Lecture 2: Central limit theorems for statistics of geometric structures

Joe Yukich

Spring school at Darmstadt, February 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 2: Central limit theorems for statistics of geometric structures

- **I Introduction**
- **II Models and Results**

  Random packing models, Growth models, Gilbert Graph

Lecture 2: Central limit theorems for statistics of geometric structures

- **I Introduction**
- **II Models and Results**

  Random packing models, Growth models, Gilbert Graph
- **III Key Idea**

  Stability of difference operator
- **IV Central limit theorems**

  Statement

  Proof
- **V Large Deviation Principles**

I Introduction

· Most models of physical systems involve particles which interact ‘locally’, inducing long-range interactions.

· We take our particles to be points, usually the realization of an i.i.d. collection of r.v. $X_i, i \geq 1$ or a homogeneous Poisson point process $\mathcal{P}_1$.

I Introduction

· Most models of physical systems involve particles which interact 'locally', inducing long-range interactions.

· We take our particles to be points, usually the realization of an i.i.d. collection of r.v. $X_i, i \geq 1$ or a homogeneous Poisson point process $\mathcal{P}_1$.

· For ease of exposition, we consider $Q_n := [-\frac{1}{2}n, \frac{1}{2}n]^d$ and let $U_i, i \leq n$ be i.i.d. uniform on Q_n.

· We let H be a generic functional defined on finite point sets.

· We are interested in the behavior of the **Poisson** functional $H(\mathcal{P}_1 \cap Q_n)$ and the **binomial** functional $H(U_1, ..., U_n)$.

I Introduction

Natural questions:

1. (LLN) When do we have $\lim_{n\to\infty} \frac{H(U_1,\ldots,U_n)}{n} = \text{constant} \quad \text{a.s.?}$

I Introduction

Natural questions:

1. (LLN) When do we have $\lim_{n\to\infty} \frac{H(U_1,...,U_n)}{n} = \text{constant} \quad \text{a.s.?}$
2. (CLT) When do we have

$$\frac{H(U_1,...,U_n) - \mathbb{E}\, H(U_1,...,U_n)}{\sqrt{\mathrm{Var} H(U_1,...,U_n)}} \xrightarrow{\mathcal{D}} N(0,1)?$$

Natural questions:

1. (LLN) When do we have $\lim_{n\to\infty} \frac{H(U_1,...,U_n)}{n} = \text{constant} \quad \text{a.s.?}$
2. (CLT) When do we have

$$\frac{H(U_1,...,U_n) - \mathbb{E}\, H(U_1,...,U_n)}{\sqrt{\mathrm{Var} H(U_1,...,U_n)}} \xrightarrow{\mathcal{D}} N(0,1)?$$

3. (Probability bounds) Seek good bounds for

$$\mathbb{P}(H(U_1,...,U_n) \geq t).$$

a. **Packing Model**. Unit volume balls $B_1, ..., B_k$ arrive sequentially and uniformly at random in the cube $Q_n := [-\frac{1}{2}n, \frac{1}{2}n]^d$. Packing rules:

· Pack ball B_1.

II Models and Results

a. **Packing Model**. Unit volume balls $B_1, ..., B_k$ arrive sequentially and uniformly at random in the cube $Q_n := [-\frac{1}{2}n, \frac{1}{2}n]^d$. Packing rules:

· Pack ball B_1.

· Pack ball $B_i, \ i > 1$, if B_i does not overlap any ball in $B_1, B_2, ..., B_{i-1}$ which has already been packed.

· Picture for $d = 1$ looks like this:

· **a. Packing Model (contd)**. Fix $k \in \{1, 2, ..., \infty\}$. If balls $B_1, ..., B_k$ have centers at points $U_1, ..., U_k \in Q_n$ with respective arrival times $\tau_1, ..., \tau_k$, then let

$$H_n(U_1, ..., U_k)$$

denote the number of packed (accepted) balls on the substrate Q_n.

· The random variable $H_n(U_1, ..., U_k)$ is the number of accepted particles in the random sequential adsorption (RSA) model. $H_n(U_1, ..., U_\infty)$ is 'packing number'.

· **a. Packing Model (contd)**. Fix $k \in \{1, 2, ..., \infty\}$. If balls $B_1, ..., B_k$ have centers at points $U_1, ..., U_k \in Q_n$ with respective arrival times $\tau_1, ..., \tau_k$, then let

$$H_n(U_1, ..., U_k)$$

denote the number of packed (accepted) balls on the substrate Q_n.

· The random variable $H_n(U_1, ..., U_k)$ is the number of accepted particles in the random sequential adsorption (RSA) model. $H_n(U_1, ..., U_\infty)$ is 'packing number'.

Rényi's Thm: $d = 1 \Rightarrow$

$$\lim_{n\to\infty} \frac{\mathbb{E}\, H_n(U_1, ..., U_\infty)}{n} = \int_0^\infty \exp(-2 \int_0^t \frac{1 - e^{-u}}{u} du) dt \sim 0.748.$$

Dvoretsky + Robbins CLT: $d = 1 \Rightarrow$

$$\frac{H_n(U_1, ..., U_\infty) - \mathbb{E}\, H_n(U_1, ..., U_\infty)}{\sqrt{\mathrm{Var} H_n(U_1, ..., U_\infty)}} \xrightarrow{\mathcal{D}} N(0, 1).$$

· **a. Packing Model (contd)** What about the case $d \geq 2$? Hurdles: lack of subadditivity, lack of independence over subsets of cube Q_n.

LLN: $d \geq 2, \alpha \in (0, \infty)$:

$$\lim_{n \to \infty} \frac{H_n(U_1, ..., U_{[\alpha n]})}{n} = C(\alpha) \quad a.s.$$

· **a. Packing Model (contd)** What about the case $d \geq 2$? Hurdles: lack of subadditivity, lack of independence over subsets of cube Q_n.

LLN: $d \geq 2, \alpha \in (0, \infty)$:

$$\lim_{n \to \infty} \frac{H_n(U_1, ..., U_{[\alpha n]})}{n} = C(\alpha) \ \ a.s.$$

CLT: $d \geq 2, \alpha \in (0, \infty)$:

$$\frac{H_n(U_1, ..., U_{[\alpha n]}) - \mathbb{E}\, H_n(U_1, ..., U_{[\alpha n]})}{\sqrt{\mathrm{Var} H_n(U_1, ..., U_{[\alpha n]})}} \xrightarrow{\mathcal{D}} N(0, 1).$$

b. Geometric Graph (Gilbert Graph)

Def. Given a finite point set $\mathcal{X}$, $r \in (0, \infty)$, put

$$C_B(\mathcal{X}, r) := \bigcup_{x \in \mathcal{X}} B_r(x).$$

When $\mathcal{X}$ is PPP we get the Boolean model. It gives rise to the geometric graph $GG_r(\mathcal{X})$: join two points x and y with an edge iff $B_{r/2}(x) \cap B_{r/2}(y) \neq \emptyset$.

b. Geometric Graph (Gilbert Graph)

Def. Given a finite point set $\mathcal{X}$, $r \in (0, \infty)$, put

$$C_B(\mathcal{X}, r) := \bigcup_{x \in \mathcal{X}} B_r(x).$$

When $\mathcal{X}$ is PPP we get the Boolean model. It gives rise to the geometric graph $GG_r(\mathcal{X})$: join two points x and y with an edge iff $B_{r/2}(x) \cap B_{r/2}(y) \neq \emptyset$.

Def. Let $H(\mathcal{X})$ be the number of edges in geometric graph $GG_r(\mathcal{X})$.

CLT: $d \geq 2, r > 0$:

$$\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)}{\sqrt{\mathrm{Var} H(\mathcal{P}_1 \cap Q_n)}} \xrightarrow{\mathcal{D}} N(0, 1).$$

c. Spatial Birth Growth Models

· The model: cells form at random locations $U_1, ..., U_k \in Q_n$ at times $\tau_1, ..., \tau_k$, respectively.

· Initially the new cell around U_i takes the form of a ball of radius $R_i \geq 0$ centered at U_i; then the cell grows radially in all directions with constant speed v.

c. Spatial Birth Growth Models

· The model: cells form at random locations $U_1, ..., U_k \in Q_n$ at times $\tau_1, ..., \tau_k$, respectively.

· Initially the new cell around U_i takes the form of a ball of radius $R_i \geq 0$ centered at U_i; then the cell grows radially in all directions with constant speed v.

· New cells form only in the *uncovered space* in $\mathbb{R}^d$.

· This models crystal growth, cavitation.

· $H(U_1, ..., U_n)$ is the volume of the region covered by the first n cells.

c. Spatial Birth Growth Models (contd)

· $H(U_1, ..., U_n)$ is the volume of the region covered by the first n cells.

CLT: $\frac{H(U_1,...,U_n) - \mathbb{E}\, H(U_1,...,U_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2).$

Chiu, Quine,...

III Key Ideas

Recall:

H: a function defined on point set of $\mathbb{R}^d$.

$Q_n := [-\frac{1}{2}n, \frac{1}{2}n]^d$

$U_i, 1 \leq i \leq n$, i.i.d. uniform on Q_n.

Goal: Seek conditions on H yielding

$$\frac{H(U_1, ..., U_n)}{n} \xrightarrow{P} \text{constant} \quad (LLN)$$

and

$$\frac{H(U_1, ..., U_n) - \mathbb{E}\, H(U_1, ..., U_n)}{\sqrt{\mathrm{Var} H(U_1, ..., U_n)}} \xrightarrow{\mathcal{D}} N(0,1) \quad (CLT)$$

III Key Ideas

- Write $Q_n := \cup_{i=1}^n Q_{n,i}$, where $Q_{n,i}$ are disjoint sub-cubes of volume 1.
- Abbreviate $\{U_1, ..., U_n\}$ by $\mathcal{U}_n$.
- In general

$$H(\mathcal{U}_n) \neq \sum_{i=1}^{n} H(\mathcal{U}_n \cap Q_{n,i}),$$

i.e., H is NOT additive.

III Key Ideas

· Write $Q_n := \cup_{i=1}^n Q_{n,i}$, where $Q_{n,i}$ are disjoint sub-cubes of volume 1.

· Abbreviate $\{U_1, ..., U_n\}$ by $\mathcal{U}_n$.

· In general

$$H(\mathcal{U}_n) \neq \sum_{i=1}^n H(\mathcal{U}_n \cap Q_{n,i}),$$

i.e., H is NOT additive.

· If H were additive, then we could deduce LLN and CLT for $H(\mathcal{U}_n)$ from the classical limit theorems.

III Key Ideas

· Restriction of H to disjoint sets does not give independence

· $H(\mathcal{U}_n \cap A)$ and $H(\mathcal{U}_n \cap B)$ are dependent in general!

III Key Ideas

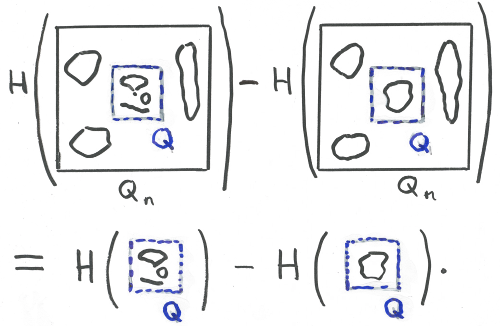
A key idea is to compare the functional $H(\mathcal{U}_n)$ with $H(\mathcal{U}_n \cup \{\mathbf{0}\})$.

Let's start by comparing $H(\mathcal{P}_1 \cap Q_n)$ with $H((\mathcal{P}_1 \cap Q_n) \cup \{\mathbf{0}\})$.

In other words, what happens to H when we insert an extra point at the origin into homogenous rate 1 Poisson input $\mathcal{P}_1$? How does H change?

III Key Ideas

Consider this situation:



This picture says that for the fixed deterministic square Q we have

$$H((\mathcal{P}_1 \cap Q_n) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q_n)$$

$$= H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q).$$

III Key Ideas: Stability of Difference Operators

The above phenomena rarely happens. Fortunately, a slightly weaker one does. Let $\mathcal{P}_1$ be unit intensity PPP on $\mathbb{R}^d$.

Def. Say that H **stabilizes** on $\mathcal{P}_1$ if there is a cube Q, $\mathrm{diam}(Q) < \infty$ a.s., such that

$$\lim_{n \to \infty} H((\mathcal{P}_1 \cap Q_n) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q_n)$$

$$= H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q).$$

III Key Ideas: Stability of Difference Operators

The above phenomena rarely happens. Fortunately, a slightly weaker one does. Let $\mathcal{P}_1$ be unit intensity PPP on $\mathbb{R}^d$.

Def. Say that H **stabilizes** on $\mathcal{P}_1$ if there is a cube Q, $\mathrm{diam}(Q) < \infty$ a.s., such that

$$\lim_{n\to\infty} H((\mathcal{P}_1 \cap Q_n) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q_n)$$

$$= H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q).$$

This condition says that the 'add-one cost' does not propagate far; it is confined to Q.

III Key Ideas: Stability of Difference Operators

Def. Say that H **stabilizes** on $\mathcal{P}_1$ if there is a cube Q, $\mathrm{diam}(Q) < \infty$ a.s., such that

$$\lim_{n\to\infty} H((\mathcal{P}_1 \cap Q_n) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q_n)$$

$$= H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q).$$

Def. $D_{\mathbf{0}} H(\mathcal{P}_1) := H(\mathcal{P}_1 \cup \{\mathbf{0}\}) - H(\mathcal{P}_1)$.

III Key Ideas: Stability of Difference Operators

Def. Say that H **stabilizes** on $\mathcal{P}_1$ if there is a cube Q, $\mathrm{diam}(Q) < \infty$ a.s., such that

$$\lim_{n\to\infty} H((\mathcal{P}_1 \cap Q_n) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q_n)$$

$$= H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q).$$

Def. $D_{\mathbf{0}}H(\mathcal{P}_1) := H(\mathcal{P}_1 \cup \{\mathbf{0}\}) - H(\mathcal{P}_1)$.

$D_{\mathbf{0}}$ is called the **first order difference operator**.

Stabilization says that the first order difference operator has a behavior which is determined by local data.

III Key Ideas: Stability of Difference Operators

Which functionals H stabilize in the above sense?

Consider the nearest neighbors graph (put an edge between every point and its nearest neighbor). Let $H(\mathcal{X})$ be the total edge length of the nearest neighbor graph on $\mathcal{X}$.

IV General CLT and Variance Asymptotics

Theorem. Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii) $(2+\epsilon)$ bounded increments:

$$\sup_{Q \subset \mathbb{R}^d,\ Q \text{ a cube}} \mathbb{E}\,|H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost" (ie. stabilizes)

IV General CLT and Variance Asymptotics

Theorem. Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii) $(2+\epsilon)$ bounded increments:

$$\sup_{Q \subset \mathbb{R}^d,\ Q \text{ a cube}} \mathbb{E}\,|H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost" (ie. stabilizes)

Then

$$\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\,H(\mathcal{P}_1 \cap Q_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

$$\frac{\mathrm{Var} H(\mathcal{P}_1 \cap Q_n)}{n} \to \sigma^2.$$

Applications

We can show that the following functionals stabilize and satisfy the CLT and variance asymptotics:

a. The number of balls accepted in the RSA packing model,

b. The volume of the occupied region in spatial birth-growth models,

Applications

We can show that the following functionals stabilize and satisfy the CLT and variance asymptotics:

a. The number of balls accepted in the RSA packing model,

b. The volume of the occupied region in spatial birth-growth models,

c. The number of edges in the random geometric graph with parameter r,

d. Total edge length of nearest neighbors graph.

IV General CLT and Variance Asymptotics

Theorem. Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii)$(2+\epsilon)$ bounded increments:

$$\sup_Q \mathbb{E}\left|H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)\right|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost".

IV General CLT and Variance Asymptotics

Theorem. Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii)$(2+\epsilon)$ bounded increments:

$$\sup_Q \mathbb{E}\,|H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost".

Then

$$\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\,H(\mathcal{P}_1 \cap Q_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

$$\frac{\mathrm{Var} H(\mathcal{P}_1 \cap Q_n)}{n} \to \sigma^2.$$

Limitations:

IV General CLT and Variance Asymptotics

Theorem. Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii)$(2+\epsilon)$ bounded increments:

$$\sup_Q \mathbb{E}\, |H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost".

Then

$$\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

$$\frac{\mathrm{Var} H(\mathcal{P}_1 \cap Q_n)}{n} \to \sigma^2.$$

Limitations:

· no formula for σ^2, no rate of normal convergence, does not address non-uniform input.

IV General CLT and Variance Asymptotics

Theorem. Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii)$(2 + \epsilon)$ bounded increments:

$$\sup_Q \mathbb{E}\left|H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)\right|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost".

Then

$$\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

$$\frac{\mathrm{Var} H(\mathcal{P}_1 \cap Q_n)}{n} \to \sigma^2.$$

Limitations:

· no formula for σ^2, no rate of normal convergence, does not address non-uniform input.

· we are unable to show that L_{TSP} stabilizes. Same for L_{MM}, L_{SMST}.

IV General CLT and Variance Asymptotics

Theorem. Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii) $(2+\epsilon)$ bounded increments:

$$\sup_{Q} \mathbb{E}\,|H((\mathcal{P}_1 \cap Q) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost".

Then

$$\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\,H(\mathcal{P}_1 \cap Q_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,\sigma^2)$$

$$\frac{\mathrm{Var} H(\mathcal{P}_1 \cap Q_n)}{n} \to \sigma^2.$$

If the add-one cost $D_{\mathbf{0}} H(\mathcal{P}_1) := H(\mathcal{P}_1 \cup \{\mathbf{0}\}) - H(\mathcal{P}_1)$ is non-degenerate, then $\sigma^2 > 0$.

IV General CLT and Variance Asymptotics

Proof of CLT. The proof involves these ingredients:

1. A stability lemma for stabilizing functionals,

2. Expressing H as a sum of martingale differences,

3. McLeish CLT (this result says that to prove a CLT it is enough to prove an L^1 WLLN).

IV General CLT and Variance Asymptotics

If H is stabilizing and trans. invariant, then there is a r.v. Δ_0 such that

$$\lim_{n\to\infty}[H((\mathcal{P}_1 \cap Q_n) \cup \mathbf{0}) - H(P_1 \cap Q_n)] = \mathbf{\Delta_0} \quad a.s.$$

IV General CLT and Variance Asymptotics

If H is stabilizing and trans. invariant, then there is a r.v. Δ_0 such that

$$\lim_{n\to\infty} [H((\mathcal{P}_1 \cap Q_n) \cup \mathbf{0}) - H(P_1 \cap Q_n)] = \Delta_{\mathbf{0}} \quad a.s.$$

Thus, if H is stabilizing and trans. invariant, then $\forall x, y \in \mathbb{R}^d$

$$H((\mathcal{P}_1 \cap Q_n) \cup x) - H(\mathcal{P}_1 \cap Q_n) \to \Delta_x \quad a.s.$$

$$H((\mathcal{P}_1 \cap Q_n) \cup y) - H(\mathcal{P}_1 \cap Q_n) \to \Delta_y \quad a.s.$$

Consequently,

$$H((\mathcal{P}_1 \cap Q_n) \cup x) - H((\mathcal{P}_1 \cap Q_n) \cup y) \to \Delta_x - \Delta_y \quad a.s.$$

i.e, replacing x by y produces a change in H which converges a.s.

Stability lemma for stabilizing functionals.

Def. $\forall x \in \mathbb{Z}^d$ let $C(x)$ be unit volume cube with center x.

Def. $\mathcal{P}_1'$ an independent copy of $\mathcal{P}_1$. $\forall x \in \mathbb{Z}^d$ let

$$\mathcal{P}_{C(x)} := \mathcal{P}_1 \setminus (\mathcal{P}_1 \cap C(x)) \cup (\mathcal{P}_1' \cap C(x)).$$

Stability lemma for stabilizing functionals.

Def. $\forall x \in \mathbb{Z}^d$ let $C(x)$ be unit volume cube with center x.

Def. $\mathcal{P}_1'$ an independent copy of $\mathcal{P}_1$. $\forall x \in \mathbb{Z}^d$ let

$$\mathcal{P}_{C(x)} := \mathcal{P}_1 \setminus (\mathcal{P}_1 \cap C(x)) \cup (\mathcal{P}_1' \cap C(x)).$$

Stability Lemma. Let H be stabilizing and trans. invariant. Then $\forall x \in \mathbb{Z}^d$ there is a r.v. $\Delta_{C(x)}$ such that as $n \to \infty$

$$H(\mathcal{P}_1 \cap Q_n) - H(\mathcal{P}_{C(x)} \cap Q_n) \to \Delta_{C(x)}.$$

Martingale Differences

· $X_i, 1 \leq i \leq n$, independent r.v.

· $\mathcal{F}_i := \sigma(X_1, ..., X_i)$.

· Let $f : \mathbb{R}^n \to \mathbb{R}$.

· Then

$$f(X_1, ..., X_n) - \mathbb{E}\, f(X_1, ..., X_n) = \sum_{i=1}^{n} D_i,$$

where

Martingale Differences

· $X_i, 1 \leq i \leq n$, independent r.v.

· $\mathcal{F}_i := \sigma(X_1, ..., X_i)$.

· Let $f : \mathbb{R}^n \to \mathbb{R}$.

· Then

$$f(X_1, ..., X_n) - \mathbb{E}\, f(X_1, ..., X_n) = \sum_{i=1}^{n} D_i,$$

where

$$\begin{aligned} D_i &:= \mathbb{E}\,(f(X_1, ..., X_n)|\mathcal{F}_i) - \mathbb{E}\,(f(X_1, ..., X_n)|\mathcal{F}_{i-1}) \\ &= \mathbb{E}\,(f(X_1, ..., X_n)|\mathcal{F}_i) - \mathbb{E}\,(f(X_1, ..., X_i', ..., X_n)|\mathcal{F}_i). \end{aligned}$$

Here X_i' is a copy of X_i.

Martingale Differences

· $X_i, 1 \leq i \leq n$, independent r.v.

· $\mathcal{F}_i := \sigma(X_1, ..., X_i)$.

· Let $f : \mathbb{R}^n \to \mathbb{R}$.

· Then

$$f(X_1, ..., X_n) - \mathbb{E}\, f(X_1, ..., X_n) = \sum_{i=1}^{n} D_i,$$

where

$$\begin{aligned} D_i &:= \mathbb{E}\,(f(X_1, ..., X_n)|\mathcal{F}_i) - \mathbb{E}\,(f(X_1, ..., X_n)|\mathcal{F}_{i-1}) \\ &= \mathbb{E}\,(f(X_1, ..., X_n)|\mathcal{F}_i) - \mathbb{E}\,(f(X_1, ..., X_i', ..., X_n)|\mathcal{F}_i). \end{aligned}$$

Here X_i' is a copy of X_i.

Seek an analogous decomposition for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)$ in terms of martingale differences. But first:

IV General CLT and Variance Asymptotics

Thm (McLeish): Let $D_i, i \geq 1$, be martingale differences such that:

· $\sup_n \mathbb{E} \max_{i \leq n}(\frac{D_i^2}{n}) < \infty$,

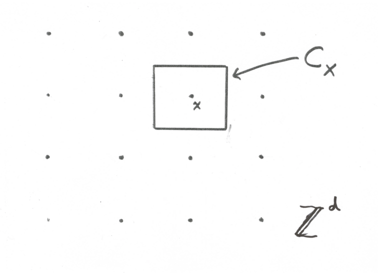
· $\frac{1}{\sqrt{n}} \max_{i \leq n} |D_i| \xrightarrow{P} 0$,

· $\frac{1}{n} \sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2$.

Then as $n \to \infty$ we have

$$\frac{\sum_{i=1}^n D_i}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2).$$

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)$



· $\mathbb{Z}^d$ ordered with dictionary ordering $\prec$.

· $\forall x \in \mathbb{Z}^d$ let $\mathcal{F}_x := \sigma(\mathcal{P}_1 \cap \cup_{y \preceq x} C_y)$.

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)$:

· $Q_n := [-\frac{1}{2}n, \frac{1}{2}n]^d$

· Write $\mathbb{Z}^d \cap Q_n = \{x_1, x_2, ..., x_n\}, \quad x_{i-1} \prec x_i$.

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)$:

· $Q_n := [-\frac{1}{2}n, \frac{1}{2}n]^d$

· Write $\mathbb{Z}^d \cap Q_n = \{x_1, x_2, ..., x_n\}, \quad x_{i-1} \prec x_i$.

· Write $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n) = \sum_{i=1}^n D_i$, with

$$\begin{aligned} D_i &:= \mathbb{E}\,(H(\mathcal{P}_1 \cap Q_n)|\mathcal{F}_{x_i}) - \mathbb{E}\,(H(\mathcal{P}_1 \cap Q_n)|\mathcal{F}_{x_{i-1}}) \\ &= \mathbb{E}\,(H(\mathcal{P}_1 \cap Q_n)|\mathcal{F}_{x_i}) - \mathbb{E}\,(H(\mathcal{P}_{C(x_i)} \cap Q_n)|\mathcal{F}_{x_i}) \\ &= \mathbb{E}\,([H(\mathcal{P}_1 \cap Q_n) - H(\mathcal{P}_{C(x_i)} \cap Q_n)] \mid \mathcal{F}_{x_i}) \end{aligned}$$

IV General CLT and Variance Asymptotics

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)$:

· $Q_n := [-\frac{1}{2}n, \frac{1}{2}n]^d$

· Write $\mathbb{Z}^d \cap Q_n = \{x_1, x_2, ..., x_n\}, \quad x_{i-1} \prec x_i$.

· Write $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n) = \sum_{i=1}^n D_i$, with

$$\begin{aligned} D_i &:= \mathbb{E}\,(H(\mathcal{P}_1 \cap Q_n)|\mathcal{F}_{x_i}) - \mathbb{E}\,(H(\mathcal{P}_1 \cap Q_n)|\mathcal{F}_{x_{i-1}}) \\ &= \mathbb{E}\,(H(\mathcal{P}_1 \cap Q_n)|\mathcal{F}_{x_i}) - \mathbb{E}\,(H(\mathcal{P}_{C(x_i)} \cap Q_n)|\mathcal{F}_{x_i}) \\ &= \mathbb{E}\,([H(\mathcal{P}_1 \cap Q_n) - H(\mathcal{P}_{C(x_i)} \cap Q_n)] \mid \mathcal{F}_{x_i}) \end{aligned}$$

To apply the McLeish CLT we need to show WLLN in L^1 sense:

$$\frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2.$$

Proof that $\frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2.$ **(**)**
(i) $\forall x \in \mathbb{Z}^d \cap Q_n$, stability lemma implies for all $1 \leq i \leq n$:

$$[H(\mathcal{P}_1 \cap Q_n) - H(\mathcal{P}_{C(x_i)} \cap Q_n)] \to \Delta_{C(x_i)} \quad a.s.$$

$(2+\epsilon)$-bounded increments condition implies uniformly in $x_i \in \mathbb{Z}^d \cap Q_n$

$$D_{x_i}^2 := (\mathbb{E}([\ldots\ldots]|\mathcal{F}_{x_i}))^2 \xrightarrow{L^1} \mathbb{E}(\Delta_{C(x_i)}|\mathcal{F}_{x_i})^2.$$

Proof that $\frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2$. **(**)**
(i) $\forall x \in \mathbb{Z}^d \cap Q_n$, stability lemma implies for all $1 \leq i \leq n$:

$$[H(\mathcal{P}_1 \cap Q_n) - H(\mathcal{P}_{C(x_i)} \cap Q_n)] \to \Delta_{C(x_i)} \quad a.s.$$

$(2+\epsilon)$-bounded increments condition implies uniformly in $x_i \in \mathbb{Z}^d \cap Q_n$

$$D_{x_i}^2 := \left(\mathbb{E}\left([......]|\mathcal{F}_{x_i}\right)\right)^2 \xrightarrow{L^1} \mathbb{E}\left(\Delta_{C(x_i)}|\mathcal{F}_{x_i}\right)^2.$$

(ii) Ergodic thm implies

$$\frac{1}{n}\sum_{x_i \in Q_n} \left(\mathbb{E}\left(\Delta_{C(x_i)}|\mathcal{F}_{x_i}\right)\right)^2 \xrightarrow{L^1} \sigma^2 := \mathbb{E}\left(\mathbb{E}\left(\Delta_{C(0)}|\mathcal{F}_0\right)\right)^2.$$

(iii) Combining (i) and (ii) we deduce the desired convergence (**).

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n :)$

Thm (McLeish): Let $D_i, i \geq 1$, be martingale differences such that:

$\cdot$ $\sup_n \mathbb{E} \max_{i \leq n} (\frac{D_i^2}{n}) < \infty$, $\frac{1}{\sqrt{n}} \max_{i \leq n} |D_i| \xrightarrow{P} 0$, $\frac{1}{n} \sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2$.

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n :)$

Thm (McLeish): Let $D_i, i \geq 1$, be martingale differences such that:

$\cdot \sup_n \mathbb{E} \max_{i \leq n} (\frac{D_i^2}{n}) < \infty$, $\frac{1}{\sqrt{n}} \max_{i \leq n} |D_i| \xrightarrow{P} 0$, $\frac{1}{n} \sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2$.

Then as $n \to \infty$ we have $\frac{\sum_{i=1}^n D_i}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$.

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n$:)

Thm (McLeish): Let $D_i, i \geq 1$, be martingale differences such that:

$\cdot \sup_n \mathbb{E} \max_{i \leq n} (\frac{D_i^2}{n}) < \infty$, $\frac{1}{\sqrt{n}} \max_{i \leq n} |D_i| \xrightarrow{P} 0$, $\frac{1}{n} \sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2$.

Then as $n \to \infty$ we have $\frac{\sum_{i=1}^n D_i}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$.

We showed the third condition. Thus we have

Martingale differences for $H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n :)$

Thm (McLeish): Let $D_i, i \geq 1$, be martingale differences such that:

$\cdot\ \sup_n \mathbb{E} \max_{i \leq n}(\frac{D_i^2}{n}) < \infty$, $\frac{1}{\sqrt{n}} \max_{i \leq n} |D_i| \xrightarrow{P} 0$, $\frac{1}{n}\sum_{i=1}^n D_i^2 \xrightarrow{L^1} \sigma^2$.

Then as $n \to \infty$ we have $\frac{\sum_{i=1}^n D_i}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$.

We showed the third condition. Thus we have

Thm (CLT). Let H be a functional on points in $\mathbb{R}^d$ which satisfies:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii) bounded increments: $\sup_Q \mathbb{E}\, |H(\mathcal{P}_1 \cap Q \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty$,

(iii) stability of 'add-one cost".

Then

$$\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \sigma^2).$$

General CLT for $H(\mathcal{P}_1 \cap Q_n)$.

Under certain moment conditions on H we have

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{H(\mathcal{P}_1 \cap Q_n) - \mathbb{E}\, H(\mathcal{P}_1 \cap Q_n)}{\sqrt{\mathrm{Var} H(\mathcal{P}_1 \cap Q_n)}} \leq x \right) - \mathbb{P}(N(0,1) \leq x) \right| \leq \sum_{i=1}^{5} \tau_i.$$

The τ_i are integrals of sums and products of first order difference operators for H:

$$D_x H(\mathcal{P}_1) := H(\mathcal{P}_1 \cup x) - H(\mathcal{P}_1)$$

as well as second order difference operators

$$D^{(2)}_{x_1,x_2} H(\mathcal{P}_1) := D_{x_1} D_{x_2}(H(\mathcal{P}_1))$$

$$= H(\mathcal{P}_1 \cup \{x_1, x_2\}) - H(\mathcal{P}_1 \cup \{x_2\}) - H(\mathcal{P}_1 \cup x_1) + H(\mathcal{P}_1).$$

The results are an improvement because they provide rates of convergence. The unwieldy terms τ_i can only be simplified when the functional H is a stabilizing functional. We discuss this in the next lecture.

Example. $X_i, i \geq 1$, i.i.d. $N(0,1)$. $S(n) = \sum_{i=1}^{n} X_i$; $\frac{S(n)}{\sqrt{n}} \stackrel{\mathcal{D}}{=} N(0,1)$.

Thus for all Borel $A \subset \mathbb{R}$ we have

$$\mathbb{P}(\frac{S(n)}{\sqrt{n}} \in A) = \frac{1}{\sqrt{2\pi}} \int_A \exp(-\frac{x^2}{2}) dx.$$

So

$$\mathbb{P}(|\frac{S(n)}{n}| \geq t) = \frac{2}{\sqrt{2\pi}} \int_{t\sqrt{n}} \exp(-\frac{x^2}{2}) dx \sim \sqrt{\frac{2}{\pi}} \frac{1}{t\sqrt{n}} \exp(-\frac{t^2 n}{2}),$$

which gives as $n \to \infty$

Example. $X_i, i \geq 1$, i.i.d. $N(0,1)$. $S(n) = \sum_{i=1}^n X_i$; $\frac{S(n)}{\sqrt{n}} \stackrel{\mathcal{D}}{=} N(0,1)$.

Thus for all Borel $A \subset \mathbb{R}$ we have

$$\mathbb{P}(\frac{S(n)}{\sqrt{n}} \in A) = \frac{1}{\sqrt{2\pi}} \int_A \exp(-\frac{x^2}{2}) dx.$$

So

$$\mathbb{P}(|\frac{S(n)}{n}| \geq t) = \frac{2}{\sqrt{2\pi}} \int_{t\sqrt{n}} \exp(-\frac{x^2}{2}) dx \sim \sqrt{\frac{2}{\pi}} \frac{1}{t\sqrt{n}} \exp(-\frac{t^2 n}{2}),$$

which gives as $n \to \infty$

$$\frac{\log \mathbb{P}(|\frac{S(n)}{n}| \geq t)}{n} \to \frac{-t^2}{2}.$$

· With small probability, $|\frac{S(n)}{n}|$ takes on relatively large values.

· $\frac{-t^2}{2}$ is the rate function.

V Large Deviation Principles

Cramer's LDP

· $X_i, i \geq 1$, i.i.d. $S(n) = \sum_{i=1}^{n} X_i$; $\mathbb{E}\, X_1 = \mu$.

· log m.g.f.

$$\Lambda(\lambda) := \log \mathbb{E}\, \exp(\lambda X_1) \geq \lambda\mu.$$

Cramer's LDP

· $X_i, i \geq 1$, i.i.d. $S(n) = \sum_{i=1}^{n} X_i$; $\mathbb{E}\, X_1 = \mu$.

· log m.g.f.

$$\Lambda(\lambda) := \log \mathbb{E}\, \exp(\lambda X_1) \geq \lambda\mu.$$

· $\Lambda(\lambda)$ is convex. Fenchel-Legendre transform (convex dual) of $\Lambda(\lambda)$ is

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda)) \geq 0.$$

· The convex dual is quadratic if X_1 is $N(\mu, \sigma^2)$.

Cramer's LDP

· Assume $\Lambda(\lambda) < \infty$. Then

(a) For all closed $F \subseteq \mathbb{R}$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{S(n)}{n} \in F) \leq - \inf_{x \in F} \Lambda^*(x).$$

(b) For all open $O \subseteq \mathbb{R}$ we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{S(n)}{n} \in O) \geq - \inf_{x \in O} \Lambda^*(x).$$

V Large Deviation Principles

Cramer's LDP

· Assume $\Lambda(\lambda) < \infty$. Then

(a) For all closed $F \subseteq \mathbb{R}$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{S(n)}{n} \in F) \leq - \inf_{x \in F} \Lambda^*(x).$$

(b) For all open $O \subseteq \mathbb{R}$ we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{S(n)}{n} \in O) \geq - \inf_{x \in O} \Lambda^*(x).$$

The sequence $\frac{S(n)}{n}$ satisfies the LDP with rate function Λ^*.

· Can one show that statistics of geometric structures also satisfy a LDP?

LDP for the TSP

· Notation: $X(n) := L_{TSP}(\mathcal{P}_1 \cap Q_n)$.

· The limit of the log m.g.f. for the sequence $X(n), n \geq 1$, is

$$\Lambda(t) := \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E} \exp(tX(n)).$$

· It is a fact that the limit exists.

LDP for the TSP

- Notation: $X(n) := L_{TSP}(\mathcal{P}_1 \cap Q_n)$.
- The limit of the log m.g.f. for the sequence $X(n), n \geq 1$, is

$$\Lambda(t) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp(tX(n)).$$

- It is a fact that the limit exists.
- Convex dual of $\Lambda(t)$ is

$$\Lambda^*(x) := \sup_{t\in\mathbb{R}} (tx - \Lambda(t)).$$

- Now we can state an LDP for $X(n) := L_{TSP}(\mathcal{P}_1 \cap Q_n)$.

Donsker-Varadhan LDP for the TSP

· $X(n) := L_{TSP}(\mathcal{P}_1 \cap Q_n)$ and $\frac{X(n)}{n} \to \gamma_{TSP}$ $a.s.$ (Lecture 1)

· $\Lambda(t) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp(tX(n)) < \infty$ and

(a) For all closed $F \subseteq \mathbb{R}$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{X(n)}{n} \in F) \leq -\inf_{x\in F} \Lambda^*(x).$$

(b) For all open $O \subseteq \mathbb{R}$ we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{X(n)}{n} \in O) \geq -\inf_{x\in O} \Lambda^*(x).$$

The sequence $\frac{X(n)}{n}$ satisfies the LDP with rate function Λ^* and

$$\Lambda^*(x) = 0 \Leftrightarrow x = \gamma_{TSP} \text{ and } \Lambda^*(x) \leq C(x - \gamma_{TSP})^2.$$

See Seppalainen + Y.

LDP for the TSP
Remark.

$\frac{X(n)}{n}$ converges exponentially to γ_{TSP}, i.e., $\forall \epsilon > 0$ there is $M := M(\epsilon) > 0$ such that $\forall n$

$$\mathbb{P}(|\frac{X(n)}{n} - \gamma_{TSP}| > \epsilon) \leq \exp(-Mn).$$

Final thoughts: LDP for the TSP

How does one prove an LDP for TSP and other statistics of large geometric structures?

Key. The sequence $X(n), n \geq 1$, satisfies near additivity in the sense that $\forall \epsilon > 0$, $\forall C > 0$, there is n_0 such that for $n \geq n_0$ and for all integers m we have

$$\mathbb{P}\left(|X(nm^d) - \sum_{i=1}^{m^d} X_i(n)| \geq \epsilon nm^d \right) \leq \exp(-Cnm^d)$$

where $X_i(n)$ are i.i.d. copies of $X(n)$.

Near additivity says that except on a set with exponentially small probability, $X(nm^d)$ can be expressed as a sum of i.i.d. random variables with a suitably small error.

THANK YOU