COMPUTO

ISSN 2824-7795

Should we correct the bias in Confidence Bands for Repeated Functional Data?

Emilie Devijver[1] CNRS, Univ. Grenoble Alpes, Grenoble INP, LIG, 38000 Grenoble, France
Adeline Samson Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Date published: 2024-05-27 Last modified: 2024-05-27

Abstract

While confidence intervals for finite quantities are well-established, constructing confidence bands for objects of infinite dimension, such as functions, poses challenges. In this paper, we explore the concept of parametric confidence bands for functional data with an orthonormal basis. Specifically, we revisit the method proposed by Sun and Loader, which yields confidence bands for the projection of the regression function in a fixed-dimensional space. This approach can introduce bias in the confidence bands when the dimension of the basis is misspecified. Leveraging this insight, we introduce a corrected, unbiased confidence band. Surprisingly, our corrected band tends to be wider than what a naive approach would suggest. To address this, we propose a model selection criterion that allows for data-driven estimation of the basis dimension, balancing the trade-off between bias and variance. Throughout the paper, we illustrate these strategies using an extensive simulation study. We conclude with an application to real data.

Keywords: functional data, repeated data, confidence band, Kac-Rice formulae, bias, dimension selection

Contents

[1]Corresponding author: emilie.devijver@univ-grenoble-alpes.fr

1 Introduction

Functional data analysis is widely used for handling complex data with smooth shapes, finding applications in diverse fields such as neuroscience (e.g., EEG data, REF), psychology (e.g., mouse-tracking data, Quinton et al. (2017)), and sensor data from daily-life activities (Jacques and Samardžić (2022)).

We consider multiple independent observations of the same function, yielding noisy functional data. To analyze such data, a common approach, typically in the parametric setting, involves projecting the data onto a functional space defined by a family of functions (Li, Qiu, and Xu (2022), Kokoszka and Reimherr (2017)). When the family serves as an orthonormal basis, e.g., Legendre (with the standard scalar product) or Fourier (with another scalar product), the projection is clearly understood, but widely used families such as splines are not orthonormal for the standard scalar product. Leveraging an approximate functional space offers a key advantage: it simplifies the inference problem to estimating coefficients, for example through methods like least squares or maximum likelihood estimation. Subsequently, the function is estimated as the mean of the functional data following projection onto the functional basis.

Measuring the uncertainty of an estimator is usually done using confidence intervals. In this paper, our focus lies specifically on providing a simultaneous confidence band for the function means, rather than point-wise confidence intervals. This task presents several challenges: the confidence band must effectively control the simultaneous functional type-I error rate, as opposed to point-wise rates; it must strike a balance between being sufficiently conservative to maintain a confidence level while not being overly so as to render it meaningless; and the method used to construct this confidence band should be computationally feasible for practical application.

Several developments have already been proposed to answer these questions. First, consider the case with only one individual (no repetition) but with many time points. Some methods study the asymptotic distribution of the infinity norm between the true function and its estimator. The asymptotic in the number of time points is studied in Hall (1991), Claeskens and Van Keilegom (2003). This approach works only for large datasets in time and is likely to be too conservative otherwise. For small samples, bootstrap methods have been developed to compute the confidence band (Neumann and Polzehl (1998), Claeskens and Van Keilegom (2003)), but with a high computational cost. Another

approach is to construct confidence bands based on the volume of the tube formula. Sun and Loader (1994) studied the tail probabilities of suprema of Gaussian random processes. This approach is based on an unbiased linear estimator of the regression function. Zhou, Shen, and Wolfe (1998) used the volume-of-tube formula for estimation by regression splines. Krivobokova, Kneib, and Claeskens (2010) used this method for the construction of confidence bands by penalized spline estimators. They proposed to mix Bayesian and frequentist approaches, to get the good properties from the Bayesian world but reducing the variability to be less conservative using the frequentist approach. The bias is considered through spline modeling, assuming sufficient knots are considered. Liebl and Reimherr (2019) have proposed a method based on random field theory and the volume-of-tube formula. They provide a band with locally varying widths using an unbiased estimator. Their method does not require the estimation of the full covariance function of the estimator, but only its diagonal. This reduces the computational time. From a practical viewpoint, Sachs, Brand, and Gabriel (2022) introduce a package to popularize simultaneous confidence bands, in the context of survival analysis.

Some papers, like ours, rely on several observations of the same function. Bunea, Ivanescu, and Wegkamp (2011) propose a threshold-type estimator and derive error bounds and simultaneous confidence bands, having an unbiased estimator. Telschow and Schwartzman (2022) propose a simultaneous confidence band based on the Gaussian kinematic formula. Again, it assumes access to an unbiased estimator of the function of interest. Note that recent extensions have been proposed, to nonstationary random field in Telschow et al. (2023) and based on conformal prediction in Diquigiovanni, Fontana, and Vantini (2022). These extensions are out of the scope of this paper, focusing on the simple functional case.

One limitation of all those approaches is that they do not generally take into account the bias of the functional estimator. Sun and Loader (1994) proposed a bias correction for a particular class of functions but left the smoothing parameter choice open, leading to an unusable estimator. In the nonparametric framework, the bias is approximated using the estimator of the second derivative of the underlying mean function (Xia (1998)). But in general, there is a lack of discussion on how to handle the bias of the functional estimator, even in the simple case of a functional space of finite dimension.

The objective of this paper is to address the bias problem in confidence band construction for a general function, utilizing a finite functional orthonormal family. Our contributions are as follows:

- we disentangle the bias issue by explicitly defining the parameter of interest within the approach of Sun and Loader (1994);
- we propose a bias correction method in a new confidence band for the function of interest;
- we illustrate this confidence band, concluding on the conservatism of the procedure;
- finally, we propose a method for selecting the dimension of the approximation space, treating it as a model selection problem, with a trade-off between conservatism and confidence level assurance.

Note that while the model selection paradigm has been extensively studied in the literature, in multivariate statistics or functional data analysis (e.g., Goepp, Bouaziz, and Nuel (Submitted), Aneiros, Novo, and Vieu (2022), Basna, Nassar, and Podgórski (2022)), it has not been explored in the context of confidence band construction.

The paper is organized as follows: Section 2 introduces the functional regression model, the considered functional family and the corresponding approximate regression models, as well as an estimator defined in the finite space, along with descriptions of the error terms. In Section 3, we propose a confidence band for the approximate regression function in the space of finite dimension, where the dimension is fixed. Section 4 proposes a strategy to construct a confidence band for the true function. This last confidence band being too conservative, Section 5 introduces a model selection criterion

to select the best confidence band, doing a trade-off between conservatism and confidence level assurance. Section 7 ends the paper by a conclusion and discussion of perspectives. The different estimation procedures are illustrated throughout the sections.

2 Statistical Model

In this paper, we consider time series as discrete measurements of functional curves. We first present the general functional regression model (Section 2.1) where the regression function belongs to a finite functional family of dimension L^*. In practice, this dimension L^* is unknown and we will work on functional space of dimension L. The regression model on the finite family of functions is presented in Section 2.2, and an estimator is proposed in Section 2.3, with a description of the error terms.

2.1 Functional regression model

Let y_{ij} be the measure at fixed time $t_j \in [a,b]$ for individual $i = 1,\dots,N$, with $j = 1,\dots,n$. We restrict ourselves to $[a,b] = [0,1]$, without loss of generality. We assume these observations are discrete-time measurements of individual curves, which are independent and noisy realisations of a common function f that belongs to a functional space. Thus for each individual i, we consider the following functional regression model

$$y_{ij} = f(t_j) + \varepsilon_{ij},$$

where $\varepsilon_{i.} = (\varepsilon_{i1},\dots,\varepsilon_{in})$ is the noise representing the individual functional variation around f. We assume that the ε_i are independent. Their distribution is detailed below.

For each individual $i = 1,\dots,N$, we denote $y_{i.} = (y_{i1},\dots,y_{in})$ the $n \times 1$ vector of observations, $t_. = (t_1,\dots,t_n)$ the $n \times 1$ vector of observation times and $f(t_.) = (f(t_1),\dots,f(t_n))$ the $n \times 1$ vector of the function f evaluated in $t_.$. We also denote $\mathbf{y} = (y_{1.},\dots,y_{N.})$ the whole matrix of observations.

Let us introduce the functional space $\mathcal{S}^{L^*} = Vect((t \mapsto B_\ell^{L^*}(t))_{1\le\ell\le L^*})$ with L^* functions $(B_\ell^{L^*})_{1\le\ell\le L^*}$ assumed to be linearly independent. Then, for any $f \in \mathcal{S}^{L^*}$, there exists a unique vector of coefficients $(\mu_\ell^{L^*})_{1\le\ell\le L^*}$ such that, for all t, $f(t) = \sum_{\ell=1}^{L^*} \mu_\ell^{L^*} B_\ell^{L^*}(t)$. The regression function f verifies the following assumption:

Definition 2.1. The function f belongs to the space $\mathcal{S}^{L^*}$ of dimension L^*. It is denoted f^{L^*} and defined as:

$$f(t) = f^{L^*}(t) = \sum_{\ell=1}^{L^*} \mu_\ell^{L^*} B_\ell^{L^*}(t).$$

Many functional spaces are available in the literature, as Splines, Fourier or Legendre families. Let us consider the space $L^2([0,1])$ with its standard scalar product $< f_1, f_2 >= \int_0^1 |f_1(t) f_2(t)| dt$, for $f_1, f_2 \in L^2([0,1])$. We introduce the following assumption:

Definition 2.2. The functional family $(t \mapsto B_\ell^{L^*}(t))_{1\le\ell\le L^*}$ is orthonormal with respect to the standard scalar product $< .,. >$.

Note that if Definition 2.2 holds, one get $\mu_\ell^{L^*} =< f^{L^*}, B_\ell^{L^*} >$ for $\ell = 1,\dots,L^*$. The Legendre family is orthonormal, the Fourier family is orthogonal for the standard scalar product (but not orthonormal), and the B-splines family is not orthogonal.

We also consider a functional noise through the following assumption.

Definition 2.3. The sequence ε_i is functional and belongs to the functional space $\mathcal{S}^{L^\varepsilon}$ of dimension L^ε. Then, there exists a sequence of coefficients $(c_{i\ell})_{1\leq\ell\leq L^\varepsilon}$ such that

$$\varepsilon_{ij} = \sum_{\ell=1}^{L^\varepsilon} c_{i\ell} B_\ell^{L^\varepsilon}(t_j).$$

We also assume that the coefficients are Gaussian: for all $i = 1, \dots, N$ and $\ell = 1, \dots, L^\varepsilon$,

$$c_{i\ell} \sim_{iid} \mathcal{N}(0, \sigma^2).$$

Definition 2.1 and Definition 2.3 imply that each curve y_i belongs to a finite family: for $j = 1, \dots, n$,

$$y_{ij} = \sum_{\ell=1}^{L^*} \mu_\ell^{L^*} B_\ell^{L^*}(t_j) + \sum_{\ell=1}^{L^\varepsilon} c_{i\ell} B_\ell^{L^\varepsilon}(t_j).$$

As the observations are recorded at discrete time points $(t_j)_{1\leq j\leq n}$, we introduce the family of functions evaluated at the discrete times of observations. For $L \in \mathbb{N}$, let us denote $\mathbf{B}^L$ the matrix of $n \times L$ with coefficient in row j and column ℓ equal to $B_\ell^L(t_j)$.

Let us introduce $c_{i.} = (c_{i1}, \dots, c_{iL^\varepsilon})$ the $L^\varepsilon \times 1$ vector. Then $\varepsilon_{i.} = \mathbf{B}^{L^\varepsilon} c_{i.}$. The vectors $y_{i.} \in \mathbb{R}^n$ are thus independent and $y_i \sim \mathcal{N}_n(f(t_.), \sigma^2 \Sigma^{L^\varepsilon})$ with $\Sigma^{L^\varepsilon} = \mathbf{B}^{L^\varepsilon}(\mathbf{B}^{L^\varepsilon})^T$.

To illustrate the model, we simulate a regression functional model with $n = 50$ points per individual and $N = 40$ individuals. In Figure 1, the function f (red curve) belongs to the Fourier (resp. Legendre and Spline) family with $L^* = 10$ and the noisy observations y_{ij} (black curves) have a functional noise in dimension $L^\varepsilon = 20$, also in the Fourier (resp. Legendre and Spline) family on the left plot (resp. middle and right).

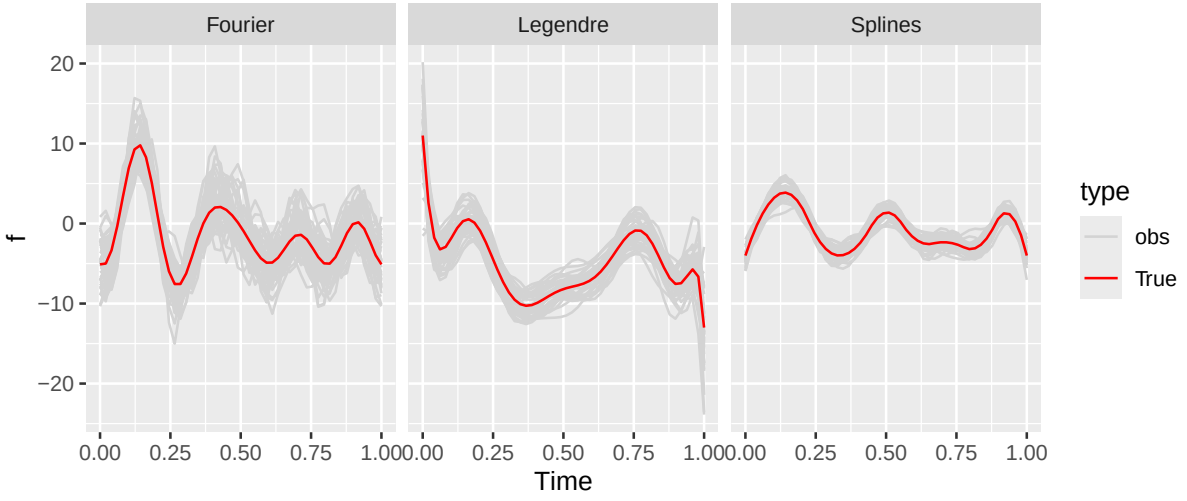

Figure 1: Illustrative example. We generate a regression functional model in the Fourier (left), Legendre (middle) and Splines (right) families. The red curve corresponds to the true function, and the gray curves correspond to noisy observations.

The objective of this paper is to construct a tight confidence bound for f^{L^*} using data $(y_{ij})_{ij}$. The main challenge is that the true dimension L^* is unknown. In the rest of the paper, we will work with a collection of models defined on a finite family of dimension L with $L \in \{L_{\min}, \dots, L_{\max}\}$, $L_{\max}$ being chosen to be sufficiently large by the user, expecting that $L^* \leq L_{\max}$. Then we will propose different strategies to choose the best bandwidths among the different collections.

First, in Section 2.2 and Section 2.3, we define for a fixed L the corresponding regression model and its estimator. Then Section 3, Section 4 and Section 5 will introduce the different bandwidths.

2.2 Approximation of the model on a finite family

Let $f^{L^*} \in \mathcal{S}^{L^*}$ with L^* unknown, and consider the space $\mathcal{S}^L$ for $L \in \{1, \dots, L_{\max}\}$ fixed. As $\mathcal{S}^L$ is a family of linearly independent functions, there always exists a unique vector μ^{L,L^*} of coefficients defining $f^{L,L^*}(t) = \sum_{\ell=1}^{L} \mu_\ell^{L,L^*} B_\ell^L(t) = B^L(t)\mu^{L,L^*}$ such that

$$f^{L,L^*} = \arg\min_{f \in \mathcal{S}^L} \{\|f^{L^*} - f\|_2^2\},$$

and if the family is orthonormal (Definition 2.2), it corresponds to the projected coefficients μ_ℓ^{L,L^*}:

$$\mu_\ell^{L,L^*} := < f^{L^*}, B_\ell^L > .$$

We can prove the following property:

Proposition 2.1. *Under Definition 2.1,*

$$f^{L^*,L^*} = f^{L^*}.$$

Moreover, if Definition 2.2 also holds, the projection coefficients verify

$$\mu_\ell^{L,L^*} = \mu_\ell^{L^*} \quad for\ \ell = 1, \dots, \min(L, L^*).$$

In practice, data are observed at discrete time, we consider the operator $\mathbf{P}^L$ defined as the matrix $\mathbf{P}^L = ((\mathbf{B}^L)^T\mathbf{B}^L)^{-1}(\mathbf{B}^L)^T$ of size $L \times n$ (this operator is a bit more complex when the functional family is not orthonormal wrt the standard scalar product). Then we define the coefficients $\underline{\mu}^{L,L^*}$ which are the coefficients of μ^{L,L^*} approximated on the vector space, denoted $\mathbf{S}^L$, defined by the matrix $\mathbf{B}^L$.

$$\underline{\mu}^{L,L^*} := \mathbf{P}^L\mathbf{B}^{L^*}\mu^{L^*}.$$

The corresponding finite approximated regression function is denoted $\underline{f}^{L,L^*}$ and is defined, for all $t \in [0, 1]$, as

$$\underline{f}^{L,L^*}(t) = B^L(t)\underline{\mu}^{L,L^*}.$$

We can prove the following properties linking L, L^* and the number of timepoints n:

Proposition 2.2. *Under Definition 2.1 and Definition 2.2,*

- *When $L \geq L^*$, $\mathbf{P}^L\mathbf{B}^{L^*}$ has L^* diagonal elements equal to 1 and other non-diagonal elements close to 0. The first L^* elements of $\underline{\mu}^{L,L^*}$ are equal to μ^{L^*} when $n > L$.*
- *When $L < L^*$, $\mathbf{P}^L\mathbf{B}^{L^*}$ has L diagonal elements equal to 1. The first L elements of $\underline{\mu}^{L,L^*}$ are different to $\mu_\ell^{L^*}$. When $n \to \infty$, $\underline{\mu}_\ell^{L,L^*} \to \mu_\ell^{L^*}$ for $\ell = 1, \dots, \min(L, L^*)$.*
- *If $n > L^*$, then $f^{L^*} = f^{L^*,L^*} = \underline{f}^{L^*,L^*}$.*

These properties are illustrated in Figure 2. The true dimension is $L^* = 11$. Three families are considered, Fourier, Legendre and Splines. The plots display the absolute difference between the coefficients $\mu_\ell^{L^*}$ and the projected coefficients $\underline{\mu}^{L,L^*}$, for different ℓ in x-axis and for different values of L and n of the y-axis, namely a case with $L < L^*$ and two values of n: $L = 7, n = 20$ and $L = 7, n = 100$; and a case with $L > L^*$ and two values of n: $L = 15, n = 20$ and $L = 15, n = 100$. The absolute difference is represented as a gradient of color, this gradient being adapted for each functional family. We can see that as Legendre (resp. Fourier) are orthonormal (resp. orthogonal) families, the differences are close to 0 when $L = 15$, whatever the values of n. When $L < L^*$, the difference is close to 0 when n is large. This property does not hold for the spline family, which is not orthogonal.

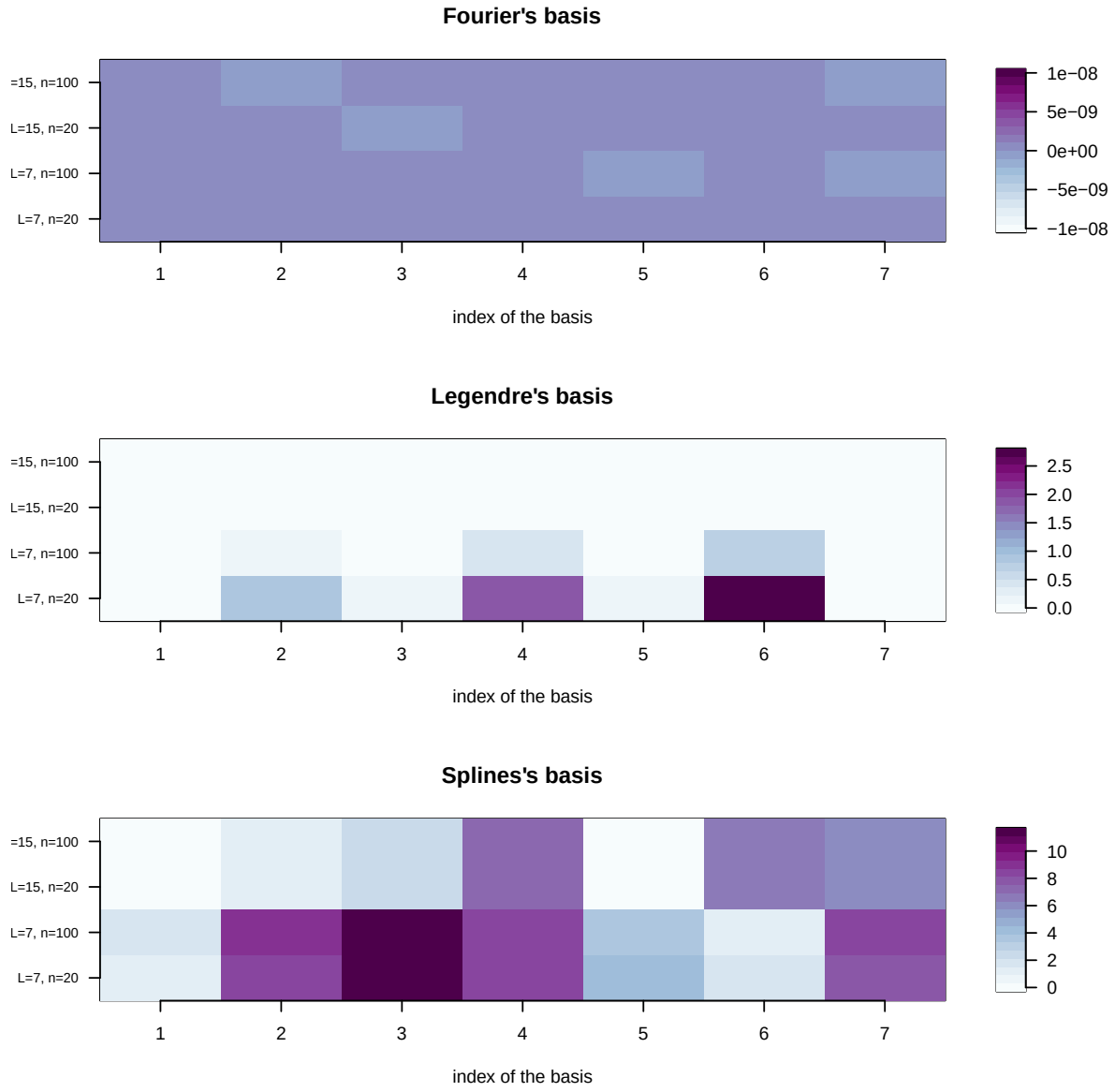

Figure 2: Illustrative example. The true dimension is 11, we generate the coefficients with three families, Fourier (which is orthogonal), Legendre (which is orthonormal) and the splines (which are not orthogonal wrt the standard scalar product). In the y-axis, two dimensions of the family (7 or 15) and two numbers of timepoints (20 or 100) are compared. We plot in x-axis the value of the absolute difference between the true coefficients and their approximations for the first 7 coefficients of the basis. The color scale is adapted to each functional basis.

2.3 Estimator

Let $L \in \{L_{\min}, \dots, L_{\max}\}$. This section presents the least square estimator of the regression function on the space of dimension L defined by the family $\mathbf{B}^L$ and discusses its error.

2.3.1 Estimation of the regression function

When considering the estimation of the regression function f^{L^*} on the space of dimension L defined by the family $\mathbf{B}^L$, we do not directly estimate f^{L^*} but its projection on this finite space, which corresponds to the projected function $\underline{f}^{L,L^*}(t)$ and its associated coefficients $(\underline{\mu}_\ell^{L,L^*})_{1\leq\ell\leq L}$.

Definition 2.4. The vector of coefficients $(\underline{\mu}_\ell^{L,L^*})_{1\leq\ell\leq L}$ is estimated by the least square estimator $\hat{\underline{\mu}}^{L,L^*}$ defined as:

$$\hat{\underline{\mu}}^{L,L^*} := \frac{1}{N}\sum_{i=1}^{N} \mathbf{P}^L y_{i.}$$

For a fixed $t \in [0,1]$, the estimator of the function $\underline{f}^{L,L^*}(t)$ is defined by:

$$\hat{\underline{f}}^{L,L^*}(t) = \sum_{\ell=1}^{L} \hat{\underline{\mu}}_\ell^{L,L^*} B_\ell^L(t) = B^L(t)\hat{\underline{\mu}}^{L,L^*}. \tag{1}$$

Equation 1 directly implies that the estimator is thus the empirical mean of the functional approximation of each individual vector of observations. Because we work with least squares estimators, we can easily study the error of estimation of $\hat{\underline{\mu}}^{L,L^*}$ and $\hat{\underline{f}}^{L,L^*}$.

Proposition 2.3. *Under Definition 2.1 and Definition 2.3, we have*

$$\hat{\underline{\mu}}^{L,L^*} \sim \mathcal{N}_L\left(\underline{\mu}^{L,L^*}, \frac{\sigma^2}{N}\Sigma_B^{L,L^\varepsilon}\right),$$

where the $L \times L$ covariance matrix $\Sigma_B^{L,L^\varepsilon}$ is defined as $\Sigma_B^{L,L^\varepsilon} := \mathbf{P}^L \Sigma^{L^\varepsilon} (\mathbf{P}^L)^T$ with $\Sigma^{L^\varepsilon} = \mathbf{B}^{L^\varepsilon}(\mathbf{B}^{L^\varepsilon})^T$.

Moreover, $B^L()\mathbf{P}^L y_i$ is a Gaussian process with mean $\underline{f}^{L,L^*}()$ and covariance function $(s,t) \mapsto \sigma^2 B^L(s)\Sigma_B^{L,L^\varepsilon}(B^L(t))^T$, and $(\hat{\underline{f}}^{L,L^*} - \underline{f}^{L,L^*})()$ is a centered Gaussian process with covariance function $C^{L,L^*} : (s,t) \mapsto \frac{\sigma^2}{N} B^L(s)\Sigma_B^{L,L^\varepsilon} B^L(t)^T$.

The proof is given in Appendix.

Figure 3 displays estimators calculated with different dimensions L. Data are generated with $L^* = 11$, $L^\varepsilon = 20$, $n = 50$ and $N = 40$. The true function and its projection $\underline{f}^{L,L^*}$ are in cyan, and the estimator $\hat{\underline{f}}^{L,L^*}$ is in red. We compute it for the three families, Legendre, Fourier and splines. In all cases, the estimators are very precise when considering the relevant function, but estimating a function of dimension L^* with a function of dimension $L < L^*$ is not consistent. Note that the performance of the estimator for the spline family is also good, even if the family is not orthonormal, because we work here at the level of the function (and not at the level of the coefficients).

Even if the estimator $\hat{\underline{f}}^{L,L^*}$ is defined on the functional space associated to $\mathbf{S}^L$, it can also be seen as an estimator of the function f^{L^*} which lies in the space $\mathcal{S}^{L^*}$. In that case, the error includes a functional approximation term due to the approximation of f^{L^*} on the space $\mathcal{S}^L$, which will be nonzero if

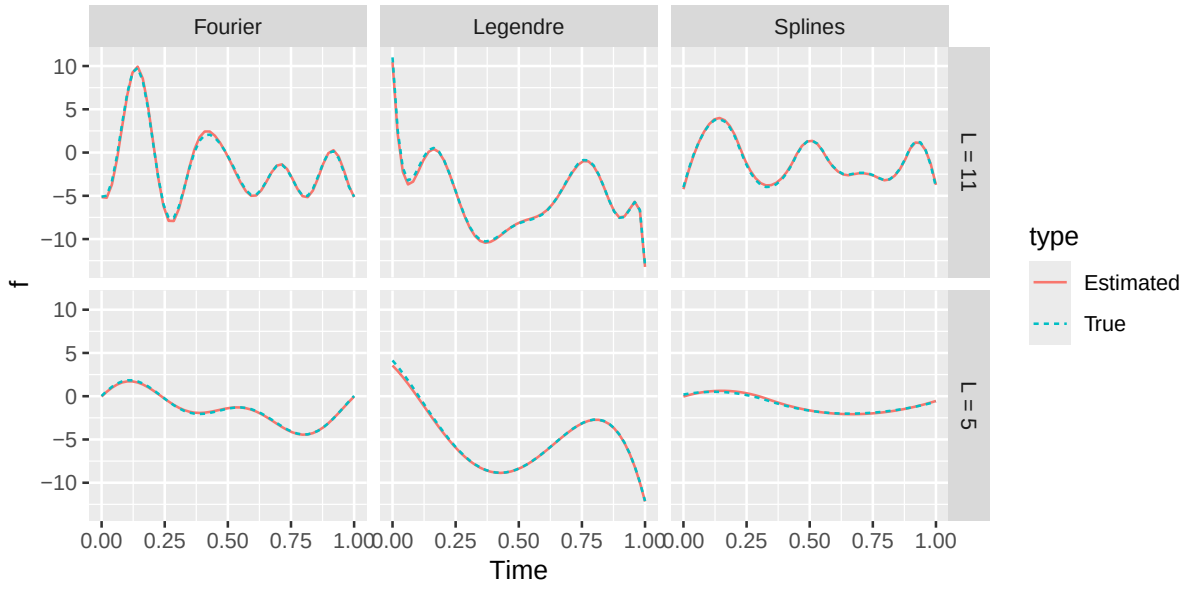

Figure 3: Illustrative example. For each family (Fourier which is orthogonal, Legendre which is orthonormal and the splines which are not orthogonal wrt the standard scalar product), we consider a function with true dimension 11 (top), and its projection on the space of dimension 5 (bottom), displayed in cyan. The estimators in dimensions 11 and 5 are displayed in red.

$L \neq L^*$. It corresponding to the bias of the estimator $\underline{\hat{f}}^{L,L^*}$, i.e. the difference between its expectation and the true f^{L^*}. Indeed, recalling that $f^{L^*} = \underline{f}^{L^*,L^*}$, the error of estimation can be decomposed into

$$\underline{\hat{f}}^{L,L^*}(t) - f^{L^*}(t) = \underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) + \underline{f}^{L,L^*}(t) - \underline{f}^{L^*,L^*}(t) =: Stat_{L,L^*}(t) + Bias_{L,L^*}(t), \tag{2}$$

The first term $Stat_{L,L^*}(t) = \underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)$ is the (unrescaled) statistics of the model. The second term $Bias_{L,L^*}(t) = \mathbb{E}(\underline{\hat{f}}^{L,L^*}(t)) - \underline{f}^{L^*,L^*}(t)$ is the bias of the estimator $\underline{\hat{f}}^{L,L^*}(t)$ when estimating the true function $\underline{f}^{L^*,L^*}(t)$.

Let us remark that this bias is different than the bias of the estimator $\underline{\hat{f}}^{L,L^*}(t)$ when estimating the projected function $\underline{f}^{L,L^*} = f^{L^*}$, which is 0.

The two terms defined in Equation 2 are more detailed in the two next subsections.

2.3.2 Statistics

The statistics of the model, $t \mapsto Stat_{L,L^*}(t) = \underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)$, is a random functional quantity which depends on the estimator $\underline{\hat{f}}^{L,L^*}$. From Proposition 2.3, for any $t \in [0,1]$, we define the centered and rescale statistics $Z_L(t)$ such that:

$$Z_L(t) := \frac{Stat_{L,L^*}(t)}{\sqrt{\mathrm{Var}(Stat_{L,L^*}(t))}} = \frac{\underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)}{\sqrt{C^{L,L^*}(t,t)}} \sim \mathcal{N}(0,1).$$

The covariance function can be estimated using the observations $y_{i.}$ as

$$\hat{C}^{L,L^*}(s,t) = \frac{1}{N-1}\sum_{i=1}^{N}(B^L(s)\mathbf{P}^L y_{i.} - \underline{\hat{f}}^{L,L^*}(s))(B^L(t)\mathbf{P}^L y_{i.} - \underline{\hat{f}}^{L,L^*}(t)).$$

2.3.3 Bias

The bias is due to the fact that the estimation is potentially performed in a different (finite) space than the space where the true function $\underline{f}^{L^*,L^*}$ lives. This is a functional bias, which is not random. It corresponds to the approximation (orthogonal projection if Definition 2.2 holds) of f^{L^*} from $\mathcal{S}^{L^*}$ to the space $\mathcal{S}^L$. It can be written as follows:

$$Bias_{L,L^*}(t) = B^L(t)\underline{\mu}^{L,L^*} - B^{L^*}(t)\mu^{L^*}.$$

Thus, we can deduce that when $L < L^*$ and if the family is orthonormal (Definition 2.2 holds),

$$Bias_{L,L^*}(t) = \sum_{\ell=1}^{L} B_\ell^L(t)\underline{\mu}_\ell^{L,L^*} - \sum_{\ell=1}^{L^*} B_\ell^{L^*}(t)\underline{\mu}_\ell^{L^*} = \sum_{\ell=L+1}^{L^*} B_\ell^{L^*}(t)\underline{\mu}_\ell^{L^*}.$$

From Proposition 2.3, we can directly deduce the following proposition:

Proposition 2.4. *Under Definition 2.1 and Definition 2.3, the mean is, for all* $t \in [0,1]$,

- *for* $L < L^*$, $Bias_{L,L^*}(t) \neq 0$,
- *for* $L \geq L^*$, $Bias_{L,L^*}(t) = 0$.

In the next section, we explain how to use this property to derive confidence bands of $\underline{f}^{L,L^*}$ and f^{L,L^*}.

3 Confidence Bands of $\underline{f}^{L,L^*}$ and f^{L,L^*} for a fixed L

The objective is to construct a confidence band for the two functions $\underline{f}^{L,L^*}$ and f^{L,L^*}, based on the observations $\mathbf{y}$, for a given value $L \in \{L_{\min}, \dots, L_{\max}\}$. The band for $\underline{f}^{L,L^*}$ enters the framework proposed by Sun and Loader (1994) which relies on an unbiased and linear estimator of the function. This is the case for the estimator $\hat{\underline{f}}^{L,L^*}$ which is an unbiased estimator of $\underline{f}^{L,L^*}$. We recall in Section 3.1 the construction of this confidence band which attains a given confidence level in a non-asymptotic setting, that is for a finite number of observations n for each individual. Then in Section 3.2, we prove that the confidence band proposed by Sun and Loader (1994) can be viewed as a confidence band for f^{L,L^*} with an asymptotic confidence level, the asymptotic framework being considered when $n \to \infty$.

3.1 Confidence band for $\underline{f}^{L,L^*}$

Let $L \in \{1, \dots, L_{\max}\}$. Consider $1-\alpha$ as a fixed confidence level. Our aim is to find a function $d^L()$ such that

$$\mathbb{P}\left(\forall t \in [0,1],\ \hat{\underline{f}}^{L,L^*}(t) - d^L(t) \leq \underline{f}^{L,L^*}(t) \leq \hat{\underline{f}}^{L,L^*}(t) + d^L(t)\right) = 1-\alpha.$$

Consider the normalized statistics $Z_L(t)$ which is a centered and reduced Gaussian process. We want to find the quantile q^L satisfying

$$q^L = \arg\min_q \left\{\mathbb{P}\left(\max_{t\in[0,1]} |Z_L(t)| \leq q\right) = 1-\alpha\right\}. \tag{3}$$

Then we can take $d^L(t) = q^L\sqrt{C^{L,L^*}(t,t)}$. The covariance function $C^{L,L^*}(t,t)$ can be replaced by its estimator $\hat{C}^{L,L^*}(t,t)$, making the distribution a Student's distribution with $N-1$ degrees of freedom. Thus, it only requires to be able to compute the critical value q^L.

This can be done following Sun and Loader (1994) who propose a confidence band for a centered Gaussian process. Their procedure is based on an unbiased linear estimator of the function of interest, which is the case for $\underline{\hat{f}}^{L,L^*}$ when we consider a band for $\underline{f}^{L,L^*}$. We recall their result in the following proposition, the computation of the value q^L is detailed thereafter.

Theorem 3.1 (Sun and Loader (1994)). *Set Definition 2.1 and Definition 2.3 and a probability $\alpha \in [0,1]$. Then, we have*

$$\mathbb{P}\left(\forall t \in [0,1], \left|\underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)\right| \leq \hat{d}^L(t)\right) = 1 - \alpha$$

with

$$\hat{d}^L(t) = \hat{q}^L \sqrt{\hat{C}^{L,L^*}(t,t)/N};$$

and $\hat{q}^L$ defined as the solution of the following equation, seen as a function of q^L:

$$\alpha = \mathbb{P}(|t_{N-1}| > q^L) + \frac{\|\tau^L\|_1}{\pi}\left(1 + \frac{(q^L)^2}{N-1}\right)^{-(N-1)/2}, \tag{4}$$

with $(\tau^L)^2(t) = \partial_{12}c(t,t) = Var(Z_L(t))'$ where we denote $\partial_{12}c(t,t)$ the partial derivatives of a function $c(t,s)$ in the first and second coordinates and then evaluated at $t = s$.

We can thus deduce a confidence band of level $1-\alpha$ for $\underline{f}^{L,L^*}$:

$$CB_1(\underline{f}^{L,L^*}) = \{\forall t \in [0,1], [\underline{\hat{f}}^{L,L^*}(t) - \hat{d}^L(t); \underline{\hat{f}}^{L,L^*}(t) + \hat{d}^L(t)]\}.$$

Dans la derniere section pour definir le critere, on a besoin de montrer que la variance de $\hat{\tau}$ est bornée par L/N. Le processus $(P^L y_i(t) - \underline{\hat{f}}^{L,L^*}(t))$ vit dans un espace de dimension L. Ca parait naturel de borner par L/N?? Mais comment le justifier ?

The value $\hat{q}^L$ is defined implicitly in Equation 4 which involves the quantity $t \mapsto \tau^L(t)$. Liebl and Reimherr (2019) propose to estimate $\tau^L(t)$, for all t, by

$$\begin{aligned}\hat{\tau}^L(t) &= \left(\widehat{Var}((U^L)'_1(t), \dots, (U^L)'_N(t)\right)^{1/2} \\ &= \left(\frac{1}{N-1}\sum_{i=1}^{N}\left((U^L)'_i(t) - \frac{1}{N}\sum_{j=1}^{N}(U^L)'_j(t)\right)^2\right)^{1/2},\end{aligned}$$

where $U_i^L(t) = (P^L y_i(t) - \underline{\hat{f}}^{L,L^*}(t))/(\hat{C}^{L,L^*}(t))^{1/2}$ and $(U^L)'_i$ is a smooth version of the differentiated function U_i^L. Then we take the L_1-norm of $\hat{\tau}^L$.

Let us describe the behavior of $\hat{d}^L$:

- $\|\hat{d}^L\|_\infty$ increases with L
- When the functions $(B_\ell^L)_{1\leq\ell\leq L}$ consists in an orthonormal family, $\|\hat{d}^L\|_\infty$ increases with L until $L = L^*$ and then $\|\hat{d}^L\|_\infty$ is constant with L.

This band is illustrated on Figure 4. It displays on the top row several functional data generated under either the Fourier family (left), Legendre (middle) or Spline (right), on the middle row the confidence bands of $\underline{f}^{L,L^*}$ for different values of $L = 3, 5$ and 11, and on the bottom row the bound $\hat{d}^L$. The true functions $\underline{f}^{L,L^*}$ are displayed in cyan and the confidence bands in purple. The bands are

very precise for each L. The behavior of $\hat{d}^L$ increases with L. As d^L can be seen as a variance, $\hat{d}^L(t)$ is larger on the boundary of the domain, as there are less observations near 0 and 1.

We also evaluate numerically the levels of the obtained confidence bands. For this, 1000 datasets are simulated, the confidence band is estimated for each of them. The empirical confidence level is then evaluated as the proportion of confidence bands that contain the true function. Table 1 presents the empirical confidence levels for different values of L and two sample sizes $n = 50$ and $n = 150$, with $N = 40$. The level is the expected one whatever the value of L, especially when $L < L^*$ and $L > L^*$ but also when $L > L^\varepsilon$. We will see in the next sections that this will not be the case for the debiased confidence band.

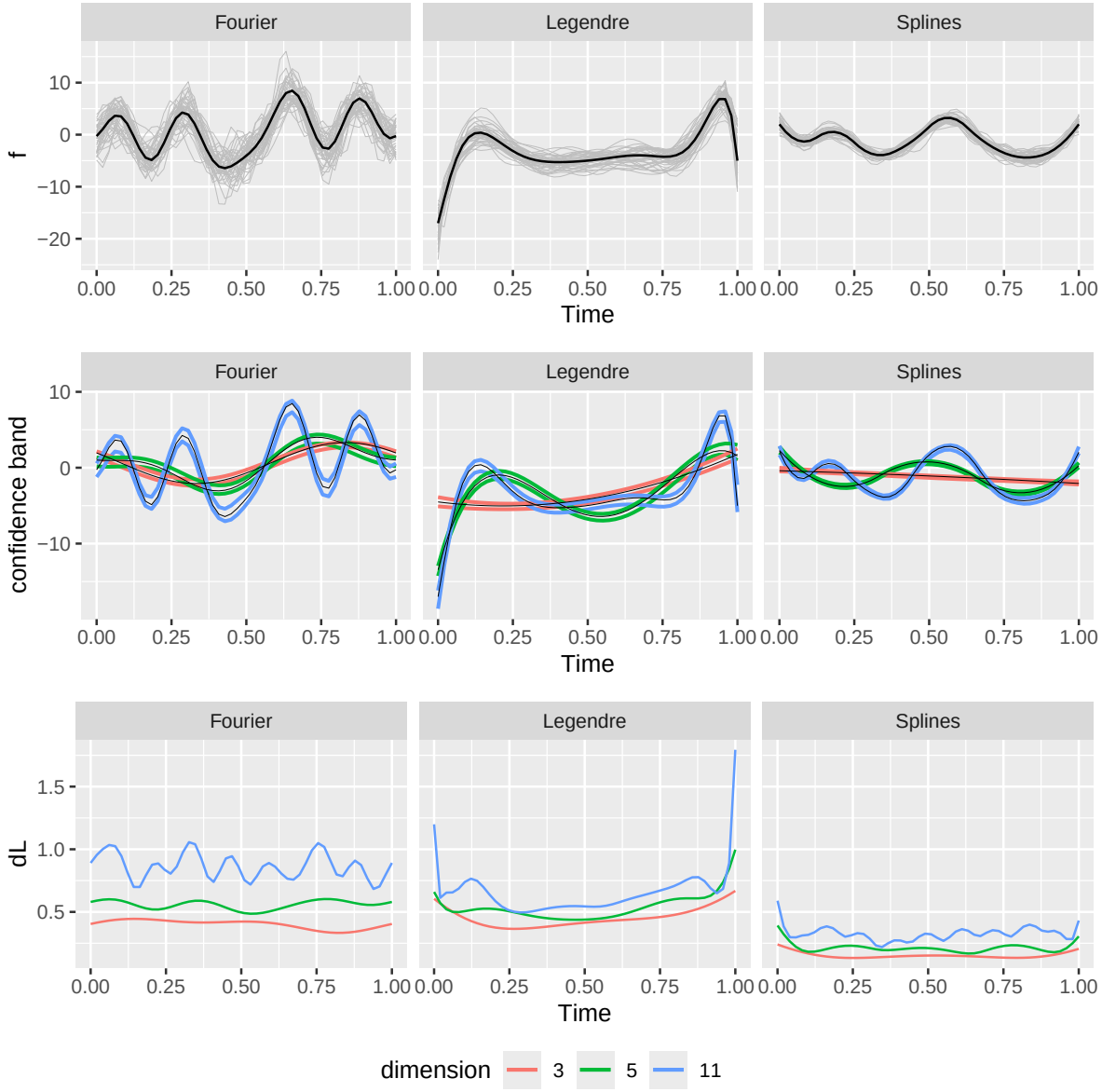

Figure 4: Illustrative example. For the three families, resp. Fourier, Legendre and the splines, we display on the top row the observed functional data, on the middle row the confidence bands for different values of L (3, 5 and 11), and on the bottom row the bound dL.

Table 1: Illustrative example. The confidence level of the confidence band is evaluated from 1000 repetitions. Confidence bands are calculated with the Legendre family, for several L in rows and several n in columns.

	50	150
3	0.972	0.957
5	0.950	0.950
11	0.944	0.952
15	0.968	0.945
21	0.938	0.943

3.2 Asymptotic confidence band for f^{L,L^*}

Note that if one works in the asymptotic framework $n \to \infty$, the previous definition of $\hat{d}^L$ induces a natural asymptotic confidence band for the function f^{L,L^*}. Indeed, we can prove that

Theorem 3.2. *Set Definition 2.1 and Definition 2.3 and a probability $\alpha \in [0,1]$. Then, we have,*

$$\lim_{n\to+\infty} \mathbb{P}\left(\forall t \in [0,1], |\underline{\hat{f}}^{L,L^*}(t) - f^{L,L^*}(t)| \leq \hat{d}^L(t)\right) = 1-\alpha,$$

with $\hat{d}^L(t) = \hat{q}^L\sqrt{\hat{C}^{L,L^*}(t,t)/N}$ and $\hat{q}^L$ defined as the solution of Equation 4.

The proof is given in Appendix.

Then a confidence band for f^{L,L^*} at the asymptotic confidence level $1-\alpha$ for a large number of observations n is given by

$$CB(f^{L,L^*}) = \{\forall t \in [0,1], [\underline{\hat{f}}^{L,L^*}(t) - \hat{d}^L(t); \underline{\hat{f}}^{L,L^*}(t) + \hat{d}^L(t)]\}.$$

We do not provide any illustration of this property, as it would be similar than the previous ones. Indeed, we notice that the asymptotic is achieved even when n is small on our examples.

4 Confidence Band of f^{L^*} by correcting the bias

The function of interest is $f^{L^*} = \underline{f}^{L^*,L^*}$, rather than $\underline{f}^{L,L^*}$. Therefore, our aim is to construct a confidence bound for f^{L^*}. However, an unbiased estimator of f^{L^*} is unavailable by definition, since the true dimension L^* is unknown. Instead, we propose to work with the estimator $\underline{\hat{f}}^{L,L^*}$ and to debias the corresponding confidence band.

To achieve this, we use the decomposition outlined in Equation 2 between the bias term and the statistical term. The idea is to bound the infinity norm of these two terms. A first strategy consists in bounding each term separately and then summing the two bounds to construct the confidence band. However, this approach tends to produce a band that is too large and conservative. The reason is that applying the infinite norms on each term before bounding them does not take into account the functional nature of the two terms. This approach is presented in details in Appendix.

A second strategy consists in keeping the functional aspect by bounding the infinity norm of the sum of the functional two terms. This approach is detailed in this section.

In Section 4.1, we first rewrite the band as a band around $\underline{f}^{L,L^*}(t)$. We then use a first subsample $\mathbf{y}^1$ to estimate the bound as defined in Section 3. A second subsample $\mathbf{y}^2$ is used to estimate the

bias term (without the infinite norm). This yields a pointwise correction of the bias, and the final confidence band is centered around $\underline{f}^{L_{\max},L^*}$. This procedure provides a collection of confidence bands, for $L = L_{\min}, \dots, L_{\max}$, each achieving a given confidence level but with varying width. Then, in Section 4.2, we propose a criterion to select the "best" band by minimizing its width.

4.1 Construction of the band of f^{L^*} for a given L

We introduce two independent sub-samples $\mathbf{y}^1$ and $\mathbf{y}^2$ of $\mathbf{y}$ of length n_1 and n_2 such that $n_1 + n_2 = n$.
On le coupe en temps ou bien on sous-echantillonne ?

We use the first sub-sample $\mathbf{y}^1$ to calculate $\hat{\underline{f}}_1^{L,L^*}(t)$, an estimator of $\underline{f}^{L,L^*}(t)$ and a functional bound denoted $\hat{d}_1^L$ that controls the bias term $\underline{f}^{L,L^*}(t) - \hat{\underline{f}}_1^{L,L^*}(t)$. This bound is defined in Section 3 applied on $\mathbf{y}^1$, for a given level α, such that:

$$\mathbb{P}\left(\forall t \in [0,1], -\hat{d}_1^L(t) \le \underline{f}^{L,L^*}(t) - \underline{f}_1^{L,L^*}(t) \le \hat{d}_1^L(t)\right) = 1 - \alpha. \tag{5}$$

Then, we need to control the bias $Bias_{L,L^*}(t) = \underline{f}^{L,L^*}(t) - f^{L^*}(t)$. Recall than when $L_{\max}$ is large enough and $n > L_{\max}$, $f^{L^*} = \underline{f}^{L_{\max},L^*}$. Therefore, we want to control the $Bias_{L,L^*}(t) = \underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}(t)$. It would be tempting to replace $Bias_{L,L^*}(t)$ by its estimation based on the second sample $\mathbf{y}^2$. But this would introduce an estimation error that we also need to control, in the same spirit than what is done in (**PCO?**). We can use again Section 3 to compute the function $\hat{d}_2^{L,L_{\max}}(t)$ on the sample $\mathbf{y}^2$, and the functional estimators $\underline{f}_2^{L,L^*}(t)$ and $\underline{f}_2^{L_{\max},L^*}(t)$ of $\underline{f}^{L,L^*}(t)$ and $\underline{f}^{L_{\max},L^*}(t)$, respectively. This allows to construct the following band for $\underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}$ for a confidence level $1 - \beta$,

$$\mathbb{P}\left(\forall t \in [0,1], -\hat{d}_2^{L,L_{\max}}(t) \le \underline{f}^{L_{\max},L^*}(t) - \underline{f}^{L,L^*}(t) - (\underline{f}_2^{L_{\max},L^*}(t) - \underline{f}_2^{L,L^*}(t)) \le \hat{d}_2^{L,L_{\max}}(t)\right) = 1 - \beta. \tag{6}$$

Combining Equation 5 and Equation 6, we can provide a debiased confidence band of $f^{L^*}(t)$.

Proposition 4.1. *Let us define*

$$\begin{aligned}
\hat{\theta}_1^L(t) &:= -\hat{d}_1^L(t) - \hat{d}_2^{L,L_{\max}}(t) + \underline{f}_2^{L_{\max},L^*}(t) - \underline{f}_2^{L,L^*}(t) \\
\hat{\theta}_2^L(t) &:= \hat{d}_1^L(t) + \hat{d}_2^{L,L_{\max}}(t) + \underline{f}_2^{L_{\max},L^*}(t) - \underline{f}_2^{L,L^*}(t),
\end{aligned}$$

where $\hat{d}_1^L(t)$ is defined on sample $\mathbf{y}^1$ by Equation 5 for a level α and $\hat{d}_2^{L,L_{\max}}(t)$ is defined on sample $\mathbf{y}^2$ by Equation 6 for a level β. Then we have

$$\mathbb{P}\left(\forall t \in [0,1], \quad \hat{\theta}_1^L(t) \le f^{L^*}(t) - \underline{f}_1^{L,L^*}(t) \le \hat{\theta}_2^L(t)\right) \ge 1 - \alpha\beta.$$

The proof is given in Appendix.

This defines a confidence band which can be defined either around $\hat{\underline{f}}_1^{L,L^*}$:

$$CB_2(\underline{f}^{L^*}) = \{\forall t \in [0,1], [\hat{\underline{f}}_1^{L,L^*}(t) + \hat{\theta}_1^L(t)\,;\, \hat{\underline{f}}_1^{L,L^*}(t) + \hat{\theta}_2^L(t)]\}$$

or around $\underline{f}_2^{L_{\max},L^*}$:

$$CB_2(\underline{f}^{L^*}) = \{\forall t \in [0,1], [\underline{f}_2^{L_{\max},L^*}(t) + \hat{\underline{f}}_1^{L,L^*}(t) - \underline{f}_2^{L,L^*}(t) - \hat{d}_1^L(t) - \hat{d}_2^{L,L_{\max}}(t)\,;\, \underline{f}_2^{L_{\max},L^*}(t) + \hat{\underline{f}}_1^{L,L^*}(t) - \underline{f}_2^{L,L^*}(t) + \hat{d}_1^L(t) + \hat{d}_2$$

The two functions $\hat{d}_1^L(t)$ and $\hat{d}_2^{L,L_{\max}}(t)$ are of the same order as they are built with the same approach. They depend on the length of the samples. To obtain the thinnest band, the best strategy is to split the sample in two sub-samples of equal length $n_1 = n_2 = n/2$.

The behavior of $\hat{d}_1^L$ has been described in Section 3. Let us describe the behavior of $\hat{d}_2^{L,L_{\max}}$:

- $\|\hat{d}_2^{L,L_{\max}}\|_\infty$ decreases with L.
- When $L > L^\varepsilon$, $\|\hat{d}_2^{L,L_{\max}}\|_\infty$ is constant with L and the probability in Equation 6 is equal to 1.
- When $L^* < L < L^\varepsilon$, $\|\hat{d}_2^{L,L_{\max}}\|_\infty$ is constant with L when the functions B_ℓ^L consists in an orthonormal family. Otherwise, the behavior is erratic.

It means that when the band defined in Proposition 4.1 is calculated for $L > L^\varepsilon$, the confidence level is $1-\alpha$ instead of $1-\alpha\beta$.

The advantages of this approach is that the bias of the band is corrected and the level for the true function f^{L^*} is guaranteed when L^ε is large. This was the main objective of the paper.

We illustrate numerically those advantages. In Figure 5, top row, we plot the confidence bands obtained for different dimensions $L \in \{3, 5, 11, 15, 21\}$ with Fourier, Legendre and Splines families and $\alpha = \beta = \sqrt{0.05} \approx 0.22$. We can see that all the confidence bands are alike. Especially, they are unbiased, even for $L = 3$. A larger dimension L provides a smoother band. On the middle and bottom rows of Figure 5, we illustrate the two terms that enter the confidence band, $t \mapsto \hat{d}_1^L(t)$ and $t \mapsto \hat{d}_2^{L,L_{\max}}(t)$. Their behavior is the same along time. The function $\hat{d}_1^L(t)$ can be seen as a variance, this is why it is larger near 0 and 1 where there are less observations. The function $\hat{d}_2^{L,L_{\max}}(t)$ is smaller than $\hat{d}_1^L(t)$ because it controls the remaining rest after the projection. Note that as expected when $L > L^\varepsilon$, $\hat{d}_2^{L,L_{\max}}(t)$ is close to 0. As explained before, the influence of L is not the same for the two functions. When L increases, $\hat{d}_1^L(t)$ increases while $\hat{d}_2^{L,L_{\max}}(t)$ decreases.

In Table 2, we simulate 1000 repeated datasets with two sample sizes $n = 50$ and $n = 150$. For each dataset, we compute the confidence band defined in Proposition 4.1 with a theoretical confidence level of $1-\alpha\beta = 0.95$ and for different values of L. Then the confidence level is approximated as the proportion of confidence bands containing the true function f. Remark that when $L < L^\varepsilon$, the level is the expected one, that is 0.95. When $L > L^\varepsilon$, the level is not more ensured, as explained before. Indeed the term $d^{L,L^{\max}}$ is mainly equal to 0, and the level is close to $1-\alpha$ instead of $1-\alpha\beta$. This is not the case for the band in Section 3, as this is due to the correction of the bias.

Table 2: Illustrative example. We display the level of confidence for the proposed confidence band, for several L in rows and several n in columns.

	50	150
3	0.960	0.963
5	0.969	0.970
11	0.961	0.970
15	0.947	0.958
21	0.780	0.799
25	0.803	0.787

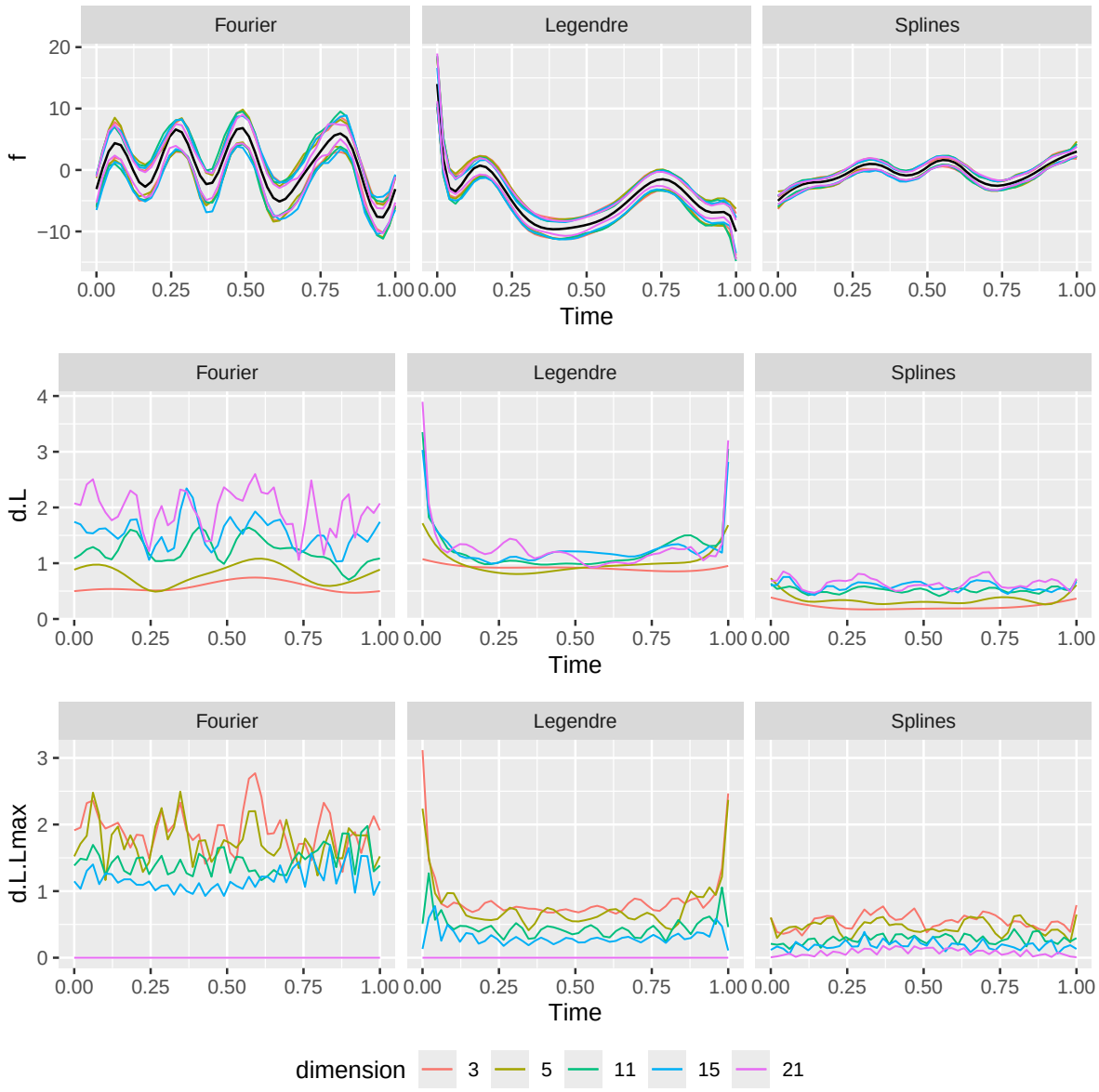

Figure 5: Illustrative example. For a given dataset, we plot several confidence bands (top row), functions dL (middle row) and dLLmax (bottom row). Bands and functions are estimated with Fourier (left column), Legendre (middle column) and Spline (right column) basis and several dimensions L (3, 5, 11, 15, 21).

4.2 Influence of L

This approach gives a collection of confidence bands for different values of L. The confidence bands have different widths for a same confidence level $1-\alpha\beta$. It is thus natural to want to select one of them. This means we want to select the best dimension L among the collection $\{1,\dots,L_{\max}\}$. We need to define what "best" means. It is quite intuitive to focus on the band that is the thinnest. Thinnest could be thought in different norms, the infinity norm, the L1 norm. Here we consider the infinity norm of the width of the confidence band, which gives a preference to smooth bands. We thus define the following criteria

$$\hat{L} = \arg\min_L \left\{ \sup_t |\hat{\theta}_2^L(t) - \hat{\theta}_1^L(t)| \right\} = \arg\min_L \left\{ \sup_t |\hat{d}^L(t) + \hat{d}^{L,L_{\max}}(t)| \right\}. \tag{7}$$

We illustrate the different terms involved in Equation 7. In Figure 6, we plot for a given dataset, the infinity norm of the width of the band $\hat{d}^L(t) + \hat{d}^{L,L_{\max}}(t)$ (top), of $\hat{d}^L(t)$ (middle) and $\hat{d}^{L,L_{\max}}(t)$ (bottom) functions obtained with the Fourier (left column), Legendre (middle column) and Spline (right column) basis. As already said, $\|\hat{d}^L\|_\infty$ increases with L while $\|\hat{d}^{L,L_{\max}}\|_\infty$ decreases (and is zero when $L > L\varepsilon$). The width of the band wrt L does not have a U-shape, as expected. It is thus difficult to minimize this criterion.

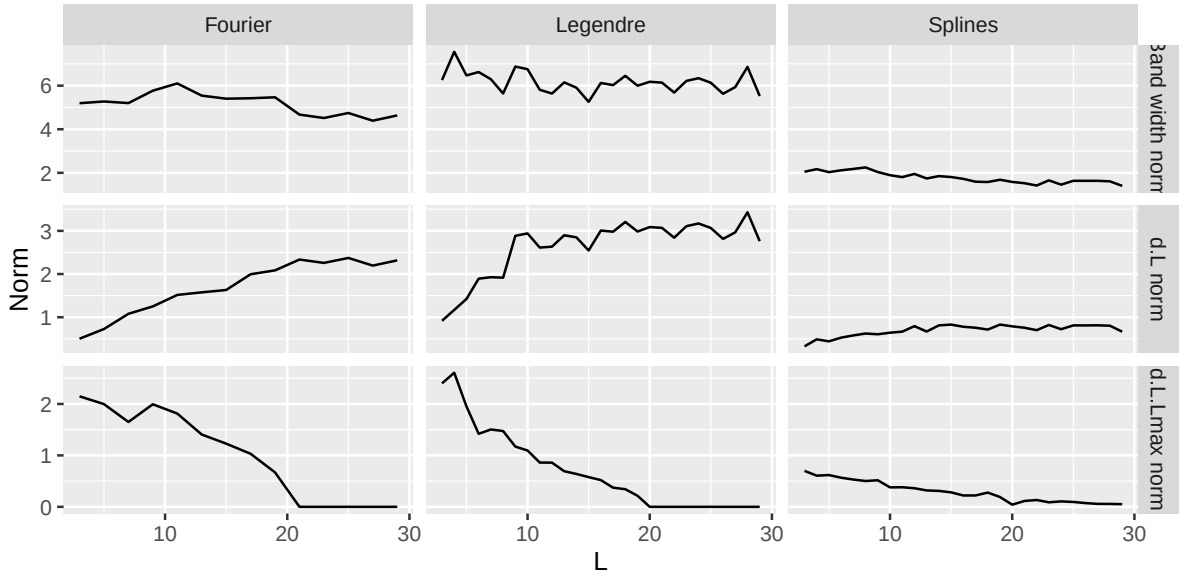

Figure 6: Illustrative example. For a given dataset, we calculate the norm of the width of the confidence band (top), of the dL function (middle) and the dLLmax function (bottom), for several dimensions L and for Fourier (left column), Legendre (middle column) and Splines (right column) basis.

We then evaluate the performance of the selection criteria. To do that, we simulate 100 repeated datasets. Confidence bands and the norm of their widths are computed for several L. We apply the selection criteria and plot the distribution of the estimated dimension $\hat{L}$ in Figure 7, for the three basis families. The estimated dimension is almost always larger than the true $L^* = 11$. Being larger is not a problem and the selected band is unbiased and has the correct level as soon as L^ε is large, which was the objective. However, the criteria has the tendency to select a (too) smooth band. It is not satisfactory not to be able to select the true dimension. This is why we propose a new approach in Section 5. In Section 4.3, we also illustrate that this band is too conservative modifier.

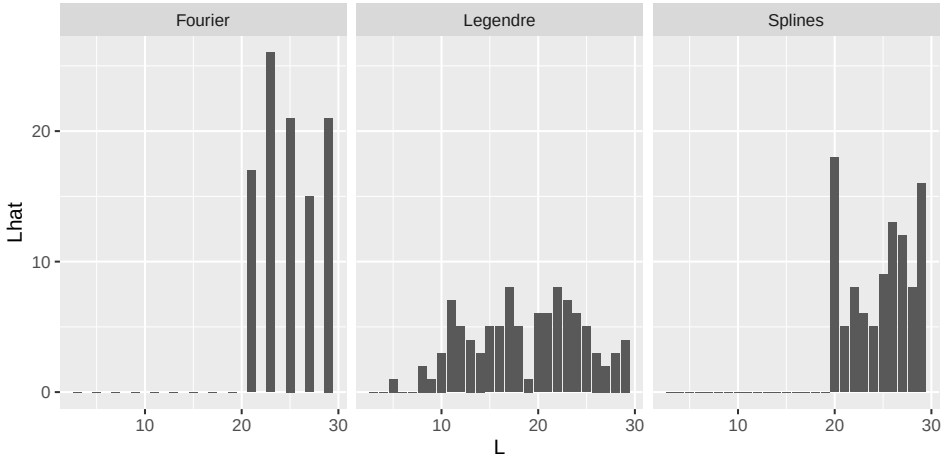

Figure 7: Illustrative example. From 100 datasets, we calculate the distribution of the estimated dimension L. The true dimension is $L^* = 11$.

4.3 Comparison with the confidence bands of Section 3

The reformulation of the band around $\underline{f}_2^{L_{\max},L^*}$ is close to the band presented in Section 3 for $L = L_{\max}$, that is a band centered around $\underline{f}^{L_{\max},L^*}$. A natural question is to understand what is the gain by doing so instead of using the band from Section 3 with $L = L_{\max}$, namely the band $[\underline{f}^{L_{\max},L^*}(t) - \hat{d}^{L_{\max}}(t); \underline{f}^{L_{\max},L^*}(t) + \hat{d}^{L_{\max}}(t)]$. To do that, we have to understand the behavior of the different terms.

It is difficult to compare theoretically the width of the two bands. We compare them with simulations. For 100 repeated datasets, we compute three different confidence bands: the band defined in Proposition 4.1 with $\hat{L}$ defined in Equation 7, the band constructed in Section 3 with $L_{\max}$ and the ideal (and not accessible) band constructed in Section 3 with the true L^*. In Figure 8, we present the boxplots of the norms of the width of the band with $\hat{L}$ (left), with $L_{\max}$ (middle) and L^* (right). The width of the confidence band with the true L^* is smaller, which is expected but unfortunately not achievable. What was not expected, but sad, is that the width of the confidence band constructed in Section 3 with some $L_{\max}$ is smaller than our band with a correction of the bias and the model selection criteria. This may be understood because we estimate two different quantities, on smaller dataset, for more conservative level ($1-\alpha$ and $1-\beta$ respectively) to achieve at the end the confidence level of $1-\alpha\beta$. The use of the two independent subsamples is mandatory to control the probability in the proof of Theorem 3.1. Therefore it is not possible to correct this problem. In the next section, we come back to the confidence bands proposed in Section 3 and propose a model selection criterion to take into account the bias.

5 Selection criteria of the best confidence band accounting for the bias

In this section, we want to propose a new heuristic criteria going back to the definition of the band itself seen as the estimation of a quantile of a certain empirical process. The criteria is inspired by model selection tools to select the best dimension L.

We work on the quantile q^L introduced in Equation 3, its oracle version q^{L^*} for the level L^* and the estimation $\hat{q}^L$. All of them are scalar, in a collection of scalars, with $L = 1, \dots, L_{\max}$. A natural criteria to choose the best L is such that the estimator $\hat{q}^L$ minimizes the quadratic error $\mathbb{E}\left(\|q^{L^*} - \hat{q}^L\|^2\right)$.

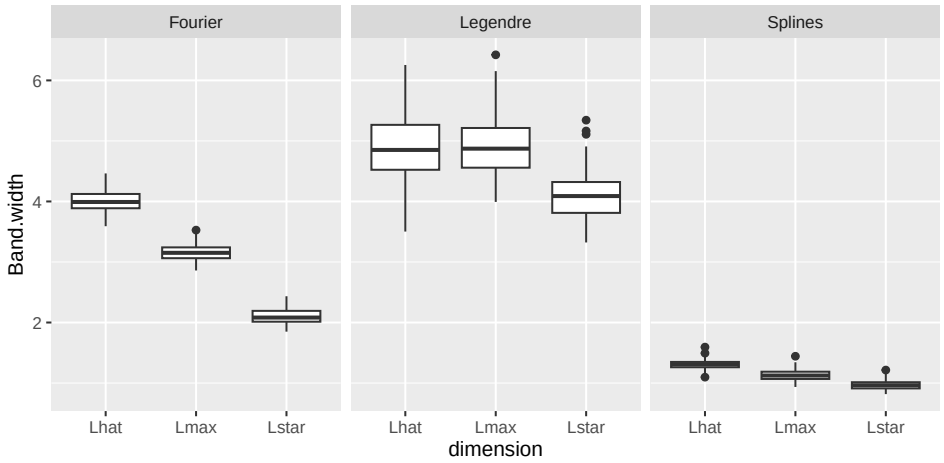

Figure 8: Illustrative example. We display within boxplots the confidence band's width over 100 repetitions for the model selected by our criterion, some fixed Lmax and the true (unknown) L^* for Fourier (left), Legendre (middle) and Splines (right) basis.

However, this quadratic error is unknown as q^{L^*} is unknown. We can not directly use it.

Instead, we study $\|\hat{q}^{L_{\max}} - \hat{q}^{L}\|^2$. We assume that $L_{\max}$ is large enough such that $\underline{f}^{L_{\max},L^*} = f^{L^*}$. While the theoretical error decreases when $L < L^*$ and increases when $L > L^*$, this approximation is still decreasing when $L > L^*$, as illustrated in Figure 9.

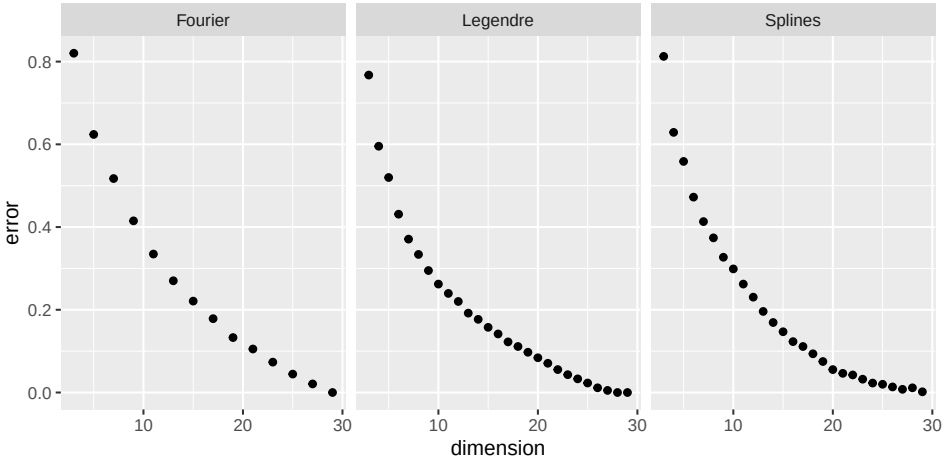

Figure 9: Illustrative example. For a given simulated dataset, we show the behavior of the error as a function of the dimension L, for Fourier (left), Legendre (middle) and Splines (right) basis.

The, we recognize a behavior similar to a bias, high with model of small dimension, and small when dimension is large. Selecting a model using this criterion will always overfit the data. Thus, we propose to penalize by the dimension, as usually in model selection criterion.

old paragraph - start

Instead, we use the classic decomposition into bias and variance:

$$\mathbb{E}\left(\|q^{L^*} - \hat{q}^{L}\|^2\right) = \|q^{L^*} - q^{L}\|^2 + \mathbb{E}\left(\|q^{L} - \hat{q}^{L}\|^2\right) =: b^L + v^L,$$

where the first term $b^L = \|q^{L^*} - q^{L}\|^2$ is the bias term and $v^L = \mathbb{E}\left(\|q^{L} - \hat{q}^{L}\|^2\right)$ is the variance term. Note that these terms are not functional and different from the functional bias defined previously.

The two terms are unknown. Let us denote $\hat{b}^L$ and $\hat{v}^L$ their estimators. We assume that $L_{\max}$ is large enough such that $\underline{f}^{L_{\max},L^*} = f^{L^*}$. Under this assumption, instead of computing the bias $\|q^{L^*} - q^L\|^2$, we focus on $\|q^{L_{\max}} - q^L\|^2$. This is not an equality because the first term decreases when $L < L^*$ and increases when $L > L^*$, while the second term is still decreasing when $L > L^*$. However, we just need the decreasing behavior of the bias which will be compensated by the increasing of the variance. We then introduce the estimator of the bias as $\hat{b}^L = \|\hat{q}^{L_{\max}} - \hat{q}^L\|^2$. For the variance term, we need to bound it. Let us remark that the variance of $\hat{q}^L$ is of order L/N. Est ce qu'on a une ref pour justifier ca ? J'ai rien trouvé. Mais pour estimer q^L, on utilise l'estimateur $\hat{\tau}^L$. Il faudrait pouvoir calculer la variance de cet estimateur pour en déduire l'ordre de celle de $\hat{q}^L$? Thus we set

$$\hat{v}^L := \lambda \frac{L}{N},$$

where $\lambda > 0$ is a tuning parameter.

old paragraph - end

A natural criteria to select the best L is then

$$crit(L) = \|\hat{q}^{L_{\max}} - \hat{q}^L\|^2 + \lambda \frac{L}{N}.$$

Then we define

$$\hat{L} = \arg\min_L crit(L),$$

and take the band centered around $\underline{f}^{\hat{L},L^*}$.

In Figure 10, we illustrate the behavior of this model selection criterion on simulated data, with $\lambda = 1$ for the three basis. We can see that $\hat{L}$ is overestimated. As we work with nested spaces, it ensures that $\hat{L}$ has the tendency to be larger than L^* and thus the confidence band is automatically unbiased.

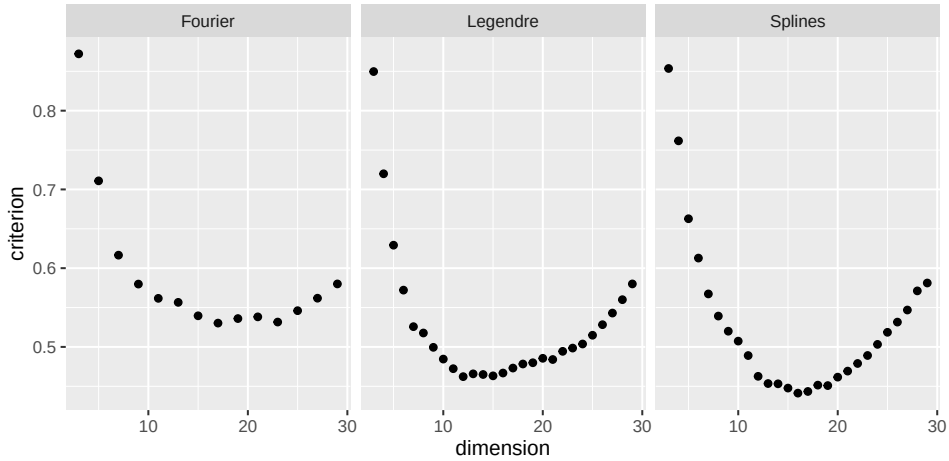

Figure 10: Illustrative example. For a given simulated dataset, we show the behavior of the criteria as a function of the dimension L, for Fourier (left), Legendre (middle) and Splines (right) basis.

In Figure 11, we test which model is selected over 100 repetitions for the three basis. The estimated dimension is equal or larger than the true $L^* = 11$. As in Figure 7, being larger is not a problem. However, the selected dimension is smaller in distribution, and closer to the true value than in Figure 7. Thus this new approach performs better. Moreover, as we then use the confidence band of Section 3, the confidence level is ensured to be the expected one.

We then show in Figure 12 that the width of the selected model is better than the width of the confidence band with a large level $L_{\max}$, which one should have used to avoid model selection.

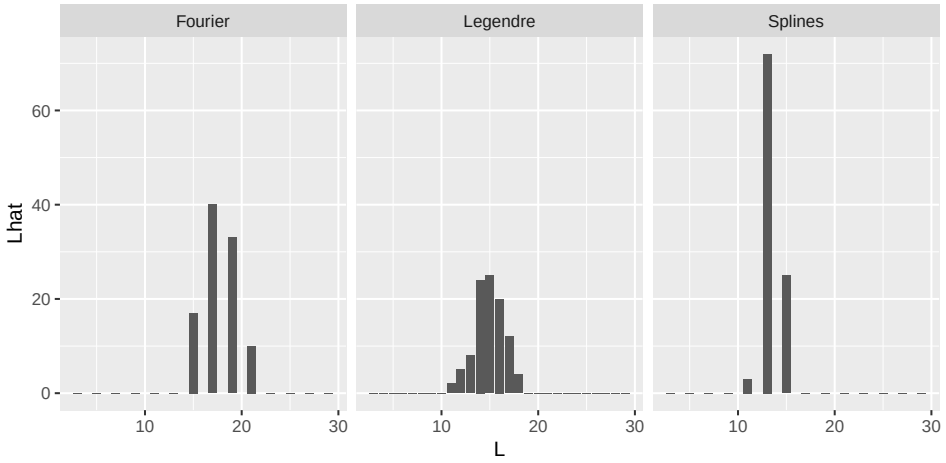

Figure 11: Illustrative example. We show the distribution of the selected model, over 100 repetitions, with the new criteria used to select a model for different basis.

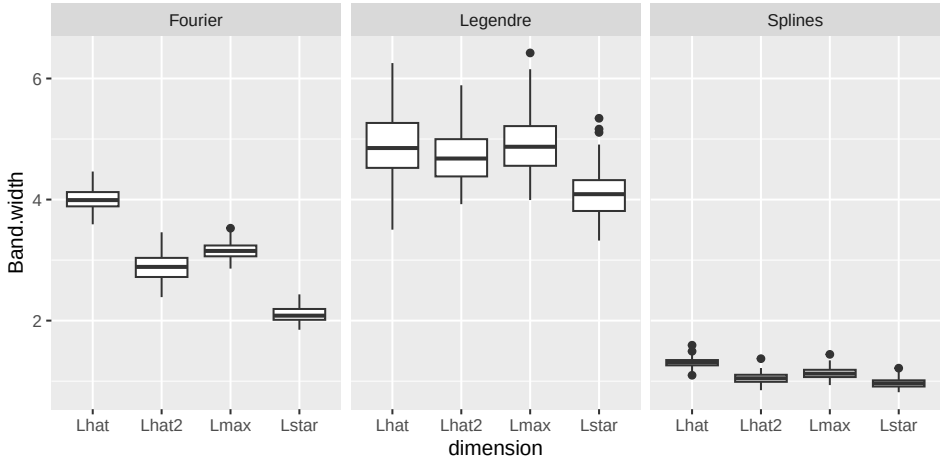

Figure 12: Illustrative example. We display within boxplots the confidence band's width over 100 repetitions for the model selected by our criterion, some fixed Lmax and the true (unknown) L^* for Fourier (left), Legendre (middle) and Splines (right) basis.

6 Real data analysis

In this section, we illustrate the proposed method on the Berkeley Growth Study data. It consists of the heights in centimeters of 39 boys at 31 ages from 1 to 18. We approximate those curves by our 3 basis, namely Legendre, Splines and Fourier. We select the level of each basis using the method introduced in Section 5.

```
[1] "model selected:6"
[1] "length L sel:1.98960434593563"
[1] "model selected:5"
[1] "length L sel:1.94736120271032"
[1] "model selected:7"
[1] "length L sel:2.07846739360277"
```

	Legendre	Splines	Fourier
Length Lmax	2.124539	2.124536	2.204230
Length selected	1.989604	1.947361	2.078467
Model selected	6.000000	5.000000	7.000000

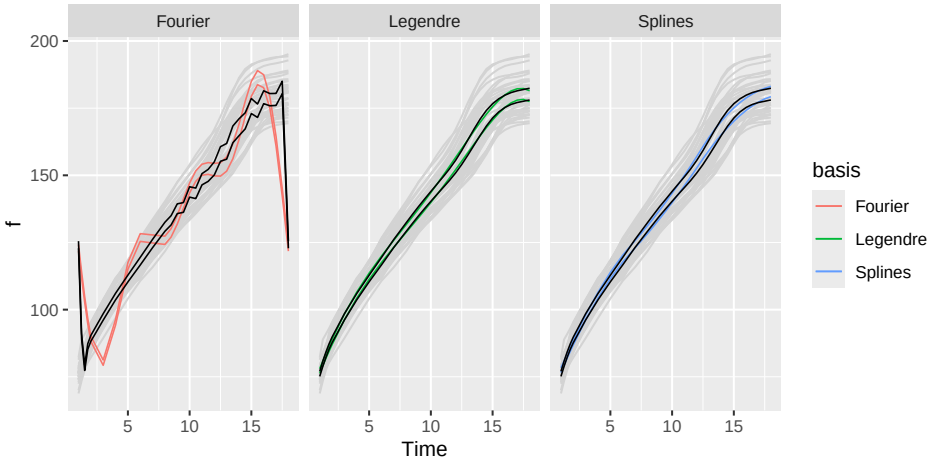

Figure 13: Real data analysis example. We display the confidence bands for Fourier (left), Legendre (middle) and Splines (right) basis on the Berkeley Growth Study data. Black curves correspond to the confidence bands with L_{max}, while colored one are the confidence bands constructed by our data-driven method.

As the data is not periodic, the Fourier basis is meaningless, and so is the associated confidence band, whatever the level considered. Both splines and Legendre basis give similar confidence bands. When analyzing the length of the bands compared with the one with $L_{\max}$ coefficients, we see that there are less smooth but also smaller, and from our empirical study we guess that it makes a trade-off between bias and variance.

7 Conclusion

This paper discusses the construction of confidence bands when considering a linear model over a functional family. Depending on the nature of the family (an orthogonal or orthonormal basis, or just a vector space), theoretical guarantees of the linear estimator are reminded and illustrated. Then, several confidence bands are proposed. First, when considering a functional family with a fixed level, we discuss the confidence band derived from Sun and Loader (1994). It is bias if the level is not high enough to approximate well the true function. Then, a new confidence band is proposed that

correct this bias. To do so, the bias is estimated and the additional randomness is controlled. A model selection criterion is proposed to select the best level. Unfortunately, the two kinds of randomness are leading to a larger confidence band, and this result is then not more interesting than the naive one, which consists in taking the largest level possible $L_{\max}$. Finally, a heuristic model selection criterion is proposed to select the level on the first confidence band, that did not correct the bias. It takes into account the bias as well as the variance, to select a moderate level. Throughout the paper, extensive experimental study on Fourier, Legendre and Spline basis have illustrated the theoretical and methodological proposition, and a real data study is proposed to conclude the paper.

The last model selection criterion is heuristic, while each term is intuitive. An interesting next step, but out of the scope of this paper, consists of a theoretical study of this criterion. No result, to our knowledge, exist for confidence band with the supremum norm. The euclidean norm is well-studied in general, but is not of interest here, where we want to ensure that the tube is valid as a whole. The supremum norm, on its side, is difficult to study theoretically. A keypoint here also is the randomness of the criterion, that has also to be taken into account, through an oracle inequality for example.

References

Aneiros, Germán, Silvia Novo, and Philippe Vieu. 2022. "Variable Selection in Functional Regression Models: A Review." *Journal of Multivariate Analysis* 188: 104871. https://doi.org/https://doi.org/10.1016/j.jmva.2021.104871.

Basna, Rani, Hiba Nassar, and Krzysztof Podgórski. 2022. "Data Driven Orthogonal Basis Selection for Functional Data Analysis." *Journal of Multivariate Analysis* 189: 104868. https://doi.org/https://doi.org/10.1016/j.jmva.2021.104868.

Bunea, Florentina, Andrada E. Ivanescu, and Marten H. Wegkamp. 2011. "Adaptive Inference for the Mean of a Gaussian Process in Functional Data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (4): 531–58. https://doi.org/https://doi.org/10.1111/j.1467-9868.2010.00768.x.

Claeskens, G., and I. Van Keilegom. 2003. "Bootstrap Confidence Bands for Regression Curves and Their Derivatives." *Ann. Stat.*

Diquigiovanni, Jacopo, Matteo Fontana, and Simone Vantini. 2022. "Conformal Prediction Bands for Multivariate Functional Data." *Journal of Multivariate Analysis* 189: 104879. https://doi.org/https://doi.org/10.1016/j.jmva.2021.104879.

Goepp, V., O. Bouaziz, and G. Nuel. Submitted. "Spline Regression with Automatic Knot Selection," Submitted.

Hall, P. 1991. "On Convergence Rates of Suprema." *Probab Theory Related Fields.*

Jacques, Julien, and Sanja Samardžić. 2022. "Analysing Cycling Sensors Data Through Ordinal Logistic Regression with Functional Covariates." *Journal of the Royal Statistical Society Series C: Applied Statistics* 71 (4): 969–86. https://doi.org/10.1111/rssc.12563.

Kokoszka, P., and M. Reimherr. 2017. *Introduction to Functional Data Analysis.* Chapman & Hall / CRC Numerical Analysis and Scientific Computing. CRC Press. https://books.google.dk/books?id=HIxIvgAACAAJ.

Krivobokova, Tatyana, Thomas Kneib, and Gerda Claeskens. 2010. "Simultaneous Confidence Bands for Penalized Spline Estimators." *Journal of the American Statistical Association* 105 (490): 852–63.

Li, Yehua, Yumou Qiu, and Yuhang Xu. 2022. "From Multivariate to Functional Data Analysis: Fundamentals, Recent Developments, and Emerging Areas." *Journal of Multivariate Analysis* 188: 104806. https://doi.org/https://doi.org/10.1016/j.jmva.2021.104806.

Liebl, D, and M. Reimherr. 2019. "Fast and Fair Simultaneous Confidence Bands for Functional Parameters."

Neumann, M., and J. Polzehl. 1998. "Simultaneous Bootstrap Confidence Bands in Nonparametric Regression." *J Nonparametr Statist.*

Quinton, J-C., E. Devijver, A. Leclercq-Samson, and A. Smeding. 2017. "Functional Mixed Effect Models for Mouse-Tracking Data in Social Cognition." In *ESCON Transfer of Knowledge Conference, Gdansk, Polland.*

Sachs, Michael C., Adam Brand, and Erin E. Gabriel. 2022. "Confidence Bands in Survival Analysis." *The British Journal of Cancer. Supplement* 127: 1636–41. https://doi.org/10.1038/s41416-022-01920-5.

Sun, Jiayang, and Clive R. Loader. 1994. "Simultaneous Confidence Bands for Linear Regression and Smoothing." *Ann. Statist.* 22 (3): 1328–45. https://doi.org/10.1214/aos/1176325631.

Telschow, Fabian J. E., Dan Cheng, Pratyush Pranav, and Armin Schwartzman. 2023. "Estimation of expected Euler characteristic curves of nonstationary smooth random fields." *The Annals of Statistics* 51 (5): 2272–97. https://doi.org/10.1214/23-AOS2337.

Telschow, Fabian J. E., and Armin Schwartzman. 2022. "Simultaneous Confidence Bands for Functional Data Using the Gaussian Kinematic Formula." *Journal of Statistical Planning and Inference* 216: 70–94. https://doi.org/https://doi.org/10.1016/j.jspi.2021.05.008.

Xia, Y. 1998. "Bias-Corrected Confidence Bands in Nonparametric Regression." *J.R. Stat. Soc. Ser. B.*

Zhou, S., X. Shen, and D. A. Wolfe. 1998. "Local Asymptotics for Regression Splines and Confidence Regions." *Ann. Statist.*

8 Appendix: proofs

8.1 Proof of Proposition Proposition 2.3

Let us prove the first point. We have

$$\mathbb{E}(\hat{\underline{\mu}}^{L,L^\star}) = (\mathbf{B}_L^T\mathbf{B}_L)^{-1}\mathbf{B}_L^T\mathbb{E}(\mathbf{y}) = (\mathbf{B}_L^T\mathbf{B}_L)^{-1}\mathbf{B}_L^T\mathbf{B}_{L^*}\mu^{L^*} =: \underline{\mu}^{L,L^\star}.$$

The theory of the linear model gives that the variance of $\hat{\underline{\mu}}^L$ is equal to $\sigma^2(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\Sigma\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}$ with $\Sigma = Diag(\Sigma_1, \dots, \Sigma_N)$ the $nN \times nN$ covariance matrix of $\mathbf{y}$. So finally, we have

$$\hat{\underline{\mu}}^{L,L^\star} \sim \mathcal{N}\left(\underline{\mu}^{L,L^\star}, \sigma^2\Sigma_B^{L,L^\varepsilon}\right).$$

Now we can easily deduce the distribution of $\hat{\underline{f}}^{L,L^*}(t)$, for each $t \in [0,1]$:

$$\hat{\underline{f}}^{L,L^*}(t) - \mathbf{f}^{L,L^*}(t) \sim \mathcal{N}\left(0, \sigma^2 B(t)\Sigma_B^{L,L^\varepsilon}B(t)^T\right).$$

8.2 Proof of Theorem Theorem 3.2

We have

$$P(\forall t \in [0,1], |\hat{\underline{f}}^{L,L^*}(t) - f^{L,L^*}(t)| \leq \hat{d}^L(t)) = P(\forall t \in [0,1], |\hat{\underline{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) + \underline{f}^{L,L^*}(t) - f^{L,L^*}(t)| \leq \hat{d}^L(t))$$

Set assumptions Definition 2.1 and Definition 2.3 and a probability $\alpha \in [0,1]$. Then, we have,

$$\lim_{n\to+\infty} P(\forall t \in [0,1], |\hat{\underline{f}}^{L,L^*}(t) - f^{L,L^*}(t)| \leq \hat{d}^L(t)) = 1 - \alpha$$

with $\hat{d}^L(t) = \hat{c}^L\sqrt{\hat{C}_L(t,t)/N}$ and $\hat{c}^L$ defined as the solution of Equation 4.

8.3 Proof of Proposition 4.1

To simplify the notations, let us denote $a(t) = \underline{f}^{L,L^*}(t) - \underline{\hat{f}}_1^{L,L^*}(t)$ and $b(t) = \underline{f}^{L_{\max},L^*}(t) - \underline{f}^{L,L^*}(t) - (\underline{f}_2^{L_{\max},L^*}(t) - \underline{f}_2^{L,L^*}(t))$. We have

$$P\left(\exists t |a(t)+b(t)| \geq \hat{d}_1^L(t) + \hat{d}_2^{L,L_{\max}}(t)\right) \leq P\left(\exists t |a(t)| + |b(t)| \geq \hat{d}_1^L(t) + \hat{d}_2^{L,L_{\max}}(t)\right)$$
$$= P\left(\exists t |a(t)| \geq \hat{d}_1^L(t)\right) P\left(\exists t |b(t)| \geq \hat{d}_2^{L,L_{\max}}(t)\right) = \alpha\beta.$$

The last equality holds thanks to the independence of the two sub-samples.

9 Heuristics of bounding separately the infinity norms of the bias term and the approximation term

Let us give the main ideas. We denote

$$A_{L,L^*} = \sup_t |Approx(t)| = \sup_t |\underline{f}^{L,L^*}(t) - \underline{f}^{L^*,L^*}(t)| = \sup_t |\underline{f}^{L,L^*}(t) - f^{L^*}(t)|,$$
$$b_{L,L^*} = \sup_t |Bias(t)| = \sup_t |\underline{\hat{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t)|,$$

the two infinity norms of the approximation and the bias terms. Recall that when $L \geq L^*$, $Approx(t) = 0$ and thus $A_{L,L^*} = 0$. Then by the triangular inequality, we get

$$\sup_t |\underline{\hat{f}}^{L,L^*}(t) - f^{L^*}(t)| \leq A_{L,L^*} + b_{L,L^*}.$$

If we find a bound M^L such that

$$P(A_{L,L^*} + b_{L,L^*} \leq M^L) \geq 1 - \alpha, \tag{8}$$

it will imply a confidence band of order $1-\alpha$ for $f^{L^*}(t)$:

$$P(\sup_t |\underline{\hat{f}}^{L,L^*}(t) - f^{L^*}(t)| \leq M^L) \geq 1 - \alpha.$$

{#eq-ML}}

The problem reduces to find M^L which bounds both the bias term and the approximation term. We bound each term in Section 9.1 and Section 9.2, respectively. The term A_{L,L^*} is deterministic. However, as it is analytically unknown, we will estimate an upper bound based on observations. To guarantee the independence between the two bounds, the sample of observations is split in two parts and the two bounds are calculated on one half of the sample. We denote $\mathbf{y}^1 = (y_1, \dots, y_{n/2})$ and $\mathbf{y}^2 = (y_{n/2+1}, \dots, y_n)$ the two sub-samples. Then we introduce the confidence band of f^{L^*} based on $\underline{f}^{L,L^*}$ in Section 9.3.

9.1 Bound of the bias term.

The bound of the bias term is calculated with the first sub-sample $\mathbf{y}^1$ of size $n/2$. The estimator $\underline{\hat{f}}_1^{L,L^*}$ is calculated on this sub-sample. The statistical bias b_{L,L^*} is the supremum of a Gaussian process.

Using the confidence band proposed in Section **??**, for a given confidence level $1-\alpha_1$, there exists an estimable function $\hat{d}_1^L()$ defined as $\hat{d}_1^L(t) = \hat{c}^L\sqrt{\frac{\hat{C}_L(t)}{n/2}}$ such that

$$P(\forall t, \hat{\underline{f}}_1^{L,L^*}(t) - \hat{d}_1^L(t) \leq \underline{f}^{L,L^*}(t) \leq \hat{\underline{f}}_1^{L,L^*}(t) + \hat{d}_1^L(t)) = 1-\alpha_1.$$

Let us denote $\hat{d}_1^L = \sup_t |\hat{d}_1^L(t)|$. Then with probability $1-\alpha_1$, we have

$$b_{L,L^*} \leq \hat{d}_1^L.$$

{eq:borne_bL}

9.2 Bound of the approximation term.

Voir comment on va présenter les deux cas selon L et L^*. Soit ici soit apres ? Ou dit on que meme si on estime A_{L,L^*}, ca ne change pas grand chose quand L est grand ??

Let us now bound the approximation term which has no explicit form. We thus need to estimate a bound of A_{L,L^*} and will use the second sub-sample $\mathbf{y}^2$ to do that.

Let us consider $L_{\max}$ large enough such that $\underline{f}^{L^*,L^*} = \underline{f}^{L_{\max},L^*}$. A natural estimator of $\underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}(t)$ is $\hat{\underline{f}}^{L,L^*}(t) - \hat{\underline{f}}^{L_{\max},L^*}(t)$. We can then decompose the approximation term by introducing its estimator:

$$\begin{aligned}|\underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}(t)| = |\underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}(t) - (\hat{\underline{f}}^{L,L^*}(t) - \hat{\underline{f}}^{L_{\max},L^*}(t))| \\ + |\hat{\underline{f}}^{L,L^*}(t) - \hat{\underline{f}}^{L_{\max},L^*}(t)|.\end{aligned}$$

The first term is a centered Gaussian process, the second term is a quantity that we can estimate and thus bound. Let us give a bound for the first term, using Section **??**. For a given level $1-\alpha_2$, there exists a function $\hat{d}^{L,L_{\max}}()$ such that

$$P\left(\forall t \in [0,1], |\underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}(t) - (\hat{\underline{f}}^{L,L^*}(t) - \hat{\underline{f}}^{L_{\max},L^*}(t))| \leq \hat{d}^{L,L_{\max}}(t)\right) = 1-\alpha_2.$$

{eq-borne_dLLmax}

Let us denote $\hat{d}^{L,L_{\max}} = \sup_t |\hat{d}^{L,L_{\max}}(t)|$ and $\|\hat{\underline{f}}^{L,L^*} - \hat{\underline{f}}^{L_{\max},L^*}\|_\infty = \sup_t |\hat{\underline{f}}^{L,L^*}(t) - \hat{\underline{f}}^{L_{\max},L^*}(t)|$. Then we can deduce that with probability $1-\beta_2$

$$A_{L,L^*} \leq \hat{d}^{L,L_{\max}} + \|\hat{\underline{f}}^{L,L^*} - \hat{\underline{f}}^{L_{\max},L^*}\|_\infty.$$

Note that when $L \geq L^*$, $ A_{L,L^*} = 0$. We thus expect the bound $\hat{d}^{L,L_{\max}} + \|\hat{\underline{f}}^{L,L^*} - \hat{\underline{f}}^{L_{\max},L^*}\|_\infty$ to be small when $L \geq L^*$. But as L^* is unknown, we can not remove it from the bound. Est ce qu'on peut voir que cette borne devient quasiment nulle quand L est grand ? Y compris quand on n'est pas dans le vrai modele ?

ajouter ici illustration 5 : étude du terme d'approximation - à regarder

9.3 Confidence band of $\underline{f}^{L^*}$ for a given L

Using the two bounds of the bias and the approximation terms allows to define a bound M^L of $\hat{\underline{f}}^{L,L^*} - f^{L^*}$. We can prove the following result:

Theorem 9.1. *With high probability, we have*

$$\forall t \in [0,1], \quad |\hat{\underline{f}}^{L,L^*}(t) - \underline{f}^{L^*,L^*}| \leq \hat{M}^L$$

with

$$\hat{M}^L := \hat{d}^L + \hat{d}^{L,L_{\max}} + \|\hat{\underline{f}}^{L,L^*} - \hat{\underline{f}}^{L_{\max},L^*}\|_\infty$$

where $\hat{d}^L$ is defined by (??) with level $\alpha/2$ using the first sub-sample $\mathbf{y}^1$ and $\hat{d}^{L,L_{\max}}$ is defined by (??) with level $\alpha/2$ using the second sub-sample $\mathbf{y}^2$.

On serait vraiment capable de prouver ce théoreme ? sur le niveau de la proba ? Est ce qu'on peut prouver vraiment la proba ? Montrer que quand $L \geq L^*$, il n'y a plus que le terme b_L et donc la proba est $1-\alpha/2$ alors qu'avant c'est $(1-\alpha/2)^2 \approx 1-\alpha$?

This theorem provides a confidence band for $\underline{f}^{L^*,L^*}$ which is based on the estimator $\hat{\underline{f}}^{L,L^*}$:

$$\left[\hat{\underline{f}}^{L,L^*}(t) - \hat{M}^L; \hat{\underline{f}}^{L,L^*}(t) + \hat{M}^L\right]$$

for all $t \in [0,1]$.