Random Euclidean Bipartite Matching Problems

Dario Trevisan

UNIVERSITÀ DEGLI STUDI DI PISA

Master Classes en probabilités at IMRA
Strasbourg, January 20-24, 2019

Summary

Setting of the problem

Let $d \geq 1$, and consider $2n \geq 1$ i.i.d. random variables

$$(X_i)_{i=1}^n, \qquad (Y_i)_{i=1}^n$$

taking values in $[0,1]^d$ with uniform distribution, i.e.,

$$P(X_i \in A) = |A| = \int_A \mathrm{d}x, \quad \text{for every } A \subseteq [0,1]^d \text{ Borel},$$

The Random Euclidean Bipartite Matching Problem is defined as the following random variational problem,

$$\min_{\sigma \in \mathcal{S}^n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|,$$

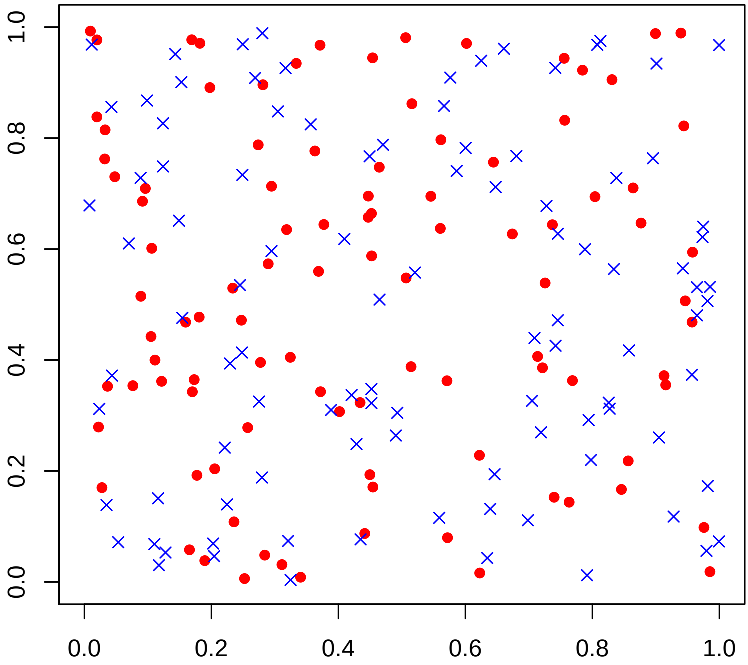
where $\mathcal{S}^n$ denotes the set of permutations over $\{1, \ldots, n\}$.

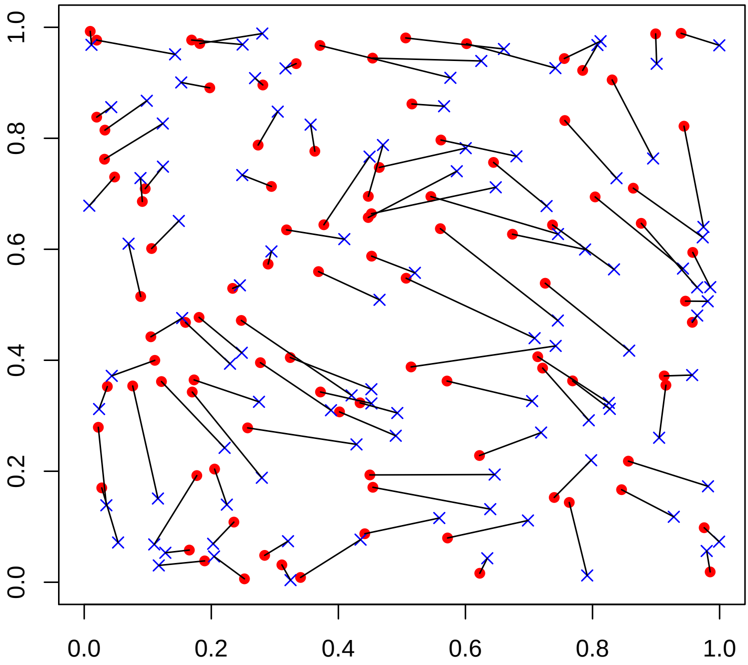
Variant: power of distance

$$\text{For } p \in (0, +\infty), \qquad B_{p,n} := \min_{\sigma \in \mathcal{S}^n} \sum_{i=1}^n |X_i - Y_{\sigma(i)}|^p.$$

Simulations

Let us look at a simulation with $n = 100$, $d = 2$, $p = 2$.


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

Related random combinatorial optimization problems

We can devise lots of variants:

1. Different domains
2. Different laws
3. Different distances (costs)
4. Different number of red/blue points: what is a matching?

Another related problem is the (monopartite) minimal matching problem:

Let $(X_i)_{i=1}^{2n}$ be i.i.d. uniform and define

$$M_{p,n} = \min_{\sigma \in \mathcal{S}^{2n}} \sum_{i=1}^{n} |X_{\sigma(i)} - X_{\sigma(n+i)}|^p,$$

that is, we forget about colouring of the points.

Exercise

Show that $\mathbb{E}[M_{p,n}] \leq \mathbb{E}[B_{p,n}]$.

Outline

Heuristics and a conjecture

Heuristics

Points are "uniformly distributed" on $[0,1]^d$
$\Rightarrow$ for each X_i we find Y_j at a distance $\approx n^{-1/d}$.

Considering the p-th power and summing upon n leads to

$$B_{p,n} \approx n \cdot \frac{1}{n^{p/d}} = n^{1-p/d}.$$

Conjecture

The limit (e.g. in probability)

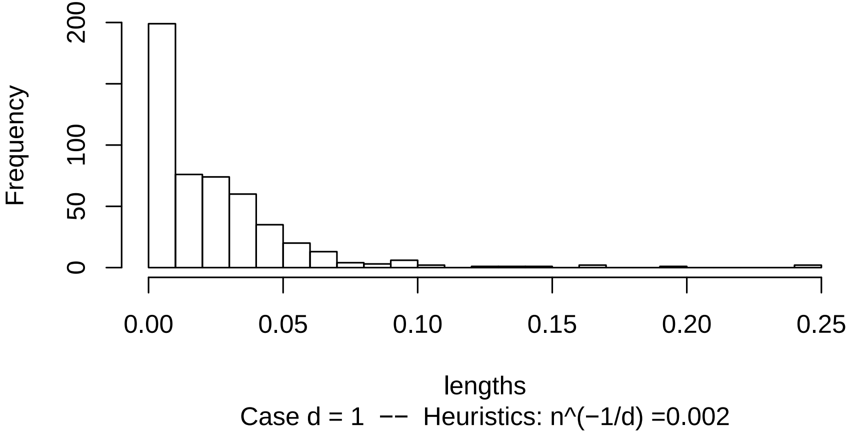
$$\lim_{n\to\infty} \frac{B_{p,n}}{n^{1-p/d}}$$

exists finite and strictly positive.

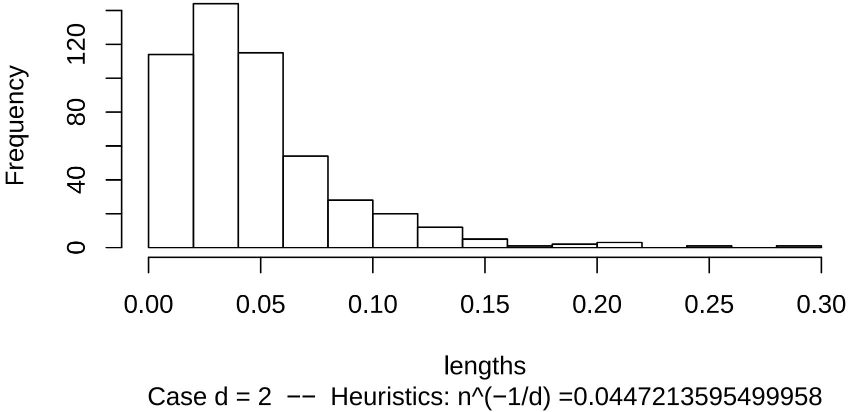
Simulations

Histograms of (simulatied) matching lengths with $p = 1$, $n = 500$ and $d \in \{1, 2, 3\}$.

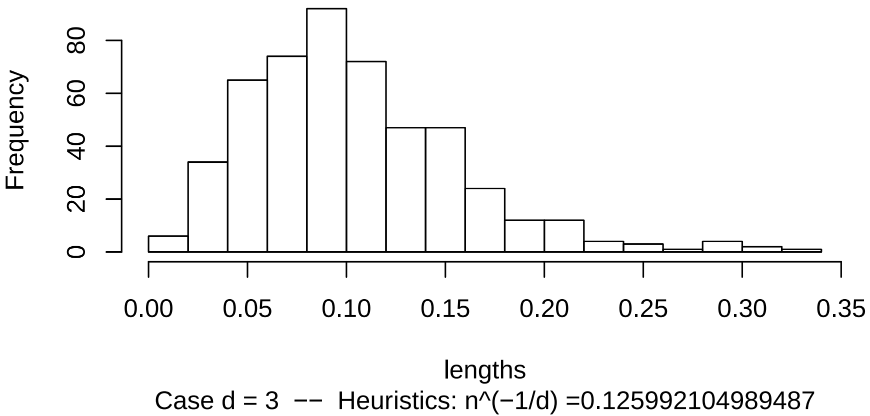
Simulations: $d = 1$, $p = 1$, $n = 500$



Simulations: $d = 2$, $p = 1$, $n = 500$



Simulations: $d = 3$, $p = 1$, $n = 500$



Outline

A rigorous lower bound

Proposition

There exists $c(d,p) > 0$ such that for every $n \geq 1$,

$$\mathbb{E}\left[B_{p,n}\right] \geq c(d,p)n^{1-p/d}.$$

We use the following lemma.

Lemma

Let $m \geq 1$ and $(X_i)_{i=1}^m$ i.i.d. on $[0,1]^d$ with uniform distribution. Then, for every $x \in [0,1]^d$,

$$\mathbb{E}\left[\min_{i=1,\ldots,m} |X_i - x|^p\right] \geq \frac{c(d,p)}{m^{p/d}},$$

where $c(d,p) > 0$ is depends on d and p only.

A rigorous lower bound

Proposition

There exists $c(d, p) > 0$ such that for every $n \geq 1$,

$$\mathbb{E}\left[B_{p,n}\right] \geq c(d, p) n^{1-p/d}.$$

We use the following lemma.

Lemma

Let $m \geq 1$ and $(X_i)_{i=1}^m$ i.i.d. on $[0, 1]^d$ with uniform distribution. Then, for every $x \in [0, 1]^d$,

$$\mathbb{E}\left[\min_{i=1,\ldots,m} |X_i - x|^p\right] \geq \frac{c(d, p)}{m^{p/d}},$$

where $c(d, p) > 0$ is depends on d and p only.

Proof of Lemma

Consider the survival (or reliability) function, i.e., for $t \in [0, \infty)$,

$$P(\min_{i=1,\ldots,m} |X_i - x|^p > t) = P(|X_1 - x| > t^{1/p})^m \quad \text{by independence}$$

$$= \mathcal{L}^d \left([0,1]^d \setminus \overline{B(x, t^{1/p})}\right)^m \quad \text{since the law is uniform}$$

$$\geq (\max\left\{(1 - \omega_d t^{d/p}), 0\right\})^m$$

with $\omega_d = |B(0,1)|$.

The general formula (for non-negative random variables Z)

$$\mathbb{E}[Z] = \int_0^\infty P(Z > t)\, \mathrm{d}t$$

gives

$$\mathbb{E}\left[\min_{i=1,\ldots,m} |X_i - x|^p\right] = \int_0^\infty P(\min_{i=1,\ldots,m} |X_i - x|^p > t)\, \mathrm{d}t$$

$$\geq \int_0^{\omega_d^{-p/d}} (1 - \omega_d t^{d/p})^m \, \mathrm{d}t = \omega_d^{-p/d} \int_0^1 (1 - u^{d/p})^m \, \mathrm{d}u$$

$$\geq \omega_d^{-p/d} \int_0^{m^{-p/d}} (1 - u^{d/p})^m \, \mathrm{d}u \geq \omega_d^{-p/d} \left(1 - \frac{1}{m}\right)^m m^{-p/d}.$$

Proof of lower bound

Given any $\sigma \in \mathcal{S}^n$,

$$|X_i - Y_{\sigma(i)}|^p \geq \min_{j=1,\dots,n} |X_i - Y_j|^p,$$

hence

$$B_{p,n} \geq \sum_{i=1}^{n} \min_{j=1,\dots,n} |X_i - Y_j|^p.$$

Taking expectation,

$$\begin{aligned}
\mathbb{E}\left[B_{p,n}\right] &\geq \mathbb{E}\left[\sum_{i=1}^{n} \min_{j=1,\dots,n} |X_i - Y_j|^p\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\min_{j=1,\dots,n} |X_i - Y_j|^p\right] \\
&= n\mathbb{E}\left[\min_{j=1,\dots,n} |X_1 - Y_j|^p\right] \quad \text{the law of } (X_i, Y_1, \dots, Y_n) \text{ does not depend on } i \\
&= n \int_{[0,1]^d} \mathbb{E}\left[\min_{j=1,\dots,n} |x - Y_j|^p\right] \mathrm{d}x \quad \text{by independence of } X_1 \text{ and } (Y_j)_j \\
&\geq n \cdot c(d,p) n^{-p/d} \quad \text{by Lemma 1.}
\end{aligned}$$

Outline

Concentration inequalities

After studying $\mathbb{E}[B_{p,n}]$, it is natural to consider $\mathrm{Var}(B_{p,n})$.

The general phenomenon of concentration of measure gives that $B_{p,n}$ is very close to its average because (as described by Talagrand)

the random variable depends in a “smooth” way on a large number of independent variables

Other (better known) examples are

- the law of large numbers, where $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_i]$
- Kolmogorov 0-1 laws, yielding e.g. that $\limsup_{n\to\infty} X_n$ is a.s. constant

Let E be a set. A function $f : E^m \to \mathbb{R}$ has bounded differences if

$$\text{for every } i \in \{1, \ldots, m\} \text{ there exists } d_i \geq 0 \text{ such that}$$
$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_m)| \leq d_i$$
$$\text{for every } x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m, x_i' \in E.$$

Lemma (Azuma-Hoeffding-McDiarmid)

Let $(E, \mathcal{E})$ be a measurable space, $f : E^m \to \mathbb{R}$ with bounded differences and and let $(X_i)_{i=1}^m$ be independent random variables with values in E. Then, $Z = f(X_1, \ldots, X_m)$ satisfies

$$P(|Z - \mathbb{E}[Z]| > r) \leq 2 \exp\left(-\frac{2r^2}{\sum_{i=1}^m d_i^2}\right) \quad \text{for every } r > 0.$$

Bounded difference as Lipschitz regularity

Consider the distance on E given by $\mathrm{d}(x, y) = 1$ if and only if $x \neq y$.

Then, the following are equivalent:

- $f : E^m \to \mathbb{R}$ has bounded differences
- for every $i = 1, \ldots, m$ and $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_m \in E$, the function

$$x_i \mapsto f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m)$$

  is Lipschitz.

Other concentration inequalities allow for Lipschitz regularity with respect to different distances (e.g., the Euclidean distance on E if $E \subseteq \mathbb{R}^d$).

Lipschitz functions are stable w.r.t. many operations, e.g. pointwise minima (with bounded constants).

Outline

Let $E = [0,1]^d$. For any $\sigma \in \mathcal{S}^n$ we argue that

$$f_\sigma(x_1, \dots, x_n, y_1, \dots, y_n) := \sum_{i=1}^{n} |x_i - y_{\sigma(i)}|^p$$

has bounded differences (with $d_i = 2d^{p/2}$). The function $\min_{\sigma \in \mathcal{S}^n} f_\sigma$ will also have bounded differences.

Let $i \in \{1, \dots, 2n\}$ and compute the difference of f_σ when evaluated at two points that may differ only for one component:

$$x_i \text{ (if } i \le n) \text{ or } y_{i-n} \text{ (if } i > n),$$

We find

$$\left| |x_i - y_{\sigma(i)}|^p - |x_i' - y_{\sigma(i)}|^p \right| \le 2d^{p/2}$$

in the case $i \le n$ and

$$\left| |x_{\sigma^{-1}(i-n)} - y_{i-n}|^p - |x_{\sigma^{-1}(i-n)} - y_{i-n}'|^p \right| \le 2d^{p/2}$$

in the case $i > n$.

We apply the concentration inequality with $m = 2n$ and the variables $(X_1, X_2, \dots, X_n, Y_1, \dots Y_n)$.

Proposition

For every $r > 0$, one has

$$P(|B_{p,n} - \mathbb{E}[B_{p,n}]| > r) \leq 2\exp\left(-\frac{r^2}{4d^p n}\right),$$

or equivalently (replacing r with $rn^{1-p/d}$),

$$P\left(\left|\frac{B_{p,n}}{n^{1-p/d}} - \frac{\mathbb{E}[B_{p,n}]}{n^{1-p/d}}\right| > r\right) \leq 2\exp\left(-\frac{r^2 n^{1-2p/d}}{4d^p}\right).$$

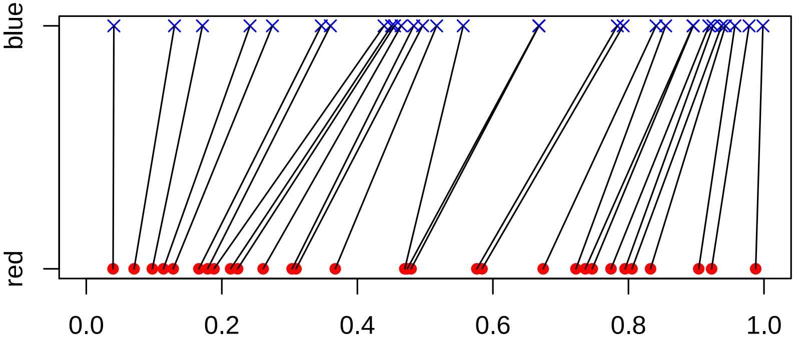
In the last inequality the right hand side is infinitesimal (summable) as $n \to \infty$ provided that $p < \frac{d}{2}$. Borel-Cantelli lemma yields:

$$\text{if } p < \frac{d}{2} \text{ and } \lim_{n\to\infty} \mathbb{E}[B_{p,n}]\, n^{1-p/d} = \beta_d \text{ exists (finite), then } B_{p,n} \to \beta_d \ P\text{-a.s.}$$

Outline

One-dimensional case

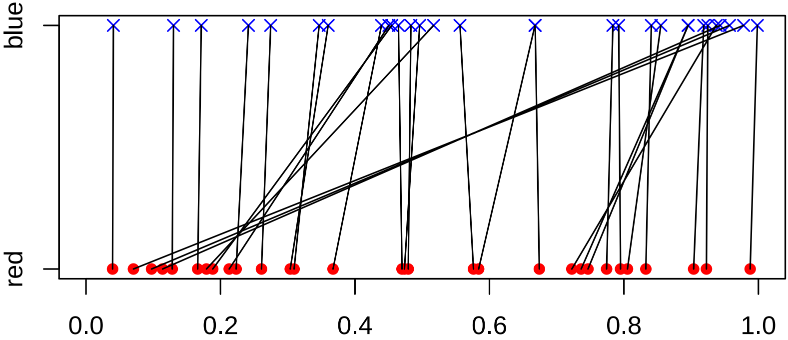
For $d = 1$ and $p \geq 1$, the bipartite matching admits a simple solution.



(Matching of $n = 30$ pairs of i.i.d. uniform points on the unit interval, $p = 1.1$)

One-dimensional case

This is not the case for $p < 1$:



(Matching of $n = 30$ pairs of i.i.d. uniform points on the unit interval, $p = 0.9$)

Outline

Order statistics

Given $n \geq 1$ (distinct) points $(z_i)_{i=1}^n$ in $\mathbb{R}$, we denote the *k-th smallest* value among them by $z_{(k)}$, e.g.

$$z_{(1)} = \min_i \{z_i\}, \quad z_{(n)} = \max_i \{z_i\},$$

and more generally, for $k \in \{1, \ldots, n-1\}$.

$$z_{(k+1)} = \min \left\{z_i : z_i > z_{(k)}\right\}.$$

Order statistics

When $z_i = X_i$ are random, write $X_{(k)}$: these are also known order statistics (of the sample).

Matching via order statistics

Proposition

Let $n \geq 1$ and let $(x_i)_{i=1}^n$, $(y_i)_{i=1}^n$ be distinct points in $\mathbb{R}$. Then there exists a unique $\sigma^\dagger \in \mathcal{S}^n$ such that

$$\left\{(x_i, y_{\sigma^\dagger(i)})\right\}_{i=1}^n = \left\{(x_{(i)}, y_{(i)})\right\}_{i=1}^n .$$

Moreover, for every $p \geq 1$,

$$\sum_{i=1}^n |x_i - y_{\sigma^\dagger(i)}|^p = \sum_{i=1}^n |x_{(i)} - y_{(i)}|^p = \min_{\sigma \in \mathcal{S}^n} \sum_{i=1}^n |x_i - y_{\sigma(i)}|^p.$$

Proof by induction using the inequality (which is just the case $n = 2$): for any $x < x'$, $y < y'$, then

$$|x - y|^p + |x' - y'|^p \leq |x - y'|^p + |x' - y|^p.$$

Outline

Law of order statistics

Lemma

Let $(Z_i)_{i=1}^n$ be i.i.d. random variables uniformly distributed on $[0,1]$. Then for every $t \in [0,1]$, $k > 1$,

$$P(Z_{(k)} > t > Z_{(k-1)}) = \binom{n}{k-1} t^{k-1}(1-t)^{n-k+1},$$

so that $Z_{(k)}$ has density $\mathrm{Beta}(k, n+1-k)$

$$f_k(t) = \frac{n!}{(k-1)!(n-k)!} t^{k-1}(1-t)^{n-k}.$$

A closed formula

Combining all these facts, we find the following result

Theorem

For $n \geq 1$,

$$\mathbb{E}[B_{2,n}] = \frac{1}{3}\frac{n}{n+1}.$$

In particular,

$$\lim_{n\to\infty} \mathbb{E}[B_{2,n}] = \frac{1}{3}.$$

This shows that the general lower bound that reads in this case

$$\mathbb{E}[B_{2,n}] \geq c(d,2)n^{-1},$$

misses in fact the correct order.

As an exercise, try to argue similarly for $B_{4,n}$.

Outline

Self-Averaging

We describe an approach (due to various authors) to prove existence of

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[B_{p,n}\right]}{n^{1-p/d}} \in (0,\infty),$$

provided that

$$0 < p < \frac{d}{2}$$

Use self-similarity of cube $[0,1]^d$:

1. Decompose it into smaller cubes
2. Find a *local matching* on each small cube
3. *Glue* them together to find a matching on the initial cube.

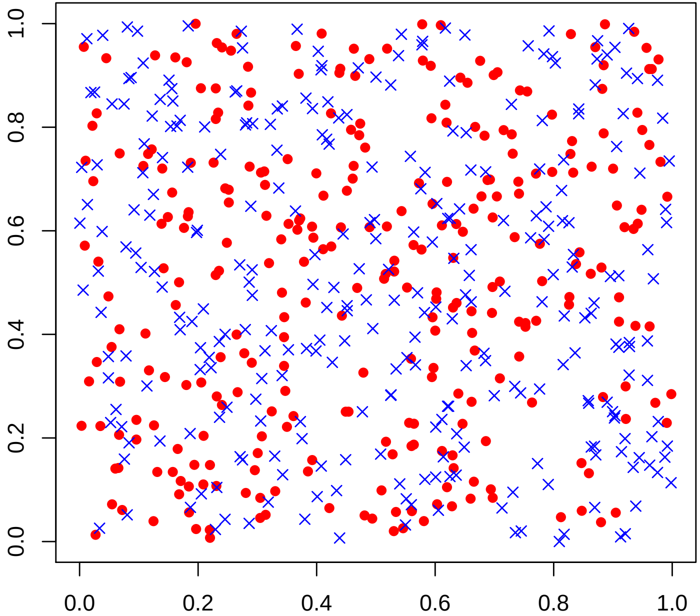
Good: cheaper to compute, one can iterate at multiple scales

Bad: on smaller cubes the number of red/blue points differ, the matching will not be globally optimal

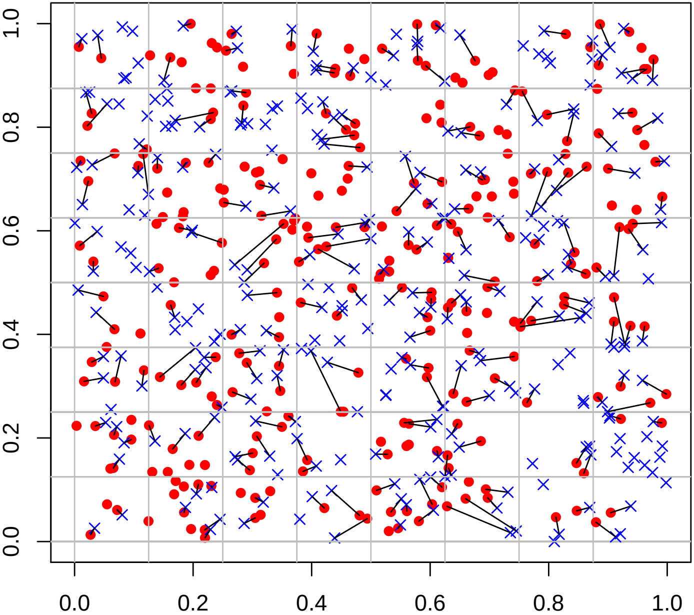
Simulations

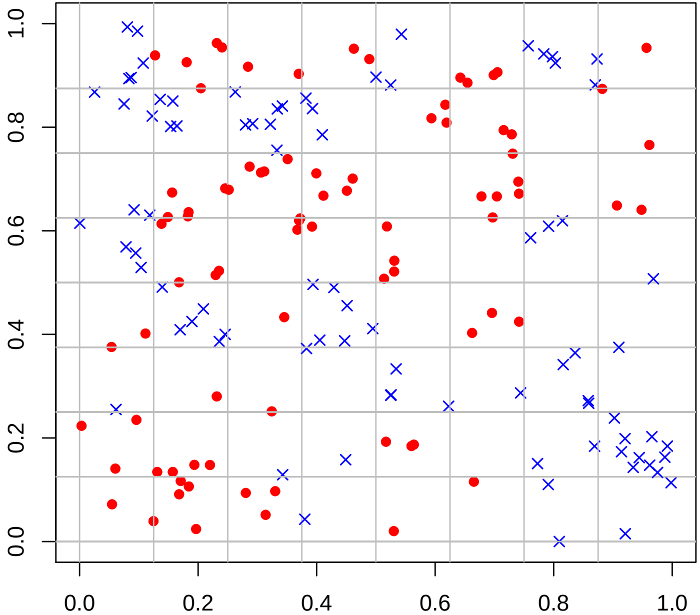
Simulating $n = 300$ pair of points:

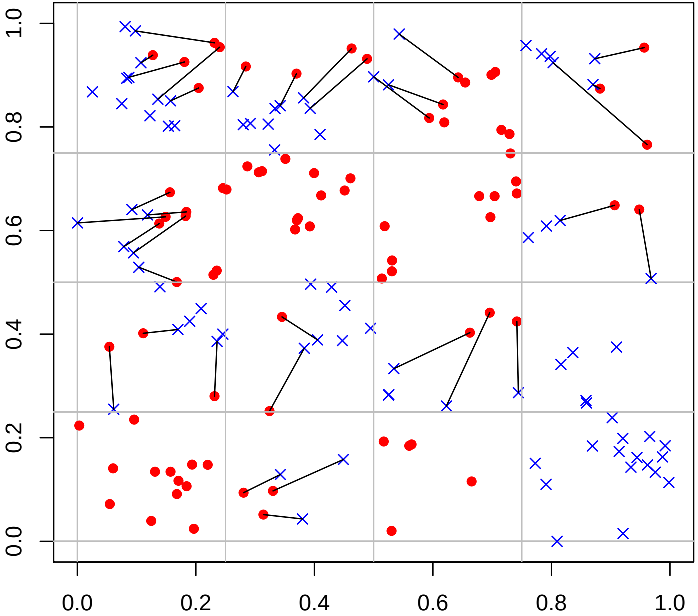
gluing local matchings at scales 2^{-k} with $k = 3, 2, 1, 0$ ($p = 1$)



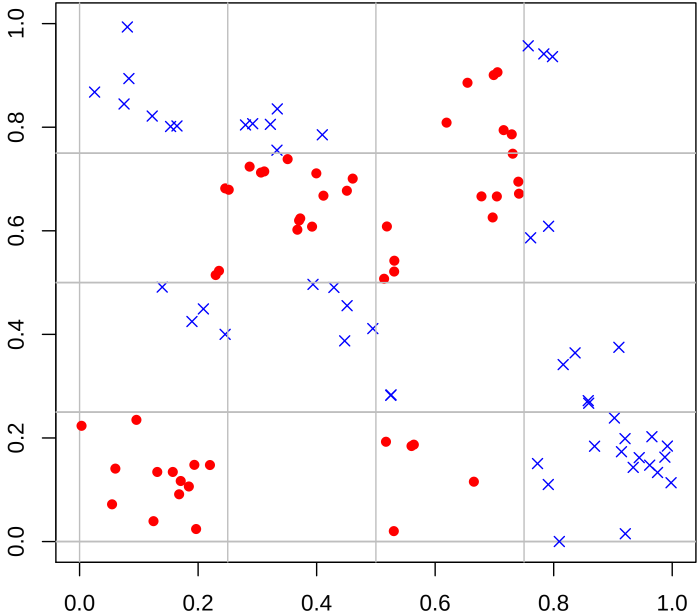
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

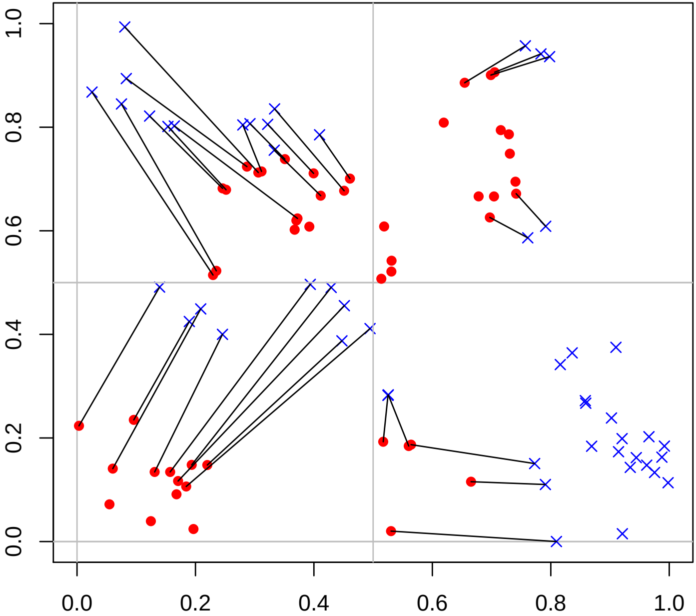

0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

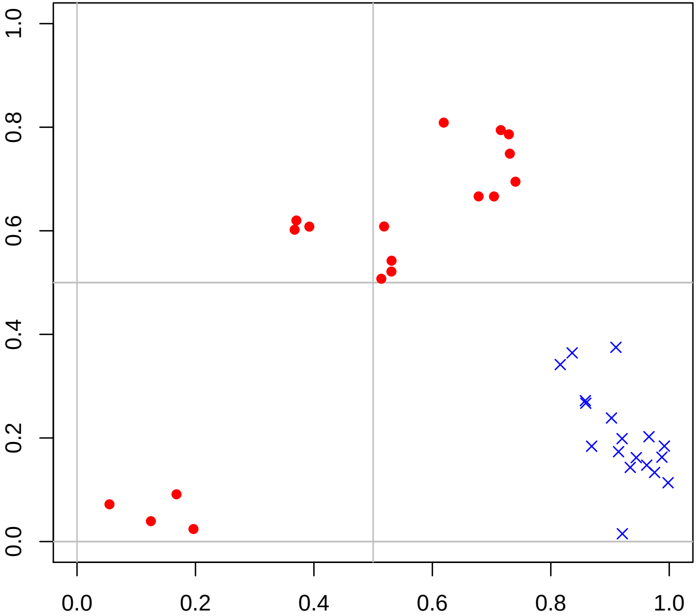


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

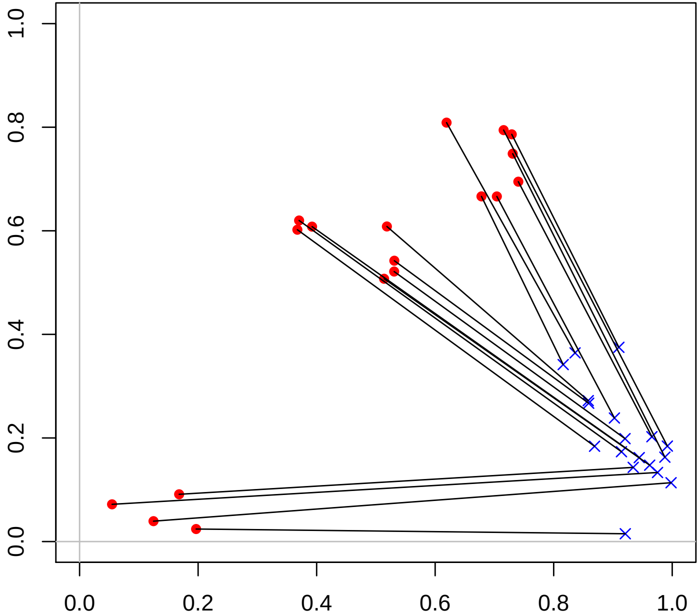

0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0



0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0



0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0



0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

Local matchings: notation

Given $n \in \mathbb{N}$, $x = (x_i)_{i=1}^n \subseteq \mathbb{R}^d$, write

$$\mathcal{I}(x; U) := \{i \in \{1, \dots, n\} \,:\, x_i \in U)\},$$

$$n(x; U) := \sharp\mathcal{I}(x; U).$$

For $m \in \mathbb{N}$, $y = (y_i)_{i=1}^m \subseteq \mathbb{R}^d$, a "local" matching in U is a relation $\sigma \subseteq \mathcal{I}(x; U) \times \mathcal{I}(y; U)$ injective functions from the set with least elements:

- if $n(x; U) \le n(y; U)$, then $\sigma : \mathcal{I}(x; U) \to \mathcal{I}(y; U)$ injective
- if $n(x; U) \ge n(y; U)$, then $\sigma : \mathcal{I}(y; U) \to \mathcal{I}(y; U)$ injective.

Let $\mathcal{S}(x, y; U)$ be set of "local" matchings in U and define the local cost

$$b_p(x, y; U) := \min_{\sigma \in \mathcal{S}(x,y;U)} \sum_{(i,j) \in \sigma} |x_i - y_j|^p.$$

Outline

Lemma

Let $n, m \in \mathbb{N}$ $x = (x_i)_{i=1}^n$, $y = (y_i)_{i=1}^m \subseteq \mathbb{R}^d$, $U \subseteq \mathbb{R}^d$ *Borel. Then,*

1. b_p *is local, i.e.,* $b_p(x, y; U) = b_p((x_i)_{i \in \mathcal{I}(x;U)}, (y_i)_{i \in \mathcal{I}(y;U)}, U)$.
2. b_p *is translation invariant: if* $v \in \mathbb{R}^d$*, then*

$$b_p(x + v, y + v; U + v) = b_p(x, y; U),$$

where $x + v = (x_i + v)_{i=1}^n$, $y + v = (y_i + v)_{i=1}^m$, $U + v = \{u + v : u \in U\}$.

3. b_p *is p-homogeneous: if* $\lambda \in (0, \infty)$*, then*

$$b_p(\lambda x, \lambda y, \lambda U) = \lambda^p b_p(x, y; U),$$

where $\lambda x = (\lambda x_i)_{i=1}^n$, $\lambda y = (\lambda y_i)_{i=1}^m$, $\lambda U = \{\lambda u : u \in U\}$.

4. b_p *is p-subadditive: for every (Borel) partition* $U = \bigcup_{k=1}^K U_k$*, one has*

$$b_p(x, y; U) \leq \sum_{k=1}^K b_p(x, y; U_k) + |n(x; U_k) - n(y; U_k)| \operatorname{diam}(U)^p,$$

where $\operatorname{diam}(U) = \sup\{|x - y| \ : \ x, y \in U)\}$.

Outline

Poissonization

Back to the random case: let $X = (X_i)_{i=1}^n$, $Y = (Y_i)_{i=1}^n$ be i.i.d. uniform on $[0,1]^d$. Then

$$n(X; U) = \sum_{i=1}^{n} I_{\{X_i \in U\}}$$

has law $\mathrm{Bin}(n, |U|)$. To obtain a self-similar model, we pick n to be also random, with Poisson law, with intensity λ, i.e., we impose that

$$P(n(X; [0,1]^d) = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{for } k \in \mathbb{N}.$$

Rigorously: fix $\lambda > 0$, let N_X be a Poisson random variables with intensity λ. Let also $(X_i)_{i=1}^{\infty}$ be i.i.d. uniform random variables in $[0,1]^d$, also independent of N_X. We define a Poisson point processes of intensity λ in $[0,1]^d$ as

$$X := (X_i)_{i=1}^{N_X}$$

(as a random variable with values in $\bigcup_{i=1}^{\infty} \mathbb{R}^{d\times i}$).

Conditionally to $\{N_X = k\}$, then $X = (X_i)_{i=1}^{k}$ i.i.d. on $[0,1]^d$ uniformly.

We consider also an independent Poisson point process of same intensity λ

$$Y = (Y_i)_{i=1}^{N_Y}.$$

Lemma

Let $\lambda > 0$ and let $X = (X_i)_{i=1}^{N_X}$ be a Poisson point process of intensity λ on $[0,1]^d$. For every $U \subseteq [0,1]^d$ Borel with $|U| > 0$, then

- the random variable
$$n(X;U) = \sharp\mathcal{I}(X;U)$$
has Poisson law of parameter $\lambda|U|$,
- conditionally to $\{n(X;U) = k\}$, the k random variables $(X_i)_{i\in\mathcal{I}(X;U)}$ are independent with uniform law on U.

Corollary

Let $Q = [a,b]^d \subseteq [0,1]^d$. Then the process

$$((X_i - a)/(b-a))_{i\in\mathcal{I}(X;Q)},$$

is a Poisson point process on $[0,1]^d$ of intensity $\lambda(b-a)^d$.

A monotonicity lemma

When we glue together local matching, we will not find exact monotonicity but an approximate one.

Lemma

Let $\alpha > 0$, $c \geq 0$, $f : [1, \infty) \to [0, \infty)$ *be continuous and such that, for every* $m \geq 1$ *integer,* $\lambda \geq 1$ *real,*

$$f(m\lambda) \leq f(\lambda) + c\lambda^{-\alpha}.$$

Then $\lim_{\lambda \to \infty} f(\lambda)$ *exists.*

Outline

A convergence result

Theorem

Let $p \in (0, d/2)$, let X_λ, Y_λ be independent Poisson point processes on $[0,1]^d$ with intensity λ. Then

$$\lim_{\lambda \to \infty} \frac{\mathbb{E}\left[b_p(X_\lambda, Y_\lambda, [0,1]^d)\right]}{\lambda^{1-p/d}} \in (0, \infty)$$

exists.

Asymptotics in $d = 2$, $p = 1$

We describe a recent approach to the case $d = 2$: one can show e.g., that

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[B_{2,n}\right]}{\log(n)} = \frac{1}{2\pi}.$$

Again, the rate $\log(n)$ differs from the heuristic $n^{1-p/d} = n^0$. The following problem is open:

$$\text{does} \quad \lim_{n\to\infty} \frac{\mathbb{E}\left[B_{1,n}\right]}{\sqrt{n\log(n)}} \quad \text{exist?}$$

We focus on the simpler bound, for $p = 1$,

$$\limsup_{n\to\infty} \frac{\mathbb{E}\left[B_{1,n}\right]}{\sqrt{n\log(n)}} < \infty.$$

Outline

Optimal transport

The main ideas comes from a more general theory of optimal transport, where one looks for efficient couplings between probability laws.

Matchings are particular (extreme) cases of couplings.

Analytical tools: duality formula

$$B_{1,n} = \max_{f \in \mathrm{Lip}(1)} \left\{ \sum_{i=1}^{n} f(X_i) - f(Y_i) \right\}$$

where Lip(1) denotes the set of $f : [0,1]^d \to \mathbb{R}$ such that $|f(x) - f(y)| \leq |x - y|$ for every $x, y \in [0,1]^d$.

Duality for general p

$$B_{p,n} = \max \left\{ \sum_{i=1}^{n} f(X_i) - g(Y_i) \ : \ f(x) - g(y) \le |x-y|^p \quad \forall x, y \right\}$$

1. $(Y_i)_{i=1}^n$ are factories (suppliers)
2. $(X_i)_{i=1}^n$ are buyers (cities)
3. $|x-y|^p$ is the cost to transport a unit of good from x to y
4. A company wants to take care of distribution of goods: buys at price $g(y)$ and sells at price $f(x)$.
5. It is competitive if $f(x) - g(y) \le |x-y|^p$.
6. The overall profit is
$$\sum_{i=1}^{n} f(X_i) - g(Y_i)$$
7. Duality reads max profit = min cost

Outline

A random elliptic PDE

We solve the random elliptic partial differential equation (PDE)

$$-\Delta u = \sum_{i=1}^{n} (\delta_{X_i} - \delta_{Y_i}) \quad \text{in } [0,1]^2,$$

where δ_x means the Dirac measure at the point x and use duality (for $p = 1$)

$$\begin{aligned}
\sum_{i=1}^{n} f(X_i) - f(Y_i) &= \int_{[0,1]^2} f \, \mathrm{d}\left(\sum_{i=1}^{n} (\delta_{X_i} - \delta_{Y_i}) \right) \\
&= -\int_{[0,1]^2} f(x) \Delta u(x) \, \mathrm{d}x \\
&= \int_{[0,1]^2} \nabla f(x) \nabla u(x) \, \mathrm{d}x \leq \int_{[0,1]^2} |\nabla u(x)| \, \mathrm{d}x
\end{aligned}$$

Taking expectation, we obtain the inequality

$$\mathbb{E}\left[B_{1,n}\right] \leq \mathbb{E}\left[\int_{[0,1]^2} |\nabla u(x)| \, \mathrm{d}x \right] \leq \mathbb{E}\left[\int_{[0,1]^2} |\nabla u(x)|^2 \, \mathrm{d}x \right]^{1/2}.$$

Outline

The advantage of the PDE with respect to the matching problem is linearity: explicit formulas e.g. via Fourier series:

$$u(x_1, x_2) = 4 \sum_{m \in \mathbb{N}^2} a_m \cos(m_1 \pi x_1) \cos(m_2 \pi x_2),$$

where $m = (m_1, m_2)$, $|m| = \left(m_1^2 + m_2^2\right)^{1/2}$, $a_{(0,0)} = 0$ and

$$a_m = \frac{4}{\pi^2 |m|^2} \sum_{i=1}^{n} \cos(m_1 \pi X_{i,1}) \cos(m_2 \pi X_{i,2}) - \cos(m_1 \pi Y_{i,1}) \cos(m_2 \pi Y_{i,2}).$$

Fourier coefficients are random variables, with mean 0 and variance

$$\mathbb{E}\left[a_m^2\right] = \frac{2n}{\left(\pi^2 |m|^2\right)^2}.$$

Energy estimate

Recall

$$\mathbb{E}\left[a_m^2\right] = \frac{2n}{\left(\pi^2 |m|^2\right)^2}.$$

Plancherel identity gives

$$\int_{[0,1]^2} |\nabla u(x)|^2 \, dx = \sum_{m \in \mathbb{N}^2} \pi^2 |m|^2 a_m^2.$$

Taking expectation we obtain the series

$$\mathbb{E}\left[\int_{[0,1]^2} |\nabla u(x)|^2 \, dx\right] = \sum_{m \in \mathbb{N}^2} \pi^2 |m|^2 \, \mathbb{E}\left[a_m^2\right]$$
$$= \frac{2n}{\pi^2} \sum_{\substack{m \in \mathbb{N}^2 \\ m \neq (0,0)}} \frac{1}{|m|^2}$$

$$\approx \frac{2n}{\pi^2} \int_1^\infty \frac{1}{r} \, dr \frac{\pi}{2} = \infty!$$

Renormalization

However, partial sums

$$\sum_{1 \leq |m| \leq a} \frac{1}{|m|^2} \approx \frac{\pi}{2} \int_1^a r \, \mathrm{d}r = \frac{\pi}{2} \log(a)$$

CLAIM: Only frequencies m with $|m| \leq n$ matter $\Rightarrow$

$$\mathbb{E}\left[B_{1,n}\right] \leq c(d)\sqrt{n \log(n)}.$$

Difficult to make rigorous!