Asymptotic distribution of the diameter of a random elliptical cloud

Yann Demichel

Joint work with A.-K. Fermin and P. Soulier

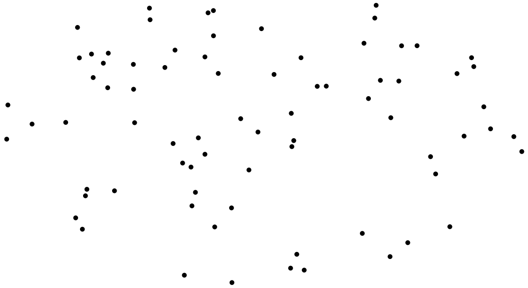
université Paris Ouest Nanterre La Défense

GDR GeoSto

Menu of the day

Random cloud

- $\mathbb{X}$ is a random vector in $\mathbb{R}^d$ with $d \geq 1$ fixed
- $n \geq 1$ is the size of the cloud
- $\{\mathbb{X}_i\}_{1 \leq i \leq n}$ are independent vectors distributed as $\mathbb{X}$

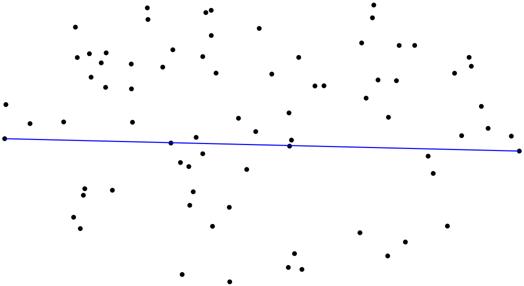


Menu of the day

Diameter of the cloud

- $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$

$$\boxed{D_n := \max_{1 \le i < j \le n} \|\mathbb{X}_i - \mathbb{X}_j\|}$$



Menu of the day

What is the asymptotic distribution of D_n when $n \to \infty$?

Answer ?

- For special cases
- Depends on the distribution of $\mathbb{X}$

Dichotomy

- Distributions supported by a bounded set
  - Distributions 'approximatively' uniform
  - Geometry of the support
- Distributions supported by an unbounded set
  - Spherically symmetric distributions

History : bounded support

- Uniform distribution supported by special planar sets (excluding balls or ellipsoids) : *Appel, Najim and Russo* (2002)
- Distributions with support included in the unit d-ball (including uniform in the d-ball, in the d-sphere, in spherical sectors) : *Mayer and Molchanov* (2007)
- Distributions supported by a polytope (included uniform or non-uniform in square, uniform in regular polygons, uniform in the unit d-cube) : *Lao* (2010)
- Distributions supported by a d-ellipsoid : *Schrempp* (2016)

Menu of the day

History : unbounded support

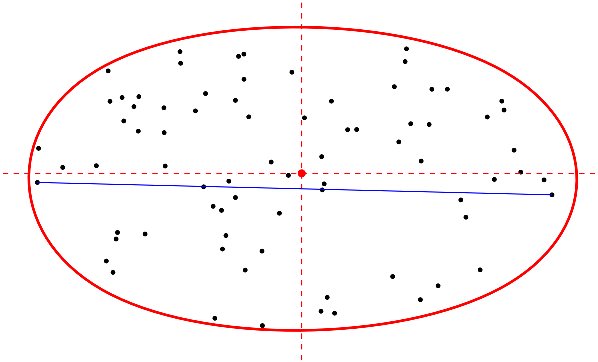
- Spherically symmetric normal distribution : *Matthews and Rukhin* (1993)
- Spherically symmetric Kotz distribution : *Henze and Klein* (1996)
- Power-tailed spherically decomposable distributions : *Henze and Lao* (2010)
- Spherically symmetric distributions : *Jammalamadaka and Janson* (2015)

  $\rightarrow$ Open question : elliptically symmetric distributions ?

Apetizer

A naive question ...

Where are the points which can achieve the diameter ?



Apetizer

... a naive answer

$$M_n := \max_{1 \le i \le n} \|\mathbb{X}_i\|$$

D_n is achieved for a pair of diametrically opposed points each of them realizing M_n

If you believe in this, you need :

- To localize the vectors with large norms
- To control the asymptotic distribution of M_n

Precisely :

- Distribution of $\|\mathbb{X}\|$?
- Distribution of $\frac{1}{\|\mathbb{X}\|}\mathbb{X}$ conditional on $\|\mathbb{X}\|$ is large ?

Today's ingredients

$$\boxed{\text{Elliptical distribution : } \mathbb{X} = R\Lambda\mathbb{U}}$$

where

- $\mathbb{U} = (U_1, \ldots, U_d)$ is uniform on the unit sphere $\mathcal{S}^{d-1}$
- Λ is an invertible $d \times d$ matrix
- R is a positive random variable independent of $\mathbb{U}$

In addition :

$$\boxed{R \text{ is in the max domain of attraction of the Gumbel distribution}}$$

Today's ingredients

R is in the max domain of attraction of the Gumbel distribution

There exists a differentiable function $\psi_R : (0, \infty) \to (0, \infty)$ such that

$$\lim_{x \to \infty} \frac{\mathbb{P}(R > x + t\psi_R(x))}{\mathbb{P}(R > x)} = \mathrm{e}^{-t}$$

locally uniformly with respect to $t \in \mathbb{R}$

Such a function ψ_R satisfies :

$$\lim_{x \to \infty} \frac{\psi_R(x + t\psi_R(x))}{\psi_R(x)} = 1 \; ; \; \lim_{x \to \infty} \psi_R'(x) = 0 \; ; \; \lim_{x \to \infty} \frac{\psi_R(x)}{x} = 0$$

Today's ingredients

Distribution of $\Lambda\mathbb{U}$

Supported by the ellipsoid $\{\Sigma u : u \in \mathcal{S}^{d-1}\}$ where

$$\Sigma := \Lambda'\Lambda$$

is (up to a constant) the covariance matrix of $\mathbb{X}$

The ellipsoid is centered at the origin and has d axes directed by the eigenvectors of Σ with semi-length the square roots of the corresponding eigenvalues

$$\lambda_1 = \cdots = \lambda_{\mathbf{m}} > \lambda_{\mathbf{m}+1} \geq \cdots \geq \lambda_d > 0$$

ordered and repeated, where $1 \leq \mathbf{m} \leq d$ is the multiplicity of the largest one. If $\mathbf{m} = d$ we have a spherical distribution.

Today's ingredients

Distribution of $\Lambda\mathbb{U}$

Up to an orthogonal transformation we may assume that

$$\Lambda = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_d})$$

Starter

Localization principle for R

If $\mathbb{V}$ is a bounded random variable then the vector $R\mathbb{V}$ has a large norm iif R is large and $\mathbb{V}$ is close to its maximum.

Therefore, when $\mathbb{X} = R\Lambda\mathbb{U}$ is large then $\|\mathbb{X}\|$ is of order $\sqrt{\lambda_1}R$ and $\mathbb{X}$ is located near the dominant eigenspace associated with λ_1 :

$$\|\mathbb{X}\| = \sqrt{\lambda_1}R\left(1 - \sum_{k=\mathbf{m}+1}^{d} \frac{\lambda_1 - \lambda_k}{\lambda_1} U_k^2\right)^{1/2}$$

Starter

Theorem [FDS, 2015]

Define the functions ψ and ϕ on $(0, \infty)$ by

$$\psi(x) = \sqrt{\lambda_1}\psi_R\Big(\frac{x}{\sqrt{\lambda_1}}\Big) \quad \text{and} \quad \phi(x) = \Big(\frac{\psi(x)}{x}\Big)^{1/2}$$

Then, as $x \to \infty$,

$$\mathbb{P}(\|\mathbb{X}\| > x) \sim C_{\mathbf{m}} \left(\phi(x)\right)^{d-\mathbf{m}} \mathbb{P}\Big(R > \frac{x}{\sqrt{\lambda_1}}\Big)$$

where

$$C_{\mathbf{m}} := \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{\mathbf{m}}{2})} 2^{(d-\mathbf{m})/2} \Bigg(\prod_{k=\mathbf{m}+1}^{d} \frac{\lambda_1}{\lambda_1 - \lambda_k} \Bigg)^{1/2}$$

In particular, $\|\mathbb{X}\|$ is also in the max domain of attraction of the Gumbel distribution

Starter

Theorem [FDS, 2015]

Define $\Theta = \frac{1}{\|\mathbb{X}\|}\mathbb{X} = (\Theta_1, \ldots, \Theta_d)$.

Then, as $x \to \infty$, conditionally on $\|\mathbb{X}\| > x$,

$$\left(\frac{\|\mathbb{X}\| - x}{\psi(x)}, \Theta_1, \ldots, \Theta_{\mathbf{m}}, \frac{\Theta_{\mathbf{m}+1}}{\phi(x)}, \ldots, \frac{\Theta_d}{\phi(x)} \right)$$

converges in distribution to

$$\left(E, \Theta^{(\mathbf{m})}, \sqrt{\tfrac{\lambda_{\mathbf{m}+1}}{\lambda_1 - \lambda_{\mathbf{m}+1}}} G_{\mathbf{m}+1}, \ldots, \sqrt{\tfrac{\lambda_d}{\lambda_1 - \lambda_d}} G_d \right)$$

where E is an exponential random variable with mean 1, $\Theta^{(\mathbf{m})}$ is uniformly distributed on $\mathcal{S}^{\mathbf{m}-1}$, $G_{\mathbf{m}+1}, \ldots, G_d$ are independent standard Gaussian random variables, and all components are independent.

Starter

$\|\mathbb{X}\|$ is in the max domain of attraction of the Gumbel distribution

Thus :

- Consider $a_n > 0$ such that $\mathbb{P}(\|\mathbb{X}\| > a_n) \sim \frac{1}{n}$
- Set $b_n = \psi(a_n)$

Then :

- $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} \frac{b_n}{a_n} = 0$
- For all $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - a_n}{b_n} \le t\right) = \mathrm{e}^{-\mathrm{e}^{-t}}$$

Starter

Corollary [FDS, 2015]

Let $c_n = \phi(a_n)$ and define the points

$$P_{n,i} = \left(\frac{\|\mathbb{X}_i\| - a_n}{b_n}, \Theta_{i,1}, \ldots, \Theta_{i,\mathbf{m}}, \frac{\Theta_{i,\mathbf{m}+1}}{c_n}, \ldots, \frac{\Theta_{i,d}}{c_n} \right)$$

Then, the point processes $\sum_{i=1}^{n} \delta_{P_{n,i}}$ converge weakly to a PPP $\sum_{i=1}^{\infty} \delta_{P_i}$ on $\mathbb{R} \times \mathcal{S}^{\mathbf{m}-1} \times \mathbb{R}^{d-\mathbf{m}}$ with

$$P_i = \left(\Gamma_i, \Theta_i^{(\mathbf{m})}, \sqrt{\frac{\lambda_{\mathbf{m}+1}}{\lambda_1 - \lambda_{\mathbf{m}+1}}} G_{i,\mathbf{m}+1}, \ldots, \sqrt{\frac{\lambda_d}{\lambda_1 - \lambda_d}} G_{i,d} \right)$$

where $\{\Gamma_i\}$ are the points of a PPP on $(-\infty, \infty]$ with mean measure $\mathrm{e}^{-t}\mathrm{d}t$, $\{\Theta_i^{(\mathbf{m})}\}$ are i.i.d. vectors uniformly distributed on $\mathcal{S}^{\mathbf{m}-1}$ and $\{G_{i,k}\}$ are i.i.d. standard Gaussian variables, all sequences being mutually independent.

Conclusion

Vectors $\mathbb{X}_i = \|\mathbb{X}_i\|\Theta_i$ with the largest norm concentrate around the dominant eigenspace in such a way that

- $\|\mathbb{X}_i\| \sim a_n + b_n\Gamma_i$ with $a_n \to \infty$ and $b_n = o(a_n)$
- The $\mathbf{m}$ first coordinates of Θ_i are uniform on $\mathcal{S}^{\mathbf{m}-1}$
- The $d - \mathbf{m}$ other coordinates of Θ_i tend to 0 with rate $c_n \to 0$ with Gaussian fluctuations

Main course

Last question

Are these large vectors always diametrically opposed ?

- If $\mathbf{m} = 1$: $\mathcal{S}^{\mathbf{m}-1}$ has only one direction
  Thus two vectors with a large norm will be on opposite sides and their distance is automatically large, typically twice as large as the norm of each one.
  We expect that D_n behaves roughly like $2a_n$

- If $\mathbf{m} > 1$: $\mathcal{S}^{\mathbf{m}-1}$ has an infinite number of directions
  Thus two vectors with a large norm can be close to each other and their distance will be typically much smaller than twice their norm.
  We expect then a corrective term when comparing D_n to $2a_n$

Main course

Theorem [FDS, 2015]

Assume that $\mathbf{m} = 1$, i.e. $\lambda_1 > \lambda_2$.
Then

$$\frac{D_n - 2a_n}{b_n} \xrightarrow{\text{(d)}} \max_{i,j\geq 1} \left\{ \Gamma_i^+ + \Gamma_j^- - \frac{1}{4} \sum_{k=2}^{d} \frac{\lambda_k}{\lambda_1 - \lambda_k} (G_{i,k}^+ + G_{j,k}^-)^2 \right\}$$

where $\{\Gamma_i^\pm\}$ are the points of a PPP with mean measure $\frac{1}{2}\mathrm{e}^{-t}\mathrm{d}t$ on $\mathbb{R}$, and $\{G_{i,k}^\pm\}$ are i.i.d. standard Gaussian variables, independent of the points $\{\Gamma_i^\pm\}$.

Main course

Theorem [FDS, 2015]

Assume that $\mathbf{m} > 1$.
Then, for all $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{D_n - 2a_n}{b_n} + d_n \leq t \right) = \mathrm{e}^{-\mathrm{e}^{-t}}$$

where

$$d_n = \frac{\mathbf{m}-1}{2} \log \frac{a_n}{b_n} - \log\log \frac{a_n}{b_n} - \log C'_{\mathbf{m}}$$

with

$$C'_{\mathbf{m}} = (2d - \mathbf{m} - 1) 2^{\mathbf{m}-4} \pi^{-1/2} \Gamma\left(\frac{\mathbf{m}}{2}\right) \left(\prod_{k=\mathbf{m}+1}^{d} \frac{\lambda_1}{\lambda_1 - \lambda_k} \right)^{-1/2}$$

Main course

Example : bivariate Gaussian variable with correlation $\rho \in (0, 1)$

$$\mathbb{X} = R\Lambda\mathbb{U} \quad \text{with} \quad R = \sqrt{\chi_2^2} \quad \text{and} \quad \Lambda'\Lambda = \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

- Eigenvalues : $\lambda_1 = 1 + \rho$ and $\lambda_2 = 1 - \rho$
- Eigenspaces : $\mathrm{span}\{(1, 1)\}$ and $\mathrm{span}\{(-1, 1)\}$
- Multiplicity : $\mathbf{m} = \begin{cases} 1 & \text{if } \rho \neq 0 \\ 2 & \text{if } \rho = 0 \end{cases}$

Main course

Set $a_n = \sqrt{(1+\rho)\log n}$ and $b_n = \sqrt{\frac{1+\rho}{2\log n}}$

- If $\rho \neq 0$ then

$$\frac{D_n - 2a_n}{b_n} \xrightarrow{(\mathrm{d})} \max_{i,j\geq 1}\left\{\Gamma_i^+ + \Gamma_j^- - \frac{1-\rho}{8\rho}(G_i^+ + G_j^-)^2\right\}$$

  where $\{\Gamma_i^\pm\}$ are the points of a PPP with mean measure $\frac{1}{2}\mathrm{e}^{-t}\mathrm{d}t$ on $\mathbb{R}$, and $\{G_i^\pm\}$ are i.i.d. standard Gaussian variables, independent of the points $\{\Gamma_i^\pm\}$.

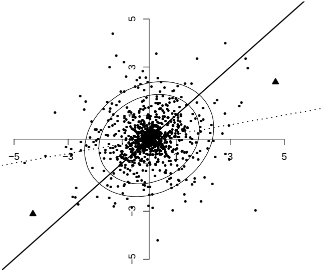
- If $\rho = 0$ then for all $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{D_n - 2a_n}{b_n} + d_n \leq t\right) = \mathrm{e}^{-\mathrm{e}^{-t}}$$

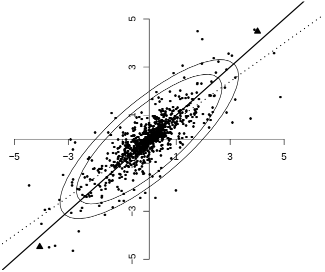
  where

$$d_n = \tfrac{1}{2}\log\log n - \log\log\log n + \log(4\sqrt{2\pi})$$

Main course



$\rho = 0.2$



$\rho = 0.8$

The two points ▲ realizing the diameter

- They concentrate around the diagonal at rate $\mathrm{O}(\log n)$
- Fluctuations are Gaussian variables with variance $\frac{1-\rho}{2\rho}$

Dessert

Possible generalizations thanks the localization principle

Distribution of $\mathbb{X}$

$$\mathbb{X} = R\lambda(\mathbb{U}) \text{ with } \lambda \text{ a bounded function}$$

The behavior of $\mathbb{X}$ given that its norm is large and then the behavior of D_n will be determined by the maxima of $\|\lambda\|$:

- If they are isolated points, we obtain results similar to the case $\mathbf{m} = 1$
- Otherwise, if $\|\lambda\|$ is constant on non empty open subsets of $\mathcal{S}^{d-1}$, we obtain results similar to the case $\mathbf{m} > 1$

Dessert

Possible generalizations thanks the localization principle

Non Euclidean diameter

Another open question in Jammalamadaka and Janson :

Asymptotic of the ℓ^q-diameter of a random spherical cloud ?

Consider :

- Spherical distribution : $\Lambda = \mathrm{I_d}$ i.e. $\mathbb{X} := R\mathbb{U}$
- The ℓ^q-diameter of the cloud :

$$D_n^{(q)} := \max_{1 \leq i < j \leq n} \|\mathbb{X}_i - \mathbb{X}_j\|_q$$

where, for $q \geq 1$, $\|x\|_q$ is the ℓ^q-norm of a vector $x \in \mathbb{R}^d$

Dessert

Non Euclidean diameter

For $d \geq 2$ and $q \geq 1$, $q \neq 2$, the maximum of the ℓ^q-norm is achieved on the Euclidean sphere $\mathcal{S}^{d-1}$ at isolated points :

- If $q \in [1, 2)$ then $\max_{u \in \mathcal{S}^{d-1}} \|u\|_q = d^{1/q-1/2}$ achieved at the 2^d diagonal points $(\pm d^{-1/2}, \ldots, \pm d^{-1/2})$
- If $q \in (2, \infty)$, then $\max_{u \in \mathcal{S}^{d-1}} \|u\|_q = 1$ achieved at the $2d$ intersections of the axes with $\mathcal{S}^{d-1}$

Therefore the localization phenomenon will occur : a spherical vector $\mathbb{X}$ such that $\|\mathbb{X}\|_q$ is large must be close to the direction of one of these maximum, and $D_n^{(q)}$ will be achieved by points which are nearly diametrically opposed along one of these directions.

Dessert

Theorem [FDS, 2015]

If $q \in [1,2)$, then

$$\frac{D_n^{(q)} - 2a_n^{(q)}}{b_n^{(q)}} \xrightarrow{\mathrm{(d)}} \max_{1 \le j \le 2^{d-1}} \max_{i,i' \ge 1} \left\{ \Gamma_{i,j}^{+} + \Gamma_{i',j}^{-} - \frac{q-1}{4} \sum_{k=1}^{d} (G_{i,j,k}^{+} + G_{i',j,k}^{-})^2 \right\}$$

where $\Gamma_{i,j}^{\pm}$ are the points of independent PPP on $(-\infty, \infty]$ with mean measure $2^{-d}\mathrm{e}^{-t}\mathrm{d}t$ and $(G_{i,j,1}^{\pm}, \ldots, G_{i,j,d}^{\pm})$ are i.i.d. Gaussian vectors with covariance matrix

$$\frac{1}{d(2-q)} \begin{pmatrix} d-1 & -1 & \ldots & -1 \\ -1 & d-1 & \ldots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & \ldots & -1 & d-1 \end{pmatrix}$$

Dessert

Theorem [FDS, 2015]

If $q \in (2, \infty)$, then

$$\frac{D_n^{(q)} - 2a_n^{(q)}}{b_n^{(q)}} \xrightarrow{\text{(d)}} \max_{1 \leq i \leq d} \left\{ \Gamma_i^+ + \Gamma_i^- \right\}$$

where $\{\Gamma_i^\pm\}$ are independent Gumbel random variables with location parameter $\log 2d$.

Thank you for your attention

Complete recipes in :

The diameter of an elliptical cloud
A.-K. Fermin, Y. Demichel and P. Soulier
Electron. Journal. Probab. 20 n°27, 1-32, 2015