Inference on Dependent Data: Contributions to Hidden Markov and Preferential Attachment Graph Models

Ibrahim Kaddouri

Under the supervision of Elisabeth Gassiat and Zacharie Naulet

December 1st, 2025

université PARIS-SACLAY

Mathématiques Orsay

Outline

What is Statistical Inference?

Statistical Model

A **statistical model** is a family of probability distributions

$$\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\},$$

all defined on a common probability space $(\Omega, \mathcal{F})$.

- An experiment draws data $X \sim \mathbb{P}_{\theta^*}$ for some *unknown* parameter $\theta^* \in \Theta$
- The statistician observes X only

What is Statistical Inference?

Statistical Model

A **statistical model** is a family of probability distributions

$$\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\},$$

all defined on a common probability space $(\Omega, \mathcal{F})$.

- An experiment draws data $X \sim \mathbb{P}_{\theta^*}$ for some *unknown* parameter $\theta^* \in \Theta$
- The statistician observes X only

Goal: Use observation X to gain knowledge about $\theta^\star$, the distribution $\mathbb{P}_{\theta^\star}$, or the structure of Θ.

What is Statistical Inference?

Statistical Model

A **statistical model** is a family of probability distributions

$$\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\},$$

all defined on a common probability space $(\Omega, \mathcal{F})$.

- An experiment draws data $X \sim \mathbb{P}_{\theta^*}$ for some *unknown* parameter $\theta^* \in \Theta$
- The statistician observes X only

Goal: Use observation X to gain knowledge about $\theta^\star$, the distribution $\mathbb{P}_{\theta^\star}$, or the structure of Θ.

Example: Gaussian Location Model

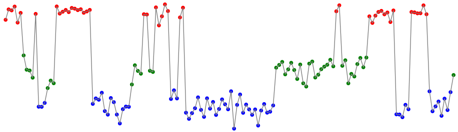
Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$

- The statistician observes $X = (X_1, \ldots, X_n)$ only
- Here, the model is $\mathcal{P} = \{\mathcal{N}(\mu, 1)^{\otimes n} : \mu \in \mathbb{R}\}$

Statistical Inference on Dependent Data

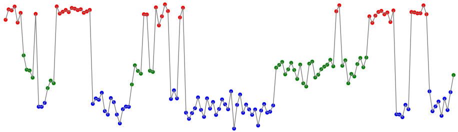
To incorporate dependence, we will consider two models of completely different nature:

Hidden Markov Model

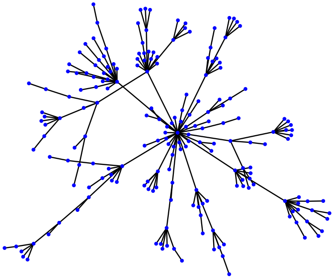


To incorporate dependence, we will consider two models of completely different nature:

Hidden Markov Model



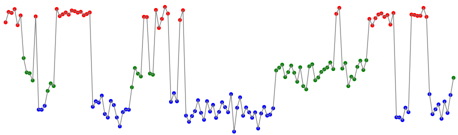
Preferential Attachment Model



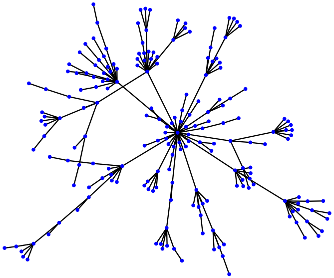
Statistical Inference on Dependent Data

To incorporate dependence, we will consider two models of completely different nature:

Hidden Markov Model



Preferential Attachment Model



Goal : Study some inference problems under both models

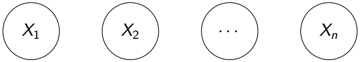
Outline

Outline

Mixture Models vs Hidden Markov Models

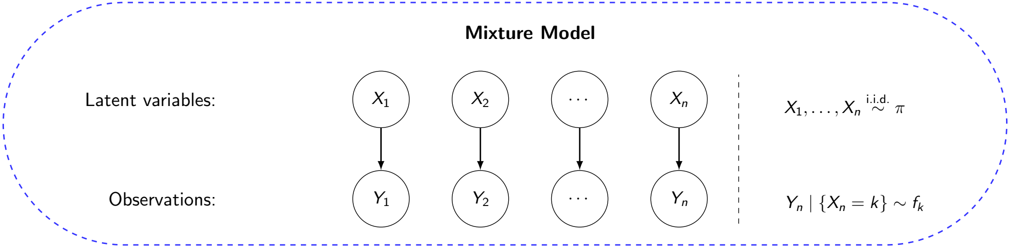
Mixture Model

Latent variables:

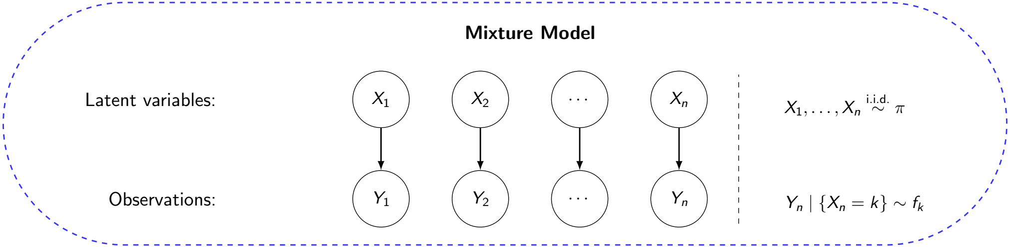


$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \pi$

Mixture Models vs Hidden Markov Models



Mixture Models vs Hidden Markov Models



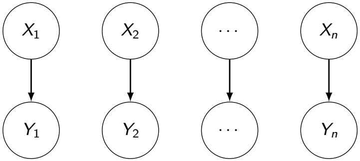
We denote the set of parameters $\theta = (K, \pi, f)$ by Θ^{ind}.

Mixture Models vs Hidden Markov Models

Mixture Model

Latent variables:

Observations:



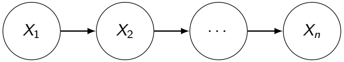
$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \pi$

$Y_n \mid \{X_n = k\} \sim f_k$

We denote the set of parameters $\theta = (K, \pi, f)$ by Θ^{ind}.

Hidden Markov Model

Latent variables:



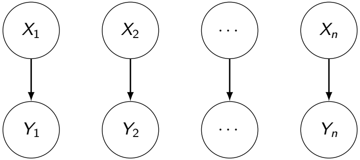
$X_1, \ldots, X_n \sim \text{Markov}(\pi, Q)$

Mixture Models vs Hidden Markov Models

Mixture Model

Latent variables:

Observations:



$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \pi$

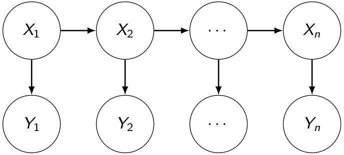
$Y_n \mid \{X_n = k\} \sim f_k$

We denote the set of parameters $\theta = (K, \pi, f)$ by Θ^{ind}.

Hidden Markov Model

Latent variables:

Observations:

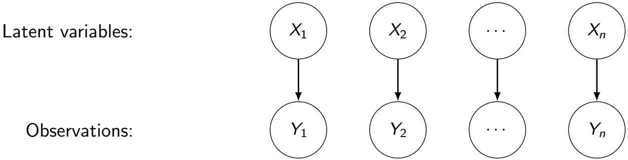


$X_1, \ldots, X_n \sim \text{Markov}(\pi, Q)$

$Y_n \mid \{X_n = k\} \sim f_k$

Mixture Models vs Hidden Markov Models

Mixture Model

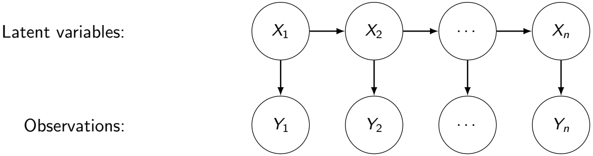


$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \pi$

$Y_n \mid \{X_n = k\} \sim f_k$

We denote the set of parameters $\theta = (K, \pi, f)$ by Θ^{ind}.

Hidden Markov Model



$X_1, \ldots, X_n \sim \text{Markov}(\pi, Q)$

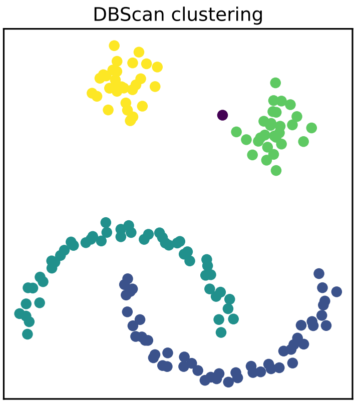
$Y_n \mid \{X_n = k\} \sim f_k$

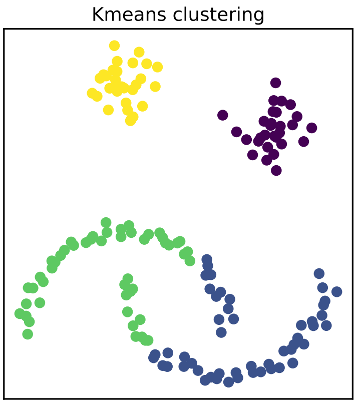
We denote the set of parameters $\theta = (K, \pi, Q, f)$ by Θ^{dep}.

Outline

What is Clustering?

Clustering identifies meaningful structure in data.

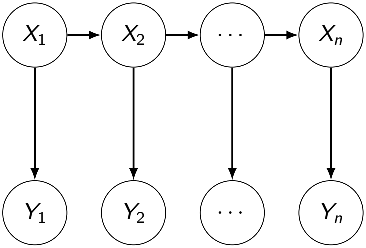




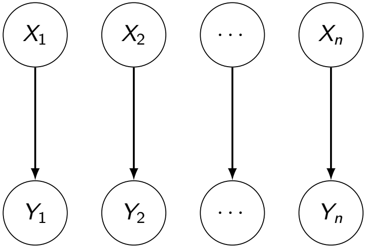
Goal: Group similar points together through a partition

What is Clustering?

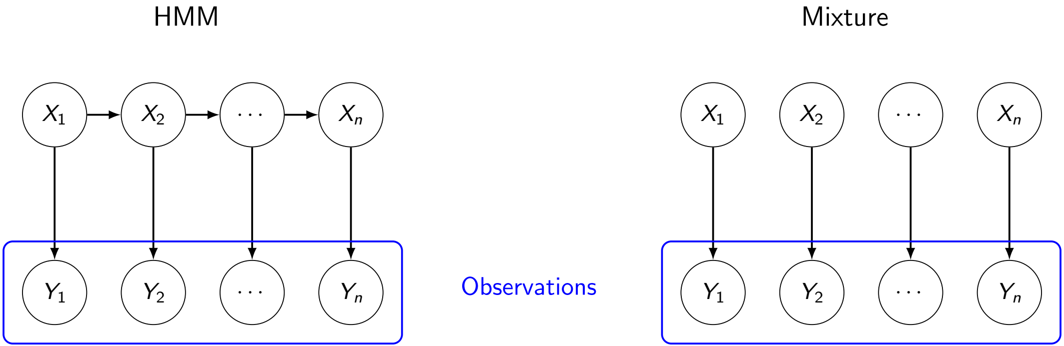
HMM



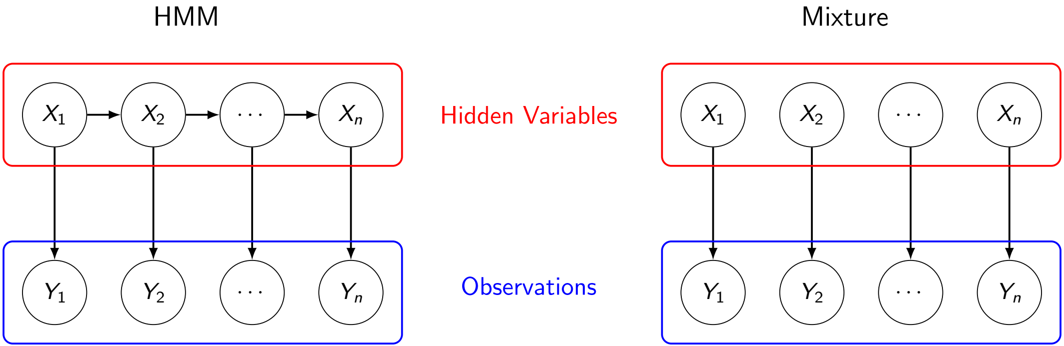
Mixture



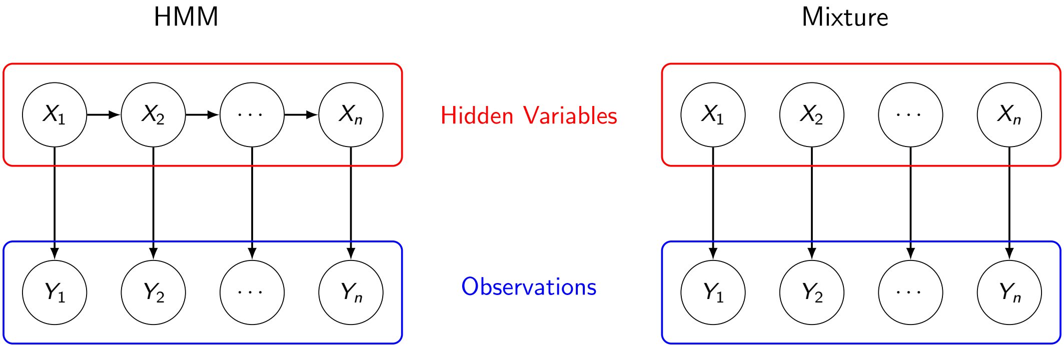
What is Clustering?



What is Clustering?

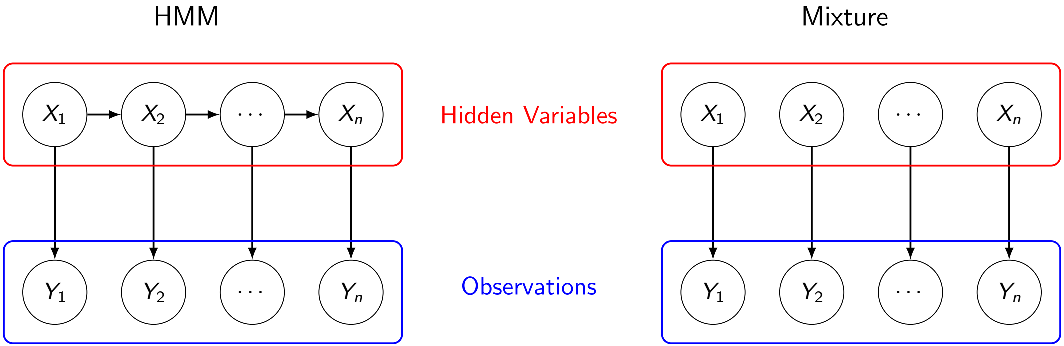


What is Clustering?



Clustering : Identifying the partition induced by the Hidden variables

What is Clustering?



Clustering : Identifying the partition induced by the Hidden variables

Warning

Clustering does not seek to identify the values of the hidden states!

Distance between Partitions

$$\ell(A, B) = \inf_{\substack{M \subseteq \mathcal{E}(A,B) \\ M \text{ is a matching}}} \left(1 - \frac{1}{n} \sum_{\{C, C'\} \in M} |C \cap C'| \right)$$

Example

Let $S = \{1, 2, 3, 4, 5\}$.

Clustering A	**Clustering** B
$C_1 = \{1, 2, 3\}$	$C'_1 = \{1, 2, 4\}$
$C_2 = \{4, 5\}$	$C'_2 = \{3, 5\}$

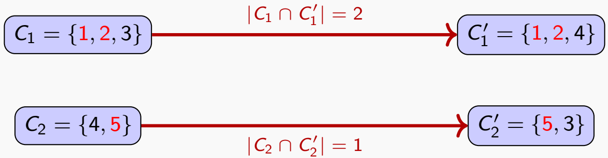
Distance between Partitions

$$\ell(A,B) = \inf_{\substack{M \subseteq \mathcal{E}(A,B) \\ M \text{ is a matching}}} \left(1 - \frac{1}{n} \sum_{\{C,C'\} \in M} |C \cap C'| \right)$$

Example

Let $S = \{1,2,3,4,5\}$.

Clustering A **Clustering** B



Matching $= \{\{C_1, C_1'\}, \{C_2, C_2'\}\}$

Value of loss $= 1 - \frac{2+1}{5} = 0.4$

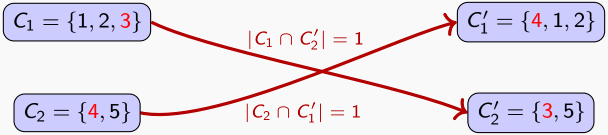
Distance between Partitions

$$\ell(A, B) = \inf_{\substack{M \subseteq \mathcal{E}(A,B) \\ M \text{ is a matching}}} \left(1 - \frac{1}{n} \sum_{\{C, C'\} \in M} |C \cap C'| \right)$$

Example

Let $S = \{1, 2, 3, 4, 5\}$.

Clustering A **Clustering** B



Matching $= \{\{C_1, C'_2\}, \{C_2, C'_1\}\}$

Value of loss $= 1 - \frac{1+1}{5} = 0.6$

Distance between Partitions

$$\ell(A, B) = \inf_{\substack{M \subseteq \mathcal{E}(A,B) \\ M \text{ is a matching}}} \left(1 - \frac{1}{n} \sum_{\{C, C'\} \in M} |C \cap C'| \right)$$

Example

Let $S = \{1, 2, 3, 4, 5\}$.

Clustering A

$C_1 = \{1, 2, 3\}$

$C_2 = \{4, 5\}$

Clustering B

$C'_1 = \{1, 2, 4\}$

$C'_2 = \{3, 5\}$

$|C_1 \cap C'_1| = 2$

Matching $= \{\{C_1, C'_1\}\}$

Value of loss $= 1 - \frac{2}{5} = 0.6$

Distance between Partitions

$$\ell(A, B) = \inf_{\substack{M \subseteq \mathcal{E}(A,B) \\ M \text{ is a matching}}} \left(1 - \frac{1}{n} \sum_{\{C, C'\} \in M} |C \cap C'|\right)$$

Example

Let $S = \{1, 2, 3, 4, 5\}$.

Clustering A

$C_1 = \{1, 2, 3\}$

$C_2 = \{4, 5\}$

Clustering B

$C'_1 = \{4, 1, 2\}$

$C'_2 = \{3, 5\}$

$|C_2 \cap C'_1| = 1$

Matching $= \{\{C_2, C'_1\}\}$

Value of loss $= 1 - \frac{1}{5} = 0.8$

Distance between Partitions

$$\ell(A,B) = \inf_{\substack{M \subseteq \mathcal{E}(A,B) \\ M \text{ is a matching}}} \left(1 - \frac{1}{n} \sum_{\{C,C'\} \in M} |C \cap C'| \right)$$

Example

Let $S = \{1,2,3,4,5\}$.

Clustering A **Clustering** B

$C_1 = \{1,2,3\}$ → $C_1' = \{1,2,4\}$, $|C_1 \cap C_1'| = 2$

$C_2 = \{4,5\}$ → $C_2' = \{5,3\}$, $|C_2 \cap C_2'| = 1$

Best matching $= \{\{C_1, C_1'\}, \{C_2, C_2'\}\}$

$$\ell(A,B) = 0.4$$

Risk of Clustering

A **clusterer** g maps the observation $(Y_1, \ldots, Y_n)$ to a partition of $[\![1, n]\!]$.

Risk of Clustering

A **clusterer** g maps the observation $(Y_1, \dots, Y_n)$ to a partition of $[\![1, n]\!]$.

Partition of the clusterer g

$$\ell\big(g(Y_1, \dots, Y_n), \quad\big)$$

Risk of Clustering

A **clusterer** g maps the observation $(Y_1, \ldots, Y_n)$ to a partition of $[\![1, n]\!]$.

Partition of the clusterer g

$$\ell\big(g(Y_1, \ldots, Y_n), \Pi_n\big)$$

Partition induced by hidden states $X_1, \ldots, X_n$

Risk of Clustering

A **clusterer** g maps the observation $(Y_1, \dots, Y_n)$ to a partition of $[\![1, n]\!]$.

Partition of the clusterer g

$$\mathbb{E}_\theta\left[\ell\big(g(Y_1, \dots, Y_n), \Pi_n\big)\right]$$

Mixture or HMM

Partition induced by hidden states $X_1, \dots, X_n$

Risk of Clustering

A **clusterer** g maps the observation $(Y_1, \dots, Y_n)$ to a partition of $[\![1, n]\!]$.

$$\mathcal{R}_n^{\text{clust}}(\theta, g) := \mathbb{E}_\theta\left[\ell\big(g(Y_1, \dots, Y_n), \Pi_n\big)\right]$$

Quantity of interest: The Bayes risk of clustering $\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g)$

Goal

- Characterize the minimizer of the clustering risk
- Quantify the magnitude of the Bayes risk
- Design some clustering procedures that perform well in the general **non-parametric** setting

What is Classification?

A **classifier** h maps the observation $(Y_1, \ldots, Y_n)$ to an element of $[\![1, K]\!]^n$.

What is Classification?

A **classifier** h maps the observation $(Y_1, \dots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

What is Classification?

A **classifier** h maps the observation $(Y_1, \ldots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

Sequence of estimated hidden states

$$\mathrm{H}\big(h(Y_1, \ldots, Y_n), \qquad \big)$$

What is Classification?

A **classifier** h maps the observation $(Y_1, \ldots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}(x, y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

Sequence of estimated hidden states

$$\mathrm{H}\big(h(Y_1, \ldots, Y_n), (X_1, \ldots, X_n)\big)$$

True hidden states $X_1, \ldots, X_n$

What is Classification?

A **classifier** h maps the observation $(Y_1, \ldots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}(x, y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

Sequence of estimated hidden states

$$\mathbb{E}_\theta\left[\mathrm{H}\big(h(Y_1, \ldots, Y_n), (X_1, \ldots, X_n)\big)\right]$$

Mixture or HMM

True hidden states $X_1, \ldots, X_n$

What is Classification?

A **classifier** h maps the observation $(Y_1, \ldots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}\left(x, y\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

$$\mathcal{R}_n^{\mathrm{class}}(\theta, h) := \mathbb{E}_\theta\left[\mathrm{H}\big(h(Y_1, \ldots, Y_n), (X_1, \ldots, X_n)\big)\right]$$

What is Classification?

A **classifier** h maps the observation $(Y_1, \ldots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

$$\mathcal{R}_n^{\text{class}}(\theta, h) := \mathbb{E}_\theta \left[\mathrm{H}\big(h(Y_1, \ldots, Y_n), (X_1, \ldots, X_n)\big) \right]$$

To any classifier h corresponds a unique clusterer g_h such that

$$g_h(Y_1, \ldots, Y_n) = \big\{ \{ i \in [\![1, n]\!] : h_i(Y_1, \ldots, Y_n) = x \} : x \in [\![1, K]\!] \big\}.$$

What is Classification?

A **classifier** h maps the observation $(Y_1, \ldots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

$$\mathcal{R}_n^{\mathrm{class}}(\theta, h) := \mathbb{E}_\theta \left[\mathrm{H}\big(h(Y_1, \ldots, Y_n), (X_1, \ldots, X_n)\big) \right]$$

To any classifier h corresponds a unique clusterer g_h such that

$$g_h(Y_1, \ldots, Y_n) = \big\{\{i \in [\![1, n]\!] : h_i(Y_1, \ldots, Y_n) = x\} : x \in [\![1, K]\!]\big\}.$$

Lemma

Let $\mathcal{S}_K$ be the group of permutations of $[\![1, K]\!]$. For $\theta \in \Theta^{\mathrm{dep}} \cup \Theta^{\mathrm{ind}}$ and any classifier $h = (h_i)_{1 \leq i \leq n}$,

$$\mathcal{R}_n^{\mathrm{class}}(\theta, h) = \mathbb{E}_\theta \left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \neq h_i(Y_1, \ldots, Y_n)_i} \right]; \quad \mathcal{R}_n^{\mathrm{clust}}(\theta, g_h) = \mathbb{E}_\theta \left[\min_{\tau \in \mathcal{S}_K} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\tau(X_i) \neq h_i(Y_1, \ldots, Y_n)} \right].$$

What is Classification?

A **classifier** h maps the observation $(Y_1, \dots, Y_n)$ to an element of $[\![1, K]\!]^n$.

Let H be the Hamming distance defined by: $\mathrm{H}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i \neq y_i}$.

$$\mathcal{R}_n^{\text{class}}(\theta, h) := \mathbb{E}_\theta \left[\mathrm{H}\big(h(Y_1, \dots, Y_n), (X_1, \dots, X_n)\big) \right]$$

To any classifier h corresponds a unique clusterer g_h such that

$$g_h(Y_1, \dots, Y_n) = \big\{ \{ i \in [\![1, n]\!] : h_i(Y_1, \dots, Y_n) = x \} : x \in [\![1, K]\!] \big\}.$$

Lemma

Let $\mathcal{S}_K$ be the group of permutations of $[\![1, K]\!]$. For $\theta \in \Theta^{\text{dep}} \cup \Theta^{\text{ind}}$ and any classifier $h = (h_i)_{1 \leq i \leq n}$,

$$\mathcal{R}_n^{\text{class}}(\theta, h) = \mathbb{E}_\theta \left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \neq h_i(Y_1, \dots, Y_n)_i} \right]; \quad \mathcal{R}_n^{\text{clust}}(\theta, g_h) = \mathbb{E}_\theta \left[\min_{\tau \in \mathcal{S}_K} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\tau(X_i) \neq h_i(Y_1, \dots, Y_n)} \right].$$

Classification is simpler to study, but how to relate both notions ?

Outline

Bayes Classifier vs Bayes Clusterer ($K > 2$)

- $h_\theta^\star$ is the Bayes classifier, minimizer of $h \mapsto \mathcal{R}_n^{\text{class}}(\theta, h)$
- $g_\theta^\star$ is the Bayes clusterer, minimizer of $g \mapsto \mathcal{R}_n^{\text{clust}}(\theta, g)$
- Recall that $g_h(Y_1, \ldots, Y_n) = \big\{\{i \in [\![1, n]\!] : h_i(Y_1, \ldots, Y_n) = x\} : x \in [\![1, K]\!]\big\}$

Bayes Classifier vs Bayes Clusterer ($K > 2$)

- $h^\star_\theta$ is the Bayes classifier, minimizer of $h \mapsto \mathcal{R}_n^{\mathrm{class}}(\theta, h)$
- $g^\star_\theta$ is the Bayes clusterer, minimizer of $g \mapsto \mathcal{R}_n^{\mathrm{clust}}(\theta, g)$
- Recall that $g_h(Y_1, \dots, Y_n) = \big\{\{i \in [\![1, n]\!] : h_i(Y_1, \dots, Y_n) = x\} : x \in [\![1, K]\!]\big\}$

Theorem (Gassiat, Kaddouri and Naulet (2025))

When $K > 2$, then for all $\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}}$ such that $f = (f_i)_{i \in [\![1,K]\!]}$, if

$$\mathbb{P}_\theta\left(\bigcup_{k=1}^{K}\left\{0 < \max_{l \neq k} \pi_l f_l(Y) < \pi_k f_k(Y) \leq \sum_{l \neq k} \pi_l f_l(Y)\right\}\right) > 0$$

then,

$$\forall n \geq 2, \quad \mathbb{P}_\theta\left(g^\star_\theta(Y_1, \dots, Y_n) \neq g_{h^\star_\theta}(Y_1, \dots, Y_n)\right) > 0.$$

Bayes Classifier vs Bayes Clusterer ($K > 2$)

- $h^\star_\theta$ is the Bayes classifier, minimizer of $h \mapsto \mathcal{R}_n^{\mathrm{class}}(\theta, h)$
- $g^\star_\theta$ is the Bayes clusterer, minimizer of $g \mapsto \mathcal{R}_n^{\mathrm{clust}}(\theta, g)$
- Recall that $g_h(Y_1, \dots, Y_n) = \big\{\{i \in [\![1, n]\!] : h_i(Y_1, \dots, Y_n) = x\} : x \in [\![1, K]\!]\big\}$

Theorem (Gassiat, Kaddouri and Naulet (2025))

When $K > 2$, then for all $\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}}$ such that $f = (f_i)_{i \in [\![1,K]\!]}$, if

$$\mathbb{P}_\theta\left(\bigcup_{k=1}^{K}\left\{0 < \max_{l \neq k} \pi_l f_l(Y) < \pi_k f_k(Y) \leq \sum_{l \neq k} \pi_l f_l(Y)\right\}\right) > 0$$

then,

$$\forall n \geq 2, \quad \mathbb{P}_\theta\left(g^\star_\theta(Y_1, \dots, Y_n) \neq g_{h^\star_\theta}(Y_1, \dots, Y_n)\right) > 0.$$

Theorem (Gassiat, Kaddouri and Naulet (2025))

In the case of dependent labels, for all $K > 2$ and all $n \geq 2$, there exists a subset $\tilde{\Theta}_{n,K} \subset \Theta^{\mathrm{dep}}$ such that

$$\forall \theta \in \tilde{\Theta}_{n,K}, \quad \mathbb{P}_\theta\left(g^\star_\theta(Y_1, \dots, Y_n) \neq g_{h^\star_\theta}(Y_1, \dots, Y_n)\right) > 0.$$

Bayes Classifier vs Bayes Clusterer ($K = 2$)

- $h^\star_\theta$ is the Bayes classifier, minimizer of $h \mapsto \mathcal{R}_n^{\mathrm{class}}(\theta, h)$
- $g^\star_\theta$ is the Bayes clusterer, minimizer of $g \mapsto \mathcal{R}_n^{\mathrm{clust}}(\theta, g)$
- Recall that $g_h(Y_1, \ldots, Y_n) = \big\{\{i \in [\![1, n]\!] : h_i(Y_1, \ldots, Y_n) = x\} : x \in [\![1, K]\!]\big\}$

Bayes Classifier vs Bayes Clusterer ($K = 2$)

- $h^\star_\theta$ is the Bayes classifier, minimizer of $h \mapsto \mathcal{R}_n^{\mathrm{class}}(\theta, h)$
- $g^\star_\theta$ is the Bayes clusterer, minimizer of $g \mapsto \mathcal{R}_n^{\mathrm{clust}}(\theta, g)$
- Recall that $g_h(Y_1, \ldots, Y_n) = \big\{\{i \in [\![1, n]\!] : h_i(Y_1, \ldots, Y_n) = x\} : x \in [\![1, K]\!]\big\}$

Theorem (Gassiat, Kaddouri and Naulet (2025))

In the case of dependent labels and when $K = 2$, there exists a subset $\tilde{\Theta} \subset \Theta^{\mathrm{dep}}$ such that

$$\left(\forall \theta \in \tilde{\Theta}\right) (\forall n \geq 2)\,, \quad \mathbb{P}_\theta\left(g^\star_\theta(Y_1, \ldots, Y_n) \neq g_{h^\star_\theta}(Y_1, \ldots, Y_n)\right) > 0.$$

Bayes Classifier vs Bayes Clusterer ($K = 2$)

- $h_\theta^\star$ is the Bayes classifier, minimizer of $h \mapsto \mathcal{R}_n^{\text{class}}(\theta, h)$
- $g_\theta^\star$ is the Bayes clusterer, minimizer of $g \mapsto \mathcal{R}_n^{\text{clust}}(\theta, g)$
- Recall that $g_h(Y_1, \ldots, Y_n) = \{\{i \in [\![1, n]\!] : h_i(Y_1, \ldots, Y_n) = x\} : x \in [\![1, K]\!]\}$

Theorem (Gassiat, Kaddouri and Naulet (2025))

In the case of dependent labels and when $K = 2$, there exists a subset $\tilde{\Theta} \subset \Theta^{\text{dep}}$ such that

$$\left(\forall \theta \in \tilde{\Theta}\right) (\forall n \geq 2), \quad \mathbb{P}_\theta \left(g_\theta^\star(Y_1, \ldots, Y_n) \neq g_{h_\theta^\star}(Y_1, \ldots, Y_n)\right) > 0.$$

Theorem (Gassiat, Kaddouri and Naulet (2025))

In the case of independent labels, if $K = 2$, then for all $\theta \in \Theta^{\text{ind}}$ and all $n \geq 2$,

$$g_\theta^\star(Y_1, \ldots, Y_n) = g_{h_\theta^\star}(Y_1, \ldots, Y_n) \quad \mathbb{P}_\theta\text{-}a.s.$$

Bayes Risk: Clustering vs Classification (I.I.D. case)

Now we intend to relate $\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g)$ to $\inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta, h)$.

Bayes Risk: Clustering vs Classification (I.I.D. case)

Now we intend to relate $\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g)$ to $\inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h)$.

Theorem (Gassiat, Kaddouri and Naulet (2025))

If $K = 2$, then for all $\theta \in \Theta^{\text{ind}}$ and $n \geq 2$,

$$(1 - \alpha_n) \inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g) \leq \inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h)$$

where $\alpha_n = 2e\sqrt{\log(2)/2n}$.

Bayes Risk: Clustering vs Classification (I.I.D. case)

Now we intend to relate $\inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g)$ to $\inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h)$.

Theorem (Gassiat, Kaddouri and Naulet (2025))

If $K = 2$, then for all $\theta \in \Theta^{\text{ind}}$ and $n \geq 2$,

$$(1-\alpha_n) \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h) \leq \inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g) \leq \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h)$$

where $\alpha_n = 2e\sqrt{\log(2)/2n}$.

Theorem (Gassiat, Kaddouri and Naulet (2025))

If $K > 2$ then for all $\theta = (K, \pi, F) \in \Theta^{\text{ind}}$ and all $n \geq 1$,

$$(1-\xi_n) \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h) - K^2 e^{-n\beta/8} \leq \inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g) \leq \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h)$$

where $\beta = \min_{j\neq k}(\pi_j + \pi_k)$ and $\xi_n = \frac{4e}{\beta}\sqrt{\log(K!)/(2n)}$.

Bayes Risk: Clustering vs Classification (Dependent case)

Now we intend to relate $\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g)$ to $\inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h)$.

Let $\Theta_\delta^{\text{dep}} = \left\{\theta = (K, \pi, Q, f) \in \Theta^{\text{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta, \quad \min_j \pi_j \geq \delta\right\}$.

Bayes Risk: Clustering vs Classification (Dependent case)

Now we intend to relate $\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g)$ to $\inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h)$.

Let $\Theta_\delta^{\text{dep}} = \left\{\theta = (K, \pi, Q, f) \in \Theta^{\text{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta, \quad \min_j \pi_j \geq \delta\right\}$.

Theorem (Gassiat, Kaddouri and Naulet (2025))

If $K = 2$, then for all $\theta \in \Theta_\delta^{\text{dep}}$ and all $n \geq 1$

$$(1 - \tilde{\alpha}_n) \inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g) \leq \inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\text{class}}(\theta, h),$$

where $\tilde{\alpha}_n = 2e\left(\frac{1-\delta}{\delta}\right)^5 \sqrt{\log(2)/2n}$.

Bayes Risk: Clustering vs Classification (Dependent case)

Now we intend to relate $\inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g)$ to $\inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta, h)$.

Let $\Theta_\delta^{\mathrm{dep}} = \{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta, \quad \min_j \pi_j \geq \delta\}$.

Theorem (Gassiat, Kaddouri and Naulet (2025))

If $K = 2$, then for all $\theta \in \Theta_\delta^{\mathrm{dep}}$ and all $n \geq 1$

$$(1 - \tilde{\alpha}_n) \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta, h) \leq \inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta, h),$$

where $\tilde{\alpha}_n = 2e\left(\frac{1-\delta}{\delta}\right)^5 \sqrt{\log(2)/2n}$.

Theorem (Gassiat, Kaddouri and Naulet (2025))

If $K > 2$, then for all $\theta \in \Theta_\delta^{\mathrm{dep}}$ and all $n \geq 1$,

$$(1 - \tilde{\xi}_n) \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta, h) - K^2 e^{-n(1-\rho_0)^2\beta^2} \leq \inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta, h),$$

where $\beta = \min_{i,j\neq k} \mathbb{P}_\theta(X_i \in \{j, k\})$, $\rho_0 = \frac{1-K\delta}{1-(K-1)\delta}$ and $\tilde{\xi}_n = \frac{5}{\beta(1-\rho_0)}\sqrt{\log(K!)/(2n)}$.

Magnitude of the Bayes risk of clustering ($K = 2$)

- $\Theta_\delta^{\mathrm{ind}} = \left\{\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}} \mid \min_j \pi_j \geq \delta\right\}$

Magnitude of the Bayes risk of clustering ($K = 2$)

- $\Theta_\delta^{\mathrm{ind}} = \left\{\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}} \mid \min_j \pi_j \geq \delta\right\}$
- $\Theta_\delta^{\mathrm{dep}} = \left\{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\right\}$

Magnitude of the Bayes risk of clustering ($K = 2$)

- $\Theta_\delta^{\mathrm{ind}} = \left\{\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}} \mid \min_j \pi_j \geq \delta\right\}$
- $\Theta_\delta^{\mathrm{dep}} = \left\{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\right\}$

For $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i \in [\![1,K]\!]}$, we define:

$$\Lambda(\theta) := \int \min_{j \in [\![1,K]\!]} \Big(\sum_{i \neq j} f_i(y) \Big)\, dy.$$

Magnitude of the Bayes risk of clustering ($K = 2$)

- $\Theta_\delta^{\mathrm{ind}} = \{\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}} \mid \min_j \pi_j \geq \delta\}$
- $\Theta_\delta^{\mathrm{dep}} = \{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\}$

For $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i \in [\![1,K]\!]}$, we define:

$$\Lambda(\theta) := \int \min_{j \in [\![1,K]\!]} \Big(\sum_{i \neq j} f_i(y) \Big)\, dy.$$

Theorem (Gassiat, Kaddouri and Naulet (2025))

- *When $K = 2$:*

$$\forall \theta \in \Theta_\delta^{\mathrm{ind}}, \quad (1 - \alpha_n)\delta \Lambda(\theta) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq (1 - \delta)\Lambda(\theta),$$

$$\forall \theta \in \Theta_\delta^{\mathrm{dep}}, \quad \frac{\delta^2(1 - \tilde{\alpha}_n)}{1 - \delta}\Lambda(\theta) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq (1 - \delta)\Lambda(\theta),$$

where $\alpha_n = 2e\sqrt{\log(2)/2n}$ and $\tilde{\alpha}_n = 2e\left(\frac{1-\delta}{\delta}\right)^5 \sqrt{\log(2)/2n}$.

Magnitude of the Bayes risk of clustering ($K > 2$)

- $\Theta_\delta^{\mathrm{ind}} = \left\{\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}} \mid \min_j \pi_j \geq \delta\right\}$

Magnitude of the Bayes risk of clustering ($K > 2$)

- $\Theta_\delta^{\mathrm{ind}} = \{\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}} \mid \min_j \pi_j \geq \delta\}$
- $\Theta_\delta^{\mathrm{dep}} = \{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\}$

Magnitude of the Bayes risk of clustering ($K > 2$)

- $\Theta_\delta^{\text{ind}} = \left\{\theta = (K, \pi, f) \in \Theta^{\text{ind}} \mid \min_j \pi_j \geq \delta\right\}$
- $\Theta_\delta^{\text{dep}} = \left\{\theta = (K, \pi, Q, f) \in \Theta^{\text{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\right\}$

For $\theta = (K, \pi, f) \in \Theta_\delta^{\text{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\text{dep}}$, where $f = (f_i)_{i \in [\![1,K]\!]}$, we define:

$$\Lambda(\theta) := \int \min_{j \in [\![1,K]\!]} \Big(\sum_{i \neq j} f_i(y) \Big)\, dy.$$

Magnitude of the Bayes risk of clustering ($K > 2$)

- $\Theta_\delta^{\mathrm{ind}} = \{\theta = (K, \pi, f) \in \Theta^{\mathrm{ind}} \mid \min_j \pi_j \geq \delta\}$
- $\Theta_\delta^{\mathrm{dep}} = \{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\}$

For $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i \in [\![1,K]\!]}$, we define:

$$\Lambda(\theta) := \int \min_{j \in [\![1,K]\!]} \Big(\sum_{i \neq j} f_i(y) \Big) \, dy.$$

Theorem (Gassiat, Kaddouri and Naulet (2025))

- *When $K > 2$ and $\theta \in \Theta_\delta^{\mathrm{ind}}$ satisfies $\delta \Lambda(\theta) \geq 4K^2 e^{-n\beta/8}$ and n large enough so that $\xi_n \leq \frac{1}{2}$:*

$$\frac{\delta}{4} \Lambda(\theta) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq (1 - (K-1)\delta) \Lambda(\theta),$$

- *When $K > 2$ and $\theta \in \Theta_\delta^{\mathrm{dep}}$ satisfies $\delta^2 \Lambda(\theta) \geq 4K^2 e^{-2n(1-\rho_0)^2 \beta^2}$ and n large enough so that $\tilde{\xi}_n \leq \frac{1}{2}$:*

$$\frac{\delta^2}{4(1-\delta)} \Lambda(\theta) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq (1 - (K-1)\delta) \Lambda(\theta),$$

where $\beta = \min_{i, j \neq k} \mathbb{P}_\theta (X_i \in \{j, k\})$ and $\rho_0 = \frac{1-K\delta}{1-(K-1)\delta}$, $\tilde{\xi}_n = \frac{5}{\beta(1-\rho_0)} \sqrt{\log(K!)/(2n)}$.

Illustration (Case $K = 2$)

Let $\delta > 0$ and $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i \in \{1,2\}}$,

$$\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$$

Main result

$\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \asymp \int f_1(y) \wedge f_2(y)\, dy$

Illustration (Case $K = 2$)

Let $\delta > 0$ and $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i \in \{1,2\}}$,

$$\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$$

Main result

$$\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \asymp \int f_1(y) \wedge f_2(y)\, dy$$

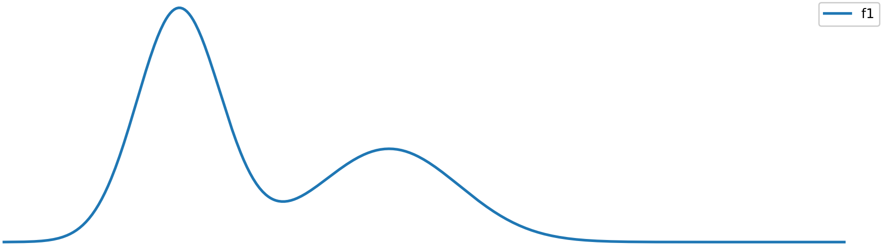


Illustration (Case $K = 2$)

Let $\delta > 0$ and $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i \in \{1,2\}}$,

$$\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$$

Main result

$\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \asymp \int f_1(y) \wedge f_2(y)\, dy$

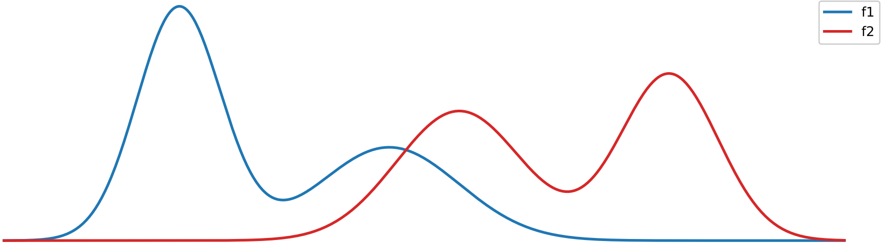


Illustration (Case $K = 2$)

Let $\delta > 0$ and $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i\in\{1,2\}}$,

$$\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$$

Main result

$\inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \asymp \int f_1(y) \wedge f_2(y)\, dy$

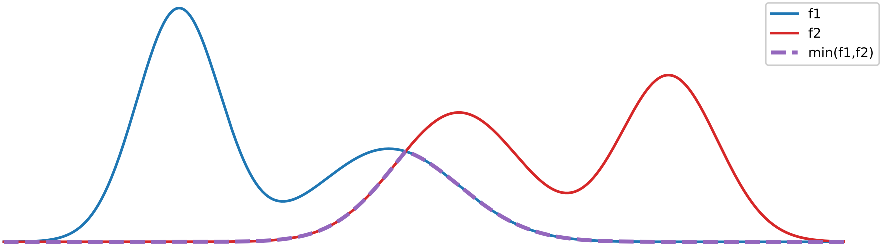


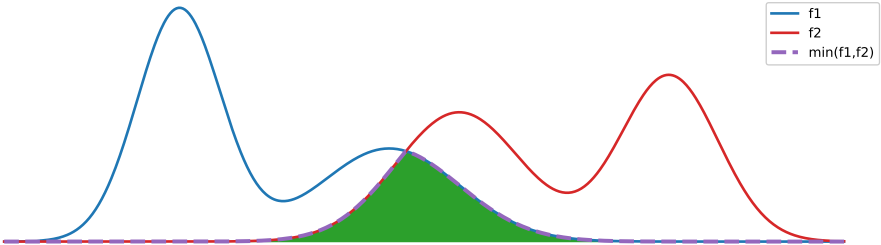
Illustration (Case $K = 2$)

Let $\delta > 0$ and $\theta = (K, \pi, f) \in \Theta_\delta^{\mathrm{ind}}$ or $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$, where $f = (f_i)_{i \in \{1,2\}}$,

$$\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$$

Main result

$\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \asymp \int f_1(y) \wedge f_2(y)\, dy$



Performance of the Bayes classifier for clustering?

- ✅ The HMM model is **identifiable** without any shape restriction

Performance of the Bayes classifier for clustering?

- The HMM model is **identifiable** without any shape restriction
- Many estimators of $\theta = (K, \pi, Q, f)$ have been proposed in the literature Anandkumar et al. (2012); Gassiat et al. (2016); De Castro et al. (2017); Abraham et al. (2022)

Performance of the Bayes classifier for clustering?

- The HMM model is **identifiable** without any shape restriction
- Many estimators of $\theta = (K, \pi, Q, f)$ have been proposed in the literature Anandkumar et al. (2012); Gassiat et al. (2016); De Castro et al. (2017); Abraham et al. (2022)
- $h^\star_\theta$ has an explicit expression $h^\star_\theta(Y_1, \dots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_\theta(X_i = a \mid Y_1, \dots, Y_n)\right)$

Performance of the Bayes classifier for clustering?

- The HMM model is **identifiable** without any shape restriction
- Many estimators of $\theta = (K, \pi, Q, f)$ have been proposed in the literature Anandkumar et al. (2012); Gassiat et al. (2016); De Castro et al. (2017); Abraham et al. (2022)
- $h^\star_\theta$ has an explicit expression $h^\star_\theta(Y_1, \ldots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_\theta(X_i = a \mid Y_1, \ldots, Y_n)\right)$
- A natural classifier is the plug-in classifier: $h^\star_{\hat\theta}(Y_1, \ldots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_{\hat\theta}(X_i = a \mid Y_1, \ldots, Y_n)\right)$

Performance of the Bayes classifier for clustering?

- The HMM model is **identifiable** without any shape restriction
- Many estimators of $\theta = (K, \pi, Q, f)$ have been proposed in the literature Anandkumar et al. (2012); Gassiat et al. (2016); De Castro et al. (2017); Abraham et al. (2022)
- $h^\star_\theta$ has an explicit expression $h^\star_\theta(Y_1, \dots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_\theta(X_i = a \mid Y_1, \dots, Y_n)\right)$
- A natural classifier is the plug-in classifier: $h^\star_{\hat\theta}(Y_1, \dots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_{\hat\theta}(X_i = a \mid Y_1, \dots, Y_n)\right)$

What is the performance of $g_{h^\star_{\hat\theta}}$?

Performance of the Bayes classifier for clustering?

- The HMM model is **identifiable** without any shape restriction
- Many estimators of $\theta = (K, \pi, Q, f)$ have been proposed in the literature Anandkumar et al. (2012); Gassiat et al. (2016); De Castro et al. (2017); Abraham et al. (2022)
- $h^\star_\theta$ has an explicit expression $h^\star_\theta(Y_1, \dots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_\theta(X_i = a \mid Y_1, \dots, Y_n)\right)$
- A natural classifier is the plug-in classifier: $h^\star_{\hat\theta}(Y_1, \dots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_{\hat\theta}(X_i = a \mid Y_1, \dots, Y_n)\right)$

What is the performance of $g_{h^\star_{\hat\theta}}$?

Theorem (Gassiat, Kaddouri and Naulet (2025))

Let $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$ *where* $f = (f_i)_{i \in [\![1,K]\!]}$. *Assume*

- $(f_i)_{i \in [\![1,K]\!]}$ *are compactly supported and* $C^\star = \int \frac{dy}{\sum_{i \in [\![1,K]\!]} f_i(y)} < \infty$
- $(f_i)_{i \in [\![1,K]\!]}$ *are linearly independent and belong to* $C^s(\mathbb{R})$
- *Assume the HMM is stationary, the matrix Q is full-rank and aperiodic*

Performance of the Bayes classifier for clustering?

- The HMM model is **identifiable** without any shape restriction
- Many estimators of $\theta = (K, \pi, Q, f)$ have been proposed in the literature Anandkumar et al. (2012); Gassiat et al. (2016); De Castro et al. (2017); Abraham et al. (2022)
- $h^\star_\theta$ has an explicit expression $h^\star_\theta(Y_1, \ldots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_\theta(X_i = a \mid Y_1, \ldots, Y_n)\right)$
- A natural classifier is the plug-in classifier: $h^\star_{\hat\theta}(Y_1, \ldots, Y_n) = \left(\arg\max_{a \in [\![1,K]\!]} \mathbb{P}_{\hat\theta}(X_i = a \mid Y_1, \ldots, Y_n)\right)$

What is the performance of $g_{h^\star_{\hat\theta}}$?

Theorem (Gassiat, Kaddouri and Naulet (2025))

Let $\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}}_\delta$ where $f = (f_i)_{i \in [\![1,K]\!]}$. Assume

- *$(f_i)_{i \in [\![1,K]\!]}$ are compactly supported and $C^\star = \int \frac{dy}{\sum_{i \in [\![1,K]\!]} f_i(y)} < \infty$*
- *$(f_i)_{i \in [\![1,K]\!]}$ are linearly independent and belong to $C^s(\mathbb{R})$*
- *Assume the HMM is stationary, the matrix Q is full-rank and aperiodic*

There exists a sequence of randomized estimators $(\hat\theta_n)_{n \geq 1}$ such that

$$\mathbb{E}_\theta[\mathcal{R}^{\mathrm{clust}}_n(\theta, g_{h^\star_{\hat\theta}})] - \inf_{g \in \mathcal{G}_n} \mathcal{R}^{\mathrm{clust}}_n(\theta, g) = \mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{\frac{s}{2s+1}}\right).$$

Illustration

- **First example:** A sample of size $n = 5 \cdot 10^4$ generated from two gaussian mixtures: $\frac{1}{2}\left(\mathcal{N}(1.7, 0.2) + \mathcal{N}(7, 0.15)\right)$ and $\frac{1}{2}\left(\mathcal{N}(3.5, 0.2) + \mathcal{N}(5, 0.4)\right)$, and $Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$.

Illustration

- **First example:** A sample of size $n = 5 \cdot 10^4$ generated from two gaussian mixtures: $\frac{1}{2}\left(\mathcal{N}(1.7, 0.2) + \mathcal{N}(7, 0.15)\right)$ and $\frac{1}{2}\left(\mathcal{N}(3.5, 0.2) + \mathcal{N}(5, 0.4)\right)$, and $Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$.

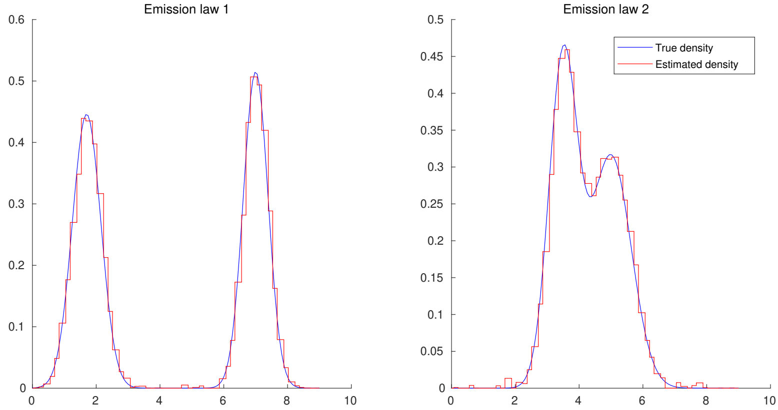
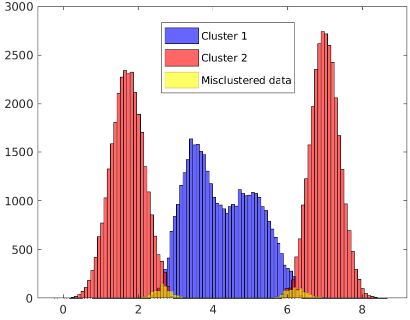


Figure: Least squares density estimation

Illustration

- **First example:** A sample of size $n = 5 \cdot 10^4$ generated from two gaussian mixtures: $\frac{1}{2}\left(\mathcal{N}(1.7, 0.2) + \mathcal{N}(7, 0.15)\right)$ and $\frac{1}{2}\left(\mathcal{N}(3.5, 0.2) + \mathcal{N}(5, 0.4)\right)$, and $Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$.



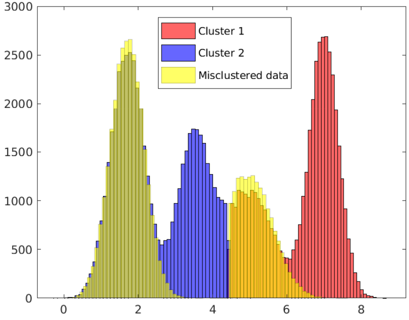


Figure: Histograms of the clusters. Left: plug-in classifier. Right: K-means.

Illustration

- **Second example:** A sample of size $n = 10^5$ generated from two gaussian mixtures : $\frac{1}{2}\left(\mathcal{N}(3, 0.6) + \mathcal{N}(7, 0.4)\right)$ and $\frac{1}{2}\left(\mathcal{N}(5, 0.3) + \mathcal{N}(9, 0.4)\right)$ and $Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$.

Illustration

- **Second example:** A sample of size $n = 10^5$ generated from two gaussian mixtures : $\frac{1}{2}\left(\mathcal{N}(3,0.6)+\mathcal{N}(7,0.4)\right)$ and $\frac{1}{2}\left(\mathcal{N}(5,0.3)+\mathcal{N}(9,0.4)\right)$ and $Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$.

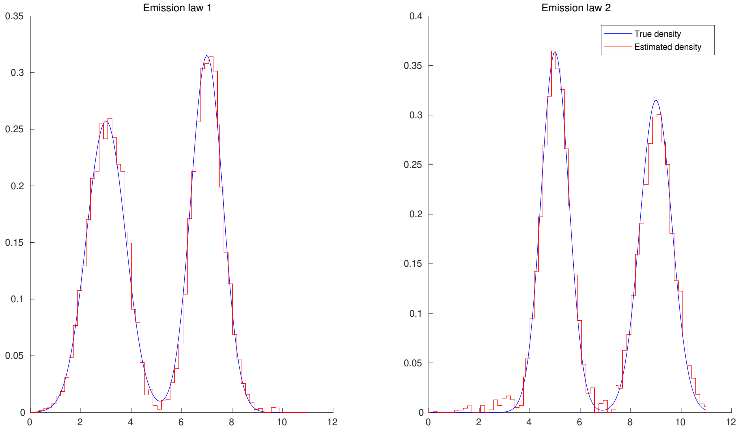


Figure: Least squares density estimation

Illustration

- **Second example:** A sample of size $n = 10^5$ generated from two gaussian mixtures :
  $\frac{1}{2}\left(\mathcal{N}(3, 0.6) + \mathcal{N}(7, 0.4)\right)$ and $\frac{1}{2}\left(\mathcal{N}(5, 0.3) + \mathcal{N}(9, 0.4)\right)$ and $Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$.

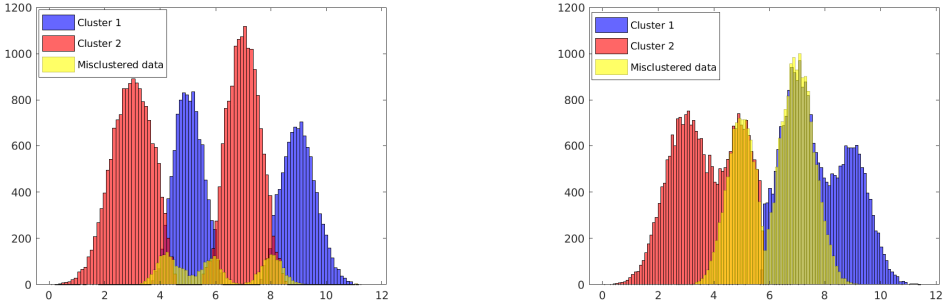


Figure: Histograms of the clusters. Left: plug-in classifier. Right: K-means.

Outline

Slowly mixing regime

- $\Theta_\delta^{\mathrm{dep}} = \left\{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\right\}$

Theorem (Gassiat, Kaddouri and Naulet (2025))

- *When $K=2$:*

$$\forall \theta \in \Theta_\delta^{\mathrm{dep}}, \quad \frac{\delta^2(1-\tilde{\alpha}_n)}{1-\delta}\Lambda(\theta) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq (1-\delta)\Lambda(\theta),$$

where

  - $\tilde{\alpha}_n = 2e\left(\frac{1-\delta}{\delta}\right)^5 \sqrt{\log(2)/2n}$
  - $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$
  - $f = (f_i)_{i \in \{1,2\}}$ *and* $\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$

Slowly mixing regime

- $\Theta_\delta^{\mathrm{dep}} = \{\theta = (K, \pi, Q, f) \in \Theta^{\mathrm{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\}$

Theorem (Gassiat, Kaddouri and Naulet (2025))

- *When $K=2$:*

$$\forall \theta \in \Theta_\delta^{\mathrm{dep}}, \quad \frac{\delta^2(1-\tilde{\alpha}_n)}{1-\delta}\Lambda(\theta) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq (1-\delta)\Lambda(\theta),$$

where

  - $\tilde{\alpha}_n = 2e\left(\frac{1-\delta}{\delta}\right)^5 \sqrt{\log(2)/2n}$
  - $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\mathrm{dep}}$
  - $f = (f_i)_{i \in \{1,2\}}$ *and* $\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$

What if δ is very small?

Slowly mixing regime

- $\Theta_\delta^{\text{dep}} = \left\{\theta = (K, \pi, Q, f) \in \Theta^{\text{dep}} \mid \min_{i,j} Q_{i,j} \geq \delta,\ \min_j \pi_j \geq \delta\right\}$

Theorem (Gassiat, Kaddouri and Naulet (2025))

- *When $K=2$:*

$$\forall \theta \in \Theta_\delta^{\text{dep}}, \quad \frac{\delta^2(1-\tilde{\alpha}_n)}{1-\delta}\Lambda(\theta) \leq \inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g) \leq (1-\delta)\Lambda(\theta),$$

where

- $\tilde{\alpha}_n = 2e\left(\frac{1-\delta}{\delta}\right)^5 \sqrt{\log(2)/2n}$
- $\theta = (K, \pi, Q, f) \in \Theta_\delta^{\text{dep}}$
- $f = (f_i)_{i \in \{1,2\}}$ *and* $\Lambda(\theta) = \int f_1(y) \wedge f_2(y) dy$

What if δ is very small?

Example

$\theta = (K, \pi, Q, f)$ where

- $\pi = (1/2, 1/2)$
- $Q = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$ with $\delta < 1/2$
- $f = (\mathcal{N}(\mu, I_d), \mathcal{N}(-\mu, I_d))$

Tighter bounds

The problem of estimation of the mean μ was recently studied in Karagulyan and Ndaoud (2024).

Tighter bounds

The problem of estimation of the mean μ was recently studied in Karagulyan and Ndaoud (2024).

Theorem (Kaddouri and Ndaoud (2025))

$$\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g) \gtrsim \begin{cases} 1 & \|\mu\|^2 \leq 2\delta, \\ \frac{\delta}{\|\mu\|^2} & 2\delta < \|\mu\|^2 \leq 1, \\ \frac{\|\mu\|}{\|\mu\|^2+1}\, \delta \exp\left(-\frac{\|\mu\|^2}{2}\right) & \|\mu\|^2 > 1, \end{cases}$$

$$\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\text{clust}}(\theta, g) \lesssim \begin{cases} 1 & \|\mu\|^2 \leq 2\delta, \\ \frac{\delta}{\|\mu\|^2}\left(\log\left(\frac{\|\mu\|^2}{2\delta}\right)+1\right) & 2\delta \leq \|\mu\|^2 \leq 1, \\ \frac{1}{\|\mu\|^2}\delta \exp\left(-\frac{\|\mu\|^2}{2}\right)\left(\log\left(\frac{\|\mu\|^2}{2\delta}\right)+1\right) & 1 \leq \|\mu\|^2 \leq 2\log\left(\frac{1}{\delta}\right), \\ \delta \exp\left[-\frac{\|\mu\|^2}{2}\left(1-\frac{1}{\|\mu\|^2}\log\left(\frac{1-\delta}{\delta}\right)\right)^2\right] & \|\mu\|^2 > 2\log\left(\frac{1}{\delta}\right). \end{cases}$$

Strong mixing regime ($\delta \approx 1$)

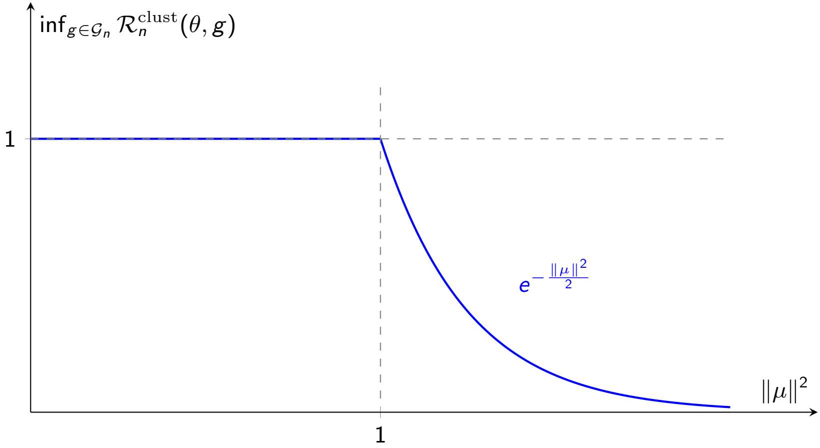


Figure: Behavior of the Bayes risk of online clustering in the **strong mixing regime**.

Slow mixing regime ($\delta << 1$)

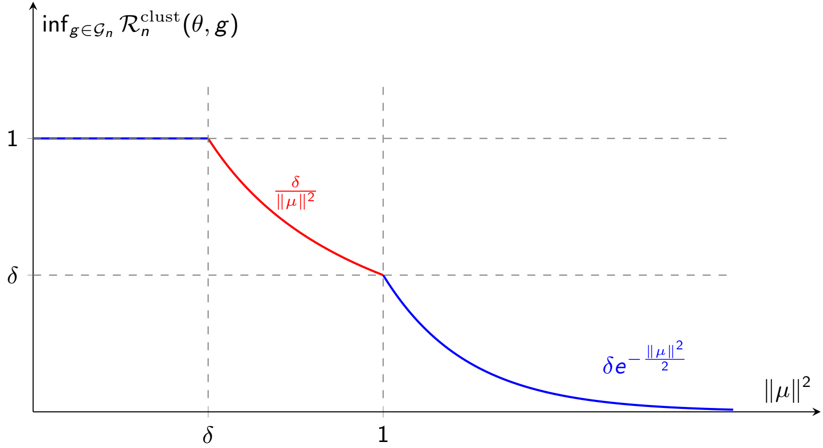


Figure: Behavior of the Bayes risk of clustering in the **slow mixing regime**

Outline

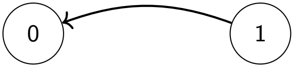
Outline

Linear preferential attachment model

Node	deg	prob
0	0	0

Figure: A preferential attachment graph.

Linear preferential attachment model

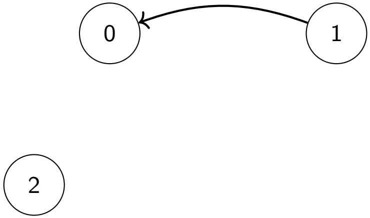


Node	deg	prob
0	1	$1/2$
1	1	$1/2$

G_1

Figure: A preferential attachment graph.

Linear preferential attachment model



Node	deg	prob
0	1	1/2
1	1	1/2

Figure: A preferential attachment graph.

Linear preferential attachment model

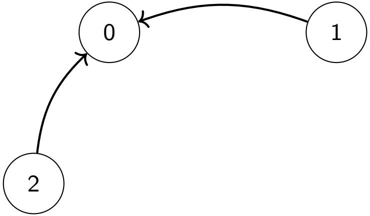


Figure: A preferential attachment graph.

G_2

Node	deg	prob
0	2	2/4
1	1	1/4
2	1	1/4

Linear preferential attachment model

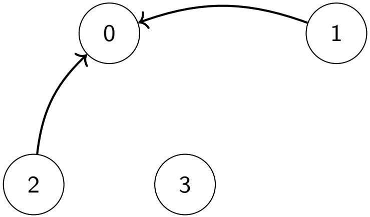


Figure: A preferential attachment graph.

Node	deg	prob
0	2	$2/4$
1	1	$1/4$
2	1	$1/4$

Linear preferential attachment model

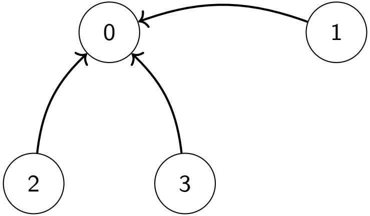


Figure: A preferential attachment graph.

$\boxed{G_3}$

Node	deg	prob
0	3	$3/6$
1	1	$1/6$
2	1	$1/6$
3	1	$1/6$

Linear preferential attachment model

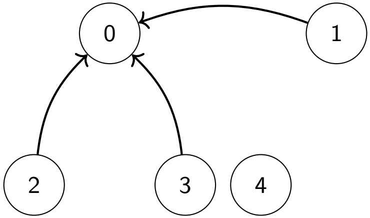


Figure: A preferential attachment graph.

Node	deg	prob
0	3	$3/6$
1	1	$1/6$
2	1	$1/6$
3	1	$1/6$

Linear preferential attachment model

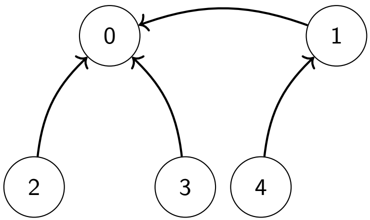


Figure: A preferential attachment graph.

G_4

Node	deg	prob
0	4	3/8
1	2	2/8
2	1	1/8
3	1	1/8
4	1	1/8

Linear preferential attachment model

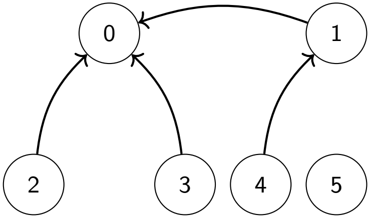


Figure: A preferential attachment graph.

Node	deg	prob
0	4	3/8
1	2	2/8
2	1	1/8
3	1	1/8
4	1	1/8

Linear preferential attachment model

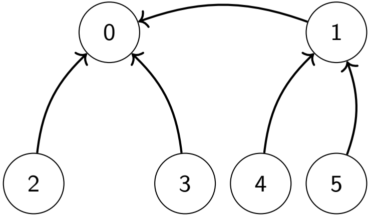


Figure: A preferential attachment graph.

G_5

Node	deg	prob
0	4	3/8
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/10

Linear preferential attachment model

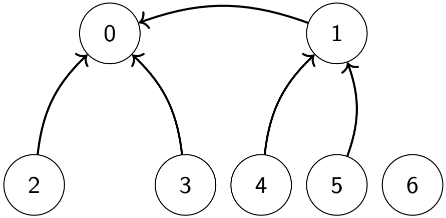
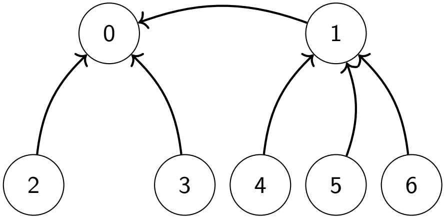


Figure: A preferential attachment graph.

Node	deg	prob
0	4	3/8
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/10

Linear preferential attachment model



G_6

Node	deg	prob
0	4	3/8
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	1	1/12

Figure: A preferential attachment graph.

Linear preferential attachment model

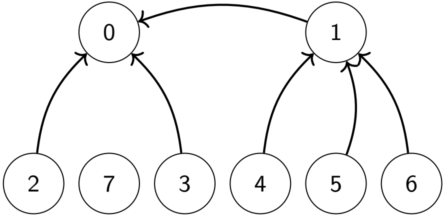


Figure: A preferential attachment graph.

Node	deg	prob
0	4	3/8
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	1	1/12

Linear preferential attachment model

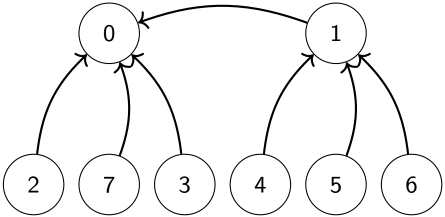


Figure: A preferential attachment graph.

G_7

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	1	1/14
7	1	1/14

Linear preferential attachment model

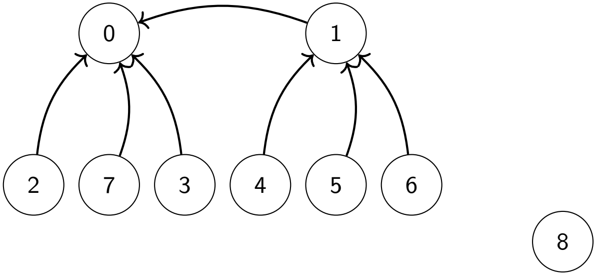


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	1	1/14
7	1	1/14

Linear preferential attachment model

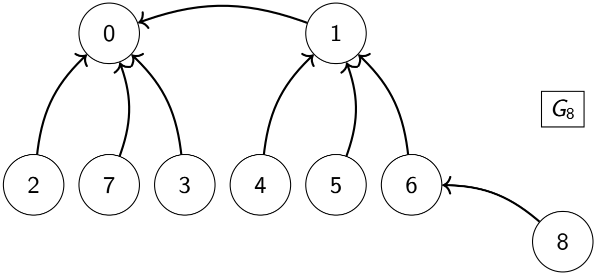


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/16

Linear preferential attachment model

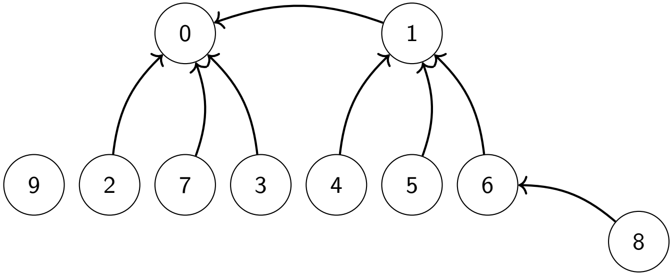


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/16

Linear preferential attachment model

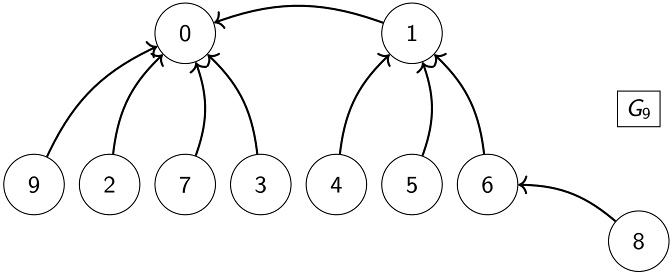


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/18
9	1	1/18

Linear preferential attachment model

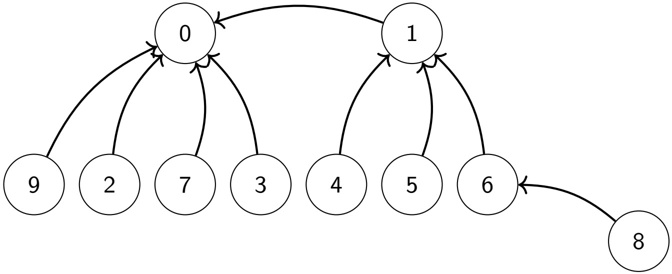


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/18
9	1	1/18

Linear preferential attachment model

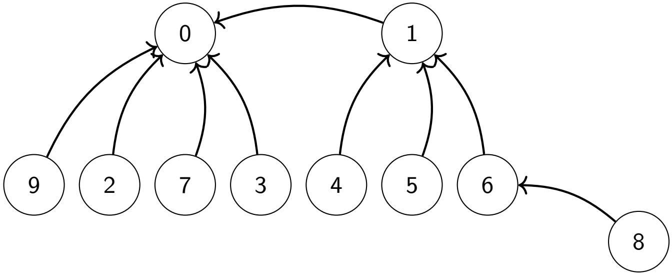


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/18
9	1	1/18

- The sequence of graphs $(G_0, G_1, G_2, ...)$ is built sequentially.

Linear preferential attachment model

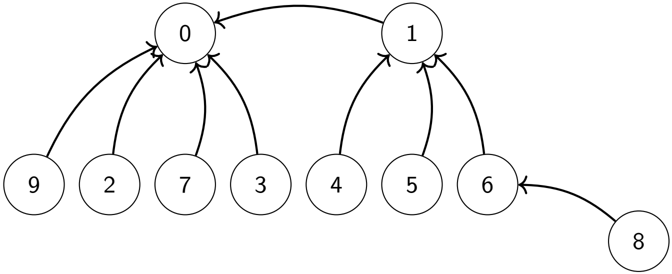


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/18
9	1	1/18

- The sequence of graphs $(G_0, G_1, G_2, \ldots)$ is built sequentially.
- Graph G_t has vertices $\{0, \ldots, t\}$.

Linear preferential attachment model

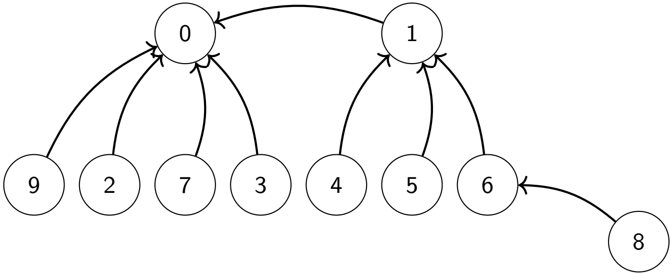


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/18
9	1	1/18

- The sequence of graphs $(G_0, G_1, G_2, \ldots)$ is built sequentially.
- Graph G_t has vertices $\{0, \ldots, t\}$.
- Graph G_t is constructed from G_{t-1} by connecting a new vertex t:

Linear preferential attachment model

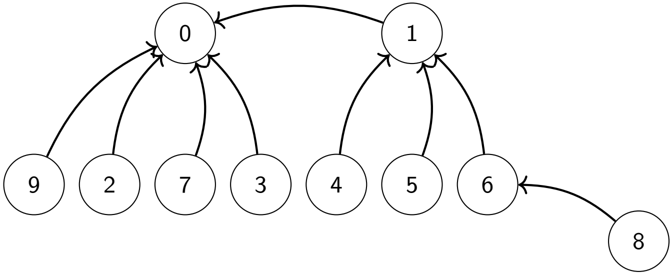


Figure: A preferential attachment graph.

Node	deg	prob
0	4	4/10
1	3	3/10
2	1	1/8
3	1	1/8
4	1	1/10
5	1	1/12
6	2	2/16
7	1	1/16
8	1	1/18
9	1	1/18

- The sequence of graphs $(G_0, G_1, G_2, \ldots)$ is built sequentially.
- Graph G_t has vertices $\{0, \ldots, t\}$.
- Graph G_t is constructed from G_{t-1} by connecting a new vertex t:

$$\mathbb{P}(\textcircled{t} \to \textcircled{v} \mid G_{t-1}) = \frac{\mathrm{d}_v(G_{t-1})}{\sum_{u=0}^{t-1} \mathrm{d}_u(G_{t-1})}$$

Affine preferential attachment model

$$\mathbb{P}(\textcircled{t} \to \textcircled{v} \mid G_{t-1}) = \frac{\mathrm{d}_v(G_{t-1})}{\sum_{u=0}^{t-1} \mathrm{d}_u(G_{t-1})}$$

Affine preferential attachment model

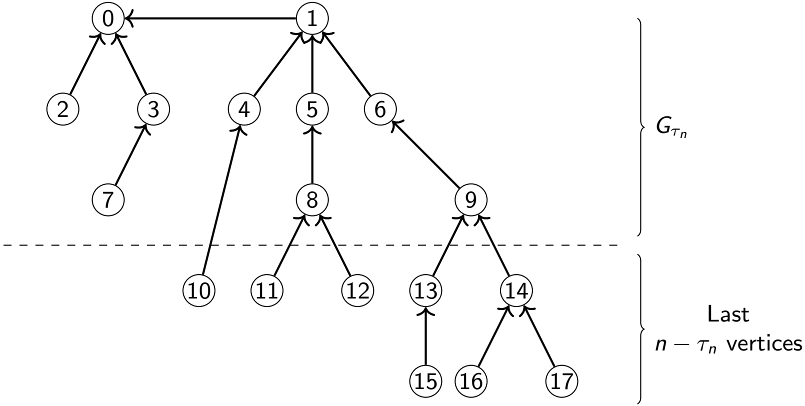
$$\mathbb{P}(\textcircled{t} \to \textcircled{v} \mid G_{t-1}) = \frac{\mathrm{d}_v(G_{t-1}) + \delta(t)}{\sum_{u=0}^{t-1} \mathrm{d}_u(G_{t-1}) + \delta(t)}$$

$$\delta(t) = \delta_0 \, \mathbf{1}_{t \le \tau_n} + \delta_1 \, \mathbf{1}_{t > \tau_n}, \qquad \delta_0, \delta_1 > -1, \quad \tau_n \le n.$$

Affine preferential attachment model

$$\mathbb{P}(\textcircled{t} \to \textcircled{v} \mid G_{t-1}) = \frac{\mathrm{d}_v(G_{t-1}) + \delta(t)}{\sum_{u=0}^{t-1} \mathrm{d}_u(G_{t-1}) + \delta(t)}$$

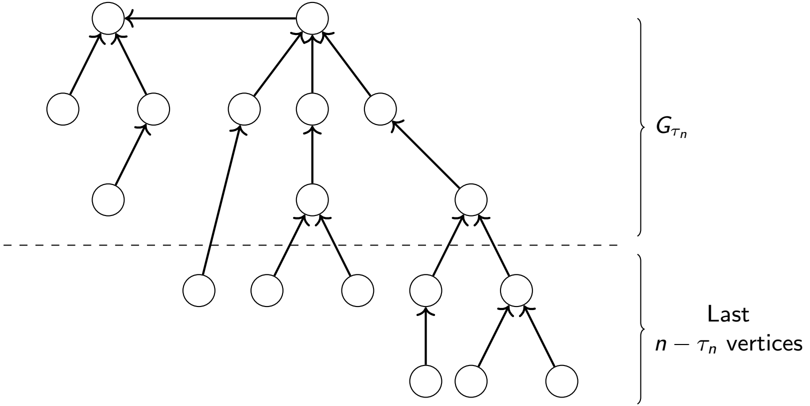
$$\delta(t) = \delta_0 \, \mathbf{1}_{t \le \tau_n} + \delta_1 \, \mathbf{1}_{t > \tau_n}, \qquad \delta_0, \delta_1 > -1, \quad \tau_n \le n.$$



Affine preferential attachment model

$$\mathbb{P}(\textcircled{t} \to \textcircled{v} \mid G_{t-1}) = \frac{\mathrm{d}_v(G_{t-1}) + \delta(t)}{\sum_{u=0}^{t-1} \mathrm{d}_u(G_{t-1}) + \delta(t)}$$

$$\delta(t) = \delta_0 \, \mathbf{1}_{t \le \tau_n} + \delta_1 \, \mathbf{1}_{t > \tau_n}, \qquad \delta_0, \delta_1 > -1, \;\; \tau_n \le n.$$



Unique observation: Final unlabeled graph

Outline

Late change-point detection: A hypothesis testing problem

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with} \quad c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

What is the minimal time needed to detect the change-point after it occurs?

Late change-point detection: A hypothesis testing problem

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with} \quad c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

What is the minimal time needed to detect the change-point after it occurs?

For a labeled graph G_n, we denote

- $u(G_n)$ the unlabeled graph induced by G_n
- $\mathbb{P}_0^n$ the distribution of graph G_n under (H_0)
- $\mathbb{P}_1^n$ the distribution of graph G_n under (H_1)

Late change-point detection: A hypothesis testing problem

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^\gamma \rfloor \quad \text{with} \quad c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

What is the minimal time needed to detect the change-point after it occurs?

For a labeled graph G_n, we denote

- $u(G_n)$ the unlabeled graph induced by G_n
- $\mathbb{P}_0^n$ the distribution of graph G_n under (H_0)
- $\mathbb{P}_1^n$ the distribution of graph G_n under (H_1)

Theorem (Bet et al. (2025))

For $\alpha > 0$, There exists a sequence of tests $(\psi_n)_{n \geq 0}$ such that

Late change-point detection: A hypothesis testing problem

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \le \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with} \quad c > 0 \quad \text{and} \quad \gamma \in (0, 1)\,.$$

What is the minimal time needed to detect the change-point after it occurs?

For a labeled graph G_n, we denote

- $u(G_n)$ the unlabeled graph induced by G_n
- $\mathbb{P}_0^n$ the distribution of graph G_n under (H_0)
- $\mathbb{P}_1^n$ the distribution of graph G_n under (H_1)

Theorem (Bet et al. (2025))

For $\alpha > 0$, There exists a sequence of tests $(\psi_n)_{n \ge 0}$ such that

- $\mathbb{P}_0^n(\psi_n(u(G_n)) \neq 0) \le (1 + \mathrm{o}(1))\alpha$

Late change-point detection: A hypothesis testing problem

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with} \quad c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

What is the minimal time needed to detect the change-point after it occurs?

For a labeled graph G_n, we denote

- $u(G_n)$ the unlabeled graph induced by G_n
- $\mathbb{P}_0^n$ the distribution of graph G_n under (H_0)
- $\mathbb{P}_1^n$ the distribution of graph G_n under (H_1)

Theorem (Bet et al. (2025))

For $\alpha > 0$, There exists a sequence of tests $(\psi_n)_{n \geq 0}$ such that

- $\mathbb{P}_0^n\left(\psi_n(u(G_n)) \neq 0\right) \leq (1 + \mathrm{o}(1))\alpha$
- $\mathbb{P}_1^n\left(\psi_n(u(G_n)) = 0\right) = \begin{cases} \mathrm{o}(1) & \textit{when } \gamma > 1/2 \\ Cte + \mathrm{o}(1) & \textit{when } \gamma = 1/2 \end{cases}$

Late change-point detection: A hypothesis testing problem

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with} \quad c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

What is the minimal time needed to detect the change-point after it occurs?

For a labeled graph G_n, we denote

- $u(G_n)$ the unlabeled graph induced by G_n
- $\mathbb{P}_0^n$ the distribution of graph G_n under (H_0)
- $\mathbb{P}_1^n$ the distribution of graph G_n under (H_1)

Theorem (Bet et al. (2025))

For $\alpha > 0$, There exists a sequence of tests $(\psi_n)_{n \geq 0}$ such that

- $\mathbb{P}_0^n\,(\psi_n(u(G_n)) \neq 0) \leq (1 + \mathrm{o}(1))\alpha$
- $\mathbb{P}_1^n\,(\psi_n(u(G_n)) = 0) = \begin{cases} \mathrm{o}(1) & \textit{when } \gamma > 1/2 \\ Cte + \mathrm{o}(1) & \textit{when } \gamma = 1/2 \end{cases}$
- $\mathbb{P}_1^n\,(\psi_n(u(G_n)) = 0) \geq (1 + \mathrm{o}(1))(1 - \alpha)$ *when* $\gamma < 1/2$

Fundamental limits of detection

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

Fundamental limits of detection

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with } c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

Fundamental limits of detection

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^\gamma \rfloor \quad \text{with } c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

Conjecture (Bet et al. (2025))

When $\gamma \leq 1/2$ there is no sequence of tests $(\psi_n)_{n \geq 0}$ such that

$$\mathbb{P}_0^n\left(\psi_n(u(G_n)) \neq 0\right) \xrightarrow[n \to +\infty]{} 0 \quad \textit{and} \quad \mathbb{P}_1\left(\psi_n(u(G_n)) = 0\right) \xrightarrow[n \to +\infty]{} 0.$$

Fundamental limits of detection

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^\gamma \rfloor \quad \text{with } c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

Conjecture (Bet et al. (2025))

When $\gamma \leq 1/2$ *there is no sequence of tests* $(\psi_n)_{n \geq 0}$ *such that*

$$\mathbb{P}_0^n\left(\psi_n(u(G_n)) \neq 0\right) \xrightarrow[n \to +\infty]{} 0 \quad \textit{and} \quad \mathbb{P}_1\left(\psi_n(u(G_n)) = 0\right) \xrightarrow[n \to +\infty]{} 0.$$

Theorem (Kaddouri, Naulet and Gassiat (2025))

If $\delta_0 > 0$ *and* $\Delta_n = o\left(n^{1/3}\right)$ *or if* $\delta_0 = 0$ *and* $\Delta_n = o\left(\frac{n^{1/3}}{\log(n)}\right)$*, then for every sequence* $(\psi_n)_n$ *of tests:*

Fundamental limits of detection

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with } c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

Conjecture (Bet et al. (2025))

When $\gamma \leq 1/2$ there is no sequence of tests $(\psi_n)_{n \geq 0}$ such that

$$\mathbb{P}_0^n \left(\psi_n(u(G_n)) \neq 0\right) \xrightarrow[n \to +\infty]{} 0 \quad \textit{and} \quad \mathbb{P}_1 \left(\psi_n(u(G_n)) = 0\right) \xrightarrow[n \to +\infty]{} 0.$$

Theorem (Kaddouri, Naulet and Gassiat (2025))

If $\delta_0 > 0$ and $\Delta_n = o\left(n^{1/3}\right)$ or if $\delta_0 = 0$ and $\Delta_n = o\left(\frac{n^{1/3}}{\log(n)}\right)$, then for every sequence $(\psi_n)_n$ of tests:

$$\mathbb{P}_0^n \left(\psi_n(u(G_n)) \neq 0\right) \xrightarrow[n \to +\infty]{} 0 \implies \mathbb{P}_1^n \left(\psi_n(u(G_n)) = 0\right) \xrightarrow[n \to +\infty]{} 1.$$

Fundamental limits of detection

$$(\mathrm{H}_0) \quad (\exists \delta_0 > -m)\, \delta(t) = \delta_0 \qquad \text{vs} \qquad (\mathrm{H}_1) \quad (\exists \delta_1 \neq \delta_0 > -m)\, \delta(t) = \delta_0 \mathbf{1}_{t \leq \tau_n} + \delta_1 \mathbf{1}_{t > \tau_n}$$

We focus here on the case of late change-point detection, i.e.

$$\tau_n = n - \Delta_n, \quad \Delta_n = \lfloor c n^{\gamma} \rfloor \quad \text{with } c > 0 \quad \text{and} \quad \gamma \in (0,1)\,.$$

Conjecture (Bet et al. (2025))

When $\gamma \leq 1/2$ *there is no sequence of tests* $(\psi_n)_{n \geq 0}$ *such that*

$$\mathbb{P}_0^n\left(\psi_n(u(G_n)) \neq 0\right) \xrightarrow[n \to +\infty]{} 0 \quad \textit{and} \quad \mathbb{P}_1\left(\psi_n(u(G_n)) = 0\right) \xrightarrow[n \to +\infty]{} 0.$$

Theorem (Kaddouri, Naulet and Gassiat (2025))

If $\delta_0 > 0$ *and* $\Delta_n = o\left(n^{1/3}\right)$ *or if* $\delta_0 = 0$ *and* $\Delta_n = o\left(\frac{n^{1/3}}{\log(n)}\right)$, *then for every sequence* $(\psi_n)_n$ *of tests:*

$$\mathbb{P}_0^n\left(\psi_n(u(G_n)) \neq 0\right) \xrightarrow[n \to +\infty]{} 0 \implies \mathbb{P}_1^n\left(\psi_n(u(G_n)) = 0\right) \xrightarrow[n \to +\infty]{} 1.$$

Errors of type *I* and *II* can not be controlled simultaneously.

Outline

Conclusion & perspectives

- ✅ Comprehensive characterization of clustering in the general **non-parametric** setting

Conclusion & perspectives

- Comprehensive characterization of clustering in the general **non-parametric** setting
- Clarifying the dependence on the **model parameters** including the **slowly mixing regime**

Conclusion & perspectives

- ✅ Comprehensive characterization of clustering in the general **non-parametric** setting
- ✅ Clarifying the dependence on the **model parameters** including the **slowly mixing regime**
- ✅ **Partial proof** of the conjecture (later fully established by Du et al. (2025))

Conclusion & perspectives

- ✓ Comprehensive characterization of clustering in the general **non-parametric** setting
- ✓ Clarifying the dependence on the **model parameters** including the **slowly mixing regime**
- ✓ **Partial proof** of the conjecture (later fully established by Du et al. (2025))
- ✓ ...

Conclusion & perspectives

- ✓ Comprehensive characterization of clustering in the general **non-parametric** setting
- ✓ Clarifying the dependence on the **model parameters** including the **slowly mixing regime**
- ✓ **Partial proof** of the conjecture (later fully established by Du et al. (2025))
- ✓ ...
- ? Developing **non plug-in** clustering approaches

Conclusion & perspectives

- ✅ Comprehensive characterization of clustering in the general **non-parametric** setting
- ✅ Clarifying the dependence on the **model parameters** including the **slowly mixing regime**
- ✅ **Partial proof** of the conjecture (later fully established by Du et al. (2025))
- ✅ ...
- ❓ Developing **non plug-in** clustering approaches
- ❓ Establishing **optimal excess risk** bounds

Conclusion & perspectives

- ✅ Comprehensive characterization of clustering in the general **non-parametric** setting
- ✅ Clarifying the dependence on the **model parameters** including the **slowly mixing regime**
- ✅ **Partial proof** of the conjecture (later fully established by Du et al. (2025))
- ✅ ...
- ❓ Developing **non plug-in** clustering approaches
- ❓ Establishing **optimal excess risk** bounds
- ❓ Handling **general dependent data** and model misspecification

Conclusion & perspectives

- ✅ Comprehensive characterization of clustering in the general **non-parametric** setting
- ✅ Clarifying the dependence on the **model parameters** including the **slowly mixing regime**
- ✅ **Partial proof** of the conjecture (later fully established by Du et al. (2025))
- ✅ ...
- ❓ Developing **non plug-in** clustering approaches
- ❓ Establishing **optimal excess risk** bounds
- ❓ Handling **general dependent data** and model misspecification
- ❓ **Minimax** point of view

Publications

Elisabeth Gassiat, Ibrahim Kaddouri and Zacharie Naulet (2025).
Clustering risk in Non-parametric Hidden Markov and I.I.D. Models.
Ann. Statist., to appear.

Ibrahim Kaddouri, Zacharie Naulet and Elisabeth Gassiat (2025).
On the impossibility of detecting a late change-point in the preferential attachment random graph model.
Bernoulli, to appear.

Ibrahim Kaddouri and Mohamed Ndaoud (2025).
Clustering under the slowly mixing Hidden Markov Model.
To appear.

References I

Animashree Anandkumar, Daniel Hsu, and Sham M. Kakade, A Method of Moments for Mixture Models and Hidden Markov Models, in *Proceedings of the 25th Annual Conference on Learning Theory*, Shie Mannor, Nathan Srebro, and Robert C. Williamson (eds.), Proceedings of Machine Learning Research, vol. 23, PMLR, Edinburgh, Scotland, 25–27 Jun 2012, pp. 33.1–33.34.

E. Gassiat, A. Cleynen, and S. Robin, Inference in finite state space nonparametric hidden Markov models and applications, *Statistics and Computing*, vol. 26, no. 1-2, pp. 61–71, 2016. doi:10.1007/s11222-014-9523-8.

Kweku Abraham, Ismaël Castillo, and Elisabeth Gassiat, Multiple testing in nonparametric hidden Markov models: An empirical Bayes approach, *J. Mach. Learn. Res.*, vol. 23, no. 1, pp. 4061–4117, 2022.

Yohann De Castro, Élisabeth Gassiat, and Sylvain Le Corff, Consistent Estimation of the Filtering and Marginal Smoothing Distributions in Nonparametric Hidden Markov Models, *IEEE Trans. Inform. Theory*, vol. 63, no. 8, pp. 4758–4777, 2017. doi:10.1109/TIT.2017.2696959.

Hang Du, Shuyang Gong, and Jiaming Xu, A Proof of The Changepoint Detection Threshold Conjecture in Preferential Attachment Models, in *Proceedings of Thirty Eighth Conference on Learning Theory*, edited by Nika Haghtalab and Ankur Moitra, PMLR, vol. 291, pp. 1559–1563, 2025.

References II

Hang Du, Shuyang Gong, and Jiaming Xu, A Proof of The Changepoint Detection Threshold Conjecture in Preferential Attachment Models, in *Proceedings of Thirty Eighth Conference on Learning Theory*, edited by Nika Haghtalab and Ankur Moitra, PMLR, vol. 291, pp. 1559–1563, 2025.

Gianmarco Bet, Kay Bogerd, Rui M. Castro, and Remco van der Hofstad, Detecting a late changepoint in the preferential attachment model, *Bernoulli*, to appear, 2025.

Ibrahim Kaddouri, Zacharie Naulet, and Elisabeth Gassiat, On the impossibility of detecting a late change-point in the preferential attachment random graph model, *Bernoulli*, to appear, 2025.

Elisabeth Gassiat, Ibrahim Kaddouri, and Zacharie Naulet, Clustering risk in Non-parametric Hidden Markov and I.I.D. Models, *Ann. Statist.*, to appear, 2025.

Vahe Karagulyan and Mohamed Ndaoud, Adaptive Mean Estimation in the Hidden Markov sub-Gaussian Mixture Model, *arXiv preprint arXiv:2406.12446*, 2024.

Ibrahim Kaddouri, and Mohamed Ndaoud, Clustering under the slowly mixing Hidden Markov Model, to appear, 2025.

Nicolas Verzelen, Magalie Fromont, Matthieu Lerasle, and Patricia Reynaud-Bouret, Optimal change-point detection and localization, *Ann. Statist.*, vol. 51, no. 4, pp. 1586–1610, 2023.

Thank you for your attention!

Bayes classifier vs Bayes clusterer ($K = 2$)

$g_\theta^\star(Y_{1:n}) = \pi_n \circ \tilde{h}_\theta(Y_{1:n})$ a.e, where

$$\tilde{h}_\theta(Y_{1:n}) \in \underset{h=(h_i)_{i\in[n]}}{\arg\min}\ \mathbb{E}_\theta\left[\min_{\tau\in\mathcal{S}_2}\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{h_i(Y_{1:n})\neq\tau(X_i)}\middle| Y_{1:n}\right].$$

Given that:

$$\mathbb{E}_\theta\left[\min_{\tau\in\mathcal{S}_2}\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{h_i(Y_{1:n})\neq\tau(X_i)}\middle| Y_{1:n}\right] = \frac{1}{2} - \frac{1}{n}\mathbb{E}_\theta\left[\left|\sum_{i=1}^n \mathbf{1}_{h_i(Y_{1:n})\neq X_i} - \frac{n}{2}\right|\middle| Y_{1:n}\right]$$

one gets $\tilde{h}_\theta(Y_{1:n}) \in \arg\max_{h=(h_i)_{i\in[n]}} \mathbb{E}_\theta\left[\left|\sum_{i=1}^n \mathbf{1}_{h_i(Y_{1:n})\neq X_i} - \frac{n}{2}\right|\middle| Y_{1:n}\right]$.

The result is a direct consequence of this lemma.

Lemma

Let $(Z_i)_{i\in[n]}$ be a sequence of independent Bernoulli random variables such that $Z_i \sim \mathcal{B}(p_i)$ where $p_i \in \{\alpha_i, 1-\alpha_i\}$. Then the maximum $\max_{(p_i)_{i\in[n]}} \mathbb{E}\left[\left|\sum_{i\in[n]} Z_i - \frac{n}{2}\right|\right]$ is reached at $(p_i)_{i\in[n]} = (\alpha_i \wedge (1-\alpha_i))_{i\in[n]}$ and $(p_i)_{i\in[n]} = (\alpha_i \vee (1-\alpha_i))_{i\in[n]}$.

Bayes classifier vs Bayes clusterer ($K > 2$)

We define

$$(\forall i \in [n])\,(\forall k \in \mathbb{X}) \quad \alpha_k^{(Y_i)} = \frac{\nu_k f_k(Y_i)}{\sum_{j=1}^{J} \nu_j f_j(Y_i)}$$

and $h_\theta^\star(Y_{1:n}) = \left(h_{\theta,i}^\star(Y_i)\right)_{i\in[n]}$ where $(\forall i \in [n])$ $h_{\theta,i}^\star(Y_i) = \arg\max_{k\in\mathbb{X}} \alpha_k^{(Y_i)}$.
Consider the event :

$$A_n = \bigcup_{j=1}^{J} \bigcap_{i=1}^{n} \left\{ \max_{k\neq j} \nu_k f_k(Y_i) < \nu_j f_j(Y_i) \right\}.$$

Then,

$$A_n \subset \left\{ \pi_n\left(\left(h_{\theta,i}^\star(Y_i)\right)_{i\in[n]} \right) = \pi_n\left((1,\dots,1)\right) \right\}.$$

Note that:

$$A_n \cap \left\{ \mathbb{E}_\theta\left[\ell(\pi_n(X_{1:n}), \pi_n((1,..,1))) \middle| Y_{1:n} \right] > \mathbb{E}_\theta\left[\ell(\pi_n(X_{1:n}), \pi_n((1,..,1,2))) \middle| Y_{1:n} \right] \right\} \subset \{g_\theta^\star(Y_{1:n}) \neq \pi_n \circ h_\theta^\star(Y_{1:n})$$

Generic lemma for lower-bounds

Define

- $U_{n,\tau}(h) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\tau(X_i)\neq h_i(Y_{1:n})}$.
- $\hat{p}_\tau(h) := \mathbb{E}_\theta\left[U_{n,\tau}(h) \mid Y_{1:n}\right]$.
- $\hat{\tau}_h$ denote a $Y_{1:n}$-measurable permutation satisfying: $\hat{p}_{\hat{\tau}_h}(h) = \mathbb{E}_\theta\left[U_{n,\hat{\tau}_h}(h) \mid Y_{1:n}\right] = \min_\tau \mathbb{E}_\theta\left[U_{n,\tau}(h) \mid Y_{1:n}\right] = \min_\tau \hat{p}_\tau(h)$.

We let $N_j := \sum_{i=1}^{n} \mathbf{1}_{\{X_i=j\}}$ and $N_{(1)} \leq N_{(2)} \leq \cdots \leq N_{(J)}$ denote the order statistics of $(N_1, \ldots, N_J)$.

Proposition

A generic lower bound that works for any latent model (i.i.d. or HMM). For all classifiers h, all ε, all η and all θ

$$\begin{aligned}
\mathbb{E}_\theta\left[\min_\tau U_{n,\tau}(h)\right] \geq\ & \mathbb{E}_\theta\left[\min_\tau \mathbb{E}_\theta[U_{n,\tau}(h) \mid Y_{1:n}]\right] \\
& - \mathbb{E}_\theta\left[\mathbb{E}_\theta\left[\max_\tau(-U_{n,\tau}(h) + \hat{p}_\tau(h)) \mid Y_{1:n}\right]\mathbf{1}_{\{\hat{p}_{\hat{\tau}_h}(h)\geq\varepsilon\}}\right] \\
& - \mathbb{E}_\theta\left[\mathbb{P}_\theta(U_{n,\hat{\tau}_h}(h) > \eta \mid Y_{1:n})\mathbf{1}_{\{\hat{p}_{\hat{\tau}_h}(h)<\varepsilon\}}\right] \\
& - \mathbb{P}_\theta\left(N_{(1)} + N_{(2)} < 2n\eta\right).
\end{aligned}$$

Tight bound on the gap $\inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta,h) - \inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta,g)$

Let $\varepsilon_{n,\theta} = \frac{1}{2} - \inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta,h)$.

Theorem

When $J=2$ *and* $\theta\in\Theta^{\mathrm{ind}}$, $\inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta,h)=0$ *if and only if* $\inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta,g)=0$*. If* $\inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta,h)\neq 0$ *then the difference between the Bayes risks satisfies*

$$\inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta,h) - \inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta,g) \leq \min\left(\frac{(1-4\varepsilon_{n,\theta}^2)^{\frac{n}{2}}}{\frac{n}{2}\log\left(\frac{1+2\varepsilon_{n,\theta}}{1-2\varepsilon_{n,\theta}}\right)}, \sqrt{\frac{\pi}{2n}}\right).$$

Furthermore, there exists a universal constant $B>0$ *such that for all* $n\geq 100$ *and all* $\theta\in\Theta^{\mathrm{ind}}$

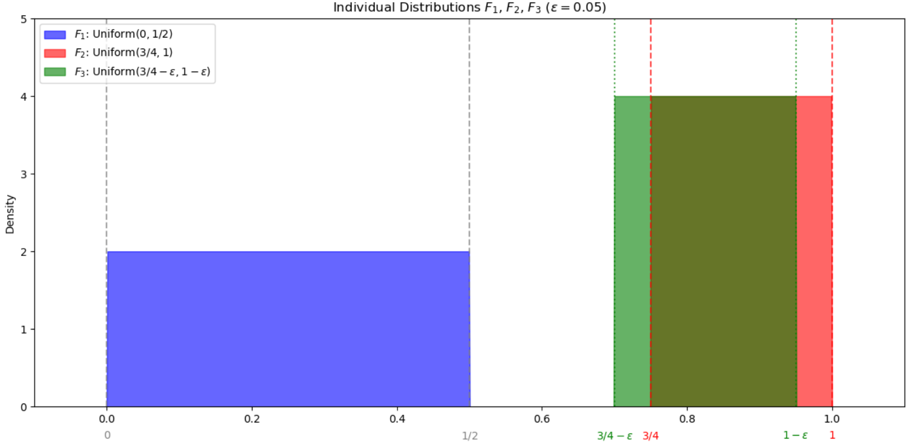
$$\inf_{h\in\mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta,h) - \inf_{g\in\mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta,g) \geq B\min\left(\frac{(1-4\varepsilon_{n,\theta}^2)^{\frac{n}{2}\left(1+\frac{6.8}{1\vee\sqrt{n}\varepsilon_{n,\theta}}\right)}}{\frac{n}{2}\log\left(\frac{1+2\varepsilon_{n,\theta}}{1-2\varepsilon_{n,\theta}}\right)}, \frac{1}{\sqrt{n}}\right).$$

Example where risks $\inf_{h\in\mathcal{H}_n}\mathcal{R}_n^{\text{class}}(\theta,h)$ and $\inf_{g\in\mathcal{G}_n}\mathcal{R}_n^{\text{clust}}(\theta,g)$ are not equivalent

$$F = (1-2\eta)\,F_1 + \eta F_2 + \eta F_3$$

where $F_1 = U\,(0,1/2)$, $F_2 = U\,(3/4,1)$ and $F_3 = U\,(3/4-\varepsilon, 1-\varepsilon)$.

- $\inf_{h\in\mathcal{H}_n}\mathcal{R}_n^{\text{class}}(\theta,h) \asymp \eta(\frac{1}{4}-\varepsilon)$
- $\inf_{g\in\mathcal{G}_n}\mathcal{R}_n^{\text{clust}}(\theta,g) \lesssim n\eta^2 + n\eta\varepsilon$



Improving the Excess risk

Lemma

For all $0 < \gamma < 1/2$ *and all* $\theta \in \Theta^{\mathrm{dep}}$

$$\mathcal{R}_n^{\mathrm{class}}(\theta, h_{\hat{\theta}}) \leq \frac{\inf_{h \in \mathcal{H}_n} \mathcal{R}_n^{\mathrm{class}}(\theta, h)}{1/2 - \gamma} + \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_\theta \left(\left\| \phi_{\theta,i|n} - \phi_{\hat{\theta},i|n} \right\|_{\mathrm{TV}} > \gamma \right) . \quad (1)$$

where $\phi_{\theta,i|n} = \mathbb{P}_\theta \left(X_i \in . \mid Y_{1:n} \right)$ *and* $h_{\hat{\theta}}$ *is the plug-in classifier.*

Comparison with the problem of segmentation/change-point localization

- Model: n i.i.d. observations $Y_i = \theta_i + \varepsilon_i$
- There exists an integer $0 < K < n-1$, a vector of integers $\tau^* = (\tau_1^*, \dots, \tau_K^*)$ satisfying

$$1 = \tau_0^* < \tau_1^* < \cdots < \tau_K^* < \tau_{K+1}^* = n+1,$$

  a vector $\mu = (\mu_1, \dots, \mu_{K+1}) \in \{\mu, -\mu\}^{K+1}$ satisfying $\mu_k \neq \mu_{k+1}$ for all $k \in \{1, \dots, K\}$ such that

$$\theta_i = \sum_{k=1}^{K+1} \mu_k \mathbf{1}_{\tau_{k-1}^* < i \leq \tau_k^*}.$$

- In Verzelen et al. (2023), an estimator $\hat{\tau}$ is built on a large probability event $\mathcal{A}$ such that

$$\mathbb{E}\left[\|\tau^\star - \hat{\tau}\|_1 \mathbf{1}_{\mathcal{A}}\right] \lesssim K\left(\frac{1}{\|\mu\|^2} \wedge e^{-c\|\mu\|^2}\right)$$

- In the slowly mixing regime, the expected number of changes is $K = n\delta$.
- The bound obtained in the slowly mixing regime:

$$\inf_{g \in \mathcal{G}_n} \mathcal{R}_n^{\mathrm{clust}}(\theta, g) \leq \delta\left(\frac{1}{\|\mu\|^2} \wedge e^{-\frac{\|\mu\|^2}{2}}\right)$$

Contiguity

We define the risk of a test ψ_n as:

$$R_n(\psi_n) = \mathbb{E}_0\left[\psi_n\right] + \mathbb{E}_1\left[1 - \psi_n\right]$$

Let $\mathrm{A_n}$ and $\mathrm{B_n}$ be two events such that $\mathrm{B_n}$ is a large probability event.

$$\mathbb{P}_1^n\left(\mathrm{A}_n\right) \leq \mathbb{P}_1^n\left(\mathrm{B}_n^c\right) + \mathbb{P}_0^n\left(\mathrm{A}_n\right)^{1/2}\mathbb{E}_{\mathbb{P}_0^n}\left[\left(\frac{d\mathbb{P}_1^n}{d\mathbb{P}_0^n}\right)^2\mathbf{1}_{\mathrm{B}_n}\right]^{1/2} \leq o(1) + \mathbb{P}_0^n\left(\mathrm{A}_n\right)^{1/2}\sup_n \mathbb{E}_{\mathbb{P}_0^n}\left[\left(\frac{d\mathbb{P}_1^n}{d\mathbb{P}_0^n}\right)^2\mathbf{1}_{\mathrm{B}_n}\right]^{1/2}$$

If $\sup_n \mathbb{E}_{\mathbb{P}_0^n}\left[\left(\frac{d\mathbb{P}_1^n}{d\mathbb{P}_0^n}\right)^2\mathbf{1}_{\mathrm{B}_n}\right] < \infty$, then :

$$\mathbb{P}_0^n\left(\mathrm{A}_n\right) \xrightarrow[n\to+\infty]{} 0 \implies \mathbb{P}_1^n\left(\mathrm{A}_n\right) \xrightarrow[n\to+\infty]{} 0$$

The type I and type II errors can not be controlled at the same time.

Idea of the proof: Intermediate problem

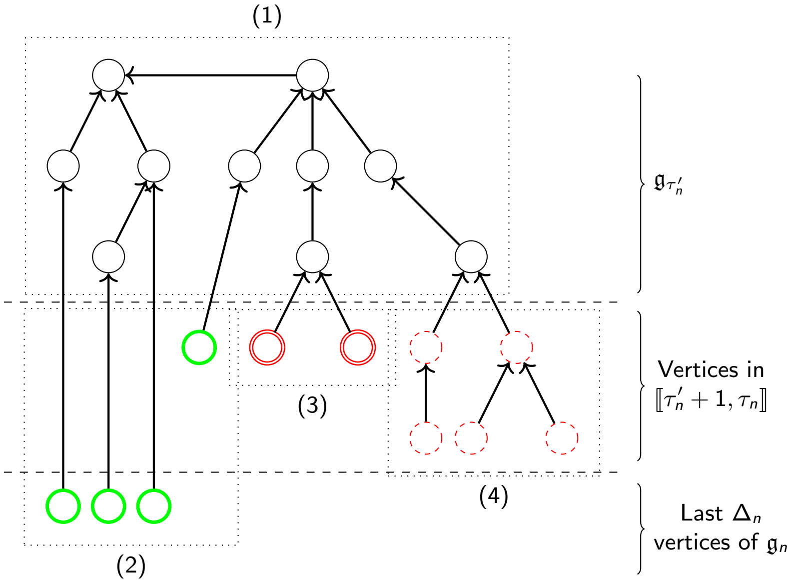


Figure: Typical preferential attachment graph $\mathfrak{g}_n$ with $m = 1$ when $\Delta_n = o(n^{1/3})$. Only green nodes are permuted.

What happens in the remaining regime ?

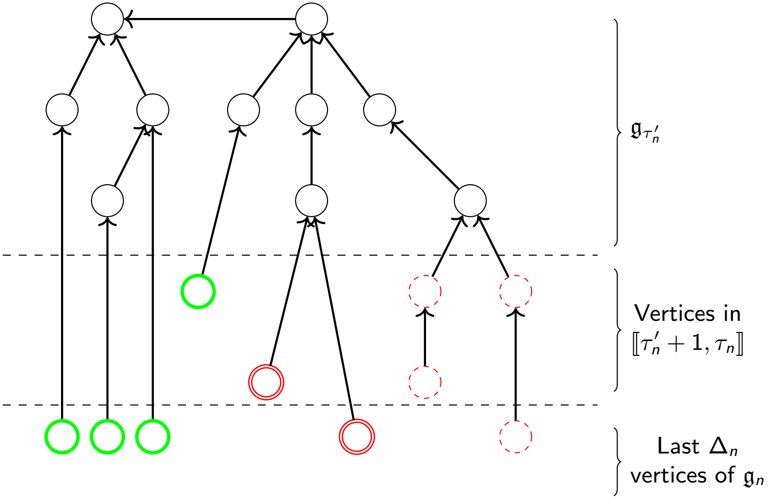


Figure: Typical preferential attachment graph $\mathfrak{g}_n$with $m = 1$ when $n^{1/3} \lesssim \Delta_n \lesssim n^{1/2}$.

The improvement

Their proof builds on and refines our approach, introducing the idea of **interpolation**.
Let $\mathbb{P}_n$ the distribution of the unlabeled graph under the null hypothesis, and let $\mathbb{Q}_{n,\tau_n}$ be its distribution under the alternative. Consider the interpolation:

$$\mathbb{P}_n = \mathbb{Q}_{n,n} \;\to\; \mathbb{Q}_{n,n-1} \;\to\; \mathbb{Q}_{n,n-2} \;\to\; \cdots \;\to\; \mathbb{Q}_{n,\tau_n}.$$

Applying the triangle inequality along this path yields

$$\mathrm{TV}(\mathbb{P}_n, \mathbb{Q}_{n,\tau_n}) \le \sum_{k=1}^{n-\tau_n} \mathrm{TV}(\mathbb{Q}_{n,n-k+1}, \mathbb{Q}_{n,n-k}) \,.$$

By the data-processing inequality,

$$\mathrm{TV}(\mathbb{P}_n, \mathbb{Q}_{n,\tau_n}) \le \sum_{k=1}^{n-\tau_n} \mathrm{TV}(\mathbb{P}_{n-k+1}, \mathbb{Q}_{n-k+1,n-k}) \,.$$

Hence, the task is to prove that

$$\forall n' \in [\![\tau_n + 1, n]\!], \quad \mathrm{TV}(\mathbb{P}_{n'}, \mathbb{Q}_{n',n'-1}) = o\Big(\tfrac{1}{n-\tau_n}\Big) \,.$$

Without loss of generality, it suffices to consider the simplified case $n' = n$ and $\tau_n = n - 1$, where the change-point occurs exactly one step before the terminal time.

Tighter bounds: Online setting

Theorem

The Bayes risk of online clustering satisfies

$$\inf_{\mathfrak{g}\in\mathcal{G}^{\text{on}}} \mathcal{R}_n^{\text{clust,on}}(\theta,\delta,\mathfrak{g}) \gtrsim \begin{cases} 1, & \|\mu\|^2 \le 2\delta, \\ \frac{\delta}{\|\mu\|^2}\left(\log\left(\frac{\|\mu\|^2}{2\delta}\right)+1\right), & 2\delta < \|\mu\|^2 \le \log\frac{1}{\delta}, \\ \frac{1}{\|\mu\|}\exp\left(-\frac{\|\mu\|^2}{2}\left(1+\frac{\log\left(\frac{1-\delta}{\delta}\right)}{2\|\mu\|^2}\right)^2\right), & \|\mu\|^2 > \log\frac{1}{\delta}, \end{cases}$$

$$\inf_{\mathfrak{g}\in\mathcal{G}^{\text{on}}} \mathcal{R}_n^{\text{clust,on}}(\theta,\delta,\mathfrak{g}) \le \begin{cases} \frac{1}{2}, & \|\mu\|^2 \le 2\delta, \\ 4\frac{\delta}{\|\mu\|^2}\left(\log\left(\frac{\|\mu\|^2}{2\delta}\right)+1\right), & 2\delta < \|\mu\|^2 \le \log\frac{1}{\delta}, \\ \frac{1}{\sqrt{2\pi}\|\mu\|}\exp\left(-\frac{\|\mu\|^2}{2}\right), & \|\mu\|^2 > \log\frac{1}{\delta}. \end{cases}$$

Slow mixing online regime ($\delta << 1$)

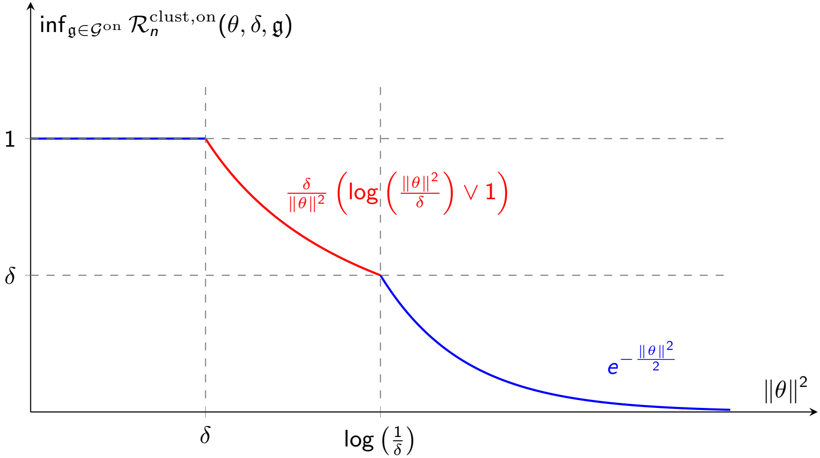


Figure: Behavior of the Bayes risk of online clustering in the slowly mixing regime