Bayesian Statistics course

February 15^{th} 2024

A.Y. 2023/2024

BAYESIAN MODELS FOR FUNCTIONAL DATA WITH WEARABLE APPLICATIONS

University of Nottingham
UK | CHINA | MALAYSIA

Riccardo Corradin
Yordan Raykov

POLITECNICO MILANO 1863

UNIVERSITÀ DEGLI STUDI DI MILANO

Gabriele Bruni
Stefania Colombo
Silvia D'Amicantonio
Chiara Facco
Luca Meanti
Michele Sparviero

PRESENTATION OVERVIEW

INTRODUCTION

Project Objective

Examine different methods for ***Scalar-on-Function Regression*** techniques.

Contribute to Parkinson's Research, advancing collective knowledge and potentially inspiring innovative approaches.

DATASET

Dataset

Our analysis employs the ***Personalized Parkinson Project***, a study conducted on ***650 patients*** affected by Parkinson's Disease and carried out at the Radboud University Medical Center in Nijmegen, ***Netherlands***.

DATASET

Dataset

Our analysis employs the ***Personalized Parkinson Project***, a study conducted on ***650 patients*** affected by Parkinson's Disease and carried out at the Radboud University Medical Center in Nijmegen, ***Netherlands***.

Specifically

During a 2-year follow-up period, starting from October 1^{st} 2017 to December 31^{st} 2023, partecipants undergo ***three extensive annual assessments*** and continuously wear a ***multi-sensor investigational research device***.

DATASET

We work with a piece of data with **160 patients** associated to variables like:

- Impact of the ***tremor*** on daily activities;
- Level of ***activities of daily living***;
- ***Year*** of onset of the ***first symptoms***;
- ***Age*** and ***gender*** of the patients;
- ***Medication*** group;
- Z-scored retrained ***classifier*** for the ***tremor***;
- ***WearingOff***, i.e the ***decreasing effectiveness*** of treatments.

PREPROCESSING

We preprocess our dataset to handle its **complexity**:

PREPROCESSING

We preprocess our dataset to handle its **complexity**:

1. **Smoothing stage**:
   - ***Kalman Filter***;
   - ***Savitzky–Golay (Savgol) Filter***.

PREPROCESSING

We preprocess our dataset to handle its **complexity**:

1. **Smoothing stage**:
   - ***Kalman Filter***;
   - ***Savitzky–Golay (Savgol) Filter***.

2. **Registration stage**:
   Extraction of the ***Karcher mean***, i.e. the "average curve" to which other curves are aligned at patient level.

THE MODEL

Scalar-on-Function Regression Model

$$Y_i = \mu + \int_T x_i(t)\,\beta(t)\,dx + \varepsilon_i\,, \quad i = 1, \ldots, n, \quad t \in T$$

where:

- Y_i is the scalar response;
- $\mu \in \mathbb{R}$;
- $T \subset \mathbb{R}^d$ is a product of compact intervals of $\mathbb{R}$;
- $x_i(t) \in L^1(T)$ is the functional predictor;
- β is a separable functional coefficient on T;
- $\varepsilon_1, \ldots, \varepsilon_n$ are independent random variables with a $\mathcal{N}(0, \sigma^2)$ distribution.

STARTING POINT

First step: develop and test the model on **synthetic data**, generated as follows:

- K Gaussian kernel;
- $\boldsymbol{\mu} = \mathbf{0}$;
- $\boldsymbol{x_1}, \dots, \boldsymbol{x_n} \sim GP(\boldsymbol{\mu}, K)$;
- σ^2 is known;
- $\beta(t) \approx \beta_{true}(t) = \sin\left(\frac{\pi}{4} t\right), \;\; t \in T$.

UNCONSTRAINED MODEL

Unconstrained Model

$$
\begin{aligned}
Y_i \mid \beta &\overset{ind}{\sim} \mathcal{N}\left(\mu + \int_T x_i(t)\,\beta(t)\,dx,\ \sigma^2\right) \\
\beta &\sim \mathrm{GP}\,(0, K)
\end{aligned}
$$

where:

- $\sigma^2 \in \mathbb{R}^+$;
- K is the continuous covariance function with a Gaussian kernel.

UNCONSTRAINED MODEL

Given the ***Likelihood*** and the ***Prior*** described above, the **Posterior** and the **Marginal distribution of Y** are:

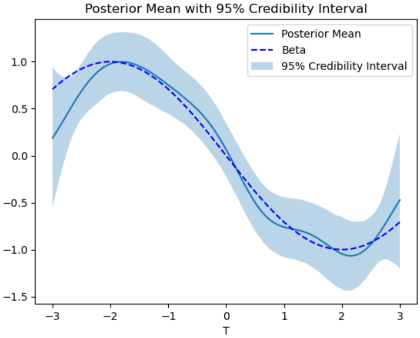
$$\beta \,|\, Y \;\sim\; \mathrm{GP}(m, K^*)$$
$$Y \;\sim\; \mathcal{N}_n(0, \Sigma + \sigma^2 \mathbf{I}_n)$$

where:

- $m(t) = Lx(t)' \left(\Sigma + \sigma^2 \mathbf{I}_n\right)^{-1} Y$;
- $K^*(s,t) = K(s,t) - Lx(s)' (\Sigma + \sigma^2 \mathbf{I}_n)^{-1} Lx(t)$;
- $\Sigma \in \mathbb{R}^{n \times n}$.

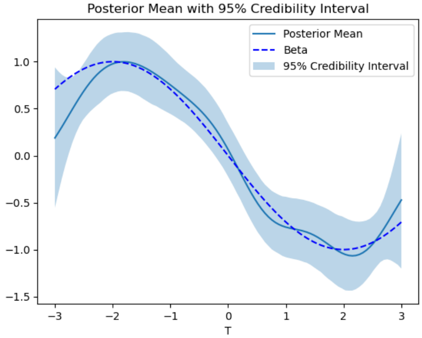
Additional details here

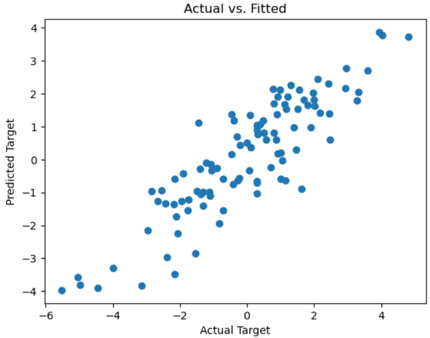
UNCONSTRAINED MODEL



Results obtained using ***synthetic*** data.

UNCONSTRAINED MODEL





Results obtained using ***synthetic*** data.

UNCONSTRAINED MODEL

Approaching the Constrained Model

The ***functional response*** assumes ***null value in certain time intervals***.

Let $T_0 \subset T$ be the union of such intervals, we constrain the ***functional coefficient*** β_t to be ***zero*** in T_0, namely:

$$\beta(t) = \beta_t \approx 0 \, \forall \, t \in T_0.$$

CONSTRAINED MODEL

RKHS in a nutshell

Let X be an arbitrary set and H a Hilbert space of real-valued functions on X. H is called a ***Reproducing Kernel Hilbert Space*** (RKHS) with kernel K if:

$$\forall f \in H \quad \forall x \in X \quad \langle f(\cdot)\,,\, K(x,\cdot)\rangle = f(x)$$

(reproducing property)

CONSTRAINED MODEL

- Denote H the RKHS with kernel K, define $H^0 := \{f \in H\colon f\,|_{T_0} = 0\}$ and let $\boldsymbol{P}$ be the ***orthogonal projection*** of H onto H^0.

CONSTRAINED MODEL

- Denote H the RKHS with kernel K, define $H^0 := \{f \in H\colon f\,|_{T_0} = 0\}$ and let $\boldsymbol{P}$ be the ***orthogonal projection*** of H onto H^0.
- H can be identified with an Hilbert space $\mathcal{H}$ of real-valued functions on T such that the ***map*** $\boldsymbol{\Theta}$ is a ***linear isometry***:

$$\begin{aligned} \Theta\colon \mathcal{H} &\to H \\ g(t) &\mapsto \Theta(g)(t) := \mathbb{E}[g\beta_t] \end{aligned}$$

CONSTRAINED MODEL

- Denote H the RKHS with kernel K, define $H^0 := \{f \in H \colon f\,|_{T_0} = 0\}$ and let $\boldsymbol{P}$ be the ***orthogonal projection*** of H onto H^0.
- H can be identified with an Hilbert space $\mathcal{H}$ of real-valued functions on T such that the ***map*** $\boldsymbol{\Theta}$ is a ***linear isometry***:

$$\begin{aligned} \Theta \colon \mathcal{H} &\to H \\ g(t) &\mapsto \Theta(g)(t) := \mathbb{E}[g\beta_t] \end{aligned}$$

- Hence $\boldsymbol{\beta_t^0}$ is the ***projection*** in $\mathcal{H}^{\mathbf{0}} = \Theta^{-1}(H^0)$ of $\beta_t \in \mathcal{H}$.
Namely,

$$\beta_t = \beta_t^0 + \beta_t^1 \quad a.s. \quad \text{with} \quad \beta_t^0 \in \mathcal{H}^0, \beta_t^1 \in \mathcal{H}^1 := (\mathcal{H}^0)^\perp$$

CONSTRAINED MODEL

Constrained Model

The model is:

$$
\begin{aligned}
Y_i \,|\, \beta^0,\ \mu,\ \sigma^2 &\overset{ind}{\sim} \mathcal{N}\left(\int_T \mu + x_i(t)\beta^0(t)\,dt, \sigma^2\right) \\
\beta^0 | \, \sigma^2, T_0 &\sim GP(0, \sigma^2 K^0) \\
p(\mu, \sigma^2, T_0) &\propto \frac{1}{\sigma^2} p(T_0)
\end{aligned}
$$

CONSTRAINED MODEL

Constrained Model

The model is:

$$Y_i \,|\, \beta^0,\, \mu,\, \sigma^2 \stackrel{ind}{\sim} \mathcal{N}\left(\int_T \mu + x_i(t)\beta^0(t)\,dt, \sigma^2\right)$$
$$\beta^0 |\, \sigma^2, T_0 \sim GP(0, \sigma^2 K^0)$$
$$p(\mu, \sigma^2, T_0) \propto \frac{1}{\sigma^2} p(T_0)$$

The **Prior for T_0** is:

$$p(T_0) := e^{-\alpha r(T_0)}, \quad \alpha \in \mathbb{R}^+$$

CONSTRAINED MODEL

Posterior of β and conditional of Y

$$\beta^0 \,|\, Y, \mu, \sigma^2, T_0 \sim \mathrm{GP}(m, K^*)$$
$$Y \,|\, \mu, \sigma^2, T_0 \sim \mathcal{N}_n\left(\mu \mathbf{1}_n,\, \sigma^2 M\right)$$

where:

- $m(t) = L^0 x(t)' \, M^{-1} \, (Y - \mu \mathbf{1}_n)$;
- $K^*(s,t) = \sigma^2 \left[K^0(s,t) - L^0 x(s)' \, M^{-1} L^0 x(t) \right]$;
- $M = \Sigma^0 + \mathbf{I}_n$;
- $Y = (Y_1, ..., Y_n)$.

Additional details here.

CONSTRAINED MODEL

Metropolis Hastings

- T is a ***discrete grid*** of $\boldsymbol{g}$ time instants.

CONSTRAINED MODEL

Metropolis Hastings

- T is a ***discrete grid*** of $\boldsymbol{g}$ time instants.
- T_0 is a ***vector of entries*** $\{0,1\}^g$ such that $T_0[i] = 1$ if $T[i] \in T_0$.

CONSTRAINED MODEL

Metropolis Hastings

- T is a ***discrete grid*** of $\boldsymbol{g}$ time instants.
- T_0 is a ***vector of entries*** $\{0,1\}^g$ such that $T_0[i] = 1$ if $T[i] \in T_0$.

➥ **Algorithm**:
- Set $\boldsymbol{g}$ and α and compute $\mathbf{k}$;
- Initialize the chain with $\mathbf{T_0^{(0)}}$ a random combination of $\boldsymbol{g}$ zeros and ones with exactly $\mathbf{k}$ runs;
- While $i < 5.000$: generate a candidate $\mathbf{T_0^{new}}$ for $T_0^{(i+1)}$, by randomly choosing if:
  - → ***removing/adding*** a 1 in a run's end;
  - → ***splitting*** a run of ones by setting to 0 a random internal point.

  Accept it with rate:

$$\text{rate} = \frac{p(Y|\, T_0^{new})p(T_0^{new})}{p(Y|\, T_0^{(i)})p(T_0^{(i)})}$$

CONSTRAINED MODEL

Metropolis Hastings - customized

- T is a ***discrete grid*** of **g** time instants;
- T_0 is a ***vector of entries*** $\{0,1\}^g$ such that $T_0[i] = 1$ if $T[i] \in T_0$.

➥ **Algorithm**:

- Set $\boldsymbol{g}$ and α and compute $\mathbf{k}$;
- Initialize the chain with $\mathbf{T_0^{(0)}} = (0, 1, 0, 1, ..., 0, 1)$;
- While $i < 5.000$: generate a candidate $\mathbf{T_0^{new}}$ by randomly choosing if:
  - → ***removing/adding*** a 1 in a run's end;
  - → ***splitting*** a run of ones by setting to 0 a random internal point.

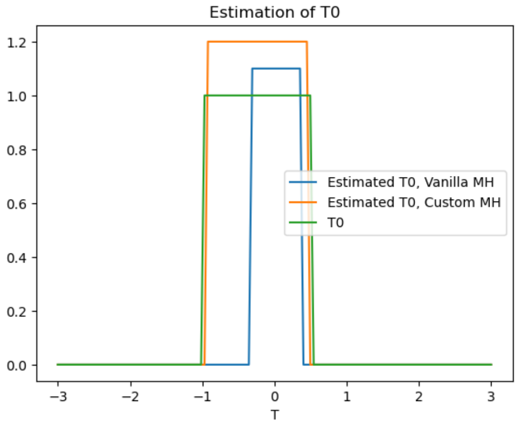
  Accept it if:

$$\text{rate} = \frac{p(Y|T_0^{new})p(T_0^{new})}{p(Y|T_0^{(i)})p(T_0^{(i)})} > \gamma^{(i)}, \quad \gamma^{(i)} = 1.0 - 0.02 \cdot \frac{(s-A)}{s}$$

where A is the number of acceptances in the last s steps.

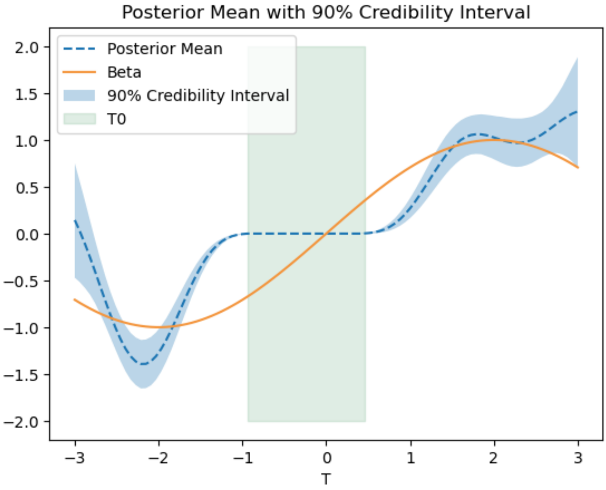
CONSTRAINED MODEL

Efficiency of the ***Customized MH*** (500 iterations; burn-in 100) versus the ***RWMH*** (2000 iterations; burn-in 1000).



Results obtained using ***synthetic*** data.

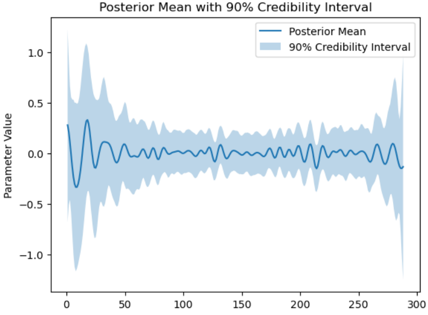
CONSTRAINED MODEL



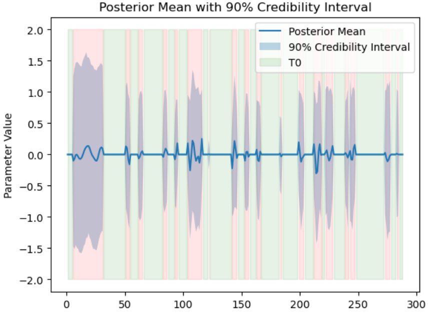
Results obtained using ***synthetic*** data.

APPLICATION ON REAL-WORLD DATA

Unconstrained Model



Constrained Model

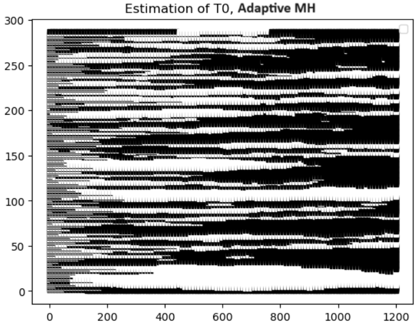


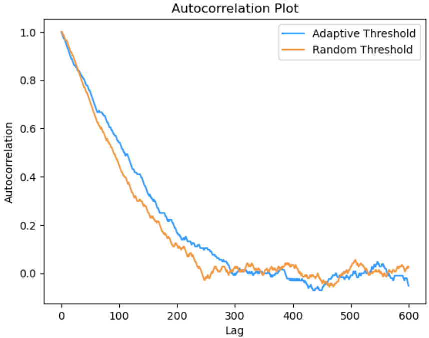
Results obtained using ***preprocessed real-world*** data.

- Y_i : Parkinson's severity for patient i;
- To classify patients, we use a **logit transformation**;
- $x_i(t)$: tremor probability (Z-score) for patient i;
- $i \in \{1, ..., 159\}$, $t \in T$.

APPLICATION ON REAL-WORLD DATA

Convergence diagnostics





Results obtained using ***preprocessed real-world*** data.

MIXED EFFECT MODEL

Mixed model with random effect on Gender

$$Y_{ij} = \int_T x_{ij}(t)\beta(t)\,dt + z_{ij}\theta_j + \theta_0 + \varepsilon_i, \quad i = 1, \ldots, 159,\ t \in T$$

$$\beta \mid \mu, \sigma^2, T_0 \sim GP(0, \sigma^2 K^0)$$

$$\theta_j \mid \mu_s, \sigma_s^2 \sim \mathcal{N}(\mu_s, \sigma_s^2)$$

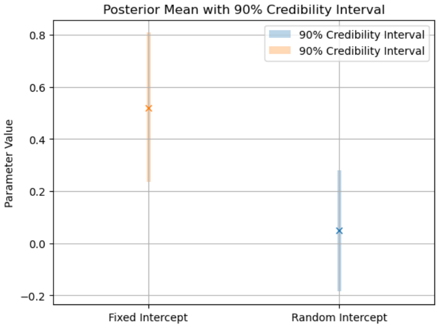
$$\theta_0 \mid \mu_0, \sigma_0^2 \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

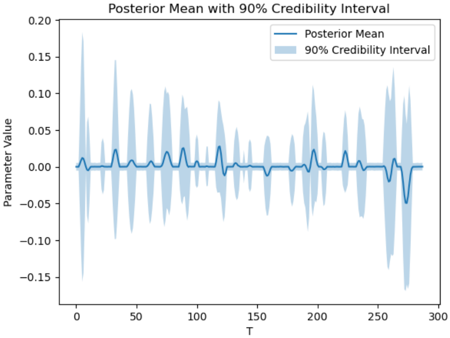
$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

where:

- $x_{ij}(t)$: tremor probability for the i^{th} patient, belonging to group j;
- Y_{ij} : severity of the i^{th} patient, belonging to group j;
- z_{ij}: grouping variable for groups $\{$men, women$\}$.

MIXED EFFECT MODEL





Results obtained using ***preprocessed real-world*** data.

MIXED EFFECT MODEL

Mixed model with random effect on Levodopa treatments

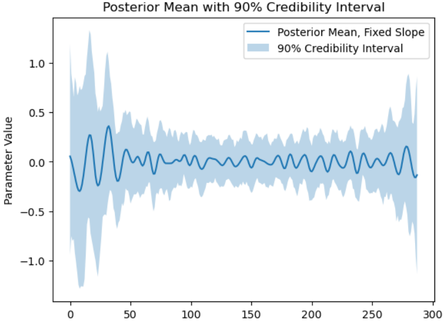
$$Y_{ij} = \int_T x_{ij}(t)\theta(t)\,dt + \int_T x_{ij}(t)\gamma_j(t)\,dt + \varepsilon_i, \qquad i = 1, \ldots, 136, t \in T$$

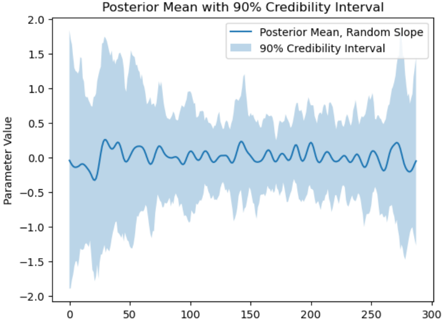
$$\begin{aligned} \theta \,|\, \sigma^2, T_0 &\sim GP(0, \sigma^2 K^0) \\ \gamma_j | \sigma_0^2 &\sim GP(0, \sigma_0^2 K^0), \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

where:

- $x_{ij}(t)$: tremor probability for the i^{th} patient, belonging to group j;
- Y_{ij}: severity of the i^{th} patient, belonging to group j;
- j represents the grouping induced by different numbers of doses received.

MIXED EFFECT MODEL





Results obtained using ***preprocessed real-world*** data.

CONCLUSIONS

Tables of **results**:

Models with Analytical Posterior	Accuracy	LPML
Unconstrained	62.3%	-206.61
Constrained	70.4%	-223.37
Constrained, Customized MH	70.0%	-218.18

Models with No Analytical Posterior	Accuracy	LPML
Unconstrained, Random Slope, 1 Group	62.5%	-188.82
Unconstrained, Random Slope, 2 Groups	65.4%	-204.68
Constrained, Random Intercept	67.3%	-167.08

CONCLUSIONS

Possible **extensions** of the models:

- ***Multiple*** Scalar-on-Function regression;
- Incorporation of ***more complex*** hierarchical structures;
- Use of ***Metropolis-within-Gibbs*** to update the estimate of parameters in Mixed Effect model;
- ***Further enhancement*** of estimation methods for T_0.

CONCLUSIONS

Overall, the **results** presented are ***satisfactory*** in terms of:

- Accuracy;
- Information obtained.

Drawbacks:
High computational costs impose many ***limits*** to our research.

ACKNOWLEDGMENTS and QUESTIONS

Thank you all for having attended this presentation.

We appreciate your attention and are available
for any ***questions*** you may have.

REFERENCES

B. R. Bloem, W. J. Marks Jr., A. L. Silva de Lima, M. L. Kuijf, T. van Laar, B. P. F. Jacobs, M. M. Verbeek, R. C. Helmich, B. P. van de Warrenburg, L. J. W. Evers, J. intHout, T. van de Zande, T. M. Snyder, R. Kapur and M. J. Meinders, ***The Personalized Perkinson Project: examining disease progression through broad biomarkers in early Parkinson's disease***, BCM Neurology, 2023.

Philip T. Reiss, Jeff Goldsmith, Han Lin Shang, R. Todd Ogden, ***Methods for Scalar-on Function Regression***, International Statistical Review, 2017.

Christophone Abraham, ***An informative prior distribution on functions with application to functional regression***, Statistica Neerlandica, 2023.

J.Q. Shi, B. Wang, ***Curve prediction and clustering with mixtures of Gaussian process functional regression models***, Springer Science, 2008.

DETAILS

- $m(t) = Lx(t)' \, (\Sigma + \sigma^2 \mathbf{I}_n)^{-1} \, Y;$
- $K^*(s,t) = K(s,t) - Lx(s)' \, (\Sigma + \sigma^2 \mathbf{I}_n)^{-1} \, Lx(t).$

where:

- $Lx_i(t) = \int_T K(t,s) \, x_i(s) \, ds \qquad i = 1, \ldots, n, \qquad t \in T \, ;$
- $Y = (Y_1, \ldots, Y_n)';$
- $\Sigma_{ij} = R(x_i, x_j) \qquad i, j = 1, \ldots, n \, ;$
- $R(x_i, x_j) = \int_T \int_T K(s,t) \, x_i(s) \, x_j(t) \, ds \, dt \qquad i, j = 1, \ldots, n \qquad t \in T \, .$

Main slide here

DETAILS

- $m(t) = L^0 x(t)' \, M^{-1} \, (Y - \mu \mathbf{1}_n)$;
- $K^*(s,t) = \sigma^2 \left[K^0(s,t) - L^0 x(s)' \, M^{-1} L^0 x(t) \right]$.

where:

- $L^0 x_i(t) = \int_T K^0(t,s) \, x_i(s) \, ds, \quad i = 1, \ldots, n, \quad t \in T$;
- $Y = (Y_1, \ldots, Y_n)'$;
- $M = \Sigma^0 + \mathbf{I}_n$;
- $\Sigma^0_{ij} = \int_T \int_T K^0(s,t) x_i(s) x_j(t) ds dt, \;\; i,j = 1, \ldots, n, \quad t \in T$.

Main slide here

DETAILS

The idea is to find β_t^0 as the projection in $\mathcal{H}^0$ of $\beta_t \in \mathcal{H}$. Thanks to the isometry Θ, the following commutative diagram is well-defined:

$$\begin{array}{ccc} \mathcal{H} & \xrightarrow{\Theta} & H \\ \downarrow{\scriptstyle \mathcal{P}} & & \downarrow{\scriptstyle P} \\ \mathcal{H}^0 & \xrightarrow{\Theta} & H^0 \end{array}$$

Then we can decompose the random process β_t:

$$\beta_t = \mathcal{P}\beta_t + \mathcal{Q}\,\beta_t = \beta_t^0 + \beta_t^1 \quad a.s.$$

Finally, for how we define H^0 and as a consequence of the reproducity property:

$$\forall f \in H^0, \forall \tau \in T^0 \quad f(\tau) = 0 \iff \langle f(\tau), K_\tau(s)\rangle = 0$$

allowing to deduce:

$$H^1 = \overline{span\{K_\tau : \tau \in T_0\}}^H.$$

DETAILS

Let $\mathcal{S}_k = \{T_0 \in T, r(T_0) = k\}$ and c_k its cardinality, we have:

$$\mathbb{P}(r(T_0) = k) = \sum_{T_0 \in \mathcal{S}_k} p(T_0) = \frac{c_k e^{-\alpha k}}{\sum_{i=1}^{g} c_i e^{-\alpha i}} \tag{1}$$

As initial value of $r(T_0)$ we use the expected number of runs, namely we fix α and g and take the k that maximize (1).