Towards a robust vision of geometric inference

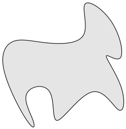
Claire Brécheteau

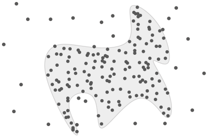
Université Paris-Sud 11, Laboratoire de Mathématiques d'Orsay and Inria Select Team – Inria Saclay, Data Shape Team

Under the supervision of
Pascal Massart (Université Paris-Sud 11) and Frédéric Chazal (Inria Saclay)

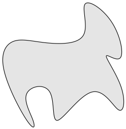
September 24, 2018

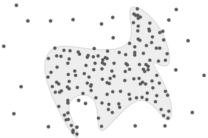
Geometric Inference : Recover geometric information from a point cloud sampled around some shape.





Geometric Inference : Recover geometric information from a point cloud sampled around some shape.





Global setting :

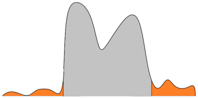
- $(\mathscr{X}, \delta)$, metric space
- P, probability distribution supported on $\mathscr{X}$

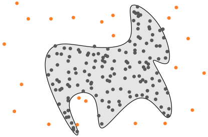
That is, $(\mathscr{X}, \delta, P)$ is a metric-measure space.

- Q, probability distribution (close to P somehow)
- $\mathbb{X}_n = \{X_1, X_2, \ldots, X_n\}$, n-sample from Q

Robustness :

1. Robustness to outliers or Trimming :

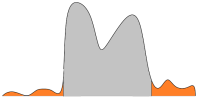


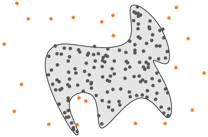


Getting rid of a proportion $1-\eta$ of the probability (resp. of the data-points).

Robustness :

1. Robustness to outliers or Trimming :



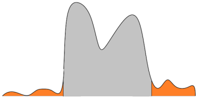


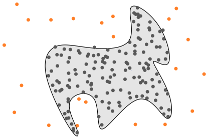
Getting rid of a proportion $1-\eta$ of the probability (resp. of the data-points).

$$t^*_\eta \in \arg\min_t \inf_{\eta\tilde{P}\leq P} \tilde{P}\gamma(t,.)$$

Robustness :

1. Robustness to outliers or Trimming :





Getting rid of a proportion $1-\eta$ of the probability (resp. of the data-points).

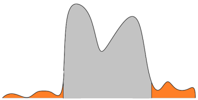
$$t^*_\eta \in \arg\min_t \inf_{\eta\tilde{P}\leq P} \tilde{P}\gamma(t,.)$$

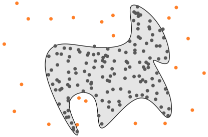
P_η such that :

$$\inf_{\eta\tilde{P}\leq P} \tilde{P}\gamma(t^*_\eta,.) = P_\eta\gamma(t^*_\eta,.)$$

Robustness :

1. Robustness to outliers or Trimming :





Getting rid of a proportion $1-\eta$ of the probability (resp. of the data-points).

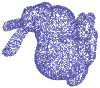
$$t_\eta^* \in \arg\min_t \inf_{\eta\tilde{P}\leq P} \tilde{P}\gamma(t,.)$$

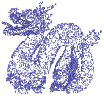
P_η such that :

$$\inf_{\eta\tilde{P}\leq P} \tilde{P}\gamma(t_\eta^*,.) = P_\eta\gamma(t_\eta^*,.)$$

2. Stability (e.g. according to a Wasserstein metric W_p).
Small $W_p(P,Q) \rightsquigarrow$ Roughly the same geometric information in P and Q.

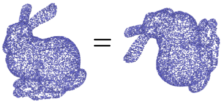
- How to compare two datasets ?

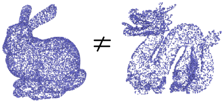
 $=$ 

 $\neq$ 

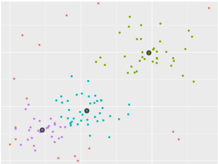
Main questions

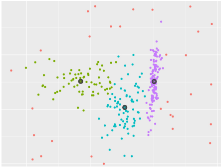
- How to compare two datasets ?



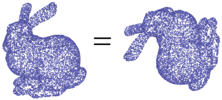


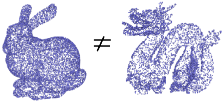
- How to make clusters from a dataset ?



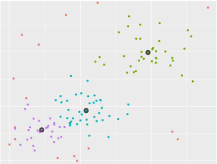


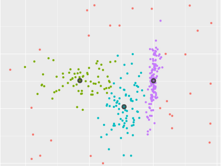
- How to compare two datasets ?



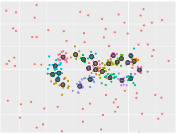


- How to make clusters from a dataset ?

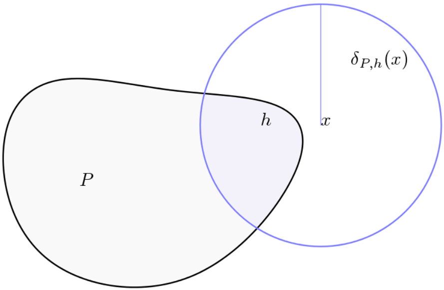




- How to infer the distance to a compact set, with a fixed budget ?

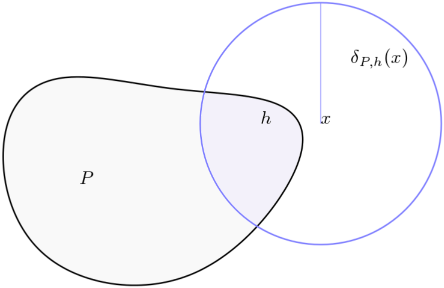


A multifunction tool : the distance-to-measure function



$$\delta_{P,h}(x) = \inf\left\{r > 0 \mid P\left(\overline{\mathrm{B}}(x,r)\right) > h\right\}$$

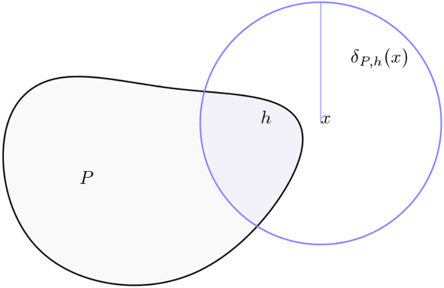
A definition for the DTM



$$\delta_{P,h}(x) = \inf\left\{r > 0 \mid P\left(\overline{\mathrm{B}}(x,r)\right) > h\right\}$$

The distance-to-measure (DTM) [Chazal, Cohen-Steiner, Mérigot 09'] is defined for all $x \in \mathscr{X}$ and $h \in [0,1]$ by :

$$\mathrm{d}_{P,h}(x) = \frac{1}{h}\int_0^h \delta_{P,l}(x)\mathrm{d}l$$



$$\delta_{P,h}(x) = \inf\left\{r > 0 \mid P\left(\overline{\mathrm{B}}(x,r)\right) > h\right\}$$

The distance-to-measure (DTM) [Chazal, Cohen-Steiner, Mérigot 09'] is defined for all $x \in \mathscr{X}$ and $h \in [0,1]$ by :

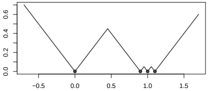
$$\mathrm{d}^{(p)}_{P,h}(x) = \left(\frac{1}{h}\int_0^h \delta^{p}_{P,l}(x)\mathrm{d}l\right)^{\frac{1}{p}}$$

The DTM for Stable – Geometric Inference

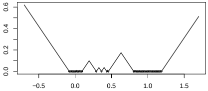
1. When $h = 0$, $\mathrm{d}_{P,0} = \mathrm{d}_{\mathcal{X}}$.
2. $\left\| \mathrm{d}_{P,h}^{(p)} - \mathrm{d}_{Q,h}^{(p)} \right\|_{\infty} \leq h^{-\frac{1}{p}} W_p(P,Q)$[Chazal, Cohen-Steiner, Mérigot 09'].

The DTM for Stable – Geometric Inference

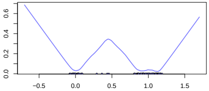
1. When $h = 0$, $\mathrm{d}_{P,0} = \mathrm{d}_{\mathscr{X}}$.
2. $\left\| \mathrm{d}^{(p)}_{P,h} - \mathrm{d}^{(p)}_{Q,h} \right\|_\infty \leq h^{-\frac{1}{p}} W_p(P,Q)$[Chazal, Cohen-Steiner, Mérigot 09'].



Distance to $\mathscr{X}$



Distance to $\mathbb{X}_n$

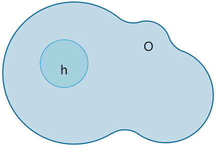


DTM with $h = 0.2$

The DTM contains information

Theorem (Brécheteau)

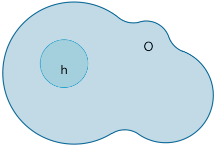
P_O (uniform distribution on O) can be recovered from $\mathrm{d}_{P_O,h}$ provided that h is small enough and O regular enough.



The DTM contains information

Theorem (Brécheteau)

P_O (uniform distribution on O) can be recovered from $\mathrm{d}_{P_O,h}$ provided that h is small enough and O regular enough.

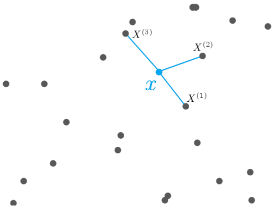


Theorem (Brécheteau)

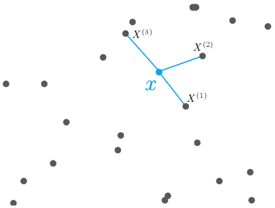
P can be recovered from $(\mathrm{d}_{P,h})_{h\in[0,1]}$ provided that $(\mathcal{X},\delta) = (\mathbb{R}^d, \|\cdot\|)$.

The DTM, an implementable tool

1. $\kappa = nh$
2. Empirical distribution $P_n = \sum_{i=1}^{n} \frac{1}{n}\delta_{X_i}$
3. $X^{(1)}$, $X^{(2)}$,... $X^{(\kappa)}$: κ nearest neighbours of x in $\mathbb{X}_n$



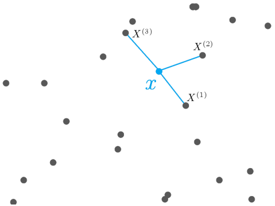
1. $\kappa = nh$
2. Empirical distribution $P_n = \sum_{i=1}^{n} \frac{1}{n} \delta_{X_i}$
3. $X^{(1)}$, $X^{(2)}$,... $X^{(\kappa)}$: κ nearest neighbours of x in $\mathbb{X}_n$



$$\mathrm{d}_{P_n,h}(x) = \frac{1}{\kappa} \sum_{i=1}^{\kappa} \delta\left(X^{(i)}, x\right)$$

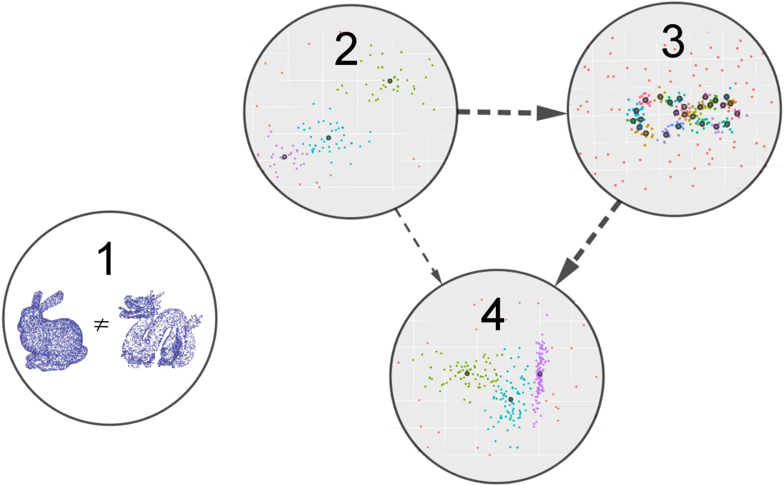
$\rightsquigarrow$ Easy implementation of the DTM at a point x in practice !

1. $\kappa = nh$
2. Empirical distribution $P_n = \sum_{i=1}^{n} \frac{1}{n}\delta_{X_i}$
3. $X^{(1)}$, $X^{(2)}$,... $X^{(\kappa)}$: κ nearest neighbours of x in $\mathbb{X}_n$

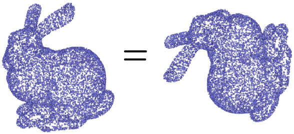


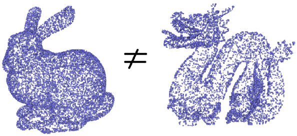
$$\mathrm{d}_{P_n,h}^{(p)}(x) = \left(\frac{1}{\kappa} \sum_{i=1}^{\kappa} \delta^{p}\left(X^{(i)}, x\right) \right)^{\frac{1}{p}}$$

$\rightsquigarrow$ Easy implementation of the DTM at a point x in practice !


1
≠
2
3
4

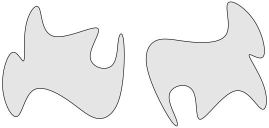
1) A statistical test of isomorphism between mm-spaces



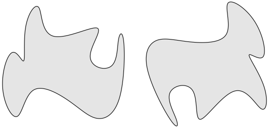


A statistical test of isomorphism between mm-spaces

Two mm-spaces $(\mathscr{X}, \delta, P)$ and $(\mathscr{Y}, \delta', P')$ are isomorphic [Gromov 81'] if :
$\exists \phi : \mathscr{X} \mapsto \mathscr{Y}$ a one-to-one isometry, s.t. for all Borel set A, $P'(\phi(A)) = P(A)$.



Two mm-spaces $(\mathscr{X}, \delta, P)$ and $(\mathscr{Y}, \delta', P')$ are isomorphic [Gromov 81'] if :
$\exists \phi : \mathscr{X} \mapsto \mathscr{Y}$ a one-to-one isometry, s.t. for all Borel set A, $P'(\phi(A)) = P(A)$.

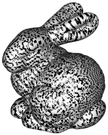


How to build a test of level $\alpha > 0$ to test the null hypothesis

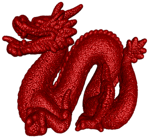
$$H_0 : \text{“}(\mathscr{X}, \delta, P) \text{ and } (\mathscr{Y}, \delta', P') \text{ are isomorphic”?}$$

vs

$$H_1 : \text{“}(\mathscr{X}, \delta, P) \text{ and } (\mathscr{Y}, \delta', P') \text{ are not isomorphic”?}$$



$(\mathscr{X}, \delta, P)$

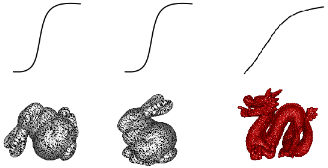


$(\mathscr{Y}, \delta', P')$

The Gromov-Wasserstein distance [Mémoli 10'] GW is a metric such that $GW((\mathscr{X},\delta,P),(\mathscr{Y},\delta',P'))=0$ iff the mm-spaces are isomorphic.
⚠ Too high computational cost.

Definition

The DTM-signature, $\mathrm{d}_{P,h}(P)$ is the distribution of $\mathrm{d}_{P,h}(X)$ when $X \sim P$.



Theorem (Brécheteau)

$$W_1\left(\mathrm{d}_{P,h}(P),\mathrm{d}_{P',h}\left(P'\right)\right) \leq \frac{1}{h} GW\left(\mathscr{X},\mathscr{Y}\right)$$

Bootstrap approximation

Definition

Defined by $\mathrm{d}_{P_N,h}(P_n)$ with P_N from $(X_1, X_2, \ldots, X_N)$ and P_n from $(X_1, X_2, \ldots, X_n)$.

Statistic :

$$T = \sqrt{n} W_1 \left(\mathrm{d}_{P_N,h}(P_n), \mathrm{d}_{P'_N,h}(P'_n)\right)$$

Subsampling distribution :

$$\mathscr{L}^*(P) = \mathscr{L}^*\left(\sqrt{n} W_1 \left(\mathrm{d}_{P_N,h}(P_n^*), \mathrm{d}_{P_N,h}(P_n^{*\prime})\right) | P_N\right)$$

Under hypothesis H_0, $\mathscr{L}(T)$ is approximated with $\mathscr{L}^* = \frac{1}{2}\mathscr{L}^*(P) + \frac{1}{2}\mathscr{L}^*(P')$.

Definition

Defined by $\mathrm{d}_{P_N,h}(P_n)$ with P_N from $(X_1, X_2, \ldots, X_N)$ and P_n from $(X_1, X_2, \ldots, X_n)$.

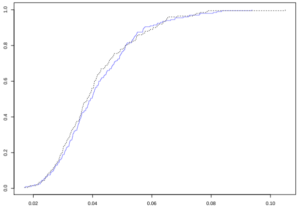
Statistic :

$$T = \sqrt{n} W_1 \left(\mathrm{d}_{P_N,h}(P_n), \mathrm{d}_{P'_N,h}(P'_n)\right)$$

Subsampling distribution :

$$\mathscr{L}^*(P) = \mathscr{L}^*\left(\sqrt{n} W_1\left(\mathrm{d}_{P_N,h}(P_n^*), \mathrm{d}_{P_N,h}(P_n^{*\prime})\right) | P_N\right)$$

Under hypothesis H_0, $\mathscr{L}(T)$ is approximated with $\mathscr{L}^* = \frac{1}{2}\mathscr{L}^*(P) + \frac{1}{2}\mathscr{L}^*(P')$.



Cdf of $\mathscr{L}(T)$ and $\mathscr{L}^*(P)$ (Bunny)
$N = 10000,\ n = 100,\ h = 0.1$

The error of type I

Test :

$$\phi_{N,n,h} = \mathbb{1}_{T \geq \hat{q}_{\alpha,N,n,h}}$$

with $\hat{q}_{\alpha,N,n,h}$, the α-quantile of $\mathscr{L}^*$.

Theorem (Brécheteau)

If P is supported on compact subsets of $\mathbb{R}^d$; $\mathscr{L}(\|\mathbb{G}_{P,h} - \mathbb{G}'_{P,h}\|_1)$ is atomless ; $n \sim N^{\frac{1}{\rho}}$;

- *in the general case, if $\rho > \frac{\max\{d,2\}}{2}$,*
- *in the (a,b)-standard case, if $\rho > 1$,*

then $\mathbb{P}_{(P,P)}\left(\phi_{N,n,h}\right) \to \alpha$, when $N \to \infty$.

$\mathbb{G}_{P,h}$ and $\mathbb{G}'_{P,h}$ independent Gaussian processes with covariance kernel $\kappa(s,t) = F_{d_{P,h}(P)}(s)\left(1 - F_{d_{P,h}(P)}(t)\right)$ for $s \leq t$.

The error of type II

$n \sim N^{\frac{1}{\rho}}$ with $\rho > 1$

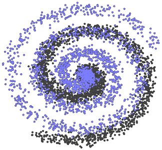
$\mathscr{X}$, $\mathscr{Y}$: compact subsets of $\mathbb{R}^d$.

Theorem (Brécheteau)

There is $n_{P,P'}$ such that $\forall n \geq n_{P,P'}$,

$$\mathbb{P}_{(P,P')}\left(1-\phi_{N,n,h}\right) \leq 4\exp\left(-\frac{W_1^2\left(\mathrm{d}_{P,h}(P),\mathrm{d}_{P',h}(P')\right)}{3\max\left\{\mathrm{Diam}_P^2,\mathrm{Diam}_{P'}^2\right\}}n\right)$$

$N = 2000$ points ; $\alpha = 0.05$, $h = 0.05$, $n = 20$.

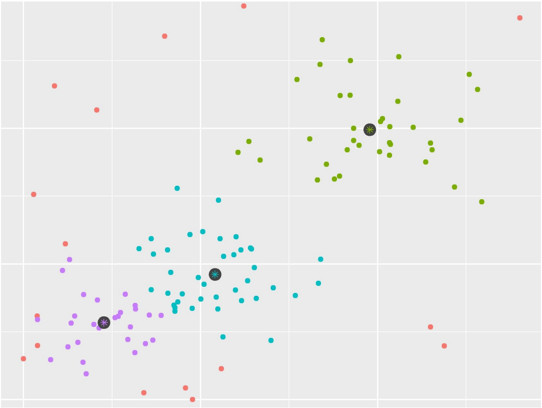


Comparison to the spiral with shape parameter 10 (grey).

spiral shape parameter	15	20	30	40	100
type I error **DTM**	0.050	0.049	0.051	0.044	0.051
type II error **DTM**	0.475	0.116	0.013	0.023	0.015
type II error **KS**	0.232	0.598	0.535	0.586	0.578

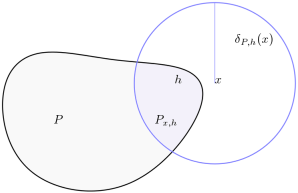
Type I and type II error approximations

2) Bregman trimmed clustering



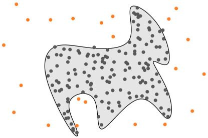
The DTM, a tool for trimming

$P_{x,h}$: restriction of P to the ball of P-mass h

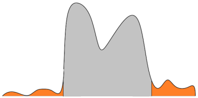


$$\mathrm{d}^2_{P,h}(x) = P_{x,h}\|.-x\|^2$$
$$= \inf_{h\tilde{P}\leq P} \tilde{P}\gamma(x,.)$$

with $\gamma(x,.) = \|.-x\|^2$.



For the DTM :
$1-h \leftrightarrow 1-\eta$



$$\mathrm{d}^2_{P,\eta} : x \mapsto \inf_{\eta\tilde{P}\leq P} \tilde{P}\|.-x\|^2$$

Minimizer x^* : Trimmed barycenter

The DTM a tool for Trimming k-means

codebook $\mathbf{c} = (c_1, c_2, \ldots, c_k)$

$$\gamma(\mathbf{c}, .) = \min_{j \in 1..k} \|. - c_j\|^2$$

$$\mathrm{B}(x, r) \leftrightarrow \bigcup_{i \in 1..k} \mathrm{B}\left(c_i, r\right).$$

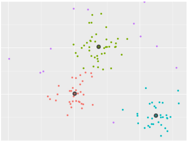
codebook $\mathbf{c} = (c_1, c_2, \ldots, c_k)$

$$\gamma(\mathbf{c}, .) = \min_{j \in 1..k} \|. - c_j\|^2$$

$$\mathrm{B}(x, r) \leftrightarrow \bigcup_{i \in 1..k} \mathrm{B}\left(c_i, r\right).$$

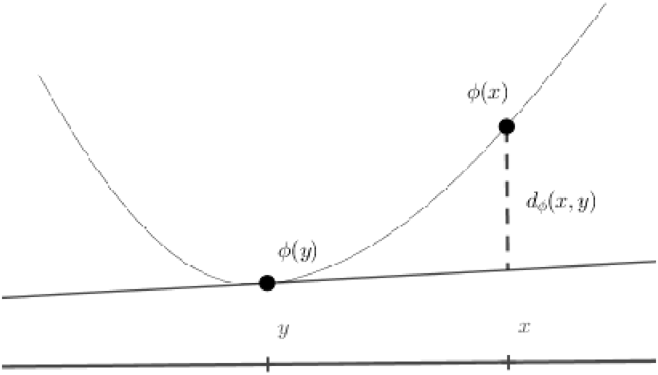
$$\mathrm{d}^2_{P,\eta} : \mathbf{c} \mapsto \inf_{\eta \tilde{P} \leq P} \tilde{P} \min_{j \in 1..k} \|. - c_j\|^2$$

Minimizer $\mathbf{c}^*$: Optimal codebook for Trimmed k-means [Cuesta et al. 97']



$\Omega \subset \mathbb{R}^d$ convex set, $\phi : \Omega \to \mathbb{R}$ strictly convex and $\mathscr{C}^1$. Bregman divergence

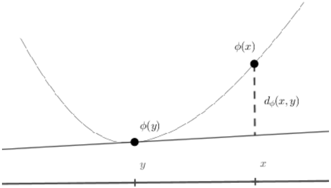
$$d_\phi : (x, y) \mapsto \phi(x) - \phi(y) - \langle \nabla_y \phi, x - y \rangle, \forall x, y \in \Omega$$



An example of trimmed clustering with a Bregman divergence

$\Omega \subset \mathbb{R}^d$ convex set, $\phi : \Omega \to \mathbb{R}$ strictly convex and $\mathscr{C}^1$. Bregman divergence

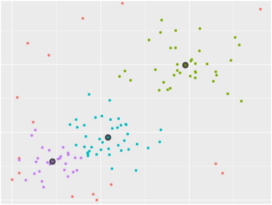
$$\mathrm{d}_\phi : (x,y) \mapsto \phi(x) - \phi(y) - \langle \nabla_y \phi, x - y \rangle, \ \forall x, y \in \Omega$$



Poisson distribution :

$$-\log\left(p_\theta(x)\right) = -\log\left(\frac{\theta^x}{x!}e^{-\theta}\right)$$
$$= \mathrm{d}_\phi(x,\theta) + C(x),$$

for $\phi(x) = x\log(x) - x$,
and $\mathrm{d}_\phi(x,c) = x\log\left(\frac{x}{c}\right) - (x-c)$.



Definition

Bregman h-trimmed variation given $\mathbf{c}$ - or - Bregman-divergence-to-measure :

$$\mathrm{d}^2_{\phi,P,\eta}(\mathbf{c}) = \inf_{\eta\tilde{P}\leq P} \tilde{P} \min_{j\in 1..k} \mathrm{d}_\phi(.,c_j)$$

Definition

A Bregman h-trimmed k-optimal codebook $\mathbf{c}^*$ is any minimizer $\mathbf{c}$ of the criterion $\mathrm{d}_{\phi,P,\eta}(\mathbf{c})$.

Theorem (Brécheteau, Fischer and Levrard)

Assume that ϕ is $\mathscr{C}^2$ and strictly convex and $F_0 = \overline{\mathrm{Conv}(\mathrm{Supp}(P))} \subset \overset{o}{\Omega}$.
Then, the minimum $\mathbf{c}^*$ exists.

$\hat{\mathbf{c}}_n$: minimizer of $\mathrm{d}_{\phi,P_n,\eta}$.

Theorem (Brécheteau, Fischer and Levrard)

If P is continuous, $P\|.\|^p < \infty$ for some $p > 2$, ϕ is $\mathscr{C}_2$ on Ω, $F_0 = \overline{\mathrm{Conv}(\mathrm{Supp}(P))} \subset \Omega^\circ$ and $\mathbf{c}^*$ is the unique minimizer of $\mathrm{d}_{\phi,P,\eta}$, then :

$$\lim_{n\to+\infty} \mathrm{dist}(\hat{\mathbf{c}}_n, \mathbf{c}^*) = 0 \text{ a.e.}$$

and

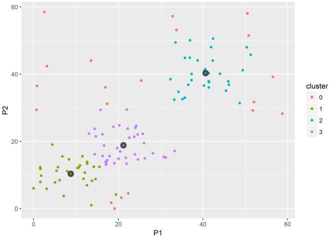
$$\lim_{n\to+\infty} \mathrm{d}_{\phi,P,\eta}(\hat{\mathbf{c}}_n) = \mathrm{d}_{\phi,P,\eta}(\mathbf{c}^*) \text{ a.e.}$$

This convergence holds at a parametric rate $\frac{1}{\sqrt{n}}$:

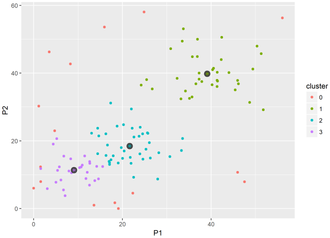
Theorem (Brécheteau, Fischer and Levrard)

Assume that $P\|.\|^p < \infty$. Then, for n large enough, with probability greater than $1 - n^{-\frac{p}{2}} - 2e^{-x}$, we have

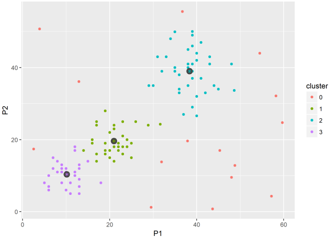
$$\mathrm{d}_{\phi,P,\eta}(\hat{\mathbf{c}}_n) - \mathrm{d}_{\phi,P,\eta}(\mathbf{c}^*) \leq \frac{C_P}{\eta\sqrt{n}}(1 + \sqrt{x}).$$



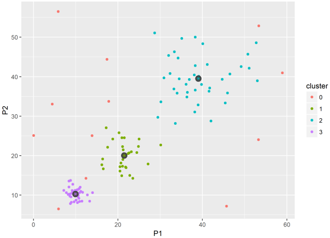
Gaussian mixture



Poisson mixture



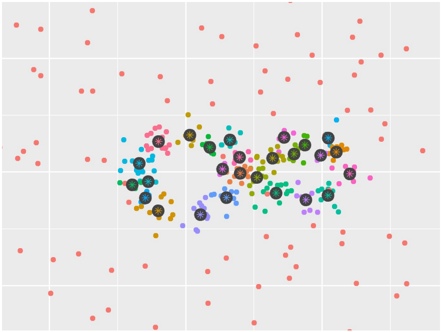
Binomial mixture



Gamma mixture

Clustering associated to the selected parameter η - dimension 2

3) Distance to a compact set inference, with a quantization point of view



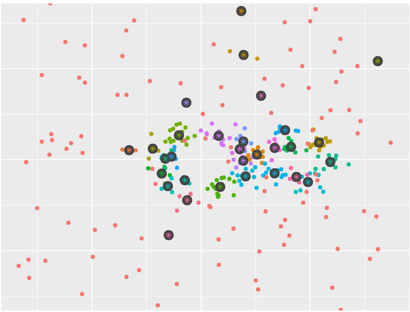
Given P, Q or $\mathbb{X}_n$:
Find $\mathbf{c}$ and $\boldsymbol{\omega}$ such that the k-power function

$$x \mapsto \min_{j \in 1..k} \|x - c_j\|^2 + \omega_j^2$$

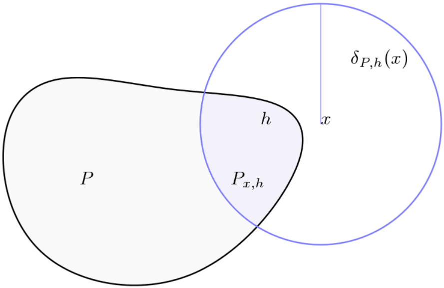
is a good approximation of the square of the distance to $\mathscr{X}$,

$$\mathrm{d}^2_{\mathscr{X}} : x \mapsto \min_{y \in \mathscr{X}} \|x - y\|^2$$

Trimmed k-means does not work...

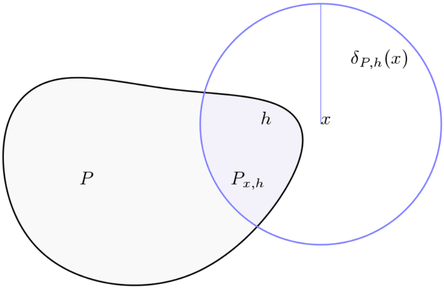


The DTM, an alternative definition as a power distance when $p = 2$ [Chazal, Cohen-Steiner, Mérigot 09']



$$
\begin{aligned}
d^2_{P,h}(x) &= P_{x,h}\|.-x\|^2\\
&= \inf_{h\tilde{P}\leq P} \tilde{P}\|.-x\|^2\\
&= \|m(P_{x,h})-x\|^2 + v(P_{x,h})\\
&= \inf_{h\tilde{P}\leq P} \|m(\tilde{P})-x\|^2 + v(\tilde{P})
\end{aligned}
$$

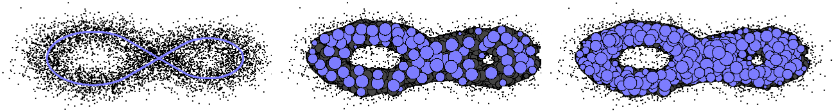
Notation : Mean $m(P)$, Variance $v(P)$.



$$d^2_{P,h}(x) = P_{x,h}\|.-x\|^2$$
$$= \inf_{h\tilde{P}\leq P} \tilde{P}\|.-x\|^2$$
$$= \|m(P_{x,h})-x\|^2 + v(P_{x,h})$$
$$= \inf_{h\tilde{P}\leq P} \|m(\tilde{P})-x\|^2 + v(\tilde{P})$$

Notation : Mean $m(P)$, Variance $v(P)$.

$\rightsquigarrow$ Sublevel sets of the DTM : union of balls.
Approximation : κ-witnessed distance [Guibas Morozov Mérigot 11'] n balls



Set $\phi_{P,h}$ the function defined on $\mathbb{R}^d$ by

$$\phi_{P,h} : x \mapsto \|x\|^2 - \mathrm{d}^2_{P,h}(x). \tag{1}$$

[Chazal, Cohen-Steiner, Mérigot 09'] The map $\phi_{P,h}$ is convex.
The Bregman-divergence associate to $\phi_{P,h}$ satisfies for $x, t \in \mathbb{R}^d$:

$$\mathrm{d}_{\phi_{P,h}}(x,t) = \|x - m(P_{t,h})\|^2 + v(P_{t,h}) - \mathrm{d}^2_{P,h}(x)$$

$$\min_{j\in 1..k} \mathrm{d}_{\phi_{P,h}}(.,t_j) = \left(\min_{j\in 1..k} \|x - m(P_{t,h})\|^2 + v(P_{t,h})\right) - \mathrm{d}^2_{P,h}(x)$$

$\rightsquigarrow$ Bregman clustering with $\mathrm{d}_{\phi_{P,h}}$!
Rq : Theory $1-\eta = 0$ (For practice $1-\eta \in [0,1)$) !

$$\begin{aligned}\mathbf{t}^* &\in \arg\min_{\mathbf{t}} P \min_{j\in 1..k} d_{\phi_{P,h}}(., t_j)\\ &= \arg\min_{\mathbf{t}} P \min_{j\in 1..k} \|. - m(P_{t_j,h})\|^2 + v(P_{t_j,h}) - d^2_{P,h}(.)\end{aligned}$$

Definition

The k-power distance-to-measure (k-PDTM) $d_{P,h,k}$ is defined for $x \in \mathbb{R}^d$ by :

$$d^2_{P,h,k}(x) = \min_{j\in 1..k} \|x - m(P_{t_j^*,h})\|^2 + v(P_{t_j^*,h})$$

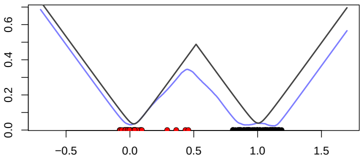
$$\omega(c) = \inf\left\{\omega > 0 \mid \forall x \in \mathbb{R}^d,\ \|x - c\|^2 + \omega^2 \geq \mathrm{d}^2_{P,h}(x)\right\}$$

Theorem (Brécheteau and Levrard)

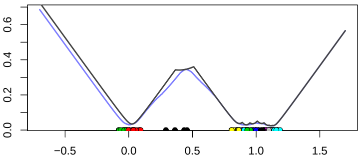
$$\mathrm{d}^2_{P,h,k}(x) = \min_{j \in 1..k} \|x - c_j^*\|^2 + \omega^2(c_j^*)$$

for

$$\mathbf{c}^* \in \arg\min\left\{P \min_{j \in 1..k} \|. - c_j\|^2 + \omega^2(c_j)\right\}$$



k-PDTM, $k = 2$ centres



k-PDTM, $k = 10$ centres

Proposition

If $\mathrm{Supp}(P) \subset \mathrm{B}(0,K)$*, and* $Q\|.\| < \infty$*,*
then $P\left|\mathrm{d}^2_{Q,h,k}(.) - \mathrm{d}^2_{P,h}(.)\right|$ *is bounded from above by*

$$3\|\mathrm{d}^2_{Q,h} - \mathrm{d}^2_{P,h}\|_{\infty,\mathrm{B}(0,K)} + P\left(\mathrm{d}^2_{P,h,k}(.) - \mathrm{d}^2_{P,h}(.)\right) + 4W_1(P,Q)\sup_{s\in\mathbb{R}^d}\|m(P_{s,h})\|$$

with $P\left(\mathrm{d}^2_{P,h,k}(.) - \mathrm{d}^2_{P,h}(.)\right)$ of order $k^{-\frac{2}{d'}}$ for a "d'-dimensional distribution".

$\text{Supp}(P) = \mathcal{X} \subset \text{B}(0, K)$
$X_i = Y_i + Z_i$, Y_i and Z_i all independent, $Y_i \sim P$, Z_i sub-Gaussian with variance $\sigma^2 \leq K^2$
$Q_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$.

Theorem (Brécheteau and Levrard)

For every $p > 0$, with probability larger than $1 - 10n^{-p}$, we have

$$\left| P\text{d}^2_{Q_n,h,k}(.) - \text{d}^2_{Q,h,k}(.) \right| \leq C\sqrt{kd}\frac{K^2((p+1)\log(n))^{\frac{3}{2}}}{h\sqrt{n}} + C\frac{K\sigma}{\sqrt{h}}.$$

$\mathrm{Supp}(P) = \mathscr{X} \subset \mathrm{B}(0, K)$
$X_i = Y_i + Z_i$, Y_i and Z_i all independent, $Y_i \sim P$, Z_i sub-Gaussian with variance $\sigma^2 \leq K^2$
$Q_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$.

Theorem (Brécheteau and Levrard)

For every $p > 0$, with probability larger than $1 - 10n^{-p}$, we have

$$\left|P\mathrm{d}^2_{Q_n,h,k}(.) - \mathrm{d}^2_{Q,h,k}(.)\right| \leq C\sqrt{kd}\frac{K^2((p+1)\log(n))^{\frac{3}{2}}}{h\sqrt{n}} + C\frac{K\sigma}{\sqrt{h}}.$$

$\rightsquigarrow$ optimize in k the quantity

$$\frac{C\sqrt{k}K^2((p+1)\log(n))^{\frac{3}{2}}}{h\sqrt{n}} + C_{P,h}k^{-\frac{2}{d'}}.$$

Approximation of the k-PDTM from point clouds

$\mathrm{Supp}(P) = \mathscr{X} \subset \mathrm{B}(0, K)$
$X_i = Y_i + Z_i$, Y_i and Z_i all independent, $Y_i \sim P$, Z_i sub-Gaussian with variance $\sigma^2 \leq K^2$
$Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$.

Theorem (Brécheteau and Levrard)

For every $p > 0$, with probability larger than $1 - 10n^{-p}$, we have

$$\left|P\mathrm{d}^2_{Q_n,h,k}(.) - \mathrm{d}^2_{Q,h,k}(.)\right| \leq C\sqrt{kd}\frac{K^2((p+1)\log(n))^{\frac{3}{2}}}{h\sqrt{n}} + C\frac{K\sigma}{\sqrt{h}}.$$

$\rightsquigarrow$ optimize in k the quantity

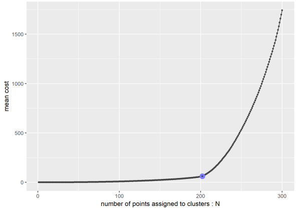
$$\frac{C\sqrt{k}K^2((p+1)\log(n))^{\frac{3}{2}}}{h\sqrt{n}} + C_{P,h}k^{-\frac{2}{d'}}.$$

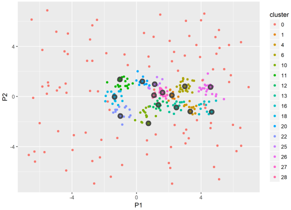
Optimal choice $k \sim n^{\frac{d'}{d'+4}}$.

Theorem (Brécheteau and Levrard)

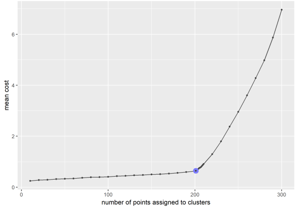
Assumption : $\forall x \in \mathscr{X},\ P(\mathrm{B}(x,r)) \geq C(P)r^{d'} \wedge 1.$
Set $\Delta_P^2 = P\mathrm{d}_{Q,h,k}^2(.)$, *then,*

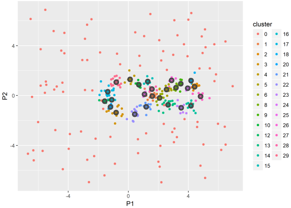
$$\sup_{x\in\mathbb{R}^d} |\mathrm{d}_{Q,h,k}(x) - \mathrm{d}_{\mathscr{X}}(x)| \leq C(P)^{-\frac{1}{d'+2}} \Delta_P^{\frac{2}{d'+2}} + 2\Delta_P + W_2(P,Q)h^{-\frac{1}{2}}.$$





k-PDTM $(\eta = 1)$





Trimmed k-PDTM $(\eta < 1)$

Summing up

Method	New tool
Isomorphism Test	DTM-signature $\mathrm{d}_{P,h}(P)$
Bregman clustering	Bregman divergence-to-measure $\mathbf{c} \mapsto \mathrm{d}_{\phi,P,\eta}(\mathbf{c})$
Quantization & $\mathrm{d}_{\mathscr{X}}$ inference	k-PDTM $x \mapsto \mathrm{d}_{P,h,k}(x)$

Method	New tool
Isomorphism Test	DTM-signature $\mathrm{d}_{P,h}(P)$
Bregman clustering	Bregman divergence-to-measure $\mathbf{c} \mapsto \mathrm{d}_{\phi,P,\eta}(\mathbf{c})$
Quantization & $\mathrm{d}_{\mathscr{X}}$ inference	k-PDTM $x \mapsto \mathrm{d}_{P,h,k}(x)$

Future work :
Non asymptotic statistics for studying

$$\hat{t}_\eta \in \arg\min_t \inf_{\eta \tilde{P_n} \leq P_n} \tilde{P_n}\gamma(t,.)$$

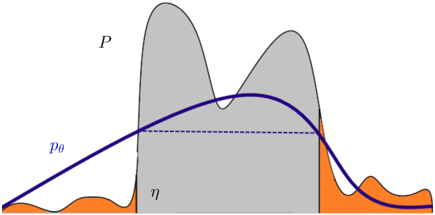
Thank you !

The DTM and the trimmed log-likelihood

$\mathscr{F} = \{P_\theta\}_\theta$; P_θ with density p_θ.

$\gamma(\theta, .) = -\log(p_\theta(.))$

$\mathrm{B}(x, r) \leftrightarrow$ upper-level set of p_θ.



$P^{\mathcal{F}}_{\theta,\eta}$

$$\mathrm{d}^{\mathscr{F}}_{P,\eta} : \theta \mapsto \inf_{\eta \tilde{P} \leq P} \tilde{P} - \log(p_\theta(.))$$

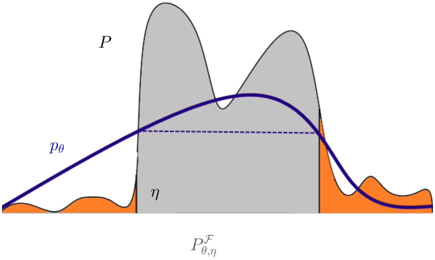
Minimizer θ^* : Trimmed log-likelihood maximizer

$\mathscr{F} = \{P_\theta\}_\theta$; P_θ with density p_θ.

$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$

$\gamma(\boldsymbol{\theta}, .) = \min_{j \in 1..k} -\log(p_{\theta_j}(.))$

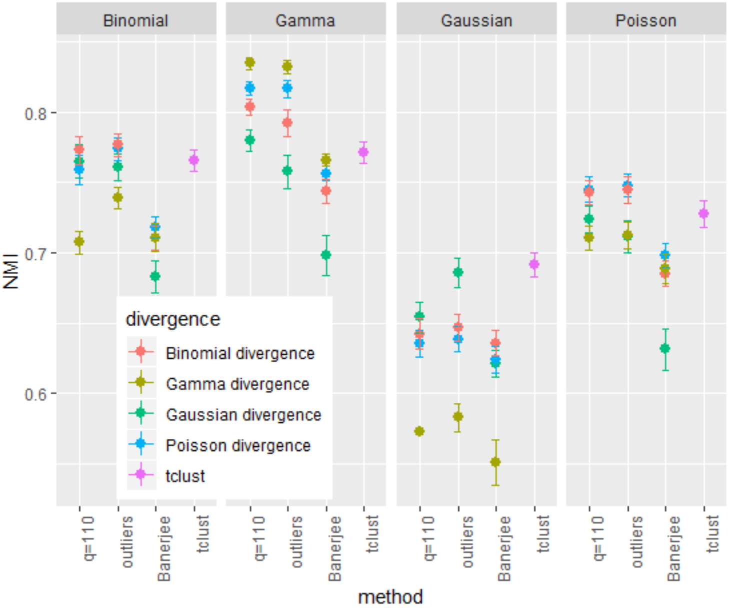
$\mathrm{B}(x, r) \leftrightarrow$ union of upper-level set of the p_{θ_j}s.



$$\mathrm{d}^{\mathscr{F}}_{P,\eta} : \boldsymbol{\theta} \mapsto \inf_{\eta\tilde{P} \leq P} \tilde{P} \min_{j \in 1..k} -\log(p_{\theta_j}(.))$$

Minimizer $\boldsymbol{\theta}^*$: Optimal codebook for some trimmed clustering...

Experiments



Robust heteroscedastic Gaussian clustering