ROBUST ESTIMATION ON A PARAMETRIC MODEL WITH TESTS

MATHIEU SART

ABSTRACT. We are interested in the problem of robust parametric estimation of a density from i.i.d observations. By using a practice-oriented procedure based on robust tests, we build an estimator for which we establish non-asymptotic risk bounds with respect to the Hellinger distance under mild assumptions on the parametric model. We prove that the estimator is robust even for models for which the maximum likelihood method is bound to fail. We also evaluate the performance of the estimator by carrying out numerical simulations for which we observe that the estimator is very close to the maximum likelihood one when the model is regular enough and contains the true underlying density.

1. Introduction

Consider $n$ independent and identically random variables $X_1,\ldots,X_n$ defined on an abstract probability space $(\Omega,\mathcal{E},\mathbb{P})$ with values in the measured space $(\mathbb{X},\mathcal{F},\mu)$. We suppose that the distribution of $X_i$ admits a density $s$ with respect to $\mu$ and aim at estimating $s$ by using a parametric approach.

When the unknown density $s$ is assumed to belong to a parametric model $\mathcal{F} = \{f_\theta, \theta \in \Theta\}$ of densities, a traditional method to estimate $s = f_{\theta_0}$ is the maximum likelihood one. It is indeed well known that the maximum likelihood estimator (m.l.e for short) possesses nice statistical properties such as consistency and asymptotic efficiency when the model $\mathcal{F}$ is regular enough. However, it is also well known that this estimator breaks down for many models $\mathcal{F}$ of interest and counter examples may be found in Pitman (1979); Ferguson (1982); Le Cam (1990); Birgé (2006) among other references.

Another drawback of the m.l.e lies in the fact that it is not robust. This means that if $s$ lies in a small neighbourhood of the model $\mathcal{F}$ but not in it, the m.l.e may perform poorly. Several kinds of robust estimators have been suggested in the literature to overcome this issue. We can cite the well known $L$ and $M$ estimators (which includes the class of minimum divergences estimators of Basu et al. (1998)) and the class of estimators built from a preliminary non-parametric estimator (such as the minimum Hellinger distance estimators introduced in Beran (1977) and the related estimators of Lindsay (1994); Basu and Lindsay (1994)).

In this paper, we focus on estimators built from robust tests. This approach, which begins in the 1970s with the works of Lucien Lecam and Lucien Birgé (Le Cam (1973, 1975); Birgé (1983, 1984a,b)), has the nice theoretical property to yield robust estimators under weak assumptions on the model $\mathcal{F}$. A key modern reference on this topic is Birgé (2006). The recent papers Birgé
Baraud and Birgé (2009); Baraud (2011, 2012); Sart (2013a, b) show that increasing attention is being paid to this kind of estimator. Their main interest is to provide general theoretical results in various statistical settings (such as general model selection theorems) which are usually unattainable by the traditional procedures (such as those based on the minimization of a penalized contrast).

For our statistical issue, the procedures using tests are based on the pairwise comparison of the elements of a thin discretisation \( \mathcal{F}_{\text{dis}} \) of \( \mathcal{F} \), that is, a finite or countable subset \( \mathcal{F}_{\text{dis}} \) of \( \mathcal{F} \) such that for all function \( f \in \mathcal{F} \), the distance between \( f \) and \( \mathcal{F}_{\text{dis}} \) is small (in a suitable sense).

As a result, their complexities are of order the square of the cardinality of \( \mathcal{F}_{\text{dis}} \). Unfortunately, this cardinality is often very large, making the construction of the estimators difficult in practice.

The aim of this paper is to develop a faster way of using tests to build an estimator when the cardinality of \( \mathcal{F}_{\text{dis}} \) is large.

From a theoretical point of view, the estimator we propose possesses similar statistical properties than those proved in Birgé (2006); Baraud (2011). Under mild assumptions on \( \mathcal{F} \), we build an estimator \( \hat{s} = f_{\hat{\theta}} \) of \( s \) such that

\[
\mathbb{P} \left[ C h^2(s, f_{\hat{\theta}}) \geq \inf_{\theta \in \Theta} h^2(s, f_\theta) + \frac{d}{n} + \xi \right] \leq e^{-n\xi} \quad \text{for all } \xi > 0,
\]

where \( C \) is a positive number depending on \( \mathcal{F} \), \( h \) the Hellinger distance and \( d \) such that \( \Theta \subset \mathbb{R}^d \). We recall that the Hellinger distance is defined on the cone \( L_1^+ (X, \mu) \) of non-negative integrable functions on \( X \) with respect to \( \mu \) by

\[
h^2(f, g) = \frac{1}{2} \int_X \left( \sqrt{f(x)} - \sqrt{g(x)} \right)^2 d\mu(x) \quad \text{for all } f, g \in L_1^+ (X, \mu).
\]

Let us make some comments on (1). When \( s \) does belong to the model \( \mathcal{F} \), the estimator achieves a quadratic risk of order \( n^{-1} \) with respect to the Hellinger distance. Besides, there exists \( \theta_0 \in \Theta \) such that \( s = f_{\theta_0} \) and we may then derive from (1) the rate of convergence of \( \hat{\theta} \) to \( \theta_0 \). In general, we do not suppose that the unknown density belongs to the model but rather use \( \mathcal{F} \) as an approximate class (sieve) for \( s \). Inequality (1) shows then that the estimator \( \hat{s} = f_{\hat{\theta}} \) cannot be strongly influenced by small departures from the model. As a matter of fact, if \( \inf_{\theta \in \Theta} h^2(s, f_\theta) \leq n^{-1} \), which means that the model is slightly misspecified, the quadratic risk of the estimator \( \hat{s} = f_{\hat{\theta}} \) remains of order \( n^{-1} \). This can be interpreted as a robustness property.

The preceding inequality (1) is interesting because it proves that our estimator is robust and converges at the right rate of convergence when the model is correct. However, the constant \( C \) depends on several parameters on the model such as the size of \( \Theta \). It is thus far from obvious that such an estimator can be competitive against more traditional estimators (such as the m.l.e).

In this paper, we try to give a partial answer for our estimator by carrying out numerical simulations. When a very thin discretisation \( \mathcal{F}_{\text{dis}} \) is used, the simulations show that our estimator is very close to the m.l.e when the model is regular enough and contains \( s \). More precisely, the larger is the number of observations \( n \), the closer they are, suggesting that our estimator inherits the efficiency of the m.l.e. Of course, this does not in itself constitute a proof but this allows to indicate what kind of results can be expected. A theoretical connection between estimators
built from tests (with the procedure described in Baraud (2011)) and the m.l.e will be found in a future paper of Yannick Baraud and Lucien Birgé.

In the present paper, we consider the problem of estimation on a single model. Nevertheless, when the statistician has at disposal several candidate models for \( s \), a natural issue is model selection. In order to address it, one may associate to each of these models the estimator resulting from our procedure and then select among those estimators by means of the procedure of Baraud (2011). By combining his Theorem 2 with our risk bounds on each individual estimator, we obtain that the selected estimator satisfies an oracle-type inequality.

We organize this paper as follows. We begin with a glimpse of the results in Section 2. We then present a procedure and its associated theoretical results to deal with models parametrized by an unidimensional parameter in Section 3. We evaluate its performance in practice by carrying out numerical simulations in Section 4. We work with models parametrized by a multidimensional parameter in Sections 5 and 6. The proofs are postponed to Section 6. Some technical results about the practical implementation of our procedure devoted to the multidimensional models are delayed to Section 7.

Let us introduce some notations that will be used all along the paper. The number \( x \vee y \) (respectively \( x \wedge y \)) stands for \( \max(x,y) \) (respectively \( \min(x,y) \)) and \( x_+ \) stands for \( x \vee 0 \). We set \( \mathbb{N}^* = \mathbb{N} \setminus \{0\} \). The vector \((\theta_1, \ldots, \theta_d)\) of \( \mathbb{R}^d \) is denoted by the bold letter \( \theta \). Given a set of densities \( \mathcal{F} = \{ f_\theta, \theta \in \Theta \} \), for all \( A \subset \Theta \), the notation \( \operatorname{diam} A \) stands for \( \sup_{\theta, \theta' \in A} h^2(f_\theta, f_{\theta'}) \).

The cardinality of a finite set \( A \) is denoted by \( |A| \). 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 indicator function of a subset \( A \) is denoted by \( 1_A \). The notations \( C,C',C'' \ldots \) are for the constants. The constants \( C,C',C'' \ldots \) may change from line to line.

2. An overview of the paper

2.1. Assumption. In this paper, we shall deal with sets of densities \( \mathcal{F} = \{ f_\theta, \theta \in \Theta \} \) indexed by a rectangle

\[
\Theta = \prod_{j=1}^d [m_j, M_j]
\]

of \( \mathbb{R}^d \). A such set will be called model. From now on, we consider models satisfying the following assumption.

**Assumption 1.** There exist positive numbers \( \alpha_1, \ldots, \alpha_d, R_1, \ldots, R_d, R_1', \ldots, R_d' \) such that for all \( \theta = (\theta_1, \ldots, \theta_d), \theta' = (\theta'_1, \ldots, \theta'_d) \in \Theta = \prod_{j=1}^d [m_j, M_j] \)

\[
\sup_{j \in \{1, \ldots, d\}} R_j |\theta_j - \theta'_j|^{\alpha_j} \leq h^2(f_\theta, f_{\theta'}) \leq \sup_{j \in \{1, \ldots, d\}} R_j |\theta_j - \theta'_j|^{\alpha_j}.
\]

This assumption allows to connect a (quasi) distance between the parameters to the Hellinger one between the corresponding densities. A similar assumption may be found in Theorem 5.8 of Chapter 1 of Ibragimov and Has’minskii (1981) to prove results on the maximum likelihood estimator. They require however that the application \( \theta \mapsto f_\theta(x) \) is continuous for \( \mu \) almost all \( x \).
to ensure the existence and the consistency of the m.l.e. Without this additional assumption, the m.l.e may not exist as shown by the translation model

\[ \mathcal{F} = \{ f_\theta, \theta \in [-1, 1] \} \text{ where } f_\theta(x) = \begin{cases} \frac{1}{4\sqrt{|x-\theta|}}1_{[-1,1]}(x-\theta) & \text{for all } x \in \mathbb{R} \setminus \{\theta\} \\ 0 & \text{for } x = \theta \end{cases} \]

for which Assumption 1 holds with \( \alpha_1 = 1/2 \).

Under suitable regularity conditions on the model, Theorem 7.6 of Chapter 1 of Ibragimov and Has'minskii (1981) shows that this assumption is fulfilled with \( \alpha_1 = \cdots = \alpha_d = 2 \). Other kinds of sufficient conditions implying Assumption 1 may be found in this book (see the beginning of Chapter 5 and Theorem 1.1 of Chapter 6). Other examples and counter-examples are given in Chapter 7 of Dacunha-Castelle (1978). Several models of interest satisfying this assumption will appear later in the paper.

2.2. Risk bound. In this paper, the risk bound we get for our estimator \( \hat{s} \) is similar to the one we would get by the procedures of Birgé (2006); Baraud (2011). More precisely:

**Theorem 1.** Suppose that Assumption 1 holds. We can build an estimator \( \hat{s} = f_{\hat{\theta}} \) such that for all \( \xi > 0 \),

\[
P \left[ Ch^2(s, f_{\hat{\theta}}) \geq h^2(s, \mathcal{F}) + \frac{d}{n} + \xi \right] \leq e^{-n\xi}
\]

where \( C > 0 \) depends on \( \sup_{1 \leq j \leq d} \bar{R}_j/R_j \) and \( \min_{1 \leq j \leq d} \alpha_j \).

We deduce from this risk bound that if \( s = f_{\theta_0} \) belongs to the model \( \mathcal{F} \), the estimator \( \hat{\theta} \) converges to \( \theta_0 \) and the random variable \( h^2(s, f_{\hat{\theta}}) \) is of order \( n^{-1} \). Besides, we may then derive from Assumption 1 that there exist positive numbers \( a, b_j \) such that

\[
P \left[ n^{1/\alpha_j} | \hat{\theta}_j - \theta_{0,j} | \geq \xi \right] \leq ae^{-n\xi} \quad \text{for all } j \in \{1, \ldots, d\} \text{ and } \xi > 0.
\]

Precisely, \( a = e^d \) and \( b_j = CR_j \). We emphasize here that this exponential inequality on \( \hat{\theta}_j \) is non-asymptotic but that the constants \( a, b_j \) are unfortunately far from optimal.

As explained in the introduction, there is no assumption on the true underlying density \( s \), which means that the model \( \mathcal{F} \) may be misspecified. In particular, when the squared Hellinger distance between the unknown density and the model \( \mathcal{F} \) is of order \( n^{-1} \), the random variable \( h^2(s, f_{\hat{\theta}}) \) remains of order \( n^{-1} \). This shows that the estimator \( \hat{s} \) possesses robustness properties.

2.3. Numerical complexity. The main interest of our procedures with respect to those of Birgé (2006); Baraud (2011) lies in their numerical complexity. More precisely, we shall prove the proposition below.

**Proposition 2.** Under Assumption 1, we can build an estimator \( \hat{s} \) satisfying (2) in less than

\[
4nC^{d/\alpha} \left[ \prod_{j=1}^{d} \left( 1 + (\bar{R}_j/R_j)^{1/\alpha_j} \right) \right] \left[ \sum_{j=1}^{d} \max \left\{ 1, \log \left( (n\bar{R}_j/d)^{1/\alpha_j} (M_j - m_j) \right) \right\} \right]
\]
operations. In the above inequality, $C$ is a constant larger than 1 (independent of $n$ and the model $\mathcal{F}$) and $\bar{\alpha}$ stands for the harmonic mean of $\alpha$, that is

$$\frac{1}{\bar{\alpha}} = \frac{1}{d} \sum_{j=1}^{d} \frac{1}{\alpha_j}.$$

If we are interested in the complexity when $n$ is large, we may deduce that this upper-bound is asymptotically equivalent to $C'n \log n$ where

$$C' = 4(d/\bar{\alpha})C^d/\bar{\alpha} \prod_{j=1}^{d} \left( 1 + \frac{\overline{R_j}}{\underline{R_j}} \right)^{1/\alpha_j}.$$

This constant is of reasonable size when $d$, $1/\bar{\alpha}$, $(\overline{R_j}/\underline{R_j})^{1/\alpha_j}$ are not too large.

Remark. The constant $C'$ does not depend only on the model but also on its parametrisation.

As a matter of fact, in the uniform model

$$\mathcal{F} = \{ f_{\theta}, \theta \in [m_1, M_1] \} \quad \text{where} \quad f_{\theta} = \theta^{-1} \mathbb{1}_{[0,\theta]}$$

we can compute explicitly the Hellinger distance

$$h^2(f_{\theta}, f_{\theta'}) = \frac{|\theta' - \theta|}{(\sqrt{\theta} + \sqrt{\theta'}) \sqrt{\max(\theta, \theta')}}$$

and bound it from above and from below by

$$\frac{1}{2M_1} |\theta' - \theta| \leq h^2(f_{\theta}, f_{\theta'}) \leq \frac{1}{2m_1} |\theta' - \theta|.$$ 

Now, if we parametrise $\mathcal{F}$ as

$$\mathcal{F} = \{ f_{e^t}, t \in [\log m_1, \log M_1] \},$$

then the Hellinger becomes $h^2(f_{e^t}, f_{e^{t'}}) = 1 - e^{-|t' - t|/2}$, and we can bound it from above and from below by

$$\frac{1 - \sqrt{m_1/M_1}}{\log(M_1/m_1)} |t' - t| \leq h^2(f_{e^t}, f_{e^{t'}}) \leq \frac{1}{2} |t' - t|.$$ 

Assumption 1 is satisfied in both case but with different values of $R_1$ and $\overline{R_1}$. When $M_1/m_1$ is large, the second parametrisation is much more interesting since it leads to a smaller constant $C'$.

3. Models parametrized by an unidimensional parameter

We now describe our procedure when the parametric model $\mathcal{F}$ is indexed by an interval $\Theta = [m_1, M_1]$ of $\mathbb{R}$. Throughout this section, Assumption 1 is supposed to be fulfilled. For the sake of simplicity, the subscripts of $m_1, M_1$ and $\alpha_1$ are omitted.
3.1. Basic ideas. We begin to detail the heuristics on which is based our procedure. We assume in this section that \( s \) belongs to the model \( \mathcal{F} \), that is, there exists \( \theta_0 \in \Theta = [m, M] \) such that \( s = f_{\theta_0} \). The starting point is the existence for all \( \theta, \theta' \in \Theta \) of a measurable function \( T(\theta, \theta') \) of the observations \( X_1, \ldots, X_n \) such that

1. For all \( \theta, \theta' \in \Theta \), \( T(\theta, \theta') = -T(\theta', \theta) \).
2. There exists \( \kappa > 0 \) such that if \( \mathbb{E}[T(\theta, \theta')] \) is non-negative, then
   \[
   h^2(s, f_\theta) > \kappa h^2(f_\theta, f_{\theta'}).
   \]
3. For all \( \theta, \theta' \in \Theta \), \( T(\theta, \theta') \) and \( \mathbb{E}[T(\theta, \theta')] \) are close (in a suitable sense).

For all \( \theta \in \Theta \), \( r > 0 \), let \( \mathcal{B}(\theta, r) \) be the Hellinger ball centered at \( \theta \) with radius \( r \), that is
\[
\mathcal{B}(\theta, r) = \{ \theta' \in \Theta : h(f_\theta, f_{\theta'}) \leq r \}.
\]

For all \( \theta, \theta' \in \Theta \), we deduce from the first point that either \( T(\theta, \theta') \) is non-negative, or \( T(\theta', \theta) \) is non-negative. It is likely that it follows from 2. and 3. that in the first case \( \theta_0 \in \Theta \setminus \mathcal{B}(\theta, \kappa^{1/2} h(f_\theta, f_{\theta'})) \) while in the second case \( \theta_0 \in \Theta \setminus \mathcal{B}(\theta', \kappa^{1/2} h(f_\theta, f_{\theta'})) \).

These sets may be interpreted as confidence sets for \( \theta_0 \).

The main idea is to build a decreasing sequence (in the sense of inclusion) of intervals \( (\Theta_i)_i \). Set \( \theta^{(1)} = m \), \( \theta^{(1)} = M \), and \( \Theta_1 = [\theta^{(1)}, \theta^{(1)}] \) (which is merely \( \Theta \)). If \( T(\theta^{(1)}, \theta^{(1)}) \) is non-negative, we consider a set \( \Theta_2 \) such that
\[
\Theta_1 \setminus \mathcal{B} \left( \theta^{(1)}, \kappa^{1/2} h(f_{\theta^{(1)}}, f_{\theta^{(1)}}) \right) \subset \Theta_2 \subset \Theta_1
\]
while if \( T(\theta^{(1)}, \theta^{(1)}) \) is non-positive, we consider a set \( \Theta_2 \) such that
\[
\Theta_1 \setminus \mathcal{B} \left( \theta^{(1)}, \kappa^{1/2} h(f_{\theta^{(1)}}, f_{\theta^{(1)}}) \right) \subset \Theta_2 \subset \Theta_1.
\]

The set \( \Theta_2 \) may thus also be interpreted as a confidence set for \( \theta_0 \). Thanks to Assumption 1, we can define \( \Theta_2 \) as an interval \( \Theta_2 = [\theta_0, \theta_0] \).

We then repeat the idea to build an interval \( \Theta_3 = [\theta^{(2)}, \theta^{(2)}] \) included in \( \Theta_2 \) and containing either
\[
\Theta_3 \supset \Theta_2 \setminus \mathcal{B} \left( \theta^{(2)}, \kappa^{1/2} h(f_{\theta^{(2)}}, f_{\theta^{(2)}}) \right) \quad \text{or} \quad \Theta_3 \supset \Theta_2 \setminus \mathcal{B} \left( \theta^{(2)}, \kappa^{1/2} h(f_{\theta^{(2)}}, f_{\theta^{(2)}}) \right)
\]
according to the sign of \( T(\theta^{(2)}, \theta^{(2)}) \).

By induction, we build a decreasing sequence of such intervals \( (\Theta_i)_i \). We now consider an integer \( N \) large enough so that the length of \( \Theta_N \) is small enough. We then define the estimator \( \hat{\theta} \) as the center of the set \( \Theta_N \) and estimate \( s \) by \( f_{\hat{\theta}} \).
3.2. Definition of the test. The test $T(\theta, \theta')$ we use in our estimation strategy is the one of Baraud (2011) applied to two suitable densities of the model. More precisely, let $\mathcal{T}$ be the functional defined for all $g, g' \in \mathbb{L}^1(X, \mu)$ by

$$\mathcal{T}(g, g') = \frac{1}{n} \sum_{i=1}^{n} \frac{g'(X_i) - g(X_i)}{\sqrt{g(X_i)} + g'(X_i)} + \frac{1}{2} \int_{X} \sqrt{g(x) + g'(x)} \left( \sqrt{g'(x)} - \sqrt{g(x)} \right) \, d\mu(x)$$

where the convention $0/0 = 0$ is in use.

We consider $t \in (0, 1]$ and $\epsilon = t(\mathbb{R}^n)^{-1/\alpha}$. We then define the finite sets

$$\Theta_{\text{dis}} = \{ m + k\epsilon, k \in \mathbb{N}, k \leq (M - m)e^{-1} \}, \quad \mathcal{F}_{\text{dis}} = \{ f_\theta, \theta \in \Theta_{\text{dis}} \}$$

and the map $\pi$ on $[m, M]$ by

$$\pi(x) = m + \lfloor(x - m)/\epsilon\rfloor \epsilon \quad \text{for all} \ x \in [m, M]$$

where $\lfloor \cdot \rfloor$ denotes the integer part. We then define $T(\theta, \theta')$ by

$$T(\theta, \theta') = \mathcal{T}(f_{\pi(\theta)}, f_{\pi(\theta')}) \quad \text{for all} \ \theta, \theta' \in [m, M].$$

The aim of the parameter $t$ is to tune the thinness of the net $\mathcal{F}_{\text{dis}}$.

3.3. Procedure. We shall build a decreasing sequence $(\Theta_i)_{i \geq 1}$ of intervals of $\Theta = [m, M]$ as explained in Section 3.1. Let $\kappa > 0$, and for all $\theta, \theta' \in [m, M]$ such that $\theta' < \theta$, let $\bar{r}(\theta, \theta')$, $\underline{r}(\theta, \theta')$ be two positive numbers such that

$$[m, M] \cap [\theta + \bar{r}(\theta, \theta'), \theta'] \subset \mathcal{B}(\theta, \kappa^{1/2}h(f_\theta, f_{\theta'}))$$

and

$$[m, M] \cap [\theta', \theta - \underline{r}(\theta, \theta')] \subset \mathcal{B}(\theta', \kappa^{1/2}h(f_\theta, f_{\theta'}))$$

where we recall that $\mathcal{B}(\theta, \kappa^{1/2}h(f_\theta, f_{\theta'}))$ and $\mathcal{B}(\theta', \kappa^{1/2}h(f_\theta, f_{\theta'}))$ are the Hellinger balls defined by (3).

We set $\theta^{(1)} = m$, $\theta'(1) = M$ and $\Theta_1 = [\theta^{(1)}, \theta'(1)]$. We define the sequence $(\Theta_i)_{i \geq 1}$ by induction. When $\Theta_i = [\theta^{(i)}, \theta'(i)]$, we set

$$\theta^{(i+1)} = \begin{cases} \theta^{(i)} + \min \left\{ \bar{r}(\theta^{(i)}, \theta'(i)), \frac{\theta'(i) - \theta^{(i)}}{2} \right\} & \text{if} \ T(\theta^{(i)}, \theta'(i)) \geq 0 \\ \theta^{(i)} & \text{otherwise} \end{cases}$$

$$\theta'^{(i+1)} = \begin{cases} \theta'^{(i)} - \min \left\{ \underline{r}(\theta^{(i)}, \theta'(i)), \frac{\theta'(i) - \theta^{(i)}}{2} \right\} & \text{if} \ T(\theta^{(i)}, \theta'^{(i)}) \leq 0 \\ \theta'^{(i)} & \text{otherwise} \end{cases}$$

We then define $\Theta_{i+1} = [\theta^{(i+1)}, \theta'^{(i+1)}]$.

The role of conditions (5) and (6) is to ensure that $\Theta_{i+1}$ is big enough to contain one of the two confidence sets

$$\Theta_i \setminus \mathcal{B}\left(\theta^{(i)}, \kappa^{1/2}h(f_{\theta^{(i)}}, f_{\theta'(i)})\right) \quad \text{and} \quad \Theta_i \setminus \mathcal{B}\left(\theta'^{(i)}, \kappa^{1/2}h(f_{\theta^{(i)}}, f_{\theta'(i)})\right).$$

The parameter $\kappa$ allows to tune the level of these confidence sets. There is a minimum in the definitions of $\theta^{(i+1)}$ and $\theta'^{(i+1)}$ in order to guarantee the inclusion of $\Theta_{i+1}$ in $\Theta_i$.

We now consider a positive number $\eta$ and build these intervals until their lengths become smaller than $\eta$. The estimator we consider is then the center of the last interval built. This
parameter \( \eta \) stands for a measure of the accuracy of the estimation and must be small enough to get a suitable risk bound for our estimator.

The algorithm is the following.

**Algorithm 1**

1: \( \theta \leftarrow m \), \( \theta' \leftarrow M \)
2: **while** \( \theta' - \theta > \eta \) **do**
3: Compute \( r = \min \{ \bar{r}(\theta, \theta'), (\theta' - \theta)/2 \} \)
4: Compute \( r' = \min \{ r(\theta, \theta'), (\theta' - \theta)/2 \} \)
5: Compute Test = \( T(\theta, \theta') \)
6: **if** Test \( \geq 0 \) **then**
7: \( \theta \leftarrow \theta + r \)
8: **end if**
9: **if** Test \( \leq 0 \) **then**
10: \( \theta' \leftarrow \theta' - r' \)
11: **end if**
12: **end while**
13: **Return**: \( \hat{\theta} = (\theta + \theta')/2 \)

### 3.4. Risk bound.

The following theorem specify the values of the parameters \( t, \kappa, \eta \) that allow to control the risk of the estimator \( \hat{s} = f_{\hat{\theta}} \).

**Theorem 3.** Suppose that Assumption 1 holds. Set

\[
\kappa = \left( 1 + \sqrt{\frac{2 + \sqrt{2}}{2 - \sqrt{2}}} \right)^{-2}.
\]

Assume that \( t \in (0, 1], \kappa \in (0, \kappa], \eta \in [\epsilon, (\bar{R}n)^{-1/2}] \) and that \( \bar{r}(\theta, \theta'), r(\theta, \theta') \) are such that (5) and (6) hold.

Then, for all \( \xi > 0 \), the estimator \( \hat{\theta} \) built in Algorithm 1 satisfies

\[
\mathbb{P} \left[ C h^2(s, f_{\hat{\theta}}) \geq h^2(s, \mathcal{F}) + \frac{1}{n} + \xi \right] \leq e^{-n\xi}
\]

where \( C > 0 \) depends only on \( \kappa, t, \alpha, \bar{R}/R \).

A slightly sharper risk bound may be found in the proof of this theorem.

### 3.5. Choice of \( \bar{r}(\theta, \theta') \) and \( r(\theta, \theta') \).

These parameters are chosen by the statistician. They do not change the risk bound given by Theorem 3 (provided that (5) and (6) hold) but affect the speed of the procedure. The larger they are, the faster the procedure is. There are three different situations.
First case: the Hellinger distance \( h(f_{\theta}, f_{\theta'}) \) can be made explicit. We have thus an interest in defining them as the largest numbers for which (5) and (6) hold, that is

\[
\bar{r}(\theta, \theta') = \sup \left\{ r > 0, \ [m, M] \cap [\theta, \theta + r] \subset B(\theta, \kappa^{1/2}h(f_{\theta}, f_{\theta'})) \right\}
\]

(8)

\[
\underline{r}(\theta, \theta') = \sup \left\{ r > 0, \ [m, M] \cap [\theta' - r, \theta'] \subset B(\theta', \kappa^{1/2}h(f_{\theta}, f_{\theta'})) \right\}.
\]

(9)

Second case: the Hellinger distance \( h(f_{\theta}, f_{\theta'}) \) can be quickly evaluated numerically but the computation of (8) and (9) is difficult. We may then define them by

\[
r(\theta, \theta') = \bar{r}(\theta, \theta') = \left( \frac{\kappa}{\bar{R}} \right)^{1/\alpha} h(f_{\theta}, f_{\theta'})^{1/\alpha}.
\]

(10)

One can verify that (5) and (6) hold. When the model is regular enough and \( \alpha = 2 \), the value of \( \bar{R} \) can be calculated by using Fisher information (see for instance Theorem 7.6 of Chapter 1 of Ibragimov and Has'minskii (1981)).

Third case: the computation of the Hellinger distance \( h(f_{\theta}, f_{\theta'}) \) involves the numerical computation of an integral and this computation is slow. An alternative definition is then

\[
\underline{r}(\theta, \theta') = \bar{r}(\theta, \theta') = \left( \frac{\kappa \bar{R}}{R} \right)^{1/\alpha} (\theta' - \theta).
\]

(11)

As in the second point, one can check that (5) and (6) hold. Note however that the computation of the test also involves in most cases the numerical computation of an integral (see (4)). This third case is thus mainly devoted to models for which this numerical integration can be avoided, as for the translation models \( \mathcal{F} = \{ f(-\theta), \ \theta \in [m, M] \} \) with \( f \) even, \( \mathbb{X} = \mathbb{R} \) and \( \mu \) the Lebesgue measure (the second term of (4) is 0 for these models).

We can upper-bound the numerical complexity of the algorithm when \( \bar{r}(\theta, \theta') \) and \( \underline{r}(\theta, \theta') \) are large enough. Precisely, we prove the proposition below.

**Proposition 4.** Suppose that the assumptions of Theorem 3 hold and that \( \underline{r}(\theta, \theta'), \bar{r}(\theta, \theta') \) are larger than

\[
(\kappa \bar{R}/\bar{R})^{1/\alpha} (\theta' - \theta).
\]

(12)

Then, the number of tests computed to build the estimator \( \hat{\theta} \) is smaller than

\[
1 + \max \left\{ \left( \frac{\bar{R}/(\kappa \bar{R})}{1/\log 2} \right)^{1/\alpha}, 1/\log 2 \right\} \log \left( \frac{M - m}{\eta} \right).
\]

It is worthwhile to notice that this upper-bound does not depend on \( t \), that is the size of the net \( \mathcal{F}_{\text{dis}} \) contrary to the preceding procedures based on tests. Obviously, the parameter \( \eta \) is involved in this upper-bound, but the whole point is that it grows slowly with \( 1/\eta \), which allows to use the procedure with \( \eta \) very small.

4. Simulations for unidimensional models

In what follows, we carry out a simulation study in order to evaluate more precisely the performance of our estimator. We simulate samples \((X_1, \ldots, X_n)\) with density \( \mathcal{S} \) and use our procedure to estimate \( \mathcal{S} \).
4.1. Models. Our simulation study is based on the following models.

Example 1. \( \mathcal{F} = \{ f_\theta, \theta \in [0.01, 100] \} \) where
\[
f_\theta(x) = \theta e^{-\theta x} 1_{[0, +\infty)}(x) \quad \text{for all } x \in \mathbb{R}.
\]

Example 2. \( \mathcal{F} = \{ f_\theta, \theta \in [-100, 100] \} \) where
\[
f_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(x-\theta)^2}{2} \right) \quad \text{for all } x \in \mathbb{R}.
\]

Example 3. \( \mathcal{F} = \{ f_\theta, \theta \in [0.01, 100] \} \) where
\[
f_\theta(x) = \frac{x}{\theta^2} \exp \left( -\frac{x^2}{2\theta^2} \right) 1_{[0, +\infty)}(x) \quad \text{for all } x \in \mathbb{R}.
\]

Example 4. \( \mathcal{F} = \{ f_\theta, \theta \in [-10, 10] \} \) where
\[
f_\theta(x) = \frac{1}{\pi (1 + (x-\theta)^2)} \quad \text{for all } x \in \mathbb{R}.
\]

Example 5. \( \mathcal{F} = \{ f_\theta, \theta \in [0.01, 10] \} \) where \( f_\theta = \theta^{-1} 1_{[0,\theta]} \).

Example 6. \( \mathcal{F} = \{ f_\theta, \theta \in [-10, 10] \} \) where
\[
f_\theta(x) = \frac{1}{(x-\theta+1)^2} 1_{[\theta, +\infty)}(x) \quad \text{for all } x \in \mathbb{R}.
\]

Example 7. \( \mathcal{F} = \{ f_\theta, \theta \in [-10, 10] \} \) where \( f_\theta = 1_{[\theta-1/2, \theta+1/2]} \).

Example 8. \( \mathcal{F} = \{ f_\theta, \theta \in [-1, 1] \} \) where
\[
f_\theta(x) = \frac{1}{4\sqrt{|x-\theta|}} 1_{[-1,1]}(x-\theta) \quad \text{for all } x \in \mathbb{R} \setminus \{\theta\}
\]
and \( f_\theta(\theta) = 0 \).

In these examples, we shall mainly compare our estimator to the maximum likelihood one. In examples 1, 2, 3, 5 and 6, the m.l.e \( \hat{\theta}_{\text{me}} \) can be made explicit and is thus easy to compute. Finding the m.l.e is more delicate for the problem of estimating the location parameter of a Cauchy distribution, since the likelihood function may be multimodal. We refer to Barnett (1966) for a discussion of numerical methods devoted to the maximization of the likelihood. In our simulation study, we avoid the issues of the numerical algorithms by computing the likelihood at \( 10^6 \) equally spaced points between \( \max(-10, \hat{\theta}-1) \) and \( \min(10, \hat{\theta}+1) \) (where \( \hat{\theta} \) is our estimator) and at \( 10^6 \) equally spaced points between \( \max(-10, \hat{\theta}_{\text{median}}-1) \) and \( \min(10, \hat{\theta}_{\text{median}}+1) \) where \( \hat{\theta}_{\text{median}} \) is the median. We then select among these points the one for which the likelihood is maximal. In Example 5, we shall also compare our estimator to the minimum variance unbiased estimator defined by
\[
\hat{\theta}_{\text{mvub}} = \frac{n+1}{n} \max_{1 \leq i \leq n} X_i.
\]
In Example 7, we shall compare our estimator to
\[
\tilde{\theta} = \frac{1}{2} \left( \max_{1 \leq i \leq n} X_i + \min_{1 \leq i \leq n} X_i \right).
\]
In the case of Example 8, the likelihood is infinite at each observation and the maximum likelihood method fails. We shall then compare our estimator to the median and the empirical mean but also to the maximum spacing product estimator $\hat{\theta}_{\text{mspe}}$ (m.s.p.e for short). This estimator was introduced by Cheng and Amin (1983); Ranneby (1984) to deal with statistical models for which the likelihood is unbounded. The m.s.p.e is known to possess nice theoretical properties such as consistency and asymptotic efficiency and precise results on the performance of this estimator may be found in Cheng and Amin (1983); Ranneby (1984); Ekström (1998); Shao and Hahn (1999); Ghost and Jammalamadaka (2001); Anatolyev and Kose nok (2005) among other references. This last method involves the problem of finding a global maximum of the maximum product function on $\Theta = [-1, 1]$ (which may be multimodal). We compute it by considering $2 \times 10^5$ equally spaced points between $-1$ and 1 and by calculating for each of these points the function to maximize. We then select the point for which the function is maximal. Using more points to compute the m.s.p.e would give more accurate results, especially when $n$ is large, but we are limited by the computer.

4.2. Implementation of the procedure. Our procedure involves several parameters that must be chosen by the statistician.

Choice of $t$. This parameter tunes the thinness of the net $\mathcal{F}_{\text{dis}}$. When the model is regular enough and contains $s$, a good choice of $t$ seems to be $t = 0$ (that is $\Theta_{\text{dis}} = \Theta$, $\mathcal{F}_{\text{dis}} = \mathcal{F}$ and $T(\theta, \theta') = \mathcal{T}(f_\theta, f_{\theta'})$), since then the simulations suggest that our estimator is very close to the m.l.e when the model is true (with large probability). In the simulations, we take $t = 0$.

Choice of $\eta$. We take $\eta$ small: $\eta = (M - m)/10^8$.

Choice of $\kappa$. This constant influences the level of the confidence sets and thus the time of construction of the estimator: the larger is $\kappa$, the faster is the procedure. We take arbitrary $\kappa = \bar{\kappa}/2$.

Choice of $r(\theta, \theta')$ and $\bar{r}(\theta, \theta')$. In examples 1,2,3,5, and 7, we define them by (8) and (9). In examples 4 and 6, we define them by (10). In the first case, $\alpha = 2$ and $R = 1/16$, while in the second case, $\alpha = 1$ and $R = 1/2$. In the case of Example 8, we use (11) with $\alpha = 1/2$, $R = 0.17$ and $\bar{R} = 1/\sqrt{2}$.

4.3. Simulations when $s \in \mathcal{F}$. We begin to simulate $N$ samples $(X_1, \ldots, X_n)$ when the true density $s$ belongs to the model $\mathcal{F}$. They are generated according to the density $s = f_1$ in examples 1,3,5 and according to $s = f_0$ in examples 2,4,6,7,8.

We evaluate the performance of an estimator $\hat{\theta}$ by computing it on each of the $N$ samples. Let $\hat{\theta}^{(i)}$ be the value of this estimator corresponding to the $i$-th sample and let

$$\hat{R}_N(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} h^2(s, f_{\hat{\theta}^{(i)}}) \quad \text{and} \quad \hat{\text{std}}_N(\hat{\theta}) = \sqrt{\frac{1}{N - 1} \sum_{i=1}^{N} \left( h^2(s, f_{\hat{\theta}^{(i)}}) - \hat{R}_N(\hat{\theta}) \right)^2}.$$  

The risk $\mathbb{E}[h^2(s, f_{\hat{\theta}})]$ of the estimator $\hat{\theta}$ is thus estimated by $\hat{R}_N(\hat{\theta})$. More precisely, if $Q_c$ denotes the $c/2$ quantile of a standard Gaussian distribution,

$$\left[ \hat{R}_N(\hat{\theta}) - Q_c \frac{\hat{\text{std}}_N(\hat{\theta})}{\sqrt{N}}, \hat{R}_N(\hat{\theta}) + Q_c \frac{\hat{\text{std}}_N(\hat{\theta})}{\sqrt{N}} \right]$$
is a confidence interval for $E\left[h^2(s, f)\right]$ with asymptotic confidence level $c$. We also introduce

$$\widehat{R}_{N,\text{rel}}(\tilde{\theta}) = \frac{\widehat{R}_N(\tilde{\theta})}{\widehat{R}_N(\hat{\theta})} - 1$$

in order to make the comparison of our estimator $\hat{\theta}$ and the estimator $\tilde{\theta}$ easier. When $R_{\text{rel}}(\tilde{\theta})$ is negative our estimator is better than $\tilde{\theta}$ whereas if $R_{\text{rel}}(\tilde{\theta})$ is positive, our estimator is worse than $\hat{\theta}$. More precisely, if $R_{\text{rel}}(\tilde{\theta}) = \alpha$, the risk of our estimator corresponds to the one of $\tilde{\theta}$ reduced of $100|\alpha|\%$ when $\alpha < 0$ and increased of $100\alpha\%$ when $\alpha > 0$.

The results are gathered below.

| Example | $R_{10^6}(\theta)$ | $R_{10^6}(\hat{\theta}_{\text{mle}})$ | $R_{10^6,\text{rel}}(\hat{\theta}_{\text{mle}})$ | std$_{10^6}(\theta)$ | std$_{10^6}(\hat{\theta}_{\text{mle}})$ |
|---------|---------------------|--------------------------------------|------------------------------------------|----------------------|-----------------------------|
| Example 1 | 0.0130 | 0.0051 | 0.0025 | 0.0017 | 0.0013 |
|          | 0.0129 | 0.0051 | 0.0025 | 0.0017 | 0.0013 |
|          | 6 $\cdot$ 10$^{-4}$ | 10$^{-5}$ | 7 $\cdot$ 10$^{-7}$ | $-8$ $\cdot$ 10$^{-9}$ | 2 $\cdot$ 10$^{-9}$ |
| Example 2 | 0.0123 | 0.0050 | 0.0025 | 0.0017 | 0.0012 |
|          | 0.0123 | 0.0050 | 0.0025 | 0.0017 | 0.0012 |
|          | 5 $\cdot$ 10$^{-10}$ | 9 $\cdot$ 10$^{-10}$ | $-2$ $\cdot$ 10$^{-9}$ | $-2$ $\cdot$ 10$^{-9}$ | $-3$ $\cdot$ 10$^{-9}$ |
| Example 3 | 0.0170 | 0.0070 | 0.0035 | 0.0023 | 0.0018 |
|          | 0.0170 | 0.0070 | 0.0035 | 0.0023 | 0.0018 |
| Example 4 | 0.0192 | 0.0073 | 0.0036 | 0.0024 | 0.0018 |
|          | 0.0192 | 0.0073 | 0.0036 | 0.0024 | 0.0018 |
| Example 5 | 0.0255 | 0.0083 | 0.0039 | 0.0025 | 0.0018 |
|          | 0.0468 | 0.0192 | 0.0096 | 0.0064 | 0.0048 |
|          | 0.0476 | 0.0196 | 0.0099 | 0.0066 | 0.0050 |
|          | 0.0350 | 0.0144 | 0.0073 | 0.0049 | 0.0037 |
|          | -0.0160 | -0.0202 | -0.0287 | -0.0271 | -0.0336 |
|          | 0.3390 | 0.3329 | 0.3215 | 0.3243 | 0.3148 |

| Example | $R_{10^8}(\theta)$ | $R_{10^8}(\hat{\theta}_{\text{mle}})$ | $R_{10^8,\text{rel}}(\hat{\theta}_{\text{mle}})$ | std$_{10^8}(\theta)$ | std$_{10^8}(\hat{\theta}_{\text{mle}})$ |
|---------|---------------------|--------------------------------------|------------------------------------------|----------------------|-----------------------------|
| Example 1 | 0.0529 | 0.0223 | 0.0112 | 0.0075 | 0.0056 |
|          | 0.0453 | 0.0192 | 0.0098 | 0.0066 | 0.0049 |
|          | 0.0316 | 0.0132 | 0.0067 | 0.0045 | 0.0034 |
ROBUST ESTIMATION ON A PARAMETRIC MODEL WITH TESTS

Example 6

\[
\begin{array}{cccccc}
\hat{R}_{10^6}(\theta) & 0.0504 & 0.0197 & 0.0098 & 0.0065 & 0.0049 \\
\hat{R}_{10^6}(\tilde{\theta}_{\text{mle}}) & 0.0483 & 0.0197 & 0.0099 & 0.0066 & 0.0050 \\
\hat{R}_{10^6,\text{rel}}(\tilde{\theta}_{\text{mle}}) & 0.0436 & -0.0019 & -0.0180 & -0.0242 & -0.0263 \\
\hat{\text{std}}_{10^6}(\theta) & 0.0597 & 0.0233 & 0.0115 & 0.0076 & 0.0057 \\
\hat{\text{std}}_{10^6}(\tilde{\theta}_{\text{mle}}) & 0.0467 & 0.0195 & 0.0099 & 0.0066 & 0.0050 \\
\end{array}
\]

Example 7

\[
\begin{array}{cccccc}
\hat{R}_{10^6}(\theta) & 0.0455 & 0.0193 & 0.0098 & 0.0066 & 0.0050 \\
\hat{R}_{10^6}(\tilde{\theta}') & 0.0454 & 0.0192 & 0.0098 & 0.0066 & 0.0050 \\
\hat{R}_{10^6,\text{rel}}(\tilde{\theta}') & 0.0029 & 0.0029 & 0.0031 & 0.0028 & 0.0030 \\
\hat{\text{std}}_{10^6}(\theta) & 0.0416 & 0.0186 & 0.0096 & 0.0065 & 0.0049 \\
\hat{\text{std}}_{10^6}(\tilde{\theta}') & 0.0415 & 0.0185 & 0.0096 & 0.0065 & 0.0049 \\
\end{array}
\]

Example 8

\[
\begin{array}{cccccc}
\hat{R}_{10^4}(\theta) & 0.050 & 0.022 & 0.012 & 0.008 & 0.006 \\
\hat{R}_{10^4}(\tilde{\theta}_{\text{mean}}) & 0.084 & 0.061 & 0.049 & 0.043 & 0.039 \\
\hat{R}_{10^4}(\tilde{\theta}_{\text{median}}) & 0.066 & 0.036 & 0.025 & 0.019 & 0.017 \\
\hat{R}_{10^4}(\tilde{\theta}_{\text{mspe}}) & 0.050 & 0.022 & 0.012 & 0.008 & 0.006 \\
\hat{R}_{10^4,\text{rel}}(\tilde{\theta}_{\text{mean}}) & -0.40 & -0.64 & -0.76 & -0.82 & -0.85 \\
\hat{R}_{10^4,\text{rel}}(\tilde{\theta}_{\text{median}}) & -0.25 & -0.39 & -0.54 & -0.59 & -0.65 \\
\hat{\text{std}}_{10^4}(\theta) & 0.054 & 0.025 & 0.013 & 0.009 & 0.007 \\
\hat{\text{std}}_{10^4}(\tilde{\theta}_{\text{mean}}) & 0.045 & 0.032 & 0.025 & 0.022 & 0.020 \\
\hat{\text{std}}_{10^4}(\tilde{\theta}_{\text{median}}) & 0.052 & 0.032 & 0.020 & 0.016 & 0.014 \\
\hat{\text{std}}_{10^4}(\tilde{\theta}_{\text{mspe}}) & 0.051 & 0.025 & 0.014 & 0.009 & 0.007 \\
\end{array}
\]

In the first four examples, the risk of our estimator is very close to the maximum likelihood estimator one, whatever the value of \( n \). In Example 5, our estimator slightly improves the maximum likelihood estimator but is worse than the minimum variance unbiased estimator. In Example 6, the risk of our estimator is larger than the one of the m.l.e when \( n = 10 \) but is slightly smaller as soon as \( n \) becomes larger than 25. In Example 7, the risk of our estimator is 0.3% larger than the one of \( \tilde{\theta}' \). In Example 8, our estimator significantly improves the empirical mean and the median. Its risk is comparable to the one of the m.s.p.e (we omit in this example the value of \( \hat{R}_{10^4,\text{rel}}(\tilde{\theta}_{\text{mspe}}) \) because it is influenced by the procedure we have used to build the m.s.p.e).

When the model is regular enough, these simulations show that our estimation strategy provides an estimator whose risk is almost equal to the one of the maximum likelihood estimator. Moreover, our estimator seems to work rather well in a model where the m.l.e does not exist (case of Example 8). Remark that contrary to the maximum likelihood method, our procedure does not involve the search of a global maximum.

We now bring to light the connection between our estimator and the m.l.e when the model is regular enough (that is in the first four examples). Let for \( c \in \{0.99, 0.999, 1\} \), \( q_c \) be the \( c \)-quantile of the random variable \( |\hat{\theta} - \tilde{\theta}_{\text{mle}}| \), and \( \hat{q}_c \) be the empirical version based on \( N \) samples (\( N = 10^6 \) in examples 1,2,3 and \( N = 10^4 \) in Example 4). The results are the following.
Example 1
\[ \hat{q}_{0.99}^{0.99}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7} \]
\[ \hat{q}_{0.99}^{0.99}, 0.07, 0.06, 0.005, 10^{-7} \]
\[ \hat{q}_{1}^{0.99}, 1.9, 0.3, 10^{-7} \]

Example 2
\[ \hat{q}_{0.99}^{0.99}, 2 \cdot 10^{-7}, 3 \cdot 10^{-7}, 3 \cdot 10^{-7}, 3 \cdot 10^{-7}, 3 \cdot 10^{-7} \]
\[ \hat{q}_{0.99}^{0.99}, 3 \cdot 10^{-7}, 3 \cdot 10^{-7}, 3 \cdot 10^{-7}, 3 \cdot 10^{-7}, 3 \cdot 10^{-7} \]
\[ \hat{q}_{1}^{0.99}, 0.03, 0.12, 0.01, 0.007, 10^{-7} \]

Example 3
\[ \hat{q}_{0.99}^{0.99}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7} \]
\[ \hat{q}_{0.99}^{0.99}, 0.03, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7} \]
\[ \hat{q}_{1}^{0.99}, 0.38, 0.12, 0.01, 10^{-7} \]

Example 4
\[ \hat{q}_{0.99}^{0.99}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6} \]
\[ \hat{q}_{0.99}^{0.99}, 3 \cdot 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6} \]
\[ \hat{q}_{1}^{0.99}, 1.5, 0.1, 10^{-6} \]

This array shows that with large probability our estimator is very close to the m.l.e. This probability is quite high for small values of $n$ and even more for larger values of $n$. This explains why the risks of these two estimators are very close in the first four examples. Note that the value of $\eta$ prevents the empirical quantile from being lower than something of order $10^{-7}$ according to the examples (in Example 4, the value of $10^{-6}$ is due to the way we have built the m.l.e).

4.4. **Speed of the procedure.** For the sake of completeness, we specify below the number of tests that have been calculated in the preceding examples.

| Example | $n = 10$ | $n = 25$ | $n = 50$ | $n = 75$ | $n = 100$ |
|---------|----------|----------|----------|----------|-----------|
| Example 1 | 77 (1.4) | 77 (0.9) | 77 (0.7) | 77 (0.6) | 77 (0.5) |
| Example 2 | 293 (1) | 294 (0.9) | 295 (0.9) | 295 (0.9) | 295 (0.9) |
| Example 3 | 89 (0.75) | 90 (0.5) | 90 (0.5) | 90 (0.5) | 90 (0.5) |
| Example 4 | 100 (3.5) | 100 (0.5) | 100 (0.001) | 100 (0.001) | 100 (0.001) |
| Example 5 | 460 (3) | 461 (1) | 462 (0.6) | 462 (0.4) | 462 (0.3) |
| Example 6 | 687 (0) | 687 (0) | 687 (0) | 687 (0) | 687 (0) |
| Example 7 | 412 (8) | 419 (8) | 425 (8) | 429 (8) | 432 (8) |
| Example 8 | 173209 (10) | 173212 (0) | 173212 (0.9) | 173206 (12) | 173212 (0.3) |

Figure 1. Number of tests computed averaged over $10^6$ samples for examples 1 to 7 and over $10^4$ samples for example 8. The corresponding standard deviations are in brackets.

4.5. **Simulations when $s \not\in \mathcal{F}$.** In Section 4.3, we were in the favourable situation where the true distribution $s$ belonged to the model $\mathcal{F}$, which may not hold true in practice. We now work with random variables $X_1, \ldots, X_n$ simulated according to a density $s \not\in \mathcal{F}$ to illustrate the robustness properties of our estimator.

We begin with an example proposed in Birgé (2006). We generate $X_1, \ldots, X_n$ according to the density

$$ s(x) = 10 \left[ (1 - 2n^{-1})1_{[0,1/10]}(x) + 2n^{-1}1_{[9/10,1]}(x) \right] \quad \text{for all } x \in \mathbb{R} $$
and compare our estimator to the maximum likelihood estimator for the uniform model

\begin{equation}
\mathcal{F} = \{f_\theta, \theta \in [0.01, 10]\} \quad \text{where} \quad f_\theta = \theta^{-1} 1_{[0,\theta]}.
\end{equation}

It is worthwhile to notice that \( h^2(s, \mathcal{F}) = \mathcal{O}(n^{-1}) \), which means that \( s \) is close to \( \mathcal{F} \) when \( n \) is large, and that our estimator still satisfies \( \mathbb{E}[h^2(s, \hat{\theta})] = \mathcal{O}(n^{-1}) \). Contrary to our estimator, the outliers make the m.l.e unstable as shown in the array below.

\begin{tabular}{ |c|c|c|c|c|c| }
\hline
 & \( n = 10 \) & \( n = 25 \) & \( n = 50 \) & \( n = 75 \) & \( n = 100 \) \\
\hline
\( \hat{R}_N(\theta) \) & 0.20 & 0.06 & 0.03 & 0.02 & 0.015 \\
\( \hat{R}_N(\hat{\theta}_{\text{mle}}) \) & 0.57 & 0.56 & 0.56 & 0.56 & 0.57 \\
\hline
\end{tabular}

\textbf{Figur e 2.} Risks for simulated data averaged over \( 10^4 \) samples.

We now propose a second example based on the mixture of two uniform laws. We use the same statistical model \( \mathcal{F} \) but we modify the distribution of the observations. We take \( p \in (0, 1) \) and define the true underlying density by

\[ s_p(x) = (1 - p)f_1(x) + pf_2(x) \quad \text{for all} \quad x \in \mathbb{R}. \]

Set \( p_0 = 1 - 1/\sqrt{2} \). One can check that

\[ H^2(s_p, \mathcal{F}) = \begin{cases} 
H^2(s_p, f_1) & \text{if} \ p \leq p_0, \\
H^2(s_p, f_2) & \text{if} \ p > p_0,
\end{cases} \]

\[ = \begin{cases} 
1 - \sqrt{2 - p}/\sqrt{2} & \text{if} \ p \leq p_0, \\
1 - (\sqrt{2 - p + \sqrt{p}})/2 & \text{if} \ p > p_0,
\end{cases} \]

which means that the best estimator of \( \mathcal{F} \) is \( f_1 \) when \( p < p_0 \) and \( f_2 \) when \( p > p_0 \).

We now compare our estimator \( \hat{\theta} \) to the m.l.e \( \hat{\theta}_{\text{mle}} \). For a lot of values of \( p \), we simulate \( N \) samples of \( n \) random variables with density \( s_p \) and investigate the behaviour of the estimator \( \hat{\theta} \in \{\hat{\theta}, \hat{\theta}_{\text{mle}}\} \) by computing the function

\[ \hat{R}_{p,n,N}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} h^2(s_p, f_{\hat{\theta}(p,i)}) \]

where \( \hat{\theta}(p,i) \) is the value of the estimator \( \hat{\theta} \) corresponding to the \( i \)-th sample whose density is \( s_p \).

We draw below the functions \( p \mapsto \hat{R}_{p,n,N}(\hat{\theta}), p \mapsto \hat{R}_{p,n,N}(\hat{\theta}_{\text{mle}}) \) and \( p \mapsto H^2(s_p, \mathcal{F}) \) for \( n = 100 \) and then for \( n = 10^4 \).
We observe that the m.l.e is rather good when $p \geq p_0$ and very poor when $p < p_0$. This can be explained by the fact that the m.l.e $\hat{\theta}_{mle}$ is close to 2 as soon as the number $n$ of observations is large enough. The shape of the function $p \mapsto \hat{R}_{p,n,5000}(\hat{\theta})$ is quite more satisfying since it looks more like the function $p \mapsto H^2(s_p, \mathcal{F})$. The lower figure suggests that $\hat{R}_{p,n,N}(\hat{\theta})$ converges to $H^2(s_p, \mathcal{F})$ when $n, N$ go to infinity except on a small neighbourhood before $p_0$.

5. Models parametrized by a multidimensional parameter

5.1. Assumption. We now deal with models $\mathcal{F} = \{f_{\theta}, \theta \in \Theta\}$ indexed by a rectangle

$$\Theta = \prod_{j=1}^{d} [m_j, M_j]$$

of $\mathbb{R}^d$ with $d$ larger than 2. Assumption 1 is supposed to be fulfilled all along this section.
5.2. Definition of the test. As previously, our estimation strategy is based on the existence for all \( \theta, \theta' \in \Theta \) of a measurable function \( T(\theta, \theta') \) of the observations possessing suitable statistical properties. The definition of this functional is the natural extension of the one we have proposed in Section 3.2.

Let for \( j \in \{1, \ldots, d\}, t_j \in (0, d^{1/\alpha_j}] \) and \( \epsilon_j = t_j (\ln n)^{-1/\alpha_j} \). We then define the finite sets

\[
\Theta_{\text{dis}} = \left\{(m_1 + k_1 \epsilon_1, \ldots, m_d + k_d \epsilon_d), \forall j \in \{1, \ldots, d\}, \ k_j \leq (M_j - m_j)\epsilon_j^{-1}\right\}
\]

\[
\mathcal{F}_{\text{dis}} = \{f_\theta : \theta \in \Theta_{\text{dis}}\}.
\]

Let \( \pi \) be the map defined on \( \prod_{j=1}^d [m_j, M_j] \) by

\[
\pi(x) = (m_1 + [(x_1 - m_1)/\epsilon_1] \epsilon_1, \ldots, m_d + [(x_d - m_d)/\epsilon_d] \epsilon_d) \quad \text{for all } x = (x_1, \ldots, x_d) \in \prod_{j=1}^d [m_j, M_j]
\]

where \([\cdot] \) is the integer part. We then define \( T(\theta, \theta') \) for all \( \theta, \theta' \in \Theta \) by

\[
T(\theta, \theta') = T(f_\pi(\theta), f_\pi(\theta')) \quad \text{for all } \theta, \theta' \in \Theta = \prod_{j=1}^d [m_j, M_j]
\]

where \( T \) is the functional given by (4).

5.3. Basic ideas. For the sake of simplicity, we first consider the case \( d = 2 \). We shall build a decreasing sequence \( (\Theta_i)_i \) of rectangles by induction. When there exists \( \theta_0 \in \Theta \) such that \( s = f_{\theta_0} \), these rectangles \( \Theta_i \) can be interpreted as confidence sets for \( \theta_0 \). Their construction is strongly inspired from the heuristics of Section 3.1.

We set \( \Theta_1 = \Theta \). Assume that \( \Theta_i = [a_1, b_1] \times [a_2, b_2] \) and let us explain how we can build a confidence set \( \Theta_{i+1} = [a_1, b_1] \times [a'_2, b'_2] \) with \( a'_2, b'_2 \) satisfying \( b'_2 - a'_2 < b_2 - a_2 \).

We begin to build by induction two preliminary finite sequences \( (\theta^{(j)})_{1 \leq j \leq N}, (\theta'^{(j)})_{1 \leq j \leq N} \) of elements of \( \mathbb{R}^2 \). Let \( \theta^{(1)} = (a_1, b_1) \) be the bottom left-hand corner of \( \Theta_i \) and \( \theta'^{(1)} = (a_1, b_2) \) be the top left-hand corner of \( \Theta_i \). Let \( \bar{r}_1(\theta^{(1)}), \bar{r}_2(\theta^{(1)}), \bar{r}_1(\theta'^{(1)}), \bar{r}_2(\theta'^{(1)}) \) be positive numbers such that the rectangles

\[
\mathcal{R}_1 = [a_1, a_1 + \bar{r}_1(\theta^{(1)}), \theta^{(1)}, \theta^{(1)}] \times [a_2, a_2 + \bar{r}_2(\theta^{(1)}), \theta^{(1)}, \theta^{(1)}]
\]

\[
\mathcal{R}'_1 = [a_1, a_1 + \bar{r}_1(\theta'^{(1)}), \theta'^{(1)}, \theta'^{(1)}] \times [b_2, b_2 - \bar{r}_2(\theta'^{(1)}), \theta'^{(1)}, \theta'^{(1)}]
\]

are respectively included in the Hellinger balls

\[
\mathcal{B}(\theta^{(1)}, \kappa^{1/2} h(f_{\theta^{(1)}}, f_{\theta'^{(1)}})) \quad \text{and} \quad \mathcal{B}(\theta'^{(1)}, \kappa^{1/2} h(f_{\theta^{(1)}}, f_{\theta'^{(1)}}))
\]

See (3) for the precise definition of these balls.

We define \( \theta^{(2)}, \theta'^{(2)} \in \mathbb{R}^2 \) as follows

\[
\theta^{(2)} = \begin{cases} 
\theta^{(1)} + (\bar{r}_1(\theta^{(1)}), \theta^{(1)}), 0 
\theta^{(1)} 
\end{cases} \quad \text{if } T(\theta^{(1)}, \theta^{(1)}) \geq 0
\]

\[
\text{otherwise}
\]

\[
\theta'^{(2)} = \begin{cases} 
\theta'^{(1)} + (\bar{r}_1(\theta'^{(1)}), \theta^{(1)}), 0 
\theta^{(1)} 
\end{cases} \quad \text{if } T(\theta^{(1)}, \theta^{(1)}) \leq 0
\]

\[
\text{otherwise}.
\]
Here is an illustration.

\[ \theta^{(1)} = \theta^{(2)} \]

**Figure 4.** Construction of \( \theta^{(2)} \) and \( \theta'^{(2)} \) when \( T(\theta^{(1)}, \theta'^{(1)}) > 0 \).

It is worthwhile to notice that in this figure, the heuristics of Section 3.1 suggest that \( \theta_0 \) belongs to \( \Theta_i \setminus R_1 \).

Now, if the first component of \( \theta^{(2)} = (\theta^{(2)}_1, \theta^{(2)}_2) \) is larger than \( b_1 \), that is \( \theta^{(2)}_1 \geq b_1 \), we set \( N = 1 \) and stop the construction of the vectors \( \theta^{(i)}, \theta'^{(i)} \). Similarly, if \( \theta'^{(2)}_1 \geq b_1 \), we set \( N = 1 \) and stop the construction of the \( \theta^{(i)}, \theta'^{(i)} \).

If \( \theta^{(2)}_1 < b_1 \) and \( \theta'^{(2)}_1 < b_1 \), we consider positive numbers \( \bar{r}_1(\theta^{(2)}, \theta'^{(2)}), \bar{r}_2(\theta^{(2)}, \theta'^{(2)}), \bar{r}_1(\theta'^{(2)}, \theta^{(2)}), \bar{r}_2(\theta'^{(2)}, \theta^{(2)}) \) such that the rectangles

\[
R_2 = [\theta_1^{(2)}, \theta_1^{(2)} + \bar{r}_1(\theta^{(2)}, \theta'^{(2)})] \times [a_2, a_2 + \bar{r}_2(\theta^{(2)}, \theta'^{(2)})]
\]

\[
R'_2 = [\theta_1^{(2)} + \bar{r}_1(\theta'^{(2)}, \theta^{(2)}), \theta^{(2)}] \times [b_2 - \bar{r}_2(\theta'^{(2)}, \theta^{(2)}), b_2]
\]

are respectively included in the Hellinger balls

\[
\mathcal{B}(\theta^{(2)}, \kappa^{1/2} h(f_{\theta^{(2)}}, f_{\theta'^{(2)}})) \quad \text{and} \quad \mathcal{B}(\theta'^{(2)}, \kappa^{1/2} h(f_{\theta^{(2)}}, f_{\theta'^{(2)}})).
\]

We then define \( \theta^{(3)}, \theta'^{(3)} \in \mathbb{R}^2 \) by

\[
\theta^{(3)} = \begin{cases} 
\theta^{(2)} + (\bar{r}_1(\theta^{(2)}, \theta'^{(2)}), 0) & \text{if } T(\theta^{(2)}, \theta'^{(2)}) \geq 0 \\
\theta^{(2)} & \text{otherwise}
\end{cases}
\]

\[
\theta'^{(3)} = \begin{cases} 
\theta'^{(2)} + (\bar{r}_1(\theta'^{(2)}, \theta^{(2)}), 0) & \text{if } T(\theta^{(2)}, \theta'^{(2)}) \leq 0 \\
\theta'^{(2)} & \text{otherwise}.
\end{cases}
\]

If \( \theta^{(3)}_1 \geq b_1 \) or if \( \theta'^{(3)}_1 \geq b_1 \) we stop the construction and set \( N = 2 \). In the contrary case, we repeat this step to build the vectors \( \theta^{(4)} \) and \( \theta'^{(4)} \).
We repeat these steps until the construction stops. Let \( N \) be the integer for which \( \theta_1^{(N+1)} \geq b_1 \) or \( \theta_1^{(N+1)} \geq b_1 \). We then define

\[
\begin{align*}
    a_2' &= \begin{cases} 
        a_2 + \min_{1 \leq j \leq N} \bar{r}_2(\theta^{(j)}, \theta'^{(j)}) & \text{if } \theta_1^{(N+1)} \geq b_1 \\
        a_2 & \text{otherwise}
    \end{cases} \\
    b_2' &= \begin{cases} 
        b_2 - \min_{1 \leq j \leq N} \bar{r}_2(\theta'^{(j)}, \theta^{(j)}) & \text{if } \theta_1^{(N+1)} \geq b_1 \\
        b_2 & \text{otherwise}
    \end{cases}
\end{align*}
\]

and set \( \Theta_{i+1} = [a_1, b_1] \times [a_2', b_2'] \).

In this figure, the set

\[
\Theta_i \setminus \left( R_1 \cup R_2 \cup R'_3 \cup R_4 \cup R_5 \right)
\]

is a confidence set for \( \theta_0 \). The set \( \Theta_{i+1} \) is the smallest rectangle containing this confidence set.

**Remark 1.** We define \( \Theta_{i+1} \) as a rectangle to make the procedure easier to implement.

**Remark 2.** By using a similar strategy, we can also build a confidence set \( \Theta_{i+1} \) of the form \( \Theta_{i+1} = [a_1', b_1'] \times [a_2', b_2'] \) where \( a_1', b_1' \) are such that \( b_1' - a_1' < b_1 - a_1 \).

We shall build the rectangles \( \Theta_i \) until their diameters become sufficiently small. The estimator we shall consider will be the center of the last rectangle built.

**5.4. Procedure.** In the general case, that is when \( d \geq 2 \), we build a finite sequence of rectangles \( (\Theta_i)_i \) of \( \Theta = \prod_{j=1}^d [m_j, M_j] \). We consider \( \kappa > 0 \) and for all rectangle \( C = \prod_{j=1}^d [a_j, b_j] \subset \Theta \), vectors \( \theta, \theta' \in C \), and integers \( j \in \{1, \ldots, d\} \), we introduce positive numbers \( \bar{r}_{C,j}(\theta, \theta') \), \( \bar{r}_{\cdot,j}(\theta, \theta') \) such that

\[
C \cap \prod_{j=1}^d [\theta_j - \bar{r}_{C,j}(\theta, \theta'), \theta_j + \bar{r}_{C,j}(\theta, \theta')] \subset B(\theta, \kappa^{1/2} h(f_{\theta}, f_{\theta'})).
\]
We also consider for all \( j \in \{1, \ldots, d\} \), \( R_{\cdot, j} \geq R_j \) such that
\[
20 \quad h^2(f_\theta, f_{\theta'}) \geq \sup_{1 \leq j \leq d} R_{\cdot, j} |\theta_j - \theta'_j|^{\alpha_j} \quad \text{for all } \theta, \theta' \in \mathcal{C}.
\]

We finally consider for all \( j \in \{1, \ldots, d\} \), an one-to-one map \( \psi_j \) from \( \{1, \ldots, d-1\} \) into \( \{1, \ldots, d\} \setminus \{j\} \).

We set \( \Theta_1 = \Theta \). Given \( \Theta_i \), we define \( \Theta_{i+1} \) by using the following algorithm.

**Algorithm 2** Definition of \( \Theta_{i+1} \) from \( \Theta_i 

**Require:** \( \Theta_i = \prod_{j=1}^d [a_j, b_j] \)

1. Choose \( k \in \{1, \ldots, d\} \) such that
\[
21 \quad R_{\Theta_i,k}(b_k - a_k)^{\alpha_k} = \max_{1 \leq j \leq d} R_{\Theta_i,j}(b_j - a_j)^{\alpha_j}
\]

2. \( \theta = (\theta_1, \ldots, \theta_d) \leftarrow (a_1, \ldots, a_d) \), \( \theta' = (\theta'_1, \ldots, \theta'_d) \leftarrow \theta \) and \( \theta'_{\cdot} \leftarrow b_k \)

3. \( \varepsilon_j \leftarrow \rho_{\Theta_i,j}(\theta, \theta') \) and \( \varepsilon'_j \leftarrow \rho_{\Theta_i,j}(\theta', \theta) \) for all \( j \neq k \)

4. \( \varepsilon_k \leftarrow (b_k - a_k)/2 \) and \( \varepsilon'_k \leftarrow (b_k - a_k)/2 \)

5. repeat
   6. Test \( \leftarrow T(\theta, \theta') \)
   7. For all \( j \), \( \bar{\tau}_j \leftarrow \rho_{\Theta_i,j}(\theta, \theta') \), \( \bar{\tau}'_j \leftarrow \rho_{\Theta_i,j}(\theta', \theta) \), \( \tau_j' \leftarrow \bar{\tau}_{\Theta_i,j}(\theta', \theta) \)
   8. if \( \text{Test} \geq 0 \)
      9. \( \varepsilon_{\psi_k(1)} \leftarrow \bar{\rho}_{\psi_k(1)} \)
    10. \( \varepsilon_{\psi_k(j)} \leftarrow \min(\varepsilon_{\psi_k(j)}, \bar{\rho}_{\psi_k(j)}) \) for all \( j \in \{2, \ldots, d-1\} \)
    11. \( \varepsilon_k \leftarrow \min(\varepsilon_k, \bar{\rho}_k) \)
    12. \( J \leftarrow \{1 \leq j \leq d-1, \theta_{\psi_k(j)} + \varepsilon_{\psi_k(j)} < b_{\psi_k(j)}\} \)
    13. if \( J \neq \emptyset \)
        14. \( j_{\text{min}} \leftarrow \min J \)
        15. \( \theta_{\psi_k(j)} \leftarrow a_{\psi_k(j)} \) for all \( j \leq j_{\text{min}} - 1 \)
        16. \( \theta_{\psi_k(j_{\text{min}})} \leftarrow \theta_{\psi_k(j_{\text{min}})} + \varepsilon_{\psi_k(j_{\text{min}})} \)
    17. else
        18. \( j_{\text{min}} \leftarrow d \)
        19. end if
    20. end if
   21. if \( \text{Test} \leq 0 \)
      22. \( \varepsilon_{\psi_k(1)}^' \leftarrow \bar{\rho}_{\psi_k(1)}^' \)
      23. \( \varepsilon_{\psi_k(j)}^' \leftarrow \min(\varepsilon_{\psi_k(j)}^', \bar{\rho}_{\psi_k(j)}^') \) for all \( j \in \{2, \ldots, d-1\} \)
      24. \( \varepsilon_k^' \leftarrow \min(\varepsilon_k^', \bar{\rho}_k^') \)
      25. \( J' \leftarrow \{1 \leq j' \leq d-1, \theta_{\psi_k(j')} + \varepsilon_{\psi_k(j')} < b_{\psi_k(j')}\} \)
      26. if \( J' \neq \emptyset \)
          27. \( j_{\text{min}}' \leftarrow \min J' \)
          28. \( \theta_{\psi_k(j)}' \leftarrow a_{\psi_k(j)} \) for all \( j \leq j_{\text{min}}' - 1 \)
          29. \( \theta_{\psi_k(j_{\text{min}}')}' \leftarrow \theta_{\psi_k(j_{\text{min}}')} + \varepsilon_{\psi_k(j_{\text{min}}')}' \)
      30. else
          31. \( j_{\text{min}}' \leftarrow d \)
We now consider $d$ positive numbers $\eta_1, \ldots, \eta_d$ and use the algorithm below to build our estimator $\hat{\theta}$.

**Algorithm 3**

1. Set $a_j = m_j$ and $b_j = M_j$ for all $j \in \{1, \ldots, d\}$
2. $i \leftarrow 0$
3. while There exists $j \in \{1, \ldots, d\}$ such that $b_j - a_j > \eta_j$ do
4. $i \leftarrow i + 1$
5. Build $\Theta_i$ and set $a_1, \ldots, a_d, b_1, \ldots, b_d$ such that $\prod_{j=1}^{d} [a_j, b_j] = \Theta_i$
6. end while
7. Return: $\hat{\theta} = \left( \frac{a_1 + b_1}{2}, \ldots, \frac{a_d + b_d}{2} \right)$

The parameters $\kappa, t_j, \eta_j, \overline{\kappa}_j(\theta, \theta'), \mathcal{L}_{\overline{\kappa}_j}(\theta, \theta')$ can be interpreted as in dimension 1. We have introduced a new parameter $R_{\kappa, j}$ whose role is to control more accurately the Hellinger distance in order to make the procedure faster. Sometimes, the computation of this parameter is difficult in practice, in which case we can avoid it by proceeding as follows. For all $\theta, \theta' \in \Theta$,

$$h^2(f_\theta, f_{\theta'}) \geq \sup_{1 \leq j \leq d} \frac{R}{R_j} |\theta_j - \theta'_j|^{\alpha_j}$$

where $R = \min_{1 \leq j \leq d} R_j$, which means that we can always assume that $R_j$ is independent of $j$. Choosing $R_{\kappa, j} = R$ simplifies the only line where this parameter is involved (line 1 of Algorithm 2). It becomes

$$(b_k - a_k)^{\alpha_k} = \max_{1 \leq j \leq d} (b_j - a_j)^{\alpha_j}$$

and $k$ can be calculated without computing $R$.

**5.5. Risk bound.** Suitable values of the parameters lead to a risk bound for our estimator $\hat{\theta}$.

**Theorem 5.** Suppose that Assumption 1 holds. Let $\bar{\kappa}$ be defined by (7), and assume that $\kappa \in (0, \bar{\kappa})$, and for all $j \in \{1, \ldots, d\}$, $t_j \in (0, d^{1/\alpha_j}]$,

$$e_j = t_j (R_j n)^{-1/\alpha_j}, \quad \eta_j \in [e_j, d^{1/\alpha_j} (R_j n)^{-1/\alpha_j}]$$
Suppose that for all rectangle $C$, $\theta, \theta' \in C$, the numbers $\bar{r}_{C,j}(\theta, \theta')$, $\underline{r}_{C,j}(\theta, \theta')$, are such that (15) holds.

Then, for all $\xi > 0$, the estimator $\hat{\theta}$ built by Algorithm 3 satisfies

$$P \left[ Ch^2(s, f_{\hat{\theta}}) \geq h^2(s, \theta) + \frac{d}{n} + \xi \right] \leq e^{-n\xi}$$

where $C > 0$ depends only on $\kappa$, $(R_j/R_j)_{1 \leq j \leq d}$, $(\alpha_j)_{1 \leq j \leq d}$, $(t_j)_{1 \leq j \leq d}$.

Remark. A look at the proof of the theorem shows that Theorem 1 ensues from this theorem when $t_j = d^{1/\alpha_j}$ and $\eta_j = \epsilon_j$.

5.6. Choice of $\bar{r}_{C,j}(\theta, \theta')$ and $\underline{r}_{C,j}(\theta, \theta')$. The parameters $\bar{r}_{C,j}(\theta, \theta')$, $\underline{r}_{C,j}(\theta, \theta')$ are involved in the procedure and must be calculated. They may be chosen arbitrary provided that the rectangle

$$C \cap \prod_{j=1}^{d} [\theta_j - \underline{r}_{C,j}(\theta, \theta'), \theta_j + \bar{r}_{C,j}(\theta, \theta')]$$

is included in the Hellinger ball $B(\theta, \kappa^{1/2} h(f_{\theta}, f_{\theta'}))$. Indeed, the theoretical properties of the estimator given by the preceding theorem does not depend on these values.

However, the numerical complexity of the algorithm strongly depends on these parameters. The algorithm computes less tests when $\bar{r}_{C,j}(\theta, \theta')$, $\underline{r}_{C,j}(\theta, \theta')$ are large and we have thus an interest in defining them as the largest numbers possible. In the cases where a direct computation of these numbers is difficult, we may use a similar strategy that the one adopted in the unidimensional case (Section 3.5).

First way. We may consider $(\overline{R}_{C,1}, \ldots, \overline{R}_{C,d}) \in \prod_{j=1}^{d} (0, \overline{R}_j)$ such that

$$h^2(f_{\theta}, f_{\theta'}) \leq \sup_{1 \leq j \leq d} \overline{R}_{C,j} |\theta_j - \theta'_j|^{\alpha_j} \quad \text{for all } \theta, \theta' \in C$$

(17)

and define them by

$$\bar{r}_{C,j}(\theta, \theta') = \underline{r}_{C,j}(\theta, \theta') = \left( (\kappa/\overline{R}_{C,j}) h^2(f_{\theta}, f_{\theta'}) \right)^{1/\alpha_j}.$$ 

(18)

One can verify that this definition implies (15).

Second way. An alternative definition that does not involve the Hellinger distance is

$$\bar{r}_{C,j}(\theta, \theta') = \underline{r}_{C,j}(\theta, \theta') = \left( \kappa/\overline{R}_{C,j} \sup_{1 \leq k \leq d} \overline{R}_{C,k} |\theta'_k - \theta_k|^{\alpha_k} \right)^{1/\alpha_j}.$$ 

(19)

Similarly, one can check that (15) holds.

The complexity of our procedure can be upper-bounded as soon as $\bar{r}_{C,j}(\theta, \theta')$ and $\underline{r}_{C,j}(\theta, \theta')$ are large enough.
Proposition 6. Suppose that the assumptions of Theorem 5 are fulfilled and that for all \( j \in \{1, \ldots, d\} \), all rectangle \( C, \theta, \theta' \in C \), the numbers \( \overline{R}_{C,j}(\theta, \theta') \), \( \underline{R}_{C,j}(\theta, \theta') \) are larger than

\[
(20) \quad \left( \frac{\kappa / \overline{R}_{C,j}}{\sup_{1 \leq k \leq d} R_{C,j}^*|\theta' - \theta|^{\alpha_k}} \right)^{1/\alpha_j}
\]

where the \( \overline{R}_{C,j} \) and \( \underline{R}_{C,j} \) are respectively such that (16) and (17) hold and such that \( \overline{R}_{C,j} \geq \overline{R}_j \) and \( \underline{R}_{C,j} \leq \underline{R}_j \).

Then, the number of tests computed to build the estimator \( \hat{\theta} \) is smaller than

\[
4 \prod_{j=1}^{d} \left( 1 + (R_j/(\kappa \overline{R}_j))^{1/\alpha_j} \right) \left[ \sum_{j=1}^{d} \max \left\{ 1, \log \left( \frac{M_j - m_j}{\eta_j} \right) \right\} \right].
\]

6. Simulations for multidimensional models

In this section, we complete the simulation study of Section 4 by dealing with multidimensional models.

6.1. Models. We propose to work with the following models.

Example 1. \( \mathcal{F} = \{ f(m,\sigma), (m, \sigma) \in [-5, 5] \times [1/5, 5] \} \) where

\[
f(m, \sigma)(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x - m)^2}{2\sigma^2} \right) \quad \text{for all } x \in \mathbb{R}.
\]

Example 2. \( \mathcal{F} = \{ f(m,\sigma), (m, \sigma) \in [-5, 5] \times [1/5, 5] \} \) where

\[
f(m, \sigma)(x) = \frac{\sigma}{\pi ((x - m)^2 + \sigma^2)} \quad \text{for all } x \in \mathbb{R}.
\]

Example 3. \( \mathcal{F} = \{ f(a,b), (a, b) \in [0.6, 10] \times [0.1, 20] \} \) where

\[
f(a,b)(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} 1_{[0, +\infty)}(x) \quad \text{for all } x \in \mathbb{R}
\]

where \( \Gamma \) is the Gamma function.

Example 4. \( \mathcal{F} = \{ f(a,b), (a, b) \in [0.7, 20] \times [0.7, 20] \} \) where

\[
f(a,b)(x) = \frac{1}{B(a,b)} x^{a-1} (1 - x)^{b-1} 1_{[0,1]}(x) \quad \text{for all } x \in \mathbb{R}.
\]

and where \( B(a,b) \) is the Beta function.

Example 5. \( \mathcal{F} = \{ f(m,\lambda), (m, \lambda) \in [-1, 1] \times [1/5, 5] \} \) where

\[
f(m, \lambda)(x) = \lambda e^{-\lambda(x-m)} 1_{[m, +\infty)}(x) \quad \text{for all } x \in \mathbb{R}.
\]

Example 6. \( \mathcal{F} = \{ f(m,r), (m, r) \in [-0.5, 0.5] \times [0.1, 2] \} \) where

\[
f(m,r)(x) = r^{-1} 1_{[m,m+r]}(x) \quad \text{for all } x \in \mathbb{R}.
\]
We shall use our procedure with \( t_j = 0 \) for all \( j \in \{1, \ldots, d\} \) (that is \( \Theta_{\text{dis}} = \Theta, \; \mathcal{F}_{\text{dis}} = \mathcal{F} \) and \( T(\theta, \theta') = T(f_\theta, f_{\theta'}) \)) and with \( \kappa = 0.9\hat{\kappa} \), \( \eta_j = (M_j - m_j)10^{-6} \). In order to avoid technicalities, we delay to Section 8 the values of \( \hat{R}_{\text{c},j}(\theta, \theta'), \; \mathcal{L}_{\text{c},j}(\theta, \theta') \) that have been chosen in this simulation study.

6.2. Simulations when \( s \in \mathcal{F} \). We simulate \( N = 10^4 \) independent samples \((X_1, \ldots, X_n)\) according to a density \( s \in \mathcal{F} \) and use our procedure to estimate \( s \) on each of the samples. In examples 1, 2, 5, 6 the density is \( s = f_{(0,1)} \), in Example 3, \( s = f_{(2,3)} \) and in Example 4, \( s = f_{(3,4)} \). The results are the following.

| Example | \( R_{10^4}(\theta) \) | \( \hat{R}_{10^4}(\hat{\theta}_{\text{mle}}) \) | \( \hat{R}_{10^4, \text{rel}}(\hat{\theta}_{\text{mle}}) \) | \( \text{std}_{10^4}(\theta) \) | \( \text{std}_{10^4}(\hat{\theta}_{\text{mle}}) \) |
|---------|----------------|----------------|----------------|----------------|----------------|
| Example 1 | 0.011 | 0.0052 | 0.0034 | 0.0025 |
| & 10^{-4} | 6 \cdot 10^{-5} | -5 \cdot 10^{-8} | 3 \cdot 10^{-8} |
| Example 2 | 0.011 | 0.0052 | 0.0034 | 0.0026 |
| & 10^{-8} | 10^{-8} | -10^{-9} | 4 \cdot 10^{-8} |
| Example 3 | 0.011 | 0.0052 | 0.0034 | 0.0026 |
| & 2 \cdot 10^{-4} | 2 \cdot 10^{-5} | 10^{-7} | 10^{-7} |
| Example 4 | 0.011 | 0.0052 | 0.0034 | 0.0026 |
| & 2 \cdot 10^{-4} | 10^{-5} | 2 \cdot 10^{-7} | 2 \cdot 10^{-7} |
| Example 5 | 0.025 | 0.012 | 0.0082 | 0.0063 |
| & 0.025 | 0.012 | 0.0083 | 0.0063 |
| | 0.020 | -0.0020 | -0.0073 | 0.0012 |
| Example 6 | 0.040 | 0.019 | 0.013 | 0.0098 |
| & 0.039 | 0.020 | 0.013 | 0.010 |
| | 0.010 | -0.015 | -0.018 | -0.016 |
| | 0.033 | 0.016 | 0.011 | 0.0080 |
| | 0.027 | 0.014 | 0.0093 | 0.0069 |
The risk of our estimator is very close to the one of the m.l.e. In the first four examples they are even almost indistinguishable. As in dimension 1, this can be explained by the fact that the first four models are regular enough to ensure that our estimator is very close to the maximum likelihood one.

To see this, let for $c \in \{0.99, 0.999, 1\}$, $q_c$ be the $c$-quantile of the random variable

$$\max \left\{ |\hat{\theta}_1 - \tilde{\theta}_{\text{mle},1}|, |\hat{\theta}_2 - \tilde{\theta}_{\text{mle},2}| \right\}$$

and $\hat{q}_c$ be the empirical version based on $10^4$ samples. These empirical quantiles are very small as shown in the array below.

| Example 1 | $n = 25$ | $n = 50$ | $n = 75$ | $n = 100$ |
|-----------|----------|----------|----------|-----------|
| $q_{0.99}$ | $9 \times 10^{-7}$ | $9 \times 10^{-7}$ | $9 \times 10^{-7}$ | $9 \times 10^{-7}$ |
| $q_{0.999}$ | $0.023$ | $10^{-6}$ | $9 \times 10^{-7}$ | $10^{-6}$ |
| $\hat{q}_1$ | $0.22$ | $0.072$ | $10^{-6}$ | $10^{-6}$ |

| Example 2 | $n = 25$ | $n = 50$ | $n = 75$ | $n = 100$ |
|-----------|----------|----------|----------|-----------|
| $q_{0.99}$ | $4 \times 10^{-7}$ | $4 \times 10^{-7}$ | $4 \times 10^{-7}$ | $4 \times 10^{-7}$ |
| $q_{0.999}$ | $5 \times 10^{-7}$ | $5 \times 10^{-7}$ | $5 \times 10^{-7}$ | $5 \times 10^{-7}$ |
| $\hat{q}_1$ | $5 \times 10^{-7}$ | $5 \times 10^{-7}$ | $5 \times 10^{-7}$ | $5 \times 10^{-7}$ |

| Example 3 | $n = 25$ | $n = 50$ | $n = 75$ | $n = 100$ |
|-----------|----------|----------|----------|-----------|
| $q_{0.99}$ | $7 \times 10^{-7}$ | $7 \times 10^{-7}$ | $7 \times 10^{-7}$ | $7 \times 10^{-7}$ |
| $q_{0.999}$ | $9 \times 10^{-7}$ | $8 \times 10^{-7}$ | $8 \times 10^{-7}$ | $8 \times 10^{-7}$ |
| $\hat{q}_1$ | $1.5$ | $0.29$ | $10^{-6}$ | $9 \times 10^{-7}$ |

| Example 4 | $n = 25$ | $n = 50$ | $n = 75$ | $n = 100$ |
|-----------|----------|----------|----------|-----------|
| $q_{0.99}$ | $10^{-6}$ | $10^{-6}$ | $10^{-6}$ | $10^{-6}$ |
| $q_{0.999}$ | $2 \times 10^{-6}$ | $10^{-6}$ | $10^{-6}$ | $10^{-6}$ |
| $\hat{q}_1$ | $1.6$ | $0.27$ | $2 \times 10^{-6}$ | $2 \times 10^{-6}$ |

6.3. **Simulations when $s \not\in \mathcal{F}$**. Contrary to the maximum likelihood estimator, our estimator possesses robustness properties. The goal of this section is to illustrate them.

Suppose that we observe $n = 100$ i.i.d random variables $X_1, \ldots, X_n$ from which we wish to estimate their distribution by using a Gaussian model

$$\mathcal{F} = \{ f_{(m,\sigma)}, (m, \sigma) \in [-10, 10] \times [0.5, 10] \}$$

where $f_{(m,\sigma)}$ is the density of a Gaussian random variable with mean $m$ and variance $\sigma^2$. The preceding section shows that when the unknown underlying density $s$ belongs to $\mathcal{F}$, our estimator is as good as the m.l.e. We now consider $p \in [0, 1]$ and define $s = s_p$ where

$$s_p(x) = (1 - p)f_{(-5,1)} + pf_{(5,1)} \quad \text{for all } x \in \mathbb{R}.$$ 

This density belongs to the model only if $p = 0$ or $p = 1$ and we are interested in comparing our estimator to the m.l.e when $p \neq 0$ and $p \neq 1$. 
We then proceed as in Section 4.5. For a lot of values of \( p \in [0, 1] \), we simulate \( N = 1000 \) samples of 100 random variables with density \( s_p \) and measure the quality of the estimator \( \tilde{\theta} \) by

\[
\widehat{R}_{p,N}(\tilde{\theta}) = \frac{1}{1000} \sum_{i=1}^{1000} h^2(s_p, f_{\tilde{\theta}(p,i)})
\]

where \( \tilde{\theta}(p,i) \) is the value of \( \tilde{\theta} \) corresponding to the \( i \)th sample whose density is \( s_p \). We compute this function for \( \tilde{\theta} \in \{ \tilde{\theta}_{mle}, \hat{\theta} \} \) and obtain the graph below.

![Graph](image.png)

**Figure 6.** Red: \( p \mapsto H^2(s_p, \mathcal{F}) \). Blue: \( p \mapsto \widehat{R}_{p,1000}(\tilde{\theta}) \). Green: \( p \mapsto \widehat{R}_{p,1000}(\tilde{\theta}_{mle}) \).

This figure shows that the risk of our estimator is smaller than the one of the m.l.e when \( p \) is close to 0 or 1 (says \( p \leq 0.2 \) or \( p \geq 0.8 \)) and is similar otherwise. For the Gaussian model, our estimator may thus be interpreted as a robust version of the m.l.e.

7. Proofs

7.1. A preliminary result. In this section, we show a result that will allow us to prove Theorems 3 and 5. Given \( \Theta' \subset \Theta \), we recall that in this paper that \( \text{diam}\Theta' \) stands for

\[
\text{diam}\Theta' = \sup_{\theta, \theta' \in \Theta'} h^2(f_\theta, f_{\theta'}).
\]

**Theorem 7.** Suppose that Assumption 1 holds. Let \( \kappa \in (0, \bar{\kappa}) \), \( N \in \mathbb{N}^* \) and let \( \Theta_1 \ldots \Theta_N \) be \( N \) non-empty subsets of \( \Theta \) such that \( \Theta_1 = \Theta \). For all \( j \in \{1, \ldots, d\} \), let \( t_j \) be an arbitrary number of \( (0, d^{1/\alpha_j}] \) and

\[
\epsilon_j = t_j (\overline{R}_j n)^{-1/\alpha_j}.
\]

Assume that for all \( i \in \{1, \ldots, N-1\} \), there exists \( L_i \geq 1 \) such that for all \( \ell \in \{1, \ldots, L_i\} \) there exist two elements \( \theta^{(i,\ell)} \neq \theta^{(i,\ell)} \) of \( \Theta_i \) such that

\[
\Theta_i \setminus \bigcup_{\ell=1}^{L_i} B^{(i,\ell)} \subset \Theta_{i+1} \subset \Theta_i
\]
where \(B^{(i,\ell)}\) is the set defined by
\[
B^{(i,\ell)} = \begin{cases} 
A^{(i,\ell)} & \text{if } T(\theta^{(i,\ell)}, \theta'(i,\ell)) > 0 \\
A'(i,\ell) & \text{if } T(\theta^{(i,\ell)}, \theta'(i,\ell)) < 0 \\
A^{(i,\ell)} \cup A'(i,\ell) & \text{if } T(\theta^{(i,\ell)}, \theta'(i,\ell)) = 0
\end{cases}
\]
where \(A^{(i,\ell)}\) and \(A'(i,\ell)\) are the Hellinger balls defined by
\[
A^{(i,\ell)} = \{ \theta'' \in \Theta_i, h^2(f_{\theta''}, f_{\theta(i,\ell)}) \leq \kappa h^2(f_{\theta(i,\ell)}, f_{\theta'(i,\ell)}) \}
\]
and where \(T\) is the functional defined by (14). Suppose moreover that there exists \(\kappa_0 > 0\) such that
\[
\kappa_0 \text{diam}(\Theta_i) \leq \inf_{1 \leq i \leq L, \kappa} h^2(f_{\theta(i,\ell)}, f_{\theta'(i,\ell)}) \text{ for all } i \in \{1, \ldots, N - 1\}.
\]
Then, for all \(\xi > 0\),
\[
\mathbb{P} \left[ C \inf_{\theta \in \Theta_N} h^2(s, f_{\theta}) \geq h^2(s, \mathcal{F}) + \frac{D\bar{\xi}}{n} + \xi \right] \leq e^{-n\xi}
\]
where \(C > 0\) depends only on \(\kappa, \kappa_0\), where
\[
D\bar{\xi} = \max \left\{ d, \sum_{j=1}^{d} \log \left( 1 + t_j^{-1} (\frac{\alpha}{cR_j/R_{\gamma}})^{1/\alpha_j} \right) \right\}
\]
and where \(c\) depends only on \(\kappa\).

This result says that if \((\Theta_i)_{1 \leq i \leq N}\) is a finite sequence of subsets of \(\Theta\) satisfying the assumptions of the theorem, then there exists an estimator \(\hat{\theta}\) with values in \(\Theta_N\) whose risk can be upper bounded by
\[
\mathbb{P} \left[ Ch^2(s, f_{\hat{\theta}}) \geq h^2(s, \mathcal{F}) + \frac{D\bar{\xi}}{n} + \xi \right] \leq e^{-n\xi}.
\]
We shall show that algorithms 1 and 3 correspond to suitable choices of sets \(\Theta_i\).

### 7.2. Proof of Theorem 7

Let \(\theta_0 \in \Theta\) be such that
\[
h^2(s, f_{\theta_0}) \leq h^2(s, \mathcal{F}) + 1/n.
\]
Define \(C_\kappa\) such that \((1 + \sqrt{C_\kappa})^2 = \kappa^{-1}\) and \(\varepsilon \in (1/\sqrt{2}, 1)\) such that
\[
\left( 1 + \min \left( \frac{1-\varepsilon}{\sqrt{2}}, \varepsilon - \frac{1}{\sqrt{2}} \right) \right)^4 (1 + \varepsilon) + \min \left( \frac{1-\varepsilon}{\sqrt{2}}, \varepsilon - \frac{1}{\sqrt{2}} \right) = C_\kappa.
\]
We then set
\[
\beta = \min \left\{ (1 - \varepsilon)/2, \varepsilon - 1/\sqrt{2} \right\}
\]
\[
\gamma = (1 + \beta)(1 + \beta^{-1}) [1 - \varepsilon + (1 + \beta)(1 + \varepsilon)]
\]
\[
c = 24(2 + \sqrt{2}/6 (\varepsilon - 1/\sqrt{2}))/((\varepsilon - 1/\sqrt{2})^2 \cdot 10^3
\]
\[
\delta = (1 + \beta^{-1}) \left[ 1 - \varepsilon + (1 + \beta^3 (1 + \varepsilon) \right] + c(1 + \beta)^2.
\]
The proof of the theorem is based on the lemma below whose proof is delayed to Section 7.2.1.
Lemma 1. For all $\xi > 0$, there exists an event $\Omega_\xi$ such that $\mathbb{P}(\Omega_\xi) \geq 1 - e^{-n\xi}$ and on which the following assertion holds: if there exists $p \in \{1, \ldots, N - 1\}$ such that $\theta_0 \in \Theta_p$ and such that

$$\gamma h^2(s, f_{\theta_0}) + \delta \left( \frac{D_{\mathcal{F}}}{n} + \xi \right) < \beta \inf_{\ell \in \{1, \ldots, L_p\}} \left( h^2(f_{\theta_0}, f_{\theta(p,\ell)}) + h^2(f_{\theta_0}, f_{\theta(p,\ell)}) \right)$$

then $\theta_0 \in \Theta_{p+1}$.

The result of Theorem 7 is straightforward if $\theta_0 \in \Theta_N$, and we shall thus assume that $\theta_0 \not\in \Theta_N$. Set

$$p = \max \{ i \in \{1, \ldots, N - 1\}, \theta_0 \in \Theta_i \}.$$ 

Let $\theta'_0$ be any element of $\Theta_N$. Then, $\theta'_0$ belongs to $\Theta_p$ and

$$h^2(f_{\theta_0}, f_{\theta'_0}) \leq \sup_{\theta, \theta' \in \Theta_p} h^2(f_{\theta}, f_{\theta'}) \leq \kappa_0^{-1} \inf_{\ell \in \{1, \ldots, L_p\}} h^2(f_{\theta(p,\ell)}, f_{\theta(p,\ell)}) \leq 2\kappa_0^{-1} \inf_{\ell \in \{1, \ldots, L_p\}} \left( h^2(f_{\theta_0}, f_{\theta(p,\ell)}) + h^2(f_{\theta_0}, f_{\theta(p,\ell)}) \right).$$

By definition of $p$, $\theta_0 \in \Theta_p \setminus \Theta_{p+1}$. We then derive from the above lemma that on $\Omega_\xi$,

$$\beta \inf_{\ell \in \{1, \ldots, L_p\}} \left( h^2(f_{\theta_0}, f_{\theta(p,\ell)}) + h^2(f_{\theta_0}, f_{\theta(p,\ell)}) \right) \leq \gamma h^2(s, f_{\theta_0}) + \delta \frac{D_{\mathcal{F}} + n\xi}{n}.$$

Hence,

$$h^2(f_{\theta_0}, f_{\theta'_0}) \leq \frac{2}{\beta \kappa_0} \left( \gamma h^2(s, f_{\theta_0}) + \delta \frac{D_{\mathcal{F}} + n\xi}{n} \right)$$

and thus

$$h^2(s, f_{\theta_0}) \leq 2h^2(s, f_{\theta_0}) + 2h^2(f_{\theta_0}, f_{\theta'_0}) \leq \left( 2 + \frac{4\gamma}{\beta \kappa_0} \right) h^2(s, f_{\theta_0}) + \frac{4\delta}{n\beta \kappa_0} (D_{\mathcal{F}} + n\xi).$$

Since $h^2(s, f_{\theta_0}) \leq h^2(s, \mathcal{F}) + 1/n$, there exists $C > 0$ such that

$$Ch^2(s, f_{\theta'_0}) \leq h^2(s, \mathcal{F}) + \frac{D_{\mathcal{F}}}{n} + \xi \quad \text{on} \quad \Omega_\xi.$$

This concludes the proof.

7.2.1. Proof of Lemma 1 We use the claim below whose proof is postponed to Section 7.2.2

Claim 1. For all $\xi > 0$, there exists an event $\Omega_\xi$ such that $\mathbb{P}(\Omega_\xi) \geq 1 - e^{-n\xi}$ and on which, for all $f, f' \in \mathcal{F}_{\text{dis}}$,

$$(1 - \varepsilon) h^2(s, f') + \frac{\overline{T}(f, f')}{\sqrt{2}} \leq (1 + \varepsilon) h^2(s, f) + c \left( \frac{D_{\mathcal{F}} + n\xi}{n} \right)$$

(see Section 5.2 for the definition of $\mathcal{F}_{\text{dis}}$).
Consequently, by using the triangular inequality and the above inequality

\[
(1 - \varepsilon) h^2(s, f_{\pi(\theta'(p, \ell))}) \leq (1 + \varepsilon) h^2(s, f_{\pi(\theta(p, \ell))}) + c \frac{(D_{\delta} + n\xi)}{n}.
\]

Consequently, by using the triangular inequality and the above inequality

\[
(1 - \varepsilon) h^2(f_{\theta_0}, f_{\pi(\theta'(p, \ell))}) \leq (1 + \beta^{-1})(1 - \varepsilon) h^2(s, f_{\theta_0}) + (1 + \beta) h^2(f_{\theta_0}, f_{\pi(\theta(p, \ell))}) + (1 + \beta) \left[ (1 + \varepsilon) h^2(s, f_{\theta_0}) + c \frac{(D_{\delta} + n\xi)}{n} \right].
\]

Since\( h^2(s, f_{\pi(\theta'(p, \ell))}) \leq (1 + \beta^{-1}) h^2(s, f_{\theta_0}) + (1 + \beta) h^2(f_{\theta_0}, f_{\pi(\theta(p, \ell))}) \),

\[
(1 - \varepsilon) h^2(f_{\theta_0}, f_{\pi(\theta'(p, \ell))}) \leq (1 + \beta^{-1}) [1 - \varepsilon + (1 + \beta)(1 + \varepsilon)] h^2(s, f_{\theta_0}) + (1 + \beta)^2(1 + \varepsilon) h^2(f_{\theta_0}, f_{\pi(\theta(p, \ell))}) + c(1 + \beta) \frac{(D_{\delta} + n\xi)}{n}.
\]

Remark now that for all \( \theta \in \Theta \),

\[
h^2(f_{\theta}, f_{\pi(\theta)}) \leq \sup_{1 \leq j \leq d} \frac{\alpha_j}{\beta_j} \leq d/n.
\]

By using the triangular inequality,

\[
h^2(f_{\theta_0}, f_{\pi(\theta(p, \ell))}) \leq (1 + \beta) h^2(f_{\theta_0}, f_{\theta(p, \ell)}) + d(1 + \beta^{-1})/n
\]

\[
h^2(f_{\theta_0}, f_{\theta(p, \ell)}) \leq (1 + \beta) h^2(f_{\theta_0}, f_{\pi(\theta(p, \ell))}) + d(1 + \beta^{-1})/n.
\]

We deduce from these two inequalities and from (22) that

\[
(1 - \varepsilon) h^2(f_{\theta_0}, f_{\theta(p, \ell)}) \leq \gamma h^2(s, f_{\theta_0}) + (1 + \beta)^4(1 + \varepsilon) h^2(f_{\theta_0}, f_{\theta(p, \ell)}) + d(1 + \beta^{-1}) [1 - \varepsilon + (1 + \beta)^2(1 + \varepsilon)] + c(1 + \beta)^2 \frac{(D_{\delta} + n\xi)}{n}.
\]

Since \( D_{\delta} \geq d \) and \( \delta \geq 1 \),

\[
(1 - \varepsilon) h^2(f_{\theta_0}, f_{\theta(p, \ell)}) \leq \gamma h^2(s, f_{\theta_0}) + (1 + \varepsilon) h^2(f_{\theta_0}, f_{\theta(p, \ell)}).
\]

By using (21),

\[
(1 - \varepsilon) h^2(f_{\theta_0}, f_{\theta(p, \ell)}) < \beta \left( h^2(f_{\theta_0}, f_{\theta(p, \ell)}) + h^2(f_{\theta_0}, f_{\theta(p, \ell)}) \right) + (1 + \beta)^4(1 + \varepsilon) h^2(f_{\theta_0}, f_{\theta(p, \ell)}).
\]

and thus

\[
h^2(f_{\theta_0}, f_{\theta(p, \ell)}) < C_{\kappa} h^2(f_{\theta_0}, f_{\theta(p, \ell)}).
\]
Finally,
\[ h^2(f_{\theta(p,\ell)}, f_{\hat{\theta}(p,\ell)}) \leq \left( h(f_{\theta_0}, f_{\theta(p,\ell)}) + h(f_{\theta_0}, f_{\hat{\theta}(p,\ell)}) \right)^2 \]
\[ < \left( 1 + \sqrt{C\kappa} \right)^2 h^2(f_{\theta_0}, f_{\theta(p,\ell)}) \]
\[ < \kappa^{-1} h^2(f_{\theta_0}, f_{\hat{\theta}(p,\ell)}) \]

which leads to $\theta_0 \not\in A(p,\ell)$ as wished.

7.2.2. Proof of Claim 1 This claim ensues from the work of Baraud (2011). More precisely, we derive from Proposition 2 of Baraud (2011) that for all $f, f' \in \mathcal{F}_{\text{dis}},$
\[ \left( 1 - \frac{1}{\sqrt{2}} \right) h^2(s, f') + \frac{T(f, f')}{\sqrt{2}} \leq \left( 1 + \frac{1}{\sqrt{2}} \right) h^2(s, f) + \frac{T(f, f') - \mathbb{E}[T(f, f')]}{\sqrt{2}}. \]

Let $z = \varepsilon - 1/\sqrt{2} \in (0, 1 - 1/\sqrt{2})$. We define $\Omega_{\varepsilon}$ by
\[ \Omega_{\varepsilon} = \bigcap_{f, f' \in \mathcal{F}_{\text{dis}}} \left[ \frac{T(f, f') - \mathbb{E}[T(f, f')]}{z (h^2(s, f) + h^2(s, f')) + c(D_{\mathcal{F}} + n\xi)/n} \leq \sqrt{2} \right]. \]

On this event, we have
\[ (1 - \varepsilon) h^2(s, f') + \frac{T(f, f')}{\sqrt{2}} \leq (1 + \varepsilon) h^2(s, f) + c\frac{D_{\mathcal{F}} + n\xi}{n} \]
and it remains to prove that $\mathbb{P}(\Omega_{\varepsilon}^c) \leq e^{-n\xi}.$

The following claim shows that Assumption 3 of Baraud (2011) is fulfilled.

Claim 2. Let
\[ \tau = 4 \frac{2 + \frac{n\sqrt{2}}{6}}{n^2 \varepsilon^2} \]
\[ \eta_{\mathcal{F}} = \max \left\{ 3de^4, \sum_{j=1}^{d} \log \left( 1 + 2t_j^{-1} \left( (d/\bar{\alpha})(cR_j/R_j) \right)^{1/\alpha_j} \right) \right\} \]

Then, for all $r \geq 2\eta_{\mathcal{F}},$
\[ |\mathcal{F}_{\text{dis}} \cap B_h(s, r\sqrt{\tau})| \leq \exp(r^2/2) \]
where $B_h(s, r\sqrt{\tau})$ is the Hellinger ball centered at $s$ with radius $r\sqrt{\tau}$ defined by
\[ B_h(s, r\sqrt{\tau}) = \{ f \in L_1^+(\mathbb{X}, \mu), h^2(s, f) \leq r^2\tau \}. \]

We then derive from Lemma 1 of Baraud (2011) that for all $\xi > 0$ and $y^2 \geq \tau (4\eta_{\mathcal{F}}^2 + n\xi),$
\[ \mathbb{P} \left[ \sup_{f, f' \in \mathcal{F}_{\text{dis}}} \frac{(T(f, f') - \mathbb{E}[T(f, f')])}{\sqrt{2} (h^2(s, f) + h^2(s, f')) \vee y^2} \geq z \right] \leq e^{-n\xi}. \]

Notice now that $4\eta_{\mathcal{F}}^2 \leq 10^3 D_{\mathcal{F}}$ and $10^3\tau \leq c/n.$ This means that we can choose
\[ y^2 = c(D_{\mathcal{F}} + n\xi)/n, \]
which concludes the proof of Claim 1.
Proof of Claim 2. If $\mathcal{F}_{\text{dis}} \cap B_h(s, r\sqrt{\tau}) = \emptyset$, (23) holds. In the contrary case, there exists $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,d}) \in \Theta_{\text{dis}}$ such that $h^2(s, f_{\theta_0}) \leq r^2 \tau$ and thus
\[
|\mathcal{F}_{\text{dis}} \cap B_h(s, r\sqrt{\tau})| \leq |\mathcal{F}_{\text{dis}} \cap B_h(f_{\theta_0}, 2r\sqrt{\tau})|.
\]
Now,
\[
|\mathcal{F}_{\text{dis}} \cap B_h(f_{\theta_0}, 2r\sqrt{\tau})| = |\{f_{\theta, \theta} \in \Theta_{\text{dis}}, h^2(f_{\theta}, f_{\theta_0}) \leq 4r^2 \tau\}|
\leq |\{\theta \in \Theta_{\text{dis}}, \forall j \in \{1, \ldots, d\}, R_j |\theta_j - \theta_{0,j}|^{\alpha_j} \leq 4r^2 \tau\}|.
\]
Let $k_{0,j} \in \mathbb{N}$ be such that $\theta_{0,j} = m_j + k_{0,j} \varepsilon_j$. Then,
\[
|\mathcal{F}_{\text{dis}} \cap B_h(f_{\theta_0}, 2r\sqrt{\tau})| \leq \prod_{j=1}^{d} \left\{k_j \in \mathbb{N}, |k_j - k_{0,j}| \leq \left(4r^2 \tau / R_j\right)^{1/\alpha_j} \varepsilon_j^{-1}\right\}
\leq \prod_{j=1}^{d} \left(1 + 2\varepsilon_j^{-1} \left(4r^2 \tau / R_j\right)^{1/\alpha_j}\right).
\]
By using $4\tau \leq c/n$ and $\varepsilon_j = t_j^{-1}(R_j/n)^{-1/\alpha_j}$,
\[
|\mathcal{F}_{\text{dis}} \cap B_h(f_{\theta_0}, 2r\sqrt{\tau})| \leq \prod_{j=1}^{d} \left(1 + 2t_j^{-1} (r^2 c R_j / R_j)^{1/\alpha_j}\right).
\]
If $\bar{\alpha} \leq e^{-4}$, one can check that $\eta^2_{\mathcal{F}} \geq 4d / \bar{\alpha}$ (since $c \geq 1$ and $t_j^{-1} \geq d^{-1/\alpha_j}$). If now $\bar{\alpha} \geq e^{-4}$, then $\eta^2_{\mathcal{F}} \geq 3de^4 \geq 3d/\bar{\alpha}$. In particular, we always have $r^2 \geq 10(d/\bar{\alpha})$.

We derive from the weaker inequality $r^2 \geq d / \bar{\alpha}$ that
\[
|\mathcal{F}_{\text{dis}} \cap B_h(f_{\theta_0}, 2r\sqrt{\tau})| \leq \left(\frac{r^2}{d / \bar{\alpha}}\right)^{d/\bar{\alpha}} \prod_{j=1}^{d} \left(1 + 2t_j^{-1} \left((d / \bar{\alpha})(c R_j / R_j)\right)^{1/\alpha_j}\right).
\]
We then deduce from the inequalities $r^2 / (d / \bar{\alpha}) \geq 10$ and $\eta^2_{\mathcal{F}} \leq r^2 / 4$ that
\[
|\mathcal{F}_{\text{dis}} \cap B_h(f_{\theta_0}, 2r\sqrt{\tau})| \leq \exp \left(\frac{\log \left(r^2 / (d / \bar{\alpha})\right)}{r^2 / (d / \bar{\alpha}) - r^2}\right) \exp \left(\eta^2_{\mathcal{F}}\right).
\]
We then deduce from the inequalities $r^2 / (d / \bar{\alpha}) \geq 10$ and $\eta^2_{\mathcal{F}} \leq r^2 / 4$ that
\[
|\mathcal{F}_{\text{dis}} \cap B_h(f_{\theta_0}, 2r\sqrt{\tau})| \leq \exp \left(r^2 / 4\right) \exp \left(r^2 / 4\right) \leq \exp(r^2 / 2)
\]
as wished. \hfill \square

7.3. Proof of Theorem 3. This theorem ensues from the following result.

**Theorem 8.** Suppose that the assumptions of Theorem 3 holds. For all $\xi > 0$, the estimator $\hat{\theta}$ built in Algorithm 1 satisfies
\[
\mathbb{P} \left[Ch^2(s, f_{\hat{\theta}}) \geq h^2(s, \mathcal{F}) + \frac{D_{\mathcal{F}}}{n} + \xi\right] \leq e^{-n\xi}
\]
where $C > 0$ depends only on $\kappa, \bar{R}/R$, where
\[
D_{\mathcal{F}} = \max \left\{1, \log \left(1 + t^{-1} (cR/(\alpha R))^{1/\alpha}\right)\right\}
\]
and where \( c \) depends on \( \kappa \) only. Besides, if
\[
h^2(f_{\theta_2}, f_{\theta_2'}) \leq h^2(f_{\theta_1}, f_{\theta_1'}) \quad \text{for all } m \leq \theta_1 \leq \theta_2 < \theta_2' \leq \theta_1' \leq M
\]
then \( C \) depends only on \( \kappa \).

**Proof.** The theorem follows from Theorem 7 page 26 where \( \Theta_i = [\theta^{(i)}, \theta^{(i)}] \) and \( L_i = 1 \). Note that
\[
diam \Theta_i \leq \bar{R}(\theta^{(i)} - \theta^{(i)})^\alpha \leq (\bar{R}/\underline{R})h^2(f_{\theta^{(i)}}, f_{\theta^{(i)}})
\]
which implies that the assumptions of Theorem 7 are fulfilled with \( \kappa_0 = \bar{R}/\underline{R} \). Consequently,
\[
P \left[ C \inf_{\theta \in \Theta_N} h^2(s, f_\theta) \geq h^2(s, \mathcal{F}) + \frac{D_n \xi}{n} + \xi \right] \leq e^{-n\xi}
\]
where \( \Theta_N = [\theta^{(N)}, \theta^{(N)}] \) is such that \( \theta^{(N)} - \theta^{(N)} \leq \eta \). Now, for all \( \theta \in \Theta_N \),
\[
h^2(s, f_\theta) \leq 2h^2(s, f_\theta) + 2h^2(f_\theta, f_\bar{\theta}) \leq 2h^2(s, f_\theta) + 2\bar{R}\eta^\alpha
\]
hence,
\[
h^2(s, f_\theta) \leq 2 \inf_{\theta \in \Theta_N} h^2(s, f_\theta) + 2/n
\]
which establishes the first part of the theorem. The second part derives from the fact that under the additional assumption, \( diam \Theta_i \leq h^2(f_{\theta^{(i)}}, f_{\theta^{(i)}}) \), which means that the assumptions of Theorem 7 are fulfilled with \( \kappa_0 = 1 \). \( \square \)

**7.4. Proof of Proposition 4.** For all \( i \in \{1, \ldots, N - 1\}, \)
\[
\theta^{(i+1)} \in \left\{ \theta^{(i)}, \theta^{(i)} + \min \left( \bar{r}(\theta^{(i)}, \theta^{(i)}), (\theta^{(i)} - \theta^{(i)})/2 \right) \right\}
\]
\[
\theta'^{(i+1)} \in \left\{ \theta^{(i)}, \theta^{(i)} - \min \left( \underline{r}(\theta^{(i)}, \theta^{(i)}), (\theta^{(i)} - \theta^{(i)})/2 \right) \right\}.
\]
Since \( \bar{r}(\theta^{(i)}, \theta^{(i)}) \) and \( \underline{r}(\theta^{(i)}, \theta^{(i)}) \) are larger than
\[
(\kappa\bar{R}/\underline{R})^{1/\alpha}(\theta^{(i)} - \theta^{(i)}),
\]
we have
\[
\theta'^{(i+1)} - \theta^{(i+1)} \leq \max \left\{ 1 - (\kappa\bar{R}/\underline{R})^{1/\alpha}, 1/2 \right\} (\theta^{(i)} - \theta^{(i)}).
\]
By induction, we derive that for all \( i \in \{1, \ldots, N - 1\}, \)
\[
\theta'^{(i+1)} - \theta^{(i+1)} \leq \left( \max \left\{ 1 - (\kappa\bar{R}/\underline{R})^{1/\alpha}, 1/2 \right\} \right)^i (M - m).
\]
The procedure requires thus the computation of at most \( N \) tests where \( N \) is the smallest integer such that
\[
\left( \max \left\{ 1 - (\kappa\bar{R}/\underline{R})^{1/\alpha}, 1/2 \right\} \right)^N (M - m) \leq \eta
\]
that is
\[
N \geq \frac{\log ((M - m)/\eta)}{-\log \left( \max \left\{ 1 - (\kappa\bar{R}/\underline{R})^{1/\alpha}, 1/2 \right\} \right)}.
\]
We conclude by using the inequality \(-1/\log(1 - x) \leq 1/x\) for all \( x \in (0, 1) \). \( \square \)
7.5. Proofs of Proposition 6 and Theorem 5.

7.5.1. Rewriting of Algorithm 3. We rewrite the algorithm to introduce some notations that will be essential to prove Proposition 6 and Theorem 5.

Algorithm 4 Construction of $\Theta_{i+1}$ from $\Theta_i$.

Require: $\Theta_i = \prod_{j=1}^d [a^{(i)}_j, b^{(i)}_j]$
1: Choose $k^{(i)} \in \{1, \ldots, d\}$ such that

$$\max_{1 \leq j \leq d} R_{\Theta_i,j}(b^{(i)}_j - a^{(i)}_j)^{\alpha_j}.$$  

2: $\theta^{(i,1)} = (a^{(i)}_1, \ldots, a^{(i)}_d)$, $\theta^{(i,1)} = \theta^{(i,1)}$ and $\theta^{(i,1)} = b^{(i)}_k$.
3: $\varepsilon_j^{(i,0)} = \bar{r}_{\Theta_i,j}(\theta^{(i,1)}, \theta^{(i,1)})$ and $\varepsilon_j^{(i,0)} = \bar{r}_{\Theta_i,j}(\theta^{(i,1)}, \theta^{(i,1)})$ for all $j \neq k^{(i)}$
4: $\varepsilon_{k^{(i)}} = (b^{(i)}_k - a^{(i)}_k)/2$ and $\varepsilon_{k^{(i)}} = (b^{(i)}_k - a^{(i)}_k)/2$
5: for all $\ell \geq 1$ do
6: $\theta^{(i,\ell+1)} = \theta^{(i,\ell)}$ and $\theta^{(i,\ell+1)} = \theta^{(i,\ell)}$
7: if $T(\theta^{(i,\ell)}, \theta^{(i,\ell)}) \geq 0$ then
8: $\varepsilon_{\psi^{(i,\ell)}_k(1)} = \bar{r}_{\Theta_i,\psi^{(i,\ell)}_k(1)}(\theta^{(i,\ell)}, \theta^{(i,\ell)})$
9: $\varepsilon^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)} = \min(\varepsilon^{(i,\ell-1)}_{\psi^{(i,\ell)}_k(j)}, \bar{r}_{\Theta_i,\psi^{(i,\ell)}_k(j)}(\theta^{(i,\ell)}, \theta^{(i,\ell)}))$, for all $j \in \{2, \ldots, d - 1\}$
10: $\varepsilon^{(i,\ell)}_{k^{(i)}} = \min(\varepsilon^{(i,\ell-1)}_{k^{(i)}}, \bar{r}_{\Theta_i,k^{(i)}}(\theta^{(i,\ell)}, \theta^{(i,\ell)}))$
11: $\Theta^{(i,\ell)} = \left\{1 \leq j \leq d - 1, \theta^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)} + \varepsilon^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)} < b^{(i)}_k\right\}$
12: if $\Theta^{(i,\ell)} \neq \emptyset$ then
13: $\tilde{j}_{\min}^{(i,\ell)} = \min \Theta^{(i,\ell)}$
14: $\theta^{(i,\ell+1)}_{\psi^{(i,\ell)}_k(j)} = a^{(i)}_k$ for all $j \leq \tilde{j}_{\min}^{(i,\ell)} - 1$
15: $\theta^{(i,\ell+1)}_{\psi^{(i,\ell)}_k(j)} = \theta^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)} + \varepsilon^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)}$
16: else $\tilde{j}_{\min}^{(i,\ell)} = d$
17: end if
18: end if
19: if $T(\theta^{(i,\ell)}, \theta^{(i,\ell)}) \leq 0$ then
20: $\varepsilon^{(i,\ell)}_{\psi^{(i,\ell)}_k(1)} = \bar{r}_{\Theta_i,\psi^{(i,\ell)}_k(1)}(\theta^{(i,\ell)}, \theta^{(i,\ell)})$
21: $\varepsilon^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)} = \min(\varepsilon^{(i,\ell-1)}_{\psi^{(i,\ell)}_k(j)}, \bar{r}_{\Theta_i,\psi^{(i,\ell)}_k(j)}(\theta^{(i,\ell)}, \theta^{(i,\ell)}))$, for all $j \in \{2, \ldots, d - 1\}$
22: $\varepsilon^{(i,\ell)}_{k^{(i)}} = \min(\varepsilon^{(i,\ell-1)}_{k^{(i)}}, \bar{r}_{\Theta_i,k^{(i)}}(\theta^{(i,\ell)}, \theta^{(i,\ell)}))$
23: $\Theta^{(i,\ell)} = \left\{1 \leq j \leq d - 1, \theta^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)} + \varepsilon^{(i,\ell)}_{\psi^{(i,\ell)}_k(j)} < b^{(i)}_k\right\}$
24: if $\Theta^{(i,\ell)} \neq \emptyset$ then
25: $\tilde{j}_{\min}^{(i,\ell)} = \min \Theta^{(i,\ell)}$
26: $\theta^{(i,\ell+1)}_{\psi^{(i,\ell)}_k(j)} = a^{(i)}_k$ for all $j \leq \tilde{j}_{\min}^{(i,\ell)} - 1$
27: end if
28: end if
29: end if
28: \( \theta_{\ell+1}^{(i)} \) = \( \theta_{\ell}^{(i)} \) + \( \varepsilon_{\ell}^{(i)} \)
29: \( \psi_{k}^{(i)}(\eta_{\ell})_{l_{\min}}^{(i)} \) = \( \psi_{k}^{(i)}(\eta_{\ell})_{l_{\min}}^{(i)} \)
30: else
31: \( l_{\min}^{(i)} = d \)
32: end if
33: if \( l_{\min}^{(i)} = d \) or \( l_{\min}^{(i)} = d \) then
34: \( L_{i} = \ell \) and quit the loop
35: end if
36: \( a_{j}^{(i+1)} = a_{j}^{(i)} \) and \( b_{j}^{(i+1)} = b_{j}^{(i)} \) for all \( j \neq k^{(i)} \)
37: end for
38: if \( l_{\min}^{(i)} = d \) then
39: \( a_{k}^{(i+1)} = a_{k}^{(i)} + \varepsilon_{k}^{(i)}(L_{i}) \)
40: end if
41: if \( l_{\min}^{(i)} = d \) then
42: \( b_{k}^{(i+1)} = b_{k}^{(i)} - \varepsilon_{k}^{(i)}(L_{i}) \)
43: end if
44: Return: \( \Theta_{i+1} = \prod_{j=1}^{d}[a_{j}^{(i+1)}, b_{j}^{(i+1)}] \)

Algorithm 4 Rewriting of Algorithm 3.

45: \( \Theta_{1} = \prod_{j=1}^{d}[a_{j}^{(1)}, b_{j}^{(1)}] = \prod_{j=1}^{d}[m_{j}, M_{j}] \)
46: for all \( i \geq 1 \) do
47: if There exists \( j \in \{1, \ldots, d\} \) such that \( b_{j}^{(i)} - a_{j}^{(i)} \) \( \eta_{j} \) then
48: Compute \( \Theta_{i+1} \)
49: else
50: Leave the loop and set \( N = i \)
51: end if
52: end for
53: Return: \( \hat{\theta} = \left( \frac{a_{1}^{(N)} + b_{1}^{(N)}}{2}, \ldots, \frac{a_{d}^{(N)} + b_{d}^{(N)}}{2} \right) \)

7.5.2. Proof of Proposition 6. The algorithm computes \( \sum_{i=1}^{N} L_{i} \) tests. Define for all \( j \in \{1, \ldots, d\} \),

\[ I_{j} = \left\{ i \in \{1, \ldots, N\}, k^{(i)} = j \right\}. \]

Then, \( \bigcup_{j=1}^{d} I_{j} = \{1, \ldots, N\} \). Since

\[ \sum_{i=1}^{N} L_{i} \leq \sum_{j=1}^{d} |I_{j}| \sup_{i \in I_{j}} L_{i}, \]
we begin to bound \(|I_j|\) from above. For all \(i \in \{1, \ldots, N-1\},\)
\[
b^{(i+1)}_j - a^{(i+1)}_j \leq b^{(i)}_j - a^{(i)}_j - \min \left( \varepsilon^{(i,L_i)}_j, \varepsilon^{(i,L_i)}_j' \right) \quad \text{if } i \in I_j
\]
\[
b^{(i+1)}_j - a^{(i+1)}_j = b^{(i)}_j - a^{(i)}_j \quad \text{if } i \notin I_j.
\]

For all \(i \in I_j\), and all \(\ell \in \{1, \ldots, L_i\}\), we derive from (20), from the equality \(\theta^{(i,\ell)}_j = b^{(i)}_j\), \(\theta^{(i,\ell)}_j = a^{(i)}_j\) and from the inequalities \(\overline{R}_{\Theta_{i,j}} \geq R_j\) and \(\overline{R}_{\Theta_{i,j}} \leq R_j\) that
\[
\begin{align*}
\bar{r}_{\Theta_{i,j}}(\Theta^{(i,\ell)}, \Theta^{(i,\ell)}') &\geq (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} (b^{(i)}_j - a^{(i)}_j) \\
\underline{r}_{\Theta_{i,j}}(\Theta^{(i,\ell)}, \Theta^{(i,\ell)}') &\geq (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} (b^{(i)}_j - a^{(i)}_j).
\end{align*}
\]

Consequently,
\[
\min \left( \varepsilon^{(i,L_i)}_j, \varepsilon^{(i,L_i)}_j' \right) \geq \min \left\{ \frac{1}{2}, (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} (b^{(i)}_j - a^{(i)}_j) \right\}.
\]

We then have,
\[
\begin{align*}
b^{(i+1)}_j - a^{(i+1)}_j &\leq \max \left( 1/2, 1 - (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} \right) (b^{(i)}_j - a^{(i)}_j) \quad \text{when } i \in I_j \\
b^{(i+1)}_j - a^{(i+1)}_j = b^{(i)}_j - a^{(i)}_j \quad \text{when } i \notin I_j.
\end{align*}
\]

Let \(n_j\) be any integer such that
\[
\left( \max \left\{ 1/2, 1 - (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} \right\} \right)^{n_j} \leq \eta_j/(M_j - m_j).
\]

If \(|I_j| > n_j\), then for \(i = \max I_j\),
\[
\begin{align*}
b^{(i)}_j - a^{(i)}_j &\leq \left( \max \left\{ 1/2, 1 - (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} \right\} \right)^{|I_j|-1} (M_j - m_j) \\
&\leq \left( \max \left\{ 1/2, 1 - (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} \right\} \right)^{n_j} (M_j - m_j)
\end{align*}
\]

and thus \(b^{(i)}_j - a^{(i)}_j \leq \eta_j\). This is impossible because \(i \in I_j\) implies that \(b^{(i)}_j - a^{(i)}_j > \eta_j\). Consequently, \(|I_j| \leq n_j\). We then set \(n_j\) as the smallest integer larger than
\[
\frac{\log (\eta_j/(M_j - m_j))}{\log \left( \max \left\{ 1/2, 1 - (\kappa \overline{R}_j/\overline{R}_j)^{1/\alpha_j} \right\} \right)}.
\]

By using the inequality \(-1/\log(1-x) \leq 1/x\) for all \(x \in (0, 1)\), we obtain
\[
|I_j| \leq 1 + \max \left( \frac{1}{\log 2}, (\kappa \overline{R}_j/\overline{R}_j)^{-1/\alpha_j} \right) \log \left( \frac{M_j - m_j}{\eta_j} \right).
\]

We now roughly bound from above the right-hand side of this inequality:
\[
|I_j| \leq 2 \left( 1 + (\overline{R}_j/(\kappa \overline{R}_j))^{1/\alpha_j} \right) \left( 1 + \log \left( \frac{M_j - m_j}{\eta_j} \right) \right).
\]

We recall that our aim is to bound from above \(\sum_{i=1}^{N} L_i\). Thanks to (24), it remains to upper bound \(\sup_{i \in I_j} L_i\). This ensues from the following lemma.
Lemma 2. Let
\[ \mathcal{L} = \left\{ 1 \leq \ell \leq L_i, \; T(\theta^{(i,\ell)}, \theta^{(i,\ell)}_{\text{opt}}) \geq 0 \right\} \quad \text{and} \quad \mathcal{L}' = \left\{ 1 \leq \ell \leq L_i, \; T(\theta^{(i,\ell)}, \theta^{(i,\ell)}_{\text{opt}}) \leq 0 \right\}. \]
Then,
\[
|\mathcal{L}| \leq \prod_{k \in \{1, \ldots, d\} \setminus \{k^{(i)}\}} \left[ 1 + \left( \frac{\bar{R}_k}{\kappa R_k} \right)^{1/\alpha_k} \right]
\]
\[
|\mathcal{L}'| \leq \prod_{k \in \{1, \ldots, d\} \setminus \{k^{(i)}\}} \left[ 1 + \left( \frac{\bar{R}_k}{\kappa R_k} \right)^{1/\alpha_k} \right].
\]
Since \( \{1, \ldots, L_i\} \subset \mathcal{L} \cup \mathcal{L}' \), we obtain
\[
\sum_{j=1}^{d} |I_j| \sup_{i \in I_j} L_i \leq 4 \left[ \sum_{j=1}^{d} \left( 1 + \log \left( \frac{M_j - m_j}{\eta_j} \right) \right) \right] \left[ \prod_{j=1}^{d} \left( 1 + \left( \frac{\bar{R}_j}{\kappa R_j} \right)^{1/\alpha_j} \right) \right],
\]
which completes the proof.

7.5.3. Proof of Lemma 2. Without lost of generality and for the sake of simplicity, we assume that \( k^{(i)} = d \) and \( \psi_d(j) = j \) for all \( j \in \{1, \ldots, d - 1\} \). Let \( \ell_1 < \cdots < \ell_r \) be the elements of \( \mathcal{L} \). Define for all \( p \in \{1, \ldots, d - 1\} \), \( k_{p,0} = 0 \) and by induction for all integer \( m \),
\[
k_{p,m+1} = \begin{cases} \inf \left\{ k > k_{p,m}, j^{(i,k)}_{\min} > p \right\} & \text{if there exists } k \in \{k_{p,m+1}, \ldots, r\} \text{ such that } j^{(i,k)}_{\min} > p \\ r & \text{otherwise.} \end{cases}
\]
Let \( \mathcal{M}_p \) be the smallest integer \( m \) for which \( k_{p,m} = r \). Set for all \( m \in \{0, \ldots, \mathcal{M}_p - 1\} \),
\[
K_{p,m} = \{k_{p,m} + 1, \ldots, k_{p,m+1}\}.
\]
The cardinality of \( K_{p,m} \) can be upper bounded by the claim below.

Claim 3. For all \( p \in \{1, \ldots, d - 1\} \) and \( m \in \{0, \ldots, \mathcal{M}_p - 1\} \),
\[
|K_{p,m}| \leq \prod_{k=1}^{p} \left[ 1 + \left( \frac{\bar{R}_k}{\kappa R_k} \right)^{1/\alpha_k} \right].
\]
Lemma 2 follows from the equality \( \mathcal{L} = K_{d-1,0} \). The cardinality of \( \mathcal{L}' \) can be bounded from above in the same way.

Proof of Claim 3. The result is proved by induction. We begin to prove (27) when \( p = 1 \).

Let \( m \in \{0, \ldots, \mathcal{M}_1 - 1\} \). We have \( \theta_1^{(i,k_{1,m+1})} = a_1^{(i)} \) and for \( j \in \{1, \ldots, k_{1,m+1} - k_{1,m} - 1\} \),
\[
\theta_1^{(i,k_{1,m}+j+1)} \geq \theta_1^{(i,k_{1,m}+j)} + \bar{r}_{\Theta_{1,1}} \left( \theta^{(i,k_{1,m}+j)}, \theta^{(i,k_{1,m}+j)} \right).
\]
Now,
\[
\bar{r}_{\Theta_{1,1}} \left( \theta^{(i,k_{1,m}+j)}, \theta^{(i,k_{1,m}+j)} \right) \geq \left( (\kappa R_{\Theta_{1,1}}/\bar{R}_{\Theta_{1,1}})(b_d^{(i)} - a_d^{(i)}) \right)^{1/\alpha_1}.
\]
Let $\gamma$. Claim 4. For all $R$

Since $R_{\Theta,1}(b_d^{(i)} - a_d^{(i)})^{\alpha_d} \geq R_{\Theta,1}(b_1^{(i)} - a_1^{(i)})^{\alpha_1},$

\[
\bar{r}_{\Theta,1} \left( \theta^{(i,\ell_{k_1,m+j})}, \theta^{(i,\ell_{k_1,m+j})} \right) \geq \left( \kappa R_{\Theta,1}/R_{\Theta,1} \right)^{1/\alpha_1} (b_1^{(i)} - a_1^{(i)})
\]

(28)

This leads to

\[
\theta_1^{(i,\ell_{k_1,m+j+1})} \geq \theta_1^{(i,\ell_{k_1,m+j})} + \left( \kappa R_{\Theta,1}/R_{\Theta,1} \right)^{1/\alpha_1} (b_1^{(i)} - a_1^{(i)}).
\]

Moreover, $\theta_1^{(i,\ell_{k_1,m+1+1})} \leq b_1^{(i)}$ (because all the $\theta^{(i,\ell)}, \theta^{(i,\ell)}$ belong to $\Theta_1$). Consequently,

\[
a_1^{(i)} + (k_{1,m+1} - k_{1,m} - 1) \left( \kappa R_{\Theta,1}/R_{\Theta,1} \right)^{1/\alpha_1} (b_1^{(i)} - a_1^{(i)}) \leq b_1^{(i)}
\]

which shows the result for $p = 1$.

Suppose now that (27) holds for $p \in \{1, \ldots, d - 2\}$. We shall show that it also holds for $p + 1$. Let $m \in \{0, \ldots, M_{p+1} - 1\}$. We use the claim below whose proof is postponed to Section 7.5.6.

**Claim 4.** For all $m \in \{0, \ldots, M_{p+1} - 1\}$, there exists $m' \in \{0, \ldots, M_p - 1\}$ such that $k_{p,m'+1} \in K_{p+1,m}$.

The claim says that we can consider the smallest integer $m_0$ of $\{0, \ldots, M_p - 1\}$ such that $k_{p,m_0+1} > k_{p+1,m}$, and the larger integer $m_1$ of $\{0, \ldots, M_p - 1\}$ such that $k_{p,m_1+1} \leq k_{p+1,m+1}$. We define

\[
I_{m_0} = \{k_{p+1,m} + 1, \ldots, k_{p,m_0+1}\}
\]

\[
I_{m'} = \{k_{p,m'} + 1, \ldots, k_{p,m'+1}\}
\]

\[
I_{m_1+1} = \{k_{p,m_1+1} + 1, \ldots, k_{p+1,m+1}\}.
\]

We then have

\[
K_{p+1,m} = \bigcup_{m'=m_0}^{m_1+1} I_{m'}.
\]

Notice that for all $m' \in \{m_0, \ldots, m_1\}$, $I_{m'} \subset K_{p,m'}$. We consider two cases.

- If $k_{p,m_1+1} = k_{p+1,m_1+1}$, then $I_{m_1+1} = \emptyset$ and thus, by using the above inclusion and the induction assumption,

\[
|K_{p+1,m'}| \leq (m_1 - m_0 + 1) \prod_{k=1}^{p} \left[ 1 + \left( \frac{\bar{R}_k}{\kappa R_{\kappa}} \right)^{1/\alpha_k} \right].
\]

- If $k_{p,m_1+1} < k_{p+1,m_1+1}$ then $m_1 + 1 \leq M_p - 1$. Indeed, if this is not true, then $m_1 = M_p - 1$, which leads to $k_{p,m_1+1} = r$ and thus $k_{p+1,m+1} > r$. This is impossible since $k_{p+1,m+1}$ is always smaller than $r$ (by definition). Consequently, $I_{m_1+1} \subseteq K_{p,m_1+1}$ and we derive from the induction assumption

\[
|K_{p+1,m'}| \leq (m_1 - m_0 + 2) \prod_{k=1}^{p} \left[ 1 + \left( \frac{\bar{R}_k}{\kappa R_{\kappa}} \right)^{1/\alpha_k} \right].
\]
We now bound from above \( m_1 - m_0 \).

Since for all \( k \in \{ k_{p+1,m} + 1, \ldots, k_{p,m_0+1} - 1 \} \), \( j_{\text{min}}^{(i,k)} \leq p \), we have

\[
\theta_{p+1}^{(i,k_{p,m_0+1})} = \theta_{p+1}^{(i,k_{p+1,m}+1)} = a_{p+1}^{(i)}.
\]

Since \( j_{\text{min}}^{(i,k_{p,m_0+1})} = p + 1 \),

\[
\theta_{p+1}^{(i,k_{p,m_0+1}+1)} \geq \theta_{p+1}^{(i,k_{p,m_0+1})} + \bar{r} \theta_{i,p+1}^{(i,1)} (\theta^{(i,1)}, \theta^{(i,1)})
\]

and thus by using a similar argument as the one used in the proof of (28),

\[
\theta_{p+1}^{(i,k_{p,m_0+1}+1)} \geq \theta_{p+1}^{(i,k_{p,m_0+1})} + (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)})
\]

\[
\geq a_{p+1}^{(i)} + (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)}).
\]

Similarly, for all \( m' \in \{ m_0 + 1, \ldots, m_1 \} \) and \( k \in \{ k_{p,m'} + 1, \ldots, k_{p,m'+1} - 1 \} \), \( j_{\text{min}}^{(i,k)} \leq p \) and thus

\[
\theta_{p+1}^{(i,k_{p,m'+1})} = \theta_{p+1}^{(i,k_{p,m'+1}+1)}.
\]

Moreover, for all \( m' \in \{ m_0 + 1, \ldots, m_1 - 1 \} \), \( j_{\text{min}}^{(i,k_{p,m'+1})} = p + 1 \) and thus

\[
\theta_{p+1}^{(i,k_{p,m'+1}+1)} \geq \theta_{p+1}^{(i,k_{p,m'+1})} + (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)})
\]

\[
\geq a_{p+1}^{(i)} + (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)}).
\]

This leads to

\[
\theta_{p+1}^{(i,k_{p,m_1}+1)} \geq \theta_{p+1}^{(i,k_{p,m_0+1}+1)} + (m_1 - m_0 - 1) (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)})
\]

\[
\geq a_{p+1}^{(i)} + (m_1 - m_0) (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)}).
\]

There are two types of cases involved: if \( k_{p,m_1+1} = k_{p+1,m_1+1} \) and if \( k_{p,m_1+1} < k_{p+1,m_1+1} \).

- If \( k_{p,m_1+1} = k_{p+1,m_1+1} \),

\[
\theta_{p+1}^{(i,k_{p+1,m_1+1})} = \theta_{p+1}^{(i,k_{p,m_1+1})}
\]

\[
\geq a_{p+1}^{(i)} + (m_1 - m_0) (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)}).
\]

Since \( \theta_{p+1}^{(i,k_{p+1,m_1+1})} \leq b_{p+1}^{(i)} \), we have

\[
m_1 - m_0 \leq (\kappa R_{p+1}/(\kappa R_{p+1}))^{1/\alpha_{p+1}}.
\]

- If now \( k_{p,m_1+1} < k_{p+1,m_1+1} \), then (29) also holds for \( m' = m_1 \). This implies

\[
\theta_{p+1}^{(i,k_{p,m_1+1}+1)} \geq a_{p+1}^{(i)} + (m_1 - m_0 + 1) (\kappa R_{p+1}/R_{p+1})^{1/\alpha_{p+1}} (b_{p+1}^{(i)} - a_{p+1}^{(i)}).
\]
Since $j_{\min}^{(i,\ell_k)} \leq p$ for all $k \in \{k_p, m_1+1, \ldots, k_{p+1}, m_1 \}$,
\[
\theta_{p+1}^{(i,\ell_k, m+1)} = \frac{\theta_{p+1}^{(i,\ell_k, m+1)}}{	heta_{p+1}^{(i,\ell_k, m+1)}} \geq a_{p+1}^{(i)} + (m_1 - m_0 + 1) \left( \frac{\kappa R_{p+1}}{R_{p+1}} \right)^{1/\alpha_{p+1}} \left( b_{p+1}^{(i)} - a_{p+1}^{(i)} \right).
\]
Since, $\theta_{p+1}^{(i,\ell_k, m+1)} \leq b_{p+1}^{(i)}$,
\[
m_1 - m_0 + 1 \leq \left( \frac{R_{p+1}}{(\kappa R_{p+1})} \right)^{1/\alpha_{p+1}}.
\]
This ends the proof.

7.5.4. Proof of Theorem 5. The lemma and claim below show that the assumptions of Theorem 7 (page 26) are satisfied.

**Lemma 3.** For all $i \in \{1, \ldots, N-1\}$,
\[
\Theta_i \setminus \bigcup_{\ell=1}^{L_i} B^{(i,\ell)} \subset \Theta_{i+1} \subset \Theta_i.
\]

**Claim 5.** For all $i \in \{1, \ldots, N-1\}$ and $\ell \in \{1, \ldots, L_i\}$,
\[
\kappa_0 \text{diam}(\Theta_i) \leq h^2(f_{\theta^{(i,\ell)}}, f_{\theta^{(i,\ell)}})
\]
where $\kappa_0 = \inf_{1 \leq j \leq d} \frac{R_j}{\kappa_j}$. We now derive from Theorem 7 that
\[
P \left[ C \inf_{\theta \in \Theta_N} h^2(s, f_\theta) \geq h^2(s, \mathcal{F}) + \frac{D_{\mathcal{F}}}{n} + \xi \right] \leq e^{-n\xi}
\]
where $C > 0$ depends only on $\kappa, \sup_{1 \leq j \leq d} \frac{R_j}{\kappa_j}$. Consequently, with probability larger than $1 - e^{-n\xi}$,
\[
h^2(s, f_\theta) \leq 2 \inf_{\theta \in \Theta_N} h^2(s, f_\theta) + 2 \text{diam} \Theta_N
\]
\[
\leq 2C^{-1} \left( h^2(s, \mathcal{F}) + \frac{D_{\mathcal{F}}}{n} + \xi \right) + 2 \sup_{1 \leq j \leq d} \frac{R_j}{\kappa_j} \eta_j^{\alpha_j}
\]
\[
\leq 2C^{-1} \left( h^2(s, \mathcal{F}) + \frac{D_{\mathcal{F}}}{n} + \xi \right) + 2d/n
\]
\[
\leq C' \left( h^2(s, \mathcal{F}) + \frac{d}{n} + \xi \right).
\]
The theorem follows.

7.5.5. Proof of Lemma 3. Since
\[
\varepsilon_{k(i)}^{(i,\ell_k)} \leq \frac{e_{k(i)}^{(i,\ell_k)} - a_{k(i)}^{(i,\ell_k)}}{2} \quad \text{and} \quad \varepsilon_{k(i)}^{(i,\ell_k)} \leq \frac{b_{k(i)}^{(i,\ell_k)} - a_{k(i)}^{(i,\ell_k)}}{2},
\]
we have $\Theta_{i+1} \subset \Theta_i$. We now aim at proving $\Theta_i \setminus \bigcup_{\ell=1}^{L_i} B^{(i,\ell)} \subset \Theta_{i+1}$.  


We introduce the rectangles
\[
\mathcal{R}^{(i,\ell)} = \prod_{q=1}^{d} \left[ \theta_q^{(i,\ell)}, \theta_q^{(i,\ell)} + \varepsilon_q^{(i,\ell)} \right]
\]
\[
\mathcal{R}^{(i,\ell)} = \prod_{q=1}^{k(i)-1} \left[ \theta_q^{(i,\ell)}, \theta_q^{(i,\ell)} + \varepsilon_q^{(i,\ell)} \right] \times \left[ \theta_k^{(i,\ell)} - \varepsilon_k^{(i,\ell)}, \theta_k^{(i,\ell)} \right] \times \prod_{q=k(i)+1}^{d} \left[ \theta_q^{(i,\ell)}, \theta_q^{(i,\ell)} + \varepsilon_q^{(i,\ell)} \right]
\]
and we set
\[
\mathcal{R}^{(i,\ell)} = \begin{cases} 
\mathcal{R}^{(i,\ell)} & \text{if } T(\theta^{(i,\ell)}, \theta^{(i,\ell)}) > 0 \\
\mathcal{R}^{(i,\ell)} & \text{if } T(\theta^{(i,\ell)}, \theta^{(i,\ell)}) < 0 \\
\mathcal{R}^{(i,\ell)} \cup \mathcal{R}^{(i,\ell)} & \text{if } T(\theta^{(i,\ell)}, \theta^{(i,\ell)}) = 0.
\end{cases}
\]
We derive from (15) that \( \Theta_i \cap \mathcal{R}^{(i,\ell)} \subset B^{(i,\ell)} \). It is then sufficient to show
\[
\Theta_i \setminus \bigcup_{\ell=1}^{L_i} \mathcal{R}^{(i,\ell)} \subset \Theta_{i+1}.
\]
For this purpose, note that either \( T(\theta^{(i,L_i)}, \theta^{(i,L_i)}) \geq 0 \) or \( T(\theta^{(i,L_i)}, \theta^{(i,L_i)}) \leq 0 \). In what follows, we assume that \( T(\theta^{(i,L_i)}, \theta^{(i,L_i)}) \geq 0 \) but a similar proof can be made if \( T(\theta^{(i,L_i)}, \theta^{(i,L_i)}) \) is non-positive. Without lost of generality, and for the sake of simplicity, we suppose as in the proof of Lemma 2 that \( k(i) = d \) and \( \psi_d(j) = j \) for all \( j \in \{1, \ldots, d-1\} \). Let
\[
\mathcal{L} = \left\{ 1 \leq \ell \leq L_i, \ T(\theta^{(i,\ell)}, \theta^{(i,\ell)}) \geq 0 \right\}
\]
and \( \ell_1 < \cdots < \ell_r \) be the elements of \( \mathcal{L} \). We have
\[
\Theta_{i+1} = \prod_{q=1}^{d-1} \left[ a_q^{(i)}, b_q^{(i)} \right] \times \left[ a_d^{(i)} + \varepsilon_d^{(i,L_i)}, b_d^{(i)} \right]
\]
and it is sufficient to prove that
\[
\prod_{q=1}^{d-1} \left[ a_q^{(i)}, b_q^{(i)} \right] \times \left[ a_d^{(i)} + \varepsilon_d^{(i,L_i)}, b_d^{(i)} \right] \subset \bigcup_{k=1}^{r} \mathcal{R}^{(i,\ell_k)}.
\]
For this, remark that for all \( k \in \{1, \ldots, r\} \), \( \theta^{(i,\ell_k)} = a_d^{(i)} \) and thus
\[
\mathcal{R}^{(i,\ell_k)} = \prod_{q=1}^{d-1} \left[ \theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varepsilon_q^{(i,\ell_k)} \right] \times \left[ a_d^{(i)} + \varepsilon_d^{(i,\ell_k)}, b_d^{(i)} \right].
\]
By using the fact that the sequence \( (\varepsilon_d^{(i,\ell_k)})_k \) is non-increasing,
\[
\left[ a_d^{(i)} + \varepsilon_d^{(i,L_i)}, b_d^{(i)} \right] \subset \bigcap_{k=1}^{r} \left[ a_d^{(i)} + \varepsilon_d^{(i,\ell_k)}, b_d^{(i)} \right].
\]
This means that it is sufficient to show that
\[
(30) \prod_{q=1}^{d-1} \left[ a_q^{(i)}, b_q^{(i)} \right] \subset \bigcup_{k=1}^{r} \prod_{q=1}^{d-1} \left[ \theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varepsilon_q^{(i,\ell_k)} \right].
\]
Let us now define (as in the proof of Lemma 2) for all \( p \in \{1, \ldots, d-1\} \), \( k_{p,0} = 0 \) and by induction for all integer \( m \),

\[
k_{p,m+1} = \begin{cases} \inf \{k > k_{p,m}, J_{\min}^{(i,\ell_k)} > p\} & \text{if there exists } k \in \{k_{p,m} + 1, \ldots, r\} \text{ such that } \tilde{J}_{\min}^{(i,\ell_k)} > p \\ \text{otherwise.} \end{cases}
\]

Let \( \mathcal{M}_p \) be the smallest integer \( m \) such that \( k_{p,m} = r \). Let then for all \( m \in \{0, \ldots, \mathcal{M}_p - 1\} \),

\[
K_{p,m} = \{k_{p,m} + 1, \ldots, k_{p,m+1}\}.
\]

We shall use the claim below (whose proof is delayed to Section 7.5.6).

**Claim 6.** Let \( m' \in \{0, \ldots, \mathcal{M}_p+1\} \), \( p \in \{1, \ldots, d-1\} \). There exists a subset \( \mathcal{M} \) of \( \{0, \ldots, \mathcal{M}_p - 1\} \) such that

\[
K'_p = \{k_{p,m+1}, m \in \mathcal{M}\} \subset K_{p+1,m'}
\]

and

\[
\left[a^{(i)}_{p+1}, b^{(i)}_{p+1}\right] \subset \bigcup_{k \in K'_p} \left[\theta^{(i,\ell_k)}_q, \theta^{(i,\ell_k)}_q + \varepsilon^{(i,\ell_k)}_q\right].
\]

We prove by induction on \( p \) the following result. For all \( p \in \{1, \ldots, d-1\} \), and all \( m \in \{0, \ldots, \mathcal{M}_p - 1\} \),

\[
\prod_{q=1}^{p} \left[a^{(i)}_q, b^{(i)}_q\right] \subset \bigcup_{k \in K_{p,m}} \prod_{q=1}^{p} \left[\theta^{(i,\ell_k)}_q, \theta^{(i,\ell_k)}_q + \varepsilon^{(i,\ell_k)}_q\right].
\]

Note that (30) ensues from this inclusion when \( p = d - 1 \) and \( m = 0 \).

We begin to prove (31) for \( p = 1 \) and all \( m \in \{0, \ldots, \mathcal{M}_1 - 1\} \). For all \( k \in \{k_{1,m+1}, \ldots, k_{1,m+1} - 1\} \), \( \tilde{J}_{\min}^{(i,\ell_k)} \leq 1 \) and thus

\[
\theta^{(i,\ell_{k+1})}_1 \in \left\{\theta^{(i,\ell_k)}_1, \theta^{(i,\ell_k)}_1 + \varepsilon^{(i,\ell_k)}_1\right\}.
\]

This implies that the set

\[
\bigcup_{k=k_{1,m+1}}^{k_{1,m+1}+1} \left[\theta^{(i,\ell_k)}_1, \theta^{(i,\ell_k)}_1 + \varepsilon^{(i,\ell_k)}_1\right]
\]

is an interval. Now, \( \theta^{(i,\ell_{k+1})}_1 = a^{(i)}_1, \theta^{(i,\ell_{k+1})}_1 + \varepsilon^{(i,\ell_k)}_1 \geq b^{(i)}_1 \) since \( \tilde{J}_{\min}^{(i,\ell_{k+1})} > 1 \). We have

\[
[a^{(i)}_1, b^{(i)}_1] \subset \bigcup_{k=k_{1,m+1}}^{k_{1,m+1}+1} \left[\theta^{(i,\ell_k)}_1, \theta^{(i,\ell_k)}_1 + \varepsilon^{(i,\ell_k)}_1\right]
\]

which establishes (31) when \( p = 1 \).

Let now \( p \in \{1, \ldots, d - 2\} \) and assume that for all \( m \in \{0, \ldots, \mathcal{M}_p - 1\} \),

\[
\prod_{q=1}^{p} \left[a^{(i)}_q, b^{(i)}_q\right] \subset \bigcup_{k \in K_{p,m}} \prod_{q=1}^{p} \left[\theta^{(i,\ell_k)}_q, \theta^{(i,\ell_k)}_q + \varepsilon^{(i,\ell_k)}_q\right].
\]
Let \( m' \in \{0, \ldots, M_{p+1} - 1\} \). We shall show that
\[
\prod_{q=1}^{p+1} a_q^{(i)} b_q^{(i)} \subseteq \bigcup_{k \in K_{p+1, m'}} \prod_{q=1}^{p+1} \left[ \theta_q^{(i, \ell_k)} + \varepsilon_q^{(i, \ell_k)} \right].
\]

Let \( x \in \prod_{q=1}^{p+1} a_q^{(i)} b_q^{(i)} \). By using Claim 6, there exists \( m \in \{0, \ldots, M_p - 1\} \) such that
\[
x_{p+1} \in \left[ \theta_{p+1}^{(i, \ell_k, p, m+1)} + \varepsilon_{p+1}^{(i, \ell_k, p, m+1)} \right]
\]
and such that \( k_{p, m+1} \in K_{p+1, m'} \). By using the induction assumption, there exists \( k \in K_{p, m} \) such that
\[
x = (x_1, \ldots, x_p) \in \prod_{q=1}^{p} \left[ \theta_q^{(i, \ell_k)} + \varepsilon_q^{(i, \ell_k)} \right].
\]
Since \( k \in K_{p, m}, \theta_{p+1}^{(i, \ell_k)} = \theta_{p+1}^{(i, \ell_k, p, m+1)} \) and \( \varepsilon_{p+1}^{(i, \ell_k, p, m+1)} \leq \varepsilon_{p+1}^{(i, \ell_k)} \). Hence,
\[
x_{p+1} \in \left[ \theta_{p+1}^{(i, \ell_k)} + \varepsilon_{p+1}^{(i, \ell_k)} \right].
\]
We finally use the claim below to show that \( k \in K_{p+1, m'} \) which concludes the proof.

**Claim 7.** Let \( m \in \{0, \ldots, M_p - 1\} \) and \( m' \in \{0, \ldots, M_{p+1} - 1\} \). If \( k_{p, m+1} \in K_{p+1, m'} \), then \( K_{p, m} \subseteq K_{p+1, m'} \).

**7.5.6. Proof of the claims.**

*Proof of Claim 4.* The set \( \{m' \in \{0, \ldots, M_p - 1\} \mid k_{p, m'} \leq k_{p+1, m+1} \} \) is non-empty and we can thus define the largest integer \( m' \) of \( \{0, \ldots, M_p - 1\} \) such that \( k_{p, m'+1} \leq k_{p+1, m+1} \). We then have
\[
k_{p, m'} = \sup \left\{ k < k_{p, m'+1}, i_{(\ell_k)}^{(i)} > p \right\}.
\]
Since \( k_{p, m'} < k_{p+1, m+1} \),
\[
k_{p, m'} = \sup \left\{ k < k_{p+1, m+1}, i_{(\ell_k)}^{(i)} > p \right\}
\]
\[
\geq \sup \left\{ k < k_{p+1, m+1}, i_{(\ell_k)}^{(i)} > p + 1 \right\}
\]
\[
\geq k_{p+1, m}.
\]
Hence, \( k_{p, m'+1} \geq k_{p, m'} + 1 \geq k_{p+1, m} + 1 \). Finally, \( k_{p, m'+1} \in K_{p, m} \). \( \square \)

*Proof of Claim 5.* Let \( i \in \{1, \ldots, N - 1\} \) and \( \ell \in \{1, \ldots, L_i\} \). Then,
\[
diam(\Theta_i) \leq \sup_{1 \leq j \leq d} R_{\Theta_i, j} (b_j^{(i)} - a_j^{(i)})^{\alpha_j}
\]
\[
\leq \left( \sup_{1 \leq j \leq d} R_{\Theta_i, j} \right) \sup_{1 \leq j \leq d} R_{\Theta_i, j} (b_j^{(i)} - a_j^{(i)})^{\alpha_j}
\]
\[
\leq \left( \sup_{1 \leq j \leq d} R_{\Theta_i, j} \right) R_{\Theta_i, k} (b_k^{(i)} - a_k^{(i)})^{\alpha_k^{(i)}}.
\]
Now, $\theta_{k(i)}^{(i, \ell)} = a_{k(i)}^{(i)}$ and $\theta_{k(i)}^{(i, \ell)} = b_{k(i)}^{(i)}$ and thus
\[
\text{diam}(\Theta_i) \leq \left( \frac{\text{sup}_{1 \leq j \leq d} R_{\Theta_{i,j}}}{\text{sup}_{1 \leq j \leq d} R_{\Theta_{i,j}}} \right) R_{\Theta_i,k(i)} (\theta_{k(i)}^{(i, \ell)} - \theta_{k(i)}^{(i, \ell)})^{a_{k(i)}}
\leq \left( \frac{\text{sup}_{1 \leq j \leq d} R_{\Theta_{i,j}}}{\text{sup}_{1 \leq j \leq d} R_{\Theta_{i,j}}} \right) \text{sup}_{1 \leq j \leq d} R_{\Theta_{i,j}} (\theta_{j}^{(i, \ell)} - \theta_{j}^{(i, \ell)})^{a_{j}}
\leq \left( \frac{\text{sup}_{1 \leq j \leq d} R_{\Theta_{i,j}}}{\text{sup}_{1 \leq j \leq d} R_{\Theta_{i,j}}} \right) h^2(\theta_{(i, \ell)}^{(i, \ell)}, \theta_{(i, \ell)}^{(i, \ell)}).
\]
We conclude by using $\frac{\text{R}_{\Theta_{i,j}}}{\text{R}_{\Theta_{i,j}}} \leq \frac{\text{R}_j}{\text{R}_j}$. \hfill \Box

**Proof of Claim 6.** Thanks to Claim 4 (page 37), we can define the smallest integer $m_0$ of $\{0, \ldots, M_p - 1\}$ such that $k_{p,m_0+1} \in K_{p+1,m'}$, and the largest integer $m_1$ of $\{0, \ldots, M_p - 1\}$ such that $k_{p+1} \in K_{p+1,m'}$. Define now
\[
M = \{m_0, m_0 + 1, \ldots, m_1\}.
\]
Note that for all $m \in \{m_0, \ldots, m_1\}$, $k_{p,m+1} \in K_{p+1,m'}$ (this ensues from the fact that the sequence $(k_{p,m})_m$ is increasing).

Let $m \in \{0, \ldots, M_p - 1\}$ be such that $k_{p,m} \in K_{p+1,m'}$ and $k_{p,m} \neq k_{p+1,m' - 1}$. Then $\ell^{(i, k_{p,m})}_{\text{min}} \leq p + 1$ and since $\ell^{(i, k_{p,m})}_{\text{min}} > p$, we also have $\ell^{(i, k_{p,m})}_{\text{min}} = p + 1$. Consequently,
\[
\theta_{p+1}^{(i, k_{p,m})} = \theta_{p+1}^{(i, k_{p,m})} + \varepsilon_{p+1}^{(i, k_{p,m})}.
\]
Now, $\theta_{p+1}^{(i, k_{p,m})} = \theta_{p+1}^{(i, k_{p,m+1})}$ since $k_{p,m} + 1$ and $k_{p,m+1}$ belong to $K_{p,m}$. The set
\[
\left[ \theta_{p+1}^{(i, k_{p,m})}, \theta_{p+1}^{(i, k_{p,m})} + \varepsilon_{p+1}^{(i, k_{p,m})} \right] \cup \left[ \theta_{p+1}^{(i, k_{p,m+1})}, \theta_{p+1}^{(i, k_{p,m+1})} + \varepsilon_{p+1}^{(i, k_{p,m+1})} \right]
\]
is thus the interval
\[
\left[ \theta_{p+1}^{(i, k_{p,m})}, \theta_{p+1}^{(i, k_{p,m+1})} + \varepsilon_{p+1}^{(i, k_{p,m+1})} \right].
\]
We apply this argument for each $m \in \{m_0 + 1, \ldots, m_1\}$ to derive that the set
\[
I = \bigcup_{m=m_0}^{m_1} \left[ \theta_{p+1}^{(i, k_{p,m+1})}, \theta_{p+1}^{(i, k_{p,m+1})} + \varepsilon_{p+1}^{(i, k_{p,m+1})} \right]
\]
is the interval
\[
I = \left[ \theta_{p+1}^{(i, k_{p,m_0+1})}, \theta_{p+1}^{(i, k_{p,m_0+1})} + \varepsilon_{p+1}^{(i, k_{p,m_0+1})} \right].
\]
The claim is proved if we show that
\[
[a_{p+1}^{(i)}, b_{p+1}^{(i)}] \subset I.
\]
Since $I$ is an interval, it remains to prove that $a_{p+1}^{(i)} \in I$ and $b_{p+1}^{(i)} \in I$. 
We begin to show \( a_{p+1}^{(i)} \in I \) by showing that \( a_{p+1}^{(i)} = \theta_{p+1}^{(i,k_{p,m_0+1})} \). If \( k_{p+1,m'} = 0 \), then \( m' = 0 \) and \( m_0 = 0 \). Besides, since 1 and \( k_{p,1} \) belong to \( K_{p,0} \), we have \( \theta_{p+1}^{(i,k_{p,0})} = \theta_{p+1}^{(i,1)} \). Now, \( \theta_{p+1}^{(i,1)} = a_{p+1}^{(i)} \) and thus \( a_{p+1}^{(i)} \in I \). We now assume that \( k_{p+1,m'} \neq 0 \). Since \( k_{p,m_0} \leq k_{p+1,m'} \), there are two cases.

- First case: \( k_{p,m_0} = k_{p+1,m'} \). We then have \( \min(i,k_{p,m_0}) > p + 1 \) and thus \( \theta_{p+1}^{(i,k_{p,m_0})} = a_{p+1}^{(i)} \). Since \( k_{p,m_0+1} \) and \( k_{p,m_0} + 1 \) belong to \( K_{p,m_0} \), \( \theta_{p+1}^{(i,k_{p,m_0+1})} = \theta_{p+1}^{(i,k_{p,m_0}+1)} \) and thus \( \theta_{p+1}^{(i,k_{p,m_0+1})} = a_{p+1}^{(i)} \) as wished.

- Second case: \( k_{p,m_0} + 1 \leq k_{p+1,m'} \). Then, \( k_{p+1,m'} \in K_{p,m_0} \) and thus \( \theta_{p+1}^{(i,k_{p,m_0+1})} = \theta_{p+1}^{(i,k_{p+1,m'})} \).

Since \( \min(i,k_{p+1,m'}) > p + 1 \), we have \( \theta_{p+1}^{(i,k_{p+1,m'})} + \min(i,k_{p+1,m'}) \geq b_{p+1}^{(i)} \). By using the fact that the sequence \( (\varepsilon_{p+1}^{(i,k)})_k \) is decreasing, we then deduce

\[
\theta_{p+1}^{(i,k_{p,m_0+1})} \geq b_{p+1}^{(i)}
\]

and thus \( \min(i,k_{p,m_0+1}) > p + 1 \). This establishes that

\[
\theta_{p+1}^{(i,k_{p,m_0+2})} = a_{p+1}^{(i)}.
\]

Let us now show \( k_{p,m_0} + 2 \leq k_{p,m_0+1} \). Otherwise, \( k_{p,m_0} + 2 = k_{p,m_0+1} + 1 \) and thus \( k_{p,m_0} + 1 = k_{p,m_0+1} \), which means that \( k_{p,m_0} + 1 = k_{p,m_0+1} \) (we recall that \( (k_{p,m})_m \) is an increasing sequence of integers). Since we are in the case where \( k_{p,m_0} + 1 \leq k_{p+1,m'} \), we have \( k_{p,m_0} + 1 \leq k_{p+1,m'} \) which is impossible since \( k_{p,m_0} + 1 \in K_{p+1,m'} \).

Consequently, since \( k_{p,m_0} + 2 \leq k_{p,m_0+1} \), we have \( k_{p,m_0} + 2 \in K_{p,m_0} \) and thus \( \theta_{p+1}^{(i,k_{p,m_0+2})} = \theta_{p+1}^{(i,k_{p,m_0+1})} \). We then deduce from (32) that \( \theta_{p+1}^{(i,k_{p,m_0+1})} = a_{p+1}^{(i)} \) as wished.

We now show that \( b_{p+1}^{(i)} \in I \) by showing that \( \theta_{p+1}^{(i,k_{p,m_1+1})} + \varepsilon_{p+1}^{(i,k_{p,m_1+1})} \geq b_{p+1}^{(i)} \). If \( m_1 = M - 1 \),

\[
\theta_{p+1}^{(i,k_{p,m_1+1})} + \varepsilon_{p+1}^{(i,k_{p,m_1+1})} = \theta_{p+1}^{(i,L_1)} + \varepsilon_{p+1}^{(i,L_1)}.
\]

Since \( \min(i,L_1) = d \), we have \( \theta_{p+1}^{(i,L_1)} + \varepsilon_{p+1}^{(i,L_1)} \geq b_{p+1}^{(i)} \) which proves the result.

We now assume that \( m_1 < M - 1 \). We begin to prove that \( k_{p,m_1+1} = k_{p+1,m'+1} \). If this inequality does not hold, we derive from the inequality \( k_{p,m_1+1} \leq k_{p+1,m'+1} < k_{p,m_1+2} \), that \( k_{p,m_1+1} + 1 \leq k_{p+1,m'+1} \) and thus \( k_{p+1,m'+1} \in K_{p,m_1+1} \). Since \( \min(i,k_{p+1,m'+1}) > p + 1 \), we have

\[
\theta_{p+1}^{(i,k_{p+1,m'+1})} + \varepsilon_{p+1}^{(i,k_{p+1,m'+1})} \geq b_{p+1}^{(i)}.
\]

Hence,

\[
\theta_{p+1}^{(i,k_{p,m_1+1})} + \varepsilon_{p+1}^{(i,k_{p,m_1+1})} \geq b_{p+1}^{(i)} \quad \text{which implies} \quad \min(i,k_{p,m_1+1}) \geq p + 1.
\]
Since,
\[ k_{p+1,m'+1} = \inf \left\{ k > k_{p+1,m'}, \frac{i}{i_{\min}} > p + 1 \right\} \]
and \( k_{p,m+1} + 1 > k_{p+1,m'} \), we have \( k_{p+1,m'+1} \leq k_{p,m+1} + 1 \). Moreover, since \( k_{p+1,m'+1} \geq k_{p,m+1} + 1 \), we have \( k_{p,m+1} + 1 = k_{p+1,m'+1} \). Consequently,
\[ k_{p,m+2} = \inf \left\{ k > k_{p,m+1}, \frac{i}{i_{\min}} > p \right\} = k_{p+1,m'+1}. \]
This is impossible because \( k_{p+1,m'+1} < k_{p,m+2} \), which finally implies that \( k_{p,m+1} = k_{p+1,m'+1} \).

We then deduce from this equality,
\[ \frac{i}{i_{\min}} = \frac{i}{i_{\min}} > p + 1. \]
Hence \( \theta \geq b_{p+1} \) and thus \( t_{p+1} \in I \). This ends the proof.

Proof of Claim 7. We have
\[ k_{p+1,m'} = \sup \left\{ k < k_{p+1,m'+1}, \frac{i}{i_{\min}} > p + 1 \right\}. \]
Since \( k_{p,m+1} > k_{p+1,m'} \),
\[ k_{p+1,m'} = \sup \left\{ k < k_{p,m+1}, \frac{i}{i_{\min}} > p + 1 \right\} \]
\[ \leq \sup \left\{ k < k_{p,m+1}, \frac{i}{i_{\min}} > p \right\} \]
\[ \leq k_{p,m}. \]
We then derive from the inequalities \( k_{p+1,m'} \leq k_{p,m} \) and \( k_{p,m+1} \leq k_{p+1,m'+1} \) that \( K_{p,m} \subset K_{p+1,m'} \).

8. Annexe: implementation of the procedure when \( d \geq 2 \)

We carry out in the following sections the values of \( R_{\mathcal{C},j} \), \( \bar{R}_{\mathcal{C},j}(\mathbf{r}, \mathbf{r}') \) and \( \bar{R}_{\mathcal{C},j}(\mathbf{r}, \mathbf{r}') \) we have used in the simulation study of Section 6. We do not claim that they minimize the number of tests to compute. The number of tests that have been computed in the simulation study with these choices of parameters may be found in Section 8.7.

8.1. Example 1. In the case of the Gaussian model, it is worthwhile to notice that the Hellinger distance between two densities \( f_{(m, \sigma)} \) and \( f_{(m', \sigma')} \) can be made explicit:
\[ h^2(f_{(m, \sigma)}, f_{(m', \sigma')}) = 1 - \sqrt{\frac{2\sigma\sigma'}{\sigma^2 + \sigma'^2}} \exp \left( \frac{(m - m')^2}{4(\sigma^2 + \sigma'^2)} \right). \]
For all \( \xi > 0 \), a sufficient condition for that \( h^2(f_{(m, \sigma)}, f_{(m', \sigma')}) \leq \xi \) is thus
\[ \sqrt{\frac{2\sigma\sigma'}{\sigma^2 + \sigma'^2}} \geq \sqrt{1 - \xi} \quad \text{and} \quad \exp \left( \frac{(m - m')^2}{4(\sigma^2 + \sigma'^2)} \right) \geq \sqrt{1 - \xi}. \]
One then deduces that the rectangle
\[
\left[ m - 2 \frac{1 - \sqrt{2 \xi - \xi^2}}{1 - \xi} \sqrt{\log \left( \frac{1}{1 - \xi} \right)} \sigma, m + 2 \frac{1 - \sqrt{2 \xi - \xi^2}}{1 - \xi} \sqrt{\log \left( \frac{1}{1 - \xi} \right)} \sigma \right] 
\times \left[ 1 - \frac{\sqrt{2 \xi - \xi^2}}{1 - \xi} \sigma, 1 + \frac{\sqrt{2 \xi - \xi^2}}{1 - \xi} \sigma \right]
\]
is included in the Hellinger ball
\[
\{ (m', \sigma') \in \mathbb{R} \times (0, +\infty), \ h^2(f_{(m, \sigma)}, f_{(m', \sigma')}) \leq \xi \}.
\]
Given \( \theta = (m, \sigma), \ \theta' = (m', \sigma') \), we can then define \( \mathcal{L}_{c,j}(\theta, \theta') \), \( \bar{r}_{c,j}(\theta, \theta') \) by
\[
\mathcal{L}_{c}(\theta, \theta') = \left( 2 \frac{1 - \sqrt{2 \xi - \xi^2}}{1 - \xi} \sqrt{\log \left( \frac{1}{1 - \xi} \right)} \sigma, -\xi + \frac{\sqrt{2 \xi - \xi^2}}{1 - \xi} \sigma \right)
\]
\[
\bar{r}_{c}(\theta, \theta') = \left( 2 \frac{1 - \sqrt{2 \xi - \xi^2}}{1 - \xi} \sqrt{\log \left( \frac{1}{1 - \xi} \right)} \sigma, \xi + \frac{\sqrt{2 \xi - \xi^2}}{1 - \xi} \sigma \right)
\]
where \( \xi = \kappa H^2(f_{\theta}, f_{\theta'}) \).

We now consider a rectangle \( \mathcal{C} = [m_0, m_0] \times [\sigma_0, \sigma_0] \) of \( \mathbb{R} \times (0, +\infty) \) and aim at choosing \( \mathbf{R}_c = (\mathbf{R}_{c,1}, \mathbf{R}_{c,2}) \). For all \( (m, \sigma), (m', \sigma') \in \mathcal{C} \),
\[
h^2(f_{(m, \sigma)}, f_{(m', \sigma')}) = 1 - \sqrt{2 \sigma \sigma'} \sqrt{\frac{\sigma - \sigma'}{\sigma^2 + \sigma'^2}} + \sqrt{2 \sigma \sigma'} \left[ 1 - \exp \left( - \frac{(m - m')^2}{4(\sigma^2 + \sigma'^2)} \right) \right].
\]
Yet,
\[
1 - \sqrt{2 \sigma \sigma'} \leq \sqrt{\frac{(\sigma' - \sigma)^2}{\sigma^2 + \sigma'^2}} \frac{(m - m')^2}{4 \sigma^2_0 \sigma'^2_0} 
\]
and
\[
\sqrt{\frac{2 \sigma \sigma'}{\sigma^2 + \sigma'^2}} \geq \sqrt{\frac{2 \sigma_0 \sigma_0}{\sigma_0^2 + \sigma_0'^2}}.
\]
Moreover,
\[
1 - \exp \left( - \frac{(m - m')^2}{4(\sigma^2 + \sigma'^2)} \right) \geq 1 - e^{-\frac{(m_0 - m_0')^2}{2(\sigma_0^2 + \sigma_0'^2)}} (m' - m)^2.
\]
In particular, we have proved that
\[
h^2(f_{(m, \sigma)}, f_{(m', \sigma')}) \geq \max \left\{ \sqrt{\frac{2 \sigma_0 \sigma_0}{\sigma_0^2 + \sigma_0'^2}} \frac{1 - e^{-\frac{(m_0 - m_0')^2}{2(\sigma_0^2 + \sigma_0'^2)}}}{(m_0 - m_0)^2} (m' - m)^2, \frac{1}{4 \sigma_0^2} (\sigma' - \sigma)^2 \right\}
\]
which means that we can take
\[
\mathbf{R}_c = \left( \sqrt{\frac{2 \sigma_0 \sigma_0}{\sigma_0^2 + \sigma_0'^2}} \frac{1 - e^{-\frac{(m_0 - m_0')^2}{2(\sigma_0^2 + \sigma_0'^2)}}}{(m_0 - m_0)^2}, \frac{1}{4 \sigma_0^2} \right).
\]
8.2. Example 2. In the case of the Cauchy model, the Hellinger distance cannot be made explicit. However, we can use Theorem 7.6 of Ibragimov and Has’minskii (1981) (chapter 1) to show that for all \( m \in \mathbb{R}, \sigma > 0 \),
\[
  h^2(f(0,1), f(m,1)) \leq m^2/16 \quad \text{and} \quad h^2(f(0,1), f(0,\sigma)) \leq (\log^2 \sigma)/16.
\]
Now,
\[
  h(f(m,\sigma), f(m',\sigma')) \leq h(f(m,\sigma), f(m',\sigma)) + h(f(m',\sigma), f(m',\sigma')) \leq h(f(0,1), f((m'-m)/\sigma,1)) + h(f(0,1), f(0,\sigma'/\sigma)) \leq \frac{|m'-m|}{4\sigma} + \frac{\log(\sigma'/\sigma)}{4}.
\]
For all \( \xi > 0 \), one then deduces that the rectangle
\[
  \left[ m - 2\sigma \sqrt{\xi}, m + 2\sigma \sqrt{\xi} \right] \times \left[ \sigma e^{-2\sqrt{\xi}}, \sigma e^{2\sqrt{\xi}} \right]
\]
is included in the Hellinger ball
\[
  \{ (m', \sigma') \in \mathbb{R} \times (0, +\infty), h^2(f(m,\sigma), f(m',\sigma')) \leq \xi \}.
\]
This provides the values of \( \bar{r}_{C,j}(\theta, \theta') \) and \( \bar{r}_{C,j}(\theta, \theta') \): given \( C \subset \Theta \), \( \theta = (m, \sigma), \theta' = (m', \sigma') \in C \), we can take
\[
  \bar{r}_C(\theta, \theta') = \left( 2\sigma \sqrt{\kappa H^2(\theta, \theta')}, \sigma - \sigma e^{-2\sqrt{\kappa H^2(\theta, \theta')}} \right)
\]
\[
  \bar{r}_C(\theta, \theta') = \left( 2\sigma \sqrt{\kappa H^2(\theta, \theta')}, \sigma e^{2\sqrt{\kappa H^2(\theta, \theta')}} - \sigma \right).
\]
For all rectangle \( C \subset \mathbb{R} \times (0, +\infty) \) we choose \( \bar{R}_{C,1} = \bar{R}_{C,2} \). Notice that this choice allows to find easily the number \( k \) that appears at line 1 of Algorithm 2 since then the equation becomes
\[
  b_k - a_k = \max_{1 \leq j \leq 2} (b_j - a_j).
\]

8.3. Example 3. Let \( \xi > 0, a, b > 0 \) and \( C \) be the rectangle \( C = [a_1, a_2] \times [b_1, b_2] \subset (0, +\infty)^2 \). We aim at finding a rectangle \( R \) containing \( (a, b) \) such that
\[
  C \cap R \subset \{ (a', b') \in (0, +\infty)^2, h^2(f(a,b), f(a',b')) \leq \xi \}.
\]
For this, notice that for all positive numbers \( a', b' \),
\[
  h^2(f(a,b), f(a',b')) \leq 2h^2(f(a,b), f(a,b')) + 2h^2(f(a,b'), f(a',b')).
\]
Now,
\[
  h^2(f(a,b), f(a,b')) = 1 - \left( \frac{2\sqrt{bb'}}{b + b'} \right)^a.
\]
Let \( \Gamma' \) be the derivative of the Gamma function \( \Gamma \) and \( \psi \) be the derivative of the digamma function \( \Gamma'/\Gamma \). We derive from Theorem 7.6 of Ibragimov and Has’minskii (1981) that
\[
  h^2(f(a,b), f(a',b')) \leq \frac{(a' - a)^2}{8} \sup_{t \in [\min(a,a'), \max(a,a')]^2} \psi(t).
\]
The function $\psi$ being non-increasing,

$$h^2(f_{(a',b')}, f_{(a',b')}) \leq \begin{cases} 1/8 \psi(a)(a'-a)^2 & \text{if } a' \geq a \\ 1/8 \psi(a_1)(a'-a)^2 & \text{if } a' < a. \end{cases}$$

We deduce from the above inequalities that we can take

$$R = \left[ a - \sqrt{\frac{2}{\psi(a_1)}} \xi, a + \sqrt{\frac{2}{\psi(a)}} \xi \right] \times \left[ \frac{2 - \xi^2 - 2\sqrt{1 - \xi^2} \xi a}{\xi^2}, \frac{2 - \xi^2 + 2\sqrt{1 - \xi^2} \xi b}{\xi^2} \right]$$

where $\xi' = (1 - \xi/4)^{1/a}$. For all rectangle $C' \subset (0, +\infty)^2$ we define $R_{C',1} = R_{C',2}$.

### 8.4. Example 4.

As in the preceding example, we consider $\xi > 0$, $a, b > 0$ and the rectangle $C = [a_1, a_2] \times [b_1, b_2] \subset (0, +\infty)^2$. Our aim is to find a rectangle $R$ containing $(a, b)$ such that

$$C \cap R \subset \{(a', b') \in (0, +\infty)^2, h^2(f_{(a,b)}, f_{(a',b')}) \leq \xi\}.$$  

For all positive numbers $a', b'$,

$$h^2(f_{(a,b)}, f_{(a',b')}) \leq 2h^2(f_{(a,b)}, f_{(a',b')}) + 2h^2(f_{(a',b')}, f_{(a',b')}).$$

We derive from Theorem 7.6 of Ibragimov and Has'minskii (1981) that

$$h^2(f_{(a,b)}, f_{(a',b')}) \leq (b' - b)^2 \sup_{t \in [\min(b, b'), \max(b, b')]} |\psi(t) - \psi(a + t)|$$

where $\psi$ is defined in the preceding example. By using the monotony of the function $t \mapsto \psi(t) - \psi(a + t)$, we deduce that if $b' > b_1$,

$$h^2(f_{(a,b)}, f_{(a',b')}) \leq \begin{cases} 1/8 (\psi(b') - \psi(a + b)) (b' - b)^2 & \text{if } b' \geq b \\ 1/8 (\psi(b_1) - \psi(a + b_1)) (b' - b)^2 & \text{if } b' < b. \end{cases}$$

Similarly,

$$h^2(f_{(a',b')}, f_{(a',b')}) \leq \frac{(a' - a)^2}{8} \sup_{t \in [\min(a, a'), \max(a, a')]} |\psi(t) - \psi(b' + t)|.$$ 

Hence, if $a' \in [a_1, a_2]$ and $b' \in [b_1, b_2]$,

$$h^2(f_{(a',b')}, f_{(a',b')}) \leq \begin{cases} 1/8 (\psi(a) - \psi(a + b_2)) (a' - a)^2 & \text{if } a' \geq a \\ 1/8 (\psi(a_1) - \psi(a + b_2)) (a' - a)^2 & \text{if } a' < a. \end{cases}$$

We deduce from the above inequalities that we can take

$$R = \left[ a - \sqrt{\frac{2}{\psi(a_1) - \psi(a + b_2)}} \xi, a + \sqrt{\frac{2}{\psi(a) - \psi(a + b_2)}} \xi \right] \times \left[ b - \sqrt{\frac{2}{\psi(b_1) - \psi(a + b_1)}} \xi, b + \sqrt{\frac{2}{\psi(b) - \psi(a + b)}} \xi \right].$$

As in the two last examples, we take $R_{C',1} = R_{C',2}$ for all rectangle $C' \subset (0, +\infty)^2$. 
8.5. Example 5. For all \(m, m' \in \mathbb{R}, \lambda, \lambda' > 0\),

\[
h^2 (f_{(m, \lambda)}, f_{(m', \lambda')}) = \begin{cases} 
1 - \frac{2\sqrt{\lambda\lambda'}}{\lambda + \lambda'} e^{-\frac{\lambda}{\lambda + \lambda'} |m' - m|} & \text{if } m' \geq m \\
1 - \frac{2\sqrt{\lambda\lambda'}}{\lambda + \lambda'} e^{-\frac{\lambda}{\lambda + \lambda'} |m' - m|} & \text{if } m' \leq m.
\end{cases}
\]

We consider \(\xi > 0\) and aim at finding \(\mathcal{R}\) containing \((m, \lambda)\) such that

\[
\mathcal{R} \subset \{(m', \lambda') \in \mathbb{R} \times (0, +\infty), h^2 (f_{(m, \lambda)}, f_{(m', \lambda')}) \leq \xi\}.
\]

Notice that if \(m' \geq m\) and if

\[
\frac{2\sqrt{\lambda\lambda'}}{\lambda + \lambda'} \geq \sqrt{1 - \xi} \quad \text{and} \quad e^{-\frac{\lambda}{\lambda + \lambda'} (m' - m)} \geq \sqrt{1 - \xi}
\]

then \(h^2 (f_{(m, \lambda)}, f_{(m', \lambda')}) \leq \xi\). Similarly, if \(m' \leq m\) and if

\[
\frac{2\sqrt{\lambda\lambda'}}{\lambda + \lambda'} \geq \sqrt{1 - \xi} \quad \text{and} \quad e^{-\frac{\lambda}{\lambda + \lambda'} (m - m')} \geq \sqrt{1 - \xi}
\]

then \(h^2 (f_{(m, \lambda)}, f_{(m', \lambda')}) \leq \xi\). We can then take

\[
\mathcal{R} = \left[ m - \frac{1 - \xi}{1 + \xi + 2\sqrt{\xi}} \log \left( \frac{1/(1 - \xi)}{\lambda} \right), m + \frac{\log (1/(1 - \xi))}{\lambda} \right] \times \left[ \frac{1 + \xi - 2\sqrt{\xi}}{1 - \xi}, \frac{1 + \xi + 2\sqrt{\xi}}{1 - \xi} \right].
\]

Let now \(\mathcal{C}' = [\bar{m}_0, \bar{m}_0] \times [0, \bar{\lambda}_0]\) be a rectangle of \(\mathbb{R} \times (0, +\infty)\). By proceeding as in the Gaussian model, we can define \(\mathcal{R}_{\mathcal{C}}\) by

\[
\mathcal{R}_{\mathcal{C}} = (\mathcal{R}_{\mathcal{C}, 1}, \mathcal{R}_{\mathcal{C}, 2}) = \left( \frac{2\sqrt{\lambda_0 \lambda}}{\lambda + \lambda_0} \frac{1 - e^{-\frac{\lambda}{\lambda_0 + \lambda_0} (\bar{m}_0 - \bar{m}_0)}}{\bar{m}_0 - \bar{m}_0}, \frac{1}{8\lambda_0^2} \right).
\]

8.6. Example 6. For all \(m, m' \in \mathbb{R}, r, r' > 0\),

\[
h^2 (f_{(m, r)}, f_{(m', r')}) = 1 - \frac{(\min\{m + r, m' + r'\} - \max\{m, m'\})_+}{\sqrt{rr'}}
\]

where \((\cdot)_+\) is the positive part of \((\cdot)\). We consider \(\xi \in (0, \bar{k})\), and aim at finding a rectangle \(\mathcal{R}\) containing \((m, r)\) such that

\[
\mathcal{R} \subset \{(m', r') \in (0, +\infty)^2, h^2 (f_{(m, r)}, f_{(m', r')}) \leq \xi\}.
\]

For this, we begin to assume that \(m' \leq m + r\) and \(m' + r' \geq m\) to ensure that

\[
h^2 (f_{(m, r)}, f_{(m', r')}) = 1 - \frac{\min\{m + r, m' + r'\} - \max\{m, m'\}}{\sqrt{rr'}}.
\]

Several cases are involved

- If \(m' \leq m\) and \(m' + r' \geq m + r\), a sufficient condition for \(h^2 (f_{(m, r)}, f_{(m', r')}) \leq \xi\) is

\[
r' \leq \frac{1}{(1 - \xi)^2 r}.
\]

- If \(m' \geq m\) and \(m' + r' \leq m + r\), a sufficient condition is

\[
r' \geq (1 - \xi)^2 r.
\]
• If \( m' \leq m \) and if \( m' + r' \leq m + r \), a sufficient condition is

\[
m - m' \leq \left( \sqrt{r'} - (1 - \xi)\sqrt{r} \right) \sqrt{r'},
\]

which holds when

\[
r' \geq (1 - \xi/2)^2 r \quad \text{and} \quad |m' - m| \leq \xi/2 \sqrt{1 - \xi^2/2r}.
\]

• If \( m' \geq m \) and if \( m' + r' \geq m + r \), a sufficient condition is

\[
m' - m \leq \left( \sqrt{r} - (1 - \xi)\sqrt{r'} \right) \sqrt{r}.
\]

This condition is fulfilled when

\[
r' \leq \frac{1}{(1 - \xi/2)^2} r \quad \text{and} \quad |m' - m| \leq \frac{\xi}{2 - \xi} r.
\]

We can verify that if \((m', r')\) belongs to the rectangle

\[
\mathcal{R} = \left[ m - \frac{\xi\sqrt{2} - \xi}{2\sqrt{2}}, m + \frac{\xi}{2 - \xi} r \right] \times \left[ (1 - \xi/2)^2 r, \frac{r}{1 - \xi/2} \right]
\]

then \( m' \leq m + r \) and \( m' + r' \geq m \) (since \( \xi \leq \bar{\kappa} \)). The rectangle \( \mathcal{R} \) suits. For all rectangle \( \mathcal{C}' \subset (0, +\infty)^2 \) we choose in this example \( \mathcal{R}_{\mathcal{C}',1} = \mathcal{R}_{\mathcal{C}',2} \).

8.7. Speed of the procedure. By way of indication, we give below the number of tests that have been calculated in the simulation study of Section 6.

| Example  | \( n = 25 \) | \( n = 50 \) | \( n = 75 \) | \( n = 100 \) |
|----------|--------------|--------------|--------------|--------------|
| Example 1 | 1602 (117)   | 1577 (72)    | 1570 (60)    | 1567 (52)    |
| Example 2 | 2935 (90)    | 2937 (76)    | 2938 (69)    | 2938 (64)    |
| Example 3 | 9082 (1846)  | 8700 (1183)  | 8569 (934)   | 8511 (800)   |
| Example 4 | 10411 (778)  | 10272 (461)  | 10236 (357)  | 10222 (304)  |
| Example 5 | 6691 (296)   | 6699 (210)   | 6715 (175)   | 6726 (158)   |
| Example 6 | 32614 (1238) | 33949 (1211) | 34822 (1190) | 35397 (1177) |

Figure 7. Number of tests computed averaged over \( 10^4 \) samples and their corresponding standard deviations in brackets.

Acknowledgements: the author acknowledges the support of the French Agence Nationale de la Recherche (ANR), under grant Calibration (ANR 2011 BS01 010 01). We are thankful to Yannick Baraud for his valuable suggestions and careful reading of the paper.

References

Anatolyev, S. and Kosenok, G. (2005). An alternative to maximum likelihood based on spacings. *Econometric Theory*, 21(2):472–476.

Baraud, Y. (2011). Estimator selection with respect to Hellinger-type risks. *Probab. Theory Related Fields*, 151(1-2):353–401.

Baraud, Y. (2012). Estimation of the density of a determinantal process. *ArXiv e-prints*. 


Baraud, Y. and Birgé, L. (2009). Estimating the intensity of a random measure by histogram type estimators. *Probability Theory and Related Fields*, 143:239–284.

Barnett, V. D. (1966). Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. *Biometrika*, 53(1/2):151–165.

Basu, A. K., Harris, I. R., Hjort, N. L., and Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559.

Basu, A. K. and Lindsay, B. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, 46(4):683–705.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, 5(3):445–463.

Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l’estimation. *Probability Theory and Related Fields*, 65:181–237.

Birgé, L. (1984a). Stabilité et instabilité du risque minimax pour des variables indépendantes équidistribuées. *Annales de l’Institut Henri Poincaré. Probabilités et Statistique*, 20:201–223.

Birgé, L. (1984b). Sur un théorème de minimax et son application aux tests. *Probability and Mathematical Statistics*, 2:259–282.

Birgé, L. (2004). Model selection for Gaussian regression with random design. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 10(6):1039–1051.

Birgé, L. (2006). Model selection via testing: an alternative to (penalized) maximum likelihood estimators. *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, 42(3):273 – 325.

Birgé, L. (2007). Model selection for Poisson processes. In *Asymptotics: particles, processes and inverse problems*, volume 55 of *IMS Lecture Notes Monogr. Ser.*, pages 32–64. Inst. Math. Statist., Beachwood, OH.

Birgé, L. (2012). Robust tests for model selection. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. A Festschrift in Honor of Jon Wellner*, volume 9, pages 47–64. IMS Collections.

Birgé, L. (2013). Model selection for density estimation with L2-loss. *Probability Theory and Related Fields*, pages 1–42.

Cheng, R. and Amin, N. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3):394–403.

Dacunha-Castelle, D. (1978). Vitesse de convergence pour certains problèmes statistiques. In *École d’Été de Probabilités de Saint-Flour, VII (Saint-Flour, 1977)*, volume 678 of *Lecture Notes in Math.*, pages 1–172. Springer, Berlin.

Ekström, M. (1998). On the consistency of the maximum spacing method. *Journal of Statistical Planning and Inference*, 70(2):209–224.

Ferguson, T. S. (1982). An inconsistent maximum likelihood estimate. *Journal of the American Statistical Association*, 77(380):831–834.

Ghost, K. and Jammalamadaka, S. R. (2001). A general estimation method using spacings. *Journal of Statistical Planning and Inference*, 93(1):71–82.

Ibragimov, I. and Has’minskii, R. (1981). *Statistical estimation–asymptotic theory*. Applications of mathematics. Springer-Verlag.

Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, 1:38–53.
Le Cam, L. (1975). On local and global properties in the theory of asymptotic normality of experiments. In Stochastic processes and related topics (Proc. Summer Res. Inst. Statist. Inference for Stochastic Processes, Indiana Univ., Bloomington, Ind., 1974, Vol. 1; dedicated to Jerzy Neyman), pages 13–54. Academic Press, New York.

Le Cam, L. (1990). Maximum likelihood: an introduction. International Statistical Review, pages 153–171.

Lindsay, B. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. The Annals of Statistics, 22(2):1081–1114.

Pitman, E. (1979). Some basic theory for statistical inference. Monographs on applied probability and statistics. Chapman and Hall.

Ranneby, B. (1984). The maximum spacing method. an estimation method related to the maximum likelihood method. Scandinavian Journal of Statistics, 11(2):93–112.

Sart, M. (2013a). Estimation of the transition density of a markov chain. Annales de l’Institut Henri Poincaré. Probabilités et Statistique. To appear.

Sart, M. (2013b). Model selection for poisson processes with covariates. ArXiv e-prints.

Shao, Y. and Hahn, M. G. (1999). Strong consistency of the maximum product of spacings estimates with applications in nonparametrics and in estimation of unimodal densities. Annals of the Institute of Statistical Mathematics, 51(1):31–49.

Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France.

E-mail address: msart@unice.fr