Domain selection for Gaussian Processes

An application to electrocardiogram signals

Nicolás Hernández

University College London

Oct. 26th, 2023

This work...

Gabriel Martos, PhD. Associate Professor at the Department of Mathematics and Statistics of Universidad Torcuato Di Tella, Buenos Aires, Argentina.

Hernández, N., & Martos, G. (2023). Domain Selection for Gaussian Process Data: An application to electrocardiogram signals. arXiv preprint arXiv:2306.00538.

(some) issues in the High-dimensional setting...

what is the deal with having too much data?

Overfitting

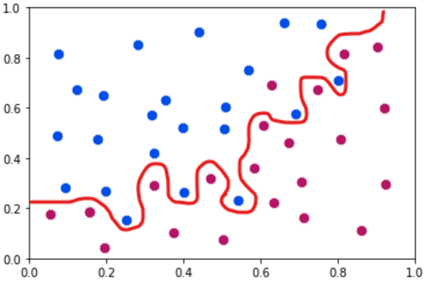
Estimation precision

Computational complexity

Interpretability

what is the deal with having too much data?

- **Overfitting**



what is the deal with having too much data?

Overfitting

- **Estimation precision**

what is the deal with having too much data?

Overfitting

Estimation precision

- **Computational complexity**

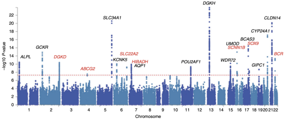
what is the deal with having too much data?

Overfitting

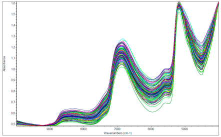
Estimation precision

Computational complexity

- **Interpretability**



➡ Genetic Data



➡ Spectral Data (Chemometrics)

➡ Health records, e.g. electrocardiograms

(some) solution to these problems...

How can we tackle these issues?

- Dimension reduction methods (e.g. PCA)

Regularisation (e.g. LASSO)

Variable selection

(some) solution to these issues...

How can we tackle these issues?

Dimension reduction methods (e.g. PCA)

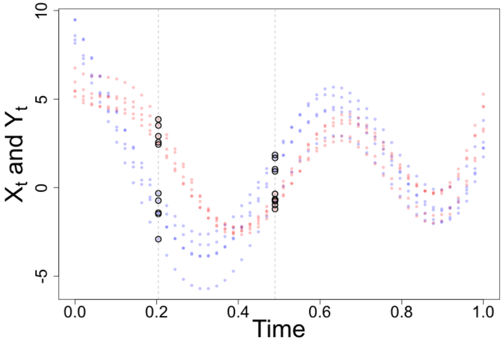
- Regularisation (e.g. LASSO)

Variable selection

How can we tackle these issues?

- Dimension reduction methods (e.g. PCA)
- Regularisation (e.g. LASSO)
- **Variable selection**

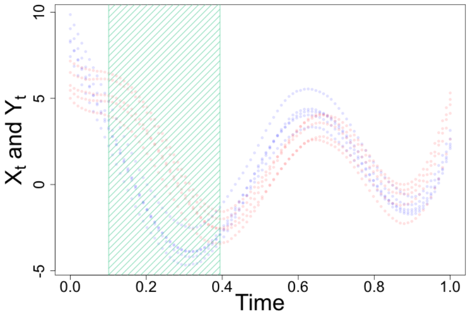
Focus on variable selection...



Select isolated variables

- find models with good prediction power,
- estimate the true "sparsity pattern".

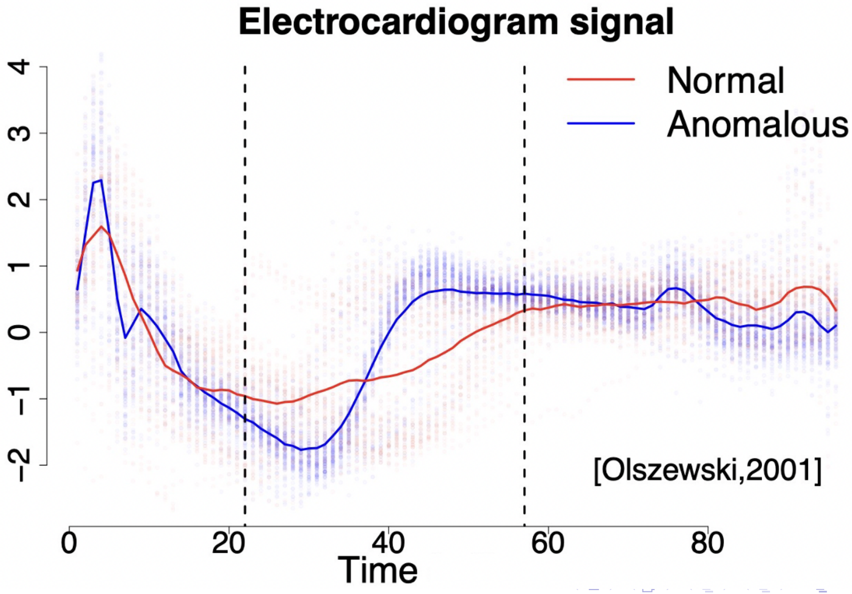
... but adding some restrictions: Domain Selection



Domain (interval) selection

- Select set of variables (intervals).
- Take advantage of the covariance structure.
- Variables are recorded almost continuously.

Motivation


Electrocardiogram signal
Normal
Anomalous
[Olszewski,2001]
Time
0
20
40
60
80
−2
−1
0
1
2
3
4

Objective: Domain Selection

For a pair of random processes, what is the region (domain) where they statistically differ the most?

- **Inference:**
  - Characterise and quantify the uncertainty around the estimation of the domain bounds.
  - Improve the power of a hypothesis test, i.e: two-sample test.
- **Prediction:** improve a classification model outcome, i.e: the misclassification error rate.
- **Computational burden:** reduce time and memory storage in the big data context. Future collection Strategy.

Objective: Domain Selection

For a pair of random processes, what is the region (domain) where they statistically differ the most?

- **Inference:**
  - Characterise and quantify the uncertainty around the estimation of the domain bounds.
  - Improve the power of a hypothesis test, i.e: two-sample test.
- **Prediction:** improve a classification model outcome, i.e: the misclassification error rate.
- **Computational burden:** reduce time and memory storage in the big data context. Future collection Strategy.

Objective: Domain Selection

For a pair of random processes, what is the region (domain) where they statistically differ the most?

- **Inference:**
  - Characterise and quantify the uncertainty around the estimation of the domain bounds.
  - Improve the power of a hypothesis test, i.e: two-sample test.
- **Prediction:** improve a classification model outcome, i.e: the misclassification error rate.
- **Computational burden:** reduce time and memory storage in the big data context. Future collection Strategy.

Objective: Domain Selection

For a pair of random processes, what is the region (domain) where they statistically differ the most?

- **Inference:**
  - Characterise and quantify the uncertainty around the estimation of the domain bounds.
  - Improve the power of a hypothesis test, i.e: two-sample test.
- **Prediction:** improve a classification model outcome, i.e: the misclassification error rate.
- **Computational burden:** reduce time and memory storage in the big data context. Future collection Strategy.

Domain Selection: How?

compute differences with a local approach

↓

Kullback-Leibler divergence

↓

Let P and Q be two distributions of a continuous random variable x. Then the KL is:

$$\text{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) log(\frac{p(x)}{q(x)}) dx$$

Domain Selection: How?

compute differences with a local approach

Kullback-Leibler divergence + **Gaussian** framework

$$\mathrm{KL}(X||Y) = \frac{1}{2}\left(\frac{\sigma_X^2}{\sigma_Y^2} - 1 + \frac{(\mu_X - \mu_Y)^2}{\sigma_Y^2} + \ln\left(\frac{\sigma_Y^2}{\sigma_X^2}\right)\right).$$

Domain Selection: How?

compute differences with a local approach

↓

Kullback-Leibler divergence + **Gaussian** framework

↓

$$\mathrm{KL}(X||Y) = \frac{1}{2}\left(\frac{\sigma_X^2}{\sigma_Y^2} - 1 + \frac{(\mu_X - \mu_Y)^2}{\sigma_Y^2} + \ln\left(\frac{\sigma_Y^2}{\sigma_X^2}\right)\right).$$

Domain Selection: How?

compute differences with a local approach

↓

Kullback-Leibler divergence + **Gaussian** framework

↓

$$\mathrm{KL}(X||Y) = \frac{1}{2}\left(\frac{\sigma_X^2}{\sigma_Y^2} - 1 + \boldsymbol{\frac{(\mu_X - \mu_Y)^2}{\sigma_Y^2}} + \ln\left(\frac{\sigma_Y^2}{\sigma_X^2}\right)\right).$$

Domain Selection: How?

compute differences with a local approach

⬇

Kullback-Leibler divergence + **Gaussian** framework

⬇

$$\mathrm{KL}(X||Y) = \frac{1}{2}\left(\frac{\boldsymbol{\sigma_X^2}}{\boldsymbol{\sigma_Y^2}} - 1 + \frac{(\mu_X - \mu_Y)^2}{\sigma_Y^2} + \mathbf{ln}\left(\frac{\boldsymbol{\sigma_Y^2}}{\boldsymbol{\sigma_X^2}}\right)\right).$$

Learning problem and local features

- Let $X(t) \sim GP(\mu_X(t), \sigma_X(t,s))$ be a GP **indexed on the compact set** $T \subset \mathbb{R}$, with:
  - $\mu_X(t) = E\{X(t)\}$ and $\sigma_X(t,s) = E\{(X(t) - \mu_X(t))(X(s) - \mu_X(s))\}$.
  - Likewise for $Y(t)$

Learn the interval where $X(t)$ and $Y(t)$ statistically differ the most.

- Estimate a compact subset $A \subset T$ with $\lambda(A) > 0$:

(A) **Differences in Mean.** $t \in A$: $|\mu_X(t) - \mu_Y(t)| > \nu$, and $t' \notin A : |\mu_X(t') - \mu_Y(t')| \leq \nu$.

Differences in Variance. $(t,s) \in A \times A$: $|\sigma_X(t,s) - \sigma_Y(t,s)| > \nu$, and $(t',s') \notin A \times A : |\sigma_X(t',s') - \sigma_Y(t',s')| \leq \nu$.

(A) and (B) simultaneously, possibly on different subsets A_μ and A_σ and for different ν_μ and ν_σ.

Learning problem and local features

- Let $X(t) \sim GP(\mu_X(t), \sigma_X(t,s))$ be a GP **indexed on the compact set** $T \subset \mathbb{R}$, with:
  - $\mu_X(t) = E\{X(t)\}$ and $\sigma_X(t,s) = E\{(X(t) - \mu_X(t))(X(s) - \mu_X(s))\}$.
  - Likewise for $Y(t)$

Learn the interval where $X(t)$ and $Y(t)$ statistically differ the most.

- Estimate a compact subset $A \subset T$ with $\lambda(A) > 0$:

  (A) **Differences in Mean.** $t \in A$: $|\mu_X(t) - \mu_Y(t)| > \nu$, and $t' \notin A : |\mu_X(t') - \mu_Y(t')| \leq \nu$.

  Differences in Variance. $(t,s) \in A \times A$: $|\sigma_X(t,s) - \sigma_Y(t,s)| > \nu$, and $(t',s') \notin A \times A : |\sigma_X(t',s') - \sigma_Y(t',s')| \leq \nu$.

  (A) and (B) simultaneously, possibly on different subsets A_μ and A_σ and for different ν_μ and ν_σ.

Learning problem and local features

- Let $X(t) \sim GP(\mu_X(t), \sigma_X(t,s))$ be a GP indexed on the compact set $T \subset \mathbb{R}$, with:
  - $\mu_X(t) = E\{X(t)\}$ and $\sigma_X(t,s) = E\{(X(t) - \mu_X(t))(X(s) - \mu_X(s))\}$.
  - Likewise for $Y(t)$

Learn the interval where $X(t)$ and $Y(t)$ statistically differ the most.

- Estimate a compact subset $A \subset T$ with $\lambda(A) > 0$:

(A) **Differences in Mean.** $t \in A$: $|\mu_X(t) - \mu_Y(t)| > \nu$, and $t' \notin A : |\mu_X(t') - \mu_Y(t')| \leq \nu$.

(B) **Differences in Variance.** $(t,s) \in A \times A$: $|\sigma_X(t,s) - \sigma_Y(t,s)| > \nu$, and $(t',s') \notin A \times A : |\sigma_X(t',s') - \sigma_Y(t',s')| \leq \nu$.

(A) and (B) simultaneously, possibly on different subsets A_μ and A_σ and for different ν_μ and ν_σ.

Learning problem and local features

- Let $X(t) \sim GP(\mu_X(t), \sigma_X(t,s))$ be a GP indexed on the compact set $T \subset \mathbb{R}$, with:
  - $\mu_X(t) = E\{X(t)\}$ and $\sigma_X(t,s) = E\{(X(t) - \mu_X(t))(X(s) - \mu_X(s))\}$.
  - Likewise for $Y(t)$

Learn the interval where $X(t)$ and $Y(t)$ statistically differ the most.

- Estimate a compact subset $A \subset T$ with $\lambda(A) > 0$:

(A) **Differences in Mean.** $t \in A$: $|\mu_X(t) - \mu_Y(t)| > \nu$, and $t' \notin A : |\mu_X(t') - \mu_Y(t')| \leq \nu$.

(B) **Differences in Variance.** $(t,s) \in A \times A$: $|\sigma_X(t,s) - \sigma_Y(t,s)| > \nu$, and $(t',s') \notin A \times A : |\sigma_X(t',s') - \sigma_Y(t',s')| \leq \nu$.

(C) **(A) and (B) simultaneously**, possibly on different subsets A_μ and A_σ and for different ν_μ and ν_σ.

The KL divergence in the GP context

- In practice GP data is recorded over a finite grid $\mathcal{T} = (t_1, \dots, t_p)$.
- In this case it corresponds to realisations of p-variate Gaussian random vectors and the KL has the following closed form:

$$\mathrm{KL}_{\mathcal{T}}(\mathbf{X}||\mathbf{Y}) \equiv \frac{1}{2}\left(\mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathcal{T},Y}^{-1}\boldsymbol{\Sigma}_{\mathcal{T},X} - \mathbf{I}_p\right) + \Delta_{\mathcal{T}}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathcal{T},Y}^{-1}\Delta_{\mathcal{T}} + \ln\left(\frac{\det \boldsymbol{\Sigma}_{\mathcal{T},Y}}{\det \boldsymbol{\Sigma}_{\mathcal{T},X}}\right)\right),$$

where:

- $\Delta_{\mathcal{T}} = \left(\boldsymbol{\mu}_{\mathcal{T},Y} - \boldsymbol{\mu}_{\mathcal{T},X}\right)$,
- $\mathrm{tr}(\boldsymbol{\Sigma})$ denote the trace of $\boldsymbol{\Sigma}$, and
- $\det(\boldsymbol{\Sigma})$ the determinant of $\boldsymbol{\Sigma}$.

Domain Selection using the local-KL divergence

- For any subset $\mathcal{A} \subseteq \mathcal{T}$, we can also define the local–KL divergence as follows:

$$\mathrm{KL}_{\mathcal{A}}(X||Y) \equiv \frac{1}{2}\left(\mathrm{tr}\left(\mathbf{\Sigma}_{\mathcal{A},Y}^{-1}\mathbf{\Sigma}_{\mathcal{A},X} - \mathbf{I}_{|\mathcal{A}|}\right) + \Delta_{\mathcal{A}}^{\mathsf{T}}\mathbf{\Sigma}_{\mathcal{A},Y}^{-1}\Delta_{\mathcal{A}} + \ln\left(\frac{\det \mathbf{\Sigma}_{\mathcal{A},Y}}{\det \mathbf{\Sigma}_{\mathcal{A},X}}\right)\right)$$

- Consider $\mathcal{A} \in \mathcal{C}_{\mathcal{T}}$, collection of all contiguous subsets from $\mathcal{T}$

- $\mathcal{A}^*(c)$, with $c \in (0,1)$, is the domain with **maximum divergence**

$$\max_{\mathcal{A}\in\mathcal{C}_{\mathcal{T}}} \mathrm{KL}_{\mathcal{A}}(X||Y), \text{ s.t. } \mathrm{len}(\mathcal{A}) \leq c\lambda(\mathcal{T}).$$

Domain Selection using the local-KL divergence

- For any subset $\mathcal{A} \subseteq \mathcal{T}$, we can also define the local–KL divergence as follows:

$$\mathrm{KL}_{\mathcal{A}}(X||Y) \equiv \frac{1}{2}\left(\mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathcal{A},Y}^{-1}\boldsymbol{\Sigma}_{\mathcal{A},X} - \mathbf{I}_{|\mathcal{A}|}\right) + \Delta_{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathcal{A},Y}^{-1}\Delta_{\mathcal{A}} + \ln\left(\frac{\det \boldsymbol{\Sigma}_{\mathcal{A},Y}}{\det \boldsymbol{\Sigma}_{\mathcal{A},X}}\right)\right)$$

- Consider $\mathcal{A} \in \mathcal{C}_{\mathcal{T}}$, collection of all contiguous subsets from $\mathcal{T}$
- $\boldsymbol{\mathcal{A}^*(c)}$, with $c \in (0,1)$, is the domain with **maximum divergence**

$$\max_{\mathcal{A}\in\mathcal{C}_{\mathcal{T}}} \mathrm{KL}_{\mathcal{A}}(X||Y), \text{ s.t. } \mathrm{len}(\mathcal{A}) \leq c\lambda(T).$$

Local-KL Divergence properties

The $\mathrm{KL}_{\mathcal{A}}(X||Y)$ is a set function that satisfies the following properties:

(a) **Non-negative:** For fixed GPs X and Y, it holds: $\mathrm{KL}_{\mathcal{A}}(X||Y) : \mathcal{P}_{\mathcal{T}} \to \mathbb{R}_0^+$ and $\mathrm{KL}_{\mathcal{A}}(X||Y) = 0$ if and only if $\mu_X(t) = \mu_Y(t)$ for all $t \in \mathcal{A}$ and $\sigma_X(t,s) = \sigma_Y(t,s)$ for all $(t,s) \in \mathcal{A} \times \mathcal{A}$.

(b) The local KL divergence is **upper bounded** (i.e. $\mathrm{KL}_{\mathcal{T}}(X||Y) < \infty$) and a **monotone** set function (i.e. for $\mathcal{A}' \subseteq \mathcal{A}$ it holds: $\mathrm{KL}_{\mathcal{A}'}(X||Y) \leq \mathrm{KL}_{\mathcal{A}}(X||Y)$)

(c) The local divergence $\mathrm{KL}_{\mathcal{A}}(X||Y)$ is a **continuous** set function in $\mathcal{C}_{\mathcal{T}}$, the collection of all contiguous subsets from the ground set $\mathcal{T}$.

Properties + $\mathcal{P}_{\mathcal{T}}$ is finite $\Rightarrow \mathcal{A}^*(c)$ exists.

Local-KL Divergence properties

The $\mathrm{KL}_{\mathcal{A}}(X||Y)$ is a set function that satisfies the following properties:

(a) **Non-negative:** For fixed GPs X and Y, it holds: $\mathrm{KL}_{\mathcal{A}}(X||Y) : \mathcal{P}_{\mathcal{T}} \to \mathbb{R}_0^+$ and $\mathrm{KL}_{\mathcal{A}}(X||Y) = 0$ if and only if $\mu_X(t) = \mu_Y(t)$ for all $t \in \mathcal{A}$ and $\sigma_X(t,s) = \sigma_Y(t,s)$ for all $(t,s) \in \mathcal{A} \times \mathcal{A}$.

(b) The local KL divergence is **upper bounded** (i.e. $\mathrm{KL}_{\mathcal{T}}(X||Y) < \infty$) and a **monotone** set function (i.e. for $\mathcal{A}' \subseteq \mathcal{A}$ it holds: $\mathrm{KL}_{\mathcal{A}'}(X||Y) \leq \mathrm{KL}_{\mathcal{A}}(X||Y)$)

(c) The local divergence $\mathrm{KL}_{\mathcal{A}}(X||Y)$ is a **continuous** set function in $\mathcal{C}_{\mathcal{T}}$, the collection of all contiguous subsets from the ground set $\mathcal{T}$.

Properties + $\mathcal{P}_{\mathcal{T}}$ is finite $\Rightarrow \mathcal{A}^*(c)$ exists.

Local-KL Divergence properties

The $\mathrm{KL}_{\mathcal{A}}(X||Y)$ is a set function that satisfies the following properties:

(a) **Non-negative:** For fixed GPs X and Y, it holds: $\mathrm{KL}_{\mathcal{A}}(X||Y) : \mathcal{P}_{\mathcal{T}} \to \mathbb{R}_0^+$ and $\mathrm{KL}_{\mathcal{A}}(X||Y) = 0$ if and only if $\mu_X(t) = \mu_Y(t)$ for all $t \in \mathcal{A}$ and $\sigma_X(t,s) = \sigma_Y(t,s)$ for all $(t,s) \in \mathcal{A} \times \mathcal{A}$.

(b) The local KL divergence is **upper bounded** (i.e. $\mathrm{KL}_{\mathcal{T}}(X||Y) < \infty$) and a **monotone** set function (i.e. for $\mathcal{A}' \subseteq \mathcal{A}$ it holds: $\mathrm{KL}_{\mathcal{A}'}(X||Y) \leq \mathrm{KL}_{\mathcal{A}}(X||Y)$)

(c) The local divergence $\mathrm{KL}_{\mathcal{A}}(X||Y)$ is a **continuous** set function in $\mathcal{C}_{\mathcal{T}}$, the collection of all contiguous subsets from the ground set $\mathcal{T}$.

Properties + $\mathcal{P}_{\mathcal{T}}$ is finite $\Rightarrow \mathcal{A}^*(c)$ exists.

Local-KL divergence estimation

- Samples recorded over the same discrete grid $\mathcal{T}$: $\mathcal{D}_X = \{\mathbf{x}_i\}_{i=1}^{n}$ and $\mathcal{D}_Y = \{\mathbf{y}_j\}_{j=1}^{m}$ drawn from $GP(\mu_X(t), \sigma_X(t,s))$ and $GP(\mu_Y(t), \sigma_Y(t,s))$.
- The **Maximum Likelihood** estimates are given by:

$$\widehat{\mu}_{\mathcal{T},X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i; \quad \widehat{\boldsymbol{\Sigma}}_{\mathcal{T},X} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \widehat{\mu}_{\mathcal{T},X})(\mathbf{x}_i - \widehat{\mu}_{\mathcal{T},X})^T,$$

(analogous expression holds for $\mu_{\mathcal{T},Y}, \boldsymbol{\Sigma}_{\mathcal{T},Y}$).

- The trace and determinant are continuous functions in the space of real symmetric matrices, then $\widehat{\mathrm{KL}}_{\mathcal{A}}(X||Y)$ inherits (fixed p), consistency, asymptotic normality and efficiency,

Local-KL divergence estimation

- Samples recorded over the same discrete grid $\mathcal{T}$: $\mathcal{D}_X = \{\mathbf{x}_i\}_{i=1}^n$ and $\mathcal{D}_Y = \{\mathbf{y}_j\}_{j=1}^m$ drawn from $GP(\mu_X(t), \sigma_X(t,s))$ and $GP(\mu_Y(t), \sigma_Y(t,s))$.
- The **Maximum Likelihood** estimates are given by:

$$\widehat{\boldsymbol{\mu}}_{\mathcal{T},X} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i; \quad \widehat{\boldsymbol{\Sigma}}_{\mathcal{T},X} = \frac{1}{n}\sum_{i=1}^n (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_{\mathcal{T},X})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_{\mathcal{T},X})^T,$$

(analogous expression holds for $\boldsymbol{\mu}_{\mathcal{T},Y}, \boldsymbol{\Sigma}_{\mathcal{T},Y}$).

- The trace and determinant are continuous functions in the space of real symmetric matrices, then $\widehat{\mathrm{KL}}_{\mathcal{A}}(X||Y)$ inherits (fixed p), consistency, asymptotic normality and efficiency,

Local-KL divergence estimation

- Samples recorded over the same discrete grid $\mathcal{T}$: $\mathcal{D}_X = \{\mathbf{x}_i\}_{i=1}^n$ and $\mathcal{D}_Y = \{\mathbf{y}_j\}_{j=1}^m$ drawn from $GP(\mu_X(t), \sigma_X(t,s))$ and $GP(\mu_Y(t), \sigma_Y(t,s))$.
- The **Maximum Likelihood** estimates are given by:

$$\widehat{\boldsymbol{\mu}}_{\mathcal{T},X} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i; \quad \widehat{\boldsymbol{\Sigma}}_{\mathcal{T},X} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_{\mathcal{T},X})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_{\mathcal{T},X})^T,$$

  (analogous expression holds for $\boldsymbol{\mu}_{\mathcal{T},Y}, \boldsymbol{\Sigma}_{\mathcal{T},Y}$).
- The trace and determinant are continuous functions in the space of real symmetric matrices, then $\widehat{\text{KL}}_{\mathcal{A}}(X||Y)$ inherits (fixed p), consistency, asymptotic normality and efficiency,

Some notes on the estimation...

- **Symmetrised** KL divergence: $\dfrac{\text{KL}(X||Y) + \text{KL}(Y||X)}{2}$

- **Regularised** covariance matrix:
  $\widehat{\boldsymbol{\Sigma}}_\eta = \eta\,\widehat{\boldsymbol{\Sigma}} + (1-\eta)\,\text{diag}(\widehat{\boldsymbol{\Sigma}}), \text{ for } \eta \in [0,1].$

- Sampling designs and misalignment: data **asynchrony** may act as a **confounding** factor when the aim is the estimation of the interval of local maximum KL divergence

- Smoothing: **low signal to noise ratio** leading to problems in the estimation of the interval of local maximum KL divergence

Some notes on the estimation...

- **Symmetrised** KL divergence: $\dfrac{\text{KL}(X||Y) + \text{KL}(Y||X)}{2}$
- **Regularised** covariance matrix:
  $\widehat{\boldsymbol{\Sigma}}_\eta = \eta\,\widehat{\boldsymbol{\Sigma}} + (1-\eta)\,\text{diag}(\widehat{\boldsymbol{\Sigma}}),\;$ for $\eta \in [0,1]$.
- Sampling designs and misalignment: data **asynchrony** may act as a **confounding** factor when the aim is the estimation of the interval of local maximum KL divergence
- Smoothing: **low signal to noise ratio** leading to problems in the estimation of the interval of local maximum KL divergence

Some notes on the estimation...

- **Symmetrised** KL divergence: $\dfrac{\mathrm{KL}(X||Y)+\mathrm{KL}(Y||X)}{2}$
- **Regularised** covariance matrix:
  $\widehat{\boldsymbol{\Sigma}}_\eta = \eta\,\widehat{\boldsymbol{\Sigma}} + (1-\eta)\,\mathrm{diag}(\widehat{\boldsymbol{\Sigma}}),$ for $\eta \in [0,1]$.
- Sampling designs and misalignment: data **asynchrony** may act as a **confounding** factor when the aim is the estimation of the interval of local maximum KL divergence
- Smoothing: **low signal to noise ratio** leading to problems in the estimation of the interval of local maximum KL divergence

Some notes on the estimation...

- **Symmetrised** KL divergence: $\dfrac{\mathrm{KL}(X||Y) + \mathrm{KL}(Y||X)}{2}$
- **Regularised** covariance matrix:
  $\widehat{\boldsymbol{\Sigma}}_\eta = \eta\,\widehat{\boldsymbol{\Sigma}} + (1-\eta)\,\mathrm{diag}(\widehat{\boldsymbol{\Sigma}}),$ for $\eta \in [0,1]$.
- Sampling designs and misalignment: data **asynchrony** may act as a **confounding** factor when the aim is the estimation of the interval of local maximum KL divergence
- Smoothing: **low signal to noise ratio** leading to problems in the estimation of the interval of local maximum KL divergence

One-shot experiment

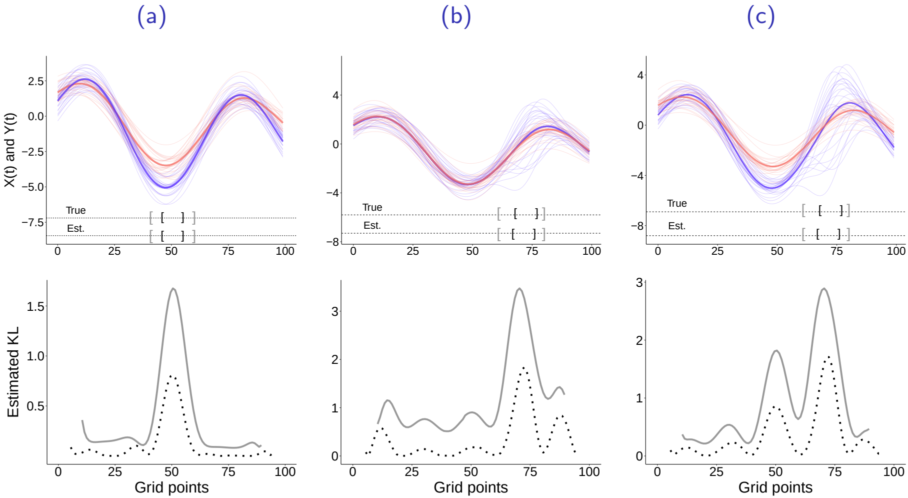


Figure: **One shot experiment:** Lower panels show the estimated local KL divergence for different interval lengths: $c = 0.1$ $(\cdots\cdots)$ and $c = 0.2$ (——)

Monte Carlo experiment

- MC replicates $= 1000$
- Sample size: $n = \{50, 100, 500, 1000\}$.
- Grid points: $p = \{50, 100, 200, 500\}$.
- c possible lengths: $(1\% - 99\%)$.
- Av. Jaccard Distance: $\text{AJD} = E\left\{ \int_0^1 \left[1 - \frac{|\mathcal{A}^*(c) \cap \widehat{\mathcal{A}}^*(c)|}{|\mathcal{A}^*(c) \cup \widehat{\mathcal{A}}^*(c)|}\right] \mathrm{d}c \right\}$

Monte Carlo results

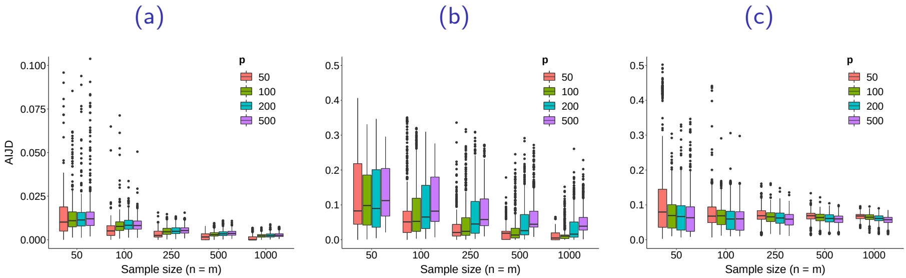
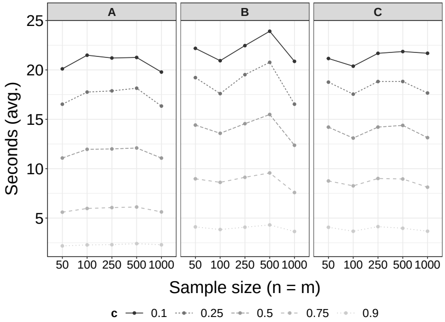


Figure: Empirical distribution of AIJD for different sample sizes n, m and grid resolution levels p.

Computational complexity



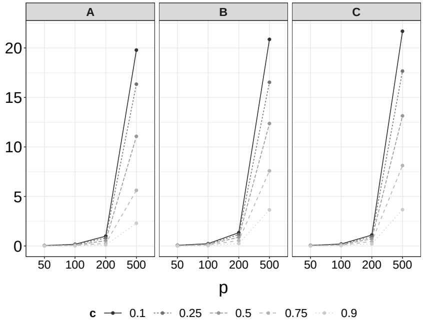


Figure: Average computational time required to estimate intervals of local maximum divergence.

Monitoring Electrocardiogram signals

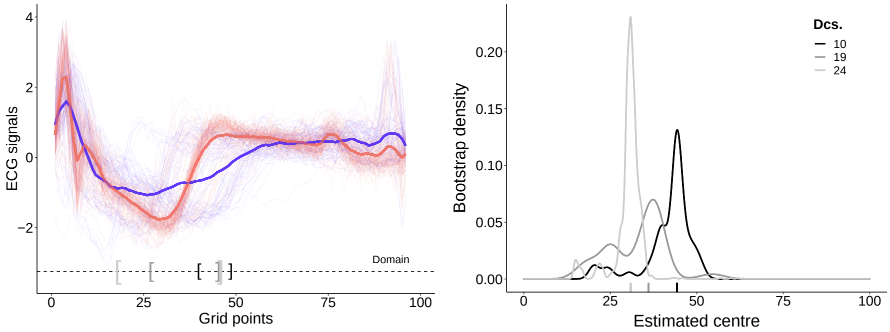


Figure: Left: ECG signals and selected domain for interval lengths 10, 19 and 24 dcs. Right: Bootstrap densities for the interval centre (same lengths)

ECG Classification

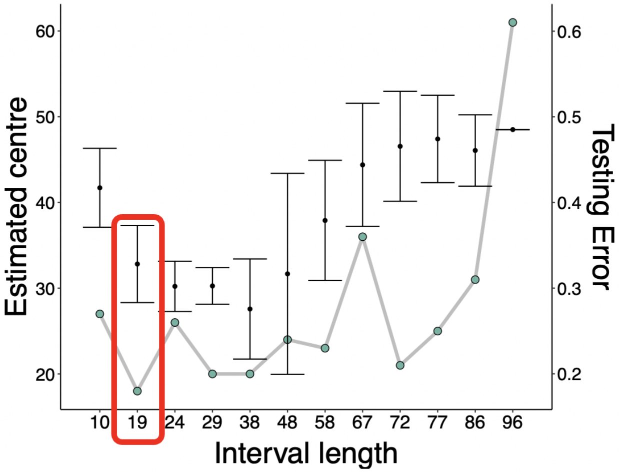


Figure: Estimated centre and classification error using a Discriminant model for different selected domains.

Summary and future work

✓ Tackle the problem of domain selection for electrocardiogram signals.

✓ Parameter of interest: interval ⇒ Covariance structure + almost continuously recorded data.

✓ Use the KL divergence and GP to develop an easy to implement algorithm for domain selection.

✓ Propose an estimator for $\mathcal{A}^*(c)$, and a nonparametric approach to assess the estimation uncertainty

⇒ Consider other model frameworks (non Gaussian), e.g.: *Wasserstein* divergence.

Summary and future work

✓ Tackle the problem of domain selection for electrocardiogram signals.

✓ Parameter of interest: interval ⇒ Covariance structure + almost continuously recorded data.

✓ Use the KL divergence and GP to develop an easy to implement algorithm for domain selection.

✓ Propose an estimator for $\mathcal{A}^*(c)$, and a nonparametric approach to assess the estimation uncertainty

⇒ Consider other model frameworks (non Gaussian), e.g.: *Wasserstein* divergence.

Summary and future work

✓ Tackle the problem of domain selection for electrocardiogram signals.

✓ Parameter of interest: interval ⇒ Covariance structure + almost continuously recorded data.

✓ Use the KL divergence and GP to develop an easy to implement algorithm for domain selection.

✓ Propose an estimator for $\mathcal{A}^*(c)$, and a nonparametric approach to assess the estimation uncertainty

⇒ Consider other model frameworks (non Gaussian), e.g.: *Wasserstein* divergence.

Summary and future work

✓ Tackle the problem of domain selection for electrocardiogram signals.

✓ Parameter of interest: interval $\Rightarrow$ Covariance structure + almost continuously recorded data.

✓ Use the KL divergence and GP to develop an easy to implement algorithm for domain selection.

✓ Propose an estimator for $\mathcal{A}^*(c)$, and a nonparametric approach to assess the estimation uncertainty

$\Rightarrow$ Consider other model frameworks (non Gaussian), e.g.: *Wasserstein* divergence.

Summary and future work

✓ Tackle the problem of domain selection for electrocardiogram signals.

✓ Parameter of interest: interval $\Rightarrow$ Covariance structure + almost continuously recorded data.

✓ Use the KL divergence and GP to develop an easy to implement algorithm for domain selection.

✓ Propose an estimator for $\mathcal{A}^*(c)$, and a nonparametric approach to assess the estimation uncertainty

$\Rightarrow$ Consider other model frameworks (non Gaussian), e.g.: *Wasserstein* divergence.

Some references

Hernández, N., & Martos, G. (2023). Domain Selection for Gaussian Process Data: An application to electrocardiogram signals. arXiv preprint arXiv:2306.00538.

Hernández, N. (2019). Statistical learning methods for functional data with applications to prediction, classification and outlier detection.

Wasserman, L., & Roeder, K. (2009). High dimensional variable selection. Annals of statistics, 37(5A), 2178.

Pini, A, & Vantini, S. Interval-wise testing for functional data. Journal of Nonparametric Statistics, 29(2):407–424, 201

Berrendero, J., Cuevas, A., & Torrecilla, J. Variable selection in functional data classification: a maxima-hunting proposal. Statistica Sinica, pages 619–638, 20

Nørgaard, L. et al. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. Applied Spectroscopy 2000 54:3, 413-419

Thank you for listening! Comments are very welcome.

(flash for arXiv pre-print)