Bernoulli **22**(3), 2016, 1617–1670
DOI: 10.3150/15-BEJ706

Robust estimation on a parametric model via testing

MATHIEU SART

Université de Lyon, Université Jean Monnet, CNRS UMR 5208 and Institut Camille Jordan, Maison de l'Université, 10 rue Tréfilerie, CS 82301, 42023 Saint-Etienne Cedex 2, France.
E-mail: mathieu.sart@univ-st-etienne.fr

We are interested in the problem of robust parametric estimation of a density from n 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 show that the estimator is robust even for models for which the maximum likelihood method is bound to fail. A numerical simulation illustrates its robustness properties. When the model is true and regular enough, we prove that the estimator is very close to the maximum likelihood one, at least when the number of observations n is large. In particular, it inherits its efficiency. Simulations show that these two estimators are almost equal with large probability, even for small values of n when the model is regular enough and contains the true density.

Keywords: parametric estimation; robust estimation; robust tests; T-estimator

1. Introduction

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

1.1. About the maximum likelihood estimator

The maximum likelihood method is one of the most widespread estimation methods to deal with this statistical setting. Indeed, it is well known that it provides estimators with nice statistical properties when the parametric model is true and regular enough.

Nevertheless, it is also recognized that it breaks down for many parametric models $\mathscr{F}$ of interest. A simple one is the translation model $\mathscr{F} = \{f(\cdot - \theta), \theta \in \Theta\}$ where $\lim_{x \to 0} f(x) = +\infty$, in which the maximum likelihood estimator (m.l.e. for short) does not exist. Other counterexamples may be found in Pitman [23], Ferguson [18], Le Cam [21], Birgé [9] among other references.

Another known defect of the m.l.e. is its lack of robustness. This means that if the assumption that s belongs to the parametric model $\mathscr{F}$ is only slightly violated, the m.l.e. may perform poorly. As an example, consider the model $\mathscr{F} = \{\theta^{-1} \mathbb{1}_{[0,\theta]}, \theta > 0\}$, in which the maximum likelihood estimator is $\hat{\theta}_{\mathrm{mle}}^{-1} \mathbb{1}_{[0,\hat{\theta}_{\mathrm{mle}}]}$ with $\hat{\theta}_{\mathrm{mle}} = \max_{1 \le i \le n} X_i$. Suppose that the true density s

does not belong to $\mathscr{F}$ but lies in a very small neighbourhood of it. For instance, assume that $s = (1-p)\mathbb{1}_{[0,1]} + p2^{-1}\mathbb{1}_{[0,2]}$ for some $p \in (0,1)$. If p is very small, the true underlying density s is very close to $\mathbb{1}_{[0,1]} \in \mathscr{F}$ and a good estimator $\hat{f}$ of s should therefore be also close to $\mathbb{1}_{[0,1]}$, at least when n is large enough and p is small enough. Nonetheless, whatever $p > 0$, the estimator $\hat{\theta}_{\mathrm{mle}}^{-1}\mathbb{1}_{[0,\hat{\theta}_{\mathrm{mle}}]}$ converges almost surely to $2^{-1}\mathbb{1}_{[0,2]}$ when n goes to infinity. It is thus a very poor estimate of s when p is small.

1.2. Alternative estimators

Several attempts have been made in the literature to overcome the difficulties of the maximum likelihood approach. When the model is regular enough, the classical notion of efficiency can be used to measure the quality of an estimator (when the model is not regular enough, the optimal rate of convergence may not be the usual root-n rate). For these models, the L-estimators commonly accomplish a good trade-off between robustness and efficiency. Some estimators have the nice feature to be simultaneously robust and asymptotically efficient. This is the case, for example, of the minimum Hellinger distance estimators introduced by Beran [8] and studied in Donoho and Liu [16], Lindsay [22] among other references. We refer to Basu et al. [7] for an introduction to these estimators.

Things become more complicated when the model is less regular and even more when the maximum likelihood estimators do not even exist. We do not know if the aforementioned estimation strategies can be adapted to cope with these models in a satisfactory way. Building a robust and optimal estimator is not straightforward in some models (where "optimal" means that it achieves the optimal rate of convergence when the model holds true). Think, for instance, about the translation model $\mathscr{F} = \{f(\cdot - \theta), \theta \in [-1,1]\}$ where

$$f(x) = \begin{cases} \dfrac{1}{4\sqrt{|x|}}\mathbb{1}_{[-1,1]}(x), & \text{for all } x \in \mathbb{R} \setminus \{0\}, \\ 0, & \text{for } x = 0. \end{cases} \tag{1}$$

The median is a natural robust estimator, but it converges slowly to the right parameter since it only reaches the rate n^{-1} whereas the optimal one is n^{-2}.

1.3. Estimation via testing

There is in the literature a more or less universal strategy of estimation that leads to robust and optimal estimators. It even manages to deal with models for which the maximum likelihood method is bound to fail. Its basic principle is to use tests to derive estimators. Historically, this idea of using tests for building estimators dates back to the 1970s with the works of Lucien Le Cam. More recently, Birgé [9] significantly extended the scope of these procedures by relating them to the problem of model selection, providing at the same time new perspectives on estimation theory. It gave birth to a series of papers; see Birgé [10–12], Baraud and Birgé [4], Baraud [2,3], Sart [26,27], Baraud et al. [5]. The main feature of these procedures is that they allow to obtain 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).

In density estimation, these papers show that under very mild assumptions on the parametric model $\mathscr{F} = \{f_\theta, \theta \in \Theta\}$, one can design an estimator $\hat{s} = f_{\hat{\theta}}$ such that

$$\mathbb{P}\left[Ch^2(s, f_{\hat{\theta}}) \geq \inf_{\theta \in \Theta} h^2(s, f_\theta) + \frac{D_{\mathscr{F}}}{n} + \xi\right] \leq \mathrm{e}^{-n\xi} \qquad \text{for all } \xi > 0, \tag{2}$$

where C is a numerical positive constant, h the Hellinger distance, and $D_{\mathscr{F}}$ measures, in some sense, the "massiveness" of $\mathscr{F}$. We recall that the Hellinger distance is defined on the cone $\mathbb{L}^1_+(\mathbb{X}, \mu)$ of non-negative integrable functions on $\mathbb{X}$ with respect to μ by

$$h^2(f, g) = \frac{1}{2} \int_{\mathbb{X}} \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 \mathrm{d}\mu(x) \qquad \text{for all } f, g \in \mathbb{L}^1_+(\mathbb{X}, \mu).$$

When s does belong to the model $\mathscr{F}$, that is, when there exists $\theta_0 \in \Theta$ such that $s = f_{\theta_0}$, the estimator $\hat{s}$ achieves a quadratic risk of order n^{-1} with respect to the Hellinger distance. Besides, if we can relate the Hellinger distance $h(f_{\theta_0}, f_\theta)$ to a distance between the parameters θ_0, θ, the convergence rate of $\hat{\theta}$ to θ_0 may be deduced from (2). For instance, when $\Theta \subset \mathbb{R}$, and when there exists $\alpha > 0$ such that $h^2(f_{\theta_0}, f_\theta) \sim |\theta_0 - \theta|^\alpha$, the estimator $\hat{\theta}$ reaches the rate $n^{-1/\alpha}$. When the model is regular enough, $h^2(f_{\theta_0}, f_\theta) \sim |\theta_0 - \theta|^2$, and the estimator $\hat{\theta}$ attains the usual root-n rate.

It is worth mentioning that one does not have to assume that the unknown density s belongs to the model, which is important since one cannot usually ensure that this is the case in practice. We rather use the model $\mathscr{F}$ as an approximating class (sieve) for s. Inequality (2) shows that the estimator $\hat{s} = f_{\hat{\theta}}$ cannot be strongly influenced by any type of small departures from the model (measured through the Hellinger distance). As a matter of fact, if $\inf_{\theta \in \Theta} h^2(s, f_\theta) \leq an^{-1}$ with $a > 0$, 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 (that is, not shared by the m.l.e.).

1.4. The purposes of this paper

One of the most annoying drawbacks of the estimators based on tests is that their practical construction is numerically very difficult. Two steps are required to build these estimators. In the first step, we discretize the model $\mathscr{F}$, that is, we build a thin net $\mathscr{F}_{\text{dis}}$ in $\mathscr{F}$ that must be finite or countable. In the second step, we use the tests to pairwise compare the elements of $\mathscr{F}_{\text{dis}}$. Therefore, the number of tests we need to compute is of the order of the square of the cardinality of $\mathscr{F}_{\text{dis}}$. Unfortunately, this cardinality is often very large, making the construction of the estimators difficult in practice. In this paper, we present a new estimation procedure based on the test designed by Baraud [2] and on an iterative construction of confidence sets. This procedure does not involve the pairwise comparison of all the elements of $\mathscr{F}_{\text{dis}}$ but only of a small (random and suitably chosen) part of them, which results in a significant reduction of the numerical complexity. In particular, this makes it possible to evaluate the quality of the estimator by means of numerical simulations in situations where the procedure of Baraud [2] would have required the computation of an intractable number of tests.

This estimation procedure outperforms the maximum likelihood one in many aspects. Similarly to the procedure of Baraud [2], the estimator $\hat{s} = f_{\hat{\theta}}$ exists in parametric models where the m.l.e. does not. We establish a risk bound akin to (2). In particular, when the model $\mathscr{F}$ is true, that is, when there exists $\theta_0 \in \Theta$ such that $s = f_{\theta_0} \in \mathscr{F}$, the estimator $\hat{\theta}$ converges to the true parameter θ_0 at the right rate of convergence. When the model is only approximately true, which means that the Hellinger distance between s and the model $\mathscr{F}$ is small, the estimator $\hat{s}$ of s still performs well.

An additional significant property of this estimator is that it essentially coincides with the m.l.e. (with large probability), when the model is true and regular enough, even when the number of observations n is small. It seems to be, in this case, as good as the m.l.e. This property was brought to light by numerical simulations in the first draft of this paper. During the revision process, an asymptotic theoretical connection between an estimator based on tests and the m.l.e. was established, for the first time, in Theorem 4 of Baraud et al. [5]. The techniques developed in this paper helped us to prove theoretically that our estimator was asymptotically very close to the m.l.e. and that it inherited in particular its nice asymptotic properties such as efficiency (at least under suitable regularity assumptions on the model $\mathscr{F}$). These regularity assumptions are however different from theirs. They may therefore hold true in some parametric models where those of Baraud et al. [5] do not.

1.5. Organization of the paper and notations

For the sake of clarity, we start by considering models parametrized by a one-dimensional parameter. In Section 2, we present our procedure and the associated theoretical results. We evaluate its performance in practice by carrying out numerical simulations in the next section. We study the multi-dimensional case in Section 4. We postpone the main proofs to Section 5 except the one of Theorem 4.1 which is quite technical and deferred to the Appendix.

We now introduce some notation that will be used all along the paper. The number $x \vee y$ stands for $\max(x, y)$ and x_+ stands for $x \vee 0$. We set $\mathbb{N}^\star = \mathbb{N} \setminus \{0\}$. The vector $(\theta_1, \dots, \theta_d)$ of $\mathbb{R}^d$ is denoted by the bold letter $\boldsymbol{\theta}$. We write indifferently $h(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'})$ or $h(\boldsymbol{\theta}, \boldsymbol{\theta}')$. 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 $\mathbb{1}_A$. The notation C, C', C'' stand for quantities independent of n. When they depend on other parameters, this dependency will be specified in the text. The values of $C, C', C'', \dots$ may change from line to line.

2. Models parametrized by a one-dimensional parameter

2.1. Assumption on the model

We start by considering sets of densities $\mathscr{F} = \{f_\theta, \theta \in \Theta\}$ indexed by a finite interval $\Theta = [m, M]$ of $\mathbb{R}$. Such a set will be called a one-dimensional model. Throughout this section, the models are assumed to satisfy the following property.

Assumption 2.1. *There exist positive numbers α, $\underline{R}$, $\overline{R}$ such that for all $\theta, \theta' \in [m, M]$,*

$$\underline{R}|\theta - \theta'|^{\alpha} \le h^2(\theta, \theta') \le \overline{R}|\theta - \theta'|^{\alpha},$$

where $h(\theta, \theta')$ stands for the Hellinger distance $h(f_\theta, f_{\theta'})$ between the two densities f_θ and $f_{\theta'}$.

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 [20] to prove results on the maximum likelihood estimator. They require, however, the application $\theta \mapsto f_\theta(x)$ to be continuous for μ-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 $\mathscr{F} = \{f(\cdot - \theta), \theta \in [-1, 1]\}$ where f is defined in the Introduction by (1) (note that Assumption 2.1 holds for this model with $\alpha = 1/2$).

Under suitable regularity conditions on the model, Theorem 7.6 of Chapter 1 of Ibragimov and Has'minskii [20] shows that this assumption is fulfilled with $\alpha = 2$. Other kinds of sufficient conditions implying Assumption 2.1 may be found in this book (see the beginning of Chapter 5 and Theorem 1.1 of Chapter 6). Other examples and counterexamples are given in Chapter 7 of Dacunha-Castelle [15]. Several models of interest satisfying this assumption will appear later in the paper.

2.2. Basic ideas

We now present the heuristic on which our estimation procedure is based. We assume in this section that s belongs to the model $\mathscr{F}$, 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 $X_1, \dots, X_n$ 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 $h^2(\theta_0, \theta) > \kappa h^2(\theta, \theta')$.
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 θ with radius r, that is,

$$\mathcal{B}(\theta, r) = \{\theta' \in \Theta, h(\theta, \theta') \le r\}. \tag{3}$$

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

$$\theta_0 \in \Theta \setminus \mathcal{B}(\theta, \kappa^{1/2} h(\theta, \theta'))$$

while in the second case

$$\theta_0 \in \Theta \setminus \mathcal{B}(\theta', \kappa^{1/2} h(\theta, \theta')).$$

These sets may be interpreted as confidence sets for θ_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 Θ). If $T(\theta^{(1)}, \theta'^{(1)})$ is non-negative, we consider a set Θ_2 such that

$$\Theta_1 \setminus \mathcal{B}\big(\theta^{(1)}, \kappa^{1/2} h\big(\theta^{(1)}, \theta'^{(1)}\big)\big) \subset \Theta_2 \subset \Theta_1$$

while if $T(\theta^{(1)}, \theta'^{(1)})$ is non-positive, we consider a set Θ_2 such that

$$\Theta_1 \setminus \mathcal{B}\big(\theta'^{(1)}, \kappa^{1/2} h\big(\theta^{(1)}, \theta'^{(1)}\big)\big) \subset \Theta_2 \subset \Theta_1.$$

The set Θ_2 may thus also be interpreted as a confidence set for θ_0. Thanks to Assumption 2.1, we can define Θ_2 as an interval $\Theta_2 = [\theta^{(2)}, \theta'^{(2)}]$.

We then repeat the construction to build an interval $\Theta_3 = [\theta^{(3)}, \theta'^{(3)}]$ included in Θ_2 such that either

$$\Theta_3 \supset \Theta_2 \setminus \mathcal{B}\big(\theta^{(2)}, \kappa^{1/2} h\big(\theta^{(2)}, \theta'^{(2)}\big)\big) \quad \text{or} \quad \Theta_3 \supset \Theta_2 \setminus \mathcal{B}\big(\theta'^{(2)}, \kappa^{1/2} h\big(\theta^{(2)}, \theta'^{(2)}\big)\big)$$

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 Θ_N is small enough. We then define the estimator $\hat{\theta}$ as the center of the set Θ_N and estimate s by $f_{\hat{\theta}}$.

2.3. Definition of the test

The test $T(\theta, \theta')$ we use in our estimation strategy is the one of Baraud [2] applied to two suitable densities of the model. More precisely, let $\overline{T}$ be the functional defined for all $g, g' \in \mathbb{L}^1_+(\mathbb{X}, \mu)$ by

$$\overline{T}\big(g, g'\big) = \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{g'(X_i)} - \sqrt{g(X_i)}}{\sqrt{g(X_i) + g'(X_i)}} + \frac{1}{2} \int_{\mathbb{X}} \sqrt{g(x) + g'(x)}\big(\sqrt{g'(x)} - \sqrt{g(x)}\big) \, \mathrm{d}\mu(x), \quad (4)$$

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

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

$$\Theta_{\mathrm{dis}} = \big\{m + k\varepsilon, k \in \mathbb{N}, k \le (M - m)\varepsilon^{-1}\big\}, \qquad \mathscr{F}_{\mathrm{dis}} = \{f_\theta, \theta \in \Theta_{\mathrm{dis}}\}$$

and the map π on $[m, M]$ by

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

where $\lfloor \cdot \rfloor$ denotes the integer part. The test $T(\theta, \theta')$ is finally defined by

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

The aim of the parameter t is to tune the thinness of the net $\mathscr{F}_{\mathrm{dis}}$. The smaller t, the thinner $\mathscr{F}_{\mathrm{dis}}$.

2.4. Estimation procedure

We shall build a decreasing sequence $(\Theta_i)_{i\geq 1}$ of intervals of $\Theta = [m, M]$ as explained in Section 2.2. Let $\kappa > 0$, and for all $\theta, \theta' \in [m, M]$ such that $\theta' > \theta$, let $\overline{r}(\theta, \theta')$, $\underline{r}(\theta, \theta')$ be two positive numbers satisfying

$$[m, M] \cap \left[\theta, \theta + \overline{r}\left(\theta, \theta'\right)\right] \subset \mathcal{B}\left(\theta, \kappa^{1/2} h\left(\theta, \theta'\right)\right), \tag{5}$$

$$[m, M] \cap \left[\theta' - \underline{r}\left(\theta, \theta'\right), \theta'\right] \subset \mathcal{B}\left(\theta', \kappa^{1/2} h\left(\theta, \theta'\right)\right), \tag{6}$$

where we recall that $\mathcal{B}(\theta, \kappa^{1/2} h(\theta, \theta'))$ and $\mathcal{B}(\theta', \kappa^{1/2} h(\theta, \theta'))$ 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

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

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

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

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

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

The parameter κ 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 Θ_{i+1} in Θ_i.

We now consider a positive number η and build these intervals until their lengths become smaller than η. The estimator is then defined as the center of the last interval. This parameter η stands for a measure of the accuracy of the estimation and must be small enough to get a suitable risk bound for the estimator. The algorithm is therefore the following.

Algorithm 1

$\theta \leftarrow m$, $\theta' \leftarrow M$
while $\theta' - \theta > \eta$ **do**
    Compute $r = \min\{\overline{r}(\theta, \theta'), (\theta' - \theta)/2\}$
    Compute $r' = \min\{\underline{r}(\theta, \theta'), (\theta' - \theta)/2\}$
    Compute Test $= T(\theta, \theta')$
    if Test ≥ 0 **then**
        $\theta \leftarrow \theta + r$

end if
if Test ≤ 0 **then**
$\theta' \leftarrow \theta' - r'$
end if
end while
Return: $\hat{\theta} = (\theta + \theta')/2$

The convergence of the algorithm is guaranteed under very mild conditions on $\overline{r}(\theta, \theta')$ and $\underline{r}(\theta, \theta')$. For instance, a sufficient condition is that the functions $\overline{r}(\cdot, \cdot)$, $\underline{r}(\cdot, \cdot)$ are positive and continuous on the set $\{(\theta, \theta'), m \leq \theta < \theta' \leq M\}$. Moreover, its numerical complexity can be bounded as soon as $\overline{r}(\theta, \theta')$ and $\underline{r}(\theta, \theta')$ are large enough as we shall see in Section 2.7.

2.5. A non-asymptotic risk bound

The following theorem specifies the values of the parameters t, κ, η that allow to control the risk of the estimator $\hat{s} = f_{\hat{\theta}}$.

Theorem 2.1. *Suppose that Assumption 2.1 holds. Set*

$$\bar{\kappa} = 3/2 - \sqrt{2}. \tag{7}$$

Assume that $t \in (0, 1]$, $\kappa \in (0, \bar{\kappa})$, $\eta \in (0, (\overline{R}n)^{-1/\alpha}]$ and that $\overline{r}(\theta, \theta')$, $\underline{r}(\theta, \theta')$ are such that (5) and (6) hold and that the algorithm converges.

Then, for all $\xi > 0$, the estimator $\hat{\theta}$ derived from Algorithm 1 satisfies

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

where $D_{\mathscr{F}} = 1 \vee \log(1 + t^{-1}((1/\alpha)(c\overline{R}/\underline{R}))^{1/\alpha})$ with c depending only on κ, and where $C > 0$ depends only on κ and $\overline{R}/\underline{R}$. Besides, if

$$h^2(\theta_2, \theta_2') \leq h^2(\theta_1, \theta_1') \qquad \text{for all } m \leq \theta_1 \leq \theta_2 < \theta_2' \leq \theta_1' \leq M$$

then C depends only on κ.

We deduce from this risk bound that if $s = f_{\theta_0}$ belongs to the model $\mathscr{F}$, the estimator $\hat{\theta}$ converges almost surely to θ_0. Besides, we may then derive from Assumption 2.1 that there exist positive numbers a, b such that

$$\mathbb{P}\left[n^{1/\alpha}|\hat{\theta} - \theta_0| \geq \xi\right] \leq a e^{-b\xi^\alpha} \qquad \text{for all } \xi > 0.$$

We emphasize here that this exponential inequality on $\hat{\theta}$ is non-asymptotic but that the numbers a and b are, unfortunately, far from optimal (since their values depend on several parameters involved in the algorithm such as t or κ). As explained in the Introduction, this theorem also shows that the estimator $\hat{s}$ possesses robustness properties with respect to the Hellinger distance.

2.6. Connection with the maximum likelihood estimator

When the model is true and regular enough, the above theorem states that $\sqrt{n}(\hat{\theta} - \theta_0)$ is sub-Gaussian (since in this case Assumption 2.1 holds with $\alpha = 2$). Actually, in favourable situations, $\hat{\theta}$ shares the nice asymptotic properties of the m.l.e., and in particular its efficiency.

Theorem 2.2. *Suppose that the model $\mathscr{F}$ satisfies the following conditions*:

(i) *There exists $\theta_0 \in (m, M)$ such that $s = f_{\theta_0} \in \mathscr{F}$.*

(ii) *The model is identifiable, that is, for all $\theta \neq \theta'$, $f_\theta \neq f_{\theta'}$.*

(iii) *For μ-almost all $x \in \mathbb{X}$, the mapping $\theta \mapsto f_\theta(x)$ is continuous and positive on $[m, M]$ and two times differentiable on (m, M). Its first and second derivatives are denoted, respectively, by $\dot{f}_\theta(x)$ and $\ddot{f}_\theta(x)$. For μ-almost all $x \in \mathbb{X}$, the function $\theta \mapsto \dot{f}_\theta(x)$ can be extended by continuity to $[m, M]$.*

(iv) *For all $\theta \in [m, M]$, the Fisher information*

$$I(\theta) = \int_{\mathbb{X}} \left(\dot{l}_\theta(x)\right)^2 f_\theta(x)\,\mathrm{d}\mu(x) \qquad \text{with } \dot{l}_\theta(x) = \frac{\partial \log f_\theta(x)}{\partial \theta}$$

is non-zero and satisfies $\sup_{\theta \in [m,M]} I(\theta) < \infty$. Moreover, $\theta \mapsto I(\theta)$ is continuous at θ_0.

(v) *The integrals $\int_{\mathbb{X}} \dot{f}_{\theta_0}(x)\,\mathrm{d}\mu(x)$, $\int_{\mathbb{X}} \ddot{f}_{\theta_0}(x)\,\mathrm{d}\mu(x)$ exist and are zero.*

(vi) *There exist two positive functions φ_1, φ_2 and two numbers $\gamma_1 > 2/3$, $\gamma_2 > 0$ such that for all $\theta, \theta' \in (m, M)$ and μ-almost all $x \in \mathbb{X}$,*

$$\left|\log f_{\theta'}(x) - \log f_\theta(x)\right| \le \varphi_1(x)\left|\theta' - \theta\right|^{\gamma_1},$$
$$\left|\ddot{l}_{\theta'}(x) - \ddot{l}_\theta(x)\right| \le \varphi_2(x)\left|\theta' - \theta\right|^{\gamma_2},$$

where $\ddot{l}_\theta(x)$ stands for the second derivative of $\theta \mapsto \log f_\theta(x)$. Moreover, $\mathbb{E}[\varphi_1^3(X_1)]$ and $\mathbb{E}[\varphi_2(X_1)]$ are finite.

Furthermore, assume the following conditions on the algorithm:

(vii) *The parameter t depends on n (one then writes $t^{(n)}$ in place of t) and $t^{(n)}$ tends to 0 in such a way that $|\log t^{(n)}| = \mathrm{o}(n)$ when n goes to infinity. The positive parameter η depends on n and is smaller than $t^{(n)}(\overline{R}n)^{-1/2}$.*

(viii) *The parameter $\kappa \in (0, \bar{\kappa})$ is chosen independently of n, the parameters $\overline{r}(\theta, \theta')$, $\underline{r}(\theta, \theta')$ are chosen in such a way that (5) and (6) hold and that the algorithm converges.*

Then Assumption 2.1 holds with $\alpha = 2$ and there exist $C > 0$ (that may depend on κ and $\underline{R}$ but not on n) and a sequence $(\zeta_n)_{n\ge1}$ in $[0, 1]$ converging to 0 such that

$$\mathbb{P}\left[\exists \tilde{\theta} \in (m, M), \sum_{i=1}^{n} \dot{l}_{\tilde{\theta}}(X_i) = 0 \text{ and } |\hat{\theta} - \tilde{\theta}| \le C\frac{t^{(n)}}{\sqrt{n}}\right] \ge 1 - \zeta_n.$$

In particular, $\hat{\theta}$ is asymptotically efficient, that is, $\sqrt{n}(\hat{\theta} - \theta_0)$ converges in distribution to a normal distribution with mean zero and variance $1/I(\theta_0)$. Moreover, if there exists $\lambda > 0$ such that $\mathbb{E}[\exp(\lambda\varphi_2(X_1))]$, $\mathbb{E}[\exp(\lambda|\dot{l}_{\theta_0}(X_1)|)]$ and $\mathbb{E}[\exp(\lambda|\ddot{l}_{\theta_0}(X_1)|)]$ are finite, then there exists $b > 0$ such that the sequence $(\zeta_n \exp(bn))_{n\ge1}$ is bounded above.

The main interest of $\hat{\theta}$ as compared to the m.l.e. when the model is regular enough lies in the fact that one usually does not know whether s belongs to the model or not. If the model is true, $\hat{\theta}$ inherits the nice asymptotic statistical properties of the m.l.e. However, it possesses robustness properties with respect to the Hellinger distance, which is definitively not the case for the m.l.e.

Remark. When the model is regular enough but does not contain the unknown density s, the theoretical properties of the estimator $\hat{\theta}$ are only guaranteed by Theorem 2.1. When $t = t^{(n)}$ depends on n and satisfies the assumptions of Theorem 2.2, the term $D_{\mathscr{F}}/n$ appearing in Theorem 2.1 converges to 0, but at a rate slower than $1/n$. It is, for instance, of the order of $\log n/n$ when $t^{(n)} = a/n^k$ with $a > 0$, $k > 0$. This deteriorates the risk bound and this could get worse since we may make this rate of convergence arbitrarily slow by playing with $t^{(n)}$. We conjecture that this phenomenon is due to technical difficulties and that the estimator remains good even when $t = t^{(n)}$ is arbitrarily small or even zero (that is, with $\mathscr{F}_{\text{dis}} = \mathscr{F}$) as suggested by the numerical simulations (in Section 3).

2.7. Numerical complexity

The numerical complexity of the estimation procedure depends on several parameters (η, κ, $\overline{r}(\theta, \theta')$, $\underline{r}(\theta, \theta')$) that must be chosen by the statistician (since they are involved in the algorithm).

The role of the parameter η is to stop the algorithm when the confidence sets are small enough. Consequently, the smaller η, the longer it takes to compute the estimator. Nevertheless, we shall see at the end of this section that the time of construction of the estimator grows slowly when η decreases.

The parameter κ tunes the level of the confidence sets, and thus also the speed of the procedure: the larger κ, the faster the procedure. Note, however, that the preceding theorems require that κ be smaller than $\bar{\kappa}$. There is no theoretical guarantee when κ is larger than $\bar{\kappa}$.

The values of the parameters $\overline{r}(\theta, \theta')$, $\underline{r}(\theta, \theta')$ do not change the theoretical statistical properties of the estimator given by Theorems 2.1 and 2.2 (provided that (5) and (6) hold) but strongly influence its construction time. The larger they are, the faster the procedure is. There are three different situations:

First case: The Hellinger distance $h(\theta, \theta')$ can be made explicit. We have thus an interest in defining them as the largest numbers for which (5) and (6) hold, that is,

$$\overline{r}(\theta, \theta') = \sup\{r > 0, [m, M] \cap [\theta, \theta + r] \subset \mathcal{B}(\theta, \kappa^{1/2} h(\theta, \theta'))\}, \tag{8}$$

$$\underline{r}(\theta, \theta') = \sup\{r > 0, [m, M] \cap [\theta' - r, \theta'] \subset \mathcal{B}(\theta', \kappa^{1/2} h(\theta, \theta'))\}. \tag{9}$$

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

$$\underline{r}(\theta, \theta') = \overline{r}(\theta, \theta') = \left((\kappa/\overline{R}) h^2(\theta, \theta')\right)^{1/\alpha}. \tag{10}$$

One can verify that (5) and (6) hold. When the model is regular enough and $\alpha = 2$, the value of $\overline{R}$ can be calculated by using Fisher information [see, for instance, Theorem 7.6 of Chapter 1 of Ibragimov and Has'minskii [20]].

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

$$\underline{r}(\theta,\theta') = \overline{r}(\theta,\theta') = (\kappa \underline{R}/\overline{R})^{1/\alpha}(\theta'-\theta). \tag{11}$$

As in the second case, 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 $\mathscr{F} = \{f(\cdot - \theta), \theta \in [m, M]\}$ with f even, $\mathbb{X} = \mathbb{R}$ and μ the Lebesgue measure (the second term of (4) is 0 for these models).

We can upper bound the numerical complexity of the algorithm when $\overline{r}(\theta,\theta')$ and $\underline{r}(\theta,\theta')$ are large enough. More precisely, we have the following.

Proposition 2.1. *Suppose that the assumptions of Theorem 2.1 hold and that $\underline{r}(\theta,\theta')$, $\overline{r}(\theta,\theta')$ are larger than*

$$(\kappa \underline{R}/\overline{R})^{1/\alpha}(\theta'-\theta). \tag{12}$$

Then the algorithm converges in less than

$$1 + \max\{(\overline{R}/(\kappa \underline{R}))^{1/\alpha}, 1/\log 2\} \log\left(\frac{M-m}{\eta}\right)$$

iterations.

This is an improvement with respect to the procedure of Baraud [2] where the number of tests computed is roughly of the order of $|\mathscr{F}_{\mathrm{dis}}|^2$, which is much larger than the above bound when ε is small enough (and $\eta = \varepsilon$).

3. Simulations for one-dimensional models

In what follows, we carry out a simulation study in order to investigate more precisely the performance of our estimator. We simulate samples $(X_1, \dots, X_n)$ with density s and use our procedure to estimate s.

3.1. Models

Our simulation study is based on the following models.

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

Example 2. $\mathscr{F} = \{f(\cdot - \theta), \theta \in [-100, 100]\}$ where f is the density of a standard Gaussian distribution.

Example 3. $\mathscr{F} = \{f(\cdot - \theta), \theta \in [-10, 10]\}$ where f is the density of a standard Cauchy distribution.

Example 4. $\mathscr{F} = \{f_\theta, \theta \in [0.01, 10]\}$ where $f_\theta = \theta^{-1}\mathbb{1}_{[0,\theta]}$.

Example 5. $\mathscr{F} = \{f_\theta, \theta \in [-10, 10]\}$ where $f_\theta(x) = \frac{1}{(x-\theta+1)^2}\mathbb{1}_{[\theta,+\infty)}(x)$ for all $x \in \mathbb{R}$.

Example 6. $\mathscr{F} = \{\mathbb{1}_{[\theta-1/2,\theta+1/2]}, \theta \in [-10, 10]\}$.

Example 7. $\mathscr{F} = \{f(\cdot - \theta), \theta \in [-1, 1]\}$ where f is defined by (1).

In these examples, we shall mainly compare our estimator with the maximum likelihood one. In Examples 1, 2, 4 and 5, the m.l.e. $\tilde{\theta}_{\text{mle}}$ 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 [6] for a discussion of numerical methods devoted to the maximization of the likelihood. In this 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 $\max(-10, \tilde{\theta}_{\text{median}} - 1)$ and $\min(10, \tilde{\theta}_{\text{median}} + 1)$ where $\tilde{\theta}_{\text{median}}$ is the median. We then select among these points the one for which the likelihood is maximal. In Example 4, we shall also compare our estimator to the estimator of the family $\{a \max_{1\le i\le n} X_i, a > 0\}$ that minimizes the Hellinger quadratic risk, that is,

$$\tilde{\theta}_{\text{best}} = \left(\frac{4n}{2n+1}\right)^{2/(2n-1)} \max_{1\le i\le n} X_i.$$

In Example 6, we shall compare our estimator to

$$\tilde{\theta}' = \frac{1}{2}\Big(\max_{1\le i\le n} X_i + \min_{1\le i\le n} X_i\Big).$$

In the case of Example 7, 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 $\tilde{\theta}_{\text{mspe}}$ (m.s.p.e. for short). This estimator was introduced by Cheng and Amin [14] and Ranneby [24] to deal with parametric models for which the likelihood is unbounded. It is known to possess nice theoretical properties when s does belong to $\mathscr{F}$. We refer, for instance, to the two aforementioned papers and to Ekström [17], Shao and Hahn [28], Ghosh and Jammalamadaka [19], Anatolyev and Kosenok [1]. This estimator is, however, not robust. In Example 4, it is, for instance, defined by $(1 + 1/n)\max_{1\le i\le n} X_i$ when all the observations X_i are positive, and is therefore highly sensitive to outliers. In Example 7, any estimator with values in $[-1, 1]$ is a m.s.p.e. when $\max_{1\le i\le n} X_i - \min_{1\le i\le n} X_i > 2$. The practical construction of the m.s.p.e. in Example 7 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 2×10^5 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 would give more accurate results, especially when n is large, but we are limited by the capacity of the computer.

3.2. Implementation of the procedure

In this simulation study, we take arbitrarily $\kappa = \bar{\kappa}/2$. We choose η very small but not too much to avoid undesirable numerical issues. More precisely, $\eta = (M - m)/10^8$ (it is small enough in view of the values of α and n).

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

It remains to choose t which tunes the thinness of the net $\mathscr{F}_{\mathrm{dis}}$. When the model is regular enough and contains s, a good choice of t seems to be $t = 0$ (that is, $\Theta_{\mathrm{dis}} = \Theta$, $\mathscr{F}_{\mathrm{dis}} = \mathscr{F}$ and $T(\theta, \theta') = \overline{T}(f_\theta, f_{\theta'})$), since then the simulations suggest that our estimator is almost equal to the m.l.e. (with large probability). In all the simulations, we take $t = 0$ (although this is not theoretically justified).

3.3. Risks when $s \in \mathscr{F}$

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

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

$$\widehat{R}_N(\tilde{\theta}) = \frac{1}{N} \sum_{i=1}^{N} h^2(s, f_{\tilde{\theta}^{(i)}}).$$

The risk $\mathbb{E}[h^2(s, f_{\tilde{\theta}})]$ of the estimator $\tilde{\theta}$ is estimated by $\widehat{R}_N(\tilde{\theta})$. We also introduce

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

in order to make the comparison of our estimator $\hat{\theta}$ and the estimator $\tilde{\theta}$ easier. When $\widehat{\mathcal{R}}_{N,\mathrm{rel}}(\tilde{\theta})$ is negative, our estimator is better than $\tilde{\theta}$, whereas if $\widehat{\mathcal{R}}_{N,\mathrm{rel}}(\tilde{\theta})$ is positive, our estimator is worse than $\tilde{\theta}$. More precisely, if $\widehat{\mathcal{R}}_{N,\mathrm{rel}}(\tilde{\theta}) = \alpha$, 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 numerical results are given in Table 1. In the first three examples, the risk of our estimator is almost equal to the one of the m.l.e., whatever the value of n. In Example 4, our estimator slightly improves the maximum likelihood estimator but has a risk 40% larger than the one of

Table 1. Risks of the estimators

		$n = 10$	$n = 25$	$n = 50$	$n = 75$	$n = 100$
Example 1	$\widehat{R}_{10^6}(\hat{\theta})$	0.0130	0.0051	0.0025	0.0017	0.0013
	$\widehat{R}_{10^6}(\tilde{\theta}_{\mathrm{mle}})$	0.0129	0.0051	0.0025	0.0017	0.0013
	$\widehat{\mathcal{R}}_{10^6,\mathrm{rel}}(\tilde{\theta}_{\mathrm{mle}})$	$6 \cdot 10^{-4}$	10^{-5}	$7 \cdot 10^{-7}$	$-8 \cdot 10^{-9}$	$2 \cdot 10^{-9}$
Example 2	$\widehat{R}_{10^6}(\hat{\theta})$	0.0123	0.0050	0.0025	0.0017	0.0012
	$\widehat{R}_{10^6}(\tilde{\theta}_{\mathrm{mle}})$	0.0123	0.0050	0.0025	0.0017	0.0012
	$\widehat{\mathcal{R}}_{10^6,\mathrm{rel}}(\tilde{\theta}_{\mathrm{mle}})$	$5 \cdot 10^{-10}$	$9 \cdot 10^{-10}$	$-2 \cdot 10^{-9}$	$-2 \cdot 10^{-9}$	$-3 \cdot 10^{-9}$
Example 3	$\widehat{R}_{10^6}(\hat{\theta})$	0.0152	0.0054	0.0026	0.0017	0.0013
	$\widehat{R}_{10^4}(\tilde{\theta}_{\mathrm{mle}})$	0.0149	0.0054	0.0026	0.0017	0.0012
	$\widehat{\mathcal{R}}_{10^4,\mathrm{rel}}(\tilde{\theta}_{\mathrm{mle}})$	−0.001	$-2 \cdot 10^{-4}$	-10^{-8}	$-3 \cdot 10^{-8}$	$9 \cdot 10^{-8}$
Example 4	$\widehat{R}_{10^6}(\hat{\theta})$	0.0468	0.0192	0.0096	0.0064	0.0048
	$\widehat{R}_{10^6}(\tilde{\theta}_{\mathrm{mle}})$	0.0476	0.0196	0.0099	0.0066	0.0050
	$\widehat{R}_{10^6}(\tilde{\theta}_{\mathrm{best}})$	0.0333	0.0136	0.0069	0.0046	0.0035
	$\widehat{\mathcal{R}}_{10^6,\mathrm{rel}}(\tilde{\theta}_{\mathrm{mle}})$	−0.0160	−0.0202	−0.0287	−0.0271	−0.0336
	$\widehat{\mathcal{R}}_{10^6,\mathrm{rel}}(\tilde{\theta}_{\mathrm{best}})$	0.4059	0.4086	0.3992	0.4025	0.3933
Example 5	$\widehat{R}_{10^6}(\hat{\theta})$	0.0504	0.0197	0.0098	0.0065	0.0049
	$\widehat{R}_{10^6}(\tilde{\theta}_{\mathrm{mle}})$	0.0483	0.0197	0.0099	0.0066	0.0050
	$\widehat{\mathcal{R}}_{10^6,\mathrm{rel}}(\tilde{\theta}_{\mathrm{mle}})$	0.0436	−0.0019	−0.0180	−0.0242	−0.0263
Example 6	$\widehat{R}_{10^6}(\hat{\theta})$	0.0455	0.0193	0.0098	0.0066	0.0050
	$\widehat{R}_{10^6}(\tilde{\theta}')$	0.0454	0.0192	0.0098	0.0066	0.0050
	$\widehat{\mathcal{R}}_{10^6,\mathrm{rel}}(\tilde{\theta}')$	0.0029	0.0029	0.0031	0.0028	0.0030
Example 7	$\widehat{R}_{10^4}(\hat{\theta})$	0.050	0.022	0.012	0.008	0.006
	$\widehat{R}_{10^4}(\tilde{\theta}_{\mathrm{mean}})$	0.084	0.061	0.049	0.043	0.039
	$\widehat{R}_{10^4}(\tilde{\theta}_{\mathrm{median}})$	0.066	0.036	0.025	0.019	0.017
	$\widehat{R}_{10^4}(\tilde{\theta}_{\mathrm{mspe}})$	0.050	0.022	0.012	0.008	0.006
	$\widehat{\mathcal{R}}_{10^4,\mathrm{rel}}(\tilde{\theta}_{\mathrm{mean}})$	−0.40	−0.64	−0.76	−0.82	−0.85
	$\widehat{\mathcal{R}}_{10^4,\mathrm{rel}}(\tilde{\theta}_{\mathrm{median}})$	−0.25	−0.39	−0.54	−0.59	−0.65

$\tilde{\theta}_{\mathrm{best}}$. In Example 5, 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 6, the risk of our estimator is 0.3% larger than the one of $\tilde{\theta}'$. In Example 7, our estimator significantly improves the empirical mean and the median. Its risk is comparable to the one of the m.s.p.e.

When the model is regular enough, these simulations show that our estimation strategy provides an estimator whose risk is very close 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 7).

Table 2. Connection with the m.l.e.

		$n = 10$	$n = 25$	$n = 50$	$n = 75$	$n = 100$
Example 1	$\hat{q}_{0.99}$	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}
	$\hat{q}_{0.999}$	0.07	10^{-7}	10^{-7}	10^{-7}	10^{-7}
	$\hat{q}_1$	1.9	0.3	0.06	0.005	10^{-7}
Example 2	$\hat{q}_{0.99}$	$2 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$
	$\hat{q}_{0.999}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$
	$\hat{q}_1$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-7}$
Example 3	$\hat{q}_{0.99}$	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
	$\hat{q}_{0.999}$	$3 \cdot 10^{-6}$	10^{-6}	10^{-6}	10^{-6}	10^{-6}
	$\hat{q}_1$	1.5	0.1	10^{-6}	10^{-6}	10^{-6}

3.4. Link with the m.l.e.

We now study numerically the connection between our estimator and the m.l.e. when the model is regular enough (that is, in the first three examples). Let for $c \in \{0.99, 0.999, 1\}$, q_c be the c-quantile of the random variable $|\hat{\theta} - \tilde{\theta}_{\mathrm{mle}}|$, and $\hat{q}_c$ be the empirical version based on N samples ($N = 10^6$ in Examples 1, 2 and $N = 10^4$ in Example 3).

Table 2 shows that with large probability, our estimator is almost equal 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 three examples. Note that the value of η prevents the empirical quantiles from being smaller than something of the order 10^{-7} according to the examples (in Example 3, the value of 10^{-6} is due to the procedure used to build the m.l.e.).

3.5. Speed of the procedure

For the sake of completeness, we specify in Table 3 the number of tests that have been calculated in the preceding examples.

We observe in Figure 1 that the number of tests computed is quite small, except for Example 7. The number of tests computed in this example is quite large because $\underline{r}(\theta, \theta')$ and $\overline{r}(\theta, \theta')$ are defined by relation (11) and $\alpha = 1/2$. The smaller α, the longer it takes to compute the estimator. Notice however that is possible to use less tests by choosing κ closer to $\bar{\kappa}$ or by using a more accurate control of the Hellinger distance $h(\theta, \theta')$.

3.6. Simulations when $s \notin \mathscr{F}$

In Section 3.3, we were in the favourable situation where the true density s belonged to the model $\mathscr{F}$, which may not hold true in practice. We now work with random variables $X_1, \dots, X_n$ simulated according to a density $s \notin \mathscr{F}$ to illustrate the robustness properties of our estimator.

Table 3. Number of tests computed averaged over 10^6 samples for Examples 1 to 6 and over 10^4 samples for Example 7. The corresponding standard deviations are in brackets

	$n = 10$	$n = 25$	$n = 50$	$n = 75$	$n = 100$
Example 1	77 (1.4)	77 (0.9)	77 (0.7)	77 (0.6)	77 (0.5)
Example 2	293 (1)	294 (1)	294 (0.9)	295 (0.9)	295 (0.9)
Example 3	100 (3.5)	100 (0.5)	100 (0.001)	100 (0)	100 (0)
Example 4	460 (3)	461 (1)	462 (0.6)	462 (0.4)	462 (0.3)
Example 5	687 (0)	687 (0)	687 (0)	687 (0)	687 (0)
Example 6	412 (8)	419 (8)	425 (8)	429 (8)	432 (8)
Example 7	173,209 (10)	173,212 (0)	173,212 (0.9)	173,206 (12)	173,212 (0.3)

We propose an example based on the mixture of two uniform laws. We use the parametric model $\mathscr{F} = \{f_\theta, \theta \in [0.01, 10]\}$ with $f_\theta = \theta^{-1}\mathbb{1}_{[0,\theta]}$, take $p \in (0, 1)$ and simulate the data according to the density

$$s_p(x) = (1 - p) f_1(x) + p f_2(x) \qquad \text{for all } x \in \mathbb{R}.$$

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

$$\begin{aligned} h^2(s_p, \mathscr{F}) &= \begin{cases} h^2(s_p, f_1), & \text{if } p \le p_0, \\ h^2(s_p, f_2), & \text{if } p \ge p_0 \end{cases} \\ &= \begin{cases} 1 - \sqrt{2-p}/\sqrt{2}, & \text{if } p \le p_0, \\ 1 - (\sqrt{2-p} + \sqrt{p})/2, & \text{if } p \ge p_0, \end{cases} \end{aligned}$$

which means that the best approximation of s_p in $\mathscr{F}$ 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. $\tilde{\theta}_{\text{mle}} = \max_{1 \le i \le n} X_i$. 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 $\tilde{\theta} \in \{\hat{\theta}, \tilde{\theta}_{\text{mle}}\}$ by computing the function

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

where $\tilde{\theta}^{(p,i)}$ is the value of the estimator $\tilde{\theta}$ corresponding to the ith sample whose density is s_p. We draw below the functions $p \mapsto \widehat{R}_{p,n,N}(\hat{\theta})$, $p \mapsto \widehat{R}_{p,n,N}(\tilde{\theta})$ and $p \mapsto h^2(s_p, \mathscr{F})$ for $n = 10^2$ and then for $n = 10^4$.

We observe in Figure 1 that the m.l.e. is rather good when $p \ge p_0$ and very poor when $p < p_0$. This can be explained by the fact that the m.l.e. $\tilde{\theta}_{\text{mle}}$ is close to 2 as soon as the number n of observations is large enough. The shape of the function $p \mapsto \widehat{R}_{p,n,5000}(\hat{\theta})$ looks more like the function $p \mapsto h^2(s_p, \mathscr{F})$. The lower figure suggests that $\widehat{R}_{p,n,N}(\hat{\theta})$ converges to $h^2(s_p, \mathscr{F})$ when n, N go to infinity except on a small neighbourhood before p_0.

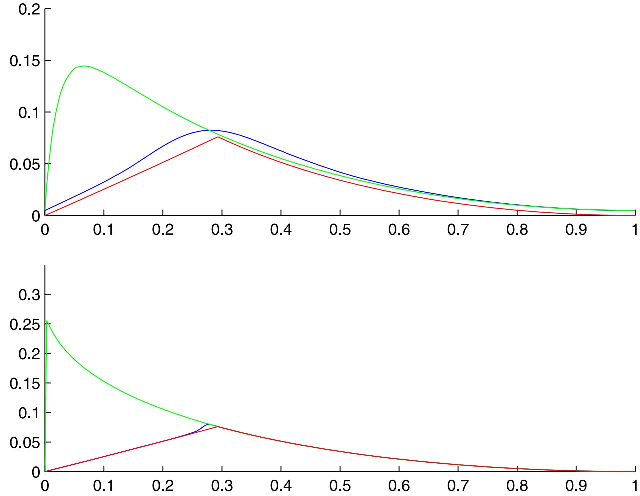

Figure 1. Red: $p \mapsto h^2(s_p, \mathscr{F})$. Blue: $p \mapsto \widehat{R}_{p,n,5000}(\hat{\theta})$. Green: $p \mapsto \widehat{R}_{p,n,5000}(\tilde{\theta}_{\text{mle}})$.

4. Models parametrized by a multi-dimensional parameter

4.1. Assumption

In the preceding sections, we have dealt with models indexed by a finite interval of $\mathbb{R}$. We now turn to the multi-dimensional case and consider models $\mathscr{F} = \{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ indexed by a rectangle $\Theta = \prod_{j=1}^d [m_j, M_j]$ of $\mathbb{R}^d$ and satisfying a multi-dimensional version of Assumption 2.1.

Assumption 4.1. *There exist positive numbers $\alpha_1, \dots, \alpha_d, \underline{R}_1, \dots, \underline{R}_d, \overline{R}_1, \dots, \overline{R}_d$ such that for all $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, $\boldsymbol{\theta}' = (\theta'_1, \dots, \theta'_d) \in \Theta = \prod_{j=1}^d [m_j, M_j]$,*

$$\sup_{j \in \{1,\dots,d\}} \underline{R}_j \left|\theta_j - \theta'_j\right|^{\alpha_j} \le h^2(\boldsymbol{\theta}, \boldsymbol{\theta}') \le \sup_{j \in \{1,\dots,d\}} \overline{R}_j \left|\theta_j - \theta'_j\right|^{\alpha_j}.$$

4.2. Definition of the test

As in the one-dimensional case, our estimation strategy is based on the existence for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ of a measurable function $T(\boldsymbol{\theta}, \boldsymbol{\theta}')$ of the observations possessing suitable statistical properties. The definition of this functional is the natural extension of the one we have proposed in Section 2.3.

Let for $j \in \{1, \dots, d\}$, $t_j \in (0, d^{1/\alpha_j}]$ and $\varepsilon_j = t_j (\overline{R} n)^{-1/\alpha_j}$. We introduce the finite sets

$$\Theta_{\text{dis}} = \left\{(m_1 + k_1\varepsilon_1, \dots, m_d + k_d\varepsilon_d), \forall j \in \{1, \dots, d\}, k_j \le (M_j - m_j)\varepsilon_j^{-1}\right\},$$

$$\mathscr{F}_{\text{dis}} = \{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_{\text{dis}}\}$$

and the map π on $\prod_{j=1}^{d}[m_j, M_j]$ by

$$\pi(\mathbf{x}) = \left(m_1 + \lfloor (x_1 - m_1)/\varepsilon_1 \rfloor \varepsilon_1, \dots, m_d + \lfloor (x_d - m_d)/\varepsilon_d \rfloor \varepsilon_d\right)$$

$$\text{for all } \mathbf{x} = (x_1, \dots, x_d) \in \prod_{j=1}^{d}[m_j, M_j],$$

where $\lfloor \cdot \rfloor$ is the integer part. We then define $T(\boldsymbol{\theta}, \boldsymbol{\theta}')$ for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ by

$$T(\boldsymbol{\theta}, \boldsymbol{\theta}') = \overline{T}(f_{\pi(\boldsymbol{\theta})}, f_{\pi(\boldsymbol{\theta}')}), \tag{13}$$

where $\overline{T}$ is given by (4).

4.3. Basic ideas

For the sake of simplicity, we first restrict ourselves to the dimension $d = 2$. The idea is to build a decreasing sequence $(\Theta_i)_i$ of rectangles by induction (in the sense of set inclusion). When there exists $\boldsymbol{\theta}_0 \in \Theta$ such that $s = f_{\boldsymbol{\theta}_0}$, these rectangles Θ_i can be interpreted as confidence sets for $\boldsymbol{\theta}_0$.

We set $\Theta_1 = \Theta$. We suppose that the rectangle Θ_i has already been built and aim at building Θ_{i+1}.

Let a_1, b_1, a_2, b_2 be such that $\Theta_i = [a_1, b_1] \times [a_2, b_2]$. For all $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta_i$, $\boldsymbol{\theta}' = (\theta_1', \theta_2') \in \Theta_i$, let $\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ be a rectangle included in Θ_i and containing a neighbourhood of $\boldsymbol{\theta}$ (for the usual topology on Θ_i) such that

$$\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{\theta}') \subset \mathcal{B}(\boldsymbol{\theta}, \kappa^{1/2} h(\boldsymbol{\theta}, \boldsymbol{\theta}')).$$

We recall that for all $\boldsymbol{\theta} \in \Theta$ and $r > 0$, $\mathcal{B}(\boldsymbol{\theta}, r) = \{\boldsymbol{\theta}' \in \Theta, h(\boldsymbol{\theta}, \boldsymbol{\theta}') \leq r\}$. Let $\mathcal{P}$ and $\mathcal{P}'$ be the two horizontal sides of the rectangle Θ_i:

$$\mathcal{P} = [a_1, b_1] \times \{a_2\},$$
$$\mathcal{P}' = [a_1, b_1] \times \{b_2\}.$$

We begin by building $L + 1$ elements $\boldsymbol{\theta}^{(\ell)} \in \mathcal{P}$ and $L + 1$ elements $\boldsymbol{\theta}'^{(\ell)} \in \mathcal{P}'$ in such a way that if $\mathcal{R}^{(\ell)}$ designates the set

$$\mathcal{R}^{(\ell)} = \begin{cases} \mathcal{R}(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}), & \text{if } T(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}) > 0, \\ \mathcal{R}(\boldsymbol{\theta}'^{(\ell)}, \boldsymbol{\theta}^{(\ell)}), & \text{if } T(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}) < 0, \\ \mathcal{R}(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}) \cup \mathcal{R}(\boldsymbol{\theta}'^{(\ell)}, \boldsymbol{\theta}^{(\ell)}), & \text{if } T(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}) = 0, \end{cases}$$

then, either

$$\mathcal{P} = \bigcup_{\ell=1}^{L} (\mathcal{R}^{(\ell)} \cap \mathcal{P}) \quad \text{or} \quad \mathcal{P}' = \bigcup_{\ell=1}^{L} (\mathcal{R}^{(\ell)} \cap \mathcal{P}'). \tag{14}$$

The rectangle Θ_{i+1} is then defined in such a way that

$$\Theta_i \Big\backslash \bigcup_{\ell=1}^{L} \mathcal{R}^{(\ell)} \subset \Theta_{i+1} \subset \Theta_i$$

and that $\Theta_{i+1} \neq \Theta_i$. Its theoretical existence is guaranteed by (14). Besides, it follows from the heuristics of Section 2.2 that Θ_{i+1} may be interpreted as a confidence set for $\boldsymbol{\theta}_0$ whenever it exists (since it contains $\Theta_i \setminus \bigcup_{\ell=1}^{L} \mathcal{R}^{(\ell)}$).

It remains to define $\boldsymbol{\theta}^{(\ell)}$ and $\boldsymbol{\theta}'^{(\ell)}$ for all $\ell \in \{1, \dots, L+1\}$. We define $\boldsymbol{\theta}^{(1)} = (a_1, a_2)$ as the bottom left corner of Θ_i and $\boldsymbol{\theta}'^{(1)} = (a_1, b_2)$ as the top left corner of Θ_i. The definition of $\boldsymbol{\theta}^{(2)}$ and $\boldsymbol{\theta}'^{(2)}$ depends on the sign of $T(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}'^{(1)})$:

- If $T(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}'^{(1)}) > 0$, we define $\boldsymbol{\theta}^{(2)}$ as the bottom right corner of $\mathcal{R}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}'^{(1)})$ and $\boldsymbol{\theta}'^{(2)} = \boldsymbol{\theta}'^{(1)}$.
- If $T(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}'^{(1)}) < 0$, we define $\boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}'^{(2)}$ as the top right corner of $\mathcal{R}(\boldsymbol{\theta}'^{(1)}, \boldsymbol{\theta}^{(1)})$.
- If $T(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}'^{(1)}) = 0$, we define $\boldsymbol{\theta}^{(2)}$ as the bottom right corner of $\mathcal{R}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}'^{(1)})$ and $\boldsymbol{\theta}'^{(2)}$ as the top right corner of $\mathcal{R}(\boldsymbol{\theta}'^{(1)}, \boldsymbol{\theta}^{(1)})$.

If $\boldsymbol{\theta}^{(2)} = (b_1, a_2)$ or if $\boldsymbol{\theta}'^{(2)} = (b_1, b_2)$, which means that one of these two points is a right corner of Θ_i, we set $L = 1$. In the contrary case, we define $\boldsymbol{\theta}^{(3)}$ either as the bottom right corner of $\mathcal{R}(\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}'^{(2)})$ or as $\boldsymbol{\theta}^{(2)}$, according to the sign of $T(\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}'^{(2)})$. Similarly, $\boldsymbol{\theta}'^{(3)}$ is either $\boldsymbol{\theta}'^{(2)}$ or the top right corner of $\mathcal{R}(\boldsymbol{\theta}'^{(2)}, \boldsymbol{\theta}^{(2)})$. If one of the points $\boldsymbol{\theta}^{(3)}$, $\boldsymbol{\theta}'^{(3)}$ is a right corner of Θ_i, we set $L = 2$. Otherwise, we build $\boldsymbol{\theta}^{(4)}$, $\boldsymbol{\theta}'^{(4)}$ and so on. More precisely, we build $\boldsymbol{\theta}^{(\ell)}$ and $\boldsymbol{\theta}'^{(\ell)}$ until that one of these two elements becomes a right corner of Θ_i. We then stop the construction and set $L = \ell - 1$. See Figure 2 for an illustration.

Remark. we define Θ_{i+1} as a rectangle to make the procedure easier to implement in practice. Note that this rectangle Θ_{i+1} is of the form $\Theta_{i+1} = [a_1, b_1] \times [a_2', b_2']$ where a_2', b_2' satisfy $b_2' - a_2' < b_2 - a_2$. We may also adapt the preceding ideas to build a confidence set Θ_{i+1} of the form $\Theta_{i+1} = [a_1', b_1'] \times [a_2, b_2]$ where a_1', b_1' satisfy $b_1' - a_1' < b_1 - a_1$.

We shall build the rectangles Θ_i until their diameters become sufficiently small. The estimator we shall consider will then be the center of the last rectangle built.

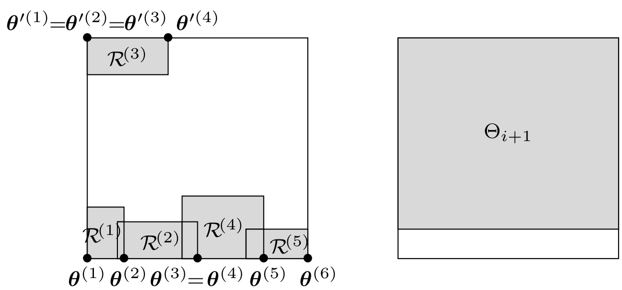

Figure 2. Illustration when $L = 5$, $T(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}'^{(i)}) > 0$ for $i \in \{1, 2, 4, 5\}$ and $T(\boldsymbol{\theta}^{(3)}, \boldsymbol{\theta}'^{(3)}) < 0$.

4.4. Estimation procedure

4.4.1. *General scheme*

In this section, we aim at designing an estimator in dimension 2 or higher. We build a finite sequence of nested rectangles $(\Theta_i)_{1\le i\le N}$ of $\mathbb{R}^d$ included in Θ by induction. These rectangles can be interpreted as confidence sets for $\boldsymbol{\theta}_0$ whenever it exists. We set $\Theta_1 = \Theta$. As long as the size of Θ_i is large enough (in a suitable sense), we use an algorithm (that we present below) to build Θ_{i+1} from Θ_i. As soon as the size of Θ_i becomes small enough, we stop the construction of the rectangles. We then denote this last rectangle by Θ_N and define our estimator $\hat{\boldsymbol{\theta}}$ as the center of Θ_N.

We now explain the general principle for constructing Θ_{i+1} from Θ_i. Let a_j, b_j be the numbers such that $\Theta_i = \prod_{j=1}^d [a_j, b_j]$ and let k be an integer in $\{1,\dots,d\}$ to be specified later. The confidence set Θ_{i+1} will be of the form

$$\Theta_{i+1} = \left(\prod_{j=1}^{k-1}[a_j, b_j]\right) \times \left[a'_k, b'_k\right] \times \left(\prod_{j=k+1}^{d}[a_j, b_j]\right) \tag{15}$$

with $a'_k, b'_k \in [a_k, b_k]$ such that $b'_k - a'_k < b_k - a_k$. In order to be a little more precise, let us consider $\kappa \in (0, \bar{\kappa})$, and, for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_i$, let $\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ be a rectangle included in Θ_i and containing a neighbourhood of $\boldsymbol{\theta}$ (for the usual topology on Θ_i) such that

$$\mathcal{R}\big(\boldsymbol{\theta}, \boldsymbol{\theta}'\big) \subset \mathcal{B}\big(\boldsymbol{\theta}, \kappa^{1/2} h\big(\boldsymbol{\theta}, \boldsymbol{\theta}'\big)\big).$$

Let $\mathcal{P}$ and $\mathcal{P}'$ be the two following opposite faces of Θ_i:

$$\mathcal{P} = \big\{(\theta_1, \dots, \theta_{k-1}, a_k, \theta_{k+1}, \dots, \theta_d), \boldsymbol{\theta} \in \Theta_i\big\},$$
$$\mathcal{P}' = \big\{(\theta_1, \dots, \theta_{k-1}, b_k, \theta_{k+1}, \dots, \theta_d), \boldsymbol{\theta} \in \Theta_i\big\}.$$

As in Section 4.3, the construction of Θ_{i+1} is based on the existence of $L+1$ elements $\boldsymbol{\theta}^{(\ell)} \in \mathcal{P}$ and $L+1$ elements $\boldsymbol{\theta}'^{(\ell)} \in \mathcal{P}'$ satisfying one of the two following relations:

$$\mathcal{P} = \bigcup_{\ell=1}^{L} \big(\mathcal{R}^{(\ell)} \cap \mathcal{P}\big) \quad \text{or} \quad \mathcal{P}' = \bigcup_{\ell=1}^{L} \big(\mathcal{R}^{(\ell)} \cap \mathcal{P}'\big), \tag{16}$$

where $\mathcal{R}^{(\ell)}$ stands for the set

$$\mathcal{R}^{(\ell)} = \begin{cases} \mathcal{R}\big(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}\big), & \text{if } T\big(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}\big) > 0, \\ \mathcal{R}\big(\boldsymbol{\theta}'^{(\ell)}, \boldsymbol{\theta}^{(\ell)}\big), & \text{if } T\big(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}\big) < 0, \\ \mathcal{R}\big(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}\big) \cup \mathcal{R}\big(\boldsymbol{\theta}'^{(\ell)}, \boldsymbol{\theta}^{(\ell)}\big), & \text{if } T\big(\boldsymbol{\theta}^{(\ell)}, \boldsymbol{\theta}'^{(\ell)}\big) = 0. \end{cases}$$

Thanks to (16), there exist $a'_k, b'_k \in [a_k, b_k]$ such that $b'_k - a'_k < b_k - a_k$ and such that the rectangle Θ_{i+1} defined by (15) satisfies

$$\Theta_i \Big\backslash \bigcup_{\ell=1}^{L} \mathcal{R}^{(\ell)} \subset \Theta_{i+1} \subset \Theta_i. \tag{17}$$

The heuristics developed in Section 2.2 show that Θ_{i+1} may be interpreted as a confidence set for $\boldsymbol{\theta}_0$ (whenever it exists). It remains to build Θ_{i+1} in a constructive way.

4.4.2. *Construction of the confidence set Θ_{i+1} from Θ_i*

We present in this section an algorithm easy to code on a computer and taking back the ideas of the preceding section to build Θ_{i+1} from Θ_i. In what follows, it is convenient to introduce positive numbers $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta}, \boldsymbol{\theta}')$, $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ such that

$$\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \Theta_i \cap \prod_{j=1}^{d} \left[\theta_j - \underline{r}_{\Theta_i,j}(\boldsymbol{\theta}, \boldsymbol{\theta}'), \theta_j + \overline{r}_{\Theta_i,j}(\boldsymbol{\theta}, \boldsymbol{\theta}')\right].$$

We recall that this set must satisfy

$$\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{\theta}') \subset \mathcal{B}\left(\boldsymbol{\theta}, \kappa^{1/2} h(\boldsymbol{\theta}, \boldsymbol{\theta}')\right). \tag{18}$$

We also consider for all $j \in \{1, \ldots, d\}$, a number $\underline{R}_{\Theta_i,j} \in [\underline{R}_j, +\infty)$ such that

$$h^2(\boldsymbol{\theta}, \boldsymbol{\theta}') \geq \sup_{1 \leq j \leq d} \underline{R}_{\Theta_i,j} \left|\theta_j - \theta'_j\right|^{\alpha_j} \qquad \text{for all } \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_i. \tag{19}$$

We finally consider for all $j \in \{1, \ldots, d\}$, a one-to-one map ψ_j from $\{1, \ldots, d-1\}$ into $\{1, \ldots, d\} \setminus \{j\}$.

We set $\Theta_1 = \Theta$. Given Θ_i, we define Θ_{i+1} by using the algorithm below. This algorithm ensues from the strategy described in the preceding section. It defines k, builds the elements $\boldsymbol{\theta}^{(\ell)}$, $\boldsymbol{\theta}'^{(\ell)}$ and, lastly returns Θ_{i+1}.

Algorithm 2 Definition of Θ_{i+1} from Θ_i

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

1: Choose $k \in \{1, \ldots, d\}$ such that

$$\underline{R}_{\Theta_i,k}(b_k - a_k)^{\alpha_k} = \max_{1 \leq j \leq d} \underline{R}_{\Theta_i,j}(b_j - a_j)^{\alpha_j}$$

2: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \leftarrow (a_1, \ldots, a_d)$

3: $\boldsymbol{\theta}' = (\theta'_1, \ldots, \theta'_d) \leftarrow (a_1, \ldots, a_{k-1}, b_k, a_{k+1}, \ldots, a_d)$

4: $\varrho_j \leftarrow \overline{r}_{\Theta_i,j}(\boldsymbol{\theta}, \boldsymbol{\theta}')$ and $\varrho'_j \leftarrow \overline{r}_{\Theta_i,j}(\boldsymbol{\theta}', \boldsymbol{\theta})$ for all $j \in \{1, \ldots, d\} \setminus \{k\}$

$\varrho_k \leftarrow (b_k - a_k)/2$ and $\varrho'_k \leftarrow (b_k - a_k)/2$
repeat
  Test $\leftarrow T(\boldsymbol{\theta}, \boldsymbol{\theta}')$
  For all $j \in \{1, \ldots, d\}$, $\overline{r}_j \leftarrow \overline{r}_{\Theta_i, j}(\boldsymbol{\theta}, \boldsymbol{\theta}')$, $\overline{r}'_j \leftarrow \overline{r}_{\Theta_i, j}(\boldsymbol{\theta}', \boldsymbol{\theta})$, $\underline{r}'_j \leftarrow \underline{r}_{\Theta_i, j}(\boldsymbol{\theta}', \boldsymbol{\theta})$
  if Test ≥ 0 **then**
    $\varrho_{\psi_k(1)} \leftarrow \overline{r}_{\psi_k(1)}$
    $\varrho_{\psi_k(j)} \leftarrow \min(\varrho_{\psi_k(j)}, \overline{r}_{\psi_k(j)})$ for all $j \in \{2, \ldots, d-1\}$
    $\varrho_k \leftarrow \min(\varrho_k, \overline{r}_k)$
    $J \leftarrow \{1 \leq j \leq d-1, \theta_{\psi_k(j)} + \varrho_{\psi_k(j)} < b_{\psi_k(j)}\}$
    if $J \neq \varnothing$ **then**
      $\mathrm{j}_{\min} \leftarrow \min J$
      $\theta_{\psi_k(j)} \leftarrow a_{\psi_k(j)}$ for all $j \leq \mathrm{j}_{\min} - 1$
      $\theta_{\psi_k(\mathrm{j}_{\min})} \leftarrow \theta_{\psi_k(\mathrm{j}_{\min})} + \varrho_{\psi_k(\mathrm{j}_{\min})}$
    else
      $\mathrm{j}_{\min} \leftarrow d$
    end if
  end if
  if Test ≤ 0 **then**
    $\varrho'_{\psi_k(1)} \leftarrow \overline{r}'_{\psi_k(1)}$
    $\varrho'_{\psi_k(j)} \leftarrow \min(\varrho'_{\psi_k(j)}, \overline{r}'_{\psi_k(j)})$ for all $j \in \{2, \ldots, d-1\}$
    $\varrho'_k \leftarrow \min(\varrho'_k, \underline{r}'_k)$
    $J' \leftarrow \{1 \leq j \leq d-1, \theta'_{\psi_k(j)} + \varrho'_{\psi_k(j)} < b_{\psi_k(j)}\}$
    if $J' \neq \varnothing$ **then**
      $\mathrm{j}'_{\min} \leftarrow \min J'$
      $\theta'_{\psi_k(j)} \leftarrow a_{\psi_k(j)}$ for all $j \leq \mathrm{j}'_{\min} - 1$
      $\theta'_{\psi_k(\mathrm{j}'_{\min})} \leftarrow \theta'_{\psi_k(\mathrm{j}'_{\min})} + \varrho'_{\psi_k(\mathrm{j}'_{\min})}$
    else
      $\mathrm{j}'_{\min} \leftarrow d$
    end if
  end if
until $\mathrm{j}_{\min} = d$ or $\mathrm{j}'_{\min} = d$
if $\mathrm{j}_{\min} = d$ **then**
  $a_k \leftarrow a_k + \varrho_k$
end if
if $\mathrm{j}'_{\min} = d$ **then**
  $b_k \leftarrow b_k - \varrho'_k$
end if
$\Theta_{i+1} \leftarrow \prod_{j=1}^{d} [a_j, b_j]$
Return: Θ_{i+1}

The parameters κ, t_j, $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$, $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ can be interpreted as in dimension 1. We have introduced a new parameter $\underline{R}_{\Theta_i,j}$ whose role is to control more accurately the Hellinger distance in order to define k. Sometimes, the computation of this parameter is difficult in practice. In this case, we can overcome this issue by remarking that for all $\boldsymbol{\theta},\boldsymbol{\theta}' \in \Theta$,

$$h^2(\boldsymbol{\theta},\boldsymbol{\theta}') \geq \sup_{1\leq j\leq d} \underline{R}\left|\theta_j - \theta_j'\right|^{\alpha_j} \qquad \text{with } \underline{R} = \min_{1\leq j\leq d} \underline{R}_j,$$

which means that we can always assume that $\underline{R}_j$ is independent of j. Choosing $\underline{R}_{\Theta_i,j} = \underline{R}$ then simplifies the only line where this parameter is involved (line 1). 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 $\underline{R}$.

4.4.3. *Construction of the estimator*

As explained in Section 4.4.1, we only build a finite number of rectangles Θ_i in order to define our estimator. We stop their construction when they become small enough. More precisely, we consider d positive numbers $\eta_1,\dots,\eta_d$ and use the following algorithm to design $\hat{\boldsymbol{\theta}}$.

Algorithm 3 Construction of the estimator

1: Set $a_j = m_j$ and $b_j = M_j$ for all $j \in \{1,\dots,d\}$
2: $i \leftarrow 0$
3: **while** there exists $j \in \{1,\dots,d\}$ such that $b_j - a_j > \eta_j$ **do**
4: $\quad i \leftarrow i+1$
5: $\quad$ Build Θ_i and set $a_1,\dots,a_d,b_1,\dots,b_d$ such that $\prod_{j=1}^d [a_j,b_j] = \Theta_i$
6: **end while**
7: **Return:**

$$\hat{\boldsymbol{\theta}} = \left(\frac{a_1+b_1}{2},\dots,\frac{a_d+b_d}{2}\right)$$

The convergence of the two preceding algorithms is guaranteed under mild conditions on $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ and $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$. We refer to Section 4.6 for more details on this point.

4.5. A risk bound

The risk of $\hat{\boldsymbol{\theta}}$ can be bounded as soon as the preceding parameters are suitably chosen.

Theorem 4.1. *Suppose that Assumption 4.1 holds with $d \geq 2$. Let $\bar{\kappa}$ be defined by (7), and assume that $\kappa \in (0,\bar{\kappa})$. Suppose that for all $j \in \{1,\dots,d\}$, $t_j \in (0, d^{1/\alpha_j}]$, $\varepsilon_j = t_j(\overline{R}_j n)^{-1/\alpha_j}$, and $\eta_j \in (0, d^{1/\alpha_j}(\overline{R}_j n)^{-1/\alpha_j}]$. Suppose moreover that for all i, $\boldsymbol{\theta},\boldsymbol{\theta}' \in \Theta_i$, the numbers $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$, $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$, are such that (18) holds and that the two preceding algorithms converge.*

Then, for all $\xi > 0$, the estimator $\hat{\boldsymbol{\theta}}$ derived from Algorithm 3 satisfies

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

where $D_{\mathscr{F}} = d \vee \sum_{j=1}^{d} \log(1 + t_j^{-1}((d/\bar{\boldsymbol{\alpha}})(c\overline{R}_j/\underline{R}_j))^{1/\alpha_j})$, *with* c *depending only on* κ, $\bar{\boldsymbol{\alpha}}$ *the harmonic mean of* $\boldsymbol{\alpha}$, *and where* $C > 0$ *depends only on* κ *and* $(\overline{R}_j/\underline{R}_j)_{1\le j\le d}$.

We can also prove that the estimator is asymptotically very close to the m.l.e. when the model $\mathscr{F}$ is regular enough and contains s. We refer to Theorem 5.2 in Section 5.1.2.

4.6. Choice of $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ and $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$

We now briefly discuss the choice of the parameters $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$, $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$. Note that they must be calculated in practice since they are involved in Algorithm 2. It turns out that the two preceding algorithms converge and that the numerical complexity of the estimation procedure can be theoretically upper bounded when $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ and $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ are larger than

$$\left(\kappa \sup_{1\le k\le d} \left\{(\underline{R}_k/\overline{R}_j)\left|\theta_k' - \theta_k\right|^{\alpha_k}\right\}\right)^{1/\alpha_j},$$

which is in particular true when they are larger than $((\kappa/\overline{R}_j)h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}))^{1/\alpha_j}$. This bound may be found in Proposition 6 of Chapter 6 of Sart [25] (it is omitted here to reduce the size of the paper). Besides, the larger $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ and $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$, the faster the convergence of the two algorithms. They should therefore be as large as possible so that (18) holds. Note that changing the values of $\overline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ and $\underline{r}_{\Theta_i,j}(\boldsymbol{\theta},\boldsymbol{\theta}')$ may influence the value of the estimator $\hat{\boldsymbol{\theta}}$ but does not modify its theoretical properties.

We refer to Sections 6 and 8 of Chapter 6 of Sart [25] for numerical simulations (the results are similar to dimension one) as well as for more information on the practical implementation of the procedure.

5. Proofs

5.1. Preliminary results on the estimation procedure

The estimators we have built in the preceding sections were based on particular sequences of subsets $(\Theta_i)_i$ of $\mathbb{R}^d$ that could be interpreted as confidence sets for the true parameter $\boldsymbol{\theta}_0$ whenever it exists. In this section, we make explicit the assumptions we need to consider on the (Θ_i) in order to ensure that the resulting estimator possesses good statistical properties.

The results of this section simultaneously cover the cases of models indexed by a one-dimensional parameter (that is, $d = 1$) and those indexed by a multi-dimensional parameter (that is, $d \ge 2$). They will allow us to prove the theoretical properties of the estimators considered in the preceding sections.

5.1.1. *A risk bound*

Theorem 5.1. *Suppose that Assumption 4.1 holds. Let* $\kappa \in (0, \bar{\kappa})$, *and let* $\Theta_1 \cdots \Theta_N$ *be* N *non-empty subsets of* Θ *such that* $\Theta_1 = \Theta$. *For all* $j \in \{1, \dots, d\}$, *let* t_j *be an arbitrary number in*

$(0, d^{1/\alpha_j}]$ and $\varepsilon_j = t_j(\overline{R}_j n)^{-1/\alpha_j}$. Assume that for all $i \in \{1, \dots, N-1\}$, there exists $L_i \geq 1$ such that for all $\ell \in \{1, \dots, L_i\}$, there exist two elements $\boldsymbol{\theta}^{(i,\ell)} \neq \boldsymbol{\theta}'^{(i,\ell)}$ of Θ_i such that

$$\Theta_i \Big\backslash \bigcup_{\ell=1}^{L_i} B^{(i,\ell)} \subset \Theta_{i+1} \subset \Theta_i, \tag{20}$$

where $B^{(i,\ell)}$ is the set defined by

$$B^{(i,\ell)} = \begin{cases} \mathcal{B}(\boldsymbol{\theta}^{(i,\ell)}, r_{i,\ell}), & \text{if } T(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}) > 0, \\ \mathcal{B}(\boldsymbol{\theta}'^{(i,\ell)}, r_{i,\ell}), & \text{if } T(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}) < 0, \\ \mathcal{B}(\boldsymbol{\theta}^{(i,\ell)}, r_{i,\ell}) \cup \mathcal{B}(\boldsymbol{\theta}'^{(i,\ell)}, r_{i,\ell}), & \text{if } T(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}) = 0, \end{cases}$$

where $r_{i,\ell}^2 = \kappa h^2(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)})$ and T the functional defined by (13). Let $\boldsymbol{\theta}_0$ be an arbitrary element of Θ such that

$$h^2(s, f_{\boldsymbol{\theta}_0}) \leq h^2(s, \mathscr{F}) + 1/n$$

and δ be a non-negative map from Θ^2 such that $\delta^2(\boldsymbol{\theta}, \boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in \Theta$ and

$$\sup_{\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_i} \delta^2(\boldsymbol{\theta}, \boldsymbol{\theta}') \leq \inf_{1 \leq \ell \leq L_i} h^2(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}) \qquad \text{for all } i \in \{1, \dots, N\}. \tag{21}$$

Then, for all $\xi > 0$,

$$\mathbb{P}\left[C \inf_{\boldsymbol{\theta} \in \Theta_N} \delta^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \geq h^2(s, \mathscr{F}) + \frac{D_{\mathscr{F}}^{(n)}}{n} + \xi\right] \leq \mathrm{e}^{-n\xi},$$

where $C > 0$ depends only on κ and where

$$D_{\mathscr{F}}^{(n)} = \max\left\{d, \sum_{j=1}^{d} \log\left(1 + t_j^{-1}\left((d/\bar{\boldsymbol{\alpha}})\left(c\overline{R}_j / \underline{R}_j^{(n)}\right)\right)^{1/\alpha_j}\right)\right\}.$$

In the definition of $D_{\mathscr{F}}^{(n)}$, $\bar{\boldsymbol{\alpha}}$ is the harmonic mean of $\boldsymbol{\alpha}$, c depends only on κ, and $\underline{R}_j^{(n)}$ is any positive number such that $\underline{R}_j^{(n)} \geq \underline{R}_j$ and such that

$$h^2(\boldsymbol{\theta}, \boldsymbol{\theta}') \geq \sup_{1 \leq j \leq d} \underline{R}_j^{(n)} \left|\theta_j - \theta_j'\right|^{\alpha_j}$$

for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ satisfying $h^2(\boldsymbol{\theta}, \boldsymbol{\theta}') \leq \frac{c}{n} \sum_{j=1}^{d} \log(1 + t_j^{-1}(M_j - m_j)(\overline{R}_j n)^{1/\alpha_j})$.

Remark. In this theorem, the sets (Θ_i), the numbers (L_i) and N as well as the elements $\boldsymbol{\theta}^{(i,\ell)}$, $\boldsymbol{\theta}'^{(i,\ell)}$ may be random.

This theorem implies Theorem 4.1. Indeed, its assumptions are fulfilled when $\underline{R}_j^{(n)} = \underline{R}_j$, when the (Θ_i) are those provided by Algorithm 2, when the elements $\boldsymbol{\theta}^{(i,\ell)}$ and $\boldsymbol{\theta}'^{(i,\ell)}$ correspond to those defined in Section 4.4 (the index i has been omitted in that section for ease of reading), and when δ^2 is defined by

$$\delta^2(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sup_{j\in\{1,\dots,d\}} \underline{R}_j \left|\theta_j - \theta'_j\right|^{\alpha_j}.$$

The fact that (20) holds follows from the fact that Θ_{i+1} has been built in such a way that (17) holds. However, this point has only be claimed and has not been proved. Its rigorous proof is quite long and is therefore postponed to the Appendix. The fact that (21) holds follows from the choice of k in Algorithm 2; see the Appendix.

The above theorem then asserts that

$$\mathbb{P}\left[C \inf_{\boldsymbol{\theta}\in\Theta_N} \sup_{j\in\{1,\dots,d\}} \underline{R}_j |\theta_{0,j} - \theta_j|^{\alpha_j} \geq h^2(s, \mathscr{F}) + \frac{D_{\mathscr{F}}}{n} + \xi\right] \leq \mathrm{e}^{-n\xi},$$

where C depends only on κ and where $D_{\mathscr{F}}$ is defined in Theorem 4.1. By using the triangular inequality, Assumption 4.1, and the fact that the estimator $\hat{\boldsymbol{\theta}}$ of Theorem 4.1 is very close to any element $\boldsymbol{\theta}$ of Θ_N (since its size is very small), we finally get

$$\mathbb{P}\left[C' h^2(s, f_{\hat{\boldsymbol{\theta}}}) \geq h^2(s, \mathscr{F}) + \frac{D_{\mathscr{F}}}{n} + \xi\right] \leq \mathrm{e}^{-n\xi},$$

where C' depends on κ and $\sup_{j\in\{1,\dots,d\}} \overline{R}_j/\underline{R}_j$.

Remark. In some models, $\underline{R}_j^{(n)}$ can be chosen much larger than $\underline{R}_j$. This refinement is omitted in Theorems 2.1 and 4.1 for ease of presentation.

5.1.2. *Connection with the maximum likelihood estimator*

In this section, we carry out the general result that link our estimator to the maximum likelihood one. In particular, it manages to deal with multi-dimensional parametric models.

We need to introduce the following notation. We define $\mathring{\Theta}$ as the interior of Θ and $l_{\boldsymbol{\theta}}(x) = \log f_{\boldsymbol{\theta}}(x)$. The gradient of the map $\boldsymbol{\theta} \mapsto \log f_{\boldsymbol{\theta}}(x)$ is denoted by $\dot{l}_{\boldsymbol{\theta}}(x)$ and its Hessian matrix by $\ddot{l}_{\boldsymbol{\theta}}(x)$. The notation $(\cdot)^T$ represents the transpose of a vector or a matrix. The Euclidean norm and its induced matrix norm are both denoted by $\|\cdot\|$. We denote the log likelihood by $L(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n l_\theta(X_i)$.

Assumption 5.1. *The following conditions are satisfied*:

(i) *Assumption* 4.1 *holds with* $\alpha_1 = \cdots = \alpha_d = 2$ *and there exists* $\boldsymbol{\theta}_0 \in \mathring{\Theta}$ *such that* $s = f_{\boldsymbol{\theta}_0} \in \mathscr{F}$.

(ii) $\mathscr{F}$ *and* κ *do not depend on* n. *The* t_j *depend on* n (*one then write* $t_j^{(n)}$ *in place of* t_j) *and are chosen in such a way that* $|\log t_j^{(n)}|/n$ *tends to* 0 *when* n *goes to infinity.*

(iii) *For μ-almost all $x \in \mathbb{X}$, the mapping $\boldsymbol{\theta} \mapsto f_{\boldsymbol{\theta}}(x)$ is positive and two times differentiable on $\mathring{\Theta}$.*

(iv) *The Fisher information matrix*

$$I(\boldsymbol{\theta}) = \int_{\mathbb{X}} \big(\dot{l}_{\boldsymbol{\theta}}(x)\big)\big(\dot{l}_{\boldsymbol{\theta}}(x)\big)^T f_{\boldsymbol{\theta}}(x)\,\mathrm{d}\mu(x)$$

exists for all $\boldsymbol{\theta} \in \mathring{\Theta}$. Moreover, the map $\boldsymbol{\theta} \mapsto I(\boldsymbol{\theta})$ is continuous and non-singular at $\boldsymbol{\theta}_0$.

(v) *The integrals $\int_{\mathbb{X}} \dot{f}_{\boldsymbol{\theta}_0}(x)\,\mathrm{d}\mu(x)$, $\int_{\mathbb{X}} \ddot{f}_{\boldsymbol{\theta}_0}(x)\,\mathrm{d}\mu(x)$ exist and are zero.*

(vi) *For all $\vartheta > 0$, there exist a neighbourhood $\Theta_0(\vartheta)$ of $\boldsymbol{\theta}_0$ (independent of n) and an event $\mathcal{A}_n(\vartheta)$ on which*

$$\sup_{\boldsymbol{\theta},\boldsymbol{\theta}' \in \Theta_0(\vartheta)} \frac{1}{n} \sum_{i=1}^{n} \frac{|\log f_{\boldsymbol{\theta}}(X_i) - \log f_{\boldsymbol{\theta}'}(X_i)|^3}{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|^2} \le \vartheta,$$

$$\frac{1}{n} \sum_{i=1}^{n} \sup_{\boldsymbol{\theta} \in \Theta_0(\vartheta)} \big\|\ddot{l}_{\boldsymbol{\theta}}(X_i) - \ddot{l}_{\boldsymbol{\theta}_0}(X_i)\big\| \le \vartheta.$$

Moreover, the maps $\vartheta \mapsto \Theta_0(\vartheta)$ and $\vartheta \mapsto \mathcal{A}_n(\vartheta)$ are non-decreasing (in the sense of set inclusion).

Theorem 5.2. *Suppose that Assumption 5.1 is fulfilled. Let δ^2 be a function satisfying the assumptions of Theorem 5.1. Let, for each $n \in \mathbb{N}^\star$, N_n be a (possibly random) positive integer and $\Theta_1, \dots, \Theta_{N_n}$ be (random) subsets satisfying the assumptions of Theorem 5.1. Let, for all $\vartheta > 0$, $\mathcal{A}'_n(\vartheta)$ be the event on which*

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \dot{l}_{\boldsymbol{\theta}_0}(X_i) \right\| \le \vartheta \quad \text{and} \quad \left\| \frac{1}{n} \sum_{i=1}^{n} \big(\ddot{l}_{\boldsymbol{\theta}_0}(X_i) - \mathbb{E}\big[\ddot{l}_{\boldsymbol{\theta}_0}(X_i)\big]\big) \right\| \le \vartheta.$$

Then there exist $\vartheta > 0$, $\xi > 0$, $n_0 \in \mathbb{N}^\star$ such that for all $n \ge n_0$:

$$\begin{aligned} &\mathbb{P}\left[\exists \tilde{\boldsymbol{\theta}} \in \mathring{\Theta}, \sum_{i=1}^{n} \dot{l}_{\tilde{\boldsymbol{\theta}}}(X_i) = 0 \text{ and } \inf_{\boldsymbol{\theta} \in \Theta_{N_n}} \delta^2(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \le \frac{120}{n} \sup_{j \in \{1,\dots,d\}} \big(t_j^{(n)}\big)^2\right] \\ &\quad \ge 1 - \big\{\mathbb{P}\big[\big(\mathcal{A}_n(\vartheta)\big)^c\big] + \mathbb{P}\big[\big(\mathcal{A}'_n(\vartheta)\big)^c\big] + \mathrm{e}^{-n\xi}\big\}. \end{aligned} \tag{22}$$

Remark that the law of large numbers implies that $\mathbb{P}[\mathcal{A}'_n(\vartheta)]$ converges to 1 when n goes to infinity. The right-hand side of inequality (22) tends therefore to 1 as soon as $\mathbb{P}[\mathcal{A}_n(\vartheta)]$ converges to 1. Moreover, under suitable assumptions, the rate of convergence of $\mathbb{P}[\mathcal{A}_n(\vartheta)]$ and $\mathbb{P}[\mathcal{A}'_n(\vartheta)]$ to 1 can be specified as in Theorem 2.2.

5.2. Proof of Theorem 5.1

Let $G:(1/\sqrt{2},1)\to(3+2\sqrt{2},+\infty)$ be the bijection defined by

$$G(x)=\frac{(1+\min((1-x)/2,x-1/\sqrt{2}))^4(1+x)+\min((1-x)/2,x-1/\sqrt{2})}{1-x-\min((1-x)/2,x-1/\sqrt{2})}.$$

Let C_κ be such that $(1+\sqrt{C_\kappa})^2=\kappa^{-1}$. Since $\kappa\in(0,\bar{\kappa})$, $C_\kappa\in(3+2\sqrt{2},+\infty)$ and there exists thus $\upsilon\in(1/\sqrt{2},1)$ such that $G(\upsilon)=C_\kappa$. We then set

$$\begin{aligned} c&=24\bigl(2+\sqrt{2}/6(\upsilon-1/\sqrt{2})\bigr)/(\upsilon-1/\sqrt{2})^2\cdot 10^3,\\ \beta_1&=\min\bigl\{(1-\upsilon)/2,\upsilon-1/\sqrt{2}\bigr\},\\ \beta_2&=(1+\beta_1)\bigl(1+\beta_1^{-1}\bigr)\bigl[1-\upsilon+(1+\beta_1)(1+\upsilon)\bigr],\\ \beta_3&=\bigl(1+\beta_1^{-1}\bigr)\bigl[1-\upsilon+(1+\beta_1)^3(1+\upsilon)\bigr]+c(1+\beta_1)^2. \end{aligned} \tag{23}$$

We need the following claim, which will be proved immediately after the present proof.

Claim 5.1. *For all $\xi>0$, there exists an event Ω_ξ such that $\mathbb{P}(\Omega_\xi)\ge 1-\mathrm{e}^{-n\xi}$ and on which, for all $f,f'\in\mathscr{F}_{\mathrm{dis}}$,*

$$(1-\upsilon)h^2\bigl(s,f'\bigr)+\frac{\overline{T}(f,f')}{\sqrt{2}}\le(1+\upsilon)h^2(s,f)+c\frac{(D_{\mathscr{F}}^{(n)}+n\xi)}{n},$$

where $D_{\mathscr{F}}^{(n)}$ is defined in Theorem 5.1 for the value of $c>0$ given by (23).

We begin by proving the following lemma.

Lemma 5.1. *For all $\xi>0$, the following assertion holds on Ω_ξ: if there exist $p\in\{1,\dots,N-1\}$ and $\ell\in\{1,\dots,L_p\}$ such that $\boldsymbol{\theta}_0\in\Theta_p$ and such that*

$$\beta_2h^2(s,f_{\boldsymbol{\theta}_0})+\beta_3\left(\frac{D_{\mathscr{F}}^{(n)}}{n}+\xi\right)<\beta_1\bigl(h^2(f_{\boldsymbol{\theta}_0},f_{\boldsymbol{\theta}^{(p,\ell)}})+h^2(f_{\boldsymbol{\theta}_0},f_{\boldsymbol{\theta}'^{(p,\ell)}})\bigr), \tag{24}$$

then $\boldsymbol{\theta}_0\notin B^{(p,\ell)}$.

Proof. Without loss of generality, we may assume that $T(\boldsymbol{\theta}^{(p,\ell)},\boldsymbol{\theta}'^{(p,\ell)})=\overline{T}(f_{\pi(\boldsymbol{\theta}^{(p,\ell)})},f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})})$ is non-negative, and prove that $\boldsymbol{\theta}_0\notin\mathcal{B}(\boldsymbol{\theta}^{(p,\ell)},r_{p,\ell})$. On the event Ω_ξ, we deduce from the claim that

$$(1-\upsilon)h^2(s,f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})})\le(1+\upsilon)h^2(s,f_{\pi(\boldsymbol{\theta}^{(p,\ell)})})+c\frac{(D_{\mathscr{F}}^{(n)}+n\xi)}{n}.$$

Consequently, by using the triangular inequality and the above inequality

$$\begin{aligned}(1-\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}) &\leq \left(1+\beta_1^{-1}\right)(1-\upsilon)h^2(s, f_{\boldsymbol{\theta}_0}) \\ &\quad + (1+\beta_1)(1-\upsilon)h^2(s, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}) \\ &\leq \left(1+\beta_1^{-1}\right)(1-\upsilon)h^2(s, f_{\boldsymbol{\theta}_0}) \\ &\quad + (1+\beta_1)\left[(1+\upsilon)h^2(s, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}) + c\frac{(D_{\mathscr{F}}^{(n)}+n\xi)}{n}\right].\end{aligned}$$

Since $h^2(s, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}) \leq (1+\beta_1^{-1})h^2(s, f_{\boldsymbol{\theta}_0}) + (1+\beta_1)h^2(f_{\boldsymbol{\theta}_0}, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})})$,

$$\begin{aligned}(1-\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}) &\leq \left(1+\beta_1^{-1}\right)\left[1-\upsilon+(1+\beta_1)(1+\upsilon)\right]h^2(s, f_{\boldsymbol{\theta}_0}) \\ &\quad + (1+\beta_1)^2(1+\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}) \\ &\quad + \frac{c(1+\beta_1)(D_{\mathscr{F}}^{(n)}+n\xi)}{n}.\end{aligned} \tag{25}$$

Remark now that for all $\boldsymbol{\theta} \in \Theta$,

$$h^2(f_{\boldsymbol{\theta}}, f_{\pi(\boldsymbol{\theta})}) \leq \sup_{1\leq j\leq d} \overline{R}_j \varepsilon_j^{\alpha_j} \leq d/n.$$

By using the triangular inequality,

$$\begin{aligned}h^2(f_{\boldsymbol{\theta}_0}, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}) &\leq (1+\beta_1)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) + d\left(1+\beta_1^{-1}\right)/n, \\ h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) &\leq (1+\beta_1)h^2(f_{\boldsymbol{\theta}_0}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}) + d\left(1+\beta_1^{-1}\right)/n.\end{aligned}$$

We deduce from these two inequalities and from (25) that

$$\begin{aligned}(1-\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) &\leq \beta_2 h^2(s, f_{\boldsymbol{\theta}_0}) + (1+\beta_1)^4(1+\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) \\ &\quad + \frac{d(1+\beta_1^{-1})[1-\upsilon+(1+\beta_1)^3(1+\upsilon)] + c(1+\beta_1)^2(D_{\mathscr{F}}^{(n)}+n\xi)}{n}.\end{aligned}$$

Since $D_{\mathscr{F}}^{(n)} \geq d$ and $\beta_3 \geq 1$,

$$\begin{aligned}(1-\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) &\leq \beta_2 h^2(s, f_{\boldsymbol{\theta}_0}) + \frac{\beta_3(D_{\mathscr{F}}^{(n)}+n\xi)}{n} \\ &\quad + (1+\beta_1)^4(1+\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}).\end{aligned}$$

By using (24),

$$\begin{aligned}(1-\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) &< \beta_1\left(h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) + h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}})\right) \\ &\quad + (1+\beta_1)^4(1+\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}})\end{aligned}$$

and thus

$$h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) < G(\upsilon)h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) \\ < C_\kappa h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}).$$

Finally,

$$h^2(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) \le \big(h(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) + h(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}})\big)^2 \\ < (1+\sqrt{C_\kappa})^2 h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) \\ < \kappa^{-1} h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}),$$

which leads to $\boldsymbol{\theta}_0 \notin \mathcal{B}(\boldsymbol{\theta}^{(p,\ell)}, r_{p,\ell})$ as wished. □

Let us return to the proof of Theorem 5.1. Since the result is straightforward when $\boldsymbol{\theta}_0 \in \Theta_N$, we assume that $\boldsymbol{\theta}_0 \notin \Theta_N$. We then set

$$p = \max\big\{i \in \{1, \dots, N-1\}, \boldsymbol{\theta}_0 \in \Theta_i\big\}$$

and consider any element $\boldsymbol{\theta}'_0$ of Θ_N. Then $\boldsymbol{\theta}'_0$ belongs to Θ_p and

$$\delta^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}'_0) \le \sup_{\boldsymbol{\theta},\boldsymbol{\theta}' \in \Theta_p} \delta^2(\boldsymbol{\theta}, \boldsymbol{\theta}') \\ \le \inf_{\ell \in \{1,\dots,L_p\}} h^2(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) \\ \le 2 \inf_{\ell \in \{1,\dots,L_p\}} \big(h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) + h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}})\big).$$

By the definition of p, $\boldsymbol{\theta}_0 \in \Theta_p \setminus \Theta_{p+1}$. We then derive from the above lemma that on Ω_ξ,

$$\beta_1 \inf_{\ell \in \{1,\dots,L_p\}} \big(h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}^{(p,\ell)}}) + h^2(f_{\boldsymbol{\theta}_0}, f_{\boldsymbol{\theta}'^{(p,\ell)}})\big) \le \beta_2 h^2(s, f_{\boldsymbol{\theta}_0}) + \beta_3 \frac{D_{\mathscr{F}}^{(n)} + n\xi}{n}.$$

Hence,

$$\delta^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}'_0) \le \frac{2}{\beta_1}\left(\beta_2 h^2(s, f_{\boldsymbol{\theta}_0}) + \beta_3 \frac{D_{\mathscr{F}}^{(n)} + n\xi}{n}\right).$$

Since $h^2(s, f_{\boldsymbol{\theta}_0}) \le h^2(s, \mathscr{F}) + 1/n$, there exists $C > 0$ depending only on κ such that

$$C\delta^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}'_0) \le h^2(s, \mathscr{F}) + \frac{D_{\mathscr{F}}^{(n)}}{n} + \xi \qquad \text{on } \Omega_\xi.$$

This concludes the proof of the theorem.

It remains to prove Claim 5.1. It actually derives from the work of Baraud [2]. More precisely, Proposition 2 of Baraud [2] says that for all $f, f' \in \mathscr{F}_{\text{dis}}$,

$$\left(1 - \frac{1}{\sqrt{2}}\right)h^2(s, f') + \frac{\overline{T}(f, f')}{\sqrt{2}} \le \left(1 + \frac{1}{\sqrt{2}}\right)h^2(s, f) + \frac{\overline{T}(f, f') - \mathbb{E}[\overline{T}(f, f')]}{\sqrt{2}}.$$

Let $z = \upsilon - 1/\sqrt{2} \in (0, 1 - 1/\sqrt{2})$. We define Ω_ξ by

$$\Omega_\xi = \bigcap_{f, f' \in \mathscr{F}_{\text{dis}}} \left[\frac{\overline{T}(f, f') - \mathbb{E}[\overline{T}(f, f')]}{z(h^2(s, f) + h^2(s, f')) + c(D_{\mathscr{F}}^{(n)} + n\xi)/n} \le \sqrt{2}\right].$$

On this event,

$$(1 - \upsilon)h^2(s, f') + \frac{\overline{T}(f, f')}{\sqrt{2}} \le (1 + \upsilon)h^2(s, f) + c\frac{D_{\mathscr{F}}^{(n)} + n\xi}{n}$$

and the inequality $\mathbb{P}(\Omega_\xi^c) \le \mathrm{e}^{-n\xi}$ will follow from Lemma 1 of Baraud [2]. Before applying this lemma, we need to check that his Assumption 3 is fulfilled. This is the purpose of the claim below.

Claim 5.2. *Let*

$$\tau = 4\frac{2 + (n\sqrt{2}/6)z}{(n^2/6)z^2},$$

$$\eta_{\mathscr{F}}^2 = \max\left\{3d\mathrm{e}^4, \sum_{j=1}^{d} \log\left(1 + 2t_j^{-1}\left((d/\bar{\boldsymbol{\alpha}})\left(c\overline{R}_j/\underline{R}_j^{(n)}\right)\right)^{1/\alpha_j}\right)\right\}.$$

Then, for all $r \ge 2\eta_{\mathscr{F}}$,

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(s, r\sqrt{\tau})\right| \le \exp\left(r^2/2\right), \tag{26}$$

where $\mathcal{B}_h(s, r\sqrt{\tau})$ *is the Hellinger ball centered at* s *with radius* $r\sqrt{\tau}$ *defined by*

$$\mathcal{B}_h(s, r\sqrt{\tau}) = \left\{f \in \mathbb{L}_+^1(\mathbb{X}, \mu), h^2(s, f) \le r^2\tau\right\}.$$

Proof. If $\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(s, r\sqrt{\tau}) = \varnothing$, (26) holds. In the contrary case, there exists $\boldsymbol{\theta}_0' = (\theta_{0,1}', \dots, \theta_{0,d}') \in \Theta_{\text{dis}}$ such that $h^2(s, f_{\boldsymbol{\theta}_0'}) \le r^2\tau$, and thus

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(s, r\sqrt{\tau})\right| \le \left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(f_{\boldsymbol{\theta}_0'}, 2r\sqrt{\tau})\right|.$$

First of all, suppose that r satisfies

$$4r^2\tau \le \frac{c}{n}\sum_{j=1}^{d} \log\left(1 + t_j^{-1}(M_j - m_j)(\overline{R}_j n)^{1/\alpha_j}\right). \tag{27}$$

Then

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(f_{\boldsymbol{\theta}'_0}, 2r\sqrt{\tau})\right| = \left|\left\{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_{\text{dis}}, h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'_0}) \leq 4r^2\tau\right\}\right|$$
$$\leq \left|\left\{\boldsymbol{\theta} \in \Theta_{\text{dis}}, \forall j \in \{1,\ldots,d\}, \underline{R}_j^{(n)}\left|\theta_j - \theta'_{0,j}\right|^{\alpha_j} \leq 4r^2\tau\right\}\right|.$$

Let $k_{0,j} \in \mathbb{N}$ be such that $\theta'_{0,j} = m_j + k_{0,j}\varepsilon_j$. Then

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(f_{\boldsymbol{\theta}'_0}, 2r\sqrt{\tau})\right| \leq \prod_{j=1}^{d}\left|\left\{k_j \in \mathbb{N}, |k_j - k_{0,j}| \leq \left(4r^2\tau/\underline{R}_j^{(n)}\right)^{1/\alpha_j}\varepsilon_j^{-1}\right\}\right|$$
$$\leq \prod_{j=1}^{d}\left(1 + 2\varepsilon_j^{-1}\left(4r^2\tau/\underline{R}_j^{(n)}\right)^{1/\alpha_j}\right).$$

It is worthwhile to notice that $10^3\tau \leq c/n$. In particular, by using the weaker inequality $4\tau \leq c/n$ and $\varepsilon_j = t_j(\overline{R}_j n)^{-1/\alpha_j}$,

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(f_{\boldsymbol{\theta}'_0}, 2r\sqrt{\tau})\right| \leq \prod_{j=1}^{d}\left(1 + 2t_j^{-1}\left(r^2 c\overline{R}_j/\underline{R}_j^{(n)}\right)^{1/\alpha_j}\right).$$

If $\bar{\boldsymbol{\alpha}} \leq \mathrm{e}^{-4}$, one can check that $\eta_{\mathscr{F}}^2 \geq 4d/\bar{\boldsymbol{\alpha}}$ (since $c \geq 1$ and $t_j^{-1} \geq d^{-1/\alpha_j}$). If now $\bar{\boldsymbol{\alpha}} \geq \mathrm{e}^{-4}$, then $\eta_{\mathscr{F}}^2 \geq 3d\mathrm{e}^4 \geq 3d/\bar{\boldsymbol{\alpha}}$. In particular, we always have $r^2 \geq 10(d/\bar{\boldsymbol{\alpha}})$.

We derive from the weaker inequality $r^2 \geq d/\bar{\boldsymbol{\alpha}}$ that

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(f_{\boldsymbol{\theta}'_0}, 2r\sqrt{\tau})\right| \leq \left(\frac{r^2}{d/\bar{\boldsymbol{\alpha}}}\right)^{d/\bar{\boldsymbol{\alpha}}}\prod_{j=1}^{d}\left(1 + 2t_j^{-1}\left((d/\bar{\boldsymbol{\alpha}})\left(c\overline{R}_j/\underline{R}_j^{(n)}\right)\right)^{1/\alpha_j}\right)$$
$$\leq \exp\left(\frac{\log(r^2/(d/\bar{\boldsymbol{\alpha}}))}{r^2/(d/\bar{\boldsymbol{\alpha}})}r^2\right)\exp\left(\eta_{\mathscr{F}}^2\right).$$

We then deduce from the inequalities $r^2/(d/\bar{\boldsymbol{\alpha}}) \geq 10$ and $\eta_{\mathscr{F}}^2 \leq r^2/4$ that

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(f_{\boldsymbol{\theta}'_0}, 2r\sqrt{\tau})\right| \leq \exp\left(r^2/4\right)\exp\left(r^2/4\right) \leq \exp\left(r^2/2\right)$$

as wished. It remains to show that this inequality remains true when (27) does not hold. In this case,

$$\left|\mathscr{F}_{\text{dis}} \cap \mathcal{B}_h(f_{\boldsymbol{\theta}'_0}, 2r\sqrt{\tau})\right| \leq |\Theta_{\text{dis}}| \leq \prod_{j=1}^{d}\left(1 + \frac{M_j - m_j}{\varepsilon_j}\right) \leq \mathrm{e}^{4r^2\tau n/c}.$$

The result follows from the inequality $4\tau n/c \leq 1/2$. □

We can now use Lemma 1 of Baraud [2] to get for all $\xi > 0$ and $y^2 \geq \tau(4\eta_{\mathscr{F}}^2 + n\xi)$,

$$\mathbb{P}\left[\sup_{f,f'\in\mathscr{F}_{\mathrm{dis}}} \frac{(\overline{T}(f,f') - \mathbb{E}[\overline{T}(f,f')])/\sqrt{2}}{(h^2(s,f)+h^2(s,f'))\vee y^2} \geq z\right] \leq \mathrm{e}^{-n\xi}.$$

Since $4\eta_{\mathscr{F}}^2 \leq 10^3 D_{\mathscr{F}}^{(n)}$ and $10^3\tau \leq c/n$, we can choose $y^2 = c(D_{\mathscr{F}}^{(n)} + n\xi)/n$, which concludes the proof of Claim 5.1.

5.3. Proof of Theorem 5.2

All along this proof, we set $\mathcal{C}_n(\vartheta) = \mathcal{A}_n(\vartheta) \cap \mathcal{A}_n'(\vartheta)$ for all $\vartheta > 0$ and we denote by λ_0 the minimum between the smallest eigenvalue of $I(\boldsymbol{\theta}_0)$ and 1. Since $I(\boldsymbol{\theta}_0)$ is invertible, $\lambda_0 \in (0,1]$.

Claim 5.3. *For all $r > 0$, there exists $\vartheta > 0$ such that, on the event $\mathcal{C}_n(\vartheta)$, there exists a solution $\tilde{\boldsymbol{\theta}} \in \mathring{\Theta}$ of the likelihood equation*

$$\frac{1}{n}\sum_{i=1}^n \dot{l}_{\tilde{\boldsymbol{\theta}}}(X_i) = 0$$

satisfying $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \leq r$.

Proof. The proof of this claim follows from classical arguments that can be found in the literature. We make them explicit for the sake of completeness. There exists a neighbourhood $\Theta_0(\lambda_0/8)$ of $\boldsymbol{\theta}_0$, such that on $\mathcal{A}_n(\lambda_0/8) \cap \mathcal{A}_n'(\lambda_0/8)$,

$$\frac{1}{n}\sum_{i=1}^n \sup_{\boldsymbol{\theta}\in\Theta_0(\lambda_0/8)} \left\|\ddot{l}_{\boldsymbol{\theta}}(X_i) - \ddot{l}_{\boldsymbol{\theta}_0}(X_i)\right\| \leq \frac{\lambda_0}{8},$$

$$\left\|\frac{1}{n}\sum_{i=1}^n \left(\ddot{l}_{\boldsymbol{\theta}_0}(X_i) - \mathbb{E}\left[\ddot{l}_{\boldsymbol{\theta}_0}(X_i)\right]\right)\right\| \leq \frac{\lambda_0}{8}.$$

Without lost of generality, we may assume that r is small enough so that $r \leq 1$ and that the ball $\{\boldsymbol{\theta} \in \Theta, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq r\}$ is a subset of $\Theta_0(\lambda_0/8)$. Let $\mathcal{S}_r$ be the d-sphere $\mathcal{S}_r = \{\boldsymbol{\theta} \in \Theta, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = r\}$ and $\vartheta = \lambda_0 r/8$. Then, on $\mathcal{A}_n'(\vartheta)$,

$$\left\|\frac{1}{n}\sum_{i=1}^n \dot{l}_{\boldsymbol{\theta}_0}(X_i)\right\| \leq \frac{r\lambda_0}{8}.$$

We now use Taylor's theorem to show that for all $\boldsymbol{\theta} \in \mathcal{S}_r$, and μ-almost all $x \in \mathbb{X}$, there exists $\boldsymbol{\theta}_x \in \Theta_0(\lambda_0/8)$ such that

$$l_{\boldsymbol{\theta}}(x) = l_{\boldsymbol{\theta}_0}(x) + \left(\dot{l}_{\boldsymbol{\theta}_0}(x)\right)^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \ddot{l}_{\boldsymbol{\theta}_x}(x)(\boldsymbol{\theta} - \boldsymbol{\theta}_0).$$

In particular, for all $\boldsymbol{\theta} \in \mathcal{S}_r$,

$$\begin{aligned}
&\left| L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \right| \\
&\quad \leq r \left\| \frac{1}{n} \sum_{i=1}^n \dot{l}_{\boldsymbol{\theta}_0}(X_i) \right\| + r^2 \left\| \frac{1}{2n} \sum_{i=1}^n \bigl(\ddot{l}_{\boldsymbol{\theta}_{X_i}}(X_i) + I(\boldsymbol{\theta}_0) \bigr) \right\| \\
&\quad \leq r \left\| \frac{1}{n} \sum_{i=1}^n \dot{l}_{\boldsymbol{\theta}_0}(X_i) \right\| + \frac{r^2}{2n} \sum_{i=1}^n \sup_{\boldsymbol{\theta}' \in \Theta_0(\lambda_0/8)} \bigl\| \ddot{l}_{\boldsymbol{\theta}'}(X_i) - \ddot{l}_{\boldsymbol{\theta}_0}(X_i) \bigr\| \\
&\qquad + r^2 \left\| \frac{1}{2n} \sum_{i=1}^n \bigl(\ddot{l}_{\boldsymbol{\theta}_0}(X_i) + I(\boldsymbol{\theta}_0) \bigr) \right\|.
\end{aligned}$$

Now, remark that

$$f_{\boldsymbol{\theta}_0}(x) \ddot{l}_{\boldsymbol{\theta}_0}(x) = \ddot{f}_{\boldsymbol{\theta}_0}(x) - \bigl(\dot{l}_{\boldsymbol{\theta}_0}(x) \bigr) \bigl(\dot{l}_{\boldsymbol{\theta}_0}(x) \bigr)^T f_{\boldsymbol{\theta}_0}(x),$$

which, together with point (v) of Assumption 5.1, yields $I(\boldsymbol{\theta}_0) = -\mathbb{E}[\ddot{l}_{\boldsymbol{\theta}_0}(X_1)]$. Therefore, on the event $\mathcal{C}_n(\vartheta) \subset \mathcal{A}_n(\lambda_0/8) \cap \mathcal{A}'_n(\vartheta)$,

$$\left| L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \right| \leq \frac{\lambda_0 r^2}{4},$$

which implies

$$\begin{aligned}
L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}_0) &\leq \frac{\lambda_0 r^2}{4} - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\
&\leq \frac{\lambda_0 r^2}{4} - \frac{\lambda_0 r^2}{2} < 0.
\end{aligned}$$

This means that for all $\boldsymbol{\theta} \in \mathcal{S}_r$, $L(\boldsymbol{\theta}) < L(\boldsymbol{\theta}_0)$ on the event $\mathcal{C}_n(\vartheta)$. In particular, this proves that there exists $\tilde{\boldsymbol{\theta}}$ in the ball $\{\boldsymbol{\theta} \in \Theta, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < r\}$ such that $\dot{L}(\tilde{\boldsymbol{\theta}}) = 0$. □

Claim 5.4. *For all $\tau \in (0,1)$, there exists a neighbourhood $\Theta_1(\tau)$ of $\boldsymbol{\theta}_0$ such that for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_1(\tau)$,*

$$\frac{1-\tau}{8} \bigl(\boldsymbol{\theta} - \boldsymbol{\theta}' \bigr)^T I(\boldsymbol{\theta}_0) \bigl(\boldsymbol{\theta} - \boldsymbol{\theta}' \bigr) \leq h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \leq \frac{1+\tau}{8} \bigl(\boldsymbol{\theta} - \boldsymbol{\theta}' \bigr)^T I(\boldsymbol{\theta}_0) \bigl(\boldsymbol{\theta} - \boldsymbol{\theta}' \bigr).$$

The proof of this claim is omitted since it is very similar to the one of Lemma 1.A of Section 31 of Borovkov [13].

Claim 5.5. *For all $\tau \in (0,1)$ and for all $r > 0$, there exist a neighbourhood $\Theta_2(\tau)$ of $\boldsymbol{\theta}_0$ (that does not depend on r) and $\vartheta > 0$ such that on $\mathcal{C}_n(\vartheta)$:*

- *There exists a solution* $\tilde{\boldsymbol{\theta}} \in \Theta_2(\tau)$ *of the likelihood equation satisfying* $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \le r$.
- *For all* $\boldsymbol{\theta} \in \Theta_2(\tau)$,

$$\left|L(\tilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}) - \tfrac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\right| < \tau(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}).$$

Proof. Let $\Theta_2(\tau)$ be a convex neighbourhood of $\boldsymbol{\theta}_0$ included in $\Theta_0(\tau\lambda_0)$. Without lost of generality, we can assume that r is small enough to ensure that the ball $\{\boldsymbol{\theta} \in \Theta, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le r\}$ is included in $\Theta_2(\tau)$.

Thanks to Claim 5.3, there exist a positive number ϑ_0 and a solution $\tilde{\boldsymbol{\theta}} \in \Theta$ of the likelihood equation satisfying $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \le r$ on $\mathcal{C}_n(\vartheta_0)$. In particular $\tilde{\boldsymbol{\theta}} \in \Theta_2(\tau)$. We then use Taylor's theorem to show that for all $\boldsymbol{\theta} \in \Theta_2(\tau)$ and μ-almost all $x \in \mathbb{X}$, there exists $\boldsymbol{\theta}_x \in \Theta_2(\tau)$ such that

$$l_{\boldsymbol{\theta}}(x) = l_{\tilde{\boldsymbol{\theta}}}(x) + \big(\dot{l}_{\tilde{\boldsymbol{\theta}}}(x)\big)^T(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + \tfrac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \ddot{l}_{\boldsymbol{\theta}_x}(x)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}).$$

Therefore,

$$\begin{aligned} &l_{\boldsymbol{\theta}}(x) - l_{\tilde{\boldsymbol{\theta}}}(x) - \big(\dot{l}_{\tilde{\boldsymbol{\theta}}}(x)\big)^T(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + \tfrac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \\ &\quad = \tfrac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \big(\ddot{l}_{\boldsymbol{\theta}_x}(x) + I(\boldsymbol{\theta}_0)\big)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}). \end{aligned}$$

We derive that

$$\begin{aligned} \left|L(\boldsymbol{\theta}) - L(\tilde{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\right| &= \left|(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \left(\frac{1}{2n}\sum_{i=1}^n \big(\ddot{l}_{\boldsymbol{\theta}_{X_i}}(X_i) + I(\boldsymbol{\theta}_0)\big)\right)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\right| \\ &\le \left(\frac{1}{2n}\sum_{i=1}^n \sup_{\boldsymbol{\theta}' \in \Theta_0(\tau\lambda_0)} \big\|\ddot{l}_{\boldsymbol{\theta}'}(X_i) - \ddot{l}_{\boldsymbol{\theta}_0}(X_i)\big\|\right)\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|^2 \\ &\quad + \left\|\frac{1}{2n}\sum_{i=1}^n \big(\ddot{l}_{\boldsymbol{\theta}_0}(X_i) + I(\boldsymbol{\theta}_0)\big)\right\| \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|^2. \end{aligned}$$

We now set $\vartheta = \min(\vartheta_0, \tau\lambda_0)$ so that $\mathcal{C}_n(\vartheta) \subset \mathcal{C}_n(\vartheta_0) \cap \mathcal{C}_n(\tau\lambda_0)$. On the event $\mathcal{C}_n(\vartheta)$, we thus have for all $\boldsymbol{\theta} \in \Theta_2(\tau)$,

$$\begin{aligned} \left|L(\boldsymbol{\theta}) - L(\tilde{\boldsymbol{\theta}}) + \tfrac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\right| &\le \tau\lambda_0\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|^2 \\ &\le \tau(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}). \end{aligned}$$

This completes the proof. □

Claim 5.6. *For all* $\tau \in (0, 1)$ *and* $r > 0$, *there exist a neighbourhood* $\Theta_3(\tau)$ *of* $\boldsymbol{\theta}_0$ (*that does not depend on* r) *and* $\vartheta > 0$ *such that on* $\mathcal{C}_n(\vartheta)$:

- *There exists a solution of the likelihood equation* $\tilde{\boldsymbol{\theta}} \in \Theta$ *satisfying* $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \le r$.

- *For all* $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_3(\tau)$,

$$\overline{T}(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \le \left(\frac{8+5\sqrt{2}}{7} + \tau\right) h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}}) - (8 - 5\sqrt{2} - \tau) h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}'}). \tag{28}$$

Proof. We introduce the function F defined on $(0, +\infty)$ by $F(x) = (\sqrt{x} - 1)/\sqrt{1+x}$ and define for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$,

$$\overline{T}_1(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) = \frac{1}{2} \int_{\mathbb{X}} \sqrt{f_{\boldsymbol{\theta}}(x) + f_{\boldsymbol{\theta}'}(x)} \left(\sqrt{f_{\boldsymbol{\theta}'}(x)} - \sqrt{f_{\boldsymbol{\theta}}(x)}\right) \mathrm{d}\mu(x),$$
$$\overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) = \frac{1}{n} \sum_{i=1}^{n} F\left(\frac{f_{\boldsymbol{\theta}'}(X_i)}{f_{\boldsymbol{\theta}}(X_i)}\right).$$

Remark that for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$,

$$\overline{T}(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) = \overline{T}_1(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) + \overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}). \tag{29}$$

We begin by bounding $\overline{T}_1(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'})$ from above. Since $f_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}'}$ are two densities, $\overline{T}_1(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'})$ is also equal to

$$\overline{T}_1(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) = \frac{1}{2} \int_{\mathbb{X}} \left(\sqrt{f_{\boldsymbol{\theta}}(x) + f_{\boldsymbol{\theta}'}(x)} - \frac{\sqrt{f_{\boldsymbol{\theta}}(x)} + \sqrt{f_{\boldsymbol{\theta}'}(x)}}{\sqrt{2}}\right) \left(\sqrt{f_{\boldsymbol{\theta}'}(x)} - \sqrt{f_{\boldsymbol{\theta}}(x)}\right) \mathrm{d}\mu(x).$$

By using the inequality

$$\left|\sqrt{a+b} - \frac{\sqrt{a} + \sqrt{b}}{\sqrt{2}}\right| \le (1 - 1/\sqrt{2}) |\sqrt{b} - \sqrt{a}| \qquad \text{for all } a, b \ge 0,$$

we get

$$\overline{T}_1(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \le (1 - 1/\sqrt{2}) h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}).$$

We then use the triangular inequality to deduce

$$\overline{T}_1(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \le (1 - 1/\sqrt{2}) \left[\left(1 + \frac{5+4\sqrt{2}}{7}\right) h^2(f_{\boldsymbol{\theta}}, f_{\tilde{\boldsymbol{\theta}}}) + \left(1 + \frac{7}{5+4\sqrt{2}}\right) h^2(f_{\boldsymbol{\theta}'}, f_{\tilde{\boldsymbol{\theta}}})\right]. \tag{30}$$

We now aim at bounding $\overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'})$ from above. We consider $\tau_0 \in (0, 1/2]$ such that

$$\frac{1+\tau_0}{1-\tau_0} \le 1 + \frac{\tau}{2}, \tag{31}$$

and define

$$\Theta_3(\tau) = \Theta_0\left(\frac{\tau \lambda_0 \sqrt{2}/64}{5\sqrt{2}/384}\right) \cap \Theta_1(\tau_0) \cap \Theta_2(\tau_0/2),$$

where we recall that $\Theta_0(\cdot)$ is given by point (vi) of Assumption 5.1 and that $\Theta_1(\cdot)$, $\Theta_2(\cdot)$ are defined in the two preceding claims. Thanks to Claim 5.5, there exists $\vartheta_0 > 0$ such that on $\mathcal{C}_n(\vartheta_0)$, there exists a solution $\tilde{\boldsymbol{\theta}}$ of the likelihood equation satisfying $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \leq r$, and such that for all $\boldsymbol{\theta} \in \Theta_2(\tau_0/2)$,

$$\left| L(\tilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \right| < \frac{\tau_0}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}). \tag{32}$$

We then set

$$\vartheta = \min\left\{ \vartheta_0, \frac{\tau \lambda_0 \sqrt{2}/64}{5\sqrt{2}/384} \right\}.$$

We shall bound $\overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'})$ on the event $\mathcal{C}_n(\vartheta)$. To this end, remark that

$$\left| F(x) - \frac{\log x}{2\sqrt{2}} \right| \leq \frac{5\sqrt{2}}{384} |\log x|^3 \qquad \text{for all } x > 0.$$

Consequently,

$$\begin{aligned} \overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) - \frac{L(\boldsymbol{\theta}')}{2\sqrt{2}} + \frac{L(\boldsymbol{\theta})}{2\sqrt{2}} &= \frac{1}{n}\sum_{i=1}^n \left[F\left(\frac{f_{\boldsymbol{\theta}'}(X_i)}{f_{\boldsymbol{\theta}}(X_i)} \right) - \frac{1}{2\sqrt{2}} \log\left(\frac{f_{\boldsymbol{\theta}'}(X_i)}{f_{\boldsymbol{\theta}}(X_i)} \right) \right] \\ &\leq \frac{5\sqrt{2}}{384n} \sum_{i=1}^n \left| \log f_{\boldsymbol{\theta}'}(X_i) - \log f_{\boldsymbol{\theta}}(X_i) \right|^3. \end{aligned} \tag{33}$$

On the event $\mathcal{C}_n(\vartheta)$, for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_3(\tau)$:

$$\begin{aligned} \frac{5\sqrt{2}}{384n} \sum_{i=1}^n \left| \log f_{\boldsymbol{\theta}}(X_i) - \log f_{\boldsymbol{\theta}'}(X_i) \right|^3 &\leq \frac{\tau\sqrt{2}}{64} \lambda_0 \left\| \boldsymbol{\theta} - \boldsymbol{\theta}' \right\|^2 \\ &\leq \frac{\tau\sqrt{2}}{64} \left(\boldsymbol{\theta} - \boldsymbol{\theta}' \right)^T I(\boldsymbol{\theta}_0) \left(\boldsymbol{\theta} - \boldsymbol{\theta}' \right). \end{aligned}$$

By using $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_1(\tau_0)$ and that $\tau_0 \leq 1/2$, we deduce from Claim 5.4 that

$$\frac{5\sqrt{2}}{384n} \sum_{i=1}^n \left| \log f_{\boldsymbol{\theta}}(X_i) - \log f_{\boldsymbol{\theta}'}(X_i) \right|^3 \leq \frac{8}{1 - 1/2} \times \frac{\tau\sqrt{2}}{64} h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}).$$

By putting this inequality into (33),

$$\begin{aligned} \overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) &\leq \frac{L(\boldsymbol{\theta}')}{2\sqrt{2}} - \frac{L(\boldsymbol{\theta})}{2\sqrt{2}} + \frac{\tau\sqrt{2}}{4} h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \\ &\leq \frac{L(\boldsymbol{\theta}') - L(\tilde{\boldsymbol{\theta}})}{2\sqrt{2}} - \frac{L(\boldsymbol{\theta}) - L(\tilde{\boldsymbol{\theta}})}{2\sqrt{2}} + \frac{\tau\sqrt{2}}{4} h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}). \end{aligned}$$

We deduce from (32),

$$\overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \leq \frac{1+\tau_0}{4\sqrt{2}}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})^T I(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}) - \frac{1-\tau_0}{4\sqrt{2}}\big(\boldsymbol{\theta}'-\tilde{\boldsymbol{\theta}}\big)^T I(\boldsymbol{\theta}_0)\big(\boldsymbol{\theta}'-\tilde{\boldsymbol{\theta}}\big) + \frac{\tau\sqrt{2}}{4} h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}).$$

Since $\boldsymbol{\theta}, \boldsymbol{\theta}'$ belong together to $\Theta_1(\tau_0)$,

$$\overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \leq \sqrt{2}\frac{1+\tau_0}{1-\tau_0} h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}}) - \sqrt{2}\frac{1-\tau_0}{1+\tau_0} h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}'}) + \frac{\tau\sqrt{2}}{4} h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}).$$

It follows from (31) that $(1-\tau_0)/(1+\tau_0) \geq 1/(1+\tau/2) \geq 1-\tau/2$, and thus

$$\overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \leq \sqrt{2}\left(1+\frac{\tau}{2}\right) h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}}) - \sqrt{2}\left(1-\frac{\tau}{2}\right) h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}'}) + \frac{\tau\sqrt{2}}{4} h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}).$$

By using the triangular inequality,

$$\frac{\tau\sqrt{2}}{4} h^2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \leq \sqrt{2}\left(\frac{\tau}{2} h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}}) + \frac{\tau}{2} h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}'})\right)$$

and hence

$$\overline{T}_2(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}'}) \leq (\sqrt{2}+\tau) h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}}) - (\sqrt{2}-\tau) h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}'}).$$

We then use (29) and (30) to complete the proof. □

We now return to the proof of Theorem 5.2. Let $\beta_1, \beta_2, \beta_3$ be the numbers given at the beginning of the proof of Theorem 5.1 (they only depend on κ). Let $\tau = 0.01$ and let $\Theta_3(\tau)$ be the set given by Claim 5.6. There exists $r_0 > 0$ such that the ball

$$\mathcal{B}(\boldsymbol{\theta}_0, r_0) = \big\{\boldsymbol{\theta} \in \Theta, h(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_0}) \leq r_0\big\}$$

is included in $\Theta_3(\tau)$. We define $\xi = r_0^2 \beta_1/(9\beta_3)$ and consider $\vartheta > 0$ so that there exists a solution $\tilde{\boldsymbol{\theta}}$ of the likelihood equation on $\mathcal{C}_n(\vartheta)$ satisfying

$$h^2(f_{\boldsymbol{\theta}_0}, f_{\tilde{\boldsymbol{\theta}}}) \leq \big[9(1+\beta_2/\beta_1)\big]^{-1} r_0^2. \tag{34}$$

We may assume (without lost of generality) that $\tilde{\boldsymbol{\theta}} \notin \Theta_{N_n}$. We then set

$$p = \max\big\{i \in \{1, \dots, N_n - 1\}, \tilde{\boldsymbol{\theta}} \in \Theta_i\big\}.$$

By the definition of p, $\tilde{\boldsymbol{\theta}} \in \Theta_p \setminus \Theta_{p+1}$. There exists $\ell \in \{1, \dots, L_p\}$ such that $\tilde{\boldsymbol{\theta}} \in B^{(p,\ell)}$ and a look at the proof of Lemma 5.1 shows that on the event $\Omega_\xi \cap \mathcal{C}_n(\vartheta)$,

$$\beta_2 h^2(f_{\boldsymbol{\theta}_0}, f_{\tilde{\boldsymbol{\theta}}}) + \beta_3 \frac{D_{\mathscr{F}}^{(n)}}{n} + \frac{r_0^2 \beta_1}{9} \geq \beta_1\big(h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}^{(p,\ell)}}) + h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}'^{(p,\ell)}})\big). \tag{35}$$

Without lost of generality, we may suppose that $T(\boldsymbol{\theta}^{(p,\ell)}, \boldsymbol{\theta}'^{(p,\ell)}) = \overline{T}(f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})})$ is non-negative and $\tilde{\boldsymbol{\theta}} \in \mathcal{B}(\boldsymbol{\theta}^{(p,\ell)}, r_{p,\ell})$. Now, by using the triangular inequality and the fact that for all $\boldsymbol{\theta} \in \Theta$,

$$h^2(f_{\boldsymbol{\theta}}, f_{\pi(\boldsymbol{\theta})}) \leq \sup_{1\leq j\leq d} \overline{R}_j \varepsilon_j^2 = \frac{1}{n} \sup_{1\leq j\leq d} \left(t_j^{(n)}\right)^2 \leq d/n,$$

we get

$$h^2(f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}, f_{\boldsymbol{\theta}_0}) \leq 3h^2(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\tilde{\boldsymbol{\theta}}}) + 3h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}_0}) + 3d/n.$$

We use (35) and then (34) to deduce

$$\begin{aligned}
h^2(f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}, f_{\boldsymbol{\theta}_0}) &\leq 3(1+\beta_2/\beta_1)h^2(f_{\boldsymbol{\theta}_0}, f_{\tilde{\boldsymbol{\theta}}}) + 3(\beta_3/\beta_1)\frac{D_{\mathscr{F}}^{(n)}}{n} + \frac{r_0^2}{3} + \frac{3d}{n} \\
&\leq \frac{2r_0^2}{3} + 3(\beta_3/\beta_1)\frac{D_{\mathscr{F}}^{(n)}}{n} + \frac{3d}{n}.
\end{aligned}$$

Since $3(\beta_3/\beta_1)D_{\mathscr{F}}^{(n)}/n + 3d/n$ tends to 0 when n goes to infinity, there exists $n_0 \in \mathbb{N}^\star$ such that for all $n \geq n_0$, $h^2(f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}, f_{\boldsymbol{\theta}_0}) \leq r_0^2$. Similarly, the bound $h^2(f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}, f_{\boldsymbol{\theta}_0}) \leq r_0^2$ also holds. In particular, $\pi(\boldsymbol{\theta}^{(p,\ell)})$ and $\pi(\boldsymbol{\theta}'^{(p,\ell)})$ belong together to $\Theta_3(\tau)$. We can therefore use (28) to get

$$\begin{aligned}
\overline{T}(f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}) \leq &\left(\frac{8+5\sqrt{2}}{7} + \tau\right)h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}) \\
&- (8-5\sqrt{2}-\tau)h^2(f_{\tilde{\boldsymbol{\theta}}}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}).
\end{aligned}$$

Since $\overline{T}(f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})})$ is non-negative, we may replace τ by its numerical value $\tau = 0.01$ to get

$$h(f_{\tilde{\boldsymbol{\theta}}}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}) \leq 1.6h(f_{\tilde{\boldsymbol{\theta}}}, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}).$$

Therefore,

$$\begin{aligned}
h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) &\leq h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\tilde{\boldsymbol{\theta}}}) + h(f_{\tilde{\boldsymbol{\theta}}}, f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}) + h(f_{\pi(\boldsymbol{\theta}'^{(p,\ell)})}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) \\
&\leq h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\tilde{\boldsymbol{\theta}}}) + 1.6h(f_{\tilde{\boldsymbol{\theta}}}, f_{\pi(\boldsymbol{\theta}^{(p,\ell)})}) + \sqrt{\frac{1}{n}\sup_{1\leq j\leq d}\left(t_j^{(n)}\right)^2} \\
&\leq h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\tilde{\boldsymbol{\theta}}}) + 1.6\left[h(f_{\tilde{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}^{(p,\ell)}}) + \sqrt{\frac{1}{n}\sup_{1\leq j\leq d}\left(t_j^{(n)}\right)^2}\right] + \sqrt{\frac{1}{n}\sup_{1\leq j\leq d}\left(t_j^{(n)}\right)^2} \\
&\leq 2.6h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\tilde{\boldsymbol{\theta}}}) + 2.6\sqrt{\frac{1}{n}\sup_{1\leq j\leq d}\left(t_j^{(n)}\right)^2}.
\end{aligned}$$

Since $\tilde{\boldsymbol{\theta}} \in \mathcal{B}(\boldsymbol{\theta}^{(p,\ell)}, r_{p,\ell})$,

$$h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) \leq 2.6\bar{\kappa}^{1/2} h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) + 2.6\sqrt{\frac{1}{n} \sup_{1 \leq j \leq d} \bigl(t_j^{(n)}\bigr)^2}.$$

By replacing $\bar{\kappa}$ by its numerical value $\bar{\kappa} = 3/2 - \sqrt{2}$,

$$h(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) \leq 10.91\sqrt{\frac{1}{n} \sup_{1 \leq j \leq d} \bigl(t_j^{(n)}\bigr)^2}.$$

Let now $\boldsymbol{\theta}$ be any element of Θ_{N_n}. Since $\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}$ belong together to Θ_p,

$$\delta^2(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \leq \sup_{\boldsymbol{\theta}', \boldsymbol{\theta}'' \in \Theta_p} \delta^2\bigl(\boldsymbol{\theta}', \boldsymbol{\theta}''\bigr) \leq h^2(f_{\boldsymbol{\theta}^{(p,\ell)}}, f_{\boldsymbol{\theta}'^{(p,\ell)}}) \leq \frac{120}{n} \sup_{1 \leq j \leq d} \bigl(t_j^{(n)}\bigr)^2.$$

Finally, we have shown that there exist $\xi > 0$, $\vartheta > 0$, $n_0 \in \mathbb{N}^\star$ such that for all $n \geq n_0$,

$$\mathbb{P}\left[\exists \tilde{\boldsymbol{\theta}} \in \Theta, \dot{L}(\tilde{\boldsymbol{\theta}}) = 0 \text{ and } \inf_{\boldsymbol{\theta} \in \Theta_{N_n}} \delta^2(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \leq \frac{120}{n} \sup_{1 \leq j \leq d} \bigl(t_j^{(n)}\bigr)^2\right] \geq \mathbb{P}\bigl[\Omega_\xi \cap \mathcal{C}_n(\vartheta)\bigr].$$

The theorem follows from $\mathbb{P}(\Omega_\xi) \geq 1 - \mathrm{e}^{-n\xi}$ and $\mathcal{C}_n(\vartheta) = \mathcal{A}_n(\vartheta) \cap \mathcal{A}'_n(\vartheta)$.

5.4. Proof of Theorem 2.1

It follows from Theorem 5.1, page 1640, where $d = 1$, $\alpha_1 = \alpha$, $\Theta_i = [\theta^{(i)}, \theta'^{(i)}]$, $L_i = 1$, $\underline{R}_1^{(n)} = \underline{R}$, $\overline{R}_1 = \overline{R}$ and $\delta^2(\theta, \theta') = \underline{R}|\theta - \theta'|^\alpha$ in the first part of the theorem and $\delta^2(\theta, \theta') = h^2(f_\theta, f_{\theta'})$ in the second part.

5.5. Proof of Proposition 2.1

For all $i \in \{1, \ldots, N-1\}$,

$$\theta^{(i+1)} \in \bigl\{\theta^{(i)}, \theta^{(i)} + \min\bigl(\overline{r}\bigl(\theta^{(i)}, \theta'^{(i)}\bigr), \bigl(\theta'^{(i)} - \theta^{(i)}\bigr)/2\bigr)\bigr\},$$
$$\theta'^{(i+1)} \in \bigl\{\theta'^{(i)}, \theta'^{(i)} - \min\bigl(\underline{r}\bigl(\theta^{(i)}, \theta'^{(i)}\bigr), \bigl(\theta'^{(i)} - \theta^{(i)}\bigr)/2\bigr)\bigr\}.$$

Since $\overline{r}(\theta^{(i)}, \theta'^{(i)})$ and $\underline{r}(\theta^{(i)}, \theta'^{(i)})$ are larger than $(\kappa \underline{R}/\overline{R})^{1/\alpha}(\theta'^{(i)} - \theta^{(i)})$,

$$\theta'^{(i+1)} - \theta^{(i+1)} \leq \max\bigl\{1 - (\kappa \underline{R}/\overline{R})^{1/\alpha}, 1/2\bigr\}\bigl(\theta'^{(i)} - \theta^{(i)}\bigr).$$

By induction, we derive that for all $i \in \{1, \ldots, N-1\}$,

$$\theta'^{(i+1)} - \theta^{(i+1)} \leq \bigl(\max\bigl\{1 - (\kappa \underline{R}/\overline{R})^{1/\alpha}, 1/2\bigr\}\bigr)^i (M - m).$$

Consequently, the algorithm converges in less than N iterations where N is the smallest integer such that

$$\big(\max\big\{1-(\kappa \underline{R}/\overline{R})^{1/\alpha}, 1/2\big\}\big)^N (M-m) \le \eta,$$

that is,

$$N \ge \frac{\log((M-m)/\eta)}{-\log[\max\{1-(\kappa \underline{R}/\overline{R})^{1/\alpha}, 1/2\}]}.$$

We conclude by using the inequality $-1/\log(1-x) \le 1/x$ for all $x \in (0,1)$.

5.6. Proof of Theorem 2.2

We shall apply Theorem 5.2, page 1643, with $\Theta_i = [\theta^{(i)}, \theta'^{(i)}]$, $L_i = 1$, $\delta^2(\theta,\theta') = \underline{R}|\theta-\theta'|^2$ and with N_n corresponding to the last step of the algorithm.

The proof that Assumption 2.1 holds can be derived from Theorem 3 of Section 31 of Borovkov [13]. Moreover, point (vi) of Assumption 5.1 is satisfied with the event $\mathcal{A}_n(\vartheta)$ on which

$$\frac{1}{n}\sum_{i=1}^n \varphi_1^3(X_i) \le 2\mathbb{E}\big[\varphi_1^3(X_1)\big] \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \varphi_2(X_i) \le 2\mathbb{E}\big[\varphi_2(X_1)\big]$$

and with

$$\Theta_0(\vartheta) = \left\{\theta \in (m,M), |\theta-\theta_0| \le \min\left\{\frac{1}{2}\left(\frac{\vartheta}{2\mathbb{E}[\varphi_1^3(X_1)]}\right)^{1/(3\gamma_1-2)}, \left(\frac{\vartheta}{2\mathbb{E}[\varphi_2(X_1)]}\right)^{1/\gamma_2}\right\}\right\}.$$

Theorem 5.2 then asserts that there exist $\vartheta > 0$, $\xi > 0$, $n_0 \in \mathbb{N}^\star$, such that for $n \ge n_0$:

$$\begin{aligned}&\mathbb{P}\left[\exists\tilde{\theta} \in (m,M), \sum_{i=1}^n \dot{l}_{\tilde{\theta}}(X_i) = 0 \text{ and } \inf_{\theta\in\Theta_{N_n}} \underline{R}(\theta-\tilde{\theta})^2 \le \frac{120}{n}\big(t^{(n)}\big)^2\right]\\ &\quad \ge 1-\big\{\mathbb{P}\big[\big(\mathcal{A}_n(\vartheta)\big)^c\big] + \mathbb{P}\big[\big(\mathcal{A}'_n(\vartheta)\big)^c\big] + \mathrm{e}^{-n\xi}\big\}.\end{aligned}$$

Recalling that $\theta'^{(N_n)} - \theta^{(N_n)} \le t^{(n)}(\overline{R}n)^{-1/2}$ and that $\hat{\theta}$ is the middle of the interval $\Theta_{N_n} = [\theta^{(N_n)}, \theta'^{(N_n)}]$,

$$\underline{R}(\hat{\theta}-\tilde{\theta})^2 \le 2\inf_{\theta\in\Theta_{N_n}} \underline{R}(\theta-\tilde{\theta})^2 + 2(\underline{R}/\overline{R})\frac{(t^{(n)})^2}{n}.$$

This shows that there exists $C > 0$ such that

$$\begin{aligned}&\mathbb{P}\left[\exists\tilde{\theta} \in (m,M), \sum_{i=1}^n \dot{l}_{\tilde{\theta}}(X_i) = 0 \text{ and } |\hat{\theta}-\tilde{\theta}| \le C\frac{t^{(n)}}{\sqrt{n}}\right]\\ &\quad \ge 1-\big\{\mathbb{P}\big[\big(\mathcal{A}_n(\vartheta)\big)^c\big] + \mathbb{P}\big[\big(\mathcal{A}'_n(\vartheta)\big)^c\big] + \mathrm{e}^{-n\xi}\big\}.\end{aligned}$$

Therefore, for all n, this probability is always larger than $1 - \zeta_n$ where

$$\zeta_n = \begin{cases} 1, & \text{if } n < n_0, \\ \min\{1, \mathbb{P}[(\mathcal{A}_n(\vartheta))^c] + \mathbb{P}[(\mathcal{A}'_n(\vartheta))^c] + \mathrm{e}^{-n\xi}\}, & \text{if } n \geq n_0. \end{cases}$$

By the law of large numbers, the two probabilities $\mathbb{P}[(\mathcal{A}_n(\vartheta))^c]$ and $\mathbb{P}[(\mathcal{A}'_n(\vartheta))^c]$ converge to 1 and therefore also the sequence $(\zeta_n)_{n\geq 1}$.

We now prove that $\hat{\theta}$ is asymptotically efficient. Let $\tilde{\theta}$ be an estimator satisfying $\sum_{i=1}^n \dot{l}_{\tilde{\theta}}(X_i) = 0$ and $|\hat{\theta} - \tilde{\theta}| \leq Ct^{(n)}/\sqrt{n}$ with probability tending to 1 when n goes to infinity. Let us consider for μ-almost all $x \in \mathbb{X}$,

$$R(x) = \int_0^1 \left(\ddot{l}_{\theta_0 + u(\tilde{\theta} - \theta_0)}(x) - \ddot{l}_{\theta_0}(x)\right) \mathrm{d}u.$$

Then

$$\dot{l}_{\tilde{\theta}}(x) = \dot{l}_{\theta_0}(x) + \ddot{l}_{\theta_0}(x)(\tilde{\theta} - \theta_0) + R(x)(\tilde{\theta} - \theta_0).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^n \dot{l}_{\tilde{\theta}}(X_i) = \frac{1}{n}\sum_{i=1}^n \dot{l}_{\theta_0}(X_i) + \frac{1}{n}\sum_{i=1}^n \ddot{l}_{\theta_0}(X_i)(\tilde{\theta} - \theta_0) + \left(\frac{1}{n}\sum_{i=1}^n R(X_i)\right)(\tilde{\theta} - \theta_0),$$

and hence

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &= \sqrt{n}(\hat{\theta} - \tilde{\theta}) + \sqrt{n}(\tilde{\theta} - \theta_0) \\ &= \sqrt{n}(\hat{\theta} - \tilde{\theta}) + \frac{(1/\sqrt{n})\sum_{i=1}^n \dot{l}_{\theta_0}(X_i) - (1/\sqrt{n})\sum_{i=1}^n \dot{l}_{\tilde{\theta}}(X_i)}{-(1/n)\sum_{i=1}^n \ddot{l}_{\theta_0}(X_i) - (1/n)\sum_{i=1}^n R(X_i)}. \end{aligned}$$

Now, with probability tending to 1 when n goes to infinity,

$$\begin{aligned} \left|\frac{1}{n}\sum_{i=1}^n R(X_i)\right| &\leq 2\mathbb{E}[\varphi_2(X_1)]|\tilde{\theta} - \theta_0|^{\gamma_2} \\ &\leq 2\mathbb{E}[\varphi_2(X_1)]\left(C\frac{t^{(n)}}{\sqrt{n}} + |\hat{\theta} - \theta_0|\right)^{\gamma_2}. \end{aligned}$$

Remark now that the term $D_{\mathscr{F}}/n$ involved in Theorem 2.1 tends to 0, which shows that $\hat{\theta}$ converges almost surely to θ_0. Therefore, $n^{-1}\sum_{i=1}^n R(X_i)$ converges to 0 in probability. Slutsky's theorem then shows that

$$\frac{(1/\sqrt{n})\sum_{i=1}^n \dot{l}_{\theta_0}(X_i) - (1/\sqrt{n})\sum_{i=1}^n \dot{l}_{\tilde{\theta}}(X_i)}{-(1/n)\sum_{i=1}^n \ddot{l}_{\theta_0}(X_i) - (1/n)\sum_{i=1}^n R(X_i)}$$

converges in distribution to $\mathcal{N}(0, 1/I(\theta_0))$. We then reuse Slutsky's theorem to prove the asymptotic efficiency of $\hat{\theta}$.

Suppose now that there exists $\lambda > 0$ such that $\mathbb{E}[\exp(\lambda\varphi_2(X_1))]$, $\mathbb{E}[\exp(\lambda|\dot{l}_{\theta_0}(X_1)|)]$ and $\mathbb{E}[\exp(\lambda|\ddot{l}_{\theta_0}(X_1)|)]$ are finite. Then the fact that $\mathbb{P}[(\mathcal{A}_n(\vartheta))^c]$ and $\mathbb{P}[(\mathcal{A}'_n(\vartheta))^c]$ go to 0 exponentially fast ensues from the following result which goes back to Cramér.

Lemma 5.2. *Let $Y_1, \dots, Y_n$ be n independent and identically distributed $\mathbb{R}$-valued random variables satisfying $\mathbb{E}[\exp(\lambda|Y_1|)] < \infty$ for some $\lambda > 0$. Then, for all $\vartheta > 0$, there exists $\sigma > 0$ such that*

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\big(Y_i - \mathbb{E}[Y_i]\big)\right| \geq \vartheta\right] \leq 2\mathrm{e}^{-\sigma n}.$$

Notice now that one can always replace φ_1 by $\varphi_1(x) = \sup_{\theta \in (m,M)} |\dot{l}_\theta(x)|$ and γ_1 by $\gamma_1 = 1$ since

$$\big|\log f_{\theta'}(x) - \log f_\theta(x)\big| \leq \Big(\sup_{\theta'' \in (m,M)} \big|\dot{l}_{\theta''}(x)\big|\Big)\big|\theta' - \theta\big|.$$

We shall show that there exists $\lambda_1 > 0$ such that $\mathbb{E}[\exp(\lambda_1 \varphi_1(X_1))] < \infty$. Since $\varphi_2(X_1)$ has also finite exponential moments, the preceding lemma will show that $\mathbb{P}[(\mathcal{A}_n(\vartheta))^c]$ goes to 0 exponentially fast. By setting $\lambda_0 = \lambda / \max\{2, 2(M-m)^{\gamma_2}\}$,

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big[\lambda_0 \sup_{\theta \in (m,M)} \big\{\big|\ddot{l}_\theta(X_1)\big|\big\}\Big]\Big] &\leq \mathbb{E}\Big[\exp\big[\lambda_0\big|\ddot{l}_{\theta_0}(X_1)\big|\big]\exp\Big[\lambda_0 \sup_{\theta \in (m,M)} \big|\ddot{l}_\theta(X_1) - \ddot{l}_{\theta_0}(X_1)\big|\Big]\Big] \\
&\leq \mathbb{E}\big[\exp\big[(\lambda/2)\big|\ddot{l}_{\theta_0}(X_1)\big|\big]\exp\big[(\lambda/2)\varphi_2(X_1)\big]\big] \\
&\leq \mathbb{E}^{1/2}\big[\exp\big(\lambda\big|\ddot{l}_{\theta_0}(X_1)\big)\big|\big]\mathbb{E}^{1/2}\big[\exp\big(\lambda\varphi_2(X_1)\big)\big] \\
&< \infty.
\end{aligned}$$

By setting $\lambda_1 = (1/2)\min\{\lambda, \lambda_0/(M-m)\}$,

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big[\lambda_1 \sup_{\theta \in (m,M)} \big|\dot{l}_\theta(X_1)\big|\Big]\Big] &\leq \mathbb{E}\Big[\exp\big[\lambda_1\big|\dot{l}_{\theta_0}(X_1)\big|\big]\exp\Big[\lambda_1 \sup_{\theta \in (m,M)} \big|\ddot{l}_\theta(X_1)\big|(M-m)\Big]\Big] \\
&\leq \mathbb{E}^{1/2}\big[\exp\big(\lambda\big|\dot{l}_{\theta_0}(X_1)\big|\big)\big]\mathbb{E}^{1/2}\Big[\exp\Big[\lambda_0 \sup_{\theta \in (m,M)} \big|\ddot{l}_\theta(X_1)\big|\Big]\Big] \\
&< \infty,
\end{aligned}$$

which completes the proof.

Appendix: Proof of Theorem 4.1

This theorem follows from Theorem 5.1 as explained in Section 5.1.1. It remains to prove that its assumptions are fulfilled, that is, that (20) and (21) hold.

For this purpose, remark that the different parameters $\underline{r}_j$, $\overline{r}_j$, ϱ_j, ϱ'_j, $\varrho_k, \dots$ that have been introduced in Algorithm 2 depend on the set Θ_i and may vary at each iteration of the until loop. We need to make explicit this dependency in order to prove rigorously (20) and (21). Unfortunately, this makes the algorithm more difficult to read.

Algorithm 4 Rewriting of Algorithm 2

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

Choose $k^{(i)} \in \{1, \dots, d\}$ such that

$$\underline{R}_{\Theta_i, k^{(i)}} \big(b_{k^{(i)}}^{(i)} - a_{k^{(i)}}^{(i)}\big)^{\alpha_{k^{(i)}}} = \max_{1 \le j \le d} \underline{R}_{\Theta_i, j} \big(b_j^{(i)} - a_j^{(i)}\big)^{\alpha_j}.$$

$\boldsymbol{\theta}^{(i,1)} = (a_1^{(i)}, \dots, a_d^{(i)})$
$\boldsymbol{\theta}'^{(i,1)} = (a_1^{(i)}, \dots, a_{k^{(i)}-1}^{(i)}, b_{k^{(i)}}^{(i)}, a_{k^{(i)}+1}^{(i)}, a_d^{(i)})$
$\varrho_j^{(i,0)} = \overline{r}_{\Theta_i, j}(\boldsymbol{\theta}^{(i,1)}, \boldsymbol{\theta}'^{(i,1)})$ and $\varrho_j'^{(i,0)} = \overline{r}_{\Theta_i, j}(\boldsymbol{\theta}'^{(i,1)}, \boldsymbol{\theta}^{(i,1)})$ for all $j \neq k^{(i)}$
$\varrho_{k^{(i)}}^{(i,0)} = (b_{k^{(i)}}^{(i)} - a_{k^{(i)}}^{(i)})/2$ and $\varrho_{k^{(i)}}'^{(i,0)} = (b_{k^{(i)}}^{(i)} - a_{k^{(i)}}^{(i)})/2$
for all $\ell \geq 1$ **do**
  if $T(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}) \geq 0$ **then**
    $\varrho_{\psi_{k^{(i)}}(1)}^{(i,\ell)} = \overline{r}_{\Theta_i, \psi_{k^{(i)}}(1)}(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)})$
    $\varrho_{\psi_{k^{(i)}}(j)}^{(i,\ell)} = \min(\varrho_{\psi_{k^{(i)}}(j)}^{(i,\ell-1)}, \overline{r}_{\Theta_i, \psi_{k^{(i)}}(j)}(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}))$, for all $j \in \{2, \dots, d-1\}$
    $\varrho_{k^{(i)}}^{(i,\ell)} = \min(\varrho_{k^{(i)}}^{(i,\ell-1)}, \overline{r}_{\Theta_i, k^{(i)}}(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}))$
    $\mathfrak{J}^{(i,\ell)} = \{1 \le j \le d-1, \theta_{\psi_{k^{(i)}}(j)}^{(i,\ell)} + \varrho_{\psi_{k^{(i)}}(j)}^{(i,\ell)} < b_{\psi_{k^{(i)}}(j)}^{(i)}\}$
    if $\mathfrak{J}^{(i,\ell)} \neq \varnothing$ **then**
      $\mathrm{j}_{\min}^{(i,\ell)} = \min \mathfrak{J}^{(i,\ell)}$
      Define $\boldsymbol{\theta}^{(i,\ell+1)}$ as

$$\begin{cases} \theta_{\psi_{k^{(i)}}(j)}^{(i,\ell+1)} = a_{\psi_{k^{(i)}}(j)}^{(i)}, & \text{for all } j < \mathrm{j}_{\min}^{(i,\ell)}, \\ \theta_{\psi_{k^{(i)}}(\mathrm{j}_{\min}^{(i,\ell)})}^{(i,\ell+1)} = \theta_{\psi_{k^{(i)}}(\mathrm{j}_{\min}^{(i,\ell)})}^{(i,\ell)} + \varrho_{\psi_{k^{(i)}}(\mathrm{j}_{\min}^{(i,\ell)})}^{(i,\ell)}, & \\ \theta_{\psi_{k^{(i)}}(j)}^{(i,\ell+1)} = \theta_{\psi_{k^{(i)}}(j)}^{(i,\ell)}, & \text{for all } j > \mathrm{j}_{\min}^{(i,\ell)}, \\ \theta_{k^{(i)}}^{(i,\ell+1)} = a_{k^{(i)}}^{(i)} & \end{cases}$$

    else
      Define $\boldsymbol{\theta}^{(i,\ell+1)} = \boldsymbol{\theta}^{(i,\ell)}$
      $\mathrm{j}_{\min}^{(i,\ell)} = d$
    end if
  end if
  if $T(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}) \leq 0$ **then**
    $\varrho_{\psi_{k^{(i)}}(1)}'^{(i,\ell)} = \overline{r}_{\Theta_i, \psi_{k^{(i)}}(1)}(\boldsymbol{\theta}'^{(i,\ell)}, \boldsymbol{\theta}^{(i,\ell)})$
    $\varrho_{\psi_{k^{(i)}}(j)}'^{(i,\ell)} = \min(\varrho_{\psi_{k^{(i)}}(j)}'^{(i,\ell-1)}, \overline{r}_{\Theta_i, \psi_{k^{(i)}}(j)}(\boldsymbol{\theta}'^{(i,\ell)}, \boldsymbol{\theta}^{(i,\ell)}))$, for all $j \in \{2, \dots, d-1\}$
    $\varrho_{k^{(i)}}'^{(i,\ell)} = \min(\varrho_{k^{(i)}}'^{(i,\ell-1)}, \underline{r}_{\Theta_i, k^{(i)}}(\boldsymbol{\theta}'^{(i,\ell)}, \boldsymbol{\theta}^{(i,\ell)}))$
    $\mathfrak{J}'^{(i,\ell)} = \{1 \le j \le d-1, \theta_{\psi_{k^{(i)}}(j)}'^{(i,\ell)} + \varrho_{\psi_{k^{(i)}}(j)}'^{(i,\ell)} < b_{\psi_{k^{(i)}}(j)}^{(i)}\}$
    if $\mathfrak{J}'^{(i,\ell)} \neq \varnothing$ **then**

$\mathrm{j}'^{(i,\ell)}_{\min} = \min \mathfrak{J}'^{(i,\ell)}$

Define $\boldsymbol{\theta}'^{(i,\ell+1)}$ as

$$\begin{cases} \theta'^{(i,\ell+1)}_{\psi_{k^{(i)}}(j)} = a^{(i)}_{\psi_{k^{(i)}}(j)}, & \text{for all } j < \mathrm{j}'^{(i,\ell)}_{\min}, \\ \theta'^{(i,\ell+1)}_{\psi_{k^{(i)}}(\mathrm{j}'^{(i,\ell)}_{\min})} = \theta'^{(i,\ell)}_{\psi_{k^{(i)}}(\mathrm{j}'^{(i,\ell)}_{\min})} + \varrho'^{(i,\ell)}_{\psi_{k^{(i)}}(\mathrm{j}'^{(i,\ell)}_{\min})}, & \\ \theta'^{(i,\ell+1)}_{\psi_{k^{(i)}}(j)} = \theta'^{(i,\ell)}_{\psi_{k^{(i)}}(j)}, & \text{for all } j > \mathrm{j}'^{(i,\ell)}_{\min}, \\ \theta'^{(i,\ell+1)}_{k^{(i)}} = b^{(i)}_{k^{(i)}} & \end{cases}$$

else
$\boldsymbol{\theta}'^{(i,\ell+1)} = \boldsymbol{\theta}'^{(i,\ell)}$
$\mathrm{j}'^{(i,\ell)}_{\min} = d$
end if
end if
if $\mathrm{j}^{(i,\ell)}_{\min} = d$ or $\mathrm{j}'^{(i,\ell)}_{\min} = d$ **then**
$L_i = \ell$ and quit the loop
end if
end for
if $\mathrm{j}^{(i,\ell)}_{\min} = d$ **then**
$a^{(i+1)}_{k^{(i)}} = a^{(i)}_{k^{(i)}} + \varrho^{(i,L_i)}_{k^{(i)}}$
end if
if $\mathrm{j}'^{(i,\ell)}_{\min} = d$ **then**
$b^{(i+1)}_{k^{(i)}} = b^{(i)}_{k^{(i)}} - \varrho'^{(i,L_i)}_{k^{(i)}}$
end if
$a^{(i+1)}_j = a^{(i)}_j$ and $b^{(i+1)}_j = b^{(i)}_j$ for all $j \neq k^{(i)}$
Return: $\Theta_{i+1} = \prod_{j=1}^d [a^{(i+1)}_j, b^{(i+1)}_j]$

Algorithm 5 Rewriting of Algorithm 3

$\Theta_1 = \prod_{j=1}^d [a^{(1)}_j, b^{(1)}_j] = \prod_{j=1}^d [m_j, M_j]$
for all $i \geq 1$ **do**
if there exists $j \in \{1, \dots, d\}$ such that $b^{(i)}_j - a^{(i)}_j > \eta_j$ **then**
Compute Θ_{i+1}
else
Leave the loop and set $N = i$
end if
end for
Return:

$$\hat{\boldsymbol{\theta}} = \left(\frac{a^{(N)}_1 + b^{(N)}_1}{2}, \dots, \frac{a^{(N)}_d + b^{(N)}_d}{2} \right)$$

We begin by proving that (21) holds.

Lemma A.1. *For all $i \in \{1, \dots, N-1\}$ and $\ell \in \{1, \dots, L_i\}$,*

$$\sup_{\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_i} \delta^2\big(\boldsymbol{\theta}, \boldsymbol{\theta}'\big) \le h^2(f_{\boldsymbol{\theta}^{(i,\ell)}}, f_{\boldsymbol{\theta}'^{(i,\ell)}}).$$

Proof. Recalling that $\underline{R}_j \le \underline{R}_{\Theta_i, j}$,

$$\begin{aligned} \sup_{\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_i} \delta^2\big(\boldsymbol{\theta}, \boldsymbol{\theta}'\big) &\le \sup_{1 \le j \le d} \underline{R}_{\Theta_i, j}\big(b_j^{(i)} - a_j^{(i)}\big)^{\alpha_j} \\ &\le \underline{R}_{\Theta_i, k^{(i)}}\big(b_{k^{(i)}}^{(i)} - a_{k^{(i)}}^{(i)}\big)^{\alpha_{k^{(i)}}}. \end{aligned}$$

Now, $\theta_{k^{(i)}}^{(i,\ell)} = a_{k^{(i)}}^{(i)}$ and $\theta_{k^{(i)}}^{\prime(i,\ell)} = b_{k^{(i)}}^{(i)}$, and thus

$$\begin{aligned} \sup_{\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_i} \delta^2\big(\boldsymbol{\theta}, \boldsymbol{\theta}'\big) &\le \underline{R}_{\Theta_i, k^{(i)}}\big(\theta_{k^{(i)}}^{\prime(i,\ell)} - \theta_{k^{(i)}}^{(i,\ell)}\big)^{\alpha_{k^{(i)}}} \\ &\le \sup_{1 \le j \le d} \underline{R}_{\Theta_i, j}\big(\theta_j^{\prime(i,\ell)} - \theta_j^{(i,\ell)}\big)^{\alpha_j} \\ &\le h^2(f_{\boldsymbol{\theta}^{(i,\ell)}}, f_{\boldsymbol{\theta}'^{(i,\ell)}}). \end{aligned}$$

□

We now show that (20) holds:

Lemma A.2. *For all $i \in \{1, \dots, N-1\}$,*

$$\Theta_i \Big\backslash \bigcup_{\ell=1}^{L_i} B^{(i,\ell)} \subset \Theta_{i+1} \subset \Theta_i.$$

Proof. Since

$$\begin{aligned} \varrho_{k^{(i)}}^{(i,L_i)} &\le \frac{b_{k^{(i)}}^{(i)} - a_{k^{(i)}}^{(i)}}{2} \quad \text{and} \\ \varrho_{k^{(i)}}^{\prime(i,L_i)} &\le \frac{b_{k^{(i)}}^{(i)} - a_{k^{(i)}}^{(i)}}{2}, \end{aligned}$$

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

$$\mathcal{R}_1^{\prime(i,\ell)} = \prod_{q=1}^{d} \big[\theta_q^{(i,\ell)}, \theta_q^{(i,\ell)} + \varrho_q^{(i,\ell)}\big],$$

$$\mathcal{R}_2'^{(i,\ell)} = \prod_{q=1}^{k^{(i)}-1} \left[\theta_q'^{(i,\ell)}, \theta_q'^{(i,\ell)} + \varrho_q'^{(i,\ell)}\right] \times \left[\theta_{k^{(i)}}'^{(i,\ell)} - \varrho_{k^{(i)}}'^{(i,\ell)}, \theta_{k^{(i)}}'^{(i,\ell)}\right]$$
$$\times \prod_{q=k^{(i)}+1}^{d} \left[\theta_q'^{(i,\ell)}, \theta_q'^{(i,\ell)} + \varrho_q'^{(i,\ell)}\right]$$

and we set

$$\mathcal{R}_3'^{(i,\ell)} = \begin{cases} \mathcal{R}_1'^{(i,\ell)}, & \text{if } T\big(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}\big) > 0, \\ \mathcal{R}_2'^{(i,\ell)}, & \text{if } T\big(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}\big) < 0, \\ \mathcal{R}_1'^{(i,\ell)} \cup \mathcal{R}_2'^{(i,\ell)}, & \text{if } T\big(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}\big) = 0. \end{cases}$$

Using that $\Theta_i \cap \mathcal{R}_1'^{(i,\ell)} \subset \mathcal{R}(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)})$, $\Theta_i \cap \mathcal{R}_2'^{(i,\ell)} \subset \mathcal{R}(\boldsymbol{\theta}'^{(i,\ell)}, \boldsymbol{\theta}^{(i,\ell)})$ together with (18) yields $\Theta_i \cap \mathcal{R}_3'^{(i,\ell)} \subset B^{(i,\ell)}$. It is then sufficient to show

$$\Theta_i \Big\backslash \bigcup_{\ell=1}^{L_i} \mathcal{R}_3'^{(i,\ell)} \subset \Theta_{i+1}.$$

Note that either $T(\boldsymbol{\theta}^{(i,L_i)}, \boldsymbol{\theta}'^{(i,L_i)}) \geq 0$ or $T(\boldsymbol{\theta}^{(i,L_i)}, \boldsymbol{\theta}'^{(i,L_i)}) \leq 0$. In what follows, we assume that $T(\boldsymbol{\theta}^{(i,L_i)}, \boldsymbol{\theta}'^{(i,L_i)}) \geq 0$ but the proof is similar if $T(\boldsymbol{\theta}^{(i,L_i)}, \boldsymbol{\theta}'^{(i,L_i)})$ is non-positive. Without lost of generality, and for the sake of simplicity, we suppose that $k^{(i)} = d$ and $\psi_d(j) = j$ for all $j \in \{1, \dots, d-1\}$. Let

$$\mathcal{L} = \left\{1 \leq \ell \leq L_i, T\big(\boldsymbol{\theta}^{(i,\ell)}, \boldsymbol{\theta}'^{(i,\ell)}\big) \geq 0\right\}$$

and $\ell_1 < \dots < \ell_r$ be the elements of $\mathcal{L}$. It is sufficient to prove that

$$\Theta_i \Big\backslash \bigcup_{\ell=1}^{L_i} \mathcal{R}_3'^{(i,\ell)} \subset \prod_{q=1}^{d-1} \big[a_q^{(i)}, b_q^{(i)}\big] \times \big[a_d^{(i)} + \varrho_d^{(i,L_i)}, b_d^{(i)}\big]. \tag{36}$$

We shall actually prove

$$\prod_{q=1}^{d-1} \big[a_q^{(i)}, b_q^{(i)}\big] \times \big[a_d^{(i)}, a_d^{(i)} + \varrho_d^{(i,L_i)}\big] \subset \bigcup_{k=1}^{r} \mathcal{R}_1'^{(i,\ell_k)},$$

which, in particular, implies (36). Remark now that for all $k \in \{1, \dots, r\}$, $\theta_d^{(i,\ell_k)} = a_d^{(i)}$, and thus

$$\mathcal{R}_1'^{(i,\ell_k)} = \prod_{q=1}^{d-1} \big[\theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varrho_q^{(i,\ell_k)}\big] \times \big[a_d^{(i)}, a_d^{(i)} + \varrho_d^{(i,\ell_k)}\big].$$

By using the fact that the sequence $(\varrho_d^{(i,\ell_k)})_k$ is non-increasing,

$$\big[a_d^{(i)}, a_d^{(i)} + \varrho_d^{(i,L_i)}\big] \subset \bigcap_{k=1}^{r} \big[a_d^{(i)}, a_d^{(i)} + \varrho_d^{(i,\ell_k)}\big].$$

This means that we only need to show

$$\prod_{q=1}^{d-1} \big[a_q^{(i)}, b_q^{(i)}\big] \subset \bigcup_{k=1}^{r} \prod_{q=1}^{d-1} \big[\theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varrho_q^{(i,\ell_k)}\big]. \tag{37}$$

Let us now define for all $p \in \{1, \dots, d-1\}$, $k_{p,0} = 0$ and by induction for all integer $\mathfrak{m}$,

$$k_{p,\mathfrak{m}+1} = \begin{cases} \inf\big\{k > k_{p,\mathfrak{m}}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p\big\}, \\ \qquad \text{if there exists } k \in \{k_{p,\mathfrak{m}}+1, \dots, r\} \text{ such that } \mathfrak{j}_{\min}^{(i,\ell_k)} > p, \\ r, \qquad \text{otherwise.} \end{cases}$$

Let $\mathfrak{M}_p$ be the smallest integer $\mathfrak{m}$ such that $k_{p,\mathfrak{m}} = r$. Let then for all $\mathfrak{m} \in \{0, \dots, \mathfrak{M}_p - 1\}$,

$$K_{p,\mathfrak{m}} = \{k_{p,\mathfrak{m}}+1, \dots, k_{p,\mathfrak{m}+1}\}.$$

We need the two following claims.

Claim A.1. *For all $\mathfrak{m} \in \{0, \dots, \mathfrak{M}_{p+1} - 1\}$, there exists $\mathfrak{m}' \in \{0, \dots, \mathfrak{M}_p - 1\}$ such that $k_{p,\mathfrak{m}'+1} \in K_{p+1,\mathfrak{m}}$.*

Proof. The set $\{\mathfrak{m}' \in \{0, \dots, \mathfrak{M}_p - 1\}, k_{p,\mathfrak{m}'+1} \le k_{p+1,\mathfrak{m}+1}\}$ is non-empty and we can thus define the largest integer $\mathfrak{m}'$ of $\{0, \dots, \mathfrak{M}_p - 1\}$ such that $k_{p,\mathfrak{m}'+1} \le k_{p+1,\mathfrak{m}+1}$. We then have

$$k_{p,\mathfrak{m}'} = \sup\big\{k < k_{p,\mathfrak{m}'+1}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p\big\}.$$

Since $k_{p,\mathfrak{m}'} < k_{p+1,\mathfrak{m}+1}$,

$$\begin{aligned} k_{p,\mathfrak{m}'} &= \sup\big\{k < k_{p+1,\mathfrak{m}+1}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p\big\} \\ &\ge \sup\big\{k < k_{p+1,\mathfrak{m}+1}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p+1\big\} \\ &\ge k_{p+1,\mathfrak{m}}. \end{aligned}$$

Hence, $k_{p,\mathfrak{m}'+1} \ge k_{p,\mathfrak{m}'} + 1 \ge k_{p+1,\mathfrak{m}} + 1$. Finally, $k_{p,\mathfrak{m}'+1} \in K_{p,\mathfrak{m}}$. □

Claim A.2. *Let $\mathfrak{m}' \in \{0, \dots, \mathfrak{M}_{p+1} - 1\}$, $p \in \{1, \dots, d-1\}$. There exists a subset $\mathcal{M}$ of $\{0, \dots, \mathfrak{M}_p - 1\}$ such that*

$$K_p' = \{k_{p,\mathfrak{m}+1}, \mathfrak{m} \in \mathcal{M}\} \subset K_{p+1,\mathfrak{m}'}$$

and

$$\left[a_{p+1}^{(i)}, b_{p+1}^{(i)}\right] \subset \bigcup_{k \in K'_p} \left[\theta_{p+1}^{(i,\ell_k)}, \theta_{p+1}^{(i,\ell_k)} + \varrho_{p+1}^{(i,\ell_k)}\right].$$

Proof. Thanks to Claim A.1, we can define the smallest integer $\mathfrak{m}_0$ of $\{0, \dots, \mathfrak{M}_p - 1\}$ such that $k_{p,\mathfrak{m}_0+1} \in K_{p+1,\mathfrak{m}'}$, and the largest integer $\mathfrak{m}_1$ of $\{0, \dots, \mathfrak{M}_p - 1\}$ such that $k_{p,\mathfrak{m}_1+1} \in K_{p+1,\mathfrak{m}'}$. Define now

$$\mathcal{M} = \{\mathfrak{m}_0, \mathfrak{m}_0 + 1, \dots, \mathfrak{m}_1\}.$$

Note that for all $\mathfrak{m} \in \{\mathfrak{m}_0, \dots, \mathfrak{m}_1\}$, $k_{p,\mathfrak{m}+1} \in K_{p+1,\mathfrak{m}'}$ (this ensues from the fact that the sequence $(k_{p,\mathfrak{m}})_{\mathfrak{m}}$ is increasing).

Let $\mathfrak{m} \in \{0, \dots, \mathfrak{M}_p - 1\}$ be such that $k_{p,\mathfrak{m}} \in K_{p+1,\mathfrak{m}'}$ and $k_{p,\mathfrak{m}} \neq k_{p+1,\mathfrak{m}'+1}$. Then $\mathfrak{j}_{\min}^{(i,\ell_{k_{p,\mathfrak{m}}})} \leq p + 1$ and since $\mathfrak{j}_{\min}^{(i,\ell_{k_{p,\mathfrak{m}}})} > p$, we get $\mathfrak{j}_{\min}^{(i,\ell_{k_{p,\mathfrak{m}}})} = p + 1$. Consequently,

$$\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}+1})} = \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}})}.$$

Now, $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}+1})} = \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}$ since $k_{p,\mathfrak{m}} + 1$ and $k_{p,\mathfrak{m}+1}$ belong together to $K_{p,\mathfrak{m}}$. The set

$$\left[\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}})}, \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}})}\right] \cup \left[\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}, \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}\right]$$

is thus the interval

$$\left[\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}}})}, \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}\right].$$

We apply this argument to each $\mathfrak{m} \in \{\mathfrak{m}_0 + 1, \dots, \mathfrak{m}_1\}$ to derive that the set

$$I = \bigcup_{\mathfrak{m}=\mathfrak{m}_0}^{\mathfrak{m}_1} \left[\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}, \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}\right]$$

is the interval

$$I = \left[\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0+1}})}, \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})}\right].$$

The claim is proved if we show that

$$\left[a_{p+1}^{(i)}, b_{p+1}^{(i)}\right] \subset I.$$

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 $a_{p+1}^{(i)} = \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0+1}})}$. If $k_{p+1,\mathfrak{m}'} = 0$, then $\mathfrak{m}' = 0$ and $\mathfrak{m}_0 = 0$. Besides, since 1 and $k_{p,1}$ belong to $K_{p,0}$, we have $\theta_{p+1}^{(i,\ell_{k_{p,1}})} = \theta_{p+1}^{(i,\ell_1)}$. Now, $\theta_{p+1}^{(i,\ell_1)} = a_{p+1}^{(i)}$ and thus $a_{p+1}^{(i)} \in I$. We now assume that $k_{p+1,\mathfrak{m}'} \neq 0$. Since $k_{p,\mathfrak{m}_0} \leq k_{p+1,\mathfrak{m}'}$, there are two cases.

- First case: $k_{p,\mathfrak{m}_0} = k_{p+1,\mathfrak{m}'}$. We then have $\mathrm{j}_{\min}^{(i,\ell_{k_{p,\mathfrak{m}_0}})} > p+1$ and thus $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0}+1})} = a_{p+1}^{(i)}$. Since $k_{p,\mathfrak{m}_0+1}$ and $k_{p,\mathfrak{m}_0}+1$ belong to $K_{p,\mathfrak{m}_0}$, $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0+1}})} = \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0}+1})}$ and thus $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0+1}})} = a_{p+1}^{(i)}$ as wished.
- Second case: $k_{p,\mathfrak{m}_0}+1 \le k_{p+1,\mathfrak{m}'}$. Then $k_{p+1,\mathfrak{m}'} \in K_{p,\mathfrak{m}_0}$, and thus

$$\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0}+1})} = \theta_{p+1}^{(i,\ell_{k_{p+1,\mathfrak{m}'}})}.$$

Since $\mathrm{j}_{\min}^{(i,\ell_{k_{p+1,\mathfrak{m}'}})} > p+1$, we have $\theta_{p+1}^{(i,\ell_{k_{p+1,\mathfrak{m}'}})} + \varrho_{p+1}^{(i,\ell_{k_{p+1,\mathfrak{m}'}})} \ge b_{p+1}^{(i)}$. By using the fact that the sequence $(\varrho_{p+1}^{(i,\ell_k)})_k$ is decreasing, we then deduce

$$\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0}+1})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0}+1})} \ge b_{p+1}^{(i)}$$

and thus $\mathrm{j}_{\min}^{(i,\ell_{k_{p,\mathfrak{m}_0}+1})} > p+1$. This proves that

$$\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0}+2})} = a_{p+1}^{(i)}. \tag{38}$$

Let us now show that $k_{p,\mathfrak{m}_0}+2 \le k_{p,\mathfrak{m}_0+1}$. If this is not true, $k_{p,\mathfrak{m}_0}+2 \ge k_{p,\mathfrak{m}_0+1}+1$, and thus $k_{p,\mathfrak{m}_0}+1 \ge k_{p,\mathfrak{m}_0+1}$ which means that $k_{p,\mathfrak{m}_0}+1 = k_{p,\mathfrak{m}_0+1}$ (we recall that $(k_{p,\mathfrak{m}})_{\mathfrak{m}}$ is an increasing sequence of integers). Since we are in the case where $k_{p,\mathfrak{m}_0}+1 \le k_{p+1,\mathfrak{m}'}$, we have $k_{p,\mathfrak{m}_0+1} \le k_{p+1,\mathfrak{m}'}$ which is impossible since $k_{p,\mathfrak{m}_0+1} \in K_{p+1,\mathfrak{m}'}$.

Therefore, we use that $k_{p,\mathfrak{m}_0}+2 \le k_{p,\mathfrak{m}_0+1}$ to get $k_{p,\mathfrak{m}_0}+2 \in K_{p,\mathfrak{m}_0}$, and thus $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0+1}})} = \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0}+2})}$. We then deduce from (38) that $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_0+1}})} = a_{p+1}^{(i)}$ as wished.

We now show that $b_{p+1}^{(i)} \in I$ by showing that $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} \ge b_{p+1}^{(i)}$. If $\mathfrak{m}_1 = \mathfrak{M}_p - 1$,

$$\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} = \theta_{p+1}^{(i,\ell_r)} + \varrho_{p+1}^{(i,\ell_r)} = \theta_{p+1}^{(i,L_i)} + \varrho_{p+1}^{(i,L_i)}.$$

Since $\mathfrak{J}^{(i,L_i)} = \varnothing$, we have $\theta_{p+1}^{(i,L_i)} + \varrho_{p+1}^{(i,L_i)} \ge b_{p+1}^{(i)}$, which proves the result.

We now assume that $\mathfrak{m}_1 < \mathfrak{M}_p - 1$. We begin to prove that $k_{p,\mathfrak{m}_1+1} = k_{p+1,\mathfrak{m}'+1}$. If this equality does not hold, we derive from the inequalities $k_{p,\mathfrak{m}_1+1} \le k_{p+1,\mathfrak{m}'+1} < k_{p,\mathfrak{m}_1+2}$, that $k_{p,\mathfrak{m}_1+1}+1 \le k_{p+1,\mathfrak{m}'+1}$ and thus $k_{p+1,\mathfrak{m}'+1} \in K_{p,\mathfrak{m}_1+1}$. Since $\mathrm{j}_{\min}^{(i,\ell_{k_{p+1,\mathfrak{m}'+1}})} > p+1$,

$$\theta_{p+1}^{(i,\ell_{k_{p+1,\mathfrak{m}'+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p+1,\mathfrak{m}'+1}})} \ge b_{p+1}^{(i)}.$$

Hence,

$$\theta_{p+1}^{(i,\ell_{(k_{p,\mathfrak{m}_1+1})+1})} + \varrho_{p+1}^{(i,\ell_{(k_{p,\mathfrak{m}_1+1})+1})} \ge b_{p+1}^{(i)} \qquad \text{which implies } \mathrm{j}_{\min}^{(i,\ell_{(k_{p,\mathfrak{m}_1+1})+1})} > p+1.$$

Since

$$k_{p+1,\mathfrak{m}'+1} = \inf\{k > k_{p+1,\mathfrak{m}'}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p+1\}$$

and $k_{p,\mathfrak{m}_1+1} + 1 > k_{p+1,\mathfrak{m}'}$, we have $k_{p+1,\mathfrak{m}'+1} \le k_{p,\mathfrak{m}_1+1} + 1$. Moreover, since $k_{p+1,\mathfrak{m}'+1} \ge k_{p,\mathfrak{m}_1+1} + 1$, we have $k_{p,\mathfrak{m}_1+1} + 1 = k_{p+1,\mathfrak{m}'+1}$. Consequently,

$$k_{p,\mathfrak{m}_1+2} = \inf\{k > k_{p,\mathfrak{m}_1+1}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p\} = k_{p+1,\mathfrak{m}'+1}.$$

This is impossible because $k_{p+1,\mathfrak{m}'+1} < k_{p,\mathfrak{m}_1+2}$, which finally implies that $k_{p,\mathfrak{m}_1+1} = k_{p+1,\mathfrak{m}'+1}$.

We then deduce from this equality,

$$\mathfrak{j}_{\min}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} = \mathfrak{j}_{\min}^{(i,\ell_{k_{p+1,\mathfrak{m}'+1}})} > p+1.$$

Hence, $\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}_1+1}})} \ge b_{p+1}^{(i)}$ and thus $b_{p+1}^{(i)} \in I$. This completes the proof. □

We now return to the proof of Lemma A.2 and prove by induction on p the following result. For all $p \in \{1,\dots,d-1\}$ and all $\mathfrak{m} \in \{0,\dots,\mathfrak{M}_p - 1\}$,

$$\prod_{q=1}^{p} [a_q^{(i)}, b_q^{(i)}] \subset \bigcup_{k \in K_{p,\mathfrak{m}}} \prod_{q=1}^{p} [\theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varrho_q^{(i,\ell_k)}]. \tag{39}$$

Note that (37) follows from this inclusion when $p = d-1$ and $\mathfrak{m} = 0$.

We begin to prove (39) for $p = 1$ and all $\mathfrak{m} \in \{0,\dots,\mathfrak{M}_1 - 1\}$. For all $k \in \{k_{1,\mathfrak{m}} + 1,\dots,k_{1,\mathfrak{m}+1} - 1\}$, $\mathfrak{j}_{\min}^{(i,\ell_k)} \le 1$, and thus

$$\theta_1^{(i,\ell_{k+1})} \in \{\theta_1^{(i,\ell_k)}, \theta_1^{(i,\ell_k)} + \varrho_1^{(i,\ell_k)}\}.$$

This implies that the set

$$\bigcup_{k=k_{1,\mathfrak{m}}+1}^{k_{1,\mathfrak{m}+1}} [\theta_1^{(i,\ell_k)}, \theta_1^{(i,\ell_k)} + \varrho_1^{(i,\ell_k)}]$$

is an interval. Now, $\theta_1^{(i,\ell_{k_{1,\mathfrak{m}}+1})} = a_1^{(i)}$, $\theta_1^{(i,\ell_{k_{1,\mathfrak{m}+1}})} + \varrho_1^{(i,\ell_{k_{1,\mathfrak{m}+1}})} \ge b_1^{(i)}$ since $\mathfrak{j}_{\min}^{(i,\ell_{k_{1,\mathfrak{m}+1}})} > 1$. Therefore,

$$[a_1^{(i)}, b_1^{(i)}] \subset \bigcup_{k=k_{1,\mathfrak{m}}+1}^{k_{1,\mathfrak{m}+1}} [\theta_1^{(i,\ell_k)}, \theta_1^{(i,\ell_k)} + \varrho_1^{(i,\ell_k)}],$$

which establishes (39) when $p = 1$.

Let now $p \in \{1,\dots,d-2\}$ and assume that for all $\mathfrak{m} \in \{0,\dots,\mathfrak{M}_p - 1\}$,

$$\prod_{q=1}^{p} [a_q^{(i)}, b_q^{(i)}] \subset \bigcup_{k \in K_{p,\mathfrak{m}}} \prod_{q=1}^{p} [\theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varrho_q^{(i,\ell_k)}].$$

Let $\mathfrak{m}' \in \{0, \ldots, \mathfrak{M}_{p+1} - 1\}$. We shall show that

$$\prod_{q=1}^{p+1} \left[a_q^{(i)}, b_q^{(i)}\right] \subset \bigcup_{k \in K_{p+1,\mathfrak{m}'}} \prod_{q=1}^{p+1} \left[\theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varrho_q^{(i,\ell_k)}\right].$$

Let $\mathbf{x} \in \prod_{q=1}^{p+1} [a_q^{(i)}, b_q^{(i)}]$. By using Claim A.2, there exists $\mathfrak{m} \in \{0, \ldots, \mathfrak{M}_p - 1\}$ such that

$$x_{p+1} \in \left[\theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}, \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})} + \varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}\right]$$

and such that $k_{p,\mathfrak{m}+1} \in K_{p+1,\mathfrak{m}'}$. By using the induction assumption, there exists $k \in K_{p,\mathfrak{m}}$ such that

$$\mathbf{x} = (x_1, \ldots, x_p) \in \prod_{q=1}^{p} \left[\theta_q^{(i,\ell_k)}, \theta_q^{(i,\ell_k)} + \varrho_q^{(i,\ell_k)}\right].$$

Since $k \in K_{p,\mathfrak{m}}$, $\theta_{p+1}^{(i,\ell_k)} = \theta_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})}$ and $\varrho_{p+1}^{(i,\ell_{k_{p,\mathfrak{m}+1}})} \le \varrho_{p+1}^{(i,\ell_k)}$. Hence,

$$x_{p+1} \in \left[\theta_{p+1}^{(i,\ell_k)}, \theta_{p+1}^{(i,\ell_k)} + \varrho_{p+1}^{(i,\ell_k)}\right].$$

We finally use the claim below to show that $k \in K_{p+1,\mathfrak{m}'}$ which concludes the proof. □

Claim A.3. *Let* $\mathfrak{m} \in \{0, \ldots, \mathfrak{M}_p - 1\}$ *and* $\mathfrak{m}' \in \{0, \ldots, \mathfrak{M}_{p+1} - 1\}$. *If* $k_{p,\mathfrak{m}+1} \in K_{p+1,\mathfrak{m}'}$, *then* $K_{p,\mathfrak{m}} \subset K_{p+1,\mathfrak{m}'}$.

Proof. We have

$$k_{p+1,\mathfrak{m}'} = \sup\left\{k < k_{p+1,\mathfrak{m}'+1}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p+1\right\}.$$

Since $k_{p,\mathfrak{m}+1} > k_{p+1,\mathfrak{m}'}$,

$$\begin{aligned} k_{p+1,\mathfrak{m}'} &= \sup\left\{k < k_{p,\mathfrak{m}+1}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p+1\right\} \\ &\le \sup\left\{k < k_{p,\mathfrak{m}+1}, \mathfrak{j}_{\min}^{(i,\ell_k)} > p\right\} \\ &\le k_{p,\mathfrak{m}}. \end{aligned}$$

We then derive from the inequalities $k_{p+1,\mathfrak{m}'} \le k_{p,\mathfrak{m}}$ and $k_{p,\mathfrak{m}+1} \le k_{p+1,\mathfrak{m}'+1}$ that $K_{p,\mathfrak{m}} \subset K_{p+1,\mathfrak{m}'}$. □

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 numerous comments and his valuable suggestions.

References

[1] Anatolyev, S. and Kosenok, G. (2005). An alternative to maximum likelihood based on spacings. *Econometric Theory* **21** 472–476. MR2179547

[2] Baraud, Y. (2011). Estimator selection with respect to Hellinger-type risks. *Probab. Theory Related Fields* **151** 353–401. MR2834722

[3] Baraud, Y. (2013). Estimation of the density of a determinantal process. *Confluentes Math*. **5** 3–21. MR3143610

[4] Baraud, Y. and Birgé, L. (2009). Estimating the intensity of a random measure by histogram type estimators. *Probab. Theory Related Fields* **143** 239–284. MR2449129

[5] Baraud, Y., Birgé, L. and Sart, M. (2014). A new method for estimation and model selection: ρ-estimation. Available at arXiv:1403.6057.

[6] Barnett, V.D. (1966). Evaluation of the maximum-likehood estimator where the likelihood equation has multiple roots. *Biometrika* **53** 151–165. MR0196838

[7] Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference*: *The Minimum Distance Approach*. *Monographs on Statistics and Applied Probability* **120**. Boca Raton, FL: CRC Press. MR2830561

[8] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann*. *Statist*. **5** 445–463. MR0448700

[9] Birgé, L. (2006). Model selection via testing: An alternative to (penalized) maximum likelihood estimators. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **42** 273–325. MR2219712

[10] Birgé, L. (2007). Model selection for Poisson processes. In *Asymptotics*: *Particles*, *Processes and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **55** 32–64. Beachwood, OH: IMS. MR2459930

[11] Birgé, L. (2013). Robust tests for model selection. In *From Probability to Statistics and Back*: *High-Dimensional Models and Processes*. *Inst*. *Math*. *Stat*. (*IMS*) *Collect*. **9** 47–64. Beachwood, OH: IMS. MR3186748

[12] Birgé, L. (2014). Model selection for density estimation with $\mathbb{L}_2$-loss. *Probab. Theory Related Fields* **158** 533–574. MR3176358

[13] Borovkov, A.A. (1998). *Mathematical Statistics*. Amsterdam: Gordon and Breach Science Publishers. MR1712750

[14] Cheng, R.C.H. and Amin, N.A.K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **45** 394–403. MR0737651

[15] Dacunha-Castelle, D. (1978). Vitesse de convergence pour certains problèmes statistiques. In *École d'Été de Probabilités de Saint-Flour*, *VII* (*Saint-Flour*, 1977). *Lecture Notes in Math*. **678** 1–172. Berlin: Springer. MR0518733

[16] Donoho, D.L. and Liu, R.C. (1988). The "automatic" robustness of minimum distance functionals. *Ann*. *Statist*. **16** 552–586. MR0947562

[17] Ekström, M. (1998). On the consistency of the maximum spacing method. *J*. *Statist*. *Plann*. *Inference* **70** 209–224. MR1649923

[18] Ferguson, T.S. (1982). An inconsistent maximum likelihood estimate. *J*. *Amer*. *Statist*. *Assoc*. **77** 831–834. MR0686406

[19] Ghosh, K. and Jammalamadaka, S.R. (2001). A general estimation method using spacings. *J*. *Statist*. *Plann*. *Inference* **93** 71–82. MR1822389

[20] Ibragimov, I.A. and Has'minskiĭ, R.Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. New York: Springer. MR0620321

[21] Le Cam, L. (1990). Maximum likelihood: An introduction. *Int*. *Stat*. *Rev*. **58** 153–171.

[22] Lindsay, B.G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann*. *Statist*. **22** 1081–1114. MR1292557

[23] Pitman, E.J.G. (1979). *Some Basic Theory for Statistical Inference*: *Monographs on Applied Probability and Statistics*. London: Chapman & Hall. MR0549771

[24] Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. *Scand*. *J*. *Stat*. **11** 93–112. MR0754972

[25] Sart, M. (2013). Estimation par tests. Ph.D. thesis, Univ. Nice Sophia Antipolis.

[26] Sart, M. (2014). Estimation of the transition density of a Markov chain. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **50** 1028–1068. MR3224298

[27] Sart, M. (2015). Model selection for Poisson processes with covariates. *ESAIM Probab*. *Stat*. **19** 204–235.

[28] 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. *Ann*. *Inst*. *Statist*. *Math*. **51** 31–49. MR1704645

Received September 2013 and revised January 2015