# MODEL SELECTION FOR POISSON PROCESSES WITH COVARIATES 

Mathieu SART ${ }^{1}$


#### Abstract

We observe $n$ inhomogeneous Poisson's processes with covariates and aim at estimating their intensities. We assume that the intensity of each Poisson's process is of the form $s(\cdot, x)$ where $x$ is a covariate and where $s$ is an unknown function. We propose a model selection approach where the models are used to approximate the multivariate function $s$. We show that our estimator satisfies an oracle-type inequality under very weak assumptions both on the intensities and the models. By using an Hellinger-type loss, we establish non-asymptotic risk bounds and specify them under several kind of assumptions on the target function $s$ such as being smooth or a product function. Besides, we show that our estimation procedure is robust with respect to these assumptions. This procedure is of theoretical nature but yields results that cannot currently be obtained by more practical ones.


Mathematics Subject Classification. 62M05, 62G05.
Received November 6, 2013. Revised June 30, 2014.

## 1. Introduction

We consider $n$ independent Poisson point processes $N_{i}$ for $i=1, \ldots, n$ indexed by a measurable space ( $\mathbb{T}$, $\mathscr{T}$ ). For each $i$, we assume that $N_{i}$ admits an intensity $s_{i}$ with respect to some reference measure $\mu$ on $(\mathbb{T}, \mathscr{T})$ such that

$$
\int_{\mathbb{T}} s_{i}(t) \mathrm{d} \mu(t)<+\infty .
$$

We suppose that these processes are related as follows: there exist a deterministic element $x_{i}$ of some measurable set $(\mathbb{X}, \mathscr{X})$ and a non-negative function $s$ on $\mathbb{T} \times \mathbb{X}$ such that $s_{i}(\cdot)=s\left(\cdot, x_{i}\right)$. Our aim is then to estimate $s$ from the observations of the pairs $\left(N_{i}, x_{i}\right)_{1 \leq i \leq n}$.

It is worth mentioning that estimating $s$ merely amounts to estimating the $n$-tuple $\left(s_{1}, \ldots, s_{n}\right)$ when $\mathbb{X}=$ $\{1, \ldots, n\}$ and $x_{i}=i$. Nevertheless, we do not restrict to $\mathbb{X}=\{1, \ldots, n\}$ and $x_{i}=i$ and we also deal with more general sets $\mathbb{X}$. Typically, we have in mind the situation where the Poisson processes model the times of failure of $n$ repairable systems whose reliability depends on external factors measured by some covariates $x_{1}, \ldots, x_{n}$. In this case, $\mathbb{T}$ corresponds to an interval of time, say $[0,1]$, and $\mathbb{X}$ to some compact subset of $\mathbb{R}^{k}$, say $[0,1]^{k}$.

In the literature, much attention has been paid to the problem of estimating the intensity of a single Poisson's process. Concerning estimation by model selection, Reynaud-Bouret [17] dealt with the $\mathbb{L}_{2}$-loss, and provided a model selection theorem for a family of linear spaces. Baraud and Birgé [4] dealt with the Hellinger's distance

[^0]and considered the case where the models consist of piecewise constants functions on a partition of $\mathbb{T}$. The models considered in [8] are more general and are sets with finite metric dimensions (in a suitable sense).

Our statistical setting includes that of Poisson's regression: if one observes $n$ independent random variables $Y_{1}, \ldots, Y_{n}$, such that $Y_{i}$ obeys to a Poisson's law with parameter $f\left(x_{i}\right)$, one can estimate $f$ by setting $\mathbb{T}=\{0\}$, $N_{i}(\{0\})=Y_{i}$. In this case, $s(0, \cdot)=f(\cdot)$ and estimating $s$ amounts to estimating $f$. This last issue has been studied in $[1-3,15]$ among other references. For the particular cases of Poisson regression and estimating the intensity of a single Poisson's process, our results recover those of [3].

If we except these cases, statistical procedures that can estimate $s$ from $n$ independent Poisson's processes with covariates are rather scarce. The only risk bounds we are aware of are due to [10] who considered the $\mathbb{L}_{2}$-loss and penalized projection estimators on linear spaces. Their approach requires that the intensity $s$ be bounded from above by a quantity that needs to be either known or suitably estimated. Besides, they impose some restrictions on the family of linear spaces in order that their estimator possesses minimax properties over classes of functions which are smooth enough.

Our approach is based on robust testing. We propose a test inspired from a variational formula in [3] and then apply the general methodology for model selection developed in [7]. This yields a general model selection theorem we present below.

Let $\mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$ be the cone of integrable and non-negative functions on $(\mathbb{T} \times \mathbb{X}, \mathscr{T} \otimes \mathscr{X})$ equipped with the product measure $M=\mu \otimes \nu_{n}$ where $\nu_{n}=n^{-1} \sum_{i=1}^{n} \delta_{x_{i}}$. In order to evaluate the risks of the estimators, we endow $\mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$ with the Hellinger-type distance $H$ defined for $f, g \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$ by

$$
\begin{aligned}
H^{2}(f, g) & =\frac{1}{2} \int_{\mathbb{T} \times \mathbb{X}}(\sqrt{f(t, x)}-\sqrt{g(t, x)})^{2} \mathrm{~d} M(t, x) \\
& =\frac{1}{2 n} \sum_{i=1}^{n} \int_{\mathbb{T}}\left(\sqrt{f\left(t, x_{i}\right)}-\sqrt{g\left(t, x_{i}\right)}\right)^{2} \mathrm{~d} \mu(t)
\end{aligned}
$$

Let now $\left(\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M), d_{2}\right)$ be the metric space of square integrable functions $f$ on $\mathbb{T} \times \mathbb{X}$ with respect to the measure $M$. Given a suitable collection $\mathbb{V}$ of models (i.e. subsets of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ which are not necessarily linear spaces) and a non-negative mapping $\Delta$ on $\mathbb{V}$ satisfying

$$
\sum_{V \in \mathbb{V}} e^{-\Delta(V)} \leq 1
$$

we build an estimator $\hat{s}$ whose risk $\mathbb{E}\left[H^{2}(s, \hat{s})\right]$ satisfies

$$
\begin{equation*}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \inf _{V \in \mathbb{V}}\left\{d_{2}^{2}(\sqrt{s}, V)+\eta_{V}^{2}+\frac{\Delta(V)}{n}\right\} \tag{1.1}
\end{equation*}
$$

where $C$ is a positive number, $d_{2}(\sqrt{s}, V)$ is the $\mathbb{L}_{2}$-distance between $\sqrt{s}$ and $V$ and $n \eta_{V}^{2}$ is the metric dimension (in a suitable sense) of $V$.

This inequality, which is a bridge between statistics and approximation theory, allows to establish new risk bounds under various kinds of assumptions on the target function $s$. For instance, a suitable choice of $(\mathbb{V}, \Delta)$ leads to an adaptive estimator $\hat{s}$ achieving the expected rates of convergence over a (very) large range of Hölder spaces including irregular ones.

We shall also consider functions $s$ defined on a subset $\mathbb{T} \times \mathbb{X}$ of a linear space with large dimension, say $\mathbb{T}=[0,1]$ and $\mathbb{X}=[0,1]^{k}$ with a large value of $k$. It is well known that in such a situation, the minimax approach based on smoothness assumptions only may lead to very slow rates of convergence. This phenomenon is known as the curse of dimensionality. In this case, an alternative approach is to assume that $s$ belongs to classes $\mathscr{F}$ of functions satisfying structural assumptions (such as the multiple index model, the generalized additive model, the multiplicative model, ...) and for which faster rates of convergence can be achieved. Very recently, this approach was developed by [14] (for the Gaussian white noise model) and by [5] (in more general
settings). Unlike [14] we shall not assume that $s$ belongs to $\mathscr{F}$ but rather consider $\mathscr{F}$ as an approximating class for $s$.

In the present paper, our point of view is closer to that developed in [5]. We shall use our new model selection theorem in conjunction with suitable families $\mathbb{V}$ of models in order to design an estimator $\hat{s}$ possessing good statistical properties with respect to many classes of functions $\mathscr{F}$ of interest. For instance, a natural situation is the one where $s(t, x)$ is a product of two unknown non-negative functions $s(t, x)=u(t) v(x)$ with $\int_{\mathbb{T}} u(t) \mathrm{d} \mu(t)=1$. In the context of reliability, this means that $v\left(x_{i}\right)$ is the averaged number of failures of system $i$ and the times of failure are, conditionally to the number of failures $N_{i}(\mathbb{T})$, distributed accordingly to the density $u$ and independently of $x_{i}$. When $s(t, x)$ is a product function where $u$ and $v$ are assumed to be smooth, we shall prove that our estimator is fully adaptive with respect to the regularities of both $u$ and $v$. We shall also consider structural assumptions on the functions $u$ and $v$ as well as parametric ones when $t$ and $x$ lie in a large dimensional space. A second natural situation is the one where the intensity $s_{i}$ of each Poisson process belongs to a parametric class of functions $\mathcal{F}_{\Theta}=\left\{f_{\theta}, \theta \in \Theta\right\}$ indexed by a subset $\Theta \subset \mathbb{R}^{k}$. This means that there exists some element $f_{\theta\left(x_{i}\right)} \in \mathcal{F}_{\Theta}$ such that $s\left(t, x_{i}\right)=f_{\theta\left(x_{i}\right)}(t)$. We shall then prove risk bounds under assumptions on the map $x \mapsto \theta(x)$.

All these results ensue from (1.1) and highlight the theoretical interest of such a general model selection theorem. The counterpart of the nice properties of our estimators is that they are very difficult to construct in practice. They should be thus considered as benchmarks for what theoretical feasible. This drawback arises in all general procedures based on robust tests (see for instance [7-9]). Recently, efforts have been made to find more practical estimation procedures based on tests. However, as far as we know, none of them can build, in a reasonable amount of time, an estimator $\hat{s}$ satisfying (1.1) for the general collections $\mathbb{V}$ considered in this paper. We are not aware of more practical estimation strategies (based on tests or not) that could yield results as general as ours.

This paper is organized as follows. The general model selection theorem can be found in Section 2. In Section 3, we study the case where $\mathscr{F}$ is a class of smooth functions, and in Section 4 the case where $\mathscr{F}$ is a class of product functions. The problem of estimating $s$ when the intensity of each Poisson process $N_{i}$ belongs to the same parametric model is dealt in Section 5. Section 6 is devoted to the proofs.

Let us introduce some notations that will be used all along the paper. We set $\mathbb{N}^{\star}=\mathbb{N} \backslash\{0\}, \mathbb{R}^{\star}=\mathbb{R} \backslash\{0\}$. The components of a vector $\boldsymbol{\theta} \in \mathbb{R}^{k}$ are denoted by $\boldsymbol{\theta}=\left(\theta_{1}, \ldots, \theta_{k}\right)$. The numbers $x \wedge y$ and $x \vee y$ stand for $\min (x, y)$ and $\max (x, y)$ respectively. For $(E, d)$ a metric space, $x \in E$ and $A \subset E$, the distance between $x$ and $A$ is denoted by $d(x, A)=\inf _{a \in A} d(x, a)$. The cardinality of a finite set $A$ is denoted by $|A|$. We use $\mathscr{F}$ as a generic notation for a family of functions of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ of special interest. The notations $C, C^{\prime}, C^{\prime \prime} \ldots$ are for constants. The constants $C, C^{\prime}, C^{\prime \prime} \ldots$ may change from line to line.

## 2. A GENERAL MODEL SELECTION THEOREM

Throughout this paper, a model $V$ is a subset of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ with bounded metric dimension, in the sense of Definition 6 of [7]. We recall this definition below.

Definition 2.1. Let $V$ be a subset of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ and $D_{V}$ a right-continuous map from $(0,+\infty)$ into $[1 / 2,+\infty)$ such that $D_{V}(\eta)=\mathcal{O}\left(\eta^{2}\right)$ when $\eta \rightarrow+\infty$. We say that $V$ has a metric dimension bounded by $D_{V}$ if for all $\eta>0$, there exists an at most countable subset $S_{V}(\eta)$ of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ such that:

- for all $f \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$, there exists $g \in S_{V}(\eta)$ such that $d_{2}(f, g) \leq \eta$.
- for all $\varphi \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ and $r \geq 2$,

$$
\left|S_{V}(\eta) \cap \mathcal{B}(\varphi, r \eta)\right| \leq \exp \left(D_{V}(\eta) r^{2}\right)
$$

where $\mathcal{B}(\varphi, r \eta)$ stands for the closed ball centered at $\varphi$ with radius $r \eta$ of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$.
Moreover, if one can choose $D_{V}$ as a constant, we say that $V$ has a finite metric dimension bounded by $D_{V}$.

This notion is more general than the dimension for linear spaces since a linear space $V$ with finite dimension (in the usual sense) has a finite metric dimension. Besides, if $V$ is not reduced to $\{0\}$ one can choose $D_{V}=\operatorname{dim} V$, what we shall do along this paper. The link with the classical definition of metric entropy may be found in Section 6.4.3 of [7]. Other models of interest with bounded metric dimension will appear later in the paper.

Given a collection of such subsets, our approach is based on model selection. We propose a selection rule based on robust testing in the spirit of the papers $[3,7]$. The test and the selection rule which are mainly abstract are postponed to Section 6. The main result is the following.

Theorem 2.2. Let $\mathbb{V}$ be an at most countable family of models $V$ with bounded metric dimension $D_{V}(\cdot)$ and $\Delta$ be a non-negative mapping on $\mathbb{V}$ such that

$$
\sum_{V \in \mathbb{V}} e^{-\Delta(V)} \leq 1
$$

There exists an estimator $\hat{s} \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$ such that, for all $\xi>0$,

$$
\mathbb{P}\left[C H^{2}(s, \hat{s}) \geq \inf _{V \in \mathbb{V}}\left\{d_{2}^{2}(\sqrt{s}, V)+\eta_{V}^{2}+\frac{\Delta(V)}{n}\right\}+\xi\right] \leq e^{-n \xi}
$$

where $C$ is a universal positive constant and where

$$
\eta_{V}=\inf \left\{\eta>0, \frac{D_{V}(\eta)}{\eta^{2}} \leq n\right\}
$$

In particular, by integrating the above inequality,

$$
\begin{equation*}
C^{\prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \inf _{V \in \mathbb{V}}\left\{d_{2}^{2}(\sqrt{s}, V)+\eta_{V}^{2}+\frac{\Delta(V)}{n}\right\} \tag{2.1}
\end{equation*}
$$

where $C^{\prime}$ is a universal positive constant.
The condition $\sum_{V \in \mathbb{V}} e^{-\Delta(V)} \leq 1$ can be interpreted as a (sub)probability on the collection $\mathbb{V}$. This a priori choice of $\Delta$ has a Bayesian flavour as well as an information-theoretic interpretation (see [6]). Note that the more complex the family $\mathbb{V}$, the larger the weights $\Delta(V)$. When $\mathbb{V}$ consists of linear spaces $V$ of finite dimensions $D_{V}$ one can take $\eta_{V}^{2}=D_{V} / n$ and hence (2.1) leads to

$$
C^{\prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \inf _{V \in \mathbb{V}}\left\{d_{2}^{2}(\sqrt{s}, V)+\frac{D_{V}+\Delta(V)}{n}\right\}
$$

When one can choose $\Delta(V)$ of order $D_{V}$, which means that the family $\mathbb{V}$ of models does not contain too many models per dimension, the estimator $\hat{s}$ achieves the best trade-off (up to a constant) between the approximation and the variance terms.

In the remaining part of this paper, we shall consider subsets $\mathscr{F} \subset \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ corresponding to various assumptions on $\sqrt{s}$ (smoothness, structural, parametric assumptions,...). For such an $\mathscr{F}$, we associate a collection $\mathbb{V}_{\mathscr{F}}$ and deduce from Theorem 2.2 a risk bound for the estimator $\hat{s}$ whenever $\sqrt{s}$ belongs or is close to $\mathscr{F}$. This bound takes the form

$$
\begin{equation*}
C^{\prime \prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \inf _{f \in \mathscr{F}}\left\{d_{2}^{2}(\sqrt{s}, f)+\varepsilon_{\mathscr{F}}(f)\right\} \tag{2.2}
\end{equation*}
$$

where

$$
\varepsilon_{\mathscr{F}}(f)=\inf _{V \in \mathbb{V}_{\mathscr{F}}}\left\{d_{2}^{2}(f, V)+\eta_{V}^{2}+\frac{\Delta(V)}{n}\right\}
$$

and we shall bound the term $\varepsilon \mathscr{F}(f)$ from above. This upper bound will mainly depend on some properties of $f$, for example smoothness ones. In this case, this result says that if $\sqrt{s}$ is irregular but sufficiently close to a smooth function $f$, the bound we get essentially corresponds to the one we would get for $f$. This can be interpreted as a robustness property.

Sometimes, several assumptions on $\sqrt{s}$ are plausible, and one does not know what class $\mathscr{F}$ should be taken. A solution is to consider $\mathfrak{F}$ a collection of such classes $\mathscr{F}$ and to use the proposition below to get an estimator whose risk satisfies (up to a remaining term) relation (2.2) simultaneously for all classes $\mathscr{F} \in \mathfrak{F}$.
Proposition 2.3. Let $\mathfrak{F}$ be an at most countable collection of subsets of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ and $\bar{\Delta}$ be a non-negative map on $\mathfrak{F}$ such that $\sum_{\mathscr{F} \in \mathfrak{F}} e^{-\bar{\Delta}(\mathscr{F})} \leq 1$. For all $\mathscr{F} \in \mathfrak{F}$, let $\mathbb{V}_{\mathscr{F}}$ be a collection of models and $\Delta_{\mathscr{F}}$ be a mapping such that the assumptions of Theorem 2.2 hold.

There exists an estimator $\hat{s}$ such that, for all $\mathscr{F} \in \mathfrak{F}$,

$$
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \inf _{f \in \mathscr{F}}\left\{d_{2}^{2}(\sqrt{s}, f)+\varepsilon_{\mathscr{F}}(f)\right\}+\frac{\bar{\Delta}(\mathscr{F})}{n}
$$

where

$$
\varepsilon_{\mathscr{F}}(f)=\inf _{V \in \mathbb{V}_{\mathscr{F}}}\left\{d_{2}^{2}(f, V)+\eta_{V}^{2}+\frac{\Delta_{\mathscr{F}}(V)}{n}\right\},
$$

and where $C$ is a universal positive constant.

## 3. Smoothness Assumptions

Let $\mathbf{I}=\prod_{j=1}^{k} I_{j}$ where the $I_{j}$ are intervals of $\mathbb{R}$ and $\boldsymbol{\alpha}=\boldsymbol{\beta}+\mathbf{p} \in(0,+\infty)^{k}$ with $\mathbf{p} \in \mathbb{N}^{k}$ and $\boldsymbol{\beta} \in(0,1]^{k}$. A function $f$ belongs to the Hölder class $\mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I})$, if there exists $L(f) \in[0,+\infty)$ such that for all $\left(x_{1}, \ldots, x_{k}\right) \in \mathbf{I}$ and all $j \in\{1, \ldots, k\}$, the functions $f_{j}(x)=f\left(x_{1}, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_{k}\right)$ admit a derivative of order $p_{j}$ satisfying

$$
\left|f_{j}^{\left(p_{j}\right)}(x)-f_{j}^{\left(p_{j}\right)}(y)\right| \leq L(f)|x-y|^{\beta_{j}} \quad \forall x, y \in I_{j}
$$

The class $\mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I})$ is said to be isotropic when the $\alpha_{j}$ are all equal, and anisotropic otherwise, in which case $\overline{\boldsymbol{\alpha}}$ given by $\overline{\boldsymbol{\alpha}}^{-1}=k^{-1} \sum_{j=1}^{k} \alpha_{j}^{-1}$ corresponds to the average smoothness of a function $f$ in $\mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I})$. We define the class of Hölderian functions on I by

$$
\mathcal{H}(\mathbf{I})=\bigcup_{\alpha \in(0,+\infty)^{k}} \mathcal{H}^{\alpha}(\mathbf{I})
$$

Assuming that $\sqrt{s}$ is Hölderian corresponds thus to the choice $\mathscr{F}=\mathcal{H}(\mathbb{T} \times \mathbb{X})$. Anisotropic classes of smoothness are of particular interest in our context since the function $s$ depends on variables $t$ and $x$ that may play very different roles.

Families of linear spaces possessing good approximation properties with respect to the elements of $\mathscr{F}$ can be found in the literature. We refer to the results of [11]. We may use these linear spaces (models) to approximate the elements of $\mathscr{F}$, and deduce from Theorem 2.2 the following result.
Corollary 3.1. Let us assume that $\mathbb{T} \times \mathbb{X}=[0,1]^{k}$ and that $\mu$ is the Lebesgue measure. There exists an estimator $\hat{s}$ such that for all $f \in \mathcal{H}\left([0,1]^{k}\right)$,

$$
\begin{equation*}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq d_{2}^{2}(\sqrt{s}, f)+L(f)^{\frac{2 k}{2 \alpha+k}} n^{-\frac{2 \bar{\alpha}}{2 \bar{\alpha}+k}}+n^{-1} \tag{3.1}
\end{equation*}
$$

where $\boldsymbol{\alpha} \in(0,+\infty)^{k}$ is such that $f \in \mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^{k}\right)$ and where $C>0$ depends only on $k$ and $\max _{1 \leq j \leq k} \alpha_{j}$.
Remark that the risk bound given by inequality (3.1) holds without any restriction on $\boldsymbol{\alpha}$. Such a generality can be obtained since our model selection theorem is valid for any collection $\mathbb{V}$ of finite dimensional linear spaces. Some restrictions on the dimensionality of the linear spaces $V \in \mathbb{V}$ (as in [10]) would prevent us to get this rate of convergence for the Hölder classes $\mathcal{H}^{\alpha}\left([0,1]^{k}\right)$ when $\min _{1 \leq j \leq k} \alpha_{j}$ is too small.

The preceding risk bound is quite satisfactory if $k$ is small but becomes worse when $k$ increases. We shall therefore consider other types of classes in the next section in order to avoid this curse of dimensionality.

## 4. FAMILIES $\mathscr{F}$ OF PRODUCT FUNCTIONS

A common way of modelling the influence of the covariates on the number of failures of $n$ systems is to assume that, for each $i \in\{1, \ldots, n\}$, the intensity of $N_{i}$, is of the form $s\left(t, x_{i}\right)=u(t) v\left(x_{i}\right)$ where $u$ is an unknown density function on $\mathbb{T}$, and $v$ some unknown non-negative function on $\mathbb{X}$. This means, that in average, the number of failures of system $i, \mathbb{E}\left[N_{i}(\mathbb{T})\right]=v\left(x_{i}\right)$, depends on $x_{i}$ through $v$ only, and conditionally to $N_{i}(\mathbb{T})=k_{i}>0$, the times of failure are distributed along $\mathbb{T}$ independently of $x_{i}$, but accordingly to the density $u$.

Let $\left(\mathbb{L}^{2}(\mathbb{T}, \mu),\|\cdot\|_{t}\right)\left(\right.$ respectively $\left.\left(\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right),\|\cdot\|_{\mathrm{x}}\right)\right)$ be the normed space of square integrable functions on $\mathbb{T}$ (respectively $\mathbb{X}$ ) with respect to $\mu$ (respectively $\nu_{n}$ ). We denote by $d_{\mathrm{t}}$ and $d_{\mathrm{x}}$ the distances associated to the norms $\|\cdot\|_{\mathrm{t}}$ and $\|\cdot\|_{\mathrm{x}}$, respectively. We shall consider the class $\mathscr{F}$ defined by

$$
\begin{equation*}
\mathscr{F}=\left\{\kappa v_{1} v_{2}, \kappa \geq 0,\left(v_{1}, v_{2}\right) \in \mathbb{L}^{2}(\mathbb{T}, \mu) \times \mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right),\left\|v_{1}\right\|_{\mathrm{t}}=\left\|v_{2}\right\|_{\mathrm{x}}=1\right\} \tag{4.1}
\end{equation*}
$$

which amounts to assuming that $s$ is of the form (or close to) a product function $u(t) v(x)$ with $u=v_{1}^{2}$ and $v=\kappa^{2} v_{2}^{2}$.

In this section, we introduce collections of models $\mathbb{V}_{1}$ and $\mathbb{V}_{2}$ in order to approximate the components $v_{1}$ and $v_{2}$ separately. Given $V_{1} \in \mathbb{V}_{1}$ to approximate $v_{1}$ and $V_{2} \in \mathbb{V}_{2}$ to approximate $v_{2}$, we approximate $v_{1} v_{2}$ by the model $V_{1} \otimes V_{2}$ defined by

$$
\begin{equation*}
V_{1} \otimes V_{2}=\left\{v_{1}^{\prime} v_{2}^{\prime},\left(v_{1}^{\prime}, v_{2}^{\prime}\right) \in V_{1} \times V_{2}\right\} \tag{4.2}
\end{equation*}
$$

The metric dimension of $V_{1} \otimes V_{2}$ is controlled as follows.
Lemma 4.1. Let $V_{1}$ and $V_{2}$ be a finite dimensional linear space of $\mathbb{L}^{2}(\mathbb{T}, \mu)$ and $\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$ respectively. The set $V_{1} \otimes V_{2}$ defined by (4.2) has a finite metric dimension bounded by

$$
D_{V_{1} \otimes V_{2}}=1.4\left(\operatorname{dim} V_{1}+\operatorname{dim} V_{2}+1\right)
$$

By using Theorem 2.2, we prove the following result.

Proposition 4.2. Let $\mathbb{V}_{1}$ (respectively $\mathbb{V}_{2}$ ) be an at most countable collection of finite dimensional linear spaces of $\mathbb{L}^{2}(\mathbb{T}, \mu)$ (respectively $\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$ ). Let, for all $i \in\{1,2\}, \Delta_{i}$ be a non-negative mapping on $\mathbb{V}_{i}$ such that

$$
\sum_{V_{i} \in \mathbb{V}_{i}} e^{-\Delta_{i}\left(V_{i}\right)} \leq 1
$$

There exists an estimator $\hat{s}$ such that, for all $\kappa v_{1} v_{2} \in \mathscr{F}$, where $\mathscr{F}$ is defined by (4.1),

$$
\begin{aligned}
& C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq d_{2}^{2}\left(\sqrt{s}, \kappa v_{1} v_{2}\right)+\inf _{V_{1} \in \mathbb{V}_{1}}\left\{\kappa^{2} d_{\mathrm{t}}^{2}\left(v_{1}, V_{1}\right)+\frac{\operatorname{dim} V_{1} \vee 1+\Delta_{1}\left(V_{1}\right)}{n}\right\} \\
&+\inf _{V_{2} \in \mathbb{V}_{2}}\left\{\kappa^{2} d_{\mathrm{x}}^{2}\left(v_{2}, V_{2}\right)+\frac{\operatorname{dim} V_{2} \vee 1+\Delta_{2}\left(V_{2}\right)}{n}\right\}
\end{aligned}
$$

where $C$ is a universal positive contant. Furthermore, $\sqrt{\hat{s}}$ belongs to $\mathscr{F}$.

Apart for the term $d_{2}^{2}\left(\sqrt{s}, \kappa v_{1} v_{2}\right)$ which corresponds to some robustness with respect to the assumption $\sqrt{s} \in \mathscr{F}$, the risk bound we get corresponds to the one we would get if we could apply a model selection theorem on the components $v_{1}$ and $v_{2}$ separately.

### 4.1. Smoothness assumptions on $v_{1}$ and $\boldsymbol{v}_{2}$.

We illustrate this proposition by setting $\mathbb{T}=[0,1]^{k_{1}}, \mathbb{X}=[0,1]^{k_{2}}, \mu$ the Lebesgue measure and

$$
\begin{equation*}
\mathscr{F}=\left\{\kappa v_{1} v_{2}, \kappa \geq 0, v_{1} \in \mathcal{H}\left([0,1]^{k_{1}}\right),\left\|v_{1}\right\|_{\mathrm{t}}=1, v_{2} \in \mathcal{H}\left([0,1]^{k_{2}}\right),\left\|v_{2}\right\|_{\mathrm{x}}=1\right\} \tag{4.3}
\end{equation*}
$$

We apply Proposition 4.2 with families $\mathbb{V}_{1}$ and $\mathbb{V}_{2}$ of linear spaces possessing good approximation properties with respect to the functions of $\mathcal{H}\left([0,1]^{k_{1}}\right)$ and $\mathcal{H}\left([0,1]^{k_{2}}\right)$, respectively. This leads to the following corollary.
Corollary 4.3. There exists an estimator $\hat{s}$ such that, for all $\kappa v_{1} v_{2} \in \mathscr{F}$, where $\mathscr{F}$ is defined by (4.3),

$$
\begin{aligned}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, \kappa v_{1} v_{2}\right)+\kappa^{\frac{2 k_{1}}{2 \alpha+k_{1}}} L\left(v_{1}\right)^{\frac{2 k_{1}}{2 \bar{\alpha}+k_{1}}} n^{-\frac{2 \bar{\alpha}}{2 \alpha+k_{1}}} \\
& +\kappa^{\frac{2 k_{2}}{2 \beta+k_{2}}} L\left(v_{2}\right)^{\frac{2 k_{2}}{2 \beta+k_{2}}} n^{-\frac{2 \bar{\beta}}{2 \beta+k_{2}}}+n^{-1}
\end{aligned}
$$

where $\boldsymbol{\alpha} \in(0,+\infty)^{k_{1}}$, is such that $v_{1} \in \mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^{k_{1}}\right)$, where $\boldsymbol{\beta} \in(0,+\infty)^{k_{2}}$ is such that $v_{2} \in \mathcal{H}^{\boldsymbol{\beta}}\left([0,1]^{k_{2}}\right)$, and where $C>0$ depends only on $k_{1}, k_{2}, \max _{1 \leq j \leq k_{1}} \alpha_{i}$, and $\max _{1 \leq j \leq k_{2}} \beta_{i}$.

In particular, if $s$ is a product function of the form $\sqrt{s}=\kappa v_{1} v_{2}$ where $v_{1} \in \mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^{k_{1}}\right)$, and $v_{2} \in \mathcal{H}^{\boldsymbol{\beta}}\left([0,1]^{k_{2}}\right)$, $\sqrt{s}$ is Hölderian with regularity $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ on $[0,1]^{k_{1}+k_{2}}$. However, the rate given by the corollary above is always faster than the one we would get by Corollary 3.1 under smoothness assumption only.

### 4.2. Mixing smoothness and structural assumptions

When $k_{2}$ is large, we may consider structural assumptions on $v_{2}$ instead of smoothness ones to improve the risk bound. Proposition 4.2 allows to consider a wide variety of situations thanks to the approximation results of [5] on composite functions. We do not present all of them for the sake of concisely. We just consider the example in which the class $\mathscr{F}$ is

$$
\begin{gather*}
\mathscr{F}=\left\{\kappa v_{1} v_{2}, \kappa \geq 0, v_{1} \in \mathcal{H}\left([0,1]^{k_{1}}\right), \boldsymbol{\theta}_{1}, \ldots, \boldsymbol{\theta}_{l} \in \mathcal{B}(0,1), g \in \mathcal{H}\left([-1,1]^{l}\right),\right.  \tag{4.4}\\
\left.\forall \boldsymbol{x} \in \mathbb{X}, v_{2}(\boldsymbol{x})=g\left(\left\langle\boldsymbol{\theta}_{1}, \boldsymbol{x}\right\rangle, \ldots,\left\langle\boldsymbol{\theta}_{l}, \boldsymbol{x}\right\rangle\right),\left\|v_{1}\right\|_{\mathrm{t}}=\left\|v_{2}\right\|_{\mathrm{x}}=1\right\}
\end{gather*}
$$

where $\mathbb{T}=[0,1]^{k_{1}}, \mu$ the Lebesgue measure and

$$
\mathbb{X}=\left\{\boldsymbol{x} \in \mathbb{R}^{k_{2}}, \sum_{j=1}^{k_{2}} x_{j}^{2} \leq 1\right\}
$$

stands for the unit ball of $\mathbb{R}^{k_{2}}$. The following corollary ensues from Proposition 4.2 and Corollary 2 of [5].
Corollary 4.4. There exists an estimator $\hat{s}$ such that, for all $\kappa v_{1} v_{2} \in \mathscr{F}$, where $\mathscr{F}$ is defined by (4.4),

$$
\begin{aligned}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, \kappa v_{1} v_{2}\right)+\kappa^{\frac{2 k_{1}}{2 \bar{\alpha}+k_{1}}} L\left(v_{1}\right)^{\frac{2 k_{1}}{2 \bar{\alpha}+k_{1}}} n^{-\frac{2 \bar{\alpha}}{2 \bar{\alpha}+k_{1}}} \\
& +\kappa^{\frac{2 l}{2 \boldsymbol{\beta}+l}} L(g)^{\frac{2 l}{2 \boldsymbol{\beta}+l}} n^{-\frac{2 \bar{\beta}}{2 \boldsymbol{\beta}+l}}+\frac{\ln \left(\kappa^{2}\|g\|_{\boldsymbol{\beta}}^{2} k_{2}^{-1}\right) \vee \ln n \vee 1}{n} k_{2}
\end{aligned}
$$

where $\boldsymbol{\alpha} \in(0,+\infty)^{k_{1}}, \boldsymbol{\beta} \in(0,+\infty)^{l}$ are such that $v_{1} \in \mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^{k_{1}}\right), g \in \mathcal{H}^{\boldsymbol{\beta}}\left([-1,1]^{l}\right)$ with $v_{2}(\boldsymbol{x})=$ $g\left(\left\langle\boldsymbol{\theta}_{1}, \boldsymbol{x}\right\rangle, \ldots,\left\langle\boldsymbol{\theta}_{l}, \boldsymbol{x}\right\rangle\right)$ and where $C>0$ depends only on $k_{1}, l$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In the above inequality, $\|g\|_{\boldsymbol{\beta}}$ stands for any positive real number such that for all $\left(x_{1}, \ldots, x_{l}\right) \in[-1,1]^{l}$ and all $j \in\{1, \ldots, l\}$, the function $g_{j}(x)=g\left(x_{1}, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_{l}\right)$ satisfies

$$
\left|g_{j}(x)-g_{j}(y)\right| \leq\|g\|_{\boldsymbol{\beta}}|x-y|^{\beta_{j} \wedge 1} \quad \forall x, y \in[-1,1] .
$$

When $\sqrt{s}$ belongs to the class $\mathscr{F}$, the risk bound of the above inequality corresponds to the one we would get if we could estimate the functions $v_{1}$ and $g$ separately. This risk bound is then better than the one we would get under smoothness assumptions on $v_{2}$ when $l<k_{2}$.

### 4.3. Examples of parametric assumptions

Theorem 2.2 also allows to deal with parametric assumptions. Hereafter, we consider a class $\mathscr{F}$ of the form

$$
\mathscr{F}=\left\{a u_{b} v_{\boldsymbol{\theta}}, a \geq 0, b \in I, \boldsymbol{\theta} \in \mathbb{R}^{k_{2}}\right\},
$$

where $I$ is an interval of $\mathbb{R},\left(u_{b}\right)_{b \in I}$ is a family of functions and $v_{\boldsymbol{\theta}}$ is defined by $v_{\boldsymbol{\theta}}(\boldsymbol{x})=\exp (\langle\boldsymbol{x}, \boldsymbol{\theta}\rangle)$ for $\boldsymbol{x} \in \mathbb{X}=\left\{\boldsymbol{x} \in \mathbb{R}^{k_{2}}, \sum_{j=1}^{k_{2}} x_{j}^{2} \leq 1\right\}$, the unit ball of $\mathbb{R}^{k_{2}}$. For each $i \in\{1, \ldots n\}$, the intensity of $N_{i}$ is thus assumed to be proportional to an element of (or an element close to) some reference parametric model $\left\{u_{b}^{2}, b \in I\right\}$. Let us give 3 examples of such models.

The Power Law Processes are Poisson's processes whose intensities are proportional to $u_{b}(t)=t^{b}$ for all $t \in \mathbb{T}=(0,1]$ and some $b \in(-1 / 2,+\infty)$. Proposed first in [12], this model is popular in reliability. Indeed, although the intensity is simple, different situations can be modelled by this model. For example, if $b=0$ each $N_{i}$ obeys to an homogeneous Poisson's process, whereas if $b>0$ (respectively $b<0$ ) the reliability of each system reduces (respectively improves) with time. In software reliability, we can cite the Goel-Okumoto's model of [13] and the $S$-Shaped's model of [18]. The former considers intensities proportional to $u_{b}(t)=e^{-b t}$ whereas the latter corresponds to $u_{b}(t)=\sqrt{t} e^{-b t}$ where $b \in[0,+\infty)$ and $t \in \mathbb{T}=[0,+\infty)$.

We consider the following assumption on the family $\left\{u_{b}, b \in I\right\}$.
Assumption 4.5. The family $\left(u_{b}\right)_{b \in I}$ is a family of non vanishing functions of $\mathbb{L}^{2}(\mathbb{T}, \mu)$ indexed by an interval $I$ of the form $\left(b_{0},+\infty\right)$. Moreover, there exist two positive non-increasing functions $\underline{\rho}, \bar{\rho}$ on $I$, such that for all $b, b^{\prime} \in I$,

$$
\underline{\rho}\left(b \vee b^{\prime}\right)\left|b-b^{\prime}\right| \leq\left\|\frac{u_{b}}{\left\|u_{b}\right\|_{\mathrm{t}}}-\frac{u_{b^{\prime}}}{\left\|u_{b^{\prime}}\right\|_{\mathrm{t}}}\right\|_{\mathrm{t}} \leq \bar{\rho}\left(b \wedge b^{\prime}\right)\left|b-b^{\prime}\right| .
$$

The purpose of the lemmas below is to show that the above assumption holds for the Duane, Goel-Okumoto and S-Shaped's models.
Lemma 4.6. Let $I=(-1 / 2,+\infty), \mathbb{T}=(0,1], \mu$ the Lebesgue's measure, and for $b \in I, u_{b}(t)=t^{b}$. Assumption 4.5 is satisfied with

$$
\underline{\rho}(u)=\bar{\rho}(u)=\frac{1}{1+2 u} \quad \text { for all } u>-1 / 2
$$

Lemma 4.7. Let $I=(0,+\infty), \mathbb{T}=[0,+\infty)$, $\mu$ the Lebesgue's measure, $k \in \mathbb{N}$, and for $b \in I$, $u_{b}(t)=t^{k / 2} e^{-b t}$. Assumption 4.5 is satisfied with

$$
\underline{\rho}(u)=\frac{1}{2 u} \quad \text { and } \quad \bar{\rho}(u)=\frac{\sqrt{k+1}}{2 u} \quad \text { for all } u>0 .
$$

All along this section, $\|\cdot\|$ denotes the standard Euclidean norm of $\mathbb{R}^{k_{2}}$

$$
\forall \boldsymbol{x} \in \mathbb{R}^{k_{2}}, \quad\|\boldsymbol{x}\|^{2}=\sum_{j=1}^{k_{2}} x_{j}^{2}
$$

and $d$ the distance induced by this norm.
Proposition 4.8. Let $\left(u_{b}\right)_{b \in I}$ be a family such that Assumption 4.5 holds. There exist $\hat{a} \geq 0, \hat{b} \in I$ and $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{k_{2}}$, such that the estimator $\hat{s}=\left(\hat{a} u_{\hat{b}} v_{\hat{\boldsymbol{\theta}}}\right)^{2}$ satisfies, for all $a \geq 0, b \in I, \boldsymbol{\theta} \in \mathbb{R}^{k_{2}}$, and $f \in \mathscr{F}$ of the form $f(t, \boldsymbol{x})=a u_{b}(t) v_{\boldsymbol{\theta}}(\boldsymbol{x})$,

$$
\begin{equation*}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq d_{2}^{2}(\sqrt{s}, f)+\frac{k_{2}(1 \vee\|\boldsymbol{\theta}\|)}{n}+\frac{C^{\prime}}{n} \tag{4.5}
\end{equation*}
$$

where $C$ is a universal positive constant and where $C^{\prime}$ depends only on $\underline{\rho}, \bar{\rho}, b_{0}$ and $b$. More precisely,

$$
C^{\prime}=\log \left[1 \vee \bar{\rho}\left(b_{0}+\frac{b-b_{0}}{b-b_{0}+1}\right)\right]+|\log (1 \wedge \underline{\rho}(1+b))|+\left|\log \left(b-b_{0}\right)\right| .
$$

Under parametric assumptions on $s$, this result says that the rate of convergence of $\hat{s}$ is of order $n^{-1}$, which is quite satisfying when $n$ is large, but may be inadequate in a non-asymptotic point of view. Indeed, the second term of the right-hand side of inequality (4.5) may be large especially when $k_{2}$ is large, says larger than $n$. This difficulty can be overcome by considering that $\boldsymbol{\theta}$ is sparse, which means that $\boldsymbol{\theta}$ is close to some (unknown) linear subspace $W$ of $\mathbb{R}^{k_{2}}$ with $\operatorname{dim} W$ small. Below, we generalize Proposition 4.8 to take account of this situation.

Proposition 4.9. Let $\left(u_{b}\right)_{b \in I}$ be a family such that Assumption 4.5 holds. Let $\mathbb{W}$ be an at most countable family of linear subspaces of $\mathbb{R}^{k_{2}}$ and let $\Delta$ be a non-negative map on $\mathbb{W}$ such that $\sum_{W \in \mathbb{W}} e^{-\Delta(W)} \leq 1$.

There exist $\hat{a} \geq 0, \hat{b} \in I$ and $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{k_{2}}$, such that the estimator $\hat{s}=\left(\hat{a} u_{\hat{b}} v_{\hat{\boldsymbol{\theta}}}\right)^{2}$ satisfies, for all $a \geq 0, b \in I$, $\boldsymbol{\theta} \in \mathbb{R}^{k_{2}}$, and $f \in \mathscr{F}$ of the form $f(t, \boldsymbol{x})=a u_{b}(t) v_{\boldsymbol{\theta}}(\boldsymbol{x})$,

$$
\begin{aligned}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}(\sqrt{s}, f)+\frac{C^{\prime}}{n} \\
& +\inf _{W \in \mathbb{W}}\left\{a^{2}\left\|u_{b}\right\|_{\mathrm{t}}^{2} e^{2\|\theta\|} d^{2}(\boldsymbol{\theta}, W)+\frac{(1 \vee \operatorname{dim} W)(1 \vee\|\boldsymbol{\theta}\|)+\Delta(W)}{n}\right\}
\end{aligned}
$$

where $C$ is a universal positive constant and where $C^{\prime}$ is given by Proposition 4.8.
For illustration purpose, let us make explicit the constant $C^{\prime}$ for the Duane's model, and let us therefore assume that there exist some unknown parameters $a, b, \boldsymbol{\theta}$ such that $s$ is of the form $\sqrt{s(t, \boldsymbol{x})}=a t^{b} \exp (\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle)$. We derive from Proposition 4.8 an estimator whose risk satisfies

$$
\begin{equation*}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \frac{(1 \vee\|\boldsymbol{\theta}\|) k_{2}+|\log (2 b+1)|}{n} \tag{4.6}
\end{equation*}
$$

where $C$ is a universal positive constant. However, if for instance $k_{2}$ is large and if most of the components of $\boldsymbol{\theta}$ are small or null, the preceding proposition can be used to improve substantially the risk of our estimators. For simplicity, assume that

$$
k_{\star}=\left|\left\{j \in\left\{1, \ldots, k_{2}\right\}, \theta_{j} \neq 0\right\}\right|
$$

is small. We then define the set $\mathcal{M}$ of all subsets of $\left\{1, \ldots, k_{2}\right\}$, and for each $m \in \mathcal{M}$, the set

$$
W_{m}=\left\{\left(y_{1}, \ldots, y_{k_{2}}\right), \forall j \notin m, y_{j}=0\right\} \subset \mathbb{R}^{k_{2}}
$$

We apply Proposition 4.9 with

$$
\mathbb{W}=\left\{W_{m}, m \in \mathcal{M}\right\} \quad \text { and } \quad \forall m \in \mathcal{M}, \Delta\left(W_{m}\right)=1+|m|+\log \binom{k_{2}}{|m|}
$$

This leads to an estimator $\hat{s}$ such that

$$
C^{\prime \prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \frac{\left(1 \vee \log k_{2} \vee\|\boldsymbol{\theta}\|\right)\left(1 \vee k_{\star}\right)+|\log (2 b+1)|}{n}
$$

which improves inequality (4.6) when $k_{\star}$ is small and $k_{2}$ large.

## 5. PARAMETRIC MODELS

In this section, we consider the natural situation where the intensity of each process $N_{i}$ belongs (or is close) to a same parametric model. Throughout this section, $n \geq 2$.

We consider a closed rectangle $\Theta$ of $\mathbb{R}^{k}$, that is a subset of $\mathbb{R}^{k}$ for which there exist $m_{1}, \ldots, m_{k} \in \mathbb{R} \cup\{-\infty\}$ and $M_{1}, \ldots, M_{k} \in \mathbb{R} \cup\{\infty\}$ such that $\Theta=\mathbb{R}^{k} \cap \prod_{i=1}^{k}\left[m_{i}, M_{i}\right]$. We denote by $\mathcal{F}=\left\{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\right\}$ a class of functions of $\mathbb{L}^{2}(\mathbb{T}, \mu)$. Our aim is then to estimate $s$ when, for each $i \in\{1, \ldots n\}$, the square root of the intensity of the

Poisson's process $N_{i}, \sqrt{s\left(\cdot, x_{i}\right)}$, is (or is close to) an element of $\mathcal{F}$. We introduce thus the class of functions $\mathscr{F}$ defined by

$$
\mathscr{F}=\left\{(t, x) \mapsto f_{\boldsymbol{u}(x)}(t), \text { where } \boldsymbol{u} \text { is a map from } \mathbb{X} \text { into } \Theta\right\}
$$

For instance, if $\mathcal{F}$ corresponds to the Duane's model (see Sect. 4.3), $\Theta$ is a closed rectangle included in $\mathbb{R} \times$ $(-1 / 2,+\infty)$ and

$$
\mathcal{F}=\left\{a t^{b},(a, b) \in \Theta\right\} .
$$

The class $\mathscr{F}$ is then the set of all functions $f$ of the form $f(t, x)=a(x) t^{b(x)}$ where $a$ and $b$ are two functions on $\mathbb{X}$ such that $(a(x), b(x)) \in \Theta$ for all $x \in \mathbb{X}$.

We consider the following assumption to deal with more general classes $\mathcal{F}$.
Assumption 5.1. The set $\Theta$ is a closed rectangle of $\mathbb{R}^{k}$. There exist $\boldsymbol{\alpha}=\left(\alpha_{j}\right)_{1 \leq j \leq k} \in(0,1]^{k}$ and $\boldsymbol{R}=$ $\left(R_{j}\right)_{1 \leq j \leq k} \in(0,+\infty)^{k}$ such that

$$
\begin{equation*}
\forall \boldsymbol{\theta}, \boldsymbol{\theta}^{\prime} \in \Theta, \quad\left\|f_{\boldsymbol{\theta}}-f_{\boldsymbol{\theta}^{\prime}}\right\|_{\mathrm{t}} \leq \sum_{j=1}^{k} R_{j}\left|\theta_{j}-\theta_{j}^{\prime}\right|^{\alpha_{j}} \tag{5.1}
\end{equation*}
$$

The aim of the lemmas below is to prove that this assumption is satisfied for the Duane, Goel-Okumoto and S-Shaped's models.
Lemma 5.2. Let $\mu$ be the Lebesgue's measure, and for all $\boldsymbol{\theta} \in \mathbb{R} \times[-1 / 2,+\infty)$,

$$
f_{\boldsymbol{\theta}}(t)=\theta_{1} t^{\theta_{2}} \quad \text { for all } t \in \mathbb{T}=(0,1]
$$

Then, for all positive numbers $r_{1}, r_{2}$, and all $\boldsymbol{\theta}, \boldsymbol{\theta}^{\prime} \in\left[-r_{1}, r_{1}\right] \times\left[-1 / 2+1 / r_{2},+\infty\right)$,

$$
\left\|f_{\boldsymbol{\theta}}-f_{\boldsymbol{\theta}^{\prime}}\right\|_{\mathrm{t}} \leq r_{2}^{1 / 2}\left|\theta_{1}-\theta_{1}^{\prime}\right|+\sqrt{2} r_{1} r_{2}^{3 / 2}\left|\theta_{2}-\theta_{2}^{\prime}\right|
$$

Lemma 5.3. Let $\mu$ be the Lebesgue's measure, and for all $k \in\{0,1\}, \boldsymbol{\theta}=\left(\theta_{1}, \theta_{2}\right) \in \mathbb{R} \times(0,+\infty)$,

$$
f_{\boldsymbol{\theta}}(t)=\theta_{1} t^{k / 2} e^{-\theta_{2} t} \quad \text { for all } t \in \mathbb{T}=(0,+\infty)
$$

Let $r_{1}, r_{2}$ be two positive numbers and let us set

$$
C_{1}(0)=\left(r_{2} / 2\right)^{1 / 2} \quad C_{2}(0)=r_{1} r_{2}^{3 / 2} / 2 \quad C_{1}(1)=r_{2} / 2 \text { and } \quad C_{2}(1)=(3 / 8)^{1 / 2} r_{1} r_{2}^{2}
$$

For all $\boldsymbol{\theta}, \boldsymbol{\theta}^{\prime} \in\left[-r_{1}, r_{1}\right] \times\left[1 / r_{2},+\infty\right)$,

$$
\left\|f_{\boldsymbol{\theta}}-f_{\boldsymbol{\theta}^{\prime}}\right\|_{\mathrm{t}} \leq C_{1}(k)\left|\theta_{1}-\theta_{1}^{\prime}\right|+C_{2}(k)\left|\theta_{2}-\theta_{2}^{\prime}\right|
$$

Remark 5.4. For the Duane's model, Lemma 5.2 shows that Assumption 5.1 is fulfilled for $\Theta=\left[-r_{1}, r_{1}\right] \times$ $\left[-1 / 2+1 / r_{2},+\infty\right)$. This will allow to obtain a risk bound when $\sqrt{s}$ is close to the class

$$
\begin{array}{r}
\mathscr{F}_{r_{1}, r_{2}}=\left\{(t, x) \mapsto a(x) t^{b(x)}, \text { where } a \text { maps } \mathbb{X} \text { into }\left[-r_{1}, r_{1}\right]\right. \text { and } \\
\left.b \text { maps } \mathbb{X} \text { into }\left[-1 / 2+1 / r_{2},+\infty\right)\right\}
\end{array}
$$

By using Proposition 2.3 with $\mathfrak{F}=\left\{\mathscr{F}_{r_{1}, r_{2}}, r_{1}, r_{2} \in \mathbb{N}^{\star}\right\}$, we can also derive a risk bound for the class

$$
\begin{aligned}
\mathscr{F}= & \bigcup_{r_{1}, r_{2} \in \mathbb{N}^{\star}} \mathscr{F}_{r_{1}, r_{2}} \\
= & \left\{(t, x) \mapsto a(x) t^{b(x)}, \text { where } a \text { maps } \mathbb{X} \text { into a compact subset of } \mathbb{R}\right. \\
& \quad \text { and } b \text { maps } \mathbb{X} \text { into a closed interval included in }(-1 / 2,+\infty)\} .
\end{aligned}
$$

### 5.1. A model selection theorem

The main theorem of Section 5 is the following.
Theorem 5.5. Suppose that Assumption 5.1 holds. Let $\mathbb{W}_{1}, \ldots, \mathbb{W}_{k}$ be $k$ families of finite dimensional linear subspaces of $\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$. Let for each $j \in\{1, \ldots, k\}, \Delta_{j}$ be a non-negative mapping on $\mathbb{W}_{j}$ such that $\sum_{W_{j} \in \mathbb{W}_{j}} e^{-\Delta_{j}\left(W_{j}\right)} \leq 1$.

There exists an estimator $\hat{s}$ such that for all map $\boldsymbol{u}=\left(u_{1}, \ldots, u_{k}\right)$ from $\mathbb{X}$ with values into $\Theta$, and $f \in \mathscr{F}$ of the form $f(t, x)=f_{\boldsymbol{u}(x)}(t)$,

$$
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq d_{2}^{2}(\sqrt{s}, f)+\sum_{j=1}^{k} \varepsilon_{j}\left(u_{j}\right)
$$

where $\varepsilon_{j}\left(u_{j}\right)$ is defined by

$$
\varepsilon_{j}\left(u_{j}\right)=\inf _{W_{j} \in \mathbb{W}_{j}}\left\{R_{j}^{2}\left(d_{\mathbf{x}}\left(u_{j}, W_{j}\right)\right)^{2 \alpha_{j}}+\frac{\left(\operatorname{dim}\left(W_{j}\right) \vee 1\right) \tau_{u, j}(n)+\Delta_{j}\left(W_{j}\right)}{n}\right\},
$$

where

$$
\tau_{\boldsymbol{u}, j}(n)=\log n+\log \left(1 \vee R_{j}\right)+\log \left(1 \vee\left\|u_{j}\right\|_{\mathrm{x}}\right),
$$

and where $C>0$ depends only on $k$ and $\alpha_{1}, \ldots, \alpha_{k}$.
Roughly speaking, this result says that the risk bound we get when $\sqrt{s}$ is of the form $\sqrt{s(t, x)}=f_{\boldsymbol{u}(x)}(t)$, corresponds to the one we would get if we could apply a model selection theorem on the components $u_{1}, \ldots, u_{k}$ separately. Each term $\varepsilon_{j}\left(u_{j}\right)$ can be controlled under structural or smoothness assumptions on $u_{j}$. For instance, if $\mathbb{X}=[0,1]^{k_{2}}$ and if $u_{j}$ is assumed to belong to the class $\mathscr{F}_{j}=\mathcal{H}\left([0,1]^{k_{2}}\right)$, a suitable choice of $\left(\mathbb{W}_{j}, \Delta_{j}\right)$ leads to

$$
C_{j} \varepsilon_{j}\left(u_{j}\right) \leq\left(R_{j} L\left(u_{j}\right)^{\alpha_{j}}\right)^{\frac{2 k_{2}}{k_{2}+2 \alpha_{j} \beta_{j}}}\left(\frac{\tau_{\boldsymbol{u}, j}(n)}{n}\right)^{\frac{2 \alpha_{j} \bar{\beta}_{j}}{2 \alpha_{j} \beta_{j}+k_{2}}}+\frac{\tau_{\boldsymbol{u}, j}(n)}{n}
$$

where $\boldsymbol{\beta}_{j}$ is such that $u_{j} \in \mathcal{H}^{\boldsymbol{\beta}_{j}}\left([0,1]^{k_{2}}\right)$ and where $C_{j}>0$ depends only on $k_{2}$ and $\boldsymbol{\beta}_{j}$. In particular, if $\alpha_{j}=1$ and if $n$ is large, $\varepsilon_{j}\left(u_{j}\right)$ is of order $(\log n / n)^{2 \bar{\beta}_{j} /\left(2 \bar{\beta}_{j}+k_{2}\right)}$. Apart from the logarithmic factor, this corresponds to the estimation rate of an Hölderian function on $[0,1]^{k_{2}}$.

The corollary below illustrates this result for the Duane's model.
Corollary 5.6. There exists an estimator $\hat{s}$ such that, for all $\boldsymbol{\alpha} \in(0,+\infty)^{k_{2}}, \boldsymbol{\beta} \in(0,+\infty)^{k_{2}}$, for all a $\in$ $\mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^{k_{2}}\right), b \in \mathcal{H}^{\boldsymbol{\beta}}\left([0,1]^{k_{2}}\right)$ satisfying $b>-1 / 2$, and for all function $f$ of the form $f(t, x)=a(x) t^{b(x)}$,

$$
\begin{aligned}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}(\sqrt{s}, f) \\
& +\left(\frac{1}{\left.1 \wedge \inf _{x \in[0,1]^{k_{2}}(2 b(x)+1)}\right)^{\frac{k_{2}}{2 \alpha+k_{2}}} L(a)^{\frac{2 k_{2}}{2 \bar{\alpha}+k_{2}}}\left(\frac{\log n}{n}\right)^{\frac{2 \bar{\alpha}}{2 \alpha+k_{2}}}}\right. \\
& +\left(\frac{1 \vee\|a\|_{\infty}^{2}}{\left.1 \wedge \inf _{x \in[0,1]^{k_{2}}(2 b(x)+1)^{3}}\right)^{\frac{k_{2}}{2 \beta+k_{2}}} L(b)^{\frac{2 k_{2}}{2 \beta+k_{2}}}\left(\frac{\log n}{n}\right)^{\frac{2 \bar{\beta}}{2 \beta+k_{2}}}}\right. \\
& +C^{\prime} \frac{\log n}{n}
\end{aligned}
$$

where $C>0$ depends on $k_{2}$, $\max _{1 \leq j \leq k} \alpha_{j}$, $\max _{1 \leq j \leq k} \beta_{j}$, and where $C^{\prime}$ depends on $L(a), L(b), \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}},\|a\|_{\infty}$, $\|b\|_{\infty}$ and $\inf _{x \in[0,1]^{k_{2}}}(2 b(x)+1)$.

### 5.2. Change point detection

In the case where the intensity $s_{i}$ of each $N_{i}$ is of the form $\sqrt{s_{i}(t)}=f_{\boldsymbol{\theta}_{i}}(t)$, a natural way to control the risk of our estimator $\hat{s}$ is to consider some assumptions on the map $i \mapsto \boldsymbol{\theta}_{i}$. This problem amounts to choosing suitable collections $\mathbb{W}_{1}, \ldots, \mathbb{W}_{k}$ to approximate functions on $\mathbb{X}=\{1, \ldots, n\}$.

In this section, we focus on the case where the map $i \mapsto \boldsymbol{\theta}_{i}$ is piecewise constant with a small number of jumps. Let $\mathcal{P}$ be the set of partitions of $\{1, \ldots, n\}$ into intervals. We aim at estimating $s$ when there exists a partition $P_{0} \in \mathcal{P}$ such that $s$ is of the form

$$
\begin{equation*}
\forall I \in P_{0}, \exists \boldsymbol{\theta}_{I} \in \Theta, \forall i \in I, \quad \sqrt{s_{i}(t)}=f_{\boldsymbol{\theta}_{I}}(t) \quad \text { for all } t \in \mathbb{T} \tag{5.2}
\end{equation*}
$$

We define for each partition $P \in \mathcal{P}$, the linear space of piecewise constant functions

$$
W_{P}=\left\{\sum_{I \in P} a_{I} \mathbb{1}_{I}, a_{I} \in \mathbb{R}\right\}
$$

and apply Theorem 5.5 with the collections and maps defined by

$$
\forall j \in\{1, \ldots, k\}, \mathbb{W}_{j}=\left\{W_{P}, P \in \mathcal{P}\right\} \text { and } \Delta_{j}\left(W_{P}\right)=|P|+\log \binom{n-1}{|P|-1}
$$

This leads to the result below.
Corollary 5.7. Assume that Assumption 5.1 and relation (5.2) hold. There exists an estimator $\hat{s}$ such that

$$
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq\left|P_{0}\right| \frac{\log n+C^{\prime}}{n}
$$

where $C>0$ depends only on $k$ and $\alpha_{1}, \ldots, \alpha_{k}$, where $C^{\prime}$ is given by

$$
C^{\prime}=\sup _{1 \leq j \leq k}\left(\log \left(1+R_{j}\right)\right)+\sup _{I \in P}\left(\log \left(1+\left\|\boldsymbol{\theta}_{I}\right\|_{\infty}\right)\right)
$$

and where $\left\|\boldsymbol{\theta}_{I}\right\|_{\infty}=\sup _{1 \leq j \leq k}\left|\left(\boldsymbol{\theta}_{I}\right)_{j}\right|$.
For illustration purpose, in the context of the Duane's model, there exist $a_{1} \ldots, a_{n} \in(0,+\infty)$, and $b_{1}, \ldots, b_{n} \in$ $(-1 / 2,+\infty)$ such that $\sqrt{s_{i}(t)}=a_{i} t^{b_{i}}$ for all $t \in(0,1]$. By combining the preceding corollary with Proposition 2.3, we build an estimator $\hat{s}$ such that

$$
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq\left(1+r_{1}+r_{2}\right) \frac{\log n+C^{\prime}}{n}
$$

where $r_{1}$ and $r_{2}$ are the numbers of jumps of the maps $i \mapsto a_{i}$ and $i \mapsto b_{i}$ respectively, where $C$ is a universal positive constant, and where $C^{\prime}$ depends on $\sup _{1 \leq i \leq n} a_{i}, \sup _{1 \leq i \leq n}\left|b_{i}\right|$ and $\inf _{1 \leq i \leq n}\left(2 b_{i}+1\right)$.

## 6. Proofs

### 6.1. Proof of Theorem 2.2

Throughout the proof, we set $\mathbf{N}=\left(N_{1}, \ldots, N_{n}\right)$ and $\mathbf{x}=\left(x_{1}, \ldots, x_{n}\right)$.

### 6.1.1. About the T-estimators

We begin to briefly recall the general strategy introduced in [7] to build estimators from tests.
Given two distinct functions $f, f^{\prime}$ of $\mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$, a test function $\psi_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})$ is a measurable function with values in $\left\{f, f^{\prime}\right\}$. The convention is that $\psi_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})=f$ means accepting $f$ whereas $\psi_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})=f^{\prime}$ means accepting $f^{\prime}$. In what follows, we need tests with the following properties. We shall build them in Section 6.1.2.
Assumption 6.1. There exist $a>0, \kappa>0$ such that for all distinct functions $f, f^{\prime} \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$ and all $z \in \mathbb{R}$, there exists a test $\psi_{f, f^{\prime}}^{(z)}(\mathbf{N}, \mathbf{x})$ satisfying

$$
\begin{align*}
& \sup _{\substack{f \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M), \kappa H(s, f) \leq H\left(f, f^{\prime}\right)}} \mathbb{P}\left[\psi_{f, f^{\prime}}^{(z)}(\mathbf{N}, \mathbf{x})=f^{\prime}\right] \leq \exp \left[-a n\left(H^{2}\left(f, f^{\prime}\right)+z\right)\right]  \tag{6.1}\\
& \sup _{\substack{f \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M), \kappa H\left(s, f^{\prime}\right) \leq H\left(f, f^{\prime}\right)}} \mathbb{P}\left[\psi_{f, f^{\prime}}^{(z)}(\mathbf{N}, \mathbf{x})=f\right] \leq \exp \left[-a n\left(H^{2}\left(f, f^{\prime}\right)-z\right)\right] . \tag{6.2}
\end{align*}
$$

We now consider an at most countable collection $\mathbb{S}$ of subsets of $\mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$. We shall assume that the sets $S \in \mathbb{S}$ are $D$-models. We recall the definition below.
Definition 6.2. An at most countable subset $S$ of $\mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$ is called a $D$-model with parameters $\bar{\eta}_{S}, \bar{D}_{S}$ and 1 if

$$
\left|S \cap \mathcal{B}\left(f, r \bar{\eta}_{S}\right)\right| \leq \exp \left[\bar{D}_{S} r^{2}\right] \quad \text { for all } r \geq 2 \text { and } f \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)
$$

where $\mathcal{B}\left(f, r \bar{\eta}_{S}\right)$ is the closed ball centered at $f$ with radius $r \bar{\eta}_{S}$ of the metric space $\left(\mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M), H\right)$.
The tests allow to select among the functions of $\cup_{S \in \mathbb{S}} S$. Precisely, the selection rule is the following.
Given a collection $\mathbb{S}$ of $D$-models, we set for all $f \in \cup_{S \in \mathbb{S}} S$,

$$
\bar{\eta}(f)=\inf \left\{\bar{\eta}_{S}, S \in \mathbb{S}, S \ni f\right\}
$$

and for all $f^{\prime} \in \cup_{S \in \mathbb{S}} S, f^{\prime} \neq f, z_{f, f^{\prime}}=\bar{\eta}\left(f^{\prime}\right)^{2}-\bar{\eta}(f)^{2}$. We define for all $f \in \cup_{S \in \mathbb{S}} S$,

$$
\mathcal{R}(f)=\left\{f^{\prime} \in \cup_{S \in S} S, \psi_{f, f^{\prime}}^{\left(z_{f, f^{\prime}}\right)}(\mathbf{N}, \mathbf{x})=f^{\prime}\right\}
$$

and consider

$$
\gamma(f)= \begin{cases}\sup \left\{H\left(f, f^{\prime}\right), f^{\prime} \in \mathcal{R}(f)\right\} & \text { if } \mathcal{R}(f) \neq \emptyset \\ 0 & \text { if } \mathcal{R}(f)=\emptyset\end{cases}
$$

Given $\varepsilon>0$, a $T_{\varepsilon}$-estimator is a measurable function $\hat{s}=\hat{s}(\mathbf{N}, \mathbf{x})$ with values in $\cup_{S \in \mathbb{S}} S$ such that

$$
\gamma(\hat{s}) \vee \varepsilon \bar{\eta}(\hat{s})=\inf _{f \in \cup S \in \mathbb{S} S}[\gamma(f) \vee \varepsilon \bar{\eta}(f)]
$$

Theorem 5 of [7] shows that such a minimizer exists almost surely and they all possess similar theoretical properties. In our framework, we can rewrite it as follows.
Theorem 6.3. Suppose that Assumption 6.1 holds. Let $\mathbb{S}$ be an at most countable collection of $D$-models such that $\bar{D}_{S} \geq 1 / 2$ for all $S \in \mathbb{S}$,

$$
\sum_{S \in \mathbb{S}} \exp \left(-\frac{a n \bar{\eta}_{S}^{2}}{21}\right) \leq 1 \quad \text { and } \quad a n \bar{\eta}_{S}^{2} \geq \frac{21 \bar{D}_{S}}{5} \quad \text { for all } S \in \mathbb{S}
$$

For all $\varepsilon \in(0,4]$, there exists almost surely a $T_{\varepsilon}$-estimator $\hat{s} \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$. Moreover, any of them satisfies

$$
\mathbb{P}\left[C H^{2}(s, \hat{s}) \geq \inf _{S \in \mathbb{S}}\left\{H^{2}(s, S)+\bar{\eta}_{S}^{2}\right\}+\xi\right] \leq e^{-n \xi} \quad \text { for all } \xi>0
$$

where $C>0$ depends only on $a, \kappa$.
It remains thus to construct the tests and the collection $\mathbb{S}$ to prove Theorem 2.2.
6.1.2. Definition of the tests

Our tests are inspired from the variational formula in [3]. Let, for all functions $f, f^{\prime}$ of $\mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$, $T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})$ be the functional

$$
\begin{aligned}
T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})= & \frac{1}{2 n} \sum_{i=1}^{n} \int_{\mathbb{T}} \sqrt{\frac{f\left(t, x_{i}\right)+f^{\prime}\left(t, x_{i}\right)}{2}}\left(\sqrt{f^{\prime}\left(t, x_{i}\right)}-\sqrt{f\left(t, x_{i}\right)}\right) \mathrm{d} \mu(t) \\
& +\frac{1}{\sqrt{2} n} \sum_{i=1}^{n} \int_{\mathbb{T}} \frac{\sqrt{f^{\prime}\left(t, x_{i}\right)}-\sqrt{f\left(t, x_{i}\right)}}{\sqrt{f\left(t, x_{i}\right)+f^{\prime}\left(t, x_{i}\right)}} \mathrm{d} N_{i}(t) \\
& -\frac{1}{2 n} \sum_{i=1}^{n} \int_{\mathbb{T}}\left(f^{\prime}\left(t, x_{i}\right)-f\left(t, x_{i}\right)\right) \mathrm{d} \mu(t)
\end{aligned}
$$

where the convention $0 / 0$ is in use. We prove the following.
Lemma 6.4. There exist positive numbers $a, b$ such that for all $z \in \mathbb{R}$ and all $f, f^{\prime} \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M)$ satisfying $4 H(s, f) \leq H\left(f, f^{\prime}\right)$,

$$
\mathbb{P}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq b z\right] \leq \exp \left[-n a\left(H^{2}\left(f, f^{\prime}\right)+z\right)\right]
$$

The proof of this lemma is deferred to Section 6.1.6 and we refer to the proof for the exact values of $a$ and $b$.
This lemma says that the functional $T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})$ can be used to construct the tests. Precisely, we set for all $z \in \mathbb{R}, f, f^{\prime} \in \mathbb{L}_{+}^{1}(\mathbb{T} \times \mathbb{X}, M), f \neq f^{\prime}$,

$$
\psi_{f, f^{\prime}}^{(z)}(\mathbf{N}, \mathbf{x})= \begin{cases}f^{\prime} & \text { if } T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})>b z \\ f & \text { if } T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})<b z\end{cases}
$$

and $\psi_{f, f^{\prime}}^{(z)}(\mathbf{N}, \mathbf{x})$ is defined arbitrarily in case of equality. Thanks to the above lemma, (6.1) holds. Note that (6.2) also holds since $T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})=-T_{f^{\prime}, f}(\mathbf{N}, \mathbf{x})$.

### 6.1.3. Construction of $\mathbb{S}$

The collection $\mathbb{S}$ is derived from $\mathbb{V}$. We shall show in Section 6.1.6 the following lemma.
Lemma 6.5. For all $\eta>0$ and $V \in \mathbb{V}$ there exists a $D$-model $\bar{S}_{V}(\eta)$ with parameters $\eta, 63 D_{V}(\eta / 2)$ and 1. Moreover,

$$
\begin{equation*}
H\left(s, \bar{S}_{V}(\eta)\right) \leq 2 \sqrt{2}\left(d_{2}(\sqrt{s}, V)+\eta\right) \tag{6.3}
\end{equation*}
$$

and for all $f \in \bar{S}_{V}(\eta)$, there exists $g \in V$ such that $\sqrt{f}=g \vee 0$.
Please note that we can assume (for the sake of simplicity and with no loss of generality), that $D_{V}$ is nonincreasing. We then set for all $V \in \mathbb{V}$,

$$
\bar{\eta}_{\bar{S}_{V}}=\left(21 \sqrt{\frac{3}{5 a}} \eta_{V}\right) \vee \sqrt{\frac{21 \Delta(V)}{n a}} \quad \text { and } \quad \bar{S}_{V}=\bar{S}_{V}\left(\bar{\eta}_{\bar{S}_{V}}\right)
$$

where $a$ is given by Lemma 6.4. Actually $a$ is very small (smaller than 1 ), which implies that $\bar{\eta}_{\bar{S}_{V}} / 2 \geq \eta_{V}$ and thus

$$
63 D_{V}\left(\bar{\eta}_{\bar{S}_{V}} / 2\right) \leq 63 D_{V}\left(\eta_{V}\right)
$$

Consequently, the set $\bar{S}_{V}$ is a $D$-model with parameters $\bar{\eta}_{\bar{S}_{V}}, \bar{D}_{\bar{S}_{V}}=63 D_{V}\left(\eta_{V}\right)$ and 1 . The collection $\mathbb{S}$ is then defined by $\mathbb{S}=\left\{\bar{S}_{V}, V \in \mathbb{V}\right\}$.

### 6.1.4. Proof of Theorem 2.2

The assumptions of Theorem 6.3 are fulfilled:

$$
a n \bar{\eta}_{\bar{S}_{V}}^{2} \geq \frac{21^{2} \times 3}{5} n \eta_{V}^{2} \geq \frac{21^{2} \times 3}{5} D_{V}\left(\eta_{V}\right) \geq \frac{21 \bar{D}_{\bar{S}_{V}}}{5}
$$

and

$$
\sum_{V \in \mathbb{V}} \exp \left(-\frac{a n \bar{\eta}_{S_{V}}^{2}}{21}\right) \leq \sum_{V \in \mathbb{V}} \exp (-\Delta(V)) \leq 1
$$

The selection rule described in Section 6.1.1 provides thus an estimator $\hat{s} \in \cup_{V \in \mathbb{V}} \bar{S}_{V}$ such that, for all $\xi>0$,

$$
\mathbb{P}\left[C H^{2}(s, \hat{s}) \geq \inf _{V \in \mathbb{V}}\left\{H^{2}\left(s, \bar{S}_{V}\right)+\bar{\eta}_{\bar{S}_{V}}^{2}\right\}+\xi\right] \leq e^{-n \xi}
$$

where $C>0$ is universal. By using inequality (6.3),

$$
H^{2}\left(s, \bar{S}_{V}\right) \leq 16\left[d_{2}^{2}(\sqrt{s}, V)+\bar{\eta}_{\bar{S}_{V}}^{2}\right]
$$

and hence

$$
\inf _{V \in \mathbb{V}}\left\{H^{2}\left(s, \bar{S}_{V}\right)+\bar{\eta}_{\bar{S}_{V}}^{2}\right\} \leq C^{\prime} \inf _{V \in \mathbb{V}}\left\{d_{2}^{2}(\sqrt{s}, V)+\eta_{V}^{2}+\frac{\Delta(V)}{n}\right\}
$$

for some universal constant $C^{\prime}>0$. Finally,

$$
\mathbb{P}\left[C^{\prime \prime} H^{2}(s, \hat{s}) \geq \inf _{V \in \mathbb{V}}\left\{d_{2}^{2}(\sqrt{s}, V)+\eta_{V}^{2}+\frac{\Delta(V)}{n}\right\}+\xi\right] \leq e^{-n \xi}
$$

where $C^{\prime \prime}=C /\left(C^{\prime} \vee 1\right)$.

### 6.1.5. Proof of Lemma 6.4

We start with the following Bennett-type's inequality which generalizes Proposition 7 of [17].
Lemma 6.6. Let $f_{1}, \ldots, f_{n}$ be $n$ bounded measurable functions. Let $\rho, v$ be positive numbers such that $\rho \geq$ $\max _{1 \leq i \leq n}\left\|f_{i}\right\|_{\infty}$ and

$$
\frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{T}} f_{i}^{2}(t) s_{i}(t) \mathrm{d} \mu(t) \leq v
$$

Then, for all $r \geq 0$,

$$
\begin{aligned}
\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n}\left[\int_{\mathbb{T}} f_{i}(t) \mathrm{d} N_{i}(t)-\mathbb{E}\left(\int_{\mathbb{T}} f_{i}(t) \mathrm{d} N_{i}(t)\right)\right] \geq r\right) & \leq \exp \left(-n \frac{v}{b^{2}} h\left(\frac{\rho r}{v}\right)\right) \\
& \leq \exp \left(-n \frac{r^{2}}{2\left(v+\frac{\rho r}{3}\right)}\right)
\end{aligned}
$$

where $h$ is the function defined for $u \in(-1,+\infty)$ by $h(u)=(1+u) \log (1+u)-u$.
Proof. By homogeneity we can assume that $\rho=1$. We assume moreover that for each $i \in\{1, \ldots, n\}, f_{i}$ is a piecewise constant function (with a finite number of pieces). There exist thus $k_{1}, \ldots, k_{n} \in \mathbb{N}^{\star}$ and a family $\left.\left(a_{i, j}\right)\right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq k_{i}}}$ of elements of $[-1,1]$ such that

$$
\forall t \in \mathbb{T}, \quad f_{i}(t)=\sum_{j=1}^{k_{i}} a_{i, j} \mathbb{1}_{A_{i, j}}(t)
$$

where the $A_{i, j}$ are measurable sets of $\mathbb{T}$ such that $A_{i, j} \cap A_{i, j^{\prime}}=\emptyset$ for all $j \neq j^{\prime}$.

Now, for all $\xi>0$,

$$
\begin{aligned}
\log \mathbb{E}\left(e^{\xi \sum_{i=1}^{n}\left[\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}-\mathbb{E}\left(\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}\right)\right]}\right) & =\sum_{i=1}^{n} \log \mathbb{E}\left(e^{\xi\left[\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}-\mathbb{E}\left(\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}\right)\right]}\right) \\
& =\sum_{i=1}^{n} \sum_{j=1}^{k_{i}} \log \mathbb{E}\left(e^{\xi a_{i, j}\left[N_{i}\left(A_{i, j}\right)-\mathbb{E}\left(N_{i}\left(A_{i, j}\right)\right)\right]}\right) \\
& =\sum_{i=1}^{n} \sum_{j=1}^{k_{i}} \mathbb{E}\left(N_{i}\left(A_{i, j}\right)\right)\left(e^{\xi a_{i, j}}-\xi a_{i, j}-1\right) .
\end{aligned}
$$

By using the monotony of the function $x \mapsto\left(e^{x}-x-1\right) / x^{2}$,

$$
\begin{aligned}
\log \mathbb{E}\left(e^{\xi \sum_{i=1}^{n}\left[\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}-\mathbb{E}\left(\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}\right)\right]}\right) & \leq \sum_{i=1}^{n} \sum_{j=1}^{k_{i}} \mathbb{E}\left(a_{i, j}^{2} N_{i}\left(A_{i, j}\right)\right)\left(e^{\xi}-\xi-1\right) \\
& \leq n v\left(e^{\xi}-\xi-1\right) .
\end{aligned}
$$

This inequality still holds when the $f_{i}$ are not piecewise constant since a measurable function can be approximated by piecewise constant functions. Indeed, there exists a sequence $\left(f_{i}^{(k)}\right)_{k \geq 1}$ of piecewise constant functions (with a finite number of jumps) such that $f_{i}^{(k)} \rightarrow f_{i}$ when $k \rightarrow+\infty$ in the space $\mathbb{L}^{2}\left(\mathbb{T}, s_{i} \mathrm{~d} \mu\right)$ and such that $\left\|f_{i}^{(k)}\right\|_{\infty} \leq 1$ whatever $k, i$. By using Fatou's lemma,

$$
\log \mathbb{E}\left(\liminf _{k \rightarrow+\infty} e^{\xi \sum_{i=1}^{n}\left[\int_{\mathrm{T}} f_{i}^{(k)} \mathrm{d} N_{i}-\int_{\mathbb{T}} f_{i}^{(k)} s_{i} \mathrm{~d} \mu_{i}\right]}\right) \leq n v\left(e^{\xi}-\xi-1\right) .
$$

Since,

$$
\mathbb{E}\left[\left|\int_{\mathbb{T}} f_{i}^{(k)} \mathrm{d} N_{i}-\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}\right|\right] \leq \int_{\mathbb{T}}\left|f_{i}^{(k)}-f_{i}\right| s_{i} \mathrm{~d} \mu \rightarrow 0
$$

one can assume (up to considering a subsequence) that $\int_{\mathbb{T}} f_{i}^{(k)} \mathrm{d} N_{i}-\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i} \rightarrow 0$ almost surely (for all $i \in\{1, \ldots n\}$ ). We then have

$$
\log \mathbb{E}\left(e^{\xi \sum_{i=1}^{n}\left[\int_{\mathbb{T}} f_{i} \mathrm{~d} N_{i}-\int_{\mathrm{T}} f_{i} s_{i} \mathrm{~d} \mu_{i}\right]}\right) \leq n v\left(e^{\xi}-\xi-1\right)
$$

as wished. The exponential inequality is then deduced from the Cramér-Chernoff's method, see Chapter 2 of [16].

Let us return to the proof of Lemma 6.4. We define the function $\zeta$ on $[0,+\infty)^{2}$ by

$$
\zeta(x, y)=\frac{1}{\sqrt{2}}\left(\sqrt{\frac{y}{x+y}}-\sqrt{\frac{x}{x+y}}\right) \quad \text { for all } x, y \in[0,+\infty)
$$

where we use the convention $0 / 0=0$. Let then

$$
\begin{aligned}
Z_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) & =T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})-\mathbb{E}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})\right] \\
& =\int_{\mathbb{T} \times \mathbb{X}} \zeta\left(f, f^{\prime}\right) \mathrm{d} M-\mathbb{E}\left(\int_{\mathbb{T} \times \mathbb{X}} \zeta\left(f, f^{\prime}\right) \mathrm{d} M\right) .
\end{aligned}
$$

We use the claim below whose proof ensues from the proofs of Propositions 2 and 3 of [3].

## Claim 6.7.

$$
\mathbb{E}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})\right] \leq\left(1+\frac{1}{\sqrt{2}}\right) H^{2}(s, f)-\left(1-\frac{1}{\sqrt{2}}\right) H^{2}\left(s, f^{\prime}\right)
$$

and

$$
\frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{T}} \zeta^{2}\left(f\left(t, x_{i}\right), f^{\prime}\left(t, x_{i}\right)\right) s\left(t, x_{i}\right) \mathrm{d} \mu(t) \leq H^{2}(s, f)+H^{2}\left(s, f^{\prime}\right)+H^{2}\left(f, f^{\prime}\right) .
$$

We derive from the first point of the claim that

$$
\begin{aligned}
\mathbb{P}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq z\right] & =\mathbb{P}\left[Z_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq z-\mathbb{E}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})\right]\right] \\
& \leq \mathbb{P}\left[Z_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq z-\left(1+\frac{1}{\sqrt{2}}\right) H^{2}(s, f)+\left(1-\frac{1}{\sqrt{2}}\right) H^{2}\left(s, f^{\prime}\right)\right] .
\end{aligned}
$$

Note that the random variable $Z\left(f, f^{\prime}\right)$ can be written as

$$
Z_{f, f^{\prime}}(\mathbf{N}, \mathbf{x})=\frac{1}{n} \sum_{i=1}^{n}\left[\int_{\mathbb{T}} \zeta\left(f\left(\cdot, x_{i}\right), f^{\prime}\left(\cdot, x_{i}\right)\right) \mathrm{d} N_{i}-\mathbb{E}\left(\int_{\mathbb{T}} \zeta\left(f\left(\cdot, x_{i}\right), f^{\prime}\left(\cdot, x_{i}\right)\right) \mathrm{d} N_{i}\right)\right] .
$$

When

$$
r=z-\left(1+\frac{1}{\sqrt{2}}\right) H^{2}(s, f)+\left(1-\frac{1}{\sqrt{2}}\right) H^{2}\left(s, f^{\prime}\right)
$$

is non-negative, we apply Lemma 6.6 with $f_{i}(\cdot)=\zeta\left(f\left(\cdot, x_{i}\right), f^{\prime}\left(\cdot, x_{i}\right)\right), \rho=1 / \sqrt{2}$ and

$$
v=H^{2}(s, f)+H^{2}\left(s, f^{\prime}\right)+H^{2}\left(f, f^{\prime}\right)
$$

to obtain

$$
\mathbb{P}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq z\right] \leq \exp \left(-\frac{n r^{2}}{2 v+\frac{r \sqrt{2}}{3}}\right)
$$

We now bound from above the right-hand side of this inequality.
For this, we begin to bound $v$ from above. We deduce from the triangular inequality and from $4 H(s, f) \leq$ $H\left(f, f^{\prime}\right)$ that

$$
\begin{aligned}
v & \leq 3 H^{2}(s, f)+3 H^{2}\left(f, f^{\prime}\right) \\
& \leq 3(1+1 / 16) H^{2}\left(f, f^{\prime}\right) .
\end{aligned}
$$

Now, we bound $r$ from below. Note that

$$
H\left(f, f^{\prime}\right) \leq H(s, f)+H\left(s, f^{\prime}\right) \leq \frac{1}{4} H\left(f, f^{\prime}\right)+H\left(s, f^{\prime}\right)
$$

and thus $H\left(s, f^{\prime}\right) \geq 3 / 4 H\left(f, f^{\prime}\right)$. This leads to

$$
\begin{aligned}
r & \geq z-\left(1+\frac{1}{\sqrt{2}}\right) \frac{1}{16} H^{2}\left(f, f^{\prime}\right)+\left(1-\frac{1}{\sqrt{2}}\right) \frac{9}{16} H^{2}\left(f, f^{\prime}\right) \\
& \geq z+C H^{2}\left(f, f^{\prime}\right)
\end{aligned}
$$

where $C=(8-5 \sqrt{2}) / 16>0$. There are two types of cases involved.

- If $z+C H^{2}\left(f, f^{\prime}\right)>0, r$ is non-negative and thus

$$
\mathbb{P}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq z\right] \leq \exp \left(-\frac{n\left(z+C H^{2}\left(f, f^{\prime}\right)\right)^{2}}{6(1+1 / 16) H^{2}\left(f, f^{\prime}\right)+\frac{\sqrt{2}}{3}\left(z+C H^{2}\left(f, f^{\prime}\right)\right)}\right)
$$

Set $C^{\prime}=9 \sqrt{2}(1+1 / 16)+C$. Then,

$$
\mathbb{P}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq z\right] \leq \exp \left(-\frac{3 n}{\sqrt{2}} \frac{\left(z+C H^{2}\left(f, f^{\prime}\right)\right)^{2}}{z+C^{\prime} H^{2}\left(f, f^{\prime}\right)}\right)
$$

One can then verify that

$$
\begin{aligned}
\frac{\left(z+C H^{2}\left(f, f^{\prime}\right)\right)^{2}}{z+C^{\prime} H^{2}\left(f, f^{\prime}\right)} & =\frac{C^{2}}{C^{\prime}} H^{2}\left(f, f^{\prime}\right)+\frac{\left(2 C^{\prime}-C\right) C}{C^{\prime 2}} z+\frac{\left(C-C^{\prime}\right)^{2} z^{2}}{C^{\prime 2}\left(z+C^{\prime} H^{2}\left(f, f^{\prime}\right)\right)} \\
& \geq \frac{C^{2}}{C^{\prime}} H^{2}\left(f, f^{\prime}\right)+\frac{\left(2 C^{\prime}-C\right) C}{C^{\prime 2}} z
\end{aligned}
$$

which implies that

$$
\begin{equation*}
\mathbb{P}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq z\right] \leq \exp \left(-\frac{3 n}{\sqrt{2}}\left(\frac{C^{2}}{C^{\prime}} H^{2}\left(f, f^{\prime}\right)+\frac{\left(2 C^{\prime}-C\right) C}{C^{\prime 2}} z\right)\right) \tag{6.4}
\end{equation*}
$$

- If now $z+C H^{2}\left(f, f^{\prime}\right) \leq 0$,

$$
\begin{aligned}
\frac{C^{2}}{C^{\prime}} H^{2}\left(f, f^{\prime}\right)+\frac{\left(2 C^{\prime}-C\right) C}{C^{\prime 2}} z & \leq\left(\frac{C^{2}}{C^{\prime}}-\frac{\left(2 C^{\prime}-C\right) C^{2}}{C^{\prime 2}}\right) H^{2}\left(f, f^{\prime}\right) \\
& \leq 0
\end{aligned}
$$

Consequently, (6.4) also holds.
We thus have proved that

$$
\mathbb{P}\left[T_{f, f^{\prime}}(\mathbf{N}, \mathbf{x}) \geq b z\right] \leq \exp \left[-n a\left(H^{2}\left(f, f^{\prime}\right)+z\right)\right]
$$

where

$$
\begin{aligned}
& a=\frac{3 C^{2}}{\sqrt{2} C^{\prime}} \simeq 4.5 \times 10^{-4} \\
& b=\frac{C C^{\prime}}{2 C^{\prime}-C} \simeq 0.029 .
\end{aligned}
$$

This ends the proof.

### 6.1.6. Proof of Lemma 6.5.

By using Proposition 7 of $[7]$, we derive from $S_{V}(\eta)$ a set $S_{V}^{\prime}(\eta) \subset V$ such that

$$
\forall \varphi \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M), \forall r \geq 2, \quad\left|S_{V}^{\prime}(\eta) \cap \mathcal{B}(\varphi, r \eta)\right| \leq \exp \left(7 D_{V}(\eta / 2) r^{2}\right)
$$

where $\mathcal{B}(\varphi, r \eta)$ is the ball centered at $\varphi$ with radius $r \eta$ of the metric space $\left(\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M), d_{2}\right)$, and such that

$$
\forall f \in V, \quad d_{2}\left(f, S_{V}^{\prime}(\eta)\right) \leq \eta
$$

Proposition 12 of [7] (applied with $T=S_{V}^{\prime}(\eta)$, the cone $M_{0}$ of non-negative functions of $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M), M^{\prime}=$ $\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ and $\bar{\pi}$ defined by $\left.\bar{\pi}(f)=f \vee 0\right)$ provides a subset $S_{V}^{\prime \prime}(\eta)$ such that the functions $f \in S_{V}^{\prime \prime}(\eta)$ are non negative, such that

$$
\forall f \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M), \forall r \geq 2, \quad\left|S_{V}^{\prime \prime}(\eta) \cap \mathcal{B}(f, r \eta)\right| \leq \exp \left(63 D_{V}(\eta) r^{2}\right)
$$

and such that

$$
\text { for all non-negative function } f \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M), \quad d_{2}\left(f, S_{V}^{\prime \prime}(\eta)\right) \leq 4 d_{2}\left(f, S_{V}^{\prime}(\eta)\right)
$$

The lemma holds with $\bar{S}_{V}(\eta)=\left\{\sqrt{f}, f \in S_{V}^{\prime \prime}(\eta)\right\}$.

### 6.2. Proof of Lemma 4.1

The proof of this proposition requires the following elementary lemma.
Lemma 6.8. Let $f, f^{\prime} \in \mathbb{L}^{2}(\mathbb{T}, \mu)$ and $g, g^{\prime} \in \mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$ such that $\|f\|_{\mathrm{t}}=\left\|f^{\prime}\right\|_{\mathrm{t}}=1$ and $\|g\|_{\mathrm{x}}=\left\|g^{\prime}\right\|_{\mathrm{x}}=1$. Let $\kappa, \kappa^{\prime} \in \mathbb{R}$. Then,

$$
d_{2}^{2}\left(\kappa f g, \kappa^{\prime} f^{\prime} g^{\prime}\right)=\left(\kappa-\kappa^{\prime}\right)^{2}+\kappa \kappa^{\prime}\left(d_{\mathrm{t}}^{2}\left(f, f^{\prime}\right)+d_{\mathrm{x}}^{2}\left(g, g^{\prime}\right)-1 / 2 d_{\mathrm{t}}^{2}\left(f, f^{\prime}\right) d_{\mathrm{x}}^{2}\left(g, g^{\prime}\right)\right)
$$

Let $\eta>0$. In this proof, we say that a set $S(\eta)$ is a $\eta$-net of a set $V$ in a metric space $(E, d)$ if, for all $y \in V$, there exists $x \in S(\eta)$ such that $d(x, y) \leq \eta$.

Let us denote by $S_{1}$ (respectively $S_{2}$ ) the unit sphere of $V_{1}$ (respectively $V_{2}$ ). Let $S_{1}(\eta) \subset S_{1}$ (respectively $S_{2}(\eta) \subset S_{2}$ ) be a $\eta$-net of $S_{1}$ (respectively $S_{2}$ ) such that

$$
\begin{align*}
& \forall f \in V_{1}, \quad \forall r \geq 0, \quad\left|S_{1}(\eta) \cap \mathcal{B}_{t}(f, r \eta)\right| \leq(2 r+1)^{\operatorname{dim} V_{1}}  \tag{6.5}\\
& \forall g \in V_{2}, \quad \forall r \geq 0, \quad\left|S_{2}(\eta) \cap \mathcal{B}_{\mathbf{x}}(g, r \eta)\right| \leq(2 r+1)^{\operatorname{dim} V_{2}} \tag{6.6}
\end{align*}
$$

where $\mathcal{B}_{t}(f, r \eta)$ and $\mathcal{B}_{\mathbf{x}}(g, r \eta)$ are the closed balls centered at $f$ and $g$ with radius $r \eta$ of the metric spaces $\left(\mathbb{L}^{2}(\mathbb{T}, \mu), d_{\mathrm{t}}\right)$ and $\left(\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right), d_{\mathrm{x}}\right)$ respectively. We refer to Lemma 4 of $[7]$ for the existence of these nets. Let now

$$
S(\eta)=\bigcup_{k \in \mathbb{N}^{\star}}\left\{\frac{k \eta}{\sqrt{2}} f g,(f, g) \in S_{1}\left(\frac{1}{\sqrt{2} k}\right) \times S_{2}\left(\frac{1}{\sqrt{2} k}\right)\right\}
$$

First of all, $S(\eta)$ is a $\eta$-net of $V$. Indeed, let $\varphi \in V$. We can write $\varphi(t, x)=\kappa f(t) g(x)$ where $\kappa \geq 0, f \in S_{1}$ and $g \in S_{2}$. Let us define

$$
k=\inf \left\{i \in \mathbb{N}^{\star}, i \geq \sqrt{2} \kappa / \eta\right\}
$$

and let $\left(f^{\prime}, g^{\prime}\right) \in S_{1}(1 /(\sqrt{2} k)) \times S_{2}(1 /(\sqrt{2} k))$ such that

$$
d_{\mathrm{t}}\left(f, f^{\prime}\right) \leq \frac{1}{\sqrt{2} k} \quad \text { and } \quad d_{\mathrm{x}}\left(g, g^{\prime}\right) \leq \frac{1}{\sqrt{2} k}
$$

By using Lemma 6.8, the mapping $\varphi^{\prime}(t, x)=\frac{k \eta}{\sqrt{2}} f^{\prime}(t) g^{\prime}(x)$ is such that

$$
d_{2}\left(\varphi, \varphi^{\prime}\right) \leq \eta
$$

which proves that $S(\eta)$ is a $\eta$-net of $V$.
According to Definition 2.1, we now consider $\varphi \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ and $r \geq 2$ and aim at bounding from above the cardinality of the set $S(\eta) \cap \mathcal{B}(\varphi, r \eta)$ (where we recall that $\mathcal{B}(\varphi, r \eta)$ is the closed ball centered at $\varphi$ with radius $r \eta$ of the metric space $\left.\left(\mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M), d_{2}\right)\right)$.

For this purpose, we begin to assume that $\varphi$ belongs to $S(\eta)$, which implies that the function can be written as $\varphi(t, x)=\kappa f(t) g(x)$. We introduce

$$
\mathcal{K}=\left\{\frac{k \eta}{\sqrt{2}}, k \in \mathbb{N}^{\star},\left|\frac{k \eta}{\sqrt{2}}-\kappa\right| \leq r \eta\right\}
$$

and for all $\kappa^{\prime} \in \mathcal{K}$,

$$
\mathcal{C}\left(\kappa^{\prime}\right)=\left(S_{1}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{t}}\left(f, 6 r^{2} \frac{\eta}{2 \kappa^{\prime}}\right)\right) \times\left(S_{2}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{x}}\left(g, 6 r^{2} \frac{\eta}{2 \kappa^{\prime}}\right)\right) .
$$

Let

$$
T(\eta)=\left\{\kappa^{\prime} f^{\prime} g^{\prime}, \kappa^{\prime} \in \mathcal{K},\left(f^{\prime}, g^{\prime}\right) \in \mathcal{C}\left(\kappa^{\prime}\right)\right\}
$$

We shall prove that

$$
\begin{equation*}
S(\eta) \cap \mathcal{B}(\kappa f g, r \eta) \subset T(\eta) \tag{6.7}
\end{equation*}
$$

We then upper-bound the cardinality of $S(\eta) \cap \mathcal{B}(\kappa f g, r \eta)$ by bounding from above the cardinality of $T(\eta)$.
Let $\varphi^{\prime} \in S(\eta) \cap \mathcal{B}(\kappa f g, r \eta)$. There exist $\kappa^{\prime}, f^{\prime}$ and $g^{\prime}$ such that $\varphi^{\prime}=\kappa^{\prime} f^{\prime} g^{\prime}$ and we derive from Lemma 6.8 that

$$
\left(\kappa-\kappa^{\prime}\right)^{2} \leq d_{2}^{2}\left(\varphi, \varphi^{\prime}\right) \leq r^{2} \eta^{2}
$$

which implies that $\kappa^{\prime} \in \mathcal{K}$. We now distinguish several cases.

- Suppose that

$$
\left(\int_{\mathbb{T}} f(t) f^{\prime}(t) \mathrm{d} \mu(t)\right)\left(\int_{\mathbb{X}} g(x) g^{\prime}(x) \mathrm{d} \nu_{n}(x)\right)<0
$$

We then have $d_{2}^{2}\left(\varphi, \varphi^{\prime}\right) \geq \kappa^{2}+\kappa^{\prime 2}$. Since $\kappa \geq \eta / \sqrt{2}, \kappa^{\prime} \leq \kappa+r \eta$ and $r \geq 2$,

$$
\frac{\kappa^{\prime}}{\kappa} \leq 1+\sqrt{2} r \leq \frac{3}{2} r
$$

Thus, $d_{2}^{2}\left(\varphi, \varphi^{\prime}\right) \geq 4 \kappa^{\prime 2} /\left(9 r^{2}\right)$. Since $f, f^{\prime} \in S_{1}$ and $g, g^{\prime} \in S_{2}$,

$$
\left\|f-f^{\prime}\right\|_{\mathrm{t}}^{2} \leq 4 \quad \text { and } \quad\left\|g-g^{\prime}\right\|_{\mathrm{x}}^{2} \leq 4
$$

and thus

$$
\begin{aligned}
& \left\|f-f^{\prime}\right\|_{\mathrm{t}}^{2} \leq \frac{9 r^{2}}{\kappa^{\prime 2}} d_{2}^{2}\left(\varphi, \varphi^{\prime}\right) \leq \frac{9 r^{4}}{\kappa^{\prime 2}} \eta^{2} \\
& \left\|g-g^{\prime}\right\|_{\mathrm{x}}^{2} \leq \frac{9 r^{2}}{\kappa^{\prime 2}} d_{2}^{2}\left(\varphi, \varphi^{\prime}\right) \leq \frac{9 r^{4}}{\kappa^{\prime 2}} \eta^{2}
\end{aligned}
$$

We then have

$$
f^{\prime} \in S_{1}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{t}}\left(f, 6 r^{2} \frac{\eta}{2 \kappa^{\prime}}\right) \quad \text { and } \quad g^{\prime} \in S_{2}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{x}}\left(g, 6 r^{2} \frac{\eta}{2 \kappa^{\prime}}\right)
$$

that is $\left(f^{\prime}, g^{\prime}\right) \in \mathcal{C}\left(\kappa^{\prime}\right)$ and thus $\varphi^{\prime} \in T(\eta)$.

- If now,

$$
\int_{\mathbb{T}} f(t) f^{\prime}(t) \mathrm{d} \mu(t)>0 \quad \text { and } \quad \int_{\mathbb{X}} g(x) g^{\prime}(x) \mathrm{d} \nu_{n}(x)>0
$$

then $d_{\mathrm{t}}^{2}\left(f, f^{\prime}\right) \leq 2$ and $d_{\mathrm{x}}^{2}\left(g, g^{\prime}\right) \leq 2$. We then derive from Lemma 6.8 and from the elementary inequality

$$
\frac{1}{2}\left(y_{1}+y_{2}\right) \leq y_{1}+y_{2}-\frac{1}{2} y_{1} y_{2} \quad \text { for all } y_{1}, y_{2} \in[0,2]
$$

that

$$
\left(\kappa-\kappa^{\prime}\right)^{2}+\frac{\kappa \kappa^{\prime}}{2}\left(d_{\mathrm{t}}^{2}\left(f, f^{\prime}\right)+d_{\mathrm{x}}^{2}\left(g, g^{\prime}\right)\right) \leq d_{2}^{2}\left(\varphi, \varphi^{\prime}\right) \leq r^{2} \eta^{2}
$$

Hence,

$$
d_{\mathrm{t}}^{2}\left(f, f^{\prime}\right)+d_{\mathrm{x}}^{2}\left(g, g^{\prime}\right) \leq \frac{2 r^{2} \eta^{2}}{\kappa \kappa^{\prime}}
$$

By using the inequality $\kappa^{\prime} / \kappa \leq 3 / 2 r$ proved in the first point, we deduce

$$
f^{\prime} \in S_{1}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{t}}\left(f, 2 \sqrt{3} r^{3 / 2} \frac{\eta}{2 \kappa^{\prime}}\right) \quad \text { and } \quad g^{\prime} \in S_{2}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{x}}\left(g, 2 \sqrt{3} r^{3 / 2} \frac{\eta}{2 \kappa^{\prime}}\right)
$$

Since $2 \sqrt{3} r^{3 / 2} \leq 6 r^{2}$ (because $r \geq 2$ ), we have $\left(f^{\prime}, g^{\prime}\right) \in \mathcal{C}\left(\kappa^{\prime}\right)$ and thus $\varphi^{\prime} \in T(\eta)$.

- Finally, assume that

$$
\int_{\mathbb{T}} f(t) f^{\prime}(t) \mathrm{d} \mu(t)<0 \quad \text { and } \quad \int_{\mathbb{X}} g(x) g^{\prime}(x) \mathrm{d} \nu_{n}(x)<0
$$

Note that the function $\varphi^{\prime}$ can also be written as $\varphi^{\prime}=\kappa^{\prime}\left(-f^{\prime}\right)\left(-g^{\prime}\right)$. We then deduce from the second point that

$$
-f^{\prime} \in S_{1}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{t}}\left(f, 2 \sqrt{3} r^{3 / 2} \frac{\eta}{2 \kappa^{\prime}}\right) \text { and }-g^{\prime} \in S_{2}\left(\frac{\eta}{2 \kappa^{\prime}}\right) \cap \mathcal{B}_{\mathrm{x}}\left(g, 2 \sqrt{3} r^{3 / 2} \frac{\eta}{2 \kappa^{\prime}}\right)
$$

and thus $\left(-f^{\prime},-g^{\prime}\right) \in \mathcal{C}\left(\kappa^{\prime}\right)$. Hence, $\varphi^{\prime} \in T(\eta)$ as wished.
We thus have proved (6.7) and therefore

$$
|S(\eta) \cap \mathcal{B}(\kappa f g, r \eta)| \leq \sum_{\kappa^{\prime} \in \mathcal{K}}\left|\mathcal{C}\left(\kappa^{\prime}\right)\right|
$$

Now, note that $|\mathcal{K}| \leq 2 \sqrt{2} r+1$. By using (6.5) and (6.6), for all $\kappa^{\prime}$,

$$
\left|\mathcal{C}\left(\kappa^{\prime}\right)\right| \leq\left(12 r^{2}+1\right)^{\operatorname{dim} V_{1}+\operatorname{dim} V_{2}}
$$

Consequently, we have proved

$$
\forall \varphi \in S(\eta), \forall r \geq 2, \quad|S(\eta) \cap \mathcal{B}(\varphi, r \eta)| \leq(2 \sqrt{2} r+1)\left(12 r^{2}+1\right)^{\operatorname{dim} V_{1}+\operatorname{dim} V_{2}}
$$

Let us recall that we must to upper bound the cardinality of $S(\eta) \cap \mathcal{B}(\varphi, r \eta)$ for all $\varphi \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$. For this, if $\varphi \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$, may be $|S(\eta) \cap \mathcal{B}(\varphi, r \eta)|=0$. If not, there exists $\varphi^{\prime} \in S(\eta) \cap \mathcal{B}(\varphi, r \eta)$ and thus

$$
|S(\eta) \cap \mathcal{B}(\varphi, r \eta)| \leq\left|S(\eta) \cap \mathcal{B}\left(\varphi^{\prime}, 2 r \eta\right)\right|
$$

Therefore, for all $\varphi \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$,

$$
\forall r \geq 2, \quad|S(\eta) \cap \mathcal{B}(\varphi, r \eta)| \leq(4 \sqrt{2} r+1)\left(48 r^{2}+1\right)^{\operatorname{dim} V_{1}+\operatorname{dim} V_{2}}
$$

The conclusion follows from the elementary inequalities

$$
\forall r \geq 2, \quad 4 \sqrt{2} r+1 \leq e^{1.4 r^{2}} \text { and } 48 r^{2}+1 \leq e^{1.4 r^{2}}
$$

### 6.3. Proof of Proposition 4.2

For all pair $\left(V_{1}, V_{2}\right) \in \mathbb{V}_{1} \times \mathbb{V}_{2}$, we define the set $V$ by relation (4.2). Let then $\mathbb{V}$ be the collection of all $V$ when $\left(V_{1}, V_{2}\right)$ varies among $\mathbb{V}_{1} \times \mathbb{V}_{2}$. Let $\bar{\Delta}$ be the mapping on $\mathbb{V}$ defined by

$$
\bar{\Delta}(V)=\Delta_{1}\left(V_{1}\right)+\Delta_{2}\left(V_{2}\right)
$$

when $V$ corresponds to $\left(V_{1}, V_{2}\right)$. We apply Theorem 2.2 with $\mathbb{V}$ and $\bar{\Delta}$ to derive an estimator $\hat{s}$ such that

$$
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq \inf _{V \in \mathbb{V}}\left\{d_{2}^{2}(\sqrt{s}, V)+\frac{\operatorname{dim} V_{1}+\operatorname{dim} V_{2}+1+\Delta_{1}\left(V_{1}\right)+\Delta_{2}\left(V_{2}\right)}{n}\right\}
$$

Let $\kappa v_{1} v_{2} \in \mathscr{F}$, and let $\left(v_{1}^{\prime}, v_{2}^{\prime}\right) \in V_{1} \times V_{2}$ such that $\left\|v_{1}^{\prime}\right\|_{\mathrm{t}}=\left\|v_{2}^{\prime}\right\|_{\mathrm{x}}=1$. The preceding inequality implies

$$
\begin{aligned}
C^{\prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, \kappa v_{1} v_{2}\right)+\kappa^{2} d_{2}^{2}\left(v_{1} v_{2}, v_{1}^{\prime} v_{2}^{\prime}\right) \\
& +\frac{\operatorname{dim} V_{1} \vee 1+\operatorname{dim} V_{2} \vee 1+\Delta_{1}\left(V_{1}\right)+\Delta_{2}\left(V_{2}\right)}{n}
\end{aligned}
$$

where $C^{\prime}=C / 2$. By using Lemma 6.8 (Sect. 6.2),

$$
d_{2}^{2}\left(v_{1} v_{2}, v_{1}^{\prime} v_{2}^{\prime}\right) \leq d_{\mathrm{t}}^{2}\left(v_{1}, v_{1}^{\prime}\right)+d_{\mathrm{x}}^{2}\left(v_{2}, v_{2}^{\prime}\right)
$$

By taking the infimum over all $v_{1}^{\prime}$ and $v_{2}^{\prime}$,

$$
\begin{aligned}
C^{\prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, \kappa v_{1} v_{2}\right)+\kappa^{2} d_{\mathrm{t}}^{2}\left(v_{1}, S_{1}\right)+\frac{\operatorname{dim} V_{1} \vee 1+\Delta_{1}\left(V_{1}\right)}{n} \\
& +\kappa^{2} d_{\mathrm{x}}^{2}\left(v_{2}, S_{2}\right)+\frac{\operatorname{dim} V_{2} \vee 1+\Delta_{2}\left(V_{2}\right)}{n}
\end{aligned}
$$

where $S_{1}$ and $S_{2}$ are the unit spheres of $V_{1}$ and $V_{2}$, respectively.
Now, remark that

$$
d_{\mathrm{t}}\left(v_{1}, S_{1}\right) \leq 2 d_{\mathrm{t}}\left(v_{1}, V_{1}\right) \quad \text { and } \quad d_{\mathrm{x}}\left(v_{2}, S_{2}\right) \leq 2 d_{\mathrm{x}}\left(v_{2}, V_{2}\right)
$$

Indeed, if $w_{1}$ is the projection of $v_{1}$ on $V_{1}$, then

$$
\begin{aligned}
d_{\mathrm{t}}\left(v_{1}, S_{1}\right) & \leq\left\|v_{1}-\frac{w_{1}}{\left\|w_{1}\right\|_{\mathrm{t}}}\right\|_{\mathrm{t}} \\
& \leq\left\|v_{1}-w_{1}\right\|_{\mathrm{t}}+\left\|w_{1}-\frac{w_{1}}{\left\|w_{1}\right\|_{\mathrm{t}}}\right\|_{\mathrm{t}}
\end{aligned}
$$

Now,

$$
\left\|w_{1}-\frac{w_{1}}{\left\|w_{1}\right\|_{\mathrm{t}} \|_{\mathrm{t}}}\right\|=\left|1-\frac{1}{\left\|w_{1}\right\|_{\mathrm{t}}}\right|\left\|w_{1}\right\|_{\mathrm{t}}=\left|\left\|w_{1}\right\|_{\mathrm{t}}-1\right|
$$

Since $\left\|v_{1}\right\|_{\mathrm{t}}=1$,

$$
\left\|w_{1}-\frac{w_{1}}{\left\|w_{1}\right\|_{\mathrm{t}} \|_{\mathrm{t}}}=\left|\left\|w_{1}\right\|_{\mathrm{t}}-\left\|v_{1}\right\|_{\mathrm{t}}\right| \leq\right\| v_{1}-w_{1} \|_{\mathrm{t}}
$$

Since $\left\|v_{1}-w_{1}\right\|_{\mathrm{t}}=d_{\mathrm{t}}\left(v_{1}, V_{1}\right)$, we have $d_{\mathrm{t}}\left(v_{1}, S_{1}\right) \leq 2 d_{\mathrm{t}}\left(v_{1}, V_{1}\right)$. The proof of the inequality $d_{\mathrm{x}}\left(v_{2}, S_{2}\right) \leq$ $2 d_{\mathrm{x}}\left(v_{2}, V_{2}\right)$ is similar.

The conclusion follows.

### 6.4. Proof of Lemma 4.6

For all $b, b^{\prime} \in(-1 / 2,+\infty)$,

$$
\int_{0}^{1}\left(\sqrt{2 b+1} t^{b}-\sqrt{2 b^{\prime}+1} t^{b^{\prime}}\right)^{2} \mathrm{~d} t=\frac{4\left(b-b^{\prime}\right)^{2}}{\left(1+b+b^{\prime}\right)\left(\sqrt{2 b+1}+\sqrt{2 b^{\prime}+1}\right)^{2}}
$$

and thus

$$
\frac{\left(b-b^{\prime}\right)^{2}}{\left(1+2\left(b \vee b^{\prime}\right)\right)^{2}} \leq \int_{0}^{1}\left(\sqrt{2 b+1} t^{b}-\sqrt{2 b^{\prime}+1} t^{b^{\prime}}\right)^{2} \mathrm{~d} t \leq \frac{\left(b-b^{\prime}\right)^{2}}{\left(1+2\left(b \wedge b^{\prime}\right)\right)^{2}}
$$

which ends the proof.

### 6.5. Proof of Lemma 4.7

For $b>0$, we define

$$
g_{b}(t)=\frac{f_{b}(t)}{\left\|f_{b}\right\|_{\mathrm{t}}}=\frac{2^{k / 2} b^{1 / 2+k / 2}}{\sqrt{k / 2 k!}} t^{k / 2} e^{-b t}
$$

For all $b, b^{\prime}>0$,

$$
\begin{aligned}
\frac{1}{2} \int_{0}^{\infty}\left(g_{b}(t)-g_{b^{\prime}}(t)\right)^{2} \mathrm{~d} t & =1-\frac{\left(2 \sqrt{b b^{\prime}}\right)^{k+1}}{\left(b+b^{\prime}\right)^{k+1}} \\
& =\frac{\sum_{j=0}^{k}\left(b+b^{\prime}\right)^{k-j}\left(2 \sqrt{b b^{\prime}}\right)^{j}}{\left(b+b^{\prime}\right)^{k+1}\left(\sqrt{b}+\sqrt{b^{\prime}}\right)^{2}}\left(b-b^{\prime}\right)^{2} \\
& =\frac{1}{\left(b+b^{\prime}\right)\left(\sqrt{b}+\sqrt{b^{\prime}}\right)^{2}} \sum_{j=0}^{k}\left(\frac{2 \sqrt{b b^{\prime}}}{b+b^{\prime}}\right)^{j}\left(b-b^{\prime}\right)^{2} .
\end{aligned}
$$

Consequently,

$$
\frac{1}{8\left(b \vee b^{\prime}\right)^{2}}\left(b-b^{\prime}\right)^{2} \leq \frac{1}{2} \int_{0}^{\infty}\left(g_{b}(t)-g_{b^{\prime}}(t)\right)^{2} \mathrm{~d} t \leq \frac{k+1}{8\left(b \wedge b^{\prime}\right)^{2}}\left(b-b^{\prime}\right)^{2}
$$

which concludes the proof.

### 6.6. Proof of Proposition 4.9

We generalize Lemma 4.1 for some new spaces. The proof of the following lemma is analogous to the one of Lemma 4.1 and will not be detailed.

Lemma 6.9. Let $V_{1}$ and $V_{2}$ be subsets of the unit spheres of $\mathbb{L}^{2}(\mathbb{T}, \mu)$ and $\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$ respectively. For each $i \in\{1,2\}$, we assume that there exist positive numbers $\underline{\rho_{i}}, \bar{\rho}_{i}$, a subset $W_{\mathrm{i}}$ of a finite dimensional normed linear space $\left(\bar{W}_{\mathrm{i}},|\cdot|_{\mathrm{i}}\right)$ and a surjective map $\Phi_{i}$ from $W_{\mathrm{i}}$ onto $\bar{V}_{i}$ such that:

$$
\begin{align*}
& \forall(x, y) \in W_{1}, \quad \underline{\rho_{1}}|x-y|_{1} \leq d_{\mathrm{t}}\left(\Phi_{1}(x), \Phi_{1}(y)\right) \leq \bar{\rho}_{1}|x-y|_{1}  \tag{6.8}\\
& \forall(x, y) \in W_{2}, \quad \underline{\rho_{2}}|x-y|_{2} \leq d_{\mathbf{x}}\left(\Phi_{2}(x), \Phi_{2}(y)\right) \leq \bar{\rho}_{2}|x-y|_{2} . \tag{6.9}
\end{align*}
$$

The set

$$
V=\left\{\kappa v_{1} v_{2},\left(v_{1}, v_{2}\right) \in V_{1} \times V_{2}, \kappa \in[0,+\infty)\right\}
$$

has a finite metric dimension bounded by

$$
D_{V}=C\left[1+\log \left(1+\frac{\bar{\rho}_{1}}{\underline{\rho_{1}}}\right) \operatorname{dim} \bar{W}_{1}+\log \left(1+\frac{\bar{\rho}_{2}}{\underline{\rho_{2}}}\right) \operatorname{dim} \bar{W}_{2}\right]
$$

where $C$ is a universal constant.
Lemma 6.10. Let for all $r, R \in\left(b_{0},+\infty\right)$, such that $R>r, V_{\mathrm{t}}(r, R)$ be the set defined by

$$
V_{\mathrm{t}}(r, R)=\left\{\frac{u_{b}}{\left\|u_{b}\right\|_{\mathrm{t}}}, b \in[r, R]\right\}
$$

Condition (6.8) holds with $\operatorname{dim} \bar{W}_{1}=1, \underline{\rho_{1}}=\underline{\rho}(R)$ and $\overline{\rho_{1}}=\bar{\rho}(r)$.
Lemma 6.11. For all positive number $\rho$ and $W \in \mathbb{W}$, let

$$
V_{2}(W, \rho)=\left\{\frac{v_{\boldsymbol{\theta}}}{\left\|v_{\boldsymbol{\theta}}\right\|_{\mathrm{x}}}, \boldsymbol{\theta} \in W,\|\boldsymbol{\theta}\| \leq \rho\right\}
$$

There exists a finite dimensional normed linear space $\left(\bar{W}_{2},|\cdot|_{2}\right)$ and a map $\Phi_{2}$ from $\bar{W}_{2}$ onto $V_{2}(W, \rho)$ such that condition (6.9) holds with $\operatorname{dim} \bar{W}_{2} \leq \operatorname{dim} W, \underline{\rho_{2}}=e^{-6 \rho}$ and $\bar{\rho}_{2}=e^{6 \rho}$.

Proof of Lemma 6.11. For any integers $i, j \in \mathbb{N}^{\star}$, let us denote by $\varphi_{i, j}$ the linear form on $\mathbb{R}^{k_{2}}$ defined by $\varphi_{i, j}(\boldsymbol{\theta})=\left\langle x_{i}-x_{j}, \boldsymbol{\theta}\right\rangle$ where $\langle\cdot, \cdot\rangle$ is the standard scalar product on $\mathbb{R}^{k_{2}}$. Let $W_{1}=\cap_{i \neq j} \operatorname{Ker} \varphi_{i, j}$ and let $W_{2}$ such that $W=W_{1} \oplus W_{2}$ and such that $\langle u, v\rangle=0$ for all $(u, v) \in W_{1} \times W_{2}$. Since the functions of $\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$ are defined $\nu_{n}$-almost everywhere, the set $V_{2}(W, \rho)$ can be written as

$$
V_{2}(W, \rho)=\Phi_{2}\left(\left\{\boldsymbol{\theta} \in W_{2},\|\boldsymbol{\theta}\| \leq \rho\right\}\right) \quad \text { where } \quad \Phi_{2}(\boldsymbol{\theta})=\frac{v_{\boldsymbol{\theta}}}{\left\|v_{\boldsymbol{\theta}}\right\|_{\mathrm{x}}}
$$

Indeed, let $\boldsymbol{\theta} \in W$ written as $\boldsymbol{\theta}=\boldsymbol{\theta}_{1}+\boldsymbol{\theta}_{2}$ where $\boldsymbol{\theta}_{1} \in W_{1}$ and $\boldsymbol{\theta}_{2} \in W_{2}$. Then, for all $j \in\{1, \ldots, n\}$,

$$
\begin{aligned}
\frac{v_{\boldsymbol{\theta}}\left(x_{j}\right)}{\left\|v_{\boldsymbol{\theta}}\right\|_{\mathrm{x}}} & =\frac{\exp \left(\left\langle x_{j}, \boldsymbol{\theta}\right\rangle\right)}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} \exp \left(2\left\langle x_{i}, \boldsymbol{\theta}\right\rangle\right)}} \\
& =\frac{\exp \left(\left\langle x_{j}, \boldsymbol{\theta}_{1}\right\rangle+\left\langle x_{j}, \boldsymbol{\theta}_{2}\right\rangle\right)}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} \exp \left(2\left\langle x_{i}, \boldsymbol{\theta}_{1}\right\rangle+2\left\langle x_{j}, \boldsymbol{\theta}_{2}\right\rangle\right)}} \\
& =\frac{v_{\boldsymbol{\theta}_{1}}\left(x_{j}\right)}{\left\|v_{\boldsymbol{\theta}_{1}}\right\|_{\mathrm{x}}}
\end{aligned}
$$

and thus $\Phi_{2}(\boldsymbol{\theta})=\Phi_{2}\left(\boldsymbol{\theta}_{1}\right), \nu_{n}$-almost everywhere.
For all $x \in \mathbb{X}$, let $\Psi_{x}$ be the function defined from $\mathbb{X}$ into $\mathbb{R}$ by $\Psi_{x}(\boldsymbol{\theta})=\Phi_{2}(\boldsymbol{\theta})(x)=v_{\boldsymbol{\theta}}(x) /\left\|v_{\boldsymbol{\theta}}\right\|_{\mathrm{x}}$. We derive from some calculus that the differential of $\Psi_{x}$ at the point $\boldsymbol{\theta} \in W_{2}$, denoted by $d \Psi_{x}(\boldsymbol{\theta})$, is

$$
\forall h \in \mathbb{R}^{k_{2}}, \quad d \Psi_{x}(\boldsymbol{\theta}) \cdot h=\frac{\frac{1}{n} \sum_{i=1}^{n} \exp \left(2\left\langle\boldsymbol{\theta}, x_{i}\right\rangle+\langle\boldsymbol{\theta}, x\rangle\right)\left(\left\langle x-x_{i}, h\right\rangle\right)}{\left(\frac{1}{n} \sum_{i=1}^{n} \exp \left(2\left\langle\boldsymbol{\theta}, x_{i}\right\rangle\right)\right)^{3 / 2}}
$$

In particular, we have

$$
\forall h \in \mathbb{R}^{k_{2}}, \quad \frac{e^{-6 \rho}}{n} \sum_{i=1}^{n}\left|\left\langle x-x_{i}, h\right\rangle\right| \leq\left|d \Psi_{x}(\boldsymbol{\theta}) \cdot h\right| \leq \frac{e^{6 \rho}}{n} \sum_{i=1}^{n}\left|\left\langle x-x_{i}, h\right\rangle\right|
$$

If we endow $W_{2}$ with the norm $|\cdot|_{2}$ defined by

$$
\forall \boldsymbol{\theta} \in W_{2}, \quad|\boldsymbol{\theta}|_{2}=\sqrt{\frac{1}{n} \sum_{i=1}^{n}\left(\frac{1}{n} \sum_{j=1}^{n}\left|\left\langle x_{i}-x_{j}, \boldsymbol{\theta}\right\rangle\right|\right)^{2}}
$$

the mean value theorem leads to

$$
\forall\left(\boldsymbol{\theta}_{1}, \boldsymbol{\theta}_{2}\right) \in W_{2}, \quad e^{-6 \rho}\left|\boldsymbol{\theta}_{1}-\boldsymbol{\theta}_{2}\right|_{2} \leq d_{\mathbf{x}}\left(\Phi_{2}\left(\boldsymbol{\theta}_{1}\right), \Phi_{2}\left(\boldsymbol{\theta}_{2}\right)\right) \leq e^{6 \rho}\left|\boldsymbol{\theta}_{1}-\boldsymbol{\theta}_{2}\right|_{2}
$$

which concludes the proof.
We now prove Proposition 4.9. We derive from Lemma 6.9 that for all $\varrho \geq 1$, all $r, R>b_{0}$ and all $W \in \mathbb{W}$, the set

$$
V(r, R, W, \varrho)=\left\{a u_{b} v_{\boldsymbol{\theta}}, a \in[0,+\infty), b \in[r, R], \boldsymbol{\theta}^{\prime} \in W,\left\|\boldsymbol{\theta}^{\prime}\right\| \leq \varrho\right\}
$$

has a metric dimension bounded by

$$
C D_{V(r, R, W, \varrho)}=1+\varrho \operatorname{dim} W+\log \left(1+\frac{\bar{\rho}(r)}{\underline{\rho}(R)}\right)
$$

for some universal positive constant $C$.
Let us define the collection $\mathbb{V}$ by

$$
\mathbb{V}=\left\{V\left(b_{0}+1 / r, b_{0}+R, W, \varrho\right), W \in \mathbb{W}, r, R, \varrho \in \mathbb{N}^{\star}\right\}
$$

and the map $\bar{\Delta}$ on $\mathbb{V}$ by

$$
\bar{\Delta}(V(r, R, W, \varrho))=\Delta(W)+\log \left(2 R^{2}\right)+\log \left(2 r^{2}\right)+\log \left(2 \varrho^{2}\right)
$$

We apply Theorem 2.2 with $(\mathbb{V}, \bar{\Delta})$ to build an estimator $\hat{s}$. For all $W \in \mathbb{W}, \varrho, r, R \in \mathbb{N}^{\star}, \boldsymbol{\theta}^{\prime} \in W$ such that $\left\|\boldsymbol{\theta}^{\prime}\right\| \leq \varrho, a \in[0,+\infty), b \in\left[b_{0}+1 / r, b_{0}+R\right]$, this estimator satisfies

$$
\begin{aligned}
C^{\prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, a u_{b} v_{\boldsymbol{\theta}^{\prime}}\right) \\
& +\frac{1+\varrho \operatorname{dim} W+\log \left(1+\frac{\bar{\rho}\left(b_{0}+1 / r\right)}{\underline{\rho}\left(b_{0}+R\right)}\right)+\Delta(W)+\log r+\log R+\log \varrho}{n}
\end{aligned}
$$

where $C^{\prime}$ is another universal positive constant.
We may roughly upper-bound the right-hand side of this inequality to get

$$
\begin{aligned}
C^{\prime \prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, a u_{b} v_{\boldsymbol{\theta}^{\prime}}\right) \\
& +\frac{1+\varrho \operatorname{dim} W+\log \left(1 \vee \bar{\rho}\left(b_{0}+1 / r\right)\right)+\left|\log \left(1 \wedge \underline{\rho}\left(b_{0}+R\right)\right)\right|+\Delta(W)}{n} \\
& +\frac{\log r+\log R+\log \varrho}{n}
\end{aligned}
$$

for some universal positive constant $C^{\prime \prime}$.
In particular, for all $W \in \mathbb{W}, \boldsymbol{\theta}^{\prime} \in W, a \in[0,+\infty)$ and $b \in I$, we may use this inequality with $R=\inf \{i \in$ $\left.\mathbb{N}^{\star}, i \geq b-b_{0}\right\}, r=\inf \left\{i \in \mathbb{N}^{\star}, i \geq 1 /\left(b-b_{0}\right)\right\}, \varrho=\inf \left\{i \in \mathbb{N}^{\star}, i \geq\left\|\boldsymbol{\theta}^{\prime}\right\|\right\}$ to derive

$$
\begin{aligned}
C^{\prime \prime \prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, a u_{b} v_{\boldsymbol{\theta}^{\prime}}\right)+\frac{\left(1 \vee\left\|\boldsymbol{\theta}^{\prime}\right\|\right)(1 \vee \operatorname{dim} W)+\Delta(W)}{n} \\
& +\frac{1}{n}\left\{\log \left[1 \vee \bar{\rho}\left(b_{0}+\frac{b-b_{0}}{b-b_{0}+1}\right)\right]+|\log (1 \wedge \underline{\rho}(1+b))|+\left|\log \left(b-b_{0}\right)\right|\right\}
\end{aligned}
$$

where $C^{\prime \prime \prime}$ is a universal positive constant.

Now, by using the triangular inequality, we have for all $\boldsymbol{\theta} \in \mathbb{R}^{k_{2}}$,

$$
\begin{aligned}
d_{2}^{2}\left(\sqrt{s}, a u_{b} v_{\boldsymbol{\theta}^{\prime}}\right) & \leq 2\left(d_{2}^{2}\left(\sqrt{s}, a u_{b} v_{\boldsymbol{\theta}}\right)+d_{2}^{2}\left(a u_{b} v_{\boldsymbol{\theta}}, a u_{b} v_{\boldsymbol{\theta}^{\prime}}\right)\right. \\
& \leq 2\left(d_{2}^{2}\left(\sqrt{s}, a u_{b} v_{\boldsymbol{\theta}}\right)+a^{2}\left\|u_{b}\right\|_{\mathrm{t}}^{2} d_{\mathrm{x}}^{2}\left(v_{\boldsymbol{\theta}}, v_{\boldsymbol{\theta}^{\prime}}\right)\right)
\end{aligned}
$$

Some calculus shows that $d_{\mathrm{x}}\left(v_{\boldsymbol{\theta}}, v_{\boldsymbol{\theta}^{\prime}}\right) \leq e^{\|\boldsymbol{\theta}\| \vee\left\|\boldsymbol{\theta}^{\prime}\right\|}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^{\prime}\right\|$.
Consequently, for all $a \in[0,+\infty), b \in I, \boldsymbol{\theta} \in \mathbb{R}^{k}, W \in \mathbb{W}$, we obtain (by taking $\boldsymbol{\theta}^{\prime}$ the projection of $\boldsymbol{\theta}$ on $W$ ),

$$
\begin{aligned}
C^{\prime \prime \prime \prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}\left(\sqrt{s}, a u_{b} v_{\boldsymbol{\theta}}\right)+a^{2}\left\|u_{b}\right\|_{\mathrm{t}}^{2} e^{2\|\boldsymbol{\theta}\|} d^{2}(\boldsymbol{\theta}, W)+\frac{(1 \vee \operatorname{dim} W)(1 \vee\|\boldsymbol{\theta}\|)+\Delta(W)}{n} \\
& +\frac{1}{n}\left\{\log \left[1 \vee \bar{\rho}\left(b_{0}+\frac{b-b_{0}}{b-b_{0}+1}\right)\right]+|\log (1 \wedge \underline{\rho}(1+b))|+\left|\log \left(b-b_{0}\right)\right|\right\}
\end{aligned}
$$

where $C^{\prime \prime \prime \prime}$ is a universal positive constant. We conclude by taking the infimum over all $W \in \mathbb{W}$.

### 6.7. Proofs of Lemmas 5.2 and 5.3

Proof of Lemma 5.2. We derive from some calculus that for all $\theta_{2}, \theta_{2}^{\prime} \in\left[-1 / 2+1 / r_{2},+\infty\right)$,

$$
\int_{0}^{1}\left(t^{\theta_{2}}-t^{\theta_{2}^{\prime}}\right)^{2} \mathrm{~d} t=\frac{2\left(\theta_{2}-\theta_{2}^{\prime}\right)^{2}}{\left(1+2 \theta_{2}\right)\left(1+\theta_{2}+\theta_{2}^{\prime}\right)\left(1+2 \theta_{2}^{\prime}\right)} \leq 2 r_{2}^{3}\left(\theta_{2}-\theta_{2}^{\prime}\right)^{2}
$$

Hence, for all $\left(\theta_{1}, \theta_{2}\right),\left(\theta_{1}^{\prime}, \theta_{2}^{\prime}\right) \in\left[-r_{1}, r_{1}\right] \times\left[-1 / 2+1 / r_{2},+\infty\right)$,

$$
\begin{aligned}
\sqrt{\int_{0}^{1}\left(\theta_{1} t^{\theta_{2}}-\theta_{1}^{\prime} t^{\theta_{2}^{\prime}}\right)^{2} \mathrm{~d} t} & \leq \sqrt{\int_{0}^{1}\left(\theta_{1}-\theta_{1}^{\prime}\right)^{2} t^{2 \theta_{2}^{\prime}} \mathrm{d} t}+\sqrt{\int_{0}^{1} \theta_{1}^{2}\left(t^{\theta_{2}}-t^{\left.\theta_{2}^{\prime}\right)^{2} \mathrm{~d} t}\right.} \\
& \leq \frac{\left|\theta_{1}-\theta_{1}^{\prime}\right|}{\sqrt{2 \theta_{2}^{\prime}+1}}+\frac{\sqrt{2}\left|\theta_{1}\right|\left|\theta_{2}-\theta_{2}^{\prime}\right|}{\sqrt{\left(1+2 \theta_{2}\right)\left(1+\theta_{2}+\theta_{2}^{\prime}\right)\left(1+2 \theta_{2}^{\prime}\right)}} \\
& \leq r_{2}^{1 / 2}\left|\theta_{1}-\theta_{1}^{\prime}\right|+\sqrt{2} r_{1} r_{2}^{3 / 2}\left|\theta_{2}-\theta_{2}^{\prime}\right|
\end{aligned}
$$

This ends the proof.
Proof of Lemma 5.3. We derive from some calculus that for all $\theta_{2}, \theta_{2}^{\prime} \geq 1 / r_{2}$,

$$
\int_{0}^{\infty}\left(t^{k / 2} e^{-\theta_{2} t}-t^{k / 2} e^{-\theta_{2}^{\prime} t}\right)^{2} \mathrm{~d} t=\frac{k!}{2^{k+1}}\left(\frac{1}{\theta_{2}^{k}}+\frac{1}{\theta_{2}^{\prime k}}-\frac{2^{2+k}}{\left(\theta_{2}+\theta_{2}^{\prime}\right)^{k}}\right)
$$

If $k=0$,

$$
\int_{0}^{\infty}\left(t^{k / 2} e^{-\theta_{2} t}-t^{k / 2} e^{-\theta_{2}^{\prime} t}\right)^{2} \mathrm{~d} t=\frac{\left(\theta_{2}^{\prime}-\theta_{2}\right)^{2}}{2 \theta_{2}^{\prime 2} \theta_{2}+2 \theta_{2}^{\prime} \theta_{2}^{2}} \leq \frac{r_{2}^{3}}{4}\left(\theta_{2}^{\prime}-\theta_{2}\right)^{2}
$$

while if $k=1$,

$$
\begin{aligned}
\int_{0}^{\infty}\left(t^{k / 2} e^{-\theta_{2} t}-t^{k / 2} e^{-\theta_{2}^{\prime} t}\right)^{2} \mathrm{~d} t & =\frac{\theta_{2}^{\prime 2}+4 \theta_{2}^{\prime} \theta_{2}+\theta_{2}^{2}}{4 \theta_{2}^{\prime 2} \theta_{2}^{2}\left(\theta_{2}^{\prime}+\theta_{2}\right)^{2}}\left(\theta_{2}^{\prime}-\theta_{2}\right)^{2} \\
& \leq \frac{3 r_{2}^{4}}{8}\left(\theta_{2}^{\prime}-\theta_{2}\right)^{2}
\end{aligned}
$$

Hence, for all $\left(\theta_{1}, \theta_{2}\right),\left(\theta_{1}^{\prime}, \theta_{2}^{\prime}\right) \in\left[-r_{1}, r_{1}\right] \times\left[1 / r_{2},+\infty\right)$,

$$
\begin{aligned}
\sqrt{\int_{0}^{\infty}\left(\theta_{1} t^{k / 2} e^{-\theta_{2} t}-\theta_{1}^{\prime} t^{k / 2} e^{-\theta_{2}^{\prime} t}\right)^{2} \mathrm{~d} t} & \leq \sqrt{\int_{0}^{\infty}\left(\theta_{1}-\theta_{1}^{\prime}\right)^{2} t^{k} e^{-2 \theta_{2}^{\prime} t} \mathrm{~d} t} \\
& +\left|\theta_{1}\right| \sqrt{\int_{0}^{\infty}\left(t^{k / 2} e^{-\theta_{2} t}-t^{k / 2} e^{-\theta_{2}^{\prime} t}\right)^{2} \mathrm{~d} t}
\end{aligned}
$$

Now,

$$
\sqrt{\int_{0}^{\infty}\left(\theta_{1}-\theta_{1}^{\prime}\right)^{2} t^{k} e^{-2 \theta_{2}^{\prime} t} \mathrm{~d} t}=\frac{\sqrt{k!}\left|\theta_{1}-\theta_{1}^{\prime}\right|}{2^{(k+1) / 2} \theta_{2}^{\prime(k+1) / 2}}
$$

which ends the proof.

### 6.8. Proof of Theorem 5.5

We start with the following proposition.
Proposition 6.12. Suppose that Assumption 5.1 holds. Let for all $j \in\{1, \ldots, k\}, W_{j}$ be a linear subspace of $\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$ with finite dimension and $Z_{j}$ be a bounded subset of $W_{j}$. Let then $\rho \in[0,+\infty)^{k}$ such that for all $j \in\{1, \ldots, k\}, Z_{j} \subset \mathcal{B}_{\mathrm{x}}\left(0, \rho_{j}\right)=\left\{g \in \mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right),\|g\|_{\mathrm{x}} \leq \rho_{j}\right\}$. Let $m_{1}, \ldots, m_{k} \in \mathbb{R} \cup\{-\infty\}$ and $M_{1}, \ldots, M_{k} \in$ $\mathbb{R} \cup\{\infty\}$ be such that $\Theta=\mathbb{R}^{k} \cap \prod_{i=1}^{k}\left[m_{i}, M_{i}\right]$ and let $\pi$ be the map defined on $\mathbb{R}^{k}$ by

$$
\pi(\boldsymbol{x})=\left(\left(x_{1} \vee m_{1}\right) \wedge M_{1}, \ldots,\left(x_{k} \vee m_{k}\right) \wedge M_{k}\right) \quad \text { for all } \boldsymbol{x}=\left(x_{1}, \ldots, x_{k}\right) \in \mathbb{R}^{k}
$$

Let for all $\boldsymbol{u} \in \prod_{j=1}^{k} Z_{j}, g_{\boldsymbol{u}}$ be the function defined by

$$
g_{\boldsymbol{u}(x)}(t)=f_{\pi(\boldsymbol{u}(x))}(t) \quad \text { for all }(t, x) \in \mathbb{T} \times \mathbb{X}
$$

Then, the set $V$ defined by

$$
V=\left\{g_{\boldsymbol{u}}, \boldsymbol{u} \in \prod_{j=1}^{k} Z_{j}\right\}
$$

has a metric dimension bounded by

$$
D_{V}(\eta)=\frac{1}{2} \vee \frac{1}{4} \sum_{j=1}^{k} \log \left(1+2\left(\frac{k R_{j}}{\eta}\right)^{1 / \alpha_{j}} \rho_{j}\right) \operatorname{dim}\left(W_{j}\right)
$$

Proof of Proposition 6.12. As in the proof of Lemma 4.1, we say that a set $S(\eta)$ is a $\eta$-net of a set $V$ in a metric space $(E, d)$ if, for all $y \in V$, there exists $x \in S(\eta)$ such that $d(x, y) \leq \eta$.

Let $\eta>0$ and for $j \in\{1, \ldots, k\}$,

$$
\eta_{j}=\left(\frac{\eta}{k R_{j}}\right)^{1 / \alpha_{j}}
$$

Let $Z_{j}^{\prime}\left(\eta_{j}\right)$ be a maximal subset of $Z_{j}$ such that $d_{\mathrm{x}}(x, y)>\eta_{j}$ for all $x \neq y \in Z_{j}^{\prime}(\eta)$. This is a $\eta_{j}$-net of $Z_{j}$ such that

$$
\left|Z_{j}^{\prime}\left(\eta_{j}\right)\right| \leq\left|Z_{j}^{\prime}\left(\eta_{j}\right) \cap \mathcal{B}_{\mathbf{x}}\left(0, \rho_{j}\right)\right|
$$

and by using Lemma 4 of [7],

$$
\begin{equation*}
\left|Z_{j}^{\prime}\left(\eta_{j}\right)\right| \leq\left(\frac{2 \rho_{j}}{\eta_{j}}+1\right)^{\operatorname{dim} W_{j}} \tag{6.10}
\end{equation*}
$$

To prove the proposition, we begin to show that the set

$$
S(\eta)=\left\{g_{\boldsymbol{u}}, \boldsymbol{u} \in \prod_{j=1}^{k} Z_{j}^{\prime}\left(\eta_{j}\right)\right\}
$$

is a $\eta$-net of $V$.
Let $f \in V$ be the function of the form $f(t, x)=g_{\boldsymbol{u}(x)}(t)=f_{\pi(\boldsymbol{u}(x))}(t)$ and for all $j \in\{1, \ldots, k\}$, let $v_{j} \in Z_{j}^{\prime}\left(\eta_{j}\right)$ such that $d_{\mathrm{x}}\left(u_{j}, v_{j}\right) \leq \eta_{j}$. We define $\boldsymbol{v}=\left(v_{1}, \ldots, v_{k}\right)$ and $g \in S(\eta)$ by $g(t, x)=g_{\boldsymbol{v}(x)}(t)=f_{\pi(\boldsymbol{v}(x))}(t)$. Then,

$$
\begin{aligned}
\|f-g\|_{2}^{2} & =\frac{1}{n} \sum_{i=1}^{n}\left\|f_{\pi\left(\boldsymbol{u}\left(x_{i}\right)\right)}(\cdot)-f_{\pi\left(\boldsymbol{v}\left(x_{i}\right)\right)}(\cdot)\right\|_{\mathrm{t}}^{2} \\
& \leq \frac{1}{n} \sum_{i=1}^{n}\left(\sum_{j=1}^{k} R_{j}\left|u_{j}\left(x_{i}\right)-v_{j}\left(x_{i}\right)\right|^{\alpha_{j}}\right)^{2}
\end{aligned}
$$

By using the Cauchy-Schwarz's inequality,

$$
\begin{aligned}
\|f-g\|_{2}^{2} & \leq \frac{1}{n} \sum_{i=1}^{n} k\left(\sum_{j=1}^{k} R_{j}^{2}\left|u_{j}\left(x_{i}\right)-v_{j}\left(x_{i}\right)\right|^{2 \alpha_{j}}\right) \\
& \leq k \sum_{j=1}^{k} R_{j}^{2}\left(\frac{1}{n} \sum_{i=1}^{n}\left|u_{j}\left(x_{i}\right)-v_{j}\left(x_{i}\right)\right|^{2 \alpha_{j}}\right)
\end{aligned}
$$

By using the concavity of the map $x \mapsto x^{\alpha_{j}}$,

$$
\begin{aligned}
\|f-g\|_{2}^{2} & \leq k \sum_{j=1}^{k} R_{j}^{2}\left(\frac{1}{n} \sum_{i=1}^{n}\left|u_{j}\left(x_{i}\right)-v_{j}\left(x_{i}\right)\right|^{2}\right)^{\alpha_{j}} \\
& \leq k \sum_{j=1}^{k} R_{j}^{2} d_{\mathrm{x}}^{2 \alpha_{j}}\left(u_{j}, v_{j}\right) \\
& \leq \eta^{2}
\end{aligned}
$$

as wished.
We now consider $r \geq 2, \varphi \in \mathbb{L}^{2}(\mathbb{T} \times \mathbb{X}, M)$ and aim at bounding from above the cardinality of $S(\eta) \cap \mathcal{B}(\varphi, r \eta)$. We have,

$$
|S(\eta) \cap \mathcal{B}(\varphi, r \eta)| \leq \prod_{j=1}^{k}\left|Z_{j}^{\prime}\left(\eta_{j}\right)\right|
$$

By using (6.10),

$$
\begin{aligned}
|S(\eta) \cap \mathcal{B}(\varphi, r \eta)| & \leq \prod_{j=1}^{k}\left(2\left(\frac{k R_{j}}{\eta}\right)^{1 / \alpha_{j}} \rho_{j}+1\right)^{\operatorname{dim}\left(W_{j}\right)} \\
& \leq \exp \left(\frac{1}{4} \sum_{j=1}^{k} \operatorname{dim}\left(W_{j}\right) \log \left(2\left(\frac{k R_{j}}{\eta}\right)^{1 / \alpha_{j}} \rho_{j}+1\right) r^{2}\right)
\end{aligned}
$$

This ends the proof.

Lemma 6.13. Let $V$ be a set with metric dimension bounded by $D_{V}$. Assume that there exist $k \in \mathbb{N}^{\star}, \boldsymbol{a}, \boldsymbol{b} \in$ $[0,+\infty)^{k}$ such that $\max _{1 \leq j \leq k} a_{j} \geq 1, \min _{1 \leq j \leq k} b_{j} \geq 1$ and such that

$$
D_{V}(\eta) \leq \frac{1}{2} \vee \sum_{j=1}^{k} a_{j} \log \left(1+\frac{b_{j}}{\eta}\right) \quad \text { for all } \eta>0
$$

Then,

$$
\eta_{V}=\inf \left\{\eta>0, \frac{D_{V}(\eta)}{\eta^{2}} \leq n\right\}
$$

can be upper bounded by

$$
C \eta_{V}^{2} \leq \frac{\sum_{j=1}^{k} a_{j} \log \left(1+b_{j}\right)}{n}+\frac{\sum_{j=1}^{k} a_{j}}{n} \log \left(1+\frac{n}{\sum_{j=1}^{k} a_{j}}\right)
$$

where $C$ is a universal positive constant.
Proof of Lemma 6.13. For all $\eta>0$,

$$
\frac{1}{2} \vee \sum_{j=1}^{k} a_{j} \log \left(1+\frac{b_{j}}{\eta}\right) \leq \begin{cases}2 \sum_{j=1}^{k} a_{j} \log \left(2 b_{j}\right)+2\left(\sum_{j=1}^{k} a_{j}\right) \log \left(\frac{1}{\eta}\right) & \text { if } \eta<1 \\ 2 \sum_{j=1}^{k} a_{j} \log \left(2 b_{j}\right) & \text { otherwise }\end{cases}
$$

The larger $D_{V}$, the larger $\eta_{V}$. Consequently, without lost of generality we can assume that

$$
D_{V}(\eta)= \begin{cases}2 \sum_{j=1}^{k} a_{j} \log \left(2 b_{j}\right)+2\left(\sum_{j=1}^{k} a_{j}\right) \log \left(\frac{1}{\eta}\right) & \text { if } \eta<1 \\ 2 \sum_{j=1}^{k} a_{j} \log \left(2 b_{j}\right) & \text { otherwise }\end{cases}
$$

Remark now that for all $\alpha, \beta, y>0$, the equation

$$
\alpha+\beta \log x=\frac{y}{2 x^{2}}
$$

has only one positive solution $x$ given by

$$
x^{2}=\frac{y}{\beta L\left(\frac{e^{\frac{2 \alpha}{\beta}}}{\beta} y\right)},
$$

where $L$ is the Lambert's function, defined as being the inverse function of $t \mapsto t e^{t}$. Consequently, by setting

$$
\alpha=\sum_{j=1}^{k} a_{j} \log \left(2 b_{j}\right) \quad \text { and } \quad \beta=\sum_{j=1}^{k} a_{j}
$$

we derive that the positive number $\eta$ defined by

$$
\eta^{2}= \begin{cases}\frac{\beta}{n} L\left(\frac{e^{\frac{2 \alpha}{\beta}}}{\beta} n\right) & \text { if } n>2 \alpha \\ \frac{2 \alpha}{n} & \text { if } n \leq 2 \alpha\end{cases}
$$

is such that $D_{V}(\eta)=n \eta^{2}$. In particular, $\eta_{V} \leq \eta$. The conclusion ensues from some elementary inequalities on the Lambert's function.

We derive from this lemma the following result.

Lemma 6.14. Under the notations and assumptions of Proposition 6.12, there exists a universal positive constant $C$ such that

$$
C \eta_{V}^{2} \leq \frac{1}{n} \sum_{j=1}^{k}\left(\frac{\operatorname{dim}\left(W_{j}\right) \vee 1}{\alpha_{j}}\right)\left(\log \left(1+k R_{j} \rho_{j}^{\alpha_{j}}\right)+\log n\right) .
$$

Proof. We can upper bound $D_{V}(\eta)$ as follows.

$$
\begin{aligned}
D_{V}(\eta) & =\frac{1}{2} \vee \frac{1}{4} \sum_{j=1}^{k} \log \left(1+2\left(\frac{k R_{j}}{\eta}\right)^{1 / \alpha_{j}} \rho_{j}\right) \operatorname{dim}\left(W_{j}\right) \\
& \leq \frac{1}{2} \vee \frac{1}{4} \sum_{j=1}^{k} \frac{1}{\alpha_{j}} \log \left(1+2^{\alpha_{j}} \frac{k R_{j}}{\eta} \rho_{j}^{\alpha_{j}}\right) \operatorname{dim}\left(W_{j}\right) .
\end{aligned}
$$

We then use Lemma 6.13 with $a_{j}=\alpha_{j}^{-1}\left(1 \vee \operatorname{dim} W_{j}\right)$ and $b_{j}=1 \vee\left(2^{\alpha_{j}} k R_{j} \rho_{j}^{\alpha_{j}}\right)$ (we recall that $\alpha_{j} \leq 1$ ). There exists thus a universal constant $C^{\prime}$ such that

$$
C^{\prime} \eta_{V}^{2} \leq \frac{1}{n} \sum_{j=1}^{k}\left(\frac{\operatorname{dim}\left(W_{j}\right) \vee 1}{\alpha_{j}}\right)\left[\log \left(1+1 \vee\left(2^{\alpha_{j}} k R_{j} \rho_{j}^{\alpha_{j}}\right)\right)+\log \left(1+\frac{n}{\sum_{i=1}^{k}\left(\frac{\operatorname{dim}\left(W_{i}\right) \vee 1}{\alpha_{i}}\right)}\right)\right] .
$$

We now roughly upper-bound the right-hand side of this inequality to end the proof.
Let us return to the proof of Theorem 5.5. For all $W_{j} \in \mathbb{W}_{j}, \rho_{j} \in \mathbb{N}^{\star}$, we introduce the set $Z_{j}\left(W_{j}, \rho_{j}\right)=$ $W_{j} \cap \mathcal{B}_{\mathbf{x}}\left(0, \rho_{j}\right)$ where $\mathcal{B}_{\mathbf{x}}\left(0, \rho_{j}\right)$ is the closed ball centered at 0 with radius $\rho_{j}$ of the metric space $\left(\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right), d_{\mathbf{x}}\right)$. For all $\boldsymbol{W}=\left(W_{1}, \ldots, W_{k}\right) \in \prod_{j=1}^{k} \mathbb{W}_{j}, \boldsymbol{\rho}=\left(\rho_{1}, \ldots, \rho_{k}\right) \in\left(\mathbb{N}^{\star}\right)^{k}$, we define

$$
V(\boldsymbol{W}, \boldsymbol{\rho})=\left\{(t, x) \mapsto f_{\pi(u(x))}(t), \boldsymbol{u} \in \prod_{j=1}^{k} Z_{j}\left(W_{j}, \rho_{j}\right)\right\} .
$$

We then define

$$
\mathbb{V}=\left\{V(\boldsymbol{W}, \boldsymbol{\rho}), \boldsymbol{W} \in \prod_{j=1}^{k} \mathbb{W}_{j}, \boldsymbol{\rho} \in\left(\mathbb{N}^{\star}\right)^{k}\right\}
$$

and we define the map $\Delta$ on $\mathbb{V}$ by

$$
\Delta(V(\boldsymbol{W}, \boldsymbol{\rho}))=\sum_{j=1}^{k}\left(\Delta_{j}\left(W_{j}\right)+\log \left(2 \rho_{j}^{2}\right)\right)
$$

We apply Theorem 2.2 with $(\mathbb{V}, \Delta)$ to build an estimator $\hat{s}$. For all $\boldsymbol{W}=\left(W_{1}, \ldots, W_{k}\right) \in \prod_{j=1}^{k} \mathbb{W}_{j}, \boldsymbol{\rho}=$ $\left(\rho_{1}, \ldots, \rho_{k}\right) \in\left(\mathbb{N}^{\star}\right)^{k}$,

$$
\begin{equation*}
C \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq d_{2}^{2}(\sqrt{s}, V(\boldsymbol{W}, \boldsymbol{\rho}))+\eta_{V(\boldsymbol{W}, \boldsymbol{\rho})}^{2}+\frac{\Delta(V(\boldsymbol{W}, \boldsymbol{\rho}))}{n} \tag{6.11}
\end{equation*}
$$

where $C$ is a universal positive constant. We then derive from Lemma 6.14 that there exists a universal positive constant $C^{\prime}$ such that

$$
C^{\prime} \eta_{V(\boldsymbol{W}, \boldsymbol{\rho})}^{2} \leq \frac{1}{n} \sum_{j=1}^{k}\left(\frac{\operatorname{dim}\left(W_{j}\right) \vee 1}{\alpha_{j}}\right)\left(\log \left(1+k R_{j} \rho_{j}^{\alpha_{j}}\right)+\log n\right) .
$$

In particular, for all function $f \in \mathscr{F}$ of the form $f(t, x)=f_{\boldsymbol{u}(x)}(t)$, for all $\boldsymbol{W} \in \prod_{j=1}^{k} \mathbb{W}_{j}$, for all map $\boldsymbol{v}=$ $\left(v_{1}, \ldots, v_{k}\right) \in \prod_{j=1}^{k} W_{j}$, such that for all $j \in\{1, \ldots, k\},\left\|v_{j}\right\|_{\mathrm{x}} \leq\left\|u_{j}\right\|_{\mathrm{x}}$, and for all function $g$ of the form $g(t, x)=f_{\pi(\boldsymbol{v}(x))}(t)$, inequality (6.11) used with $\rho_{j}=\inf \left\{i \in \mathbb{N}^{\star}, i \geq\left\|u_{j}\right\|_{\mathrm{x}}\right\}$ leads to

$$
\begin{aligned}
C^{\prime \prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq & d_{2}^{2}(\sqrt{s}, f)+d_{2}^{2}(f, g)+\frac{\sum_{j=1}^{k}\left(\Delta_{j}\left(W_{j}\right)+\log \left(1+\left\|u_{j}\right\|_{\mathrm{x}}\right)\right)}{n} \\
& +\frac{1}{n} \sum_{j=1}^{k}\left(\frac{\operatorname{dim}\left(W_{j}\right) \vee 1}{\alpha_{j}}\right)\left[\log \left(1+k R_{j}\left(1+\left\|u_{j}\right\|_{\mathrm{x}}\right)^{\alpha_{j}}\right)+\log n\right]
\end{aligned}
$$

where $C^{\prime \prime}$ is a universal positive constant.
By using Assumption 5.1 and the Cauchy-Schwarz's inequality,

$$
d_{2}^{2}(f, g) \leq k \sum_{j=1}^{k} R_{j}^{2}\left\|u_{j}-v_{j}\right\|_{\mathrm{x}}^{2 \alpha_{j}}
$$

We then choose $v_{j}$ as being the projection of $u_{j}$ on $W_{j}$ in the space $\mathbb{L}^{2}\left(\mathbb{X}, \nu_{n}\right)$, and take the infimum over all $\boldsymbol{W} \in \prod_{j=1}^{k} \mathbb{W}_{j}$ to conclude.

### 6.9. Proof of Corollary 5.6

Let $r_{1}, r_{2} \in \mathbb{N}^{\star}$. We may apply Theorem 5.5 with collections $\mathbb{W}_{1}, \mathbb{W}_{2}$ provided by Proposition 1 of [5]. This yields an estimator $\tilde{s}$ such that for all $a \in \mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^{k_{2}}\right)$ with values into $\left[-r_{1}, r_{1}\right]$, for all $b \in \mathcal{H}^{\boldsymbol{\beta}}\left([0,1]^{k_{2}}\right)$ with values into $\left[-1 / 2+1 / r_{2},+\infty\right)$,

$$
C \mathbb{E}\left[H^{2}(s, \tilde{s})\right] \leq d_{2}^{2}(\sqrt{s}, f)+\varepsilon_{1}(a)+\varepsilon_{2}(b)
$$

where

$$
\begin{aligned}
C_{1} \varepsilon_{1}(a) \leq & \left(r_{2}^{1 / 2} L(a)\right)^{\frac{2 k_{2}}{k_{2}+2 \alpha}}\left(\frac{\log n+\log \left(1 \vee r_{2}^{1 / 2}\right)+\log \left(1 \vee\|a\|_{\mathrm{x}}\right)}{n}\right)^{\frac{2 \bar{\alpha}}{2 \alpha+k_{2}}} \\
& +\frac{\log n+\log \left(1 \vee r_{2}^{1 / 2}\right)+\log \left(1 \vee\|a\|_{\mathrm{x}}\right)}{n} \\
C_{2} \varepsilon_{2}(b) \leq & \left(\sqrt{2} r_{1} r_{2}^{3 / 2} L(b)\right)^{\frac{2 k_{2}}{k_{2}+2 \beta}}\left(\frac{\log n+\log \left(1 \vee \sqrt{2} r_{1} r_{2}^{3 / 2}\right)+\log \left(1 \vee\|b\|_{\mathrm{x}}\right)}{n}\right)^{\frac{2 \bar{\beta}}{2 \beta+k_{2}}} \\
& +\frac{\log n+\log \left(1 \vee \sqrt{2} r_{1} r_{2}^{3 / 2}\right)+\log \left(1 \vee\|b\|_{\mathrm{x}}\right)}{n}
\end{aligned}
$$

where $C>0$ is universal, where $C_{1}>0$ depends only on $k_{2}$, $\max _{1 \leq j \leq k_{2}} \alpha_{j}$, and where $C_{2}>0$ depends only on $k_{2}, \max _{1 \leq j \leq k_{2}} \beta_{j}$.

The above estimator depends on $r_{1}$ and $r_{2}$. We can then use Proposition 2.3, to derive that there exists an estimator $\hat{s}$, such that for all $r_{1}, r_{2} \in \mathbb{N}^{\star}$, for all $a \in \mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^{k_{2}}\right)$ with values into $\left[-r_{1}, r_{1}\right]$, for all $b \in \mathcal{H}^{\boldsymbol{\beta}}\left([0,1]^{k_{2}}\right)$ with values into $\left[-1 / 2+1 / r_{2},+\infty\right)$,

$$
C^{\prime \prime} \mathbb{E}\left[H^{2}(s, \hat{s})\right] \leq d_{2}^{2}(\sqrt{s}, f)+\varepsilon_{1}(a)+\varepsilon_{2}(b)+\frac{\log r_{1}+\log r_{2}}{n}
$$

where $C^{\prime \prime}>0$ is universal.

In particular, if we choose $r_{1}$ as being the smallest integer larger than $\|a\|_{\infty}$ and $r_{2}$ as being the smallest integer larger than $2 /\left(\inf _{x \in[0,1]^{k_{2}}}(2 b(x)+1)\right)$, we get

$$
C_{1}^{\prime} \varepsilon_{1}(a) \leq\left(\sqrt{1+\frac{2}{\inf _{x \in[0,1]^{k_{2}}}(2 b(x)+1)}} L(a)\right)^{\frac{2 k_{2}}{k_{2}+2 \bar{\alpha}}}\left(\frac{\log n}{n}\right)^{\frac{2 \bar{\alpha}}{2 \bar{\alpha}+k_{2}}}+C_{1}^{\prime \prime} \frac{\log n}{n}
$$

where $C_{1}^{\prime}>0$ depends only on $k_{2}, \max _{1 \leq j \leq k_{2}} \alpha_{j}$ and where $C_{1}^{\prime \prime}$ depends only on $k_{2}, \overline{\boldsymbol{\alpha}},\|a\|_{\infty}, L(a),\|b\|_{\infty}$ and $\inf _{x \in[0,1]^{k_{2}}}(2 b(x)+1)$. We then use

$$
1+\frac{2}{\inf _{x \in[0,1]^{k_{2}}}(2 b(x)+1)} \leq \frac{3}{1 \wedge \inf _{x \in[0,1]^{k_{2}}}(2 b(x)+1)}
$$

to get

We can bound from above $\varepsilon_{2}(b)$ in a similar fashion.

Acknowledgements. Many thanks to Yannick Baraud for his suggestions, comments and careful reading of the paper. We gratefully acknowledge an anonymous referee for its comments on an earlier draft of this paper.

## References

[1] A. Antoniadis and T. Sapatinas, Wavelet shrinkage for natural exponential families with quadratic variance functions. Biometrika 88 (2001) 805-820.
[2] A. Antoniadis, P. Besbeas and T. Sapatinas, Wavelet shrinkage for natural exponential families with cubic variance functions. Sankhyā. Indian J. Stat., Series A 63 (2001) 309-327.
[3] Y. Baraud, Estimator selection with respect to Hellinger-type risks. Probab. Theory Relat. Fields 151 (2011) 353-401.
[4] Y. Baraud and L. Birgé, Estimating the intensity of a random measure by histogram type estimators. Probab. Theory Relat. Fields 143 (2009) 239-284.
[5] Y. Baraud and L. Birgé, Estimating composite functions by model selection. Ann. Inst. Henri Poincaré, Probab. Stat. 50 (2014) 285-314.
[6] A.R. Barron and T.M. Cover, Minimum complexity density estimation. IEEE Trans. Inform. Theory 37 (1991) 1034-1054.
[7] L. Birgé, Model selection via testing: an alternative to (penalized) maximum likelihood estimators. Ann. Inst. Henri Poincaré, Probab. Stat. 42 (2006) 273-325.
[8] L. Birgé, Model selection for Poisson processes. In Asymptotics: particles, processes and inverse problems. In vol. 55 of $I M S$ Lect. Notes Monogr. Ser. Inst. Math. Statist., Beachwood, OH (2007) 32-64.
[9] L. Birgé, Model selection for density estimation with $\mathbb{L}_{2}$-loss. Probab. Theory Relat. Fields (2013) 1-42.
[10] F. Comte, S. Gaiffas and A. Guilloux, Adaptive estimation of the conditional intensity of marker-dependent counting processes. Ann. Inst. Henri Poincaré, Probab. Stat. 47 (2011) 1171-1196.
[11] W. Dahmen, R. DeVore and K. Scherer, Multi-dimensional spline approximation. SIAM J. Numer. Anal. 17 (1980) 380-402.
[12] J.T. Duane, Learning curve approach to reliability monitoring. IEEE Trans. Aerospace 2 (1964) 563-566.
[13] A.L. Goel and K. Okumoto, Time-dependent error-detection rate model for software reliability and other performance measures. IEEE Trans. Reliab. R-28 (1979) 206-211.
[14] A. Juditsky, O. Lepski and A. Tsybakov, Nonparametric estimation of composite functions. Ann. Stat. 37 (2009) 1360-1404.
[15] K. Krishnamurthy, M. Raginsky and R. Willett, Multiscale photon-limited spectral image reconstruction. SIAM J. Imaging Sci. 3 (2010) 619-645.
[16] P. Massart, Concentration inequalities and model selection. École d'été de Probabilités de Saint-Flour. In vol. 1896 of Lect. Notes Math. Springer, Berlin/Heidelberg (2003).
[17] P. Reynaud-Bouret, Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. Probab. Theory Relat. Fields 126 (2003) 103-153.
[18] S. Yamada, O. Mitsuru and S. Osaki, S-shaped reliability growth modeling for software error detection. IEEE Trans. Reliab. 32 (1983) 475-484.


[^0]:    Keywords and phrases. Adaptive estimation, model selection, Poisson processes, T-estimator.
    ${ }^{1}$ Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France. mathieu.sart@unice.fr

