Clustering under the slowly mixing Gaussian Hidden Markov Model

Ibrahim Kaddouri
Joint work with Mohamed Ndaoud

Hidden Markov Model

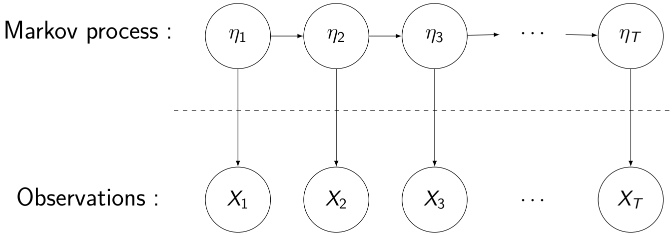


Figure: A Hidden Markov Model.

Latent (unobserved) variables $(\eta_k)_k$ form a Markov chain.

Hidden Markov Model

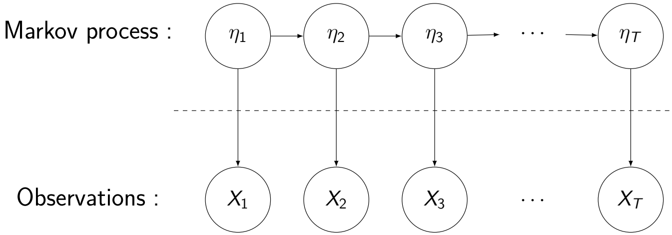


Figure: A Hidden Markov Model.

Latent (unobserved) variables $(\eta_k)_k$ form a Markov chain.
Observations $(X_k)_k$ are independent conditionnally to $(\eta_k)_k$.

A simple Hidden Markov Model

One observes $X_1, \ldots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta \eta_i + \xi_i$$

A simple Hidden Markov Model

One observes $X_1, \dots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta\eta_i + \xi_i$$

where

- $\xi_1, \dots, \xi_n$ are standard Gaussian random vectors

A simple Hidden Markov Model

One observes $X_1, \ldots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta\eta_i + \xi_i$$

where

- $\xi_1, \ldots, \xi_n$ are standard Gaussian random vectors
- $(\eta_i)_{1 \le i \le n}$ is a Markov chain with two states -1 and 1

A simple Hidden Markov Model

One observes $X_1, \ldots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta\eta_i + \xi_i$$

where

- $\xi_1, \ldots, \xi_n$ are standard Gaussian random vectors
- $(\eta_i)_{1 \leq i \leq n}$ is a Markov chain with two states -1 and 1
- The transition matrix ensures $Q = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$

Here, δ is allowed to depend on n and to be very small.

A simple Hidden Markov Model

One observes $X_1, \dots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta\eta_i + \xi_i$$

where

- $\xi_1, \dots, \xi_n$ are standard Gaussian random vectors
- $(\eta_i)_{1 \leq i \leq n}$ is a Markov chain with two states -1 and 1
- The transition matrix ensures $Q = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$

Here, δ is allowed to depend on n and to be very small.

Some natural questions arise in this setting:

A simple Hidden Markov Model

One observes $X_1, \ldots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta \eta_i + \xi_i$$

where

- $\xi_1, \ldots, \xi_n$ are standard Gaussian random vectors
- $(\eta_i)_{1 \le i \le n}$ is a Markov chain with two states -1 and 1
- The transition matrix ensures $Q = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$

Here, δ is allowed to depend on n and to be very small.

Some natural questions arise in this setting:

- Estimation of θ

A simple Hidden Markov Model

One observes $X_1, \ldots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta\eta_i + \xi_i$$

where

- $\xi_1, \ldots, \xi_n$ are standard Gaussian random vectors
- $(\eta_i)_{1 \le i \le n}$ is a Markov chain with two states -1 and 1
- The transition matrix ensures $Q = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$

Here, δ is allowed to depend on n and to be very small.

Some natural questions arise in this setting:

- Estimation of θ
- Estimation of δ

A simple Hidden Markov Model

One observes $X_1, \dots, X_n \in \mathbb{R}^d$ such that

$$X_i = \theta \eta_i + \xi_i$$

where

- $\xi_1, \dots, \xi_n$ are standard Gaussian random vectors
- $(\eta_i)_{1 \le i \le n}$ is a Markov chain with two states -1 and 1
- The transition matrix ensures $Q = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}$

Here, δ is allowed to depend on n and to be very small.

Some natural questions arise in this setting:

- Estimation of θ
- Estimation of δ
- Clustering

The problem of estimation

The problem of estimating θ has already been studied recently in Karagulyan **and** Ndaoud 2024. Let

$$M(n, d, \delta, t) = \inf_{\hat{\theta}(X_{1:n})} \sup_{\|\theta\|=t} \mathbb{E}_\theta \left[\min \left\{ \|\hat{\theta}(X_{1:n}) - \theta\|, \|\hat{\theta}(X_{1:n}) + \theta\| \right\} \right]$$

The problem of estimation

The problem of estimating θ has already been studied recently in Karagulyan **and** Ndaoud 2024. Let

$$M(n,d,\delta,t) = \inf_{\hat{\theta}(X_{1:n})} \sup_{\|\theta\|=t} \mathbb{E}_\theta \left[\min \left\{ \|\hat{\theta}(X_{1:n}) - \theta\|, \|\hat{\theta}(X_{1:n}) + \theta\| \right\} \right]$$

It was shown in Karagulyan **and** Ndaoud 2024 that when $d \le \delta n$,

$$M(n,d,\delta,t) \asymp \begin{cases} t & \text{if } t \le \left(\frac{\delta d}{n}\right)^{1/4} \\ \frac{1}{t}\sqrt{\frac{\delta d}{n}} & \text{if } \left(\frac{\delta d}{n}\right)^{1/4} \le t \le \sqrt{\delta} \\ \sqrt{\frac{d}{n}} & \text{if } \sqrt{\delta} \le t \end{cases}$$

The problem of clustering

Let $\eta = (\eta_i)_{1 \leq i \leq n}$ and $\ell\left(\hat{\eta}, \eta\right) = \min_{\nu \in \{-1,1\}} \left|\hat{\eta} + \nu\eta\right|$ where $|.|$ is the Hamming distance. We define the risk of a clustering procedure $\hat{\eta}$ as

$$\mathcal{R}\left(\theta, \delta, \hat{\eta}\right) = \mathbb{E}\left[\ell\left(\hat{\eta}(X_{1:n}), \eta\right)\right]$$

and the Bayes risk of clustering as

$$\inf_{\hat{\eta}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right)$$

The problem of clustering

Let $\eta = (\eta_i)_{1 \leq i \leq n}$ and $\ell(\hat{\eta}, \eta) = \min_{\nu \in \{-1,1\}} |\hat{\eta} + \nu\eta|$ where $|.|$ is the Hamming distance. We define the risk of a clustering procedure $\hat{\eta}$ as

$$\mathcal{R}(\theta, \delta, \hat{\eta}) = \mathbb{E}\left[\ell\left(\hat{\eta}(X_{1:n}), \eta\right)\right]$$

and the Bayes risk of clustering as

$$\inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta})$$

This risk was studied in Gassiat, Kaddouri **and** Naulet 2023 in the strongly mixing setting.

The problem of clustering

Let $\eta = (\eta_i)_{1 \le i \le n}$ and $\ell(\hat{\eta}, \eta) = \min_{\nu \in \{-1,1\}} |\hat{\eta} + \nu\eta|$ where $|.|$ is the Hamming distance. We define the risk of a clustering procedure $\hat{\eta}$ as

$$\mathcal{R}(\theta, \delta, \hat{\eta}) = \mathbb{E}\left[\ell\left(\hat{\eta}(X_{1:n}), \eta\right)\right]$$

and the Bayes risk of clustering as

$$\inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta})$$

This risk was studied in Gassiat, Kaddouri **and** Naulet 2023 in the strongly mixing setting. It was shown that for $\tilde{\alpha}_n = \frac{C}{\delta^5 \sqrt{n}}$

$$\frac{\delta^2(1-\tilde{\alpha}_n)}{2(1-\delta)} e^{-2\|\theta\|^2} \le \inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta}) \le (1-\delta) e^{-\frac{\|\theta\|^2}{2}}$$

The problem of clustering

Let $\eta = (\eta_i)_{1\leq i\leq n}$ and $\ell(\hat{\eta}, \eta) = \min_{\nu\in\{-1,1\}} |\hat{\eta} + \nu\eta|$ where $|.|$ is the Hamming distance. We define the risk of a clustering procedure $\hat{\eta}$ as

$$\mathcal{R}(\theta, \delta, \hat{\eta}) = \mathbb{E}\left[\ell(\hat{\eta}(X_{1:n}), \eta)\right]$$

and the Bayes risk of clustering as

$$\inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta})$$

This risk was studied in Gassiat, Kaddouri **and** Naulet 2023 in the strongly mixing setting. It was shown that for $\tilde{\alpha}_n = \frac{C}{\delta^5\sqrt{n}}$

$$\frac{\delta^2(1-\tilde{\alpha}_n)}{2(1-\delta)} e^{-2\|\theta\|^2} \leq \inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta}) \leq (1-\delta)e^{-\frac{\|\theta\|^2}{2}}$$

When δ is allowed to be very small, this fails to provide a precise understanding of the interplay between the strength of the signal and the Markovian dependence measured by δ.

Two frameworks

We study the risk of clustering observations in two frameworks:

- **Offline setting:** All observations are used in the clustering procedures. Clustering rules are of the form: $\hat{\eta}(Y_{1:n}) = (\hat{\eta}_i(Y_{1:n}))_{1\leq i\leq n}$.
  The Bayes risk of online clustering will be denoted $\boxed{\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right)}$.

Two frameworks

We study the risk of clustering observations in two frameworks:

- **Offline setting:** All observations are used in the clustering procedures. Clustering rules are of the form: $\hat{\eta}(Y_{1:n}) = (\hat{\eta}_i(Y_{1:n}))_{1 \leq i \leq n}$.
  The Bayes risk of online clustering will be denoted $\inf_{\hat{\eta}^{On}} \mathcal{R}(\theta, \delta, \hat{\eta})$.
- **Online setting:** Observations are clustered sequentially, with access to the past observations only at each step. Clustering rules are of the form: $\hat{\eta}(Y_{1:n}) = (\hat{\eta}_i(Y_{1:i}))_{1 \leq i \leq n}$.
  The Bayes risk of offline clustering will be denoted $\inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta})$.

Online framework

Theorem

When $\delta \geq \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \leq \left\{ \begin{array}{l} \\ \\ \end{array} \right.$$

Online framework

Theorem

When $\delta \geq \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \leq \begin{cases} \frac{1}{2} & \text{if } \|\theta\|^2 \leq 2\delta \\ & \end{cases}$$

Online framework

Theorem

When $\delta \geq \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \leq \begin{cases} \frac{1}{2} & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{4\delta}{\|\theta\|^2}\left(\log\left(\frac{\|\theta\|^2}{2\delta}\right) + 1\right) & \textit{if } 2\delta < \|\theta\|^2 \leq \log\left(\frac{1}{\delta}\right) \\ \end{cases}$$

Online framework

Theorem

When $\delta \geq \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \leq \begin{cases} \frac{1}{2} & \text{if } \|\theta\|^2 \leq 2\delta \\ \frac{4\delta}{\|\theta\|^2}\left(\log\left(\frac{\|\theta\|^2}{2\delta}\right) + 1\right) & \text{if } 2\delta < \|\theta\|^2 \leq \log\left(\frac{1}{\delta}\right) \\ e^{-\frac{\|\theta\|^2}{2}} & \text{if } \|\theta\|^2 > \log\left(\frac{1}{\delta}\right) \end{cases}$$

Online framework

Theorem

When $\delta \geq \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}(\theta, \delta, \hat{\eta}) \leq \begin{cases} \frac{1}{2} & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{4\delta}{\|\theta\|^2}\left(\log\left(\frac{\|\theta\|^2}{2\delta}\right) + 1\right) & \textit{if } 2\delta < \|\theta\|^2 \leq \log\left(\frac{1}{\delta}\right) \\ e^{-\frac{\|\theta\|^2}{2}} & \textit{if } \|\theta\|^2 > \log\left(\frac{1}{\delta}\right) \end{cases}$$

This upper-bound is reached by this online clustering procedure:

- For $i \in [\![1, k]\!]$:

$$\hat{\eta}_i(X_{1:n}) = \text{sign}\left(\langle X_i, \theta\rangle\right)$$

- For $i > k$:

$$\hat{\eta}_i(X_{1:n}) = \text{sign}\left(\left\langle \frac{1}{k}\sum_{j=i-k+1}^{i} X_j, \theta\right\rangle\right)$$

with $k = \left\lceil \frac{2}{\|\theta\|^2}\log\left(\frac{\|\theta\|^2}{2\delta}\right)\right\rceil$.

Online framework

Proposition

When $\delta \gtrsim \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \gtrsim \left\{ \right.$$

Online framework

Proposition

When $\delta \gtrsim \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \gtrsim \begin{cases} 1 & \text{if } \|\theta\|^2 \leq 2\delta \\ & \end{cases}$$

Online framework

Proposition

When $\delta \gtrsim \frac{1}{n}$,

$$\inf_{\hat{\eta}^{On}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \gtrsim \begin{cases} 1 & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{\delta}{\|\theta\|^2}\left(\log\left(\frac{\|\theta\|^2}{2\delta}\right)+1\right) & \textit{if } 2\delta < \|\theta\|^2 \leq \log\left(\frac{1}{\delta}\right) \end{cases}$$

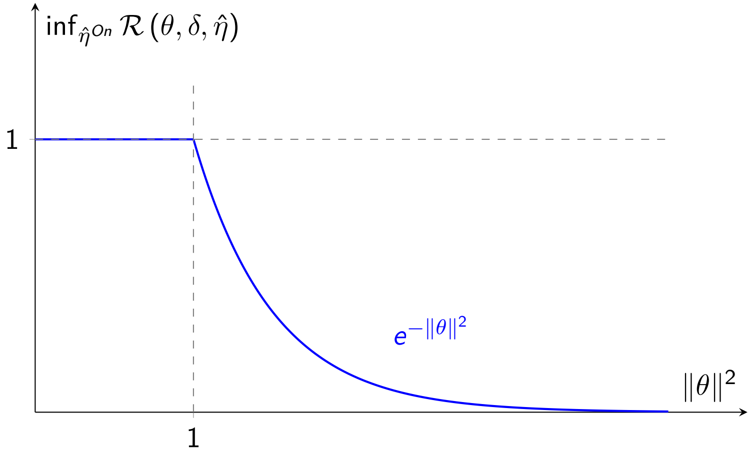
Online framework

Proposition

When $\delta \gtrsim \frac{1}{n}$,

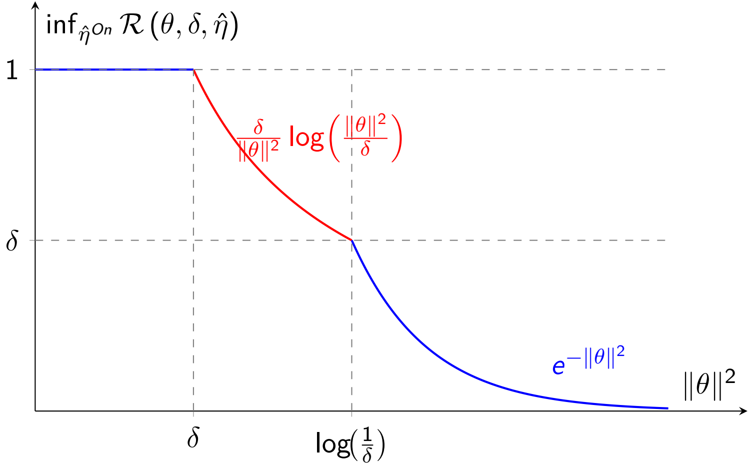
$$\inf_{\hat{\eta}^{On}} \mathcal{R}(\theta, \delta, \hat{\eta}) \gtrsim \begin{cases} 1 & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{\delta}{\|\theta\|^2}\left(\log\left(\frac{\|\theta\|^2}{2\delta}\right) + 1\right) & \textit{if } 2\delta < \|\theta\|^2 \leq \log\left(\frac{1}{\delta}\right) \\ \delta e^{-2\|\theta\|^2} & \textit{if } \|\theta\|^2 > \log\left(\frac{1}{\delta}\right) \end{cases}$$

Behavior of the Bayes risk of online clustering



Behavior of the Bayes risk of online clustering in the **strongly** mixing regime.

Behavior of the Bayes risk of online clustering



Behavior of the Bayes risk of online clustering in the **slowly** mixing regime.

Offline framework

Proposition

$$\inf_{\hat{\eta}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \lesssim \left\{ \vphantom{\begin{array}{c} a \\ b \\ c \\ d \end{array}} \right.$$

Offline framework

Proposition

$$\inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta}) \lesssim \begin{cases} \frac{1}{2} & \textit{if } \|\theta\|^2 \leq 2\delta \\ & \end{cases}$$

Proposition

$$\inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta}) \lesssim \begin{cases} \frac{1}{2} & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{2\delta}{\|\theta\|^2}\left(\log\left(\frac{\|\theta\|^2}{2\delta}\right) + 1\right) & \textit{if } 2\delta \leq \|\theta\|^2 \leq 1 \\ & \end{cases}$$

Offline framework

Proposition

$$\inf_{\hat{\eta}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \lesssim \begin{cases} \frac{1}{2} & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{2\delta}{\|\theta\|^2}\left(\log\left(\frac{\|\theta\|^2}{2\delta}\right)+1\right) & \textit{if } 2\delta \leq \|\theta\|^2 \leq 1 \\ \delta \log\left(\frac{1}{\delta}\right) e^{-\frac{\|\theta\|^2}{4}} & \textit{if } \|\theta\|^2 > 1 \end{cases}$$

Offline clustering procedure

Consider $\tilde{\eta}_{a-1}$ and $\tilde{\eta}_{a+1}$ are the online estimators defined by:

$$\tilde{\eta}_{a-1}(X_{a-k:a-1}) = \langle \frac{1}{k} \sum_{j=a-k}^{a-1} X_j, \theta \rangle, \ \tilde{\eta}_{a+1}(X_{a+1:a+k}) = \langle \frac{1}{k} \sum_{j=a+1}^{a+k} X_j, \theta \rangle$$

where $k = \left\lceil \frac{2}{\|\theta\|^2} \log\left(\frac{\|\theta\|^2}{2\delta}\right) \right\rceil$.
Consider the clustering procedure $\hat{\eta}$ such that for $a \in [\![1, n]\!]$:

$$\hat{\eta}_a(\tilde{\eta}_{a-1}, X_{a-k:a+k}, \tilde{\eta}_{a+1}) =$$
$$\begin{cases} \tilde{\eta}_{a-1} & \text{if } \tilde{\eta}_{a-1} = \tilde{\eta}_{a+1}, \|\theta\|^2 < \log\left(\frac{1}{\delta}\right) \text{ and } a \in [\![k+1, n-k]\!] \\ \operatorname{sign}\left(\langle X_a, \theta \rangle\right) & \text{else.} \end{cases}$$

Lower-bound

Proposition

When $\delta \gtrsim \frac{1}{n}$,

$$\inf_{\hat{\eta}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \gtrsim \left\{ \right.$$

Lower-bound

Proposition

When $\delta \gtrsim \frac{1}{n}$,

$$\inf_{\hat{\eta}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \gtrsim \begin{cases} 1 & \textit{if } \|\theta\|^2 \leq 2\delta \\ & \end{cases}$$

Lower-bound

Proposition

When $\delta \gtrsim \frac{1}{n}$,

$$\inf_{\hat{\eta}} \mathcal{R}(\theta, \delta, \hat{\eta}) \gtrsim \begin{cases} 1 & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{\delta}{\|\theta\|^2} & \textit{if } 2\delta < \|\theta\|^2 \leq 1 \end{cases}$$

Lower-bound

Proposition

When $\delta \gtrsim \frac{1}{n}$,

$$\inf_{\hat{\eta}} \mathcal{R}\left(\theta, \delta, \hat{\eta}\right) \gtrsim \begin{cases} 1 & \textit{if } \|\theta\|^2 \leq 2\delta \\ \frac{\delta}{\|\theta\|^2} & \textit{if } 2\delta < \|\theta\|^2 \leq 1 \\ \delta e^{-2\|\theta\|^2} & \textit{if } \|\theta\|^2 > 1 \end{cases}$$

Behavior of the Bayes risk of offline clustering

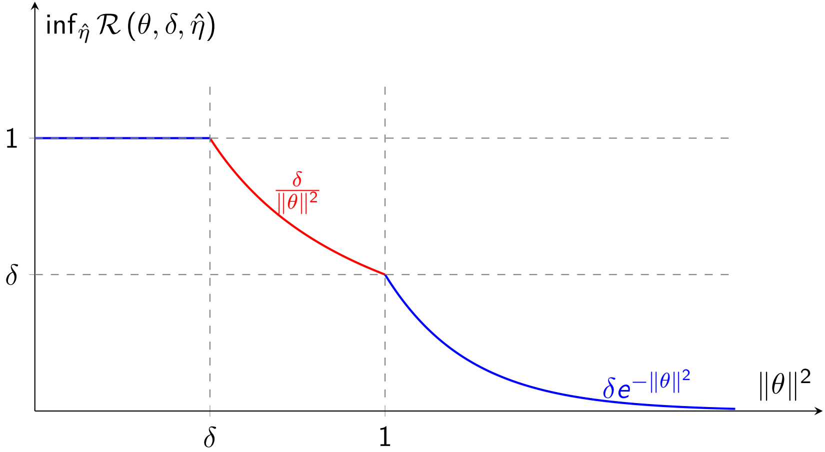


Figure: Behavior of the Bayes risk of offline clustering in the slowly mixing regime

Adaptation

WLOG, assume $n = k\ell$. For each bucket $i \in [\![1, \ell]\!]$, consider the sample mean of k observations inside the i-th bucket

$$\tilde{X}_i = \frac{1}{k} \sum_{j=k(i-1)+1}^{ki} X_j$$

Note that

$$\tilde{X}_i = \theta \bar{\eta}_i + \frac{\xi_i}{\sqrt{k}}$$

where $\bar{\eta}_i = \frac{1}{k} \sum_{j=k(i-1)+1}^{ki} \eta_j$ and ξ_i is a standard Gaussian vector.

Adaptation

WLOG, assume $n = k\ell$. For each bucket $i \in [\![1, \ell]\!]$, consider the sample mean of k observations inside the i-th bucket

$$\tilde{X}_i = \frac{1}{k} \sum_{j=k(i-1)+1}^{ki} X_j$$

Note that

$$\tilde{X}_i = \theta \bar{\eta}_i + \frac{\xi_i}{\sqrt{k}}$$

where $\bar{\eta}_i = \frac{1}{k} \sum_{j=k(i-1)+1}^{ki} \eta_j$ and ξ_i is a standard Gaussian vector.
We stack the ℓ terms in an $\mathbb{R}^{d \times \ell}$ matrix form as follows:

$$\tilde{X} = \theta \bar{\eta}^\top + \frac{\xi}{\sqrt{k}}$$

We then consider the Gram matrix of observations $\tilde{X}\tilde{X}^\top$.

Adaptation

It is easy to see that:

$$\mathbb{E}\left[\frac{1}{\ell}\tilde{X}\tilde{X}^\top\right] = \frac{\mathbb{E}\left[\|\bar{\eta}\|^2\right]}{\ell}\theta\theta^\top + \frac{1}{k}\mathbf{I}_d$$

Since our goal is adaptation up to multiplicative constants, it is natural to consider the following estimator of $\|\theta\|$:

$$\|\hat{\theta}(\ell)\| = \left\|\frac{1}{\ell}\tilde{X}\tilde{X}^\top - \frac{1}{k}\mathbf{I}_d\right\|_{op}^{1/2}$$

Proposition

There exists $c > 0$ and $C > 0$ such that for $\varepsilon > 0$ and $\ell \geq n\|\theta\|^2 \vee \frac{d}{\varepsilon^2} \vee \frac{2}{3\varepsilon}n\delta$,

$$\mathbb{P}\left(\left|\|\hat{\theta}(\ell)\|^2 - \|\theta\|^2\right| \geq C\varepsilon\frac{\ell}{n}\right) \leq e^{-\frac{c\varepsilon^2\ell}{2}}.$$

Adaptation

Define now $\mathbf{I}_\ell = \left[\|\hat{\theta}(\ell)\|^2 - \frac{5}{6}\frac{\ell}{n},\ \|\hat{\theta}(\ell)\|^2 + \frac{5}{6}\frac{\ell}{n}\right]$ and $\hat{\ell} = \min\left\{\tilde{\ell} \mid \cap_{\ell \geq \tilde{\ell}} \mathbf{I}_\ell \neq \varnothing\right\}$. Let $\tilde{\theta} \in \cap_{\ell \geq \tilde{\ell}} \mathbf{I}_\ell$.

Theorem

There exist positive absolute constants c_1, c_2 *and* c_3 *such that in the regime where* $\|\theta\|^2 \geq c_1 \left(\frac{d}{n} \vee \delta\right)$,

$$\mathbb{P}\left(\left|\tilde{\theta}^2 - \|\theta\|^2\right| \geq \frac{\|\theta\|^2}{2}\right) \leq c_2 e^{-c_3 n \|\theta\|^2}.$$

This shows that $\tilde{\theta}^2$ is an appropriate estimator of $\|\theta\|^2$ up to multiplicative constants.

Conclusion

- Clustering is easier under the slowly mixing regime.
- Unexpected behavior of the Bayes risk in some regimes.
- Many questions are still not answered: full adaptation, high dimension, estimation of δ.

References

Gassiat, Elisabeth, Ibrahim Kaddouri **and** Zacharie Naulet (2023). "Clustering and Classification Risks in Non-parametric Hidden Markov and I.I.D. Models". **in**arXiv: 2309.12238 [stat.ST].

Karagulyan, Vahe **and** Mohamed Ndaoud (2024). "Adaptive Mean Estimation in the Hidden Markov sub-Gaussian Mixture Model". **in**arXiv: 2406.12446 [stat.ST].