Wavelet thresholding estimation in a Poissonian interactions model with application to genomic data

Laure Sansonnet

Wavelet thresholding estimation in a Poissonian interactions model with application to genomic data

Laure Sansonnet

Laboratoire de Mathématiques, CNRS UMR 8628
Université Paris-Sud 11
15 rue Georges Clémenceau, 91405 Orsay Cedex, France

2011

Abstract: This paper deals with the study of dependencies between two given events modeled by point processes. In particular, we focus on the context of DNA to detect favored or avoided distances between two given motifs along a genome suggesting possible interactions at a molecular level. For this, we naturally introduce a so-called reproduction function h that allows to quantify the favored positions of the motifs and which is considered as the intensity of a Poisson process. Our first interest is the estimation of this function h assumed to be well localized. The estimator $\tilde{h}$ based on random thresholds achieves an oracle inequality. Then, minimax properties of $\tilde{h}$ on Besov balls $\mathcal{B}^s_{2,\infty}(R)$ are established. Some simulations are provided, allowing the calibration of tuning parameters from a numerical point of view and proving the good practical behavior of our procedure. Finally, our method is applied to the analysis of the influence between gene occurrences along the *E. coli* genome and occurrences of a motif known to be part of the major promoter sites for this bacterium.

Keywords: Adaptive estimation, interactions model, oracle inequalities, Poisson process, thresholding rule, U-statistics, wavelets.

MSC2010: Primary 60G55, 62G05; secondary 62G20, 62G30.

1 Introduction

The goal of the present paper is to study the dependence between two given events modeled by point processes. We propose a general statistical approach to analyze any type of interaction, for instance, interactions between neurons in neurosciences or the comprehension of bankruptcies by contagion in economics. In particular, we focus on a model to study favored or avoided distances between patterns on a strand of DNA, which is an important task in genomics.

We are first interested in the modeling of the influence between two given motifs, a motif being defined as a sequence of letters in the alphabet `{a,c,g,t}`. This alphabet represents the four nucleotides bases of DNA: adenine, cytosine, guanine and thymine. Our aim is to model the dependence between motifs in order to identify favored or avoided distances between them, suggesting possible interactions at a molecular level. Because genomes are long (some 1 million bases) and motifs of interest are short (3 up to 20 bases), motif occurrences can be viewed as points along genomes. For convenience, we work in a continuous framework and then, the occurrences of a motif along a genome are modeled by a point process lying in the interval $[0;T]$, where T is the normalized length of the studied genome and will drive the asymptotic. We add that our model focuses on only one direction of interactions, that is to say we investigate the way a first given motif influences a second one. To study the influence of the second motif on the first one, we just invert their roles in the model.

We observe the occurrences of both given motifs (we presuppose interactions between them) and we assume that their distributions are as follows. The locations of the first motif are modeled by a n-sample of uniform random variables on $[0;T]$, denoted $U_1,\dots,U_n$ and named parents. As the parameter T, the number n of parents will also drive the asymptotic. Then, each U_i gives birth independently to a Poisson process N^i with intensity the function $t \longmapsto h(t-U_i)$ with respect to the Lebesgue measure on $\mathbb{R}$ (for instance, see [17]), which models the locations of the second motif. We consequently observe the aggregated process

$$N = \sum_{i=1}^{n} N^i \quad \text{with intensity the function} \quad t \longmapsto \sum_{i=1}^{n} h(t-U_i) \tag{1.1}$$

and the points of the process N are named children. But in this model, for any child we do not observe which parent gives birth to him. The unknown function h is so-called reproduction function. Our goal is then to estimate h with the observations of the U_i's and realizations of N.

Such a modeling of locations of the first motif is linked to the work on the distribution of words in DNA sequences of Schbath and coauthors (for instance, see [31], [24] and [29]). Indeed, the first motif of interest is a rare word and is modeled by a homogeneous Poisson process N^0 on $[0;T]$. Thus, conditionally to the event "the number of points falling into $[0;T]$ is n", the points of the process N^0 (i.e. the parents) obey the same law as a n-sample of uniform random variables on $[0;T]$. Moreover, with very high probability, n is proportional to T and this constitutes the asymptotic considered in genomics, to which we will refer as the "DNA case". With our model (considering a uniform law on the parents), we can also take into consideration the cases $n \ll T$ (parents are far away with respect to each other and one can almost identify which points are the children of a given parent) and $n \gg T$ (parents are too close to each other, which leads to hard statistical problems).

If $n = 1$, the purpose is to estimate the intensity of only one Poisson process. Many adaptive methods have been proposed to deal with Poisson intensity estimation. For instance, Rudemo [30] studied data-driven histogram and kernel estimates based on the cross-validation method. Donoho [8] fitted the universal thresholding procedure proposed by Donoho and Johnstone [9] by using the Anscombe's transform. Kolaczyk [18] refined this idea by investigating the tails of the distribution of the noisy wavelet coefficients of the intensity. By using model selection, other optimal estimators have been proposed by Reynaud-Bouret [25] or Willett and Nowak [32]. Reynaud-Bouret and Rivoirard [26] proposed a data-driven thresholding procedure that is near optimal under oracle and minimax points of view, with as few support assumptions as possible (the support of the intensity h may be unknown or not finite), unlike previous methods that need to assume that the intensity has a known bounded support.

We notice that the reproduction function h can be also viewed as the intensity of a Cox process (for instance, see [5]) where the covariates are the parents $U_1,\dots,U_n$. Comte *et al.* [4] proposed an original estimator of the conditional intensity of a Cox process (more generally, a marker-dependent counting process). Using model selection methods, they prove that their estimator satisfies an oracle inequality and has minimax properties. Note that we consider here point processes on the real line. Some aspects of similar spatial processes are studied in a parametric way [23], for instance.

Some work has been done to study the statistical dependence between motif occurrences. For instance, in Gusto and Schbath's article [12], the framework consists in modeling the occurrences of two motifs by a Hawkes process (see [13]): our framework can be viewed when the support of h is in $\mathbb{R}_+$ as a very particular case of theirs. Their method, called FADO, uses maximum likelihood estimates of the coefficients of h on a Spline basis coupled with an AIC criterion. However, even if the FADO procedure is quite effective and can manage interactions between two types of events, spontaneous apparition (a child can be an orphan) and self-excitation (a child can give birth to another child), there are several drawbacks. In fact, this procedure is a parametric estimation method coupled with a classical AIC criterion which behaves poorly for complex families of models. Moreover, FADO involves sparsity issues. Indeed, our feeling is that if interaction exists, say around the distance d bases, the function h to estimate should take large values around d and if there is no biological reason for any other interaction, then h should be null anywhere else. However, if the FADO estimate takes small

values in this last situation, it does not vanish, which can result in misleading biological interpretations (see [27]). Finally, in this model, the occurrences of the first motif only depend on the past occurrences of first and second motifs. Reynaud-Bouret and Schbath [27] have proposed an alternative based on model selection principle for Hawkes processes that solves the sparsity problem. Their estimate satisfies an oracle inequality and has adaptive minimax properties with respect to certain classes of functions. But it manages only one motif whereas it is necessary to treat interaction with another type of events and the method has a high computational cost even for a small number of models. Note that Hawkes processes have a long story of parametric inference (see [22], [21] and [5]). In particular, for genomic data, Carstensen *et al.* [2] recently deal with multivariate Hawkes process models in order to model the joint occurrences of multiple transcriptional regulatory elements (TREs) along the genome that are capable of providing new insights into dependencies among elements involved in transcriptional regulation.

In this paper, the proposed model is simple. Each child comes from one parent (no orphan and no child who is a parent), that is to say we do not take into account the phenomenons of spontaneous apparition and self-excitation, contrary to Hawkes process models. But it brings novelties. To estimate the reproduction function h, we propose a nonparametric method, using a wavelet thresholding rule that will compensate sparsity issues of the FADO method. Furthermore, our model treats interaction between two types of events, with a possible influence of the past occurrences but also future occurrences. Then, there is the presence of a double asymptotic: the normalized length of the studied genome T and the number n of parents, which is not usual. In the biological context, it is not acceptable assuming to know each child's parent. Our model, via the reproduction function h, allows to quantify the favored locations of children in relation to their parent, even if one cannot attribute a child to a parent before the statistical inference. First we provide in this paper theoretical results and we derive oracle inequalities and minimax rates showing that our method achieves good theoretical performances. The proofs of these results are essentially based on concentration inequalities and on exponential and moment inequalities for U-statistics (see [6], [11] and [14]). Secondly some simulations are carried out to validate our procedure and an application on real data (*Escherichia coli* genome) is proposed. The procedure provides satisfying reconstructions, overcomes the problems raised by the FADO method and agrees with the knowledge of the considered biological mechanism. For these numerical aspects, we have used a low computational complexity cascade algorithm.

In Section 2, we define the notations and we describe the method. Then Section 2 discusses the properties of our procedure for the oracle and minimax approaches. Section 3 is devoted to the implementation of our method and provides simulations. The cascade algorithm is presented in Section 3.1. Section 4 presents the application on the complete *Escherichia coli* genome. A more technical result that is at the origin of the one stated in Section 2.3 and proofs can be found in Section 6 (Appendix).

2 General results

2.1 Notations

To estimate the reproduction function, we assume that h belongs to $\mathbb{L}_1(\mathbb{R})$ and $\mathbb{L}_\infty(\mathbb{R})$. Consequently, we can consider the decomposition of h on a particular biorthogonal wavelet basis, built by Cohen *et al.* [3], that we can describe as follows. We set $\phi = \mathbf{1}_{[0,1]}$ the analysis father wavelet. For any $r > 0$, there exist three functions ψ, $\tilde{\phi}$ and $\tilde{\psi}$ with the following properties:

- $\tilde{\phi}$ and $\tilde{\psi}$ are compactly supported,
- $\tilde{\phi}$ and $\tilde{\psi}$ belong to C^{r+1}, where C^{r+1} denotes the Hölder space of order $r+1$,
- ψ is compactly supported and is a piecewise constant function,
- ψ is orthogonal to polynomials of degree no larger than r,

- $\left\{(\phi_k,\psi_{j,k})_{j\geqslant 0,k\in\mathbb{Z}},(\tilde{\phi}_k,\tilde{\psi}_{j,k})_{j\geqslant 0,k\in\mathbb{Z}}\right\}$ is a biorthogonal family: for any $j,j'\geqslant 0$, for any $k,k'\in\mathbb{Z}$,

$$\int_{\mathbb{R}}\phi_k(x)\tilde{\psi}_{j',k'}(x)\,dx=\int_{\mathbb{R}}\psi_{j,k}(x)\tilde{\phi}_{k'}(x)\,dx=0,$$

$$\int_{\mathbb{R}}\phi_k(x)\tilde{\phi}_{k'}(x)\,dx=\mathbf{1}_{\{k=k'\}},\quad \int_{\mathbb{R}}\psi_{j,k}(x)\tilde{\psi}_{j',k'}(x)\,dx=\mathbf{1}_{\{j=j',k=k'\}},$$

where for any $x\in\mathbb{R}$,

$$\phi_k(x)=\phi(x-k),\quad \psi_{j,k}(x)=2^{j/2}\psi(2^jx-k)$$

and

$$\tilde{\phi}_k(x)=\tilde{\phi}(x-k),\quad \tilde{\psi}_{j,k}(x)=2^{j/2}\tilde{\psi}(2^jx-k).$$

On the one hand, decomposition wavelets ϕ_k and $\psi_{j,k}$ are piecewise constant functions and, on the other hand, reconstruction wavelets $\tilde{\phi}_k$ and $\tilde{\psi}_{j,k}$ are smooth functions. This implies the following wavelet decomposition of $h\in\mathbb{L}_2(\mathbb{R})$:

$$h=\sum_{k\in\mathbb{Z}}\alpha_k\tilde{\phi}_k+\sum_{j\geqslant 0}\sum_{k\in\mathbb{Z}}\beta_{j,k}\tilde{\psi}_{j,k},\tag{2.1}$$

where for any $j\geqslant 0$ and any $k\in\mathbb{Z}$,

$$\alpha_k=\int_{\mathbb{R}}h(x)\phi_k(x)\,dx,\quad \beta_{j,k}=\int_{\mathbb{R}}h(x)\psi_{j,k}(x)\,dx.$$

The Haar basis, used in practice, can be viewed as a particular biorthogonal wavelet basis, by setting $\tilde{\phi}=\phi$ and $\tilde{\psi}=\psi=\mathbf{1}_{]\frac{1}{2};1]}-\mathbf{1}_{[0;\frac{1}{2}]}$, with $r=0$ (even if the second property is not satisfied with such a choice). The Haar basis is an orthonormal basis, which is not true for general biorthogonal wavelet bases. This kind of decomposition has already been used in thresholding methods by Juditsky and Lambert-Lacroix [16], Reynaud-Bouret and Rivoirard [26], and Reynaud-Bouret *et al.* [28].

To shorten mathematical expressions, we set

$$\Lambda=\{\lambda=(j,k):j\geqslant -1,k\in\mathbb{Z}\},$$

and for any $\lambda\in\Lambda$,

$$\varphi_\lambda=\begin{cases}\phi_k & \text{if } \lambda=(-1,k)\\ \psi_{j,k} & \text{if } \lambda=(j,k) \text{ with } j\geqslant 0\end{cases},\quad \tilde{\varphi}_\lambda=\begin{cases}\tilde{\phi}_k & \text{if } \lambda=(-1,k)\\ \tilde{\psi}_{j,k} & \text{if } \lambda=(j,k) \text{ with } j\geqslant 0\end{cases}$$

and similarly

$$\beta_\lambda=\begin{cases}\alpha_k & \text{if } \lambda=(-1,k)\\ \beta_{j,k} & \text{if } \lambda=(j,k) \text{ with } j\geqslant 0\end{cases}.$$

Then (2.1) can be rewritten as

$$h=\sum_{\lambda\in\Lambda}\beta_\lambda\tilde{\varphi}_\lambda\quad\text{with}\quad \beta_\lambda=\int_{\mathbb{R}}h(x)\varphi_\lambda(x)\,dx\tag{2.2}$$

and now, we have to estimate these wavelet coefficients.

For all λ in Λ, we define $\hat{\beta}_\lambda$ an estimator of β_λ as

$$\hat{\beta}_\lambda=\frac{G(\varphi_\lambda)}{n},\quad\text{with}\quad G(\varphi_\lambda)=\int_{\mathbb{R}}\sum_{i=1}^{n}\left[\varphi_\lambda(t-U_i)-\frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]dN_t,\tag{2.3}$$

where π is the uniform distribution on $[0;T]$ and $\mathbb{E}_\pi(\varphi_\lambda(t-U))$ denotes the expectation of $\varphi_\lambda(t-U)$ where $U\sim\pi$ (an independent copy of $U_1,\ldots,U_n$). If $n=1$, we obtain the natural estimators of the β_λ's in the case of only one Poisson process on the real line (see [26]).

Lemma 2.1. *For all $\lambda = (j,k)$ in Λ,*

$$\mathbb{E}(G(\varphi_\lambda)) = n \int_{\mathbb{R}} \varphi_\lambda(x) h(x)\, dx,$$

i.e. $\hat{\beta}_\lambda$ is an unbiased estimator for β_λ. Furthermore, its variance is upper bounded as follows:

$$\mathrm{Var}(\hat{\beta}_\lambda) \leqslant C \left\{ \frac{1}{n} + \frac{1}{T} + \frac{2^{-j} n}{T^2} \right\}$$

and

$$\sup_{\lambda \in \Lambda} \mathrm{Var}(\hat{\beta}_\lambda) \leqslant C' \left\{ \frac{1}{n} + \frac{n}{T^2} \right\},$$

where C and C' depend on $\|h\|_1$, $\|h\|_\infty$, $\|\psi\|_1$ and $\|\psi\|_2$.

The behavior of the variance of the $\hat{\beta}_\lambda$'s is not usual, because two parameters n and T are involved. Nevertheless, when n is proportional to T ("DNA case" as explained in Introduction), the variance is bounded by $1/T$ up to a constant, as for the Hawkes process (see [27]). When $n \ll T$, the variance is bounded by $1/n$ up to a constant, which means that the apparition's distance between two parents is large enough to make their interactions insignificant for the statistical analysis. So in this case, our framework can be viewed as the observation of a n-sample of a Poisson process with common intensity h (see [26]). Finally, when $n \gg T$, the variance deteriorates and is only bounded by n/T^2 up to a constant, and in this case, the small apparition's distance between two parents leads to rough statistical issues hard to overcome.

2.2 Description of our method

We start assuming that h is compactly supported in $[-A; A]$, with A a positive real number. This quantity A can denote the maximal memory along DNA sequences (this is chosen by the biologists (see [12]), depending on the underlying biological process they have in mind). Furthermore, the properties of the biorthogonal wavelet bases introduced previously allow us to assume that we know a positive real number M such that the support of ψ is contained in $[-M; M]$.

First, we introduce the following deterministic subset Γ of Λ

$$\Gamma = \big\{ \lambda = (j,k) \in \Lambda : -1 \leqslant j \leqslant j_0, k \in \mathcal{K}_j \big\},$$

where j_0 a positive integer that will be fixed later and at each resolution level j, we denote $\mathcal{K}_j$ the set of integers such that the intersection of the supports of φ_λ and h is not empty, with $\lambda = (j,k)$. Straightforward computations lead to a cardinal of Γ of order 2^{j_0}.

Then, given some parameter $\gamma > 0$, we define for any $\lambda \in \Gamma$, the threshold

$$\eta_\lambda(\gamma, \Delta) = \sqrt{2\gamma j_0 \widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)} + \frac{\gamma j_0}{3} B\left(\frac{\varphi_\lambda}{n}\right) + \Delta \frac{N_{\mathbb{R}}}{n} \tag{2.4}$$

where Δ is a positive quantity and $N_{\mathbb{R}}$ is the number of points of the aggregated process N lying in $\mathbb{R}$. For theoretical results, Δ will be taken of order $\frac{j_0^2 2^{j_0/2}}{n} + \frac{j_0}{\sqrt{T}} + \frac{\sqrt{j_0 n}}{T}$ times a constant depending on γ, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$. In (2.4), we set

$$B\left(\frac{\varphi_\lambda}{n}\right) = \frac{1}{n} B(\varphi_\lambda) = \frac{1}{n} \left\| \sum_{i=1}^{n} \left[\varphi_\lambda(\cdot - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(\cdot - U)) \right] \right\|_\infty \tag{2.5}$$

and

$$\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right) = \frac{1}{n^2} \widetilde{V}(\varphi_\lambda) = \frac{1}{n^2} \left(\hat{V}(\varphi_\lambda) + \sqrt{2\gamma j_0 \hat{V}(\varphi_\lambda) B^2(\varphi_\lambda)} + 3\gamma j_0 B^2(\varphi_\lambda) \right) \tag{2.6}$$

where

$$\hat{V}(\varphi_\lambda) = \int_{\mathbb{R}} \left(\sum_{i=1}^{n} \left[\varphi_\lambda(t - U_i) - \frac{n-1}{n} \mathbb{E}_\pi(\varphi_\lambda(t-U)) \right] \right)^2 dN_t. \tag{2.7}$$

Since they only depend on the observations, the numerical values of $B(\varphi_\lambda)$, $\hat{V}(\varphi_\lambda)$ and so $\widetilde{V}(\varphi_\lambda)$ defined respectively by (2.5), (2.7) and (2.6) can be exactly computed.

We denote $\tilde{\beta}$ the estimator of $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$ associated with the previous thresholding rule:

$$\tilde{\beta} = \left(\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geqslant \eta_\lambda(\gamma,\Delta)} \mathbf{1}_{\lambda \in \Gamma} \right)_{\lambda \in \Lambda} \tag{2.8}$$

and finally, we set

$$\tilde{h} = \sum_{\lambda \in \Lambda} \tilde{\beta}_\lambda \tilde{\varphi}_\lambda \tag{2.9}$$

an estimator of h that only depends on the choice of (γ, Δ) and j_0 fixed later.

Thresholding procedures have been introduced by Donoho and Johnstone [9]. They derive from the sufficiency to keep a small amount of the coefficients to have a good estimation of the function h. The threshold $\eta_\lambda(\gamma,\Delta)$ seems to be defined in a rather complicated manner but the first term: $\sqrt{2\gamma j_0 \widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)}$ looks like the universal threshold proposed by [9] in the Gaussian regression framework, by choosing γ close to 1 and j_0 of order $\log n$. The universal threshold of [9] is defined by $\eta_\lambda = \sqrt{2\sigma^2 \log n}$, where σ^2 (assumed to be known) is the variance of each noisy wavelet coefficient. In our setting, $\mathrm{Var}(\hat{\beta}_\lambda)$ depends on h, so it is (over)estimated by $\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)$. The other terms of the threshold (2.4) are unavoidable remaining terms which allow to obtain sharp concentration inequalities.

2.3 Main result and discussions

Our main result is an oracle one. Given a collection of procedures (for example, penalization, projection or thresholding), the oracle represents the ideal "estimator" among the collection. In our setting the oracle gives, for our thresholding rule, the coefficients that have to be kept. In our framework (see [9] and [26]), the "oracle estimator" is

$$\bar{h} = \sum_{\lambda \in \Gamma} \bar{\beta}_\lambda \tilde{\varphi}_\lambda, \quad \text{with} \quad \bar{\beta}_\lambda = \hat{\beta}_\lambda \mathbf{1}_{\mathrm{Var}(\hat{\beta}_\lambda) < \beta_\lambda^2}.$$

This "estimator" is not a true estimator, of course, since it depends on h. The approach of optimal adaptation is to derive true estimators which achieve the same performance as the "oracle estimator". Our goal is now to compare the risk of $\tilde{h}$ defined in Section 2.2 to the oracle risk:

$$\mathbb{E}\left(\|\bar{h} - h\|_2^2\right) = \sum_{\lambda \in \Gamma} \mathbb{E}\left[(\hat{\beta}_\lambda \mathbf{1}_{\mathrm{Var}(\hat{\beta}_\lambda) < \beta_\lambda^2} - \beta_\lambda)^2\right] + \sum_{\lambda \notin \Gamma} \beta_\lambda^2 = \sum_{\lambda \in \Gamma} \min(\mathrm{Var}(\hat{\beta}_\lambda), \beta_\lambda^2) + \sum_{\lambda \notin \Gamma} \beta_\lambda^2.$$

Theorem 1. *We assume that $n \geqslant 2$, $j_0 \in \mathbb{N}^*$ such that $2^{j_0} \leqslant n < 2^{j_0+1}$, $\gamma > 2\log 2$ and Δ is defined in the Appendix by (6.15) and (6.16). Then the estimator $\tilde{h}$ defined in Section 2.2 satisfies*

$$\mathbb{E}\left(\|\tilde{h} - h\|_2^2\right) \leqslant C_1 \inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + \left[(\log n)^4 \times \frac{1}{n} + (\log n)^2 \times \frac{n}{T^2} \right] |m| \right\} + C_2 \left[\frac{1}{n} + \frac{n}{T^2} \right],$$

where $|m|$ is the cardinal of the set m, C_1 is a positive constant depending on γ, $\|h\|_1$, $\|h\|_\infty$, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$ and C_2 is a positive constant depending on the compact support of h, $\|h\|_1$, $\|h\|_\infty$, the compact support of ψ, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$.

As the expression between brackets is of the same order as the upper bound of $\mathrm{Var}(\hat{\beta}_\lambda)$ established in Lemma 2.1 (up to a logarithmic term), the oracle type inequality of Theorem 1 proves that the estimator $\tilde{h}$ achieves satisfying theoretical properties.

In particular, if we apply Theorem 1 with n proportional to T ("DNA case"), then the estimator $\tilde{h}$ defined in Section 2.2 satisfies

$$\mathbb{E}\left(\|\tilde{h}-h\|_2^2\right) \leqslant C_1 \inf_{m\subset\Gamma}\left\{\sum_{\lambda\notin m}\beta_\lambda^2 + \frac{(\log T)^4}{T}|m|\right\} + \frac{C_2}{T}.$$

This oracle type inequality is similar to the one obtained by Theorem 1 of [27] where the Hawkes process is considered. Since n is proportional to T, this inequality is typical of classical oracle inequalities obtained in model selection (for example, see Theorem 2.1 of [26] where only one Poisson process on the real line is considered or more generally, see [20] for density estimation).

Then, we establish a minimax result on Besov balls still with n is proportional to T. For any $R>0$ and $s\in\mathbb{R}$ such that $0<s<r+1$ (where $r>0$ denotes the wavelet smoothness parameter introduced in the description of the biorthogonal wavelet bases at the beginning of the current section), we consider the following Besov ball of radius R:

$$\mathcal{B}^s_{2,\infty}(R) = \left\{ f\in\mathbb{L}_2(\mathbb{R}) : f=\sum_{\lambda\in\Lambda}\beta_\lambda\tilde{\varphi}_\lambda, \forall j\geqslant -1, \sum_{k\in\mathcal{K}_j}\beta^2_{(j,k)}\leqslant R^2 2^{-2js}\right\}.$$

Now, let us state the upper bound of the risk of $\tilde{h}$ when h belongs to $\mathcal{B}^s_{2,\infty}(R)$.

Corollary 2.1. *Let $R>0$ and $s\in\mathbb{R}$ such that $0<s<r+1$. Assume that $h\in\mathcal{B}^s_{2,\infty}(R)$ and n is proportional to T. Then the estimator $\tilde{h}$ defined in Section 2.2 satisfies*

$$\mathbb{E}\left(\|\tilde{h}-h\|_2^2\right) \leqslant C\left(\frac{(\log T)^4}{T}\right)^{\frac{2s}{2s+1}},$$

where C is a positive constant depending on γ, the compact support of h, $\|h\|_1$, $\|h\|_\infty$, the compact support of ψ, $\|\psi\|_1$, $\|\psi\|_2$, $\|\psi\|_\infty$ and R.

The rate of the risk of $\tilde{h}$ corresponds to the minimax rate, up to the logarithmic term, for estimation of a compactly supported intensity of a Poisson process (see [25]) or for a compactly supported density when we have n i.i.d. observations (see [10]). One more time this illustrates the optimality of the procedure $\tilde{h}$ but in the minimax setting.

3 Implementation procedure

From now on we consider the context of DNA, i.e. n is proportional to T. As mentioned in Introduction, we can assume that the parents are the points of a homogeneous Poisson process N^0 on $[0;T]$ with constant intensity μ which allows to write $n\simeq\mu T$.

In this section, we specify a procedure for the computation of the family of random thresholds $(\eta_\lambda(\gamma,\Delta))_{\lambda\in\Gamma}$ to reconstruct the reproduction function h. We also provide some simulations in order to calibrate parameters from a numerical point of view and to show the robustness of our procedure.

3.1 Algorithm

We only focus on the Haar basis where

$$\phi=\tilde{\phi}=\mathbf{1}_{[0;1]} \quad\text{and}\quad \psi=\tilde{\psi}=\mathbf{1}_{]\frac{1}{2};1]}-\mathbf{1}_{[0;\frac{1}{2}]},$$

because the expression of the functions associated to this basis, that are piecewise constant functions, allows to implement simple and fast algorithms. Furthermore, considering this kind of functions is suitable for our genomic setting. In fact, according to biological studies, the reproduction function h is expected to be very irregular, with large null ranges and sudden changes at specific distances. We recall that h is assumed to be compactly supported in $[-A;A]$, with A a positive integer in practice.

We consider the thresholding rule $\tilde{h}$ defined in Section 2.2 with

$$\Gamma = \big\{\lambda = (j,k) \in \Lambda : -1 \leqslant j \leqslant j_0, k \in \mathcal{K}_j\big\},$$

and

$$\eta_\lambda(\gamma,\delta) = \sqrt{2\gamma j_0 \hat{V}\left(\frac{\varphi_\lambda}{n}\right)} + \frac{\gamma j_0}{3} B\left(\frac{\varphi_\lambda}{n}\right) + \frac{\delta}{\sqrt{T}}\frac{N_{\mathbb{R}}}{n}.$$

Observe that $\eta_\lambda(\gamma,\delta)$ slightly differs from the threshold defined in (2.4) since the parameter Δ is replaced with $\frac{\delta}{\sqrt{T}}$ (thanks to the definition (6.15) of Δ) and $\widetilde{V}(\varphi_\lambda)$ is now replaced with $\hat{V}(\varphi_\lambda)$ (there is no major difference in our simulations). The ideal choice (from a theoretical point of view) of the maximal resolution level j_0 is given by Theorem 1, that is to say j_0 is the positive integer such that $2^{j_0} \leqslant n < 2^{j_0+1}$. But we will fix $j_0 = 5$ in the sequel (in particular, to limit the computation time). The choice of the parameters γ and δ is discussed in the next subsection.

A key point of the algorithm is the computation of the quantity

$$S(\varphi_\lambda)(t) = \sum_{i=1}^{n}\left[\varphi_\lambda(t-U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right], \quad \text{for all } t \in \mathbb{R}$$

that appears in $\hat{\beta}_\lambda$, $B(\varphi_\lambda)$ and $\hat{V}(\varphi_\lambda)$. We decompose it into two parts: a random "piecewise constant" part $S_r(\varphi_\lambda) = \sum_{i=1}^{n}\varphi_\lambda(\cdot - U_i)$ and a deterministic (piecewise affine) part $(n-1)\mathbb{E}_\pi(\varphi_\lambda(t-U))$. Note that the deterministic part can easily be implemented with a low computational cost. This is not the case of the random "piecewise constant" part for which we have constructed a cascade algorithm, inspired by the pioneering work of Mallat [19]. To explain this algorithm in few words, we use the following notations: for any $j \geqslant 0$, for any $k \in \mathbb{Z}$, for any $x \in \mathbb{R}$,

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),$$

where $\phi_{j,k}$ are father wavelets and $\psi_{j,k}$ mother wavelets. We have the following relationships between wavelets at level j and wavelets at level $(j+1)$:

$$\psi_{j,k} = \frac{\sqrt{2}}{2}\big(\phi_{j+1,2k+1} - \phi_{j+1,2k}\big) \quad \text{and} \quad \phi_{j,k} = \frac{\sqrt{2}}{2}\big(\phi_{j+1,2k} + \phi_{j+1,2k+1}\big). \tag{3.1}$$

We notice that only mother wavelets and the father wavelet of level $j = 0$ (that corresponds to φ_λ with $\lambda = (-1,k)$) are used to reconstruct the signal. The cascade algorithm is implemented as follows.

1. Compute $S_r(\phi_{j_0,0})$. Since $S_r(\phi_{j_0,0})$ is a piecewise constant function, this computation gives a partition and the values of $S_r(\phi_{j_0,0})$ on the intervals of the partition.

2. Shift by $+2^{-j_0}k$ the intervals of the previous partition by keeping the same values on the partition to obtain $S_r(\phi_{j_0,k})$ for any integer k in $[-2^{j_0}A; 2^{j_0}A - 1]$.

3. For any resolution level j going from $(j_0 - 1)$ to 0, in a decreasing way, compute $S_r(\psi_{j,k})$ and $S_r(\phi_{j,k})$ with expressions (3.1). The quantities $S_r(\psi_{j,k})$ allow the reconstruction of the signal and the quantities $S_r(\phi_{j,k})$ are transitional and will be used for the computations of the lower resolution level $(j-1)$.

4. Also keep $S_r(\phi_{0,k})$ because it is used for the reconstruction of the signal.

Now, let us define our thresholding estimate of h for a practical purpose.

Step 0 Let $j_0 = 5$ and choose also positive constants γ and δ.

Step 1 Set $\Gamma = \big\{\lambda = (j,k) \in \Lambda : -1 \leqslant j \leqslant j_0, k \in \mathcal{K}_j\big\}$ and compute for any λ in Γ, $S(\varphi_\lambda)(X)$ for all points X of the process N. In the same way, also compute the coefficients $\hat{\beta}_\lambda$, $B(\varphi_\lambda)$ and $\hat{V}(\varphi_\lambda)$.

Step 2 Threshold the coefficients by setting $\tilde{\beta}_\lambda = \hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geqslant \eta_\lambda(\gamma,\delta)}$ according to the following threshold choice:

$$\eta_\lambda(\gamma,\delta) = \sqrt{2\gamma j_0 \hat{V}\left(\frac{\varphi_\lambda}{n}\right)} + \frac{\gamma j_0}{3} B\left(\frac{\varphi_\lambda}{n}\right) + \frac{\delta}{\sqrt{T}} \frac{N_{\mathbb{R}}}{n}.$$

Step 3 Reconstruct the function h by using the $\tilde{\beta}_\lambda$'s and denote

$$\tilde{h} = \sum_{\lambda \in \Lambda} \tilde{\beta}_\lambda \tilde{\varphi}_\lambda.$$

3.2 Experiments on simulated data

The programs have been coded in `Scilab 5.2` and are available upon request.

3.2.1 Choice of parameters

Now, we deal with the choice of the parameters γ and δ in our procedure from a practical point of view. The question is: how to choose the optimal parameters? We work with two testing functions denoted 'Signal1' and 'Signal2' whose definitions are given in the following table:

'Signal1'	'Signal2'
$\nu \times \mathbf{1}_{[0;1]}$	$\nu \times \frac{8}{3}\left(\mathbf{1}_{[0.5;0.625]} + \mathbf{1}_{[1;1.25]}\right)$

with ν, the children's intensity, set to 4. We fix willfully $A = 10$. Such a choice of A (remember that $[-A;A]$ is the support of h) assumes that we do not know the support of functions. We recall that $j_0 = 5$.

Given T, μ the parents' intensity and a testing function, we denote $R(\gamma,\delta)$ the quadratic risk of our procedure $\tilde{h}$ (depending on (γ,δ)) defined in Section 3.1. Of course, we aim at finding values of (γ,δ) such that this quadratic risk is minimal. The average over 100 simulations of $R(\gamma,\delta)$ is computed providing an estimation of $\mathbb{E}(R(\gamma,\delta))$. This average risk, denoted $\bar{R}(\gamma,\delta)$ and viewed as a function of the parameters (γ,δ), is plotted for $(T,\mu) \in \{(10000,0.1),(2000,0.1),(2000,0.5)\}$ and for the two signals considered previously: 'Signal1' and 'Signal2'.

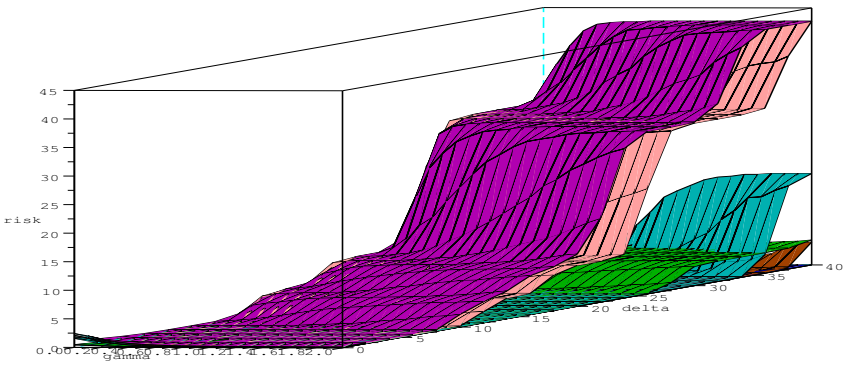

Figure 1: *The function* $(\gamma,\delta) \mapsto \bar{R}(\gamma,\delta)$ *for 'Signal1' and 'Signal2' for different values of* T *and* μ*: 'Signal1' in* ■ *and 'Signal2' in* ■ *with* $(T,\mu) = (10000,0.1)$*; 'Signal1' in* ■ *and 'Signal2' in* ■ *with* $(T,\mu) = (2000,0.1)$*; 'Signal1' in* ■ *and 'Signal2' in* ■ *with* $(T,\mu) = (2000,0.5)$.

Figure 1 displays $\bar{R}$ for 'Signal1' and 'Signal2' decomposed on the Haar basis. This figure allows to draw the following conclusion: for any $(T, \mu) \in \{(10000, 0.1), (2000, 0.1), (2000, 0.5)\}$ and for 'Signal1' or 'Signal2',

$$\bar{R}(\gamma, \delta) \approx 0$$

for many values of (γ, δ). So, we observe a kind of "plateau phenomenon".

Reconstructions of the intensities of 'Signal1' and 'Signal2' are respectively given in Figure 2 and Figure 3 with the choice $(\gamma, \delta) = (0.18, 2.4)$, a common value of several plateaus. Note the good performance of our thresholding rule, in particular for $T = 10000$ and $\mu = 0.1$ (we have $\mu T = 1000$ parents and $\mu\nu T = 4000$ children in average), which corresponds to the real case treated in Section 4. Thus, we propose to take systematically $(\gamma, \delta) = (0.18, 2.4)$ in our procedure $\tilde{h}$ defined in Section 3.1.

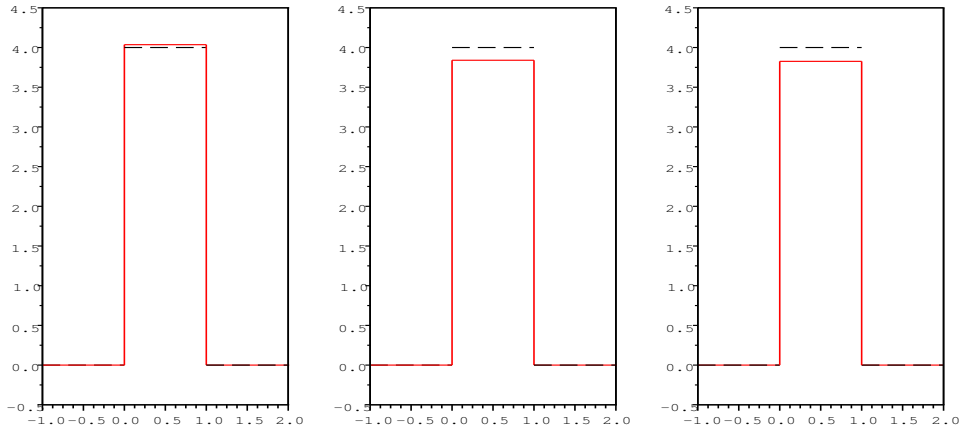

Figure 2: *Reconstructions of 'Signal1' (true: dotted line, estimate: solid line): left:* $(T, \mu) = (10000, 0.1)$*; middle:* $(T, \mu) = (2000, 0.1)$*; right:* $(T, \mu) = (2000, 0.5)$.

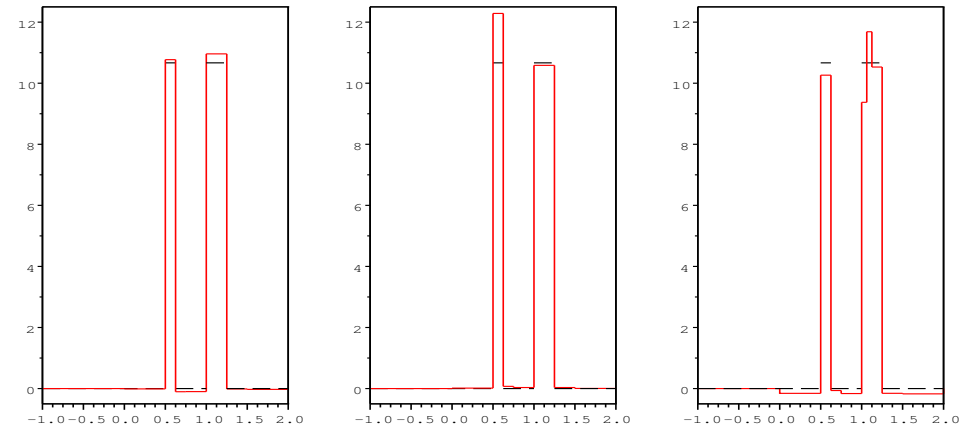

Figure 3: *Reconstructions of 'Signal2' (true: dotted line, estimate: solid line): left:* $(T, \mu) = (10000, 0.1)$*; middle:* $(T, \mu) = (2000, 0.1)$*; right:* $(T, \mu) = (2000, 0.5)$.

3.2.2 About the support of h

We are interested in the robustness of our procedure with respect to the support issue from a numerical point of view. What happens if we are wrong about the support of the function that we want to estimate? For instance, we consider the testing function denoted 'Signal3' whose definition is given in the following table:

'Signal3'
$\nu \times \frac{1}{4}\left(\mathbf{1}_{[-0.75;-0.5]} + \mathbf{1}_{[4.25;8]}\right)$

with ν, the children's intensity, set to 4.

Figure 4 displays reconstructions of 'Signal3' with different supports of h: $[-A; A]$, with $A \in \{1, 5, 10\}$. This figure shows that when we take a not large enough support ($A = 1$ or 5), we do not make large errors of approximation on $[-A; A]$. So, the procedure seems to take into account what happens beyond the chosen support. And for $A = 10$, we have a good complete reconstruction of 'Signal3'.

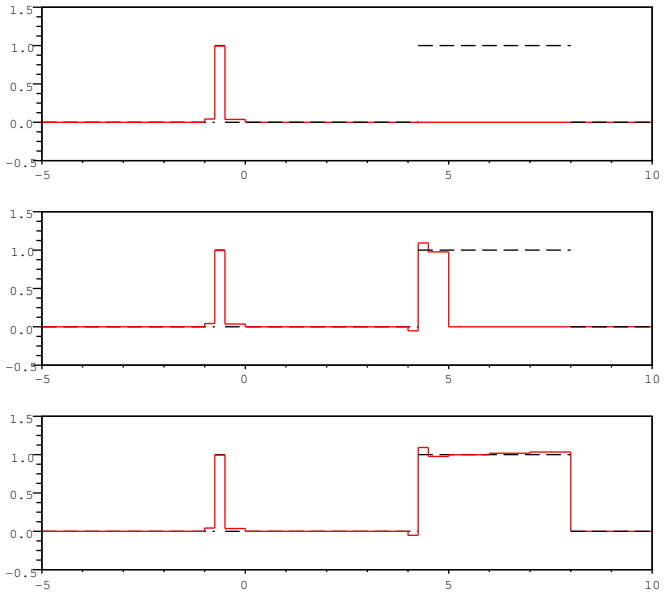

Figure 4: *Reconstructions of 'Signal3' (true: dotted line, estimate: solid line) with different supports: top:* $A = 1$*; middle:* $A = 5$*; bottom:* $A = 10$.

Finally, even if the support of the reproduction function is unknown, our method estimates correctly the signal on the chosen support, which explains the robustness of our procedure with respect to the support issue.

3.2.3 The case of spontaneous apparition

Here, we investigate the case of spontaneous apparition. Even if our model does not take into account the spontaneous apparition (i.e. children can not be orphans), we are interested by the performance of our procedure if there is a presence of orphans. On the one hand, let us give two processes: a process of intensity 'Signal1' with $\nu = 3$, $T = 10000$ and $\mu = 0.1$, to which is added a homogeneous Poisson process on $[0; T + 1]$ with intensity $\mu(4 - \nu) = 0.1$ (the orphans are viewed as a Poissonian noise). Thus, we have in average 1000 parents, 3000 children having a parent and 1000 children being orphans. On the other hand, let us give two other processes: a process of intensity 'Signal1' with $\nu = 1$ this

time, $T = 10000$ and $\mu = 0.1$, to which is added a homogeneous Poisson process on $[0; T + 1]$ with intensity $\mu(4 - \nu) = 0.3$. Thus, we have in average 1000 parents, 1000 children having a parent and 3000 children being orphans.

Reconstructions of 'Signal1' with $\nu = 3$ and $\nu = 1$ are given in Figure 5. When there is a small proportion of children being orphans, the reconstruction is still acceptable; the procedure can manage few orphans. But, when there are too many orphans, our procedure makes approximation errors, which are due to the fact that our model consists in associating any child with a parent.

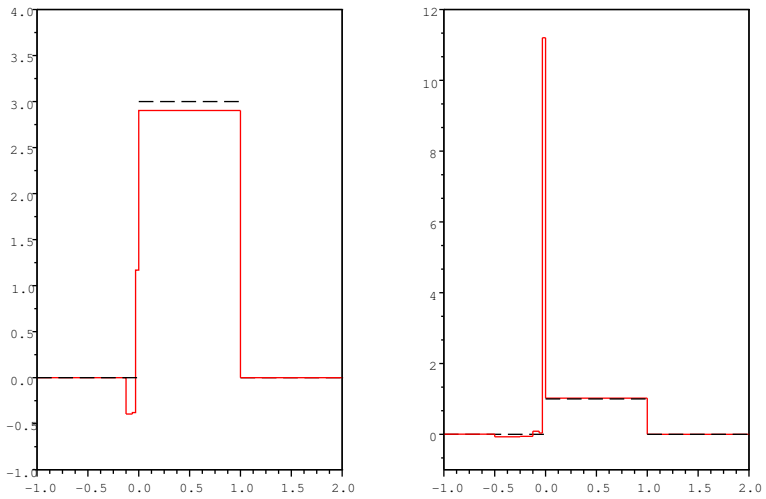

Figure 5: *Reconstructions of 'Signal1' (true: dotted line, estimate: solid line) with different values of ν: left: $\nu = 3$; right: $\nu = 1$.*

We mention that the case of spontaneous apparition is only numerical. For a more precise study of this phenomenon, we should extend our model by adding a positive constant to the intensity function $t \longmapsto \sum_{i=1}^{n} h(t - U_i)$, that would represent the orphans. This is outside the scope of this paper.

4 Applications to genomic data

As application, we are interested in the *Escherichia coli* genome. *E. coli* is an intestinal bacterium in mammals and very common in humans which is widely studied and used in genetics. More precisely, we are interested in the study of the dependence between promoter sites and genes along the complete genome of the bacterium. In particular, promoters are usually structured motifs located before the genes and not too far from them. Here, we have considered the major promoter of the bacterium *E. coli* and more precisely the word `tataat`. Most of the genes of *E. coli* should be preceded by this word at a very short distance apart. In order to validate our thresholding estimation procedure (proposed at Section 3), we hope to detect short favored distances between genes and previous occurrences of `tataat`.

For this, as in [12] we have analyzed the sequence composed of both strands of *E. coli* genome (4639221 bases); each strand being separated by 10000 artificial bases to avoid artificial dependencies between occurrences on one strand and occurrences on the other strand; we took 10000 bases for the maximal memory. It then represents a sequence of length 9288442; there are 4290 genes (we took the positions of the first base of coding sequences) and 1036 occurrences of `tataat`. For convenience, we set $T = 9289$ and so $A = 10$ (we work on a scale of $1 : 1000$). We recall that we have fixed $j_0 = 5$ and taken $(\gamma, \delta) = (0.18, 2.4)$.

First, we investigate the way the DNA motif `tataat` influences genes and so, in our model, the parents are the occurrences of `tataat` and children are the occurrences of genes. To give general insight on h, Figure 6 gives the estimator $\tilde{h}$ defined in Section 3.1 without Step 2 (no thresholding), i.e. we have kept all the estimated coefficients. We observe a peak around 0 which corresponds to what we thought about the fact that most of the genes of *E. coli* should be preceded by the word `tataat` at a very short distance apart. We also observe other peaks, for instance around 1200 bases. The biological significance of these peaks remains an open question.

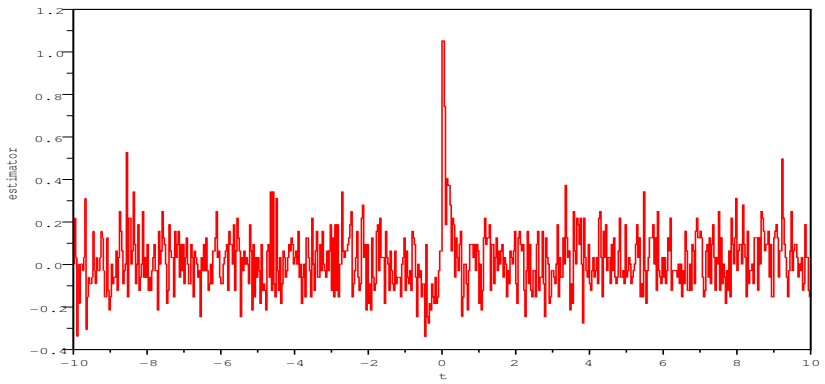

Figure 6: *Estimator, no thresholding, for E. coli data at the scale* $1:1000$ *(i.e. 1 corresponds to 1000 bases), with parents=*`tataat` *and children=genes.*

We apply the complete procedure proposed in Section 3.1 (with thresholding) and we obtain Figure 7. The shape of this estimator explains how occurrences of genes are influenced by occurrences of `tataat`. We can draw following conclusions, that coincide with the ones we could expect:

- The estimator $\tilde{h}(t) = 0$ if $t \leqslant 0$ and $t \geqslant 500$. It means that for such t's, gene occurrences seem to be uncorrelated of `tataat` occurrences.
- Conversely, if $t \in [0; 500]$, $\tilde{h}(t) > 0$, meaning that short distances are favored; smaller the distance, higher is the influence.

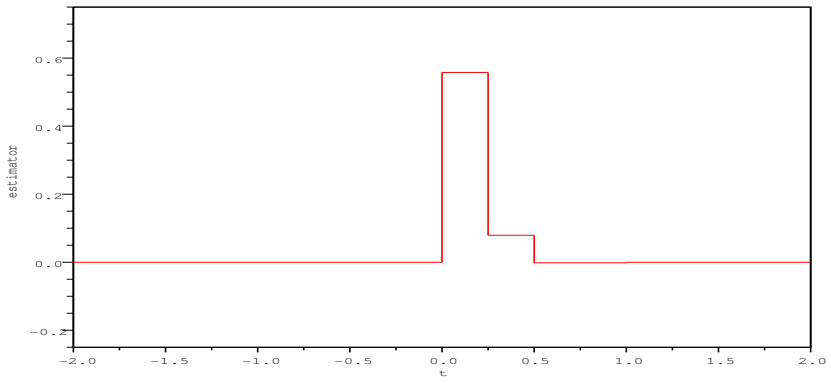

Figure 7: *Estimator* $\tilde{h}$ *defined in Section 3.1 for E. coli data at the scale* $1:1000$*, with parents=*`tataat` *and children=genes.*

Then, we investigate the way genes influences the DNA motif `tataat` and so, in our model, the parents are the occurrences of genes and children are the occurrences of `tataat`. Figure 8 gives the

estimator $\tilde{h}$ defined in Section 3.1 (with $(\gamma, \delta) = (0.72, 2.4)$). The shape of this estimator explains how occurrences of `tataat` are influenced by occurrences of genes. We can draw following conclusions, that is completely coherent with biological observations:

- When $t \leqslant -500$ and $t \geqslant 1000$, $\tilde{h}(t) = 0$. It means that for such t's, `tataat` occurrences seem to be uncorrelated of gene occurrences.
- When $t \in [-500; 0]$, $\tilde{h}(t) > 0$, meaning that there is a preference having a word `tataat` just before the occurrence of a gene. It corresponds to the same conclusions drawn from Figure 7 (second point). The motif `tataat` is part of the most common promoter sites of *E. coli* meaning that it should occur in front of the majority of the genes.
- When $t \in [0; 1000]$, $\tilde{h}(t) < 0$; occurrences of `tataat` are avoided for such distances t. Genes on the same strand do not usually overlap and they are about 1000 bases long in average: this fact can explain this conclusion.

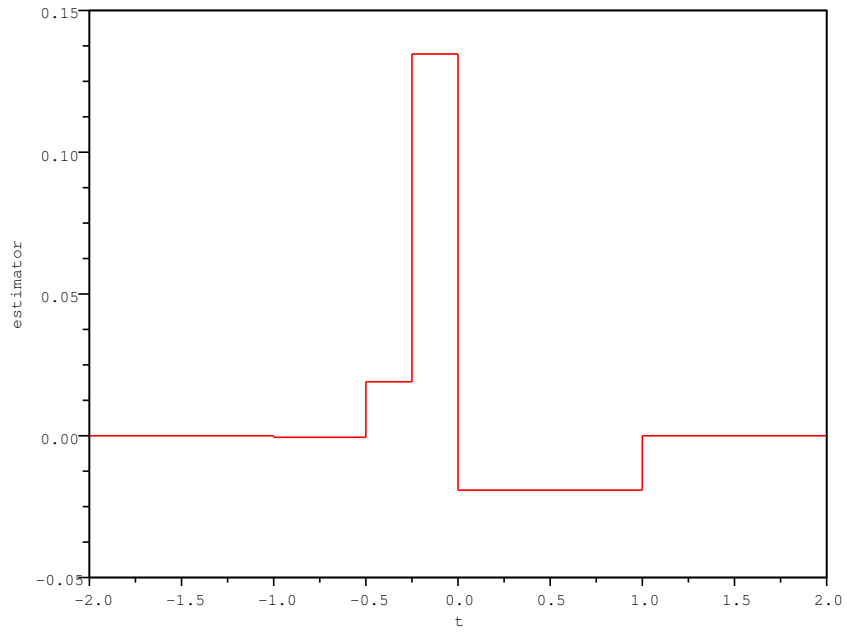

Figure 8: *Estimator $\tilde{h}$ defined in Section 3.1 for E. coli data at the scale* $1 : 1000$*, with parents=genes and children=`tataat`.*

Finally, Figure 9 presents the results of the FADO procedure [12] and Figure 10 presents the results of the Islands procedure of [27]. For the FADO procedure, we have forced the estimators to be piecewise constant to make the comparison easier. Our results agree with the ones obtained by FADO and Islands. But our method has advantage to point out that nothing significant happens after a certain distance (contrary to the FADO procedure), has advantage to treat interaction with another type of events (contrary to the Islands procedure) and has advantage to deal with the dependence on the past occurrences but also on the future occurrences (the function h is supported in $\mathbb{R}_+$ for the two other procedures). For algorithmic reason, a practical limitation of our method is that we only consider piecewise constant estimators (as for the Islands procedure), but it is enough to get a general trend on favored or avoided distances within a point process.

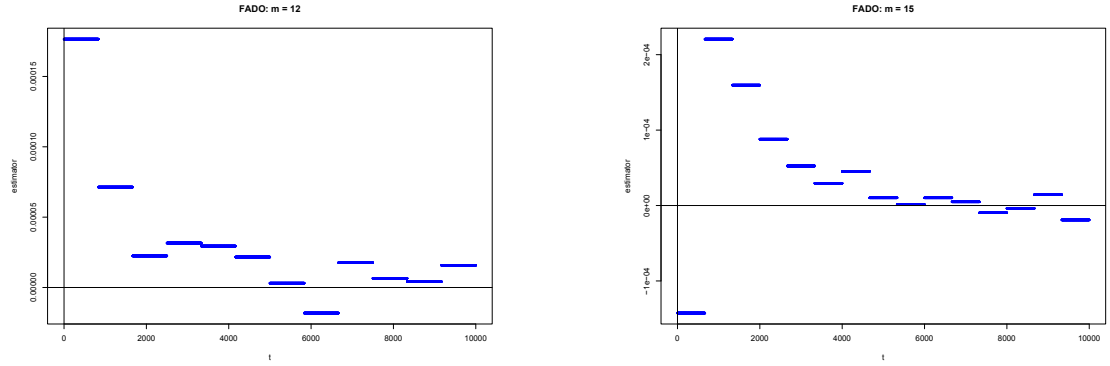

Figure 9: *FADO estimators for both E. coli datasets: left:* `tataat`*; right: genes.*

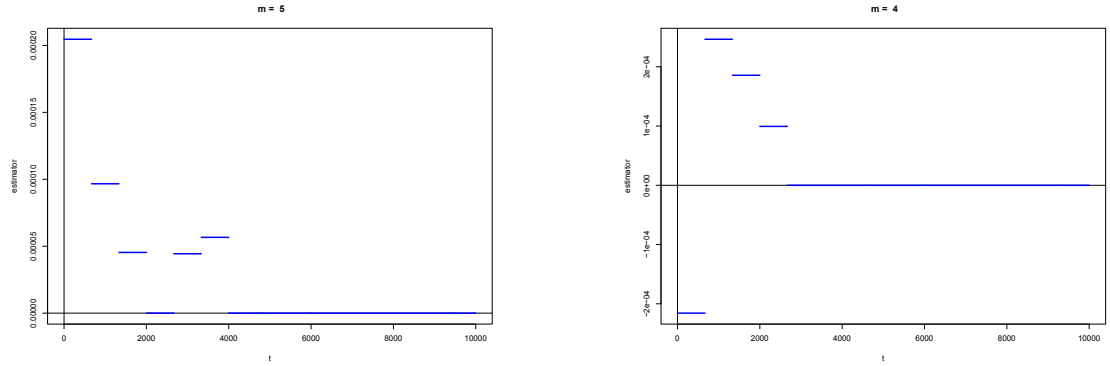

Figure 10: *Islands estimators for both E. coli datasets: left:* `tataat`*; right: genes.*

5 Conclusion

In our paper, we have investigated the dependencies between two given motifs. A random thresholding procedure has been proposed in Section 2.2. The general results of Section 2.3 have revealed the optimality of the procedure in the oracle and minimax setting. Our theoretical results have been strengthened by simulations illustrating the robustness of our procedure, despite a calibration of parameters from a practical point of view that differs from the theoretical choice. Section 4 has validated the procedure with a good detection of favored or avoided distances between occurrences of `tataat` and genes along the *E. coli* genome.

Further extensions of our model could be investigated. First, we could consider a more sophisticated model that takes into account the phenomenons of spontaneous apparition and self-excitation (as for the complete Hawkes model). But this model raises serious difficulties from the theoretical point of view. This is an exciting challenge to overcome them. Secondly, we could extend our cascade algorithm to general wavelet bases and not only to Haar bases. Finally, it is also relevant to study similar processes in the spatial framework and to connect them, for instance, to the Neymann-Scott process (see Section 6.3 of [5]), which is a stimulating topic we wish to consider.

Acknowledgments: The author wishes to thank Sophie Schbath for the two genomic data sets used in Section 4 and both her PhD advisors, Patricia Reynaud-Bouret and Vincent Rivoirard, for a wealth of smart advice and encouragement along this work.

References

[1] Boucheron, S., Bousquet, O., Lugosi, G. and Massart, P. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, Mar. 2005.

[2] Carstensen, L., Sandelin, A., Winther, O. and Hansen, N. R. Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC Bioinformatics*, 11(456), Sep. 2010.

[3] Cohen, A., Daubechies, I. and Feauveau, J. C. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.

[4] Comte, F., Gaïffas, S. and Guilloux, A. Adaptive estimation of the conditional intensity of marker-dependent counting processes. *arXiv:0810.4263v1*, Oct. 2008.

[5] Daley, D. J. and Vere-Jones, D. *An Introduction to the Theory of Point Processes - Volume I: Elementary Theory and Methods*. Probab. Appl. (N. Y.), 2003.

[6] de la Peña, V. H. and Giné, E. *Decoupling: From Dependence to Independence*. Probab. Appl. (N. Y.), 1999.

[7] de la Peña, V. H. and Montgomery-Smith, S. J. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.*, 23(2):806–816, Apr. 1995.

[8] Donoho, D. L. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Different Perspectives on Wavelets - AMS Short Course, San Antonio (Texas), 1993*, volume 47, pages 173–205. Proc. Sympos. Appl. Math., 1993.

[9] Donoho, D. L. and Johnstone, I. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[10] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539, 1996.

[11] Giné, E., Latala, R. and Zinn, J. Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II*, volume 47, pages 13–38. Progr. Probab. Birkhäuser Boston, 2000.

[12] Gusto, G. and Schbath, S. FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Stat. Appl. Genet. Mol. Biol.*, 4(1), Sep. 2005.

[13] Hawkes, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[14] Houdré, C. and Reynaud-Bouret, P. Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications*, volume 56, pages 55–69. Progr. Probab. Birkhäuser Verlag Basel, 2003.

[15] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. *Wavelets, Approximation and Statistical Applications*, volume 129. Lecture Notes in Statist., 1998.

[16] Juditsky, A. and Lambert-Lacroix, S. On minimax density estimation on $\mathbb{R}$. *Bernoulli*, 10(2):187–220, Apr. 2004.

[17] Kingman, J. F. C. *Poisson Processes*. Oxford Sci. Publ., 1993.

[18] Kolaczyk, E. D. Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statist. Sinica*, 9:119–135, 1999.

[19] Mallat, S. G. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans. Amer. Math. Soc.*, 315(1):69–87, Sep. 1989.

[20] Massart, P. *Concentration Inequalities and Model Selection, St. Flour 2003*, volume 1896. Lecture Notes in Math., 2007.

[21] Ogata, Y. and Akaik H. On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 44(1):102–107, 1982.

[22] Ozaki, T. Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.*, 31(1):145–155, 1979.

[23] Rasmussen, J. G. *Aspects of temporal and spatio-temporal processes*. PhD thesis, Aalborg University - Denmark, 2006.

[24] Reinert, G. and Schbath, S. Large compound Poisson approximations for occurrences of multiple words. In *Statistics in Molecular Biology and Genetics*, volume 33, pages 257–275. IMS Lecture Notes Monogr. Ser., 1999.

[25] Reynaud-Bouret, P. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Related Fields*, 126:103–153, 2003.

[26] Reynaud-Bouret, P. and Rivoirard, V. Near optimal thresholding estimation of a Poisson intensity on the real line. *Electron. J. Stat.*, 4:172–238, 2010.

[27] Reynaud-Bouret, P. and Schbath, S. Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.*, 38(5):2781–2822, 2010.

[28] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. Adaptive density estimation: a curse of support? *J. Statist. Plann. Inference*, 141:115–139, 2011.

[29] Robin, S., Rodolphe, F. and Schbath, S. *DNA, Words and Models*. Cambridge Univ. Press, 2005.

[30] Rudemo, M. Empirical choice of histograms and kernel density estimators. *Scand. J. Stat.*, 9:65–78, 1982.

[31] Schbath, S. Compound Poisson approximation of word counts in DNA sequences. *ESAIM Probab. Stat.*, 1:1–16, 1995.

[32] Willett, R. M. and Nowak, R. D. Multiscale Poisson intensity and density estimation. *IEEE Trans. Inform. Theory*, 53(9):3171–3187, Sep. 2007.

6 Appendix: Proof of Theorem 1

In the sequel, the values of the constants $K, K', K_0, K_1, K_2, K_3, \ldots$ may change from line to line. For the sake of clarity, the proofs are fully detailed in this appendix.

6.1 A more general result

We first give a general result stated and proved in [26].

Theorem 2 (Theorem 2.2 of [26]). *To estimate a countable family $\beta = (\beta_\lambda)_{\lambda\in\Lambda}$, such that $\|\beta\|_{\ell_2} < \infty$, we assume that a family of coefficient estimators $(\hat{\beta}_\lambda)_{\lambda\in\Gamma}$, where Γ is a known deterministic subset of Λ, and a family of possibly random thresholds $(\eta_\lambda)_{\lambda\in\Gamma}$ are available and we consider the thresholding rule*

$$\tilde{\beta} = \left(\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geqslant \eta_\lambda} \mathbf{1}_{\lambda\in\Gamma}\right)_{\lambda\in\Lambda}.$$

Let $\varepsilon > 0$ be fixed. Assume that there exist a deterministic family $(H_\lambda)_{\lambda\in\Gamma}$ and three constants $\kappa \in [0;1[$, $\omega \in [0;1]$ and $\zeta > 0$ (that may depend on ε but not on λ) with the following properties:

(A1) For all λ in Γ,

$$\mathbb{P}\left(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa\eta_\lambda\right) \leqslant \omega.$$

(A2) There exist $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$ and a constant $R > 0$ such that for all λ in Γ,

$$\left[\mathbb{E}\left(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}\right)\right]^{\frac{1}{p}} \leqslant R\max\left\{H_\lambda, H_\lambda^{\frac{1}{p}}\varepsilon^{\frac{1}{q}}\right\}.$$

(A3) There exists a constant θ such that for all λ in Γ such that $H_\lambda < \theta\varepsilon$,

$$\mathbb{P}\left(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa\eta_\lambda, |\hat{\beta}_\lambda| > \eta_\lambda\right) \leqslant H_\lambda\zeta.$$

Then the estimator $\tilde{\beta}$ satisfies

$$\frac{1-\kappa^2}{1+\kappa^2}\mathbb{E}\left(\|\tilde{\beta} - \beta\|_{\ell_2}^2\right) \leqslant \mathbb{E}\left(\inf_{m\subset\Gamma}\left\{\frac{1+\kappa^2}{1-\kappa^2}\sum_{\lambda\notin m}\beta_\lambda^2 + \frac{1-\kappa^2}{\kappa^2}\sum_{\lambda\in m}(\hat{\beta}_\lambda - \beta_\lambda)^2 + \sum_{\lambda\in m}\eta_\lambda^2\right\}\right) + LD\sum_{\lambda\in\Gamma}H_\lambda,$$

with $LD = \frac{R}{\kappa^2}\left((1+\theta^{-1/q})\omega^{1/q} + (1+\theta^{1/q})\varepsilon^{1/q}\zeta^{1/q}\right)$.

Using the previous theorem, we establish the following result that we will prove in Section 6.4.

Theorem 3. *Let $n \geqslant 1$, $j_0 \in \mathbb{N}^*$, $\gamma > 0$ and Δ defined by (6.15) and (6.16). Then the estimator $\tilde{\beta}$ defined by (2.8) in Section 2.2 satisfies*

$$\mathbb{E}\left(\|\tilde{\beta} - \beta\|_{\ell_2}^2\right) \leqslant C_1 \inf_{m\subset\Gamma}\left\{\sum_{\lambda\notin m}\beta_\lambda^2 + F(j_0,n,T)|m|\right\} + C_2R\left(e^{-\kappa_1 j_0\gamma/2} + e^{-\kappa_2 n\|h\|_1/2}\right)2^{j_0},$$

where C_1 is a positive constant depending on γ, $\|h\|_1$, $\|h\|_\infty$, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$, C_2 is a positive constant depending on the compact support of h and the compact support of ψ,

$$F(j_0,n,T) = \frac{j_0}{n} + \frac{j_0^{3/2}2^{j_0/2}}{n^{3/2}} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0^2}{T} + \frac{j_0^{3/2}2^{j_0/2}}{nT^{1/2}} + \frac{j_0 n}{T^2},$$

$$R = C_R\left\{\frac{1}{n} + \frac{2^{j_0/2}}{n^{3/2}} + \frac{2^{j_0/2}}{nT^{1/2}} + \frac{n}{T^2}\right\},$$

with C_R a positive constant depending on $\|h\|_1$, $\|h\|_\infty$, the compact support of ψ, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$ and κ_1 and κ_2 are absolute constants in $]0;1[$.

To obtain Theorem 1, we consider $n \geqslant 2$, we take j_0 the positive integer such that $2^{j_0} \leqslant n < 2^{j_0+1}$ and $\gamma > 2\log 2$ in the previous theorem. Therefore, we note that

$$\begin{aligned}
F(j_0,n,T) &= \frac{j_0}{n} + \frac{j_0^{3/2}2^{j_0/2}}{n^{3/2}} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0^2}{T} + \frac{j_0^{3/2}2^{j_0/2}}{nT^{1/2}} + \frac{j_0 n}{T^2} \\
&\leqslant K\left\{\frac{\log n}{n} + \frac{(\log n)^{3/2}n^{1/2}}{n^{3/2}} + \frac{(\log n)^4 n}{n^2} + \frac{(\log n)^2}{T} + \frac{(\log n)^{3/2}n^{1/2}}{nT^{1/2}} + \frac{(\log n)n}{T^2}\right\} \\
&\leqslant K\left\{\frac{\log n}{n} + \frac{(\log n)^{3/2}}{n} + \frac{(\log n)^4}{n} + \frac{(\log n)^2}{T} + \frac{(\log n)^{3/2}}{n^{1/2}T^{1/2}} + \frac{(\log n)n}{T^2}\right\} \\
&\leqslant K\left\{\frac{(\log n)^4}{n} + \frac{(\log n^2)n}{T^2}\right\},
\end{aligned}$$

by comparing all the terms of the right-hand side between them (for this, we distinguish the cases $n \leqslant T$ and $n \geqslant T$), with K an absolute positive constant (that changes from line to line) and

$$R \leqslant K' C_R \left\{ \frac{1}{n} + \frac{n}{T^2} \right\},$$

with K' an absolute positive constant. Moreover, $e^{-\kappa_1 j_0 \gamma/2} 2^{j_0}$ is bounded thanks to the choice of γ.

Finally, since

$$\|\tilde{h} - h\|_2^2 \leqslant K_0 \|\tilde{\beta} - \beta\|_{\ell_2}^2,$$

with K_0 a positive constant depending only on the functions that generate the biorthogonal wavelet basis, we establish Theorem 1.

6.2 Technical lemmas

Before proving Theorem 3, we establish two lemmas which we will use throughout the proof.

Lemma 6.1.

(a) For any function f in $\mathbb{L}_2(\mathbb{R})$ and for all $t \in \mathbb{R}$, $\mathrm{Var}_\pi(f(t-U)) \leqslant \frac{1}{T}\int_{\mathbb{R}} f^2(x)\,dx$, where $\mathrm{Var}_\pi(f(t-U))$ denotes the variance of $f(t-U)$ where $U \sim \pi$.

(b) For any function f in $\mathbb{L}_1(\mathbb{R})$ and for all $t \in \mathbb{R}$, $\int_{\mathbb{R}} \mathbb{E}_\pi(f(t-U))\,dt = \int_{\mathbb{R}} f(x)\,dx$.

(c) For any nonnegative function f in $\mathbb{L}_1(\mathbb{R})$ and for all $t \in \mathbb{R}$, $\mathbb{E}_\pi(f(t-U)) \leqslant \frac{1}{T}\int_{\mathbb{R}} f(x)\,dx$.

Proof.

(a) Let $f \in \mathbb{L}_2(\mathbb{R})$ and $t \in \mathbb{R}$.

$$\mathrm{Var}_\pi(f(t-U)) \leqslant \mathbb{E}_\pi(f^2(t-U)) = \frac{1}{T}\int_0^T f^2(t-u)\,du \leqslant \frac{1}{T}\int_{\mathbb{R}} f^2(x)\,dx.$$

(b) Let $f \in \mathbb{L}_1(\mathbb{R})$ and $t \in \mathbb{R}$.

$$\int_{\mathbb{R}} \mathbb{E}_\pi(f(t-U))\,dt = \mathbb{E}_\pi\left(\int_{\mathbb{R}} f(t-U)\,dt\right) = \mathbb{E}\left(\int_{\mathbb{R}} f(x)\,dx\right) = \int_{\mathbb{R}} f(x)\,dx.$$

(c) Let $f \in \mathbb{L}_1(\mathbb{R})$ such that $f \geqslant 0$ and $t \in \mathbb{R}$.

$$\mathbb{E}_\pi(f(t-U)) = \frac{1}{T}\int_0^T f(t-u)\,du \leqslant \frac{1}{T}\int_{\mathbb{R}} f(x)\,dx.$$

□

The next result is a Rosenthal type inequality for any Poisson process, that extends Lemma 6.2 of [26].

Lemma 6.2. *Let $p \geqslant 1$. Consider a Poisson process N on $(\mathbb{X}, \mathcal{X})$ a measurable space, with a finite mean measure $\nu : \mathcal{X} \mapsto \mathbb{R}_+$ and a function $\varphi : \mathbb{X} \mapsto \mathbb{R}$ which belongs to $\mathbb{L}_{2p}(\nu)$. We denote $\hat{\beta} = \int_{\mathbb{X}} \varphi(x)\,dN_x$ a natural estimator of $\beta = \int_{\mathbb{X}} \varphi(x)\,d\nu(x)$ that satisfies $\mathbb{E}(\hat{\beta}) = \beta$. Then, there exists a positive constant $C(p)$ only depending on p such that*

$$\mathbb{E}(|\hat{\beta} - \beta|^{2p}) \leqslant C(p)\left(\int_{\mathbb{X}} |\varphi(x)|^{2p}\,d\nu(x) + \left(\mathrm{Var}(\hat{\beta})\right)^p\right),$$

where $\mathrm{Var}(\hat{\beta}) = \int_{\mathbb{X}} \varphi^2(x)\,d\nu(x)$.

Proof. Let $p \geqslant 1$. Suppose $\|\varphi\|_\infty < +\infty$ first. As a Poisson process is infinitely divisible, we can write: for any positive integer k,

$$dN = \sum_{i=1}^{k} dN^i,$$

where the N^i's are mutually independent Poisson processes on $\mathbb{X}$ with mean measure ν/k. Hence,

$$\hat{\beta} - \beta = \sum_{i=1}^{k} \int_{\mathbb{X}} \varphi(x) \left(dN_x^i - k^{-1} d\nu(x) \right) = \sum_{i=1}^{k} Y_i,$$

where for any i,

$$Y_i = \int_{\mathbb{X}} \varphi(x) \left(dN_x^i - k^{-1} d\nu(x) \right).$$

So the Y_i's are i.i.d. centered variables, each of them has moments of order $2p$ and 2. We apply the classical Rosenthal's inequality (for instance, see Proposition 10.2 of [15]): there exists a positive constant $C(p)$ only depending on p such that

$$\mathbb{E}\left(\left| \sum_{i=1}^{k} Y_i \right|^{2p} \right) \leqslant C(p) \left(\sum_{i=1}^{k} \mathbb{E}(|Y_i|^{2p}) + \left(\sum_{i=1}^{k} \mathbb{E}(Y_i^2) \right)^p \right).$$

Now, we give an upper bound of the limit of $\mathbb{E}\left(\sum_{i=1}^{k} |Y_i|^\ell \right)$ for $\ell \in \{2p, 2\}$ when $k \to \infty$. Let us introduce

$$\Omega_k = \left\{ \forall i \in \{1, \dots, k\}, N_{\mathbb{X}}^i \leqslant 1 \right\},$$

where $N_{\mathbb{X}}^i$ is the number of points of N^i lying in $\mathbb{X}$. Then,

$$\begin{aligned} \mathbb{P}(\Omega_k^c) &= \mathbb{P}(\exists i \in \{1, \dots, k\}, N_{\mathbb{X}}^i \geqslant 2) \\ &\leqslant \sum_{i=1}^{k} \mathbb{P}(N_{\mathbb{X}}^i \geqslant 2) = k \sum_{j \geqslant 2} \frac{(\nu(\mathbb{X})/k)^j}{j!} e^{-\nu(\mathbb{X})/k} \\ &\leqslant k(\nu(\mathbb{X})/k)^2 = k^{-1} \nu(\mathbb{X})^2. \end{aligned}$$

On Ω_k, if $N_{\mathbb{X}}^i = 0$ (so $\int_{\mathbb{X}} \varphi(x)\, dN_x^i = 0$),

$$|Y_i|^\ell = O_k(k^{-\ell})$$

and if $N_{\mathbb{X}}^i = 1$ (so $\int_{\mathbb{X}} \varphi(x)\, dN_x^i = \varphi(T)$, where T is the point of the process N^i),

$$|Y_i|^\ell = |\varphi(T)|^\ell + O_k(k^{-1} |\varphi(T)|^{\ell-1}).$$

Consequently,

$$\begin{aligned} &\mathbb{E}\left(\sum_{i=1}^{k} |Y_i|^\ell \right) \\ &\leqslant \mathbb{E}\left[\mathbf{1}_{\Omega_k} \left(k O_k(k^{-\ell}) + \sum_{T \in N} \left[|\varphi(T)|^\ell + O_k(k^{-1} |\varphi(T)|^{\ell-1}) \right] \right) \right] + \sqrt{\mathbb{P}(\Omega_k^c)} \sqrt{\mathbb{E}\left[\left(\sum_{i=1}^{k} |Y_i|^\ell \right)^2 \right]}. \end{aligned} \tag{6.1}$$

But we have

$$\begin{aligned}\sum_{i=1}^{k}|Y_i|^\ell &\leqslant 2^{\ell-1}\sum_{i=1}^{k}\left[\left|\int_{\mathbb{X}}\varphi(x)\,dN_x^i\right|^\ell+\left(k^{-1}\int_{\mathbb{X}}|\varphi(x)|\,d\nu(x)\right)^\ell\right]\\ &\leqslant 2^{\ell-1}\left(\sum_{i=1}^{k}\|\varphi\|_\infty^\ell(N_{\mathbb{X}}^i)^\ell+k\left(k^{-1}\int_{\mathbb{X}}|\varphi(x)|\,d\nu(x)\right)^\ell\right)\\ &\leqslant 2^{\ell-1}\left(\|\varphi\|_\infty^\ell N_{\mathbb{X}}^\ell+k\left(k^{-1}\int_{\mathbb{X}}|\varphi(x)|\,d\nu(x)\right)^\ell\right).\end{aligned}$$

Thus, when $k\to\infty$, the last term in (6.1) converges to 0 since a Poisson variable has moments of every order and

$$\limsup_{k\to\infty}\mathbb{E}\left(\sum_{i=1}^{k}|Y_i|^\ell\right)\leqslant\mathbb{E}\left(\int_{\mathbb{X}}|\varphi(x)|^\ell\,dN_x\right)=\int_{\mathbb{X}}|\varphi(x)|^\ell\,d\nu(x),$$

which concludes the proof in the bounded case.
But for any function φ such that $\int_{\mathbb{X}}|\varphi(x)|^{2p}\,d\nu(x)<+\infty$, the desired upper bound is finite and we get it by approximating φ by, for instance, piecewise constant functions. □

6.3 Proof of Lemma 2.1

Let $\lambda\in\Lambda$ be fixed. $G(\varphi_\lambda)$, defined by (2.3), is a measurable function of the observations and by considering the aggregated process (1.1), we can write

$$\begin{aligned}G(\varphi_\lambda)&=\int_{\mathbb{R}}\sum_{i=1}^{n}\left[\varphi_\lambda(t-U_i)-\frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]dN_t\\ &=\int_{\mathbb{R}}\sum_{i=1}^{n}\varphi_\lambda(t-U_i)\,dN_t-(n-1)\int_{\mathbb{R}}\mathbb{E}_\pi(\varphi_\lambda(t-U))\,dN_t\\ &=\sum_{1\leqslant i,j\leqslant n}\int_{\mathbb{R}}\varphi_\lambda(t-U_i)\,dN_t^j-\sum_{1\leqslant i\neq j\leqslant n}\int_{\mathbb{R}}\mathbb{E}_\pi(\varphi_\lambda(t-U))\,dN_t^j\\ &=\sum_{i=1}^{n}\left[\int_{\mathbb{R}}\varphi_\lambda(t-U_i)\,dN_t^i+\sum_{j\neq i}\int_{\mathbb{R}}\left[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]dN_t^j\right].\end{aligned}$$

Now, we prove the first part of Lemma 2.1. We have

$$\begin{aligned}&\mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)\\ &=\sum_{i=1}^{n}\left[\int_{\mathbb{R}}\varphi_\lambda(t-U_i)h(t-U_i)\,dt+\sum_{j\neq i}\int_{\mathbb{R}}\left[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]h(t-U_j)\,dt\right].\end{aligned}$$

Write $x=t-U_i$ in the first integral. Therefore,

$$\mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)=n\int_{\mathbb{R}}\varphi_\lambda(x)h(x)\,dx+W(\varphi_\lambda),$$

where

$$W(\varphi_\lambda)=\sum_{1\leqslant i\neq j\leqslant n}\int_{\mathbb{R}}\left[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]h(t-U_j)\,dt.$$

Moreover,

$$\begin{aligned}\mathbb{E}(W(\varphi_\lambda)) &= \sum_{1\leqslant i\neq j\leqslant n}\int_{\mathbb{R}}\mathbb{E}\Big(\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]h(t-U_j)\Big)\,dt\\ &= \sum_{1\leqslant i\neq j\leqslant n}\int_{\mathbb{R}}\mathbb{E}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}(h(t-U_j))\,dt\\ &= 0.\end{aligned}$$

Finally,

$$\mathbb{E}(G(\varphi_\lambda)) = n\int_{\mathbb{R}}\varphi_\lambda(x)h(x)\,dx,$$

i.e. $\hat{\beta}_\lambda$ is an unbiased estimator for β_λ:

$$\mathbb{E}(\hat{\beta}_\lambda) = \mathbb{E}\left(\frac{G(\varphi_\lambda)}{n}\right) = \int_{\mathbb{R}}\varphi_\lambda(x)h(x)\,dx = \beta_\lambda.$$

It remains to control the variance of the estimator $\hat{\beta}_\lambda$.

$$\begin{aligned}\mathrm{Var}(G(\varphi_\lambda)) &= \mathbb{E}\left[\left(G(\varphi_\lambda)-n\int_{\mathbb{R}}\varphi_\lambda(x)h(x)\,dx\right)^2\right]\\ &= \mathbb{E}\big[(G(\varphi_\lambda)-\mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)+W(\varphi_\lambda))^2\big]\\ &= \mathbb{E}(V(\varphi_\lambda))+\mathbb{E}(W(\varphi_\lambda)^2),\end{aligned}$$

where

$$V(\varphi_\lambda) = \mathrm{Var}(G(\varphi_\lambda)|U_1,\dots,U_n).$$

We start by dealing with the first term by using technics for Poisson processes. We have

$$\begin{aligned}V(\varphi_\lambda) &= \int_{\mathbb{R}}\left(\sum_{i=1}^n\left[\varphi_\lambda(t-U_i)-\frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right)^2\sum_{j=1}^n h(t-U_j)\,dt\\ &= \int_{\mathbb{R}}\sum_{j=1}^n\left(\varphi_\lambda(t-U_j)+\sum_{i\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\right)^2 h(t-U_j)\,dt\\ &= \sum_{j=1}^n\int_{\mathbb{R}}\varphi_\lambda^2(t-U_j)h(t-U_j)\,dt\\ &\quad+2\sum_{j=1}^n\int_{\mathbb{R}}\sum_{i\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\varphi_\lambda(t-U_j)h(t-U_j)\,dt\\ &\quad+\sum_{j=1}^n\int_{\mathbb{R}}\sum_{i\neq j}\sum_{k\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]h(t-U_j)\,dt.\end{aligned}$$

In the first integral, write $x = t-U_j$. So,

$$\begin{aligned}V(\varphi_\lambda) &= n\int_{\mathbb{R}}\varphi_\lambda^2(x)h(x)\,dx+2\sum_{j=1}^n\int_{\mathbb{R}}\sum_{i\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\varphi_\lambda(t-U_j)h(t-U_j)\,dt\\ &\quad+\sum_{j=1}^n\int_{\mathbb{R}}\sum_{i\neq j}\sum_{k\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]h(t-U_j)\,dt.\end{aligned}\tag{6.2}$$

Each term can be computed by taking the expectation conditionally to U_j (in each sum) and we obtain

$$\begin{aligned}\mathbb{E}(V(\varphi_\lambda)) &= n\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + \sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i\neq j}\mathbb{E}\Big(\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\Big)\mathbb{E}(h(t-U_j))\,dt\\ &= n\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + n(n-1)\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\mathbb{E}_\pi(h(t-U))\,dt. \qquad (6.3)\end{aligned}$$

Then using equations (a) and (b) of Lemma 6.1, we have

$$\mathbb{E}(V(\varphi_\lambda)) \leqslant n\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + \frac{n(n-1)}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\int_{\mathbb{R}}h(x)\,dx. \qquad (6.4)$$

Now, we deal with the second term by using the U-statistics technics. However, $W(\varphi_\lambda)$ is a U-statistics of order 2 but it is not degenerate. So we write

$$W(\varphi_\lambda) = W_1(\varphi_\lambda) + W_2(\varphi_\lambda), \qquad (6.5)$$

with

$$\begin{aligned}W_1(\varphi_\lambda) &= \sum_{1\leqslant i\neq j\leqslant n}\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\\ &= (n-1)\sum_{i=1}^{n}\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\end{aligned}$$

and

$$W_2(\varphi_\lambda) = \sum_{1\leqslant i\neq j\leqslant n} g(U_i,U_j),$$

where

$$g(U_i,U_j) = \int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt.$$

$W_2(\varphi_\lambda)$ is a degenerate U-statistics. It is easy to verify that

$$\mathbb{E}(W(\varphi_\lambda)^2) = \mathbb{E}(W_1(\varphi_\lambda)^2) + \mathbb{E}(W_2(\varphi_\lambda)^2).$$

First we compute $\mathbb{E}(W_1(\varphi_\lambda)^2)$.

$$\begin{aligned}\mathbb{E}(W_1(\varphi_\lambda)^2) &= \mathrm{Var}(W_1(\varphi_\lambda))\\ &= n(n-1)^2\mathrm{Var}\left(\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right)\\ &= n(n-1)^2\mathrm{Var}\left(\int_{\mathbb{R}}\varphi_\lambda(t-U_1)\mathbb{E}_\pi(h(t-U))\,dt\right)\\ &\leqslant n(n-1)^2\mathbb{E}\left[\left(\int_{\mathbb{R}}|\varphi_\lambda(t-U_1)|\mathbb{E}_\pi(h(t-U))\,dt\right)^2\right]\\ &\leqslant \frac{n(n-1)^2}{T^2}\mathbb{E}\left[\left(\int_{\mathbb{R}}|\varphi_\lambda(t-U_1)|\,dt\right)^2\right]\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\\ &\leqslant \frac{n(n-1)^2}{T^2}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^2, \qquad (6.6)\end{aligned}$$

by applying inequality (c) of Lemma 6.1 with $f = h$.

It remains to compute $\mathbb{E}(W_2(\varphi_\lambda)^2)$. It is easy to see that

$$\begin{aligned}\mathbb{E}(W_2(\varphi_\lambda)^2) &= \sum_{1\leqslant i\neq j\leqslant n} \mathbb{E}\big[g(U_i,U_j)(g(U_i,U_j)+g(U_j,U_i))\big] \leqslant 2\sum_{1\leqslant i\neq j\leqslant n} \mathbb{E}\left[g(U_i,U_j)^2\right]\\ &\leqslant 2n(n-1)\mathbb{E}\left[\left(\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[h(t-U_2)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right)^2\right].\end{aligned}$$

We denote $\mathbb{E}_{(U,V)\sim\pi\otimes\pi}(g(U,V))$ the expectation of $g(U,V)$ where $U\sim\pi$ and $V\sim\pi$ are independent and $f^X(t)=f(t-X)$. Hence,

$$\begin{aligned}&\mathbb{E}(W_2(\varphi_\lambda)^2)\\ &\leqslant 2n(n-1)\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left[\left(\int_{\mathbb{R}}\big[\varphi_\lambda^U(t)-\mathbb{E}_\pi(\varphi_\lambda^U(t))\big]\big[h^V(t)-\mathbb{E}_\pi(h^V(t))\big]\,dt\right)^2\right]\\ &\leqslant 2n(n-1)\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\Bigg[\Bigg(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt-\mathbb{E}_{V\sim\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)\\ &\qquad\qquad -\mathbb{E}_{U\sim\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)+\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)\Bigg)^2\Bigg]\\ &\leqslant 2n(n-1)\Bigg\{\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left[\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)^2\right]-\mathbb{E}_{U\sim\pi}\left[\left(\mathbb{E}_{V\sim\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)\right)^2\right]\\ &\qquad\qquad -\mathbb{E}_{V\sim\pi}\left[\left(\mathbb{E}_{U\sim\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)\right)^2\right]+\left(\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)\right)^2\Bigg\}\\ &\leqslant 2n(n-1)\Bigg\{\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left[\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)^2\right]+\left(\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)\right)^2\Bigg\}.\end{aligned}$$

But,

$$\begin{aligned}\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left[\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)^2\right] &\leqslant \mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left(\int_{\mathbb{R}}(\varphi_\lambda^U)^2(t)h^V(t)\,dt\int_{\mathbb{R}}h^V(t)\,dt\right)\\ &= \mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left(\int_{\mathbb{R}}(\varphi_\lambda^U)^2(t)h^V(t)\,dt\right)\int_{\mathbb{R}}h(x)\,dx\\ &= \int_{\mathbb{R}}\mathbb{E}_\pi\big((\varphi_\lambda^U)^2(t)\big)\mathbb{E}_\pi(h^V(t))\,dt\int_{\mathbb{R}}h(x)\,dx\\ &\leqslant \frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\end{aligned}$$

and

$$\left|\mathbb{E}_{(U,V)\sim\pi\otimes\pi}\left(\int_{\mathbb{R}}\varphi_\lambda^U(t)h^V(t)\,dt\right)\right| = \left|\int_{\mathbb{R}}\mathbb{E}_\pi(\varphi_\lambda^U(t))\mathbb{E}_\pi(h^V(t))\,dt\right| \leqslant \frac{1}{T}\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\int_{\mathbb{R}}h(x)\,dx,$$

by using Lemma 6.1. So,

$$\mathbb{E}(W_2(\varphi_\lambda)^2)\leqslant 2n(n-1)\left\{\frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2+\frac{1}{T^2}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\right\}. \tag{6.7}$$

Finally, by combining inequalities (6.4), (6.6) and (6.7), we obtain the following control of the

variance of the estimator $\hat{\beta}_\lambda$:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_\lambda) &= \mathrm{Var}\left(\frac{G(\varphi_\lambda)}{n}\right)\\ &\leqslant \frac{1}{n}\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + \frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\int_{\mathbb{R}}h(x)\,dx\\ &\quad + \frac{n}{T^2}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^2 + \frac{2}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\\ &\leqslant \frac{1}{n}\int_{\mathbb{R}}\varphi_\lambda^2(x)h(x)\,dx + \frac{1}{T}\|\varphi_\lambda\|_2^2\|h\|_1 + \frac{n}{T^2}\|\varphi_\lambda\|_1^2\|h\|_1^2 + \frac{2}{T}\|\varphi_\lambda\|_2^2\|h\|_1^2.\end{aligned}$$

By using the properties of the biorthogonal wavelet bases considered in this paper, for any $\lambda = (j,k)$ in Λ, we have: $\|\varphi_\lambda\|_1 \leqslant 2^{-j/2}\max(\sqrt{2}/2, \|\psi\|_1)$ and $\|\varphi_\lambda\|_2 \leqslant \max(1, \|\psi\|_2)$, which allows us to get the purposed upper bound in Lemma 2.1.

6.4 Proof of Theorem 3

In the sequel, we will consider: $n \geqslant 1$, $T \geqslant 1$ and $j_0 = O(n)$ and we will use following notations: $M_{h,1} = \max(\|h\|_1, 1)$, $M_{h,\infty} = \max(\|h\|_\infty, 1)$, $M_{\psi,1} = \max(\|\psi\|_1, \sqrt{2}/2)$, $M_{\psi,2} = \max(\|\psi\|_2, 1)$ and $M_{\psi,\infty} = \max(\|\psi\|_\infty, \sqrt{2})$ (so that, for any $\lambda = (j,k) \in \Lambda$, we have: $\|\varphi_\lambda\|_1 \leqslant 2^{-j/2}M_{\psi,1}$, $\|\varphi_\lambda\|_2 \leqslant M_{\psi,2}$ and $\|\varphi_\lambda\|_\infty \leqslant 2^{j/2}M_{\psi,\infty}$).

We recall that A and M are positive real numbers such that h and ψ are compactly supported in $[-A; A]$ and in $[-M; M]$ respectively.

Now, to prove Theorem 3, we apply Theorem 2 and for this purpose we have to verify Assumptions: (A1), (A2) and (A3).

6.4.1 Proof of Assumption (A1)

Let $\lambda \in \Gamma$ be fixed. Remember that conditionally to the U_i's, the expression given in (1.1) is a Poisson process. We apply Lemma 6.1 of [26]: for any $\alpha > 0$, with probability larger than $1 - 2e^{-\alpha}$, conditionally to the U_i's, we have

$$\left|G(\varphi_\lambda) - n\int_{\mathbb{R}}\varphi_\lambda(x)h(x)\,dx - W(\varphi_\lambda)\right| \leqslant \sqrt{2\alpha V(\varphi_\lambda)} + \frac{\alpha}{3}B(\varphi_\lambda),$$

where $W(\varphi_\lambda)$ is defined by (6.5), $V(\varphi_\lambda) = \mathrm{Var}(G(\varphi_\lambda)|U_1,\ldots,U_n)$ and

$$B(\varphi_\lambda) = \left\|\sum_{i=1}^n\left[\varphi_\lambda(\cdot - U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(\cdot - U))\right]\right\|_\infty.$$

Unlike $B(\varphi_\lambda)$, $V(\varphi_\lambda)$ is non-observable (it depends on the unknown function h). This is the reason why, by fixing $\alpha > 0$, we estimate $V(\varphi_\lambda)$ by

$$\widetilde{V}(\varphi_\lambda) = \hat{V}(\varphi_\lambda) + \sqrt{2\alpha\hat{V}(\varphi_\lambda)B^2(\varphi_\lambda)} + 3\alpha B^2(\varphi_\lambda)$$

where

$$\hat{V}(\varphi_\lambda) = \int_{\mathbb{R}}\left(\sum_{i=1}^n\left[\varphi_\lambda(t - U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right)^2 dN_t.$$

Moreover, by Lemma 6.1 of [26], we have also: $\mathbb{P}(V(\varphi_\lambda) \geqslant \widetilde{V}(\varphi_\lambda)) \leqslant e^{-\alpha}$. So, with probability larger than $1 - 3e^{-\alpha}$,

$$\left|G(\varphi_\lambda) - n\int_{\mathbb{R}}\varphi_\lambda(x)h(x)\,dx\right| \leqslant \sqrt{2\alpha\widetilde{V}(\varphi_\lambda)} + \frac{\alpha}{3}B(\varphi_\lambda) + |W(\varphi_\lambda)|. \tag{6.8}$$

We provide a control in probability of $W_1(\varphi_\lambda)$.

$$W_1(\varphi_\lambda) = (n-1)\sum_{i=1}^{n}\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt.$$

This is a sum of i.i.d. random variables. We apply Bernstein's inequality (for instance, see Proposition 2.9 of [20]) to get that with probability larger than $1-2e^{-\alpha}$,

$$|W_1(\varphi_\lambda)| \leqslant \sqrt{2\alpha v(\varphi_\lambda)} + \frac{\alpha}{3}b(\varphi_\lambda),$$

with

$$v(\varphi_\lambda) = \mathrm{Var}(W_1(\varphi_\lambda)) \leqslant \frac{n(n-1)^2}{T^2}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^2$$

(see inequality (6.6)) and

$$\begin{aligned} b(\varphi_\lambda) &= (n-1)\sup_{u\in[0;T]}\left|\int_{\mathbb{R}}\big[\varphi_\lambda(t-u) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right| \\ &\leqslant \frac{2(n-1)}{T}\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\int_{\mathbb{R}}h(x)\,dx, \end{aligned}$$

using equations (b) and (c) of Lemma 6.1. Then, with probability larger than $1-2e^{-\alpha}$,

$$|W_1(\varphi_\lambda)| \leqslant \frac{\sqrt{2\alpha n}(n-1)}{T}\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\int_{\mathbb{R}}h(x)\,dx + \frac{2\alpha(n-1)}{3T}\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\int_{\mathbb{R}}h(x)\,dx. \tag{6.9}$$

Now it remains to control $W_2(\varphi_\lambda)$, with

$$W_2(\varphi_\lambda) = \sum_{1\leqslant i\neq j\leqslant n} g(U_i,U_j),$$

where

$$g(U_i,U_j) = \int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[h(t-U_j) - \mathbb{E}_\pi(h(t-U))\big]\,dt.$$

This is a degenerate U-statistics of order 2, we can rewrite it as

$$W_2(\varphi_\lambda) = \sum_{1\leqslant j<i\leqslant n}\mathcal{G}(U_i,U_j),$$

where

$$\mathcal{G}(U_i,U_j) = g(U_i,U_j) + g(U_j,U_i).$$

We apply Theorem 3.4 of [14] to W_2 and $-W_2$ (keeping the same notations of [14]): for all $\varepsilon_0 > 0$ ($\varepsilon_0 = 1$ for instance), with probability larger than $1 - 2\times 2.77e^{-\alpha}$,

$$|W_2(\varphi_\lambda)| \leqslant 2(1+\varepsilon_0)^{3/2}C\sqrt{\alpha} + \eta(\varepsilon_0)D\alpha + \beta(\varepsilon_0)B\alpha^{3/2} + \gamma(\varepsilon_0)A\alpha^2, \tag{6.10}$$

where

- $A = \|\mathcal{G}\|_\infty$ and by applying equality (b) of Lemma 6.1 with $f = h$, we easily have

$$A \leqslant 8\|\varphi_\lambda\|_\infty\int_{\mathbb{R}}h(x)\,dx, \tag{6.11}$$

- $C^2 = \mathbb{E}(W_2(\varphi_\lambda)^2)$ and with (6.7), we have

$$C^2 \leqslant 2n(n-1)\left\{\frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2 + \frac{1}{T^2}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\right\}, \tag{6.12}$$

- $D = \sup\left\{\mathbb{E}\left(\sum_{1\leqslant j<i\leqslant n}\mathcal{G}(U_i,U_j)a_i(U_i)b_j(U_j)\right) : \mathbb{E}\left(\sum_{i=2}^{n}a_i(U_i)^2\right)\leqslant 1, \mathbb{E}\left(\sum_{j=1}^{n-1}b_j(U_j)^2\right)\leqslant 1\right\}$,

$$D = \sup\left\{\mathbb{E}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}g(U_i,U_j)a_i(U_i)b_j(U_j)+\sum_{j=1}^{n-1}\sum_{i=j+1}^{n}g(U_j,U_i)a_i(U_i)b_j(U_j)\right) : \mathbb{E}\left(\sum_{i=2}^{n}a_i(U_i)^2\right)\leqslant 1, \mathbb{E}\left(\sum_{j=1}^{n-1}b_j(U_j)^2\right)\leqslant 1\right\}.$$

But, with the conditions on the a_i's and the b_j's, we have:

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}g(U_i,U_j)a_i(U_i)b_j(U_j)\right)\\
&\leqslant \int_{\mathbb{R}}\mathbb{E}\left(\sum_{i=2}^{n}\big|\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big||a_i(U_i)|\right)\mathbb{E}\left(\sum_{j=1}^{n-1}\big|h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big||b_j(U_j)|\right)dt\\
&\leqslant \int_{\mathbb{R}}\sqrt{(n-1)\mathrm{Var}_\pi(\varphi_\lambda(t-U))}\,\mathbb{E}\left(\sum_{j=1}^{n-1}\big|h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big||b_j(U_j)|\right)dt\\
&\leqslant \sqrt{\frac{n-1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx}\;\mathbb{E}\left(\sum_{j=1}^{n-1}\int_{\mathbb{R}}\big|h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big||b_j(U_j)|\,dt\right)\\
&\leqslant \sqrt{\frac{n-1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx}\;\mathbb{E}\left(2\sum_{j=1}^{n-1}|b_j(U_j)|\int_{\mathbb{R}}h(x)\,dx\right)\\
&\leqslant 2(n-1)\sqrt{\frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx}\int_{\mathbb{R}}h(x)\,dx,
\end{aligned}$$

using equations (a) and (b) of Lemma 6.1. Inverting the a_i's and the b_j's, the same computations apply to the second term and we obtain

$$D\leqslant 4(n-1)\sqrt{\frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx}\int_{\mathbb{R}}h(x)\,dx, \tag{6.13}$$

- $B^2 = \max\left\{\sup_{u,i}\left(\sum_{j=1}^{i-1}\mathbb{E}(\mathcal{G}(u,U_j)^2)\right), \sup_{u,j}\left(\sum_{i=j+1}^{n}\mathbb{E}(\mathcal{G}(U_i,u)^2)\right)\right\}$ and since $\mathcal{G}$ is symmetric, we have:

$$\begin{aligned}
B^2 &= \sup_{u,i}\left(\sum_{j=1}^{i-1}\mathbb{E}(\mathcal{G}(u,U_j)^2)\right)\\
&= \sup_{u}\left(\sum_{j=1}^{n-1}\mathbb{E}(\mathcal{G}(u,U_j)^2)\right)\\
&\leqslant 2\sup_{u}\left(\sum_{j=1}^{n-1}\left[\mathbb{E}(g(u,U_j)^2)+\mathbb{E}(g(U_j,u)^2)\right]\right).
\end{aligned}$$

But,

$$
\begin{aligned}
&\mathbb{E}(g(u,U_j)^2)\\
&=\mathbb{E}\left[\left(\int_{\mathbb{R}}\left[\varphi_\lambda(t-u)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\left[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\right]dt\right)^2\right]\\
&\leqslant\mathbb{E}\left[\int_{\mathbb{R}}\left[\varphi_\lambda(t-u)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]^2\left|h(t-U_j)-\mathbb{E}_\pi(h(t-U))\right|dt\int_{\mathbb{R}}\left|h(t-U_j)-\mathbb{E}_\pi(h(t-U))\right|dt\right]\\
&\leqslant 2\mathbb{E}\left[\int_{\mathbb{R}}\left[\varphi_\lambda(t-u)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]^2\left|h(t-U_j)-\mathbb{E}_\pi(h(t-U))\right|dt\right]\int_{\mathbb{R}}h(x)\,dx\\
&\leqslant\frac{4}{T}\int_{\mathbb{R}}\left[\varphi_\lambda(t-u)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]^2dt\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\\
&\leqslant\frac{16}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2
\end{aligned}
$$

and in the same way

$$
\begin{aligned}
&\mathbb{E}(g(U_j,u)^2)\\
&=\mathbb{E}\left[\left(\int_{\mathbb{R}}\left[\varphi_\lambda(t-U_j)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\left[h(t-u)-\mathbb{E}_\pi(h(t-U))\right]dt\right)^2\right]\\
&\leqslant\mathbb{E}\left[\int_{\mathbb{R}}\left[\varphi_\lambda(t-U_j)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]^2\left|h(t-u)-\mathbb{E}_\pi(h(t-U))\right|dt\int_{\mathbb{R}}\left|h(t-u)-\mathbb{E}_\pi(h(t-U))\right|dt\right]\\
&\leqslant 2\mathbb{E}\left[\int_{\mathbb{R}}\left[\varphi_\lambda(t-U_j)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]^2\left|h(t-u)-\mathbb{E}_\pi(h(t-U))\right|dt\right]\int_{\mathbb{R}}h(x)\,dx\\
&\leqslant\frac{2}{T}\int_{\mathbb{R}}\left|h(t-u)-\mathbb{E}_\pi(h(t-U))\right|dt\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\int_{\mathbb{R}}h(x)\,dx\\
&\leqslant\frac{4}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2,
\end{aligned}
$$

by using Lemma 6.1. Hence,

$$
B^2\leqslant\frac{40(n-1)}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2. \tag{6.14}
$$

Finally, by inequalities (6.8), (6.9) and (6.10) combined with (6.11), (6.14), (6.12) and (6.13), we obtain: for any $\varepsilon_0>0$, with probability larger than $1-(5+2\times 2.77)e^{-\alpha}$,

$$
\begin{aligned}
|\hat\beta_\lambda-\beta_\lambda|&\leqslant\sqrt{2\alpha\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)}+\frac{\alpha}{3}B\left(\frac{\varphi_\lambda}{n}\right)\\
&\quad+\left\{\frac{\sqrt{2\alpha n}}{T}\|\varphi_\lambda\|_1+\frac{2\alpha}{3T}\|\varphi_\lambda\|_1+2(1+\varepsilon_0)^{3/2}\sqrt{2\alpha}\sqrt{\frac{1}{T}\|\varphi_\lambda\|_2^2+\frac{1}{T^2}\|\varphi_\lambda\|_1^2}\right.\\
&\qquad\left.+4\eta(\varepsilon_0)\alpha\sqrt{\frac{1}{T}\|\varphi_\lambda\|_2^2}+\beta(\varepsilon_0)\alpha^{3/2}\sqrt{\frac{40}{nT}\|\varphi_\lambda\|_2^2}+\frac{8}{n}\gamma(\varepsilon_0)\alpha^2\|\varphi_\lambda\|_\infty\right\}\|h\|_1\\
&\leqslant\sqrt{2\alpha\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)}+\frac{\alpha}{3}B\left(\frac{\varphi_\lambda}{n}\right)\\
&\quad+\left\{2M_{\psi,1}\frac{\sqrt{\alpha n}}{T}+2\sqrt{2}M_{\psi,1}\frac{\alpha}{3T}+2\sqrt{2}(1+\varepsilon_0)^{3/2}M_{\psi,2}\sqrt{\frac{\alpha}{T}}+4(1+\varepsilon_0)^{3/2}M_{\psi,1}\frac{\sqrt{\alpha}}{T}\right.\\
&\qquad\left.+4\eta(\varepsilon_0)M_{\psi,2}\frac{\alpha}{\sqrt{T}}+\sqrt{40}\beta(\varepsilon_0)M_{\psi,2}\frac{\alpha^{3/2}}{\sqrt{nT}}+8\gamma(\varepsilon_0)M_{\psi,\infty}\frac{2^{j_0/2}\alpha^2}{n}\right\}\|h\|_1,
\end{aligned}
$$

because $2^{-j/2} \leqslant \sqrt{2}$ if $-1 \leqslant j \leqslant j_0$. We denote b the quantity between braces above.

This upper bound depends on h (via $\|h\|_1$) and this potential threshold could not be used for applications because h is unknown. So we overestimate $\|h\|_1$ by $\frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n}$ and we have a threshold that does not depend on h. So, for any value of $\kappa \in]0;1[$, by fixing $\alpha = \kappa^2 j_0\gamma$ with $\gamma > 0$, we define for all λ in Γ,

$$\eta_\lambda(\gamma,\Delta) = \sqrt{2j_0\gamma\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)} + \frac{j_0\gamma}{3}B\left(\frac{\varphi_\lambda}{n}\right) + \Delta\frac{N_{\mathbb{R}}}{n},$$

where

$$\Delta = (1+\varepsilon_0)\Bigg\{2M_{\psi,1}\frac{\sqrt{j_0\gamma n}}{T} + 2\sqrt{2}M_{\psi,1}\frac{j_0\gamma}{3T} + 2\sqrt{2}(1+\varepsilon_0)^{3/2}M_{\psi,2}\sqrt{\frac{j_0\gamma}{T}} + 4(1+\varepsilon_0)^{3/2}M_{\psi,1}\frac{\sqrt{j_0\gamma}}{T}$$
$$+ 4\eta(\varepsilon_0)M_{\psi,2}\frac{j_0\gamma}{\sqrt{T}} + \sqrt{40}\beta(\varepsilon_0)M_{\psi,2}\frac{j_0^{3/2}\gamma^{3/2}}{\sqrt{nT}} + 8\gamma(\varepsilon_0)M_{\psi,\infty}\frac{2^{j_0/2}j_0^2\gamma^2}{n}\Bigg\}.$$

Thus, for all λ in Γ,

$$\begin{aligned}
&\mathbb{P}\big(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa\eta_\lambda(\gamma,\Delta)\big)\\
&\leqslant \mathbb{P}\left(|\hat{\beta}_\lambda - \beta_\lambda| > \sqrt{2\alpha\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)} + \frac{\alpha}{3}B\left(\frac{\varphi_\lambda}{n}\right) + b\frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n}\,,\, \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} > \|h\|_1\right)\\
&\quad + \mathbb{P}\left(|\hat{\beta}_\lambda - \beta_\lambda| > \sqrt{2\alpha\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)} + \frac{\alpha}{3}B\left(\frac{\varphi_\lambda}{n}\right) + b\frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n}\,,\, \frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leqslant \|h\|_1\right)\\
&\leqslant \mathbb{P}\left(|\hat{\beta}_\lambda - \beta_\lambda| > \sqrt{2\alpha\widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)} + \frac{\alpha}{3}B\left(\frac{\varphi_\lambda}{n}\right) + b\|h\|_1\right) + \mathbb{P}\left(\frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leqslant \|h\|_1\right)\\
&\leqslant (5 + 2\times 2.77)e^{-\alpha} + \mathbb{P}\left(\frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leqslant \|h\|_1\right),
\end{aligned}$$

with

$$\begin{aligned}
\mathbb{P}\left(\frac{(1+\varepsilon_0)N_{\mathbb{R}}}{n} \leqslant \|h\|_1\right) &= \mathbb{P}\left(N_{\mathbb{R}} - n\|h\|_1 \leqslant -\frac{\varepsilon_0 n\|h\|_1}{1+\varepsilon_0}\right)\\
&\leqslant \exp\left(-g(\varepsilon_0)n\|h\|_1\right),
\end{aligned}$$

using Proposition 7 of [25] with $g(\varepsilon_0) = \frac{1}{1+\varepsilon_0}\left(\log\frac{1}{1+\varepsilon_0} - 1\right) + 1$.

Therefore, Assumption (A1) is true if we take $\omega = (5 + 2\times 2.77)e^{-\kappa^2 j_0\gamma} + \exp\left(-g(\varepsilon_0)n\|h\|_1\right)$, with $\gamma > 0$ and $\varepsilon_0 > 0$. Furthermore, the threshold (2.4) that lies at the heart of the paper is achieved by rewriting Δ by grouping the constants into one:

$$\Delta = d(\gamma, \|\psi\|_1, \|\psi\|_2, \|\psi\|_\infty)\left\{\frac{j_0^2 2^{j_0/2}}{n} + \frac{j_0}{\sqrt{T}} + \frac{\sqrt{j_0 n}}{T}\right\} \tag{6.15}$$

with

$$\begin{aligned}
&d(\gamma, \|\psi\|_1, \|\psi\|_2, \|\psi\|_\infty)\\
&= (1+\varepsilon_0)\Bigg\{2\sqrt{\gamma}M_{\psi,1} + \frac{2\sqrt{2}}{3}\gamma M_{\psi,1} + 2\sqrt{2}(1+\varepsilon_0)^{3/2}\sqrt{\gamma}M_{\psi,2} + 4(1+\varepsilon_0)^{3/2}\sqrt{\gamma}M_{\psi,1}\\
&\qquad + 4\eta(\varepsilon_0)\gamma M_{\psi,2} + \sqrt{40}\beta(\varepsilon_0)\gamma^{3/2}M_{\psi,2} + 8\gamma(\varepsilon_0)\gamma^2 M_{\psi,\infty}\Bigg\},
\end{aligned} \tag{6.16}$$

where $\beta(\varepsilon_0)$, $\gamma(\varepsilon_0)$ and $\eta(\varepsilon_0)$ are defined in [14] with $\varepsilon_0 = 1$.

6.4.2 Proof of Assumption (A2)

Let $\lambda \in \Gamma$ be fixed. For any $p \geqslant 1$,

$$\begin{aligned}\mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}) &= \mathbb{E}\left(\left|G\left(\frac{\varphi_\lambda}{n}\right) - \int_{\mathbb{R}} \varphi_\lambda(x)h(x)\,dx\right|^{2p}\right) \\ &= \frac{1}{n^{2p}}\mathbb{E}\left(\left|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n) + W(\varphi_\lambda)\right|^{2p}\right) \\ &\leqslant \frac{2^{2p-1}}{n^{2p}}\left[\mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)|^{2p}) + \mathbb{E}(|W(\varphi_\lambda)|^{2p})\right]. \qquad (6.17)\end{aligned}$$

Now, let us give an upper bound of each term of the right-hand side of the previous inequality.

We first study the first term of (6.17). We have:

$$\mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)|^{2p}) = \mathbb{E}\left[\mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)|^{2p})\,|\,U_1,\dots,U_n)\right]$$

and conditionally to the U_i's, N is a Poisson process. We apply Lemma 6.2: for any $p \geqslant 1$, there exists a positive constant $C(p)$ only depending on p such that

$$\begin{aligned}&\mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)|^{2p})\,|\,U_1,\dots,U_n) \\ &\leqslant C(p)\left(\int_{\mathbb{R}}\left|\sum_{i=1}^n\left[\varphi_\lambda(t-U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}\sum_{j=1}^n h(t-U_j)\,dt + V(\varphi_\lambda)^p\right). \qquad (6.18)\end{aligned}$$

On the one hand, we provide a control in expectation of the first term of (6.18). We have:

$$\begin{aligned}&\mathbb{E}\left(\int_{\mathbb{R}}\left|\sum_{i=1}^n\left[\varphi_\lambda(t-U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}\sum_{j=1}^n h(t-U_j)\,dt\right) \\ &= \mathbb{E}\left(\int_{\mathbb{R}}\sum_{j=1}^n\left|\varphi_\lambda(t-U_j) + \sum_{i\neq j}\left[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}h(t-U_j)\,dt\right) \\ &\leqslant 2^{2p-1}\left[\mathbb{E}\left(\int_{\mathbb{R}}\sum_{j=1}^n|\varphi_\lambda(t-U_j)|^{2p}h(t-U_j)\,dt\right)\right. \\ &\qquad\left. + \mathbb{E}\left(\int_{\mathbb{R}}\sum_{j=1}^n\left|\sum_{i\neq j}\left[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}h(t-U_j)\,dt\right)\right],\end{aligned}$$

with

$$\mathbb{E}\left(\int_{\mathbb{R}}\sum_{j=1}^n|\varphi_\lambda(t-U_j)|^{2p}h(t-U_j)\,dt\right) = n\int_{\mathbb{R}}|\varphi_\lambda(x)|^{2p}h(x)\,dx \leqslant n\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\|\varphi_\lambda\|_\infty^{2p-2}\|h\|_\infty$$

and

$$\begin{aligned}&\mathbb{E}\left(\int_{\mathbb{R}}\sum_{j=1}^n\left|\sum_{i\neq j}\left[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}h(t-U_j)\,dt\right) \\ &= \sum_{j=1}^n\int_{\mathbb{R}}\mathbb{E}\left(\left|\sum_{i\neq j}\left[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}\right)\mathbb{E}(h(t-U_j))\,dt \\ &= n\int_{\mathbb{R}}\mathbb{E}\left(\left|\sum_{i=1}^{n-1}\left[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}\right)\mathbb{E}_\pi(h(t-U))\,dt.\end{aligned}$$

By applying Rosenthal's inequality, there exists a positive constant $C(p)$ only depending on p such that

$$\begin{aligned}&\mathbb{E}\left(\left|\sum_{i=1}^{n-1}\left[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}\right)\\&\leqslant C(p)\left((n-1)\mathbb{E}\left(\left|\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right|^{2p}\right)+(n-1)^p\left[\mathrm{Var}_\pi(\varphi_\lambda(t-U))\right]^p\right).\end{aligned}$$

But,

$$\begin{aligned}\mathbb{E}\left(\left|\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right|^{2p}\right)&\leqslant K\ \left(\mathbb{E}_\pi(|\varphi_\lambda(t-U)|^{2p})+|\mathbb{E}_\pi(\varphi_\lambda(t-U))|^{2p}\right)\\&\leqslant K\ \left(\frac{1}{T}\int_0^T|\varphi_\lambda(t-u)|^{2p}\,du+\left|\frac{1}{T}\int_0^T\varphi_\lambda(t-u)\,du\right|^{2p}\right)\\&\leqslant K\ \left(\frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\ \|\varphi_\lambda\|_\infty^{2p-2}+\frac{1}{T^{2p}}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^{2p}\right),\end{aligned}$$

with K a positive constant only depending on p and using inequality (a) of Lemma 6.1,

$$\mathrm{Var}_\pi(\varphi_\lambda(t-U))\leqslant\frac{1}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx.$$

Thus,

$$\begin{aligned}&\mathbb{E}\left(\int_{\mathbb{R}}\sum_{j=1}^{n}\left|\sum_{i\neq j}\left[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}h(t-U_j)\,dt\right)\\&\leqslant n\ K\ \left(\frac{n-1}{T}\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}+\frac{n-1}{T^{2p}}\|\varphi_\lambda\|_1^{2p}+\frac{(n-1)^p}{T^p}\|\varphi_\lambda\|_2^{2p}\right)\int_{\mathbb{R}}\mathbb{E}_\pi(h(t-U))\,dt\\&\leqslant K\ \left(\frac{n^2}{T}\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}+\frac{n^2}{T^{2p}}\|\varphi_\lambda\|_1^{2p}+\frac{n^{p+1}}{T^p}\|\varphi_\lambda\|_2^{2p}\right)\|h\|_1,\end{aligned}$$

using equation (b) of Lemma 6.1.

Therefore, we have the following control of the first term of (6.18)

$$\begin{aligned}&\mathbb{E}\left(\int_{\mathbb{R}}\left|\sum_{i=1}^{n}\left[\varphi_\lambda(t-U_i)-\frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]\right|^{2p}\sum_{j=1}^{n}h(t-U_j)\,dt\right)\\&\leqslant K\left[n\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}\|h\|_\infty+\left(\frac{n^2}{T}\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}+\frac{n^2}{T^{2p}}\|\varphi_\lambda\|_1^{2p}+\frac{n^{p+1}}{T^p}\|\varphi_\lambda\|_2^{2p}\right)\|h\|_1\right],\end{aligned}\tag{6.19}$$

Now let us provide a control in expectation of the second term of (6.18), i.e $V(\varphi_\lambda)^p$. First, we recall that $V(\varphi_\lambda)=\mathrm{Var}(G(\varphi_\lambda)|U_1,\dots,U_n)$ and we remark that $\mathbb{E}(V(\varphi_\lambda)^p)\leqslant\left[\mathbb{E}(V(\varphi_\lambda)^{2p})\right]^{1/2}$ (using the Cauchy-Schwarz inequality). So, we focus on the moments of $V(\varphi_\lambda)$ of any order $m\geqslant 2$.

Let $m \geqslant 2$. According to the expression (6.2) of $V(\varphi_\lambda)$, we have:

$$
\begin{aligned}
&V(\varphi_\lambda)\\
&= n\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + 2\sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\varphi_\lambda(t-U_j)h(t-U_j)\,dt\\
&\quad+\sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i\neq j}\sum_{k\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]h(t-U_j)\,dt\\
&= n\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + 2\sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U)))\,dt\\
&\quad+2\sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_j)h(t-U_j)-\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\big]\,dt\\
&\quad+\sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i\neq j}\sum_{k\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\\
&\quad+\sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i\neq j}\sum_{k\neq j}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\\
&\qquad\qquad\times\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt.
\end{aligned}
$$

This formula provides a decomposition of $V(\varphi_\lambda)$ in a sum of degenerate U-statistics of order 0, 1, 2 and 3. Indeed

$$V(\varphi_\lambda) = \mathcal{W}_0(\varphi_\lambda) + \mathcal{W}_1(\varphi_\lambda) + \mathcal{W}_2(\varphi_\lambda) + \mathcal{W}_3(\varphi_\lambda),$$

with $\mathcal{W}_i(\varphi_\lambda)$ is a degenerate U-statistic of order i defined as follows:

$$
\begin{aligned}
\mathcal{W}_3(\varphi_\lambda) = \sum_{1\leqslant i\neq j\neq k\leqslant n}\int_{\mathbb{R}}&\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\\
&\times\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt,
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{W}_2(\varphi_\lambda)\\
&= 2\sum_{1\leqslant i\neq j\leqslant n}\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_j)h(t-U_j)-\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\big]\,dt\\
&\quad+(n-2)\sum_{1\leqslant i\neq k\leqslant n}\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\\
&\quad+\sum_{1\leqslant i\neq j\leqslant n}\int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2-\mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}_1(\varphi_\lambda) = 2(n-1)&\sum_{i=1}^{n}\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt\\
&+(n-1)\sum_{i=1}^{n}\int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2-\mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\mathbb{E}_\pi(h(t-U))\,dt\\
&+(n-1)\sum_{j=1}^{n}\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt
\end{aligned}
$$

and

$$\begin{aligned}\mathcal{W}_0(\varphi_\lambda) &= n\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + n(n-1)\int_{\mathbb{R}} \mathrm{Var}_\pi(\varphi_\lambda(t-U))\mathbb{E}_\pi(h(t-U))\,dt \\ &= \mathbb{E}(V(\varphi_\lambda)) \leqslant n\int_{\mathbb{R}} \varphi_\lambda^2(x)h(x)\,dx + \frac{n(n-1)}{T}\int_{\mathbb{R}} \varphi_\lambda^2(x)\,dx\int_{\mathbb{R}} h(x)\,dx, \end{aligned} \tag{6.20}$$

by using (6.3) and (6.4).

First, we are interested in the moments of $\mathcal{W}_1(\varphi_\lambda)$ that we write:

$$\mathcal{W}_1(\varphi_\lambda) = \mathcal{W}_{1,1}(\varphi_\lambda) + \mathcal{W}_{1,2}(\varphi_\lambda) + \mathcal{W}_{1,3}(\varphi_\lambda),$$

with:

- $\mathcal{W}_{1,1}(\varphi_\lambda) = 2(n-1)\sum_{i=1}^{n}\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt.$

We have:

$$\begin{aligned}&\mathbb{E}(|\mathcal{W}_{1,1}(\varphi_\lambda)|^m) \\ &= 2^m(n-1)^m\mathbb{E}\left(\left|\sum_{i=1}^{n}\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt\right|^m\right) \\ &\leqslant 2^m(n-1)^m \times C(m)\Bigg(n\mathbb{E}\left(\left|\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt\right|^m\right) \\ &\qquad + n^{m/2}\left[\mathrm{Var}\left(\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt\right)\right]^{m/2}\Bigg),\end{aligned}$$

using Rosenthal's inequality, where $C(m)$ is a positive constant only depending on m. But, applying Lemma 6.1,

$$\begin{aligned}&\mathbb{E}\left(\left|\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt\right|^m\right) \\ &\leqslant \frac{2^m}{T^m}\left(\int_{\mathbb{R}} |\varphi_\lambda(x)|\,dx\right)^m\left(\int_{\mathbb{R}} |\varphi_\lambda(x)|h(x)\,dx\right)^m\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Var}\left(\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt\right) \\ &= \mathrm{Var}\left(\int_{\mathbb{R}} \varphi_\lambda(t-U_1)\mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\,dt\right) \\ &\leqslant \mathbb{E}\left[\left(\int_{\mathbb{R}} |\varphi_\lambda(t-U_1)|\mathbb{E}_\pi(|\varphi_\lambda(t-U)|h(t-U))\,dt\right)^2\right] \\ &\leqslant \frac{1}{T^2}\left(\int_{\mathbb{R}} |\varphi_\lambda(x)|\,dx\right)^2\left(\int_{\mathbb{R}} |\varphi_\lambda(x)|h(x)\,dx\right)^2.\end{aligned}$$

So,

$$\begin{aligned}\mathbb{E}(|\mathcal{W}_{1,1}(\varphi_\lambda)|^m) &\leqslant 2^m(n-1)^m \times C(m)\Bigg(\frac{2^m n}{T^m}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^m\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|h(x)\,dx\right)^m \\ &\qquad + \frac{n^{m/2}}{T^m}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^m\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|h(x)\,dx\right)^m\Bigg) \\ &\leqslant K_{1,1}\,\frac{n^{3m/2}}{T^m}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^m\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|h(x)\,dx\right)^m \\ &\leqslant K_{1,1}\,\frac{n^{3m/2}}{T^m}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^{2m}\|h\|_\infty^m, \end{aligned} \tag{6.21}$$

with $K_{1,1}$ a positive constant only depending on m.

- $\mathcal{W}_{1,2}(\varphi_\lambda) = (n-1)\sum_{i=1}^n \int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2 - \mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\mathbb{E}_\pi(h(t-U))\,dt.$

We have:

$$\begin{aligned}&\mathbb{E}(|\mathcal{W}_{1,2}(\varphi_\lambda)|^m) \\ &= (n-1)^m\mathbb{E}\left(\left|\sum_{i=1}^n\int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2 - \mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\mathbb{E}_\pi(h(t-U))\,dt\right|^m\right) \\ &\leqslant n^m \times C(m)\Bigg(n\mathbb{E}\left(\left|\int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2 - \mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\mathbb{E}_\pi(h(t-U))\,dt\right|^m\right) \\ &\qquad + n^{m/2}\left[\mathrm{Var}\left(\int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2 - \mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\mathbb{E}_\pi(h(t-U))\,dt\right)\right]^{m/2}\Bigg),\end{aligned}$$

using Rosenthal's inequality, where $C(m)$ is a positive constant only depending on m. But, applying Lemma 6.1,

$$\begin{aligned}&\mathbb{E}\left(\left|\int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2 - \mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\mathbb{E}_\pi(h(t-U))\,dt\right|^m\right) \\ &\leqslant K_{1,2}\,\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m,\end{aligned}$$

with $K_{1,2}$ a positive constant only depending on m and

$$\begin{aligned}&\mathrm{Var}\left(\int_{\mathbb{R}}\Big[\big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2 - \mathrm{Var}_\pi(\varphi_\lambda(t-U))\Big]\mathbb{E}_\pi(h(t-U))\,dt\right) \\ &= \mathrm{Var}\left(\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\mathbb{E}_\pi(h(t-U))\,dt\right) \\ &\leqslant \mathbb{E}\left[\left(\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\mathbb{E}_\pi(h(t-U))\,dt\right)^2\right] \\ &\leqslant \frac{1}{T^2}\mathbb{E}\left[\left(\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\,dt\right)^2\right]\left(\int_{\mathbb{R}}h(x)\,dx\right)^2 \\ &\leqslant K_{1,2}\,\frac{1}{T^2}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\end{aligned}$$

So,

$$\begin{aligned}
&\mathbb{E}(|\mathcal{W}_{1,2}(\varphi_\lambda)|^m)\\
&\leqslant K_{1,2}\ n^m\left(\frac{n}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m+\frac{n^{m/2}}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m\right)\\
&\leqslant K_{1,2}\ \frac{n^{3m/2}}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m. \qquad (6.22)
\end{aligned}$$

- $\mathcal{W}_{1,3}(\varphi_\lambda)=(n-1)\sum_{j=1}^{n}\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt.$

We have:

$$\begin{aligned}
&\mathbb{E}(|\mathcal{W}_{1,3}(\varphi_\lambda)|^m)\\
&=(n-1)^m\mathbb{E}\left(\left|\sum_{j=1}^{n}\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right|^m\right)\\
&\leqslant (n-1)^m\times C(m)\Bigg(n\mathbb{E}\left(\left|\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\big[h(t-U_1)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right|^m\right)\\
&\qquad\qquad+n^{m/2}\left[\mathrm{Var}\left(\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\big[h(t-U_1)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right)\right]^{m/2}\Bigg),
\end{aligned}$$

using Rosenthal's inequality, where $C(m)$ is a positive constant only depending on m. But, applying Lemma 6.1,

$$\mathbb{E}\left(\left|\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\big[h(t-U_1)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right|^m\right)\leqslant\frac{2^m}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m,$$

and

$$\begin{aligned}
&\mathrm{Var}\left(\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))\big[h(t-U_1)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right)\\
&=\mathrm{Var}\left(\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))h(t-U_1)\,dt\right)\\
&\leqslant\mathbb{E}\left[\left(\int_{\mathbb{R}}\mathrm{Var}_\pi(\varphi_\lambda(t-U))h(t-U_1)\,dt\right)^2\right]\\
&\leqslant\frac{1}{T^2}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^2\mathbb{E}\left[\left(\int_{\mathbb{R}}h(t-U_1)\,dt\right)^2\right]\\
&\leqslant\frac{1}{T^2}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^2.
\end{aligned}$$

So,

$$\begin{aligned}
\mathbb{E}(|\mathcal{W}_{1,3}(\varphi_\lambda)|^m)&\leqslant(n-1)^m\times C(m)\left(\frac{2^m n}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m\right.\\
&\qquad\left.+\frac{n^{m/2}}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m\right)\\
&\leqslant K_{1,3}\ \frac{n^{3m/2}}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m, \qquad (6.23)
\end{aligned}$$

with $K_{1,3}$ a positive constant only depending on m.

Next we deal with the moments of $\mathcal{W}_2(\varphi_\lambda)$ that we write:

$$\mathcal{W}_2(\varphi_\lambda) = \mathcal{W}_{2,1}(\varphi_\lambda) + \mathcal{W}_{2,2}(\varphi_\lambda) + \mathcal{W}_{2,3}(\varphi_\lambda),$$

with:

- $\mathcal{W}_{2,1}(\varphi_\lambda) = 2 \sum_{1\leqslant i\neq j\leqslant n} \int_{\mathbb{R}} [\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]\big[\varphi_\lambda(t-U_j)h(t-U_j) - \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\big]\, dt.$

We want to use Theorem 8.1.6 of [6] (a moment inequality for U-statistics using decoupling) so we write:

$$\mathcal{W}_{2,1}(\varphi_\lambda) = 2 \sum_{1\leqslant i\neq j\leqslant n} f(U_i, U_j),$$

where

$$f(U_i, U_j) = \int_{\mathbb{R}} \big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_j)h(t-U_j) - \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\big]\, dt.$$

There exists a positive constant $C_{2,m}$ depending on m only such that

$$\begin{aligned}
&\mathbb{E}\left(\left|\sum_{1\leqslant i\neq j\leqslant n} f(U_i, U_j)\right|^m\right)\\
&\leqslant C_{2,m} n^m \mathbb{E}(|f(U_1, U_2)|^m)\\
&\leqslant C_{2,m} n^m \mathbb{E}\left(\left|\int_{\mathbb{R}} [\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))]\big[\varphi_\lambda(t-U_2)h(t-U_2) - \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\big]\, dt\right|^m\right)\\
&\leqslant C_{2,m} n^m \mathbb{E}\left[\left(\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\, dt\right)^{m/2}\right]\\
&\quad \times \mathbb{E}\left[\left(\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_2)h(t-U_2) - \mathbb{E}_\pi(\varphi_\lambda(t-U)h(t-U))\big]^2\, dt\right)^{m/2}\right]\\
&\leqslant K_{2,1}\ n^m \left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^{m/2} \left(\int_{\mathbb{R}} \varphi_\lambda^2(x) h^2(x)\, dx\right)^{m/2}\\
&\leqslant K_{2,1}\ n^m \left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^{m} \|h\|_\infty^m,
\end{aligned}$$

by applying Lemma 6.1 and setting $K_{2,1}$ a positive constant only depending on m. So,

$$\mathbb{E}(|\mathcal{W}_{2,1}(f)|^m) \leqslant K_{2,1}\ n^m \left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^{m} \|h\|_\infty^m. \tag{6.24}$$

- We may write: $\mathcal{W}_{2,2}(\varphi_\lambda) = (n-2) \sum_{1\leqslant i\neq k\leqslant n} f(U_i, U_k)$, where

$$f(U_i, U_k) = \int_{\mathbb{R}} \big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\, dt.$$

We use Theorem 8.1.6 of [6]: there exists a positive constant $C_{2,m}$ depending on m only such that

$$\begin{aligned}
&\mathbb{E}\left(\left|\sum_{1\leqslant i\neq k\leqslant n} f(U_i,U_k)\right|^m\right)\\
&\leqslant C_{2,m}n^m\mathbb{E}(|f(U_1,U_2)|^m)\\
&\leqslant C_{2,m}n^m\mathbb{E}\left(\left|\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_2)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right|^m\right)\\
&\leqslant K_{2,2}\,\frac{n^m}{T^m}\left\{\mathbb{E}\left[\left(\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\,dt\right)^{m/2}\right]\right\}^2\left(\int_{\mathbb{R}}h(x)\,dx\right)^m\\
&\leqslant K_{2,2}\,\frac{n^m}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m,
\end{aligned}$$

by applying Lemma 6.1 and setting $K_{2,2}$ a positive constant only depending on m. So,

$$\mathbb{E}(|\mathcal{W}_{2,2}(f)|^m)\leqslant K_{2,2}\,\frac{n^{2m}}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\left(\int_{\mathbb{R}}h(x)\,dx\right)^m. \tag{6.25}$$

- We may write: $\mathcal{W}_{2,3}(\varphi_\lambda)=\sum_{1\leqslant i\neq j\leqslant n} f(U_i,U_j)$, where

$$f(U_i,U_j)=\int_{\mathbb{R}}\left[\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2-\mathrm{Var}_\pi(\varphi_\lambda(t-U))\right]\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt.$$

We use Theorem 8.1.6 of [6]: there exists a positive constant $C_{2,m}$ depending on m only such that

$$\begin{aligned}
&\mathbb{E}\left(\left|\sum_{1\leqslant i\neq j\leqslant n} f(U_i,U_j)\right|^m\right)\\
&\leqslant C_{2,m}n^m\mathbb{E}(|f(U_1,U_2)|^m)\\
&\leqslant C_{2,m}n^m\mathbb{E}\left(\left|\int_{\mathbb{R}}[[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))]^2-\mathrm{Var}_\pi(\varphi_\lambda(t-U))][h(t-U_2)-\mathbb{E}_\pi(h(t-U))]\,dt\right|^m\right)\\
&\leqslant K_{2,3}\;n^m\mathbb{E}\left[\left(\int_{\mathbb{R}}\left|\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2-\mathrm{Var}_\pi(\varphi_\lambda(t-U))\right|\,dt\right)^m\right]\|h\|_\infty^m\\
&\leqslant K_{2,3}\;n^m\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\|h\|_\infty^m,
\end{aligned}$$

by applying Lemma 6.1 and setting $K_{2,3}$ a positive constant only depending on m. So,

$$\mathbb{E}(|\mathcal{W}_{2,3}(\varphi_\lambda)|^m)\leqslant K_{2,3}\;n^m\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\|h\|_\infty^m. \tag{6.26}$$

And finally, we focus on the moments of $\mathcal{W}_3(\varphi_\lambda)$ that we write: $\mathcal{W}_3(\varphi_\lambda)=\sum_{1\leqslant i\neq j\neq k\leqslant n} f(U_i,U_j,U_k)$, where

$$f(U_i,U_j,U_k)=\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_k)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt.$$

We use Theorem 8.1.6 of [6]: there exists a positive constant $C_{3,m}$ depending on m only such that

$$\begin{aligned}
&\mathbb{E}\left(\left|\sum_{1\leqslant i\neq j\neq k\leqslant n} f(U_i,U_j,U_k)\right|^m\right)\\
&\leqslant C_{3,m}n^{3m/2}\mathbb{E}(|f(U_1,U_2,U_3)|^m)\\
&\leqslant C_{3,m}n^{3m/2}\mathbb{E}\Bigg(\bigg|\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_2)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\\
&\qquad\qquad\times\big[h(t-U_3)-\mathbb{E}_\pi(h(t-U))\big]\,dt\bigg|^m\Bigg)\\
&\leqslant K_3\ n^{3m/2}\mathbb{E}\left[\left(\int_{\mathbb{R}}\Big|\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_2)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\Big|\,dt\right)^m\right]\|h\|_\infty^m,
\end{aligned}$$

by applying Lemma 6.1 and setting K_3 a positive constant only depending on m. Furthermore, using the support properties of the biorthogonal wavelet bases considered in this paper, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{\mathbb{R}}\Big|\big[\varphi_\lambda(t-U_1)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[\varphi_\lambda(t-U_2)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\Big|\,dt\right)^m\right]\\
&=\mathbb{E}\Bigg[\bigg(\int_{\mathbb{R}}\Big|\varphi_\lambda(t-U_1)\varphi_\lambda(t-U_2)-\varphi_\lambda(t-U_1)\mathbb{E}_\pi(\varphi_\lambda(t-U))-\varphi_\lambda(t-U_2)\mathbb{E}_\pi(\varphi_\lambda(t-U))\\
&\qquad\qquad+\big[\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\Big|\,dt\bigg)^m\Bigg]\\
&\leqslant K_3\left\{\mathbb{E}\left[\left(\int_{\mathbb{R}}\Big|\varphi_\lambda(t-U_1)\varphi_\lambda(t-U_2)\Big|\,dt\right)^m\right]+\mathbb{E}\left[\left(\int_{\mathbb{R}}\Big|\varphi_\lambda(t-U_1)\mathbb{E}_\pi(\varphi_\lambda(t-U))\Big|\,dt\right)^m\right]\right.\\
&\qquad\qquad\left.+\mathbb{E}\left[\left(\int_{\mathbb{R}}\big[\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\,dt\right)^m\right]\right\},
\end{aligned}$$

with:

$$\begin{aligned}
\mathbb{E}\left[\left(\int_{\mathbb{R}}\Big|\varphi_\lambda(t-U_1)\varphi_\lambda(t-U_2)\Big|\,dt\right)^m\right]&=\frac{1}{T^2}\int_0^T du_1\int_0^T du_2\left(\int_{\mathbb{R}}\big|\varphi_\lambda(t-u_1)\varphi_\lambda(t-u_2)\big|\,dt\right)^m\\
&\leqslant\frac{1}{T^2}\int_0^T du_1\int_{u_1-2M}^{u_1+2M}du_2\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\\
&\leqslant\frac{4M}{T}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\left(\int_{\mathbb{R}}\Big|\varphi_\lambda(t-U_1)\mathbb{E}_\pi(\varphi_\lambda(t-U))\Big|\,dt\right)^m\right]&=\mathbb{E}\left[\left(\int_{\mathbb{R}}\Big|\varphi_\lambda(t-U_1)\left(\frac{1}{T}\int_0^T\varphi_\lambda(t-u)\,du\right)\Big|\,dt\right)^m\right]\\
&\leqslant\frac{1}{T^m}\mathbb{E}\left[\left(\int_0^T du\int_{\mathbb{R}}\big|\varphi_\lambda(t-U_1)\varphi_\lambda(t-u)\big|\,dt\right)^m\right]\\
&\leqslant\frac{1}{T^{m+1}}\int_0^T du_1\left(\int_0^T du\int_{\mathbb{R}}\big|\varphi_\lambda(t-u_1)\varphi_\lambda(t-u)\big|\,dt\right)^m\\
&\leqslant\frac{1}{T^{m+1}}\int_0^T du_1\left(\int_{u_1-2M}^{u_1+2M}du\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m\\
&\leqslant\frac{(4M)^m}{T^m}\left(\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\right)^m
\end{aligned}$$

and

$$\begin{aligned}\int_{\mathbb{R}} \left[\mathbb{E}_\pi(\varphi_\lambda(t-U))\right]^2 dt &= \frac{1}{T^2}\int_{\mathbb{R}} dt \int_0^T du\, \varphi_\lambda(t-u) \int_0^T dv\, \varphi_\lambda(t-v) \\ &= \frac{1}{T^2}\int_0^T du \int_0^T dv \int_{\mathbb{R}} \varphi_\lambda(t-u)\varphi_\lambda(t-v)\, dt \\ &\leqslant \frac{1}{T^2}\int_0^T du \int_{u-2M}^{u+2M} dv \int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx \\ &\leqslant \frac{4M}{T}\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx.\end{aligned}$$

So,

$$\mathbb{E}(|\mathcal{W}_3(\varphi_\lambda)|^m) \leqslant K_3'\, n^{3m/2}\left[\frac{1}{T}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m + \frac{1}{T^m}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\right]\|h\|_\infty^m, \tag{6.27}$$

with K_3' a positive constant only depending on m and the compact support of ψ. Note that if we had used the same method as for the control of the moments of $W_{2,2}(\varphi_\lambda)$, we would not get the correct rate of convergence. We obtain a better rate of convergence thanks to the properties of the biorthogonal wavelet bases used here.

Thus, combining inequalities (6.20), (6.21), (6.22), (6.23), (6.24), (6.25), (6.26) and (6.27) yields

$$\begin{aligned}&\mathbb{E}(V(\varphi_\lambda)^m) \\ &\leqslant K\left\{n^m\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m \|h\|_\infty^m + \frac{n^{2m}}{T^m}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\left(\int_{\mathbb{R}} h(x)\, dx\right)^m\right. \\ &\qquad + \frac{n^{3m/2}}{T^m}\left(\int_{\mathbb{R}} |\varphi_\lambda(x)|\, dx\right)^{2m}\|h\|_\infty^m + \frac{n^{3m/2}}{T^m}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\left(\int_{\mathbb{R}} h(x)\, dx\right)^m \\ &\qquad + \frac{n^{3m/2}}{T^m}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\left(\int_{\mathbb{R}} h(x)\, dx\right)^m + n^m\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\|h\|_\infty^m \\ &\qquad + \frac{n^{2m}}{T^m}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\left(\int_{\mathbb{R}} h(x)\, dx\right)^m + n^m\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\|h\|_\infty^m \\ &\qquad \left. + n^{3m/2}\left[\frac{1}{T}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m + \frac{1}{T^m}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\, dx\right)^m\right]\|h\|_\infty^m\right\} \\ &\leqslant K\left\{\frac{n^{3m/2}}{T^m}\|\varphi_\lambda\|_1^{2m}\|h\|_\infty^m + \frac{n^{2m}}{T^m}\|\varphi_\lambda\|_2^{2m}\|h\|_1^m + \left[n^m + \frac{n^{3m/2}}{T}\right]\|\varphi_\lambda\|_2^{2m}\|h\|_\infty^m\right\},\end{aligned}$$

with K a positive constant only depending on m and the compact support of ψ. So, we obtain

$$\mathbb{E}(V(\varphi_\lambda)^p) \leqslant K\left\{\frac{n^{3p/2}}{T^p}\|\varphi_\lambda\|_1^{2p}\|h\|_\infty^p + \frac{n^{2p}}{T^p}\|\varphi_\lambda\|_2^{2p}\|h\|_1^p + \left[n^p + \frac{n^{3p/2}}{T}\right]\|\varphi_\lambda\|_2^{2p}\|h\|_\infty^p\right\}, \tag{6.28}$$

with K a positive constant only depending on p and the compact support of ψ.

To conclude for the first term of (6.17), using inequalities (6.19) and (6.28) in (6.18), we have

$$\begin{aligned}&\mathbb{E}(|G(\varphi_\lambda) - \mathbb{E}(G(\varphi_\lambda)|U_1,\dots,U_n)|^{2p}) \\ &\leqslant K\left\{\frac{n^2}{T^{2p}}\|\varphi_\lambda\|_1^{2p}\|h\|_1 + \frac{n^{p+1}}{T^p}\|\varphi_\lambda\|_2^{2p}\|h\|_1 + \frac{n^2}{T}\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}\|h\|_1 + n\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}\|h\|_\infty\right. \\ &\qquad \left. + \frac{n^{3p/2}}{T^p}\|\varphi_\lambda\|_1^{2p}\|h\|_\infty^p + \frac{n^{2p}}{T^p}\|\varphi_\lambda\|_2^{2p}\|h\|_1^p + \left[n^p + \frac{n^{3p/2}}{T}\right]\|\varphi_\lambda\|_2^{2p}\|h\|_\infty^p\right\}.\end{aligned} \tag{6.29}$$

Now, we have to focus on the second term of (6.17). Recall the definition (6.5) of $W(\varphi_\lambda)$

$$W(\varphi_\lambda) = W_1(\varphi_\lambda) + W_2(\varphi_\lambda),$$

with

$$W_1(\varphi_\lambda) = (n-1)\sum_{i=1}^{n}\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt,$$

and

$$W_2(\varphi_\lambda) = \sum_{1\leqslant i\neq j\leqslant n} g(U_i,U_j),$$

where

$$g(U_i,U_j) = \int_{\mathbb{R}} \big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[h(t-U_j) - \mathbb{E}_\pi(h(t-U))\big]\,dt.$$

So,

$$\mathbb{E}(|W(\varphi_\lambda)|^{2p}) \leqslant 2^{2p-1}\big[\mathbb{E}(|W_1(\varphi_\lambda)|^{2p}) + \mathbb{E}(|W_2(\varphi_\lambda)|^{2p})\big].$$

On the one hand, we have to control $\mathbb{E}(|W_1(\varphi_\lambda)|^{2p})$. We use Rosenthal's inequality: there exists a positive constant $C(p)$ only depending on p such that

$$\begin{aligned}
&\mathbb{E}(|W_1(\varphi_\lambda)|^{2p}) \\
&= (n-1)^{2p}\mathbb{E}\left(\left|\sum_{i=1}^{n}\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_i) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right|^{2p}\right) \\
&\leqslant (n-1)^{2p}\times C(p)\Bigg(n\mathbb{E}\left(\left|\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right|^{2p}\right) \\
&\qquad\qquad + n^p\left[\mathrm{Var}\left(\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right)\right]^p\Bigg).
\end{aligned}$$

But, applying Lemma 6.1,

$$\begin{aligned}
&\mathbb{E}\left(\left|\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right|^{2p}\right) \\
&\leqslant \frac{1}{T^{2p}}\mathbb{E}\left(\left|\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\,dt\right|^{2p}\right)\left(\int_{\mathbb{R}} h(x)\,dx\right)^{2p} \\
&\leqslant \frac{2^{2p}}{T^{2p}}\left(\int_{\mathbb{R}} |\varphi_\lambda(x)|\,dx\right)^{2p}\left(\int_{\mathbb{R}} h(x)\,dx\right)^{2p}
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Var}\left(\int_{\mathbb{R}} \big[\varphi_\lambda(t-U_1) - \mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\mathbb{E}_\pi(h(t-U))\,dt\right) \\
&= \mathrm{Var}\left(\int_{\mathbb{R}} \varphi_\lambda(t-U_1)\mathbb{E}_\pi(h(t-U))\,dt\right) \\
&\leqslant \mathbb{E}\left[\left(\int_{\mathbb{R}} |\varphi_\lambda(t-U_1)|\mathbb{E}_\pi(h(t-U))\,dt\right)^2\right] \\
&\leqslant \frac{1}{T^2}\mathbb{E}\left[\left(\int_{\mathbb{R}} |\varphi_\lambda(t-U_1)|\,dt\right)^2\right]\left(\int_{\mathbb{R}} h(x)\,dx\right)^2 \\
&\leqslant \frac{1}{T^2}\left(\int_{\mathbb{R}} |\varphi_\lambda(x)|\,dx\right)^2\left(\int_{\mathbb{R}} h(x)\,dx\right)^2.
\end{aligned}$$

So,

$$\mathbb{E}(|W_1(\varphi_\lambda)|^{2p}) \leqslant K_1\, n^{2p}\left[\frac{n}{T^{2p}}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^{2p}+\frac{n^p}{T^{2p}}\left(\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\right)^{2p}\right]\left(\int_{\mathbb{R}}h(x)\,dx\right)^{2p}, \quad (6.30)$$

with K_1 a positive constant only depending on p.

And on the other hand, we have to control $\mathbb{E}(|W_2(\varphi_\lambda)|^{2p})$. We have:

$$W_2(\varphi_\lambda)=\sum_{1\leqslant i\neq j\leqslant n} g(U_i,U_j).$$

We use Theorem 3.3 of [11] associated with Theorem 1 of [7] (we keep the same notations of [11]). We set $h_{i,j}=\begin{cases}0 & \text{if } i=j\\ g & \text{otherwise}\end{cases}$ and we consider $(U_i^{(1)}, i=1\dots n)$ and $(U_i^{(2)}, i=1\dots n)$ two independent copies of $(U_i, i=1\dots n)$. With Theorem 3.3 of [11], there exists an universal constant K such that

$$\mathbb{E}\left(\left|\sum_{1\leqslant i,j\leqslant n} h_{i,j}(U_i^{(1)},U_j^{(2)})\right|^{2p}\right)\leqslant K^{2p}\big[(2p)^pC^{2p}+(2p)^{2p}D^{2p}+(2p)^{3p}B^{2p}+(2p)^{4p}A^{2p}\big],$$

where

- $A=\max_{i,j}\|h_{i,j}\|_\infty=\|g\|_\infty$. But, for all $(x,y)\in\mathbb{R}^2$,

$$\begin{aligned}|g(x,y)| &= \left|\int_{\mathbb{R}}\big[\varphi_\lambda(t-x)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[h(t-y)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right|\\ &\leqslant 4\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\|h\|_\infty,\end{aligned}$$

  using equality (b) of Lemma 6.1 with $f=\varphi_\lambda$. So,

$$A\leqslant 4\int_{\mathbb{R}}|\varphi_\lambda(x)|\,dx\|h\|_\infty, \quad (6.31)$$

- $C^2=\sum_{i,j}\mathbb{E}(h_{i,j}^2(U_i^{(1)},U_j^{(2)}))=\sum_{i\neq j}\mathbb{E}(g^2(U_i,U_j))$. But, for all $i\neq j$,

$$\begin{aligned}&\mathbb{E}(g^2(U_i,U_j))\\ &\leqslant \mathbb{E}\left[\left(\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]\big[h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big]\,dt\right)^2\right]\\ &\leqslant \mathbb{E}\left[\int_{\mathbb{R}}[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))]^2|h(t-U_j)-\mathbb{E}_\pi(h(t-U))|\,dt\int_{\mathbb{R}}|h(t-U_j)-\mathbb{E}_\pi(h(t-U))|\,dt\right]\\ &\leqslant 2\mathbb{E}\left[\int_{\mathbb{R}}\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\big|h(t-U_j)-\mathbb{E}_\pi(h(t-U))\big|\,dt\right]\int_{\mathbb{R}}h(x)\,dx\\ &\leqslant \frac{4}{T}\int_{\mathbb{R}}\mathbb{E}\left(\big[\varphi_\lambda(t-U_i)-\mathbb{E}_\pi(\varphi_\lambda(t-U))\big]^2\right)dt\left(\int_{\mathbb{R}}h(x)\,dx\right)^2\\ &\leqslant \frac{4}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2,\end{aligned}$$

  using Lemma 6.1. So,

$$C^2\leqslant\frac{4n(n-1)}{T}\int_{\mathbb{R}}\varphi_\lambda^2(x)\,dx\left(\int_{\mathbb{R}}h(x)\,dx\right)^2, \quad (6.32)$$

- $B^2 = \max\left[\left\|\sum_i \mathbb{E}_1(h_{i,j}^2(U_i,y))\right\|_\infty, \left\|\sum_j \mathbb{E}_2(h_{i,j}^2(x,U_j))\right\|_\infty\right]$, with

$$\mathbb{E}_1(h_{i,j}^2(U_i,y)) = \begin{cases} \mathbb{E}_\pi(g^2(U,y)) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \leqslant \frac{4}{T}\int_{\mathbb{R}} \varphi_\lambda^2(x)\,dx \left(\int_{\mathbb{R}} h(x)\,dx\right)^2$$

and

$$\mathbb{E}_1(h_{i,j}^2(x,U_j)) = \begin{cases} \mathbb{E}_\pi(g^2(x,U)) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \leqslant \frac{16}{T}\int_{\mathbb{R}} \varphi_\lambda^2(x)\,dx \left(\int_{\mathbb{R}} h(x)\,dx\right)^2,$$

using established inequalities to get equation (6.14) in the proof of the assumption (A1). So,

$$B^2 \leqslant \frac{16n}{T}\int_{\mathbb{R}} \varphi_\lambda^2(x)\,dx \left(\int_{\mathbb{R}} h(x)\,dx\right)^2, \tag{6.33}$$

- $D = \sup\left\{\mathbb{E}\Big(\sum_{i,j} h_{i,j}(U_i^{(1)},U_j^{(2)})a_i(U_i^{(1)})b_j(U_j^{(2)})\Big) : \mathbb{E}\Big(\sum_i a_i^2(U_i)\Big) \leqslant 1, \mathbb{E}\Big(\sum_j b_j^2(U_j)\Big) \leqslant 1\right\}$.
By using established inequalities to get equation (6.13) in the proof of the assumption (A1), we obtain

$$D \leqslant 2n\sqrt{\frac{1}{T}\int_{\mathbb{R}} \varphi_\lambda^2(x)\,dx}\int_{\mathbb{R}} h(x)\,dx. \tag{6.34}$$

Moreover, we use the equivalence of Theorem 3.3 of [11] and the decoupling inequality provided in Theorem 1 of [7] to obtain the following upper bound of $\mathbb{E}(|W_2(\varphi_\lambda)|^{2p})$:

$$\mathbb{E}(|W_2(\varphi_\lambda)|^{2p}) \leqslant K_2\left[\frac{n^{2p}}{T^p}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\,dx\right)^p\left(\int_{\mathbb{R}} h(x)\,dx\right)^{2p} + \frac{n^p}{T^p}\left(\int_{\mathbb{R}} \varphi_\lambda^2(x)\,dx\right)^p\left(\int_{\mathbb{R}} h(x)\,dx\right)^{2p} + \left(\int_{\mathbb{R}} |\varphi_\lambda(x)|\,dx\right)^{2p}\|h\|_\infty^{2p}\right], \tag{6.35}$$

with K_2 a positive constant only depending on p.

Finally, by using inequalities (6.29), (6.30) and (6.35) in (6.17), we obtain:

$$\begin{aligned}
&\mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}) \\
&\leqslant K\left\{\frac{1}{n^{2p-2}T^{2p}}\|\varphi_\lambda\|_1^{2p}\|h\|_1 + \frac{1}{n^{p-1}T^p}\|\varphi_\lambda\|_2^{2p}\|h\|_1 + \frac{1}{n^{2p-2}T}\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}\|h\|_1\right. \\
&\quad + \frac{1}{n^{2p-1}}\|\varphi_\lambda\|_2^2\|\varphi_\lambda\|_\infty^{2p-2}\|h\|_\infty + \frac{1}{n^{p/2}T^p}\|\varphi_\lambda\|_1^{2p}\|h\|_\infty^p + \frac{1}{T^p}\|\varphi_\lambda\|_2^{2p}\|h\|_1^p \\
&\quad + \frac{1}{n^p}\|\varphi_\lambda\|_2^{2p}\|h\|_\infty^p + \frac{1}{n^{p/2}T}\|\varphi_\lambda\|_2^{2p}\|h\|_\infty^p \quad + \quad \frac{n}{T^{2p}}\|\varphi_\lambda\|_1^{2p}\|h\|_1^{2p} + \frac{n^p}{T^{2p}}\|\varphi_\lambda\|_1^{2p}\|h\|_1^{2p} \\
&\quad + \quad \left.\frac{1}{T^p}\|\varphi_\lambda\|_2^{2p}\|h\|_1^{2p} + \frac{1}{n^pT^p}\|\varphi_\lambda\|_2^{2p}\|h\|_1^{2p} + \frac{1}{n^{2p}}\|\varphi_\lambda\|_1^{2p}\|h\|_\infty^{2p}\right\}
\end{aligned}$$

and so,

$$\begin{aligned}
&\left[\mathbb{E}\left(|\hat{\beta}_\lambda-\beta_\lambda|^{2p}\right)\right]^{\frac{1}{p}}\\
&\leqslant K\left\{\frac{1}{n^{2-2/p}T^2}\|\varphi_\lambda\|_1^2\|h\|_1^{1/p}+\frac{n}{T^2}\|\varphi_\lambda\|_1^2\|h\|_1^2+\frac{1}{n^{1/2}T}\|\varphi_\lambda\|_1^2\|h\|_\infty+\frac{1}{n^2}\|\varphi_\lambda\|_1^2\|h\|_\infty^2\right.\\
&\quad+\frac{1}{n^{1-1/p}T}\|\varphi_\lambda\|_2^2\|h\|_1^{1/p}+\frac{1}{T}\|\varphi_\lambda\|_2^2\|h\|_1+\frac{1}{T}\|\varphi_\lambda\|_2^2\|h\|_1^2+\left[\frac{1}{n}+\frac{1}{n^{1/2}T^{1/p}}\right]\|\varphi_\lambda\|_2^2\|h\|_\infty\\
&\quad\left.+\left[\frac{1}{n^{2-2/p}T^{1/p}}+\frac{1}{n^{2-1/p}}\right]\|\varphi_\lambda\|_2^{2/p}\|\varphi_\lambda\|_\infty^{2-2/p}\|h\|_\infty^{1/p}\right\},
\end{aligned}$$

with K a positive constant only depending on p and the compact support of ψ.

Recall that for any $\lambda=(j,k)\in\Lambda$, we have:

$$\|\varphi_\lambda\|_1\leqslant 2^{-j/2}M_{\psi,1},\quad \|\varphi_\lambda\|_2\leqslant M_{\psi,2}\quad\text{and}\quad \|\varphi_\lambda\|_\infty\leqslant 2^{j/2}M_{\psi,\infty}$$

We consider $1<p<\infty$ and we fix $1<q<\infty$ such that $\frac{1}{p}+\frac{1}{q}=1$, so that

$$\begin{aligned}
\left[\mathbb{E}\left(|\hat{\beta}_\lambda-\beta_\lambda|^{2p}\right)\right]^{\frac{1}{p}}&\leqslant K\left\{\frac{n}{T^2}\|h\|_1^2+\frac{1}{n^2}\|h\|_\infty^2+\frac{1}{n^{1/q}T}\|h\|_1^{1/p}+\frac{1}{T}\|h\|_1+\frac{1}{T}\|h\|_1^2\right.\\
&\quad\left.+\left[\frac{1}{n}+\frac{1}{n^{1/2}T^{1/p}}\right]\|h\|_\infty+\left[\frac{1}{n^{2/q}T^{1/p}}+\frac{1}{n^{1+1/q}}\right]2^{j_0/q}\|h\|_\infty^{1/p}\right\},
\end{aligned}$$

with K a positive constant depending on p, $\|\psi\|_1$, $\|\psi\|_2$, $\|\psi\|_\infty$ and the compact support of ψ.

Finally, choosing $p=2$, Assumption (A2) is fulfilled with

$$R=C_R\left\{\frac{1}{n}+\frac{2^{j_0/2}}{n^{3/2}}+\frac{2^{j_0/2}}{nT^{1/2}}+\frac{n}{T^2}\right\},$$

where C_R is a positive constant depending on $\|h\|_1$, $\|h\|_\infty$, $\|\psi\|_1$, $\|\psi\|_2$, $\|\psi\|_\infty$ and the compact support of ψ,

$$H_\lambda=\mathbf{1}_{\lambda\in\Gamma}$$

and $\varepsilon=1$.

6.4.3 Proof of Assumption (A3)

To shorten mathematical expressions, we denote $\eta_\lambda=\eta_\lambda(\gamma,\Delta)$ in the sequel. The following inequality:

$$\mathbb{P}\left(|\hat{\beta}_\lambda-\beta_\lambda|>\kappa\eta_\lambda,|\hat{\beta}_\lambda|>\eta_\lambda\right)\leqslant H_\lambda\zeta$$

is obvious with $\zeta=\omega$, which proves Assumption (A3) choosing $\theta=\frac{1+\varepsilon}{\varepsilon}$.

6.4.4 Completion of the proof of Theorem 3

Therefore we can apply Theorem 2: the estimator $\tilde{\beta}=\left(\hat{\beta}_\lambda\mathbf{1}_{|\hat{\beta}_\lambda|\geqslant\eta_\lambda}\mathbf{1}_{\lambda\in\Gamma}\right)_{\lambda\in\Lambda}$ satisfies

$$\begin{aligned}
\frac{1-\kappa^2}{1+\kappa^2}\mathbb{E}\left(\|\tilde{\beta}-\beta\|_{\ell_2}^2\right)&\leqslant\mathbb{E}\left(\inf_{m\subset\Gamma}\left\{\frac{1+\kappa^2}{1-\kappa^2}\sum_{\lambda\notin m}\beta_\lambda^2+\frac{1-\kappa^2}{\kappa^2}\sum_{\lambda\in m}(\hat{\beta}_\lambda-\beta_\lambda)^2+\sum_{\lambda\in m}\eta_\lambda^2\right\}\right)+LD\sum_{\lambda\in\Gamma}H_\lambda\\
&\leqslant\inf_{m\subset\Gamma}\left\{\frac{1+\kappa^2}{1-\kappa^2}\sum_{\lambda\notin m}\beta_\lambda^2+\frac{1-\kappa^2}{\kappa^2}\sum_{\lambda\in m}\mathbb{E}((\hat{\beta}_\lambda-\beta_\lambda)^2)+\sum_{\lambda\in m}\mathbb{E}(\eta_\lambda^2)\right\}+LD\sum_{\lambda\in\Gamma}H_\lambda,
\end{aligned}$$

with

- for all $\lambda = (j,k)$ in Γ,
$$\mathbb{E}((\hat{\beta}_\lambda - \beta_\lambda)^2) = \mathrm{Var}(\hat{\beta}_\lambda) \leqslant K \left\{ \frac{1}{n} + \frac{1}{T} + \frac{2^{-j}n}{T^2} \right\},$$
where K is a positive constant depending on $\|h\|_1$, $\|h\|_\infty$, $\|\psi\|_1$ and $\|\psi\|_2$ (see Lemma 2.1);
- for all $\lambda = (j,k)$ in Γ,
$$\eta_\lambda \leqslant K \left(\sqrt{j_0 \widetilde{V}\left(\frac{\varphi_\lambda}{n}\right)} + j_0 B\left(\frac{\varphi_\lambda}{n}\right) + \tilde{\Delta}\frac{N_{\mathbb{R}}}{n} \right)$$
where K depends on ε, κ, γ, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$ and
$$\tilde{\Delta} = \frac{j_0^2 2^{j_0/2}}{n} + \frac{j_0}{\sqrt{T}} + \frac{\sqrt{j_0 n}}{T};$$
- $LD = \frac{R}{\kappa^2}\big((1+\theta^{-1/2})\omega^{1/2} + (1+\theta^{1/2})\varepsilon^{1/2}\zeta^{1/2}\big) \leqslant K\ R\big(e^{-\kappa^2 j_0 \gamma/2} + \exp\left(-g(\varepsilon_0)n\|h\|_1/2\right)\big)$, where K is a positive constant depending only on ε and κ,
- $\sum_{\lambda\in\Gamma} H_\lambda = |\Gamma|$, where $|\Gamma|$ is the cardinal of the set Γ. So, we can upper bound this quantity by $K\ 2^{j_0}$, where K is a positive constant depending only on the compact support of h and the compact support of ψ.

Recall that $\varepsilon = 1$, $\kappa \in]0;1[$ will be fixed in the sequel and $\gamma > 0$, according Assumption (A1).

It remains to compute $\mathbb{E}(\eta_\lambda^2)$. Let $\lambda \in \Gamma$. We have:
$$\mathbb{E}(\eta_\lambda^2) \leqslant K \left(\frac{j_0}{n^2}\mathbb{E}(\widetilde{V}(\varphi_\lambda)) + \frac{j_0^2}{n^2}\mathbb{E}(B^2(\varphi_\lambda)) + \left\{ \frac{j_0 n}{T^2} + \frac{j_0^2}{T} + \frac{2^{j_0} j_0^4}{n^2} \right\} \frac{\mathbb{E}(N_{\mathbb{R}}^2)}{n^2} \right),$$
with K depending on ε, κ, γ, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$ and $\mathbb{E}(N_{\mathbb{R}}^2) = n\|h\|_1 + n^2\|h\|_1^2 \leqslant 2n^2 M_{h,1}^2$.
We control $\widetilde{V}(\varphi_\lambda)$ in expectation and we recall that $\alpha = \kappa^2 j_0 \gamma$.
$$\mathbb{E}(\widetilde{V}(\varphi_\lambda)) \leqslant \mathbb{E}(\hat{V}(\varphi_\lambda)) + \sqrt{2\alpha\mathbb{E}(\hat{V}(\varphi_\lambda))\mathbb{E}(B^2(\varphi_\lambda))} + 3\alpha\mathbb{E}(B^2(\varphi_\lambda)), \tag{6.36}$$
with using inequality (6.4),
$$\mathbb{E}(\hat{V}(\varphi_\lambda)) = \mathbb{E}(V(\varphi_\lambda)) \leqslant K \left\{ n\|h\|_\infty + \frac{n^2}{T}\|h\|_1 \right\},$$
where K is a positive constant depending only on $\|\psi\|_2$.

Now, we focus on $\mathbb{E}(B^2(\varphi_\lambda))$ where
$$\begin{aligned} B(\varphi_\lambda) = B(\varphi_\lambda) &= \left\| \sum_{i=1}^n \left[\varphi_\lambda(\cdot - U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(\cdot - U)) \right] \right\|_\infty \\ &= \sup_{t\in\mathbb{R}} \left| \sum_{i=1}^n \left[\varphi_\lambda(t - U_i) - \frac{n-1}{n}\mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right| \\ &\leqslant \tilde{B}(\varphi_\lambda) + \frac{1}{T}\|\varphi_\lambda\|_1, \end{aligned}$$
with $\tilde{B}(\varphi_\lambda) = \sup_{t\in\mathbb{R}} \left| \sum_{i=1}^n \left[\varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right] \right|$.
Then, we have to control $\mathbb{E}(\tilde{B}(\varphi_\lambda))$. Since it is a decomposition biorthogonal wavelet, φ_λ is a piecewise constant function and we can write:
$$\varphi_\lambda = \sum_{l=1}^N c_l \mathbf{1}_{[a_l;b_l]},$$

where $N \in \mathbb{N}^*$ and for any $l \in \{1, \dots, N\}$, $a_l, b_l, c_l \in \mathbb{R}$ and $a_l < b_l$. It is easy to see that

$$\tilde{B}(\varphi_\lambda) \leqslant \sum_{l=1}^{N} \tilde{B}(c_l \mathbf{1}_{[a_l;b_l]}) = \sum_{l=1}^{N} |c_l| \tilde{B}(\mathbf{1}_{[a_l;b_l]}).$$

It remains to compute $\mathbb{E}(\tilde{B}(\mathbf{1}_{[a;b]}))$ for some interval $[a; b]$.

$$\begin{aligned}\tilde{B}(\mathbf{1}_{[a;b]}) &= \sup_{t\in\mathbb{R}} \left| \sum_{i=1}^{n} \left[\mathbf{1}_{[a;b]}(t - U_i) - \mathbb{E}_\pi(\mathbf{1}_{[a;b]}(t - U)) \right] \right| \\ &\leqslant \sup_{B_t, t\in\mathbb{R}} \left| \sum_{i=1}^{n} \left[\mathbf{1}_{B_t}(U_i) - \mathbb{E}_\pi(\mathbf{1}_{B_t}(U)) \right] \right|,\end{aligned}$$

where for any $t \in \mathbb{R}$, $B_t = [t - b; t - a]$.
We set $\mathcal{B} = \{B_t, t \in \mathbb{R}\}$ and for every integer n, $m_n(\mathcal{B}) = \sup_{A\subset\mathbb{R}, |A|=n} |\{A \cap B_t, t \in \mathbb{R}\}|$. It is easy to see that

$$m_n(\mathcal{B}) \leqslant 1 + \frac{n(n+1)}{2}$$

and so, the VC-dimension of $\mathcal{B}$ defined by $sup\{n \geqslant 0, m_n(\mathcal{B}) = 2^n\}$ is bounded by 2 (see Definition 6.2 of [20]).
By applying Lemma 6.4 of [20], we obtain:

$$\sqrt{n}\mathbb{E}(\tilde{B}(\mathbf{1}_{[a;b]})) \leqslant \frac{K}{2}\sqrt{2},$$

where K is an absolute constant. So, for any λ in Γ,

$$\mathbb{E}(\tilde{B}(\varphi_\lambda)) \leqslant \frac{K}{\sqrt{n}}.$$

But, we want an upper bound of $\mathbb{E}(\tilde{B}^2(\varphi_\lambda))$. For this, we use Theorem 11 of [1]:

$$\left[\mathbb{E}(\tilde{B}^2(\varphi_\lambda))\right]^{1/2} \leqslant K \left\{ \mathbb{E}(\tilde{B}(\varphi_\lambda)) + \|\mathcal{M}\|_2 \right\},$$

where

$$\mathcal{M} = \max_{1\leqslant i\leqslant n} \sup_{t\in\mathbb{R}} \left| \varphi_\lambda(t - U_i) - \mathbb{E}_\pi(\varphi_\lambda(t - U)) \right|.$$

Hence,

$$\|\mathcal{M}\|_2^2 \leqslant 4\|\varphi_\lambda\|_\infty^2 \leqslant K\ 2^j,$$

with K a constant only depending on $\|\psi\|_\infty$.

Finally,

$$\begin{aligned}\mathbb{E}(B^2(\varphi_\lambda)) &\leqslant K\ \left\{ \mathbb{E}(\tilde{B}^2(\varphi_\lambda)) + \frac{2^{-j}}{T^2} \right\} \\ &\leqslant K\ \left\{ \left[\mathbb{E}(\tilde{B}(\varphi_\lambda))\right]^2 + 2^j + \frac{2^{-j}}{T^2} \right\} \\ &\leqslant K\ \left\{ \frac{1}{n} + 2^j + \frac{2^{-j}}{T^2} \right\}, \end{aligned} \tag{6.37}$$

with K a constant only depending on $\|\psi\|_1$ and $\|\psi\|_\infty$.

Then combining (6.36) and (6.37) yields

$$\mathbb{E}(\eta_\lambda^2) \leqslant K \left\{ \frac{j_0}{n} + \frac{j_0^{3/2} 2^{j_0/2}}{n^{3/2}} + \frac{j_0^2 2^{j_0}}{n^2} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0}{T} + \frac{j_0^2}{T} + \frac{j_0^{3/2} 2^{j_0/2}}{nT^{1/2}} + \frac{j_0 n}{T^2} \right\},$$

where K is a constant depending on γ, $\|h\|_1$, $\|h\|_\infty$, $\|\psi\|_1$, $\|\psi\|_2$ and $\|\psi\|_\infty$, which concludes the proof of Theorem 3 by setting

$$F(j_0, n, T) = \frac{j_0}{n} + \frac{j_0^{3/2} 2^{j_0/2}}{n^{3/2}} + \frac{j_0^4 2^{j_0}}{n^2} + \frac{j_0^2}{T} + \frac{j_0^{3/2} 2^{j_0/2}}{nT^{1/2}} + \frac{j_0 n}{T^2}.$$