H}_m^{(2)}$: 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 $m$-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 $\mathbb{E}||Y||, \mathbb{E}||Z|| < \infty$. 
Distances between \(m\)-dimensional vectors

(ii) \(\mathcal{H}_m^{(3)}\): all \(C^3\)-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)

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

where \(\mathcal{I}\) is the set of indicators of convex sets in \(\mathbb{R}^m\).
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 \( \ell \) closed half-spaces in \( \mathbb{R}^m \).

For \( m = \ell = 1 \), \( d_{\mathbb{H}_1} \) is the univariate Kolmogorov distance \( d_K \). We may consider \( d_{\mathbb{H}_\ell} \) to be a multi-dimensional generalization of \( d_K \).
Multivariate CLT without rates

Recall for \( A_i \subset W, i \in \{1, \ldots, m\} \),

\[
H_s^{(i)} := H_s^{(i)}(P_{sg}) := \sum_{x \in P_{sg} \cap A_i} \xi_s^{(i)}(x, P_{sg}), \ s \geq 1.
\]

Centered version: \( \bar{H}_s^{(i)} := H_s^{(i)} - \mathbb{E} H_s^{(i)} \)
Recall for $A_i \subset W, i \in \{1, ..., m\}$,

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

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

**Thm (Penrose; Baryshnikov + Y.)** Assume $(\xi_s^{(1)})_{s \geq 1}, ..., (\xi_s^{(m)})_{s \geq 1}$, are

(i) exponentially stabilizing, and

(ii) satisfy the $p$-moment condition for some $p > 2$.

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

$$\frac{\text{Cov}(H_s^{(i)}, H_s^{(j)})}{s} \to \sigma_{ij}, \quad s^{-1/2}(\bar{H}_s^{(1)}, ..., \bar{H}_s^{(m)}) \overset{D}{\to} N_\Sigma.$$  

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

Joseph Yukich (joint with Matthias Schulte, [Rates of Multivariate Normal Approximation](#))
Main Theorem: Rates of Multivariate Normal Convergence

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

**Theorem (Schulte + Y.)** Assume $(\xi_s^{(1)})_{s \geq 1}, \ldots, (\xi_s^{(m)})_{s \geq 1}$ are
(i) exponentially stabilizing, and
(ii) satisfy the $p$-moment condition for some $p > 6$. 

---

Joseph Yukich (joint with Matthias Schulte, Bern) (Lehigh University)
Main Theorem: Rates of Multivariate Normal Convergence

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

**Theorem (Schulte + Y.)** Assume $(\xi_s^{(1)})_{s \geq 1}, ..., (\xi_s^{(m)})_{s \geq 1}$ are
(i) exponentially stabilizing, and
(ii) satisfy the $p$-moment condition for some $p > 6$.

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

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

for $\tilde{d} \in \{d_2, d_3, d_{\text{convex}}\}$. 

Joseph Yukich (joint with Matthias Schulte, Bern) (Lehigh University)
Main Theorem: Rates of Multivariate Normal Convergence

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

**Theorem (Schulte + Y.)** Assume $(\xi_s^{(1)})_{s \geq 1}, \ldots, (\xi_s^{(m)})_{s \geq 1}$ are
(i) exponentially stabilizing, and
(ii) satisfy the $p$-moment condition for some $p > 6$.
Then there is a constant $C \in (0, \infty)$ such that

$$\tilde{d}(s^{-1/2}(\bar{H}_s^{(1)}, \ldots, \bar{H}_s^{(m)}), N_\Sigma) \leq Cs^{-1/d}, \ s \geq 1, \ (*)$$

for $\tilde{d} \in \{d_2, d_3, d_{\mathbb{H}_e}\}$. If the add-one cost satisfies

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

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

- Rates are the same for smooth and non-smooth distances
Two remarks

\[ H_s^{(i)} := H_s^{(i)}(P_{sg}) := \sum_{x \in P_{sg} \cap A_i} \xi_s^{(i)}(x, P_{sg}), \quad s \geq 1. \]

(i) A main ingredient to the proof: For all \( i, j \in \{1, \ldots, m\} \)

\[
\left| \sigma_{ij} - \frac{\text{Cov}(H_s^{(i)}, H_s^{(j)})}{s} \right| \leq Cs^{-1/d}, \quad s \geq 1.
\]
Two remarks

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

(i) A main ingredient to the proof: For all \( i, j \in \{1, \ldots, m\} \)

\[
\left| \sigma_{ij} - \frac{\text{Cov}(H_s^{(i)}, H_s^{(j)})}{s} \right| \leq Cs^{-1/d}, \quad s \geq 1.
\]

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

\[ s^{-1/2}(\bar{H}_s^{(1)}, \ldots, \bar{H}_s^{(m)}), \]

then the rates of multivariate normal convergence improve to

\[
\tilde{d}(s^{-1/2}(\bar{H}_s^{(1)}, \ldots, \bar{H}_s^{(m)}), N_{\Sigma(s)}) \leq Cs^{-1/2}, \quad s \geq 1,
\]

for \( \tilde{d} \in \{d_2, d_3, d_{\mathbb{H}_\ell}, d_{\text{convex}}\} \). Rates are not improvable in general.
· Barbour (1990), Goldstein and Rinott (1996)


· multivariate clts for vectors with certain dependency structures: Raič (2004), Goldstein and Rinott (2005), Chen, Goldstein, and Shao (2011)

· Peccati and Zheng (2010)
· Barbour (1990), Goldstein and Rinott (1996)
· multivariate clts for vectors with certain dependency structures: Raič (2004), Goldstein and Rinott (2005), Chen, Goldstein, and Shao (2011)
· Chatterjee and Meckes (2008), Meckes (2009), Barbour and Xia (2019).
Barbour (1990), Goldstein and Rinott (1996)


multivariate clts for vectors with certain dependency structures: Raič (2004), Goldstein and Rinott (2005), Chen, Goldstein, and Shao (2011)

Chatterjee and Meckes (2008), Meckes (2009), Barbour and Xia (2019).

Penrose and Wade (2008): consider the special case $\xi_{ss}^{(1)} = ... = \xi_{ss}^{(m)}$ and all sets $A_i, i \in \{1, ..., m\}$, are disjoint. They establish rate of normal convergence $O(s^{-1/(2d+\epsilon)})$, $\epsilon > 0$, wrt Kolmogorov distance in $\mathbb{R}^d$. 
· Barbour (1990), Goldstein and Rinott (1996)


· multivariate clts for vectors with certain dependency structures: Raič (2004), Goldstein and Rinott (2005), Chen, Goldstein, and Shao (2011)

· Chatterjee and Meckes (2008), Meckes (2009), Barbour and Xia (2019).

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

· Peccati and Zheng (2010)
Proof idea for $d_2$, $d_3$; Stein’s method, Malliavin calculus

- Peccati and Zheng (2010): $\eta$ a Poisson process over $(X, F)$ with intensity measure $\lambda$; $F = (F_1, \ldots, F_m)$, $m \in \mathbb{N}$, a vector of (Poisson) functionals of $\eta$, $\mathbb{E} F_i = 0$, $\Sigma = (\sigma_{i,j})_{i,j \in \{1, \ldots, m\}}$ positive definite. Put

$$
\beta_1 := \sqrt{\sum_{i,j=1}^{m} \mathbb{E} (\sigma_{i,j} - \int_X D_x F_i (-D_x L^{-1} F_j) \lambda(dx))^2}
$$

$$
\beta_2 := \int_X \mathbb{E} \left( \sum_{i=1}^{m} |D_x F_i| \right)^2 \sum_{j=1}^{m} |D_x L^{-1} F_j| \lambda(dx).
$$

Then $d_2(F, N_\Sigma)$ and $d_3(F, N_\Sigma)$ bounded by $C(m, \Sigma) \cdot (\beta_1 + \beta_2)$. 
Proof idea for $d_2, d_3$; Stein’s method, Malliavin calculus

- Peccati and Zheng (2010): $\eta$ a Poisson process over $(\mathbb{X}, \mathcal{F})$ with intensity measure $\lambda$; $F = (F_1, \ldots, F_m)$, $m \in \mathbb{N}$, a vector of (Poisson) functionals of $\eta$, $\mathbb{E} F_i = 0$, $\Sigma = (\sigma_{i,j})_{i,j \in \{1, \ldots, m\}}$ positive definite. Put

$$\beta_1 := \sqrt{\sum_{i,j=1}^{m} \mathbb{E} \left( \sigma_{i,j} - \int_{\mathbb{X}} D_x F_i (-D_x L^{-1} F_j) \lambda(dx) \right)^2}$$

$$\beta_2 := \int_{\mathbb{X}} \mathbb{E} \left( \sum_{i=1}^{m} |D_x F_i| \right)^2 \sum_{j=1}^{m} |D_x L^{-1} F_j| \lambda(dx).$$

Then $d_2(F, N_{\Sigma})$ and $d_3(F, N_{\Sigma})$ bounded by $C(m, \Sigma) \cdot (\beta_1 + \beta_2)$.

- Last, Peccati, Schulte: $\mathbb{E} |D_x L^{-1} F|^p$ and $\mathbb{E} |D_{x,y}^2 L^{-1} F|^p$ bounded by moments of difference operators.
Proof idea for $d_2$, $d_3$; Stein’s method, Malliavin calculus

- Peccati and Zheng (2010): $\eta$ a Poisson process over $(X, \mathcal{F})$ with intensity measure $\lambda$; $F = (F_1, \ldots, F_m)$, $m \in \mathbb{N}$, a vector of (Poisson) functionals of $\eta$, $\mathbb{E} F_i = 0$, $\Sigma = (\sigma_{i,j})_{i,j \in \{1, \ldots, m\}}$ positive definite. Put

$$\beta_1 := \sqrt{\sum_{i,j=1}^{m} \mathbb{E} \left( \sigma_{i,j} - \int_X D_x F_i (-D_x L^{-1} F_j) \lambda(dx) \right)^2}$$

$$\beta_2 := \int_X \mathbb{E} \left( \sum_{i=1}^{m} |D_x F_i| \right)^2 \sum_{j=1}^{m} |D_x L^{-1} F_j| \lambda(dx).$$

Then $d_2(F, N_\Sigma)$ and $d_3(F, N_\Sigma)$ bounded by $C(m, \Sigma) \cdot (\beta_1 + \beta_2)$.

- Last, Peccati, Schulte: $\mathbb{E} |D_x L^{-1} F|^p$ and $\mathbb{E} |D_{x,y}^2 L^{-1} F|^p$ bounded by moments of difference operators.

- Let $F_i := s^{-1/2} \bar{H}_s^{(i)}$; $H_s^{(i)} := \sum_{x \in P_{sg} \cap A_i} \xi_s^{(i)}(x, P_{sg})$. 

Joseph Yukich (joint with Matthias Schulte, Bern) (Lehigh University) 
Rates of Multivariate Normal Approximation for Statistics in Geometric Probability
Proof idea for $d_2, d_3$

· Schulte + Y: Let $F_i := s^{-1/2} \bar{H}_s^{(i)}$. $F_i$ is a sum of stabilizing scores. Then the integrals of moments of difference operators are $O(s^{-1/2})$. We get

$$\beta_1 \leq C \sum_{i,j=1}^{m} |\sigma_{ij} - \text{Cov}(F_i, F_j)| + O(s^{-1/2}),$$

Thus $\tilde{d}(\bar{H}_s(1), \ldots, \bar{H}_s(m), N, \Sigma) \leq Cs^{-1/2}, s \geq 1$, for $\tilde{d} \in \{d_2, d_3\}$. 

Joseph Yukich (joint with Matthias Schulte, Bern) (Lehigh University) 

Charles Stein's Influence on Probability 17
Proof idea for $d_2$, $d_3$

- Schulte + Y: Let $F_i := s^{-1/2} \bar{H}_s^{(i)}$. $F_i$ is a sum of stabilizing scores. Then the integrals of moments of difference operators are $O(s^{-1/2})$. We get

$$\beta_1 \leq C \sum_{i,j=1}^m |\sigma_{ij} - \text{Cov}(F_i, F_j)| + O(s^{-1/2}),$$

and

$$\beta_2 = O(s^{-1/2}).$$

Thus

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

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

1. **Stein:** Let $F = (F_1, \ldots, F_m)$ be a vector of Poisson functionals; let $\Sigma \in \mathbb{R}^{m \times m}$ be positive definite; $h : \mathbb{R}^m \to \mathbb{R}$.

To assess the difference $\mathbb{E} h(F) - \mathbb{E} (h(N_\Sigma))$ over a class of test functions it is enough to assess the difference

$$
\mathbb{E} \sum_{i=1}^{m} F_i \frac{\partial f_h}{\partial y_i}(F) - \frac{\partial^2 f_h}{\partial y_i^2}(F),
$$

where $f_h : \mathbb{R}^m \to \mathbb{R}$ is a solution of the multivariate Stein equation:

$$
\sum_{i=1}^{m} y_i \frac{\partial f}{\partial y_i}(y) - \frac{\partial^2 f}{\partial y_i^2}(y) = h(y) - \mathbb{E} h(N_\Sigma), \ y \in \mathbb{R}^m.
$$
Proof idea for $d_{\mathbb{H}_\ell}, d_{\text{convex}}$

- **1. Stein:** Let $F = (F_1, \ldots, F_m)$ be a vector of Poisson functionals; let $\Sigma \in \mathbb{R}^{m \times m}$ be positive definite; $h : \mathbb{R}^m \to \mathbb{R}$.

  To assess the difference $\mathbb{E} h(F) - \mathbb{E} (h(N_{\Sigma}))$ over a class of test functions it is enough to assess the difference

  $$\mathbb{E} \sum_{i=1}^m F_i \frac{\partial f_h}{\partial y_i}(F) - \frac{\partial^2 f_h}{\partial y_i^2}(F),$$

  where $f_h : \mathbb{R}^m \to \mathbb{R}$ is a solution of the multivariate Stein equation:

  $$\sum_{i=1}^m y_i \frac{\partial f}{\partial y_i}(y) - \frac{\partial^2 f}{\partial y_i^2}(y) = h(y) - \mathbb{E} h(N_{\Sigma}), \; y \in \mathbb{R}^m.$$

- Given $t \in (0, 1)$, and test function $h$, we introduce its smoothed version

  $$h_{t,\Sigma}(y) := \int_{\mathbb{R}^m} h(\sqrt{t}z + \sqrt{1-t}y) \phi_{\Sigma}(z) dz,$$

  where $\phi_{\Sigma}(z)$ is the density of $N_{\Sigma}$. 
Proof idea for $d_{\mathbb{H}_\ell}$

2. Smoothing lemma:

$$d_{\mathbb{H}_\ell}(F, N_{\Sigma}) \leq 2 \sup_{h \in \mathbb{H}_\ell} |\mathbb{E} h_{t,\Sigma}(F) - \mathbb{E} h_{t,\Sigma}(N_{\Sigma})| + 24\ell \sqrt{mt/\pi}.$$ 

So it is enough to assess the difference of expectations over a smooth class of test functions.
Proof idea for $d_{\mathbb{H}_\ell}$

2. Smoothing lemma:

$$d_{\mathbb{H}_\ell}(F, N_\Sigma) \leq 2 \sup_{h \in \mathbb{H}_\ell} |\mathbb{E} h_{t, \Sigma}(F) - \mathbb{E} h_{t, \Sigma}(N_\Sigma)| + 24\ell \sqrt{mt}/\sqrt{\pi}.$$ 

So it is enough to assess the difference of expectations over a smooth class of test functions.

3. Peccati + Zheng (Malliavin calculus on Poisson space):

$$\mathbb{E} h_{t, \Sigma}(F) - \mathbb{E} h_{t, \Sigma}(N_\Sigma) = \sum_{i,j=1}^{m} \sigma_{ij} \mathbb{E} \frac{\partial^2 f_{t,h,\Sigma}}{\partial y_i \partial y_j}(F)$$

$$- \sum_{i=1}^{m} \mathbb{E} \int_{\mathbb{X}} D_x \frac{\partial f_{t,h,\Sigma}}{\partial y_k}(F)(-D_x L^{-1} F_k) \lambda(dx).$$

Here $f_{t,h,\Sigma}$ is the sol. to the MV Stein eq. associated with $h_{t,\Sigma}$. 

Joseph Yukich (joint with Matthias Schulte, Lehigh University)
4. Good sup norm and $L^2$ bounds on the 2nd derivatives of $f_{t,h,\Sigma}$.

$$\sup_{h \in \mathbb{H}_\ell} \mathbb{E} \sum_{i,j=1}^m \left( \frac{\partial^2 f_{t,h,\Sigma}}{\partial y_i \partial y_j} (F) \right)^2$$

$$\leq \left\| \Sigma^{-1} \right\|_{op}^2 \left( m^2 (\log t)^2 d_{\mathbb{H}_{2\ell}} (F, N_\Sigma) + 444m^{23/6} \right)$$
Proof idea for $d_{\mathbb{H}_\ell}$

5. Combine steps 2, 3, 4. We get

$$d_{\mathbb{H}_\ell}(F, N_\Sigma) \leq 2 \sup_{h \in \mathbb{H}_\ell} |\mathbb{E} h_{t, \Sigma}(F') - \mathbb{E} h_{t, \Sigma}(N_\Sigma)| + 24\ell \sqrt{m} \sqrt{t}/\sqrt{\pi},$$

$$\leq C|\log t| \sqrt{d_{\mathbb{H}_{2\ell}}(F, N_\Sigma)} + 24\ell \sqrt{m} \sqrt{t}/\sqrt{\pi} + \ldots$$
Proof idea for $d_{H_\ell}$

5. Combine steps 2, 3, 4. We get

$$d_{H_\ell}(F, N_\Sigma) \leq 2 \sup_{h \in H_\ell} |\mathbb{E} h_{t,\Sigma}(F') - \mathbb{E} h_{t,\Sigma}(N_\Sigma)| + 24\ell \sqrt{m} \sqrt{t}/\sqrt{\pi},$$

$$\leq C|\log t| \sqrt{d_{H_2\ell}(F, N_\Sigma)} + 24\ell \sqrt{m} \sqrt{t}/\sqrt{\pi} + \ldots$$

- Choose $\sqrt{t}$ as the maximum of terms growing like

$$\Gamma_1, \Gamma_2, \sum_{i,j=1}^m |\sigma_{ij} - \text{Cov}(F_i, F_j)|, \ldots$$

This leads to the following theorem.
Proof idea for $d_{\mathbb{H}_\ell}$

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

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

and

$$
\mathbb{E} |D^2_{x_1, x_2} F_i|^p < \infty, \quad \lambda\text{-a.e. } x_1, x_2 \in \mathbb{R}^d.
$$
Proof idea for $d_{\mathbb{H}_\ell}$

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

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

and

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

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

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

$$
\Gamma_1 := \left[ \sum_{j,k=1}^{m} \int_X \mathbb{E}(D_x F_j)^4 \lambda(dx) \right. \\
+ 6 \int_{X^2} \left( \mathbb{E} D_{x,y}^2 F_j \right)^{4/2} \left( \mathbb{E} (D_x F_k)^4 \right)^{1/2} \lambda^2(d(x,y)) \\
\left. + 3 \int_{X^2} \left( \mathbb{E} D_{x,y}^2 F_j \right)^{4/2} \left( \mathbb{E} (D_{x,y} F_k)^4 \right)^{1/2} \lambda^2(d(x,y)) \right]^{1/2}
$$
Let $F_i := s^{-1/2} \bar{H}_s^{(i)}$; $H_s^{(i)} := \sum_{x \in P_{sg} \cap A_i} \xi^{(i)}_s(x, P_{sg})$.

Then $\Gamma_1(p) = O(s^{-1/2})$, $\Gamma_3(p) = O(s^{-1/2})$ and we get...
Proof idea for $d_{\mathbb{H}_\ell}$

Let $F_i := s^{-1/2} \bar{H}_s^{(i)}; \ H_s^{(i)} := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg})$.

Then $\Gamma_1(p) = O(s^{-1/2}), \Gamma_3(p) = O(s^{-1/2})$ and we get

**Theorem (Schulte + Y.)** Assume $(\xi_s^{(1)})_{s \geq 1}, \ldots, (\xi_s^{(m)})_{s \geq 1}$ are

(i) exponentially stabilizing, and

(ii) satisfy the $p$-moment condition for some $p > 6$. 

Joseph Yukich (joint with Matthias Schulte, Bern) (Lehigh University) 

Rates of Multivariate Normal Approximation for Statistics in Geometric Probability

Charles Stein's Influence on Probability 24 / 29
Proof idea for $d_{\mathbb{H}_\ell}$

Let $F_i := s^{-1/2} \bar{H}_s^{(i)}$; $H_s^{(i)} := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi_s^{(i)}(x, \mathcal{P}_{sg})$.

Then $\Gamma_1(p) = O(s^{-1/2})$, $\Gamma_3(p) = O(s^{-1/2})$ and we get

**Theorem (Schulte + Y.)** Assume $(\xi_s^{(1)})_{s \geq 1}, \ldots, (\xi_s^{(m)})_{s \geq 1}$ are
(i) exponentially stabilizing, and
(ii) satisfy the $p$-moment condition for some $p > 6$.

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

$$d_{\mathbb{H}_\ell}(s^{-1/2}(\bar{H}_s^{(1)}, \ldots, \bar{H}_s^{(m)}), N_{\Sigma}) \leq Cs^{-1/d}, \quad s \geq 1.$$
(i) **Multivariate statistics of $k$NN 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 $\infty$. Then for all $k_i \in \mathbb{N}$, $1 \leq i \leq m$, we have

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

for $d \in \{d_2, d_3, d_{\mathbb{H}_\ell}\}$. 
(ii) Multivariate statistics of random geometric graph. Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite point set. Put $N_s^{(i)}(\mathcal{X})$ to be the number of components of random geometric graph $G(s^{1/d}\mathcal{X}, s^{1/d}r)$ of size $i$. 
(ii) **Multivariate statistics of random geometric graph.** Let $\mathcal{X} \subset \mathbb{R}^d$ be a finite point set. Put $N_s^{(i)}(\mathcal{X})$ 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 $\infty$. When $r = \rho s^{-1/d}$ we have for all $i_j \in \mathbb{N}$, $1 \leq j \leq m$

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

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

Joseph Yukich (joint with Matthias Schulte, Bern) (Lehigh University) 

Rates of Multivariate Normal Approximation for Statistics in Geometric Probability 26 / 29
Recap

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

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

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

We have found rates of multivariate normal convergence for the vector

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

Extensions:

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

(ii) rates of multivariate normal convergence for random measures

$$\mu^{(i)}_s(\mathcal{P}_{sg}) := \sum_{x \in \mathcal{P}_{sg} \cap A_i} \xi^{(i)}_s(x, \mathcal{P}_{sg}) \delta_x, \quad A_i \subset W.$$
THANK YOU
(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 \( i \), \( 1 \leq i \leq m \).

**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 \( \infty \). We have

\[
d(s^{-1/2}(\bar{H}_s^{(1)}(\mathcal{P}_{sg}), ..., \bar{H}_s^{(m)}(\mathcal{P}_{sg})), N_\Sigma) \leq Cs^{-1/d}, \ s \geq 1,
\]

for \( d \in \{d_2, d_3, d_{\text{H}_\ell}, d_{\text{convex}}\} \).

This vector features in tests for equality of distributions.