Probabilistic Analysis of Geometric Structures

Joe Yukich

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 1: Probabilistic analysis of Euclidean optimization problems

- **I Results**

- **II Methods**

  Subadditivity

  Boundary graphs

  Concentration inequalities

- **III Open Questions**

I Results

TSP: $\mathcal{X} \subset \mathbb{R}^d$ finite point set

· $L_{TSP}(\mathcal{X})$:= length of shortest tour through $\mathcal{X}$.

$X_i, 1 \leq i \leq n$, i.i.d. with uniform density $\kappa(x)$ on $[0,1]^d$.

A typical edge in the TSP tour through $\{X_i\}_{i=1}^n$ has length of order $n^{-1/d}$:

· Beardwood, Halton, Hammersley (1959): $X_i, 1 \leq i \leq n$, i.i.d. with density $\kappa(x)$ on $[0,1]^d$. Then

$$\lim_{n \to \infty} \frac{L_{TSP}(\{X_1, ..., X_n\})}{n^{(d-1)/d}} \overset{P}{=} \gamma_{TSP}(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx$$

· How to find the shortest tour?

Karp ('76, '77): $U_i, i \geq 1$, i.i.d. uniform on $[0,1]^d$.

For all $\epsilon > 0$, there is an algorithm producing a feasible tour through $U_1, ..., U_n$ with length

$$L^{\epsilon}_{TSP}(U_1, ..., U_n)$$

such that

(i) algorithm runs in polynomial time, and

(ii) $\frac{L^{\epsilon}_{TSP}(U_1,...,U_n)}{L_{TSP}(U_1,...,U_n)} < 1 + \epsilon, \quad$ with prob. $1 - o(1)$

$\mathcal{P}_1$: rate 1 Poisson point process on $\mathbb{R}^d$

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

Laws of large numbers (BHH '59):

$$\lim_{n\to\infty} \frac{L_{TSP}(\mathcal{P}_1 \cap Q_n)}{n} \stackrel{P}{=} \gamma_{TSP}(d)$$

Rates of convergence:

$$\left| \frac{\mathbb{E}\, L_{TSP}(\mathcal{P}_1 \cap Q_n)}{n} - \gamma_{TSP}(d) \right| \leq Cn^{-1/d}.$$

Alexander '94, Redmond-Y '94.

Minimal Spanning Tree

· $\mathcal{X} \subset \mathbb{R}^d$ finite.

· $L_{MST}(\mathcal{X}) \; := $ length of minimal spanning tree (MST) through $\mathcal{X}$.

Theorem. If $X_i, 1 \leq i \leq n$, i.i.d. with density $\kappa(x)$ on $[0,1]^d$, then

$$\lim_{n \to \infty} \frac{L_{MST}(\{X_1, ..., X_n\})}{n^{(d-1)/d}} \stackrel{P}{=} \gamma_{MST}(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx$$

Aldous, Alexander, Steele, Redmond + Y.

Minimal Matching

· $\mathcal{X} \subset \mathbb{R}^d$ finite; $\mathrm{card}(\mathcal{X})$ of even parity.

Match the points in pairs so as to minimize total edge length of matching.

· $L_{MM}(\mathcal{X})$:= length of minimal matching (MM) on $\mathcal{X}$.

Theorem. If $X_i, 1 \leq i \leq n$, i.i.d. with density $\kappa(x)$ on $[0,1]^d$, then

$$\lim_{n\to\infty} \frac{L_{MM}(\{X_1, ..., X_n\})}{n^{(d-1)/d}} \stackrel{P}{=} \gamma_{MM}(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx.$$

Steele; Redmond + Y.

Steiner MST

· $\mathcal{X} \subset \mathbb{R}^d$ finite.

· A Steiner tree on $\mathcal{X}$ is a connected graph which contains $\mathcal{X}$.

The graph may include 'Steiner points', i.e., vertices other than those in $\mathcal{X}$.

· Length of Steiner MST:

$$L_{SMST}(\mathcal{X}) := \min_S \sum_{e \in S} |e|,$$

where the minimum ranges over all Steiner trees S on $\mathcal{X}$.

Theorem. If $X_i, 1 \leq i \leq n$, i.i.d. with density $\kappa(x)$ on $[0,1]^d$, then

$$\lim_{n\to\infty} \frac{L_{SMST}(\{X_1, ..., X_n\})}{n^{(d-1)/d}} \stackrel{P}{=} \gamma_{SMST}(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx$$

Redmond + Y.

Undirected k-nearest neighbors graph

· $\mathcal{X} \subset \mathbb{R}^d$ finite, $k \in \mathbb{N}$.

· $kNN(\mathcal{X})$ is the graph on $\mathcal{X}$ putting an edge between each point in $\mathcal{X}$ and each of its k-nearest neighbors.

Example ($k = 1$):

· Total edge length of $kNN(\mathcal{X})$ is $L_{kNN}(\mathcal{X})$.

Theorem. If $X_i, 1 \leq i \leq n$, i.i.d. with density $\kappa(x)$ on $[0,1]^d$, then

$$\lim_{n \to \infty} \frac{L_{kNN}(\{X_1, ..., X_n\})}{n^{(d-1)/d}}$$

$$\stackrel{P}{=} \gamma_{kNN}(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx.$$

Redmond + Y.

Voronoi tessellations

· $\mathcal{X} := \{x_1, ..., x_n\} \subset \mathbb{R}^d$.

· Consider the set of all points closer to x_i than to any other point $x_j, j \neq i$. This convex set is denoted by $C(x_i)$.

· Example:

· $C(x_i)$ is the intersection of half-spaces.

· $\{C(x_i)\}_{x_i \in \mathcal{X}}$ is the Voronoi tessellation of $\mathbb{R}^d$ induced by $\mathcal{X}$.

· $L_{VOR}(\mathcal{X})$ is the sum of the lengths of edges of the finite edges of Voronoi tessellation induced by $\mathcal{X}$.

Theorem. $X_i, 1 \le i \le n$, i.i.d. with density $\kappa(x)$ on $[0,1]^d$. Then

$$\lim_{n\to\infty} \frac{L_{VOR}(\{X_1, ..., X_n\})}{n^{(d-1)/d}} \stackrel{P}{=} \ell_{VOR}(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx.$$

McGivney + Y.

What about the surface area of the Voronoi tessellation?

· $A_{VOR}(\mathcal{X})$ is the sum of the areas of the finite faces of Voronoi tessellation induced by $\mathcal{X}$.

Theorem. $X_i, 1 \leq i \leq n$, i.i.d. with density $\kappa(x)$ on $[0,1]^3$. Then

$$\lim_{n\to\infty} \frac{A_{VOR}(\{X_1, ..., X_n\})}{n^{1/3}} \stackrel{P}{=} \alpha_{VOR} \int_{[0,1]^3} \kappa(x)^{1/3} dx.$$

Goals

Given a graph (eg. TSP, MST, MM, SMST, kNN, Vor) whose nodes are given by n i.i.d. random variables $X_1, ..., X_n$ establish the limit theory for the sum

$$L(\{X_1, ..., X_n\})$$

of the lengths of the edges of the given graph. We seek:

· SLLN

· rates of convergence

· CLT

· LDP

More generally, we seek the limit theory for $H(\{X_1, ..., X_n\})$, where $H(\cdot)$ is a general functional of input $\{X_1, ..., X_n\}$.

Question

Given a graph (eg. TSP, MST, MM, SMST, kNN, Vor) whose nodes are given by n i.i.d. random variables $X_1, ..., X_n$, what are the crucial structural elements which are essential to teasing out a common limit theory?

II Methods

- Subadditive Euclidean functionals
- Boundary functionals
- Concentration

Euclidean Functionals

$\mathcal{X} \subset \mathbb{R}^d$ a locally finite point set

Definition. A (real-valued) functional $L(\mathcal{X})$ is a **Euclidean functional** if these two conditions are satisfied:

$$L(\mathcal{X} + y) = L(\mathcal{X}) \;\; \forall y \in \mathbb{R}^d$$

$$L(\alpha \mathcal{X}) = \alpha L(\mathcal{X}) \;\; \forall \alpha \in \mathbb{R}^+.$$

Example: The total edge length of a graph on $\mathcal{X}$.

Euclidean Functionals

$\mathcal{X} \subset \mathbb{R}^d$ a locally finite point set

Definition. A (real-valued) functional $L(\mathcal{X})$ is a **Euclidean functional of order** $p \in (0, \infty)$ if these two conditions are satisfied:

$$L(\mathcal{X} + y) = L(\mathcal{X}) \;\; \forall y \in \mathbb{R}^d$$

$$L(\alpha \mathcal{X}) = \alpha^p L(\mathcal{X}) \;\; \forall \alpha \in \mathbb{R}^+.$$

Example ($p = 2$**)**: The total area of faces in Voronoi tessellation of $\mathcal{X} \subset \mathbb{R}^3$.

Subadditive Euclidean Functionals

Definition. A Euclidean functional L is **subadditive** if there is a constant $C \in (0, \infty)$ such that for all rectangles $R := R_1 \cup R_2$ and all point sets $\mathcal{X} \subset R$:

$$L(\mathcal{X}) \leq L(\mathcal{X} \cap R_1) + L(\mathcal{X} \cap R_2) + C \mathrm{diam}(R).$$

Here R_1 and R_2 are rectangles.

Example: The TSP length $L_{TSP}(\mathcal{X})$ is subadditive:

Subadditive Euclidean Functionals

Fact: $L_{MST}, L_{SMST}, L_{MM}, L_{kNN}$ are also subadditive.

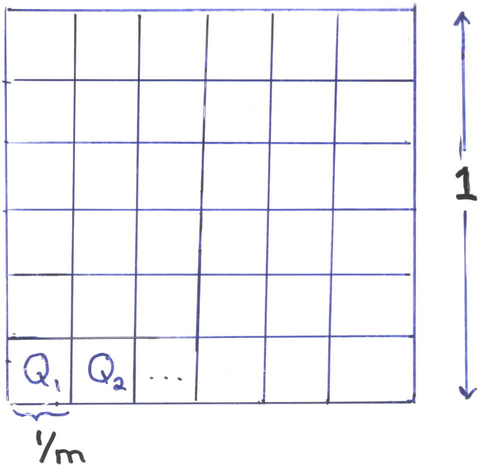
Geometric subadditivity

Let $\{Q_i\}_{i=1}^{m^d}$ be a partition of $[0,1]^d$ into subcubes of edge length m^{-1}.

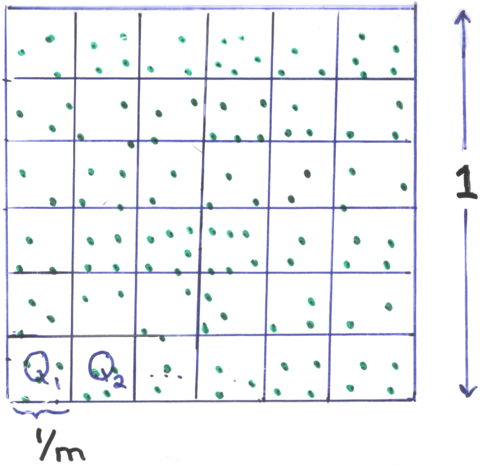
Subadditivity $L(\mathcal{X}) \leq L(\mathcal{X} \cap R_1) + L(\mathcal{X} \cap R_2) + C\mathrm{diam}(R)$ yields

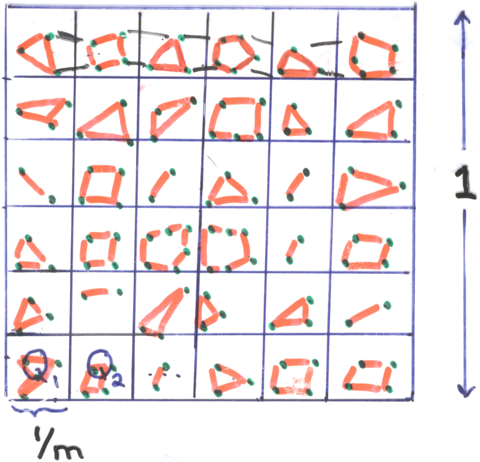
$$L(\mathcal{X} \cap [0,1]^d) \leq \sum_{i=1}^{m^d} L(\mathcal{X} \cap Q_i) + Cm^{d-1}.$$

Remark: when L is the TSP tour length and when $\mathcal{X}$ is a point set which is the realization of a homogenous Poisson point process, we are bounding the minimal tour length on points in the unit cube by a sum of identically distributed minimal tour lengths plus an error term.



Q_1
Q_2
. . .
1
1/m


Q_1
Q_2
...
1
$1/m$


1
Q1
Q2
1/m

Summary: Subadditive Euclidean functionals L satisfy for all $\mathcal{X} \subset \mathbb{R}^d$

$$L(\mathcal{X} + y) = L(\mathcal{X}) \ \ \forall y \in \mathbb{R}^d,$$

$$L(\alpha \mathcal{X}) = \alpha L(\mathcal{X}) \ \ \forall \alpha \in \mathbb{R}^+,$$

$$L(\mathcal{X} \cap [0,1]^d) \leq \sum_{i=1}^{m^d} L(\mathcal{X} \cap Q_i) + Cm^{d-1}.$$

Fact (Growth Bounds) : Let L be a subadditive Euclidean functional. Then for all $\mathcal{X} \subset [0,1]^d$ we have

$$L(\mathcal{X}) \leq C(\mathrm{card}\mathcal{X})^{(d-1)/d}.$$

Definition A Euclidean functional L is smooth if for all $\mathcal{X}_1, \mathcal{X}_2 \subset [0,1]^d$

$$|L(\mathcal{X}_1 \cup \mathcal{X}_2) - L(\mathcal{X}_2)| \le C(\mathrm{card}\mathcal{X}_1)^{(d-1)/d}.$$

Fact: L_{TSP}, L_{MST}, L_{SMST}, L_{MM}, L_{kNN} are smooth subadditive Euclidean functionals.

We show this for L_{TSP}.

Hille's 1948 Subadditive Limit Theorem. If

$$x_{m+n} \leq x_m + x_n \quad \forall m, n \in \mathbb{N}$$

then

$$\lim_{n \to \infty} \frac{x_n}{n} = \alpha$$

where

$$\alpha := \inf\{\frac{x_m}{m} : \quad m \geq 1\}.$$

Can we similarly deduce limit theorems under a geometric subadditivity condition?

Basic weak law of large numbers for smooth subadditive Euclidean functionals

Let $\mathcal{P}_n$ be a Poisson point process of intensity n on $[0,1]^d$.

Let L be a smooth subadditive Euclidean functional.

Basic Theorem for Poisson Input $\lim_{n \to \infty} \frac{\mathbb{E}\, L(\mathcal{P}_n)}{n^{(d-1)/d}} = \gamma_L(d)$.

The proof uses subadditivity and smoothness and goes as follows...

Let $Q_i, i = 1, ..., m^d$ be a partition of $[0,1]^d$ into subcubes of edge length m^{-1}.

- $\mathbb{E}\, L(\mathcal{P}_{nm^d} \cap [0,1]^d) \leq \sum_{i=1}^{m^d} \mathbb{E}\, L(\mathcal{P}_{nm^d} \cap Q_i) + Cm^{d-1}$.
- $\phi(nm^d) \leq m^d \frac{1}{m} \phi(n) + Cm^{d-1}$

- $\mathbb{E}\, L(\mathcal{P}_{nm^d} \cap [0,1]^d) \le \sum_{i=1}^{m^d} \mathbb{E}\, L(\mathcal{P}_{nm^d} \cap Q_i) + Cm^{d-1}$.
- $\phi(nm^d) \le m^d \frac{1}{m} \phi(n) + Cm^{d-1}$
- Homogenize: $\frac{\phi(nm^d)}{(nm^d)^{(d-1)/d}} \le \frac{\phi(n)}{n^{(d-1)/d}} + \frac{C}{n^{(d-1)/d}}$.
- Set $\beta := \liminf_{n\to\infty} \frac{\phi(n)}{n^{(d-1)/d}}$. Given $\epsilon > 0$, there is $n_0 \in \mathbb{N}$ such that

$$\frac{\phi(n_0)}{n_0^{(d-1)/d}} \le \beta + \epsilon; \quad \frac{C}{n_0^{(d-1)/d}} \le \epsilon.$$

· $\mathbb{E}\, L(\mathcal{P}_{nm^d} \cap [0,1]^d) \le \sum_{i=1}^{m^d} \mathbb{E}\, L(\mathcal{P}_{nm^d} \cap Q_i) + Cm^{d-1}$.

· $\phi(nm^d) \le m^d \frac{1}{m} \phi(n) + Cm^{d-1}$

· Homogenize: $\frac{\phi(nm^d)}{(nm^d)^{(d-1)/d}} \le \frac{\phi(n)}{n^{(d-1)/d}} + \frac{C}{n^{(d-1)/d}}$.

· Set $\beta := \liminf_{n\to\infty} \frac{\phi(n)}{n^{(d-1)/d}}$. Given $\epsilon > 0$, there is $n_0 \in \mathbb{N}$ such that

$$\frac{\phi(n_0)}{n_0^{(d-1)/d}} \le \beta + \epsilon; \quad \frac{C}{n_0^{(d-1)/d}} \le \epsilon.$$

Thus $\frac{\phi(n_0 m^d)}{(n_0 m^d)^{(d-1)/d}} \le \beta + 2\epsilon,\ m = 1, 2, \ldots$

Smoothness gives $\limsup_{n\to\infty} \frac{\phi(n)}{n^{(d-1)/d}} \le \beta + 2\epsilon$.

Let $\epsilon \downarrow 0$. □

To obtain:

· rates of convergence in WLLN

· WLLN over non-uniform random input

· LDP

it is useful to establish that many Euclidean functionals have an intrinsic superadditive structure.

We illustrate this with the TSP problem.

The key idea is to introduce an auxiliary TSP functional which allows ‘free’ travel on the boundary.

Boundary TSP functional

Definition. For all $\mathcal{X} \subset [0,1]^d$, let $L_{B,TSP}(\mathcal{X})$ be the length of the shortest tour through $\mathcal{X}$ where travel on the boundary of $[0,1]^d$ is 'free'.

Example.

$L_{B,TSP}(\mathcal{X})$ is superadditive with no error term. In other words for all rectangles $R := R_1 \cup R_2$ and all point sets $\mathcal{X} \subset R$ we have

$$L_{B,TSP}(\mathcal{X}) \geq L_{B,TSP}(\mathcal{X} \cap R_1) + L_{B,TSP}(\mathcal{X} \cap R_2).$$

Example.

Closeness

Definition. Let L be a Euclidean functional and let L_B be its boundary version. We say that L and L_B are 'close' if

$$|L_B(\mathcal{X}) - L(\mathcal{X})| = o((\mathrm{card}\mathcal{X})^{(d-1)/d}), \quad \mathcal{X} \subset [0,1]^d.$$

The error is negligible when compared to the growth rate.

Examples:

· $L_{B,TSP}$ and L_{TSP};

· $L_{B,MST}$ and L_{MST};

· $L_{B,MM}$ and L_{MM};

· $L_{B,SMST}$ and L_{SMST}.

Umbrella LLN

Umbrella Theorem. Assume:
· L smooth subadditive Euclidean functional
· L_B smooth superadditive Euclidean functional
· L and L_B are close, ie.,

$$|L_B(\mathcal{X}) - L(\mathcal{X})| = o((\mathrm{card}\mathcal{X})^{(d-1)/d}), \quad \mathcal{X} \subset [0,1]^d.$$

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

$$\lim_{n \to \infty} \frac{L(X_1, ..., X_n)}{n^{(d-1)/d}} = \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx \quad a.s.,$$

where $X_i, i \geq 1$, are i.i.d. with density κ.

Umbrella LLN

What about subadditive Euclidean functionals of order $p \in (0, d]$?

Umbrella LLN

What about subadditive Euclidean functionals of order $p \in (0, d]$?

Recall that these functionals are subadditive and satisfy the scaling relation:

$$L(\alpha \mathcal{X}) = \alpha^p L(\mathcal{X}) \;\; \forall \alpha \in \mathbb{R}^+.$$

Umbrella LLN

Umbrella Theorem. Assume:

· L smooth subadditive Euclidean functional of order p

· L_B smooth superadditive Euclidean functional of order p

· L and L_B are close, ie.,

$$|L_B(\mathcal{X}) - L(\mathcal{X})| = o((\mathrm{card}\mathcal{X})^{(d-p)/d}), \quad \mathcal{X} \subset [0,1]^d.$$

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

$$\lim_{n\to\infty} \frac{L(X_1, ..., X_n)}{n^{(d-p)/d}} = \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-p)/d} dx \quad a.s.,$$

where $X_i, i \geq 1$, are i.i.d. with density κ.

The umbrella theorem covers these functionals:
· total edge length of TSP

· total edge length of MST

· total edge length of MM

· total edge length of Steiner MST

· total edge length of k-nearest neighbors graph

· total edge length and surface area of Voronoi tessellation

Umbrella thm extends BHH theorem.

We sketch the proof of the umbrella theorem ($p = 1$):

Put $\mathcal{X}_n := \{X_i\}_{i=1}^n$. To prove

$$\lim_{n\to\infty} \frac{L(\mathcal{X}_n)}{n^{(d-1)/d}} = \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx \quad a.s.,$$

we will show

$$\limsup_{n\to\infty} \frac{\mathbb{E}\, L(\mathcal{X}_n)}{n^{(d-1)/d}} \le \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx,$$

$$\liminf_{n\to\infty} \frac{\mathbb{E}\, L(\mathcal{X}_n)}{n^{(d-1)/d}} \ge \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx,$$

and then we apply a concentration inequality to deduce a.s. convergence.

We will actually only prove the above inequalities for *blocked* distributions

$$\kappa(x) = \sum_{i=1}^{m^d} \alpha_i \mathbf{1}_{Q_i}(x).$$

Proof

. $\kappa(x) = \sum_{i=1}^{m^d} \alpha_i \mathbf{1}_{Q_i}(x), \quad \mathcal{X}_n := \{X_i\}_{i=1}^n$. Subadditivity gives:

$$\begin{aligned}\mathbb{E}\, L(\mathcal{X}_n) &\le \mathbb{E} \sum_{i=1}^{m^d} L(\mathcal{X}_n \cap Q_i) + Cm^{d-1} \\ &= \sum_{i=1}^{m^d} \mathbb{E}\, L(\{U_{ij}\}_{j=1}^{\mathrm{Bi}(n,\alpha_i m^d)}) + Cm^{d-1} \\ &\le \sum_{i=1}^{m^d} \mathbb{E}\, L(\{U_{ij}\}_{j=1}^{n\alpha_i m^d}) + Cm^{d-1} \\ &\quad + \sum_{i=1}^{m^d} C\mathbb{E}\, |\mathrm{Bi}(n, \alpha_i m^d) - n\alpha_i m^d|^{(d-1)/d}.\end{aligned}$$

Thus

$$\limsup_{n\to\infty} \frac{\mathbb{E}\, L(\mathcal{X}_n)}{n^{(d-1)/d}} \le \limsup_{n\to\infty} \frac{\sum_{i=1}^{m^d} \mathbb{E}\, L(\{U_{ij}\}_{j=1}^{n\alpha_i m^d})}{n^{(d-1)/d}}.$$

Thus

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{\mathbb{E}\, L(\mathcal{X}_n)}{n^{(d-1)/d}} \\
&\le \limsup_{n\to\infty} \frac{\sum_{i=1}^{m^d} \mathbb{E}\, L(\{U_{ij}\}_{j=1}^{n\alpha_i m^d})}{n^{(d-1)/d}} \\
&= \limsup_{n\to\infty} \frac{\sum_{i=1}^{m^d} m^{-1} \mathbb{E}\, L(m\{U_{ij}\}_{j=1}^{n\alpha_i m^d})}{(nm^{-d}\alpha_i)^{(d-1)/d}} \frac{(nm^{-d}\alpha_i)^{(d-1)/d}}{n^{(d-1)/d}} \\
&= \gamma_L(d) \sum_{i=1}^{m^d} m^{-1}(m^{-d}\alpha_i)^{(d-1)/d} \quad \text{(by Basic Thm for Poisson input)} \\
&= \gamma_L(d) \sum_{i=1}^{m^d} \alpha_i^{(d-1)/d} m^{-d} \\
&= \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx,
\end{aligned}
$$

which was to be proved.

We have thus proved for blocked densities κ that

$$\lim_{n\to\infty} \frac{\mathbb{E}\, L(\mathcal{X}_n)}{n^{(d-1)/d}} = \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx.$$

Using smoothness, we can extend this limit to arbitrary κ.

To deduce an a.s. result we need to show that $L(\mathcal{X}_n)$ concentrates around its mean.

Concentration

Theorem. If L is a smooth, subadditive Euclidean functional then for all $\epsilon > 0$ we have

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{L(\mathcal{X}_n) - \mathbb{E}\, L(\mathcal{X}_n)}{n^{(d-1)/d}}\right| > \epsilon\right) < \infty.$$

Combining this with the Borel-Cantelli lemma we obtain:

Corollary. If

$$\lim_{n\to\infty} \frac{\mathbb{E}\, L(\mathcal{X}_n)}{n^{(d-1)/d}} = \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx$$

then

$$\lim_{n\to\infty} \frac{L(\mathcal{X}_n)}{n^{(d-1)/d}} = \gamma_L(d) \int_{[0,1]^d} \kappa(x)^{(d-1)/d} dx \ a.s.$$

This concludes the proof of the umbrella thm.

Question: Can we obtain more precise concentration bounds?

Yes.

The answer turns on understanding what happens to a Euclidean functional when one point in the sample is changed.

Azuma-Hoeffding Inequality

Fact 1. If $X \in L^1(\Omega, \mathcal{A}, P)$, then X may be represented as a sum of martingale differences.

Proof. Given a filtration

$$(\emptyset, \Omega) := \mathcal{A}_0 \subset \mathcal{A}_1 \subset ... \subset \mathcal{A}_n := \mathcal{A}$$

we define the martingale differences

$$d_i := \mathbb{E}\,(X|\mathcal{A}_i) - \mathbb{E}\,(X|\mathcal{A}_{i-1}).$$

Then

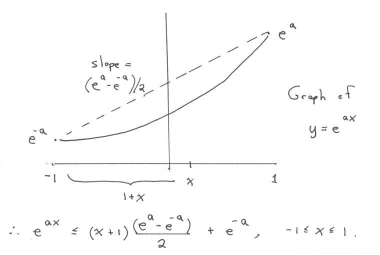
$$X - \mathbb{E}\,X = \mathbb{E}\,(X|\mathcal{A}_n) - \mathbb{E}\,(X|\mathcal{A}_0) = \sum_{i=1}^{n} d_i.$$

□

Fact 2. If $||X||_\infty \leq 1$, $\mathbb{E}\, X = 0$ then

$$\mathbb{E}\, \exp(aX) \leq \exp \frac{a^2}{2}, \quad a \in \mathbb{R}.$$

Proof. It is always the case that $\exp ax \leq \exp a^2/2 + x \sinh a$.
We prove this with a picture...



Fact 2. If $||X||_\infty \leq 1$, $\mathbb{E}\, X = 0$ then

$$\mathbb{E}\, \exp(aX) \leq \exp \frac{a^2}{2}, \quad a \in \mathbb{R}.$$

Fact 3. If $\mathbb{E}\, X = 0$ then

$$\mathbb{E}\, \exp(aX) \leq \exp(\frac{a^2||X||_\infty^2}{2}), \quad a \in \mathbb{R}.$$

Azuma-Hoeffding Inequality. $d_i := \mathbb{E}(X|\mathcal{A}_i) - \mathbb{E}(X|\mathcal{A}_{i-1})$. For all $\lambda > 0$ we have

$$\mathbb{P}(|\sum_{i=1}^{n} d_i| > \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{n} ||d_i||_\infty^2}\right).$$

Proof. Fact 3 implies for all $i = 1, 2, ..., n$ that

$$\mathbb{E}(\exp(a d_i)|\mathcal{A}_{i-1}) \leq \exp(\frac{a^2}{2}||d_i||_\infty^2). \quad (*)$$

Azuma-Hoeffding Inequality. $d_i := \mathbb{E}(X|\mathcal{A}_i) - \mathbb{E}(X|\mathcal{A}_{i-1})$. For all $\lambda > 0$ we have

$$\mathbb{P}(|\sum_{i=1}^{n} d_i| > \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{n} ||d_i||_\infty^2}\right).$$

Proof. Fact 3 implies for all $i = 1, 2, ..., n$ that

$$\mathbb{E}(\exp(ad_i)|\mathcal{A}_{i-1}) \leq \exp(\frac{a^2}{2}||d_i||_\infty^2). \quad (*)$$

We have for all $t > 0$

$$\begin{aligned} \mathbb{E}\, e^{t\sum_{i=1}^{n} d_i} &= \mathbb{E}\,\mathbb{E}(e^{t\sum_{i=1}^{n} d_i}|\mathcal{A}_{n-1}) \\ &= \mathbb{E}\,[e^{t\sum_{i=1}^{n-1} d_i}\mathbb{E}(e^{td_n}|\mathcal{A}_{n-1})] \\ &\leq \mathbb{E}\,[e^{t\sum_{i=1}^{n-1} d_i}]e^{\frac{t^2}{2}||d_n||_\infty^2} \end{aligned}$$

Iterate (*) to get

$$\mathbb{E}\, e^{t\sum_{i=1}^{n} d_i} \leq e^{\frac{t^2}{2}\sum_{i=1}^{n}||d_i||_\infty^2}.$$

Thus

$$\mathbb{E}\left(\exp(t\sum_{i=1}^{n} d_i)\right) \le \exp\left(\frac{t^2}{2}\sum_{i=1}^{n} ||d_i||_\infty^2\right).$$

Markov's inequality for the map $x \mapsto \exp(tx)$ implies for all $t > 0$ we have

$$\mathbb{P}(|\sum_{i=1}^{n} d_i| > \lambda) \le \exp(-\lambda t)\exp\left(\frac{t^2}{2}\sum_{i=1}^{n} ||d_i||_\infty^2\right).$$

Let $t = \lambda(\sum_{i=1}^{n} ||d_i||_\infty^2)^{-1}$ to conclude the proof of Azuma-Hoeffding. □

Application to TSP tour length. Put $L(n) := L_{TSP}(U_1, ...U_n)$, with U_i i.i.d. on $[0,1]^d$. Put $\mathcal{A}_i := \sigma(U_1, ..., U_i)$ and

$$d_i := \mathbb{E}\,(L(n)|\mathcal{A}_i) - \mathbb{E}\,(L(n)|\mathcal{A}_{i-1}).$$

Rhee + Talagrand:

$$||d_i||_\infty \le C(d)(n-i+1)^{-1/d}, \quad d \ge 2.$$

Azuma-Hoeffding implies $(d = 2)$:

$$\mathbb{P}(|L(n) - \mathbb{E}\,L(n)| > t) \le 2\exp(-\frac{ct^2}{\log n})$$

So for $d = 2$

$$L(n) - \mathbb{E}\,L(n) = O(\log n) \quad a.s.$$

Rk. The $\log n$ factor can be removed (Talagrand, '95).

$L(n) := L_{TSP}(U_1, ...U_n)$, with U_i i.i.d. on $[0,1]^d$.

1. (Sub-Gaussian tail bounds for minimal matching) Do we have $(d = 2)$:

$$\mathbb{P}(|L_{MM}(U_1, ...U_n) - \mathbb{E}\, L_{MM}(U_1, ...U_n)| > t) \leq C \exp(-ct^2)?$$

2. (Central limit theorem for TSP) As $n \to \infty$, do we have

$$\frac{L(n) - \mathbb{E}\, L(n)}{\sqrt{\mathrm{Var} L(n)}} \xrightarrow{\mathcal{D}} N(0,1)?$$

3. (Central limit theorem for minimal matching) As $n \to \infty$, do we have

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

4. $\gamma_{TSP}(d) =?$

THANK YOU