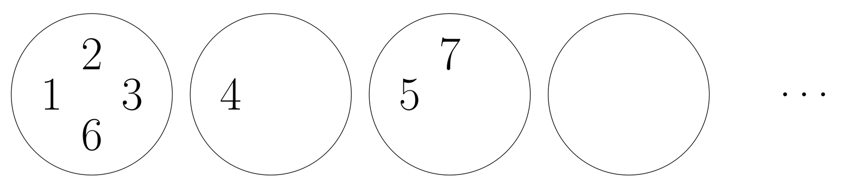
# Analysis of the maximal a posteriori partition in the Gaussian Dirichlet Process Mixture Model

Łukasz Rajkowski
l.rajkowski@mimuw.edu.pl
University of Warsaw

## The Chinese Restaurant Process

This construction can be used to define the Dirichlet Process




$\mathbb{P}(\text{new table}) \propto \alpha \qquad \mathbb{P}(\text{join table}) \propto \#\text{ sitting there}$

e.g. $\mathbb{P}\{\{1,2,4,6\},\{3\},\{5,7\}\} = \frac{\alpha}{\alpha}\cdot\frac{1}{1+\alpha}\cdot\frac{\alpha}{2+\alpha}\cdot\frac{2}{3+\alpha}\cdot\frac{\alpha}{4+\alpha}\cdot\frac{3}{5+\alpha}\cdot\frac{1}{6+\alpha}$

## The Gaussian DPMM and the MAP

The Gaussian DPMM for $n$ observations may be modelled as follows

$$\mathcal{J} \sim \mathrm{CRP}(\alpha)_n \quad \text{(the Chinese Restaurant P.)}$$

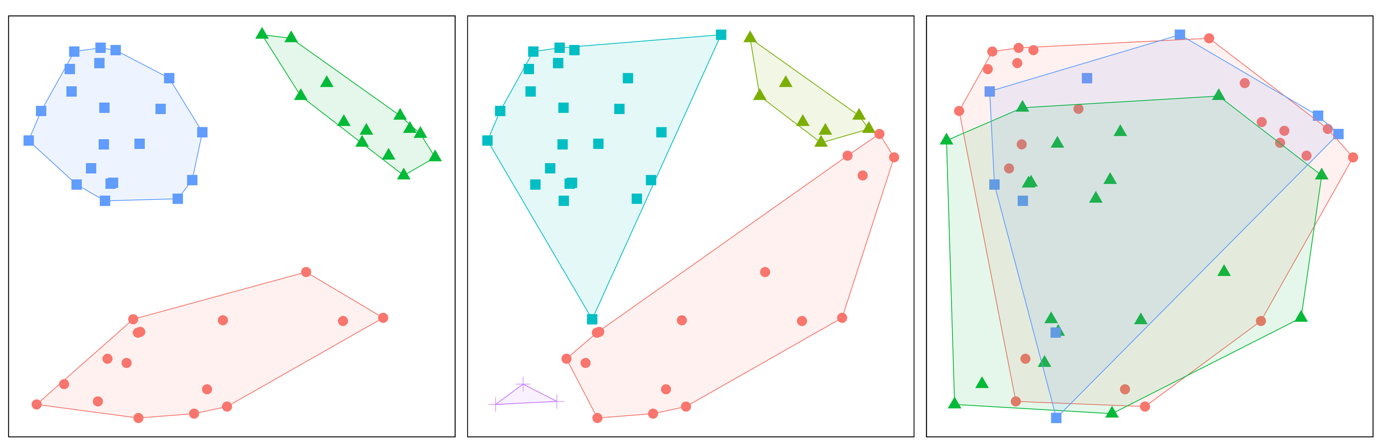
$$\boldsymbol{\theta} = (\theta_J)_{J\in\mathcal{J}} \,|\, \mathcal{J} \overset{\text{iid}}{\sim} \mathcal{N}(\vec{\mu}, T)$$

$$\boldsymbol{x}_J = (x_j)_{j\in J} \,|\, \mathcal{J}, \boldsymbol{\theta} \overset{\text{iid}}{\sim} \mathcal{N}(\theta_J, \Sigma) \quad \text{for } J \in \mathcal{J}$$

The partition that maximises the posterior probability is *the MAP partition*. It is denoted by $\hat{\mathcal{J}}(x_1,\ldots,x_n)$.

## Result 1. Convexity of the MAP

For every $n\in\mathbb{N}$ if $J_1, J_2 \in \hat{\mathcal{J}}(x_1,\ldots,x_n)$, $J_1 \neq J_2$ and $A_k$ is the convex hull of the set $\{x_i : i \in J_k\}$ for $k=1,2$ then $A_1 \cap A_2$ is an empty set or a singleton $\{x_i\}$ for some $i \le n$.



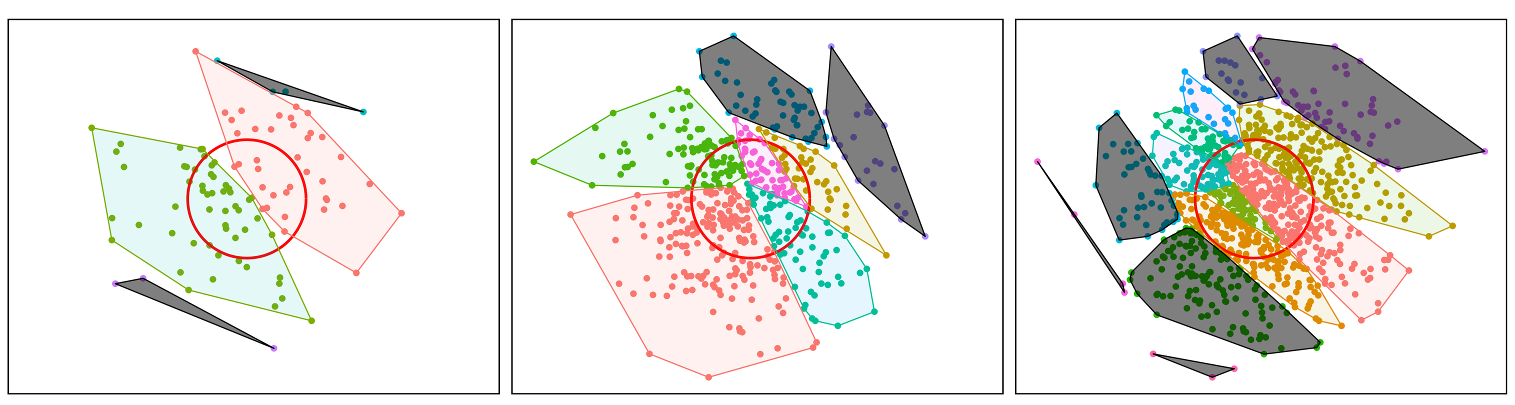
(a) This is the desired partition which is also convex. (b) This is a convex partition which is not ideal. (c) This partition is not convex and it is clearly a bad one.

## Result 2. Which clusters are big?

If $\sup_n \frac{1}{n}\sum_{i=1}^n \|x_n\|^2 < \infty$ then

$$\liminf_{n\to\infty} \min\{|J| : J \in \hat{\mathcal{J}}(x_1,\ldots,x_n), \exists_{j\in J}\|x_j\| < r\}/n > 0$$

for every $r > 0$.



(a) $n = 100$ (b) $n = 500$ (c) $n = 1000$

**Corollary.** If $\left(\frac{1}{n}\sum_{i=1}^n \|x_i\|^2\right)_{n=1}^\infty$ is bounded then for every $r > 0$ the number of clusters that intersect $B(\mathbf{0}, r)$ is bounded.

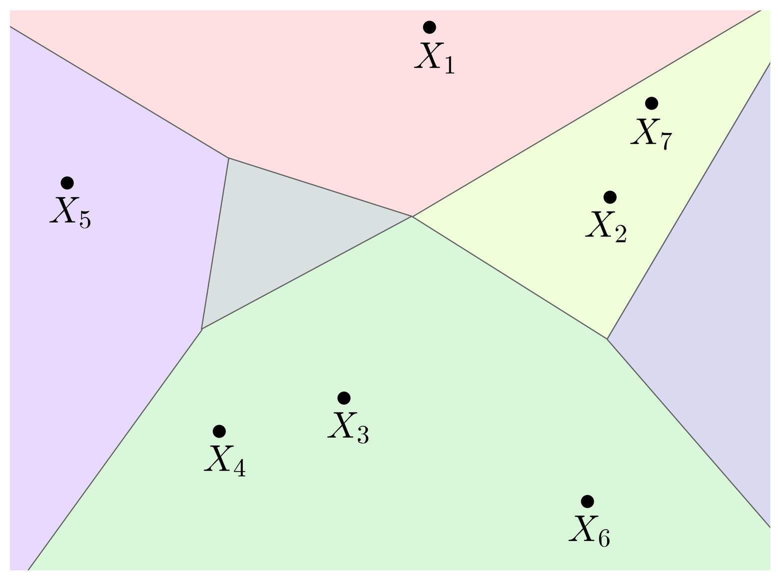
## Commentary

Let $X_1, X_2, \ldots \overset{\text{iid}}{\sim} P$ and $\hat{\mathcal{J}}_n = \hat{\mathcal{J}}(X_1,\ldots,X_n)$.

- If $\alpha = T = \Sigma = 1$ and $P = \sum_{m=0}^\infty q(1-q)^m \delta_{18^m}$, where $q = (2\cdot 18)^{-1}$, then $\mathbb{E}\,X^4 < \infty$ and almost surely $\liminf_{n\to\infty} \min\{|J| : J \in \hat{\mathcal{J}}_n\} = 1$.
- If $\alpha = T = 1$, $\Sigma < (32\ln 2)^{-1}$ and $P = \mathrm{Exp}(1)$ then $\lim_{n\to\infty} |\hat{\mathcal{J}}_n| = \infty$ almost surely.

This implies that Result 2 is not easily generalised!

## The Induced Partition




Let $\mathcal{A}$ be a fixed partition of $\mathbb{R}^d$. For $n \in \mathbb{N}$ and $A \in \mathcal{A}$ let $J_n^A = \{i \le n : X_i \in A\}$ and define a random partition of $[n]$ by

$$\mathcal{J}_n^{\mathcal{A}} = \{J_n^A \neq \emptyset : A \in \mathcal{A}\}.$$

We say that this partition of $[n]$ is *induced by* $\mathcal{A}$.

$\longrightarrow \mathcal{J}_7^{\mathcal{A}} = \{\{1\},\{2,7\},\{3,4,6\},\{5\}\}$

## The function Δ

Let $\Delta$ be the function on the space of finite families of measurable sets defined by

$$\Delta(\mathcal{A}) = \underbrace{\frac{1}{2}\sum_{A\in\mathcal{A}} P(A)\left\|\Sigma^{-1/2}\mathbb{E}(X \,|\, X\in A)\right\|^2}_{\text{variance of CEV}} + \underbrace{\sum_{A\in\mathcal{A}} P(A)\ln P(A)}_{-\text{entropy of CEV}}.$$

Consider $X_1, X_2, \ldots \overset{\text{iid}}{\sim} P$. Then

$$\sqrt[n]{\text{posterior score of } \mathcal{J}_n^{\mathcal{A}}} \approx n\exp\{\Delta(\mathcal{A})\}$$

Let $\hat{\mathcal{A}}_n$ be the family of the convex hulls of clusters in $\hat{\mathcal{J}}_n = \hat{\mathcal{J}}(X_1,\ldots,X_n)$.

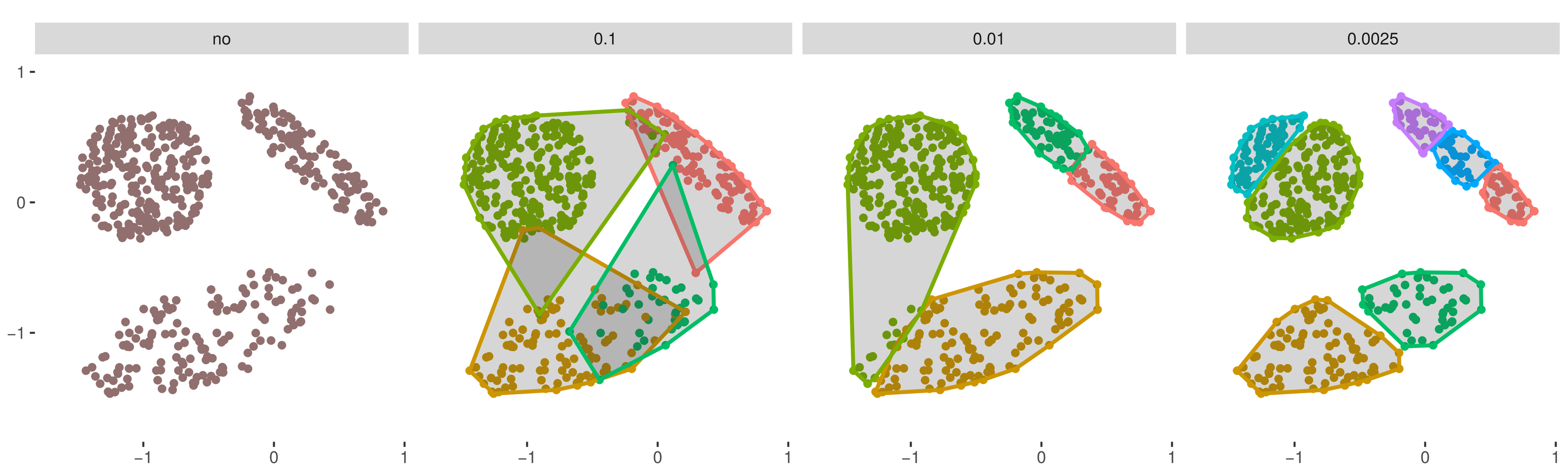
$$\sqrt[n]{\text{posterior score of } \hat{\mathcal{J}}_n} \approx n\exp\{\Delta(\hat{\mathcal{A}}_n)\}$$

## Result 3. The MAP asymptotic

Assume that $P$ has bounded support and is continuous with respect to Lebesgue measure. Then the distance between $\hat{\mathcal{A}}_n$ and the set of partitions that maximise the function $\Delta$ converges to 0.

## Result 4. The force of Σ

Assume that $P$ has bounded support and is continuous with respect to Lebesgue measure. Then for every $K \in \mathbb{N}$ there exists an $\varepsilon > 0$ such that if $\|\Sigma\| < \varepsilon$ then $|\hat{\mathcal{J}}_n| > K$ for sufficiently large $n$.




$\alpha = 1, T = Id, \Sigma = \sigma^2 Id$, where $\sigma^2 \in \{.1, .01, .0025\}$

## What's next?

1. **'Limit' result and unbounded support of $P$**
   - the possibility of small probability clusters distant from 0, unbounded # of clusters
   - no chance of convergence in Hausdorff metric, perhaps only $d_P$
2. **Prior on the covariance structure**
   - we may put Wishart distribution on covariance parameter
   - preliminary computations for induced partitions give

$$\Delta'(\mathcal{A}) = -\frac{1}{2}\sum p_A \ln\det\left(\mathrm{Var}(X \,|\, X\in A)\right) - \sum p_A \ln p_A$$

   - more difficult to relate induced partitions to the MAP

The details available on arXiv and soon in Bayesian Analysis