ITERATED MODIFIED HISTOGRAMS AS DYNAMICAL SYSTEMS

by

Alain Berlinet [1]
University of Montpellier II,

and

Gérard Biau
University of Paris VI.

Abstract. A modified histogram estimate depending on a reference density has recently been proved to have good consistency properties according to several information theoretic criteria. In the present paper, this modified histogram is studied as a dynamical system in a functional space with the reference density as initial state. Properties of stationary densities of the system are investigated. It is proved, under appropriate assumptions, that the trajectories are almost surely stationary. Indeed, this result opens up the way for optimization of the estimate with respect to the reference density. Results of simulations are presented.

Keywords. Modified histogram estimate, dynamical system, stationary density, quantile.
American Mathematical Society 1991 subject classification. 62 G 05.

[1]Corresponding author: Alain Berlinet, Department of Mathematics, UMR CNRS 5030, CC 051, University Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.
E-mail: berlinet@stat.math.univ-montp2.fr

1 Introduction

Let f and f^* be two densities with respect to the Lebesgue measure λ on the Borel line $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The *information divergence* (also called I-*divergence* or *relative entropy* or *Kullback-Leibler information number*) $D(f, f^*)$ of f with respect to f^* is defined by

$$D(f, f^*) = \begin{cases} \displaystyle\int_{\mathbb{R}} f \ln \frac{f}{f^*} \, \mathrm{d}\lambda & \text{if } f \ll f^* \\ \infty & \text{otherwise.} \end{cases}$$

It is widely used in many fields of statistics (data compression, telecommunication networks, classification, pattern recognition, neural networks, ...), particularly in decision processes based on density estimates and functional of them. Many authors have put forward its attractive properties as a criterion of accuracy which is considerably stronger than the usual L_1 criterion. However, considering convergence of estimates of a density f in the sense of Kullback-Leibler divergence causes some trouble. With standard histograms f_n the situation is even hopeless as empty cells, occuring with high probability, make $D(f, f_n)$ infinite. The modified histograms introduced by Barron [2] circumvent this problem. They are defined as follows.
Suppose that we observe independent real random variables $X_1, \ldots, X_n$ with common unknown density f.

- denote by g a known density on $\mathbb{R}$ and by ν the associated probability measure;
- define a sequence of integers $\{m_n\}_{n \geq 2}$ such that $1 < \ m_n \leq n$ and let $h_n = 1/m_n$;
- introduce a sequence of partitions $\mathcal{P}_n = \{A_{n1}, \ldots, A_{nm_n}\}$, $n \geq 2$, such that the A_{ni}'s are intervals with $\nu(A_{ni}) = h_n$ for $i = 1, \ldots, m_n$;
- finally consider, for $a_n = 1/(nh_n + 1)$, the following density estimate

  $$f_n(x) = \left[(1 - a_n) \frac{\mu_n(A_n(x))}{h_n} + a_n\right] g(x)$$

  i.e.,

  $$f_n(x) = \frac{n\mu_n(A_n(x)) + 1}{nh_n + 1} g(x), \tag{1}$$

where μ_n stands for the empirical measure associated with the sample X_1, $\ldots$, X_n and $A_n(x)$ equals A_{ni} if $x \in A_{ni}$.

Barron, Györfi, and van der Meulen [3] proved that this estimate is consistent both in information divergence and expected information divergence, *i.e.*,

$$\lim_{n \to \infty} \mathbf{E} D(f, f_n) = 0 \ \text{ and } \ \lim_{n \to \infty} D(f, f_n) = 0 \ a.s.$$

provided $D(f,g) < \infty$, $h_n \to 0$ and $nh_n \to \infty$ as n grows to infinity. It is worth noting that the condition $D(f,g) < \infty$ implies that f is absolutely continuous with respect to g. This condition can be easily fulfilled. The choice of g can be guided by some a priori information if available. For instance when the observations are measurement errors a Gaussian density is often used as initial guess g. Roughly speaking the modified histogram results from the comparison of the quantiles of g with the empirical quantiles. A suitable modification of g leads to f_n. For further results on modified histograms, we refer the reader to Berlinet, Györfi, and van der Meulen [8] who prove a central limit theorem for Kullback-Leibler divergence, Györfi, Liese, Vajda, and van der Meulen [14] and Berlinet, Vajda, and van der Meulen [9] who extend the I-divergence consistency properties to the χ^2-divergence and to more general ϕ-divergences. As they are based on histograms, the estimates defined in (1) are in general not continuous. They can estimate consistently very irregular densities. Smoothed versions improve rates of convergence for twice continuously differentiable densities (see Beirlant, Berlinet, Biau, and Vajda [8]).
Once the observations are given two parameters have to be chosen to build the modified histogram, the number of intervals m_n and the function g. Recent results obtained by Berlinet and Brunel [5], [7] show that the Kullback-Leibler cross-validation technique works well to select m_n from the data and that it is asymptotically optimal.
Now, to choose g such that $D(f,g) < \infty$ is easy in a large variety of nonparametric models (see Györfi and van der Meulen [16]). Under mild conditions the choice of g does not affect dramatically the asymptotics. Practically, however, it should reflect some a priori idea we can have on f. The sample being fixed one can wonder whether the resulting estimate is or not a better initial guess. Taking it as new function g do we get after two steps a better estimate? Is there any interest in iterating the process?
Of course there is absolutely no reason to get a better estimate in two steps (even in one step! If you are very well inspired g can be equal to f) but in most cases several steps will produce an improved estimate.
In the present paper, the modified histogram (1) is viewed as a dynamical system in a functional space with initial state the reference density g. In Section 2, some elementary properties of stationary densities of this dynamical system are studied. In Section 3 we establish, under suitable conditions, that the trajectories of the system are almost surely stationary. Results of simulations are presented. In a sake of clarity, proofs are gathered in Section 4.

2 Iteration of the modified histogram

From now on we consider that n and the sample $X_1, \ldots, X_n$ are fixed. As seen in the introduction, the estimation process (1) can be thought of as a system that, given an input density g, supplies an output density $f_n = Bg$ (the dependence of B upon n and upon the sample $X_1, \ldots, X_n$ being understood). Thus, denoting by $\mathcal{D}$ the set of densities on $\mathbb{R}$, we are led to consider the dynamical system $B : \mathcal{D} \to \mathcal{D}$ which associates to each g $(= B^0 g)$ in $\mathcal{D}$ a trajectory $\{B^p g\}_{p \geq 0}$ of densities. We call any element g of $\mathcal{D}$ satisfying the equation $Bg = g$ a *stationary density* of B. A trajectory is said to be stationary if it contains a stationary density. In that case the system no longer evolves after a finite number of iterations.
The issue we have in mind is to select, among the family of estimates $\{B^p g\}_{p \geq 0}$, a particular one, say $B^{p_0} g$, such that

$$p_0 \in \mathrm{Argmin}_{p \geq 0} \{\hat{D}(f, B^p g)\},$$

where $\hat{D}(f, B^p g)$ is some estimate of the unknown Kullback-Leibler divergence $D(f, B^p g)$ (other criteria could be used as well). This is actually a difficult optimization problem that will be briefly discussed in Section 3 but which is beyond the scope of the present work. The first question arising is about the finiteness of the trajectories. *This is the question we are dealing with in the following.* The main result of the paper is that, under mild conditions, any trajectory is stationary. In connection with recent optimality results obtained by Berlinet and Brunel [6], [7], this opens the way to simultaneous optimization of the reference density and the number of bins.
We start with an elementary remark: a density g is stationary for B if and only if $\mu_n(A_{ni}) = h_n$ for $i = 1, \ldots, m_n$. This can only happen if m_n (the number of classes) is a divisor of n (the number of observations), see Figure 1 for a schematic illustration.

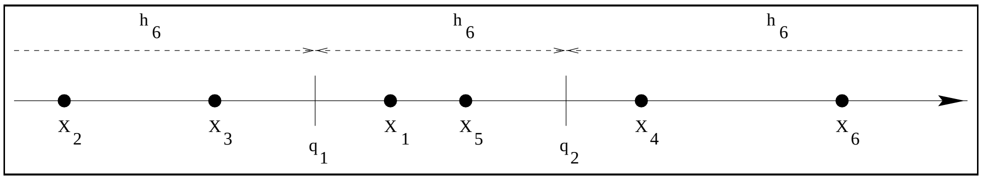

Figure 1. *Schematic illustration of a stationary density g of B ($n = 6$, $m_6 = 3$). The measure ν with density g gives the mass $h_6 = 1/m_6$ to each of the intervals* $]-\infty, q_1]$, $]q_1, q_2]$, $]q_2, \infty[$.

Thus, assuming that n is a multiple of m_n, it is easy to construct many stationary densities of B. The most "simple" are *Random Partition* or *Statistically Equivalent Blocks* (S.E.B.) estimates (Anderson [1], Bosq and

Lecoutre [12], Tukey [18]) defined as follows.
Let $t_n = [n/m_n]$ ([ ] stands for the integer part of a real number) and let $(X_{(1)}, \ldots, X_{(n)})$ be the order statistics vector of the sample $X_1, \ldots, X_n$. Construct a partition of $[X_{(1)}, X_{(n)}]$ into m_n intervals A_{ni} $(i = 1, \ldots, m_n)$ each of them containing exactly the same number of observations up to one (either t_n or $t_n + 1$). The number of observations falling into A_{ni} is denoted $t(A_{ni})$. Then a *Random Partition* or *Statistically Equivalent Blocks* (S.E.B.) estimate b_n is defined by

$$b_n(x) = \begin{cases} \dfrac{t(A_n(x))}{n\lambda(A_n(x))} & \text{if } x \in [X_{(1)}, X_{(n)}] \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 1 *Assume that m_n is a divisor of n. Then the S.E.B. estimate b_n is a stationary density of B.*

Proof. Let us denote by

$$]-\infty, q_1],\ ]q_1, q_2], \ldots,\ ]q_{m_n-1}, \infty[$$

the classes defining the estimate b_n and let ν_n be its associated probability distribution. By construction:

$$\begin{aligned} \nu_n(]-\infty, q_1]) &= \nu_n(]q_1, q_2]) = \ldots = \nu_n(]q_{m_n-1}, \infty[) \\ &= 1/m_n = h_n. \end{aligned}$$

It follows that $q_1, \ldots, q_{m_n-1}$ are the quantiles of order ih_n $(i = 1, \ldots, m_n - 1)$ of b_n. To prove that b_n is stationary, it is sufficient to show that $\mu_n(]-\infty, q_1]) = \mu_n(]q_1, q_2]) = \ldots = \mu_n(]q_{m_n-1}, \infty[) = h_n$. This is true, by the very construction of the S.E.B. estimate.

3 Asymptotic results

In this section, we give results about the asymptotics of the trajectory, that is the behavior of $B^p g$, as p tends to infinity. To avoid useless technicalities, we assume that the distribution function of the initial density g is invertible. We also assume that m_n is a divisor of n. In practice, we may restrict ourselves to $m_n[n/m_n]$ data. This does not harm too much since m_n is very small with respect to n. More precisely, the asymptotic theory reveals that m_n should be of order $n^{1/3}$ (Barron, Györfi, and van der Meulen [3]). Also, a slight modification of the estimate could allow to remove the divisibility condition. We denote, for each p in $\mathbb{N}$:

- $q_1^p, q_2^p, \ldots, q_{m_n-1}^p$ the quantiles of order ih_n $(i = 1, \ldots, m_n - 1)$ of the density $B^p g$ associated with the partition

$$\{A_{n1}^p, \ldots, A_{nm_n}^p\} \stackrel{\text{def}}{=} \{]-\infty, q_1^p],]q_1^p, q_2^p], \ldots,]q_{m_n-1}^p, \infty[\};$$

- $\alpha_1^p, \ldots, \alpha_{m_n}^p$ the (*Barron*) *coefficients* associated with the partition above, namely:

$$\alpha_i^p = \frac{n\mu_n(A_{ni}^p) + 1}{nh_n + 1} \ , \quad i = 1, \ldots, m_n.$$

With these notations, the dynamical system B reads

$$B^{p+1}g = \sum_{i=0}^{m_n - 1} \alpha_{i+1}^p \mathbf{1}_{]q_i^p, q_{i+1}^p]} B^p g,$$

where $\mathbf{1}_A$ is the characteristic function of a set A and where we used the convention $q_0^p = -\infty$ and $q_{m_n}^p = \infty$. Thus, the study of the sequence $\{B^p g\}_{p \geq 0}$ reduces to the study of the two sequences $\{(q_1^p, \ldots, q_{m_n - 1}^p)\}_{p \geq 0}$ and $\{(\alpha_1^p, \ldots, \alpha_{m_n}^p)\}_{p \geq 0}$. Whenever $m_n \geq 3$, we need the following assumption.

Assumption 1 (HLi, $\mathrm{i} = 2, \ldots, \mathrm{m_n} - 1$) *For p large enough,*

$$q_i^p > X_{\left(\frac{in}{m_n} - 1\right)}.$$

Assumption (HLi) means that, for p large enough, each value of q_i^p falling to the left of $X_{\left(\frac{in}{m_n}\right)}$ is actually "close" to this observation. The following theorem, whose proof is deferred to Section 4, is essential (see Figure 2 for an illustration):

Theorem 1 *Assume that m_n divides n and that the condition*

$$(\mathrm{H}) = (\mathrm{HL2}) + \ldots + (\mathrm{HLm_n} - 1)$$

is satisfied whenever $m_n \geq 3$. Then each sequence $\{q_i^p\}_{p \geq 0}$ $(i = 1, \ldots, m_n - 1)$ is almost surely stationary: in a finite number of iterations each sequence reaches almost surely a limit point belonging to the empirical quantile interval of order ih_n, namely $[X_{\left(\frac{in}{m_n}\right)}, X_{\left(\frac{in}{m_n} + 1\right)}[$.

It is clear that Condition (H) is necessary for the convergence of the trajectory $\{B^p g\}_{p \geq 0}$. Theorem 1 states its sufficiency. We never encountered in our simulations a situation in which convergence does not occur so that we conjecture that Condition (H) is satisfied for a large class $\mathcal{G}$ of functions g whatever the sample can be. Conditions guaranteeing that some function g belongs to $\mathcal{G}$ are currently under study.
We developed an algorithm automatically generating the sequence of quantiles (see Figure 2 and Biau [10] for details). The code was implemented with the mathematical macro-language *Matlab* and is available on request. Let us now state a corollary which is an easy consequence of Theorem 1.

Corollary 1 *Under the same conditions as in Theorem 1:*

- *Each sequence of coefficients* $\{\alpha_i^p\}_{p\geq 0}$ *(*$i = 1, \ldots, m_n$*) reaches almost surely the value 1 in a finite number of iterations.*
- *The sequence of densities* $\{B^p g\}_{p\geq 0}$ *is almost surely stationary.*

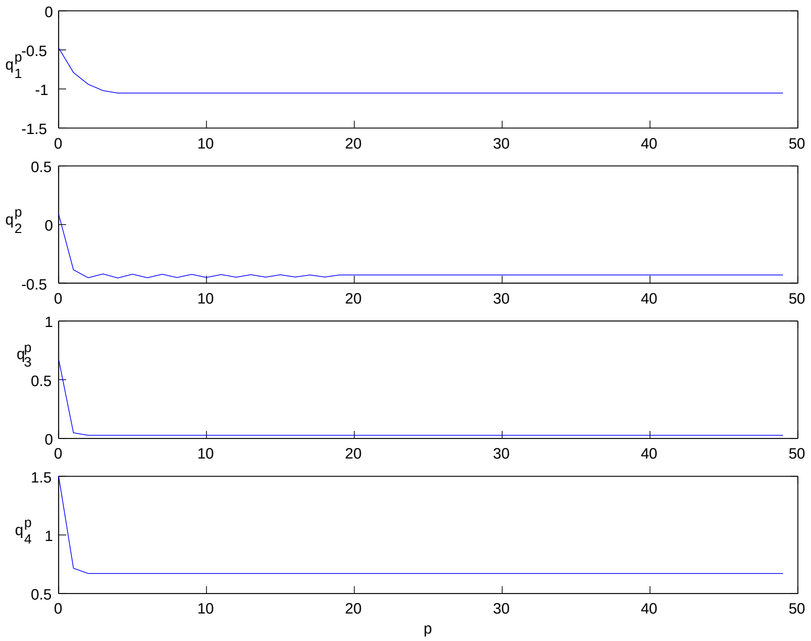

Figure 2. *Evolution of quantiles obtained with an initial density of Gumbel type, the density to be estimated being Gaussian* $\mathcal{N}(0,1)$*.* $n = 100$, $m_n = 5$, $n/m_n = 20$.

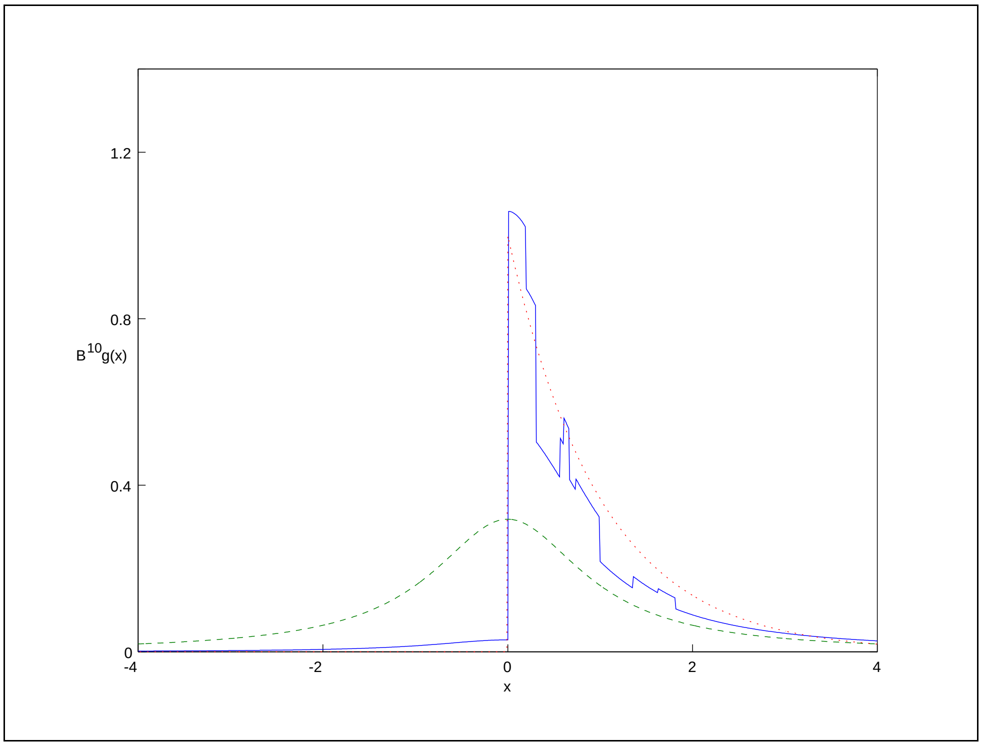

Figure 3. CONTINUOUS LINE: *limit stationary density obtained with an initial density of Cauchy type, the density to be estimated being exponential with parameter* 1. DASHED LINE: *initial Cauchy density.* DOTTED LINE: *exponential density to be estimated.* $n = 60$, $m_n = 4$, $n/m_n = 15$.

The last assertion of Corollary 1 is illustrated in Figure 3.

Let us come back to the optimization problem set at the beginning of Section 2. The optimization of $\hat{D}(f, B^p g)$ is out of the scope of the present paper but we give two examples of the behavior of the true values of $D(f, B^p g)$ as p varies. We select an estimate $B^{p_0} g$ among the family $\{B^p g\}_{p \geq 0}$ satisfying

$$p_0 \in \operatorname{Argmin}_{p \geq 0} \{D(f, B^p g)\} .$$

Figure 4 reports the values of $D(f, B^p g)$ when f is exponential with parameter 1, g is Cauchy (TOP ($n = 100$, $m_n = 4$)), f is Gaussian $\mathcal{N}(8, 1)$ and g is Gumbel (BOTTOM ($n = 100$, $m_n = 5$)). In the first case, we select the eighth estimate out of 42, whereas in the second, we select the third out of 4. Note that in the second example the first two iterates are worse than g. In practice, the values $D(f, B^p g)$ are unknown and have to be estimated (Györfi and van der Meulen [15]). Recently, Berlinet and Brunel [5], [7], proposed an automatic procedure for selecting the asymptotically optimal bandwidth h_n. The results of these authors could be coupled with our iterative approach to design an optimal "ready for use" estimate. As suggested by a referee, we also compared in terms of Kullback-Leiber divergence and L_1 error the performance of the selected estimate $B^{p_0}g$ with the standard kernel estimate of f defined with a Gaussian $\mathcal{N}(0,1)$ kernel K and a smoothing parameter $h > 0$ as

$$f_{n,h} = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big)\,.$$

The density to be estimated is a mixture with equal weights of a uniform density on $]0,1[$ and of a Gaussian $\mathcal{N}(1,1)$ density.
For the iterated modified histogram, the setup is as follows. We start with a Gaussian $\mathcal{N}(0,1)$ and we select an estimate B^{p_0} such that

$$p_0 \in \mathrm{Argmin}_{p\geq 0}\{D(f, B^p g)\} \quad \text{for the Kullback-Leibler error criterion}$$

and

$$p_0 \in \mathrm{Argmin}_{p\geq 0}\Big\{\int |f - B^p g|\,\mathrm{d}\lambda\Big\} \quad \text{for the } L_1 \text{ error criterion.}$$

We fix $n = 100$ and optimize at the first step m_n among the divisors of 100. As for the kernel estimate, we decided to work with the best bandwidth h^* pertaining to each criteria, that is

$$h^* = \mathrm{Argmin}_{h>0}\{D(f, f_{n,h})\} \quad \text{for the Kullback-Leibler criterion}$$

and

$$h^* = \mathrm{Argmin}_{h>0}\Big\{\int |f - f_{n,h}|\,\mathrm{d}\lambda\Big\} \quad \text{for the } L_1 \text{ criterion.}$$

The boxplots of our simulation results (100 repetitions) are given in Figure 5 and 6 and typical realizations in Figure 7 and 8. The iterated modified histograms show some ability to estimate non smooth densities.

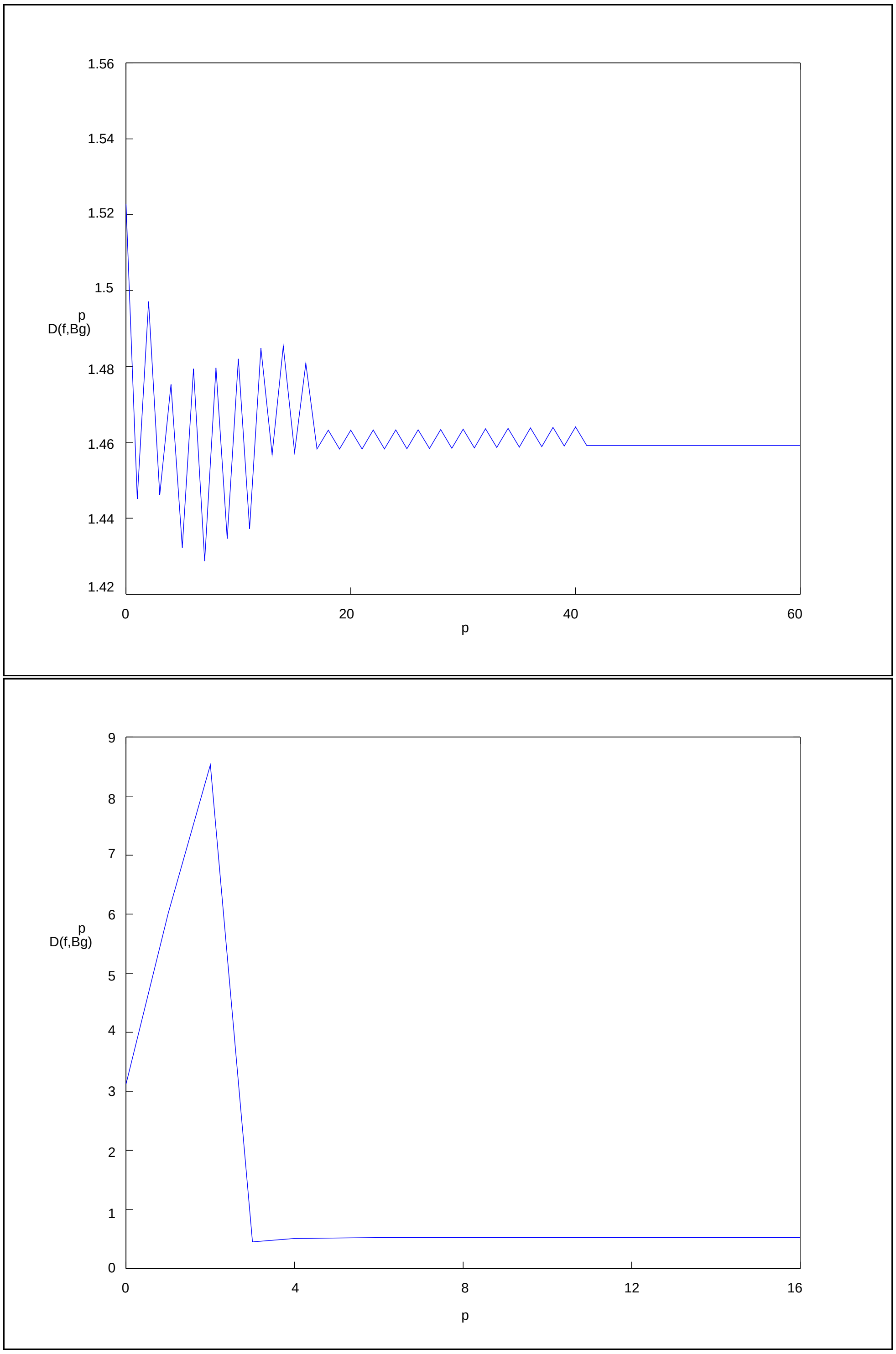

Figure 4. (See text).

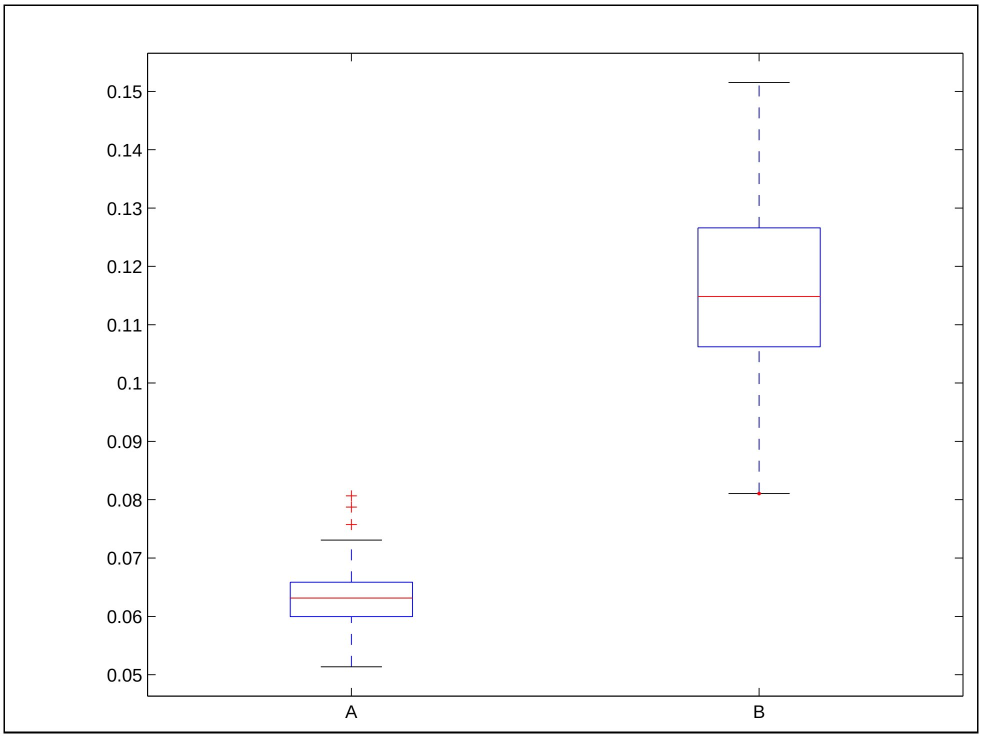

Figure 5. Kullback-Leibler divergence error for the iterated modified histogram (A) and the kernel estimate (B).

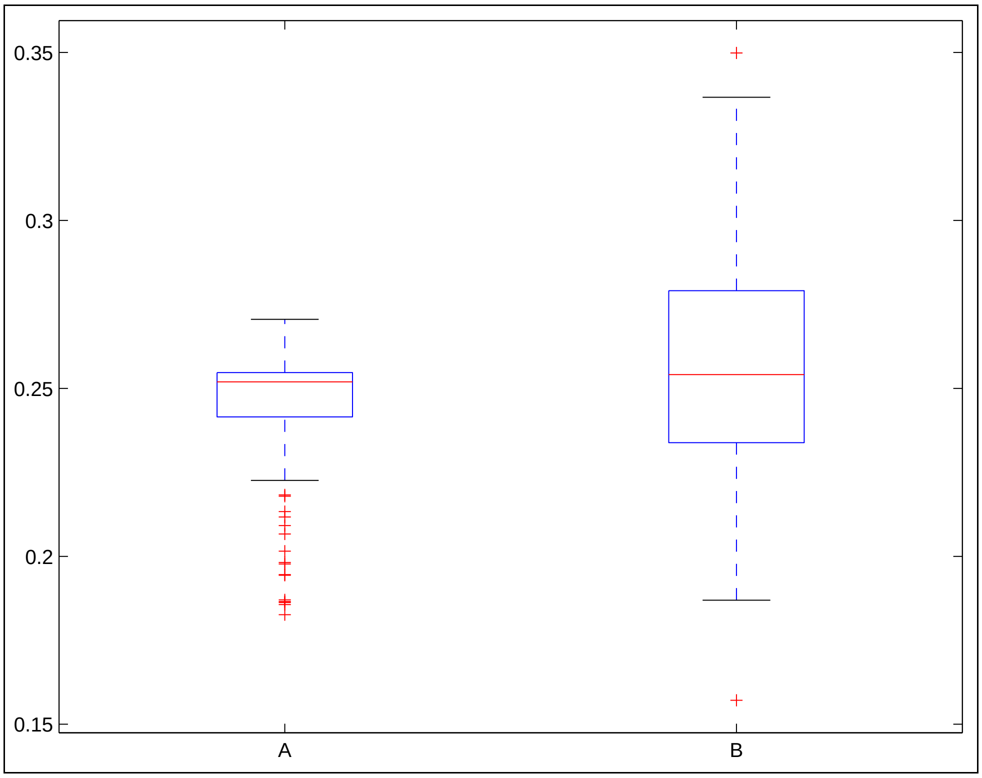

Figure 6. L_1 error for the iterated modified histogram (A) and the kernel estimate (B).

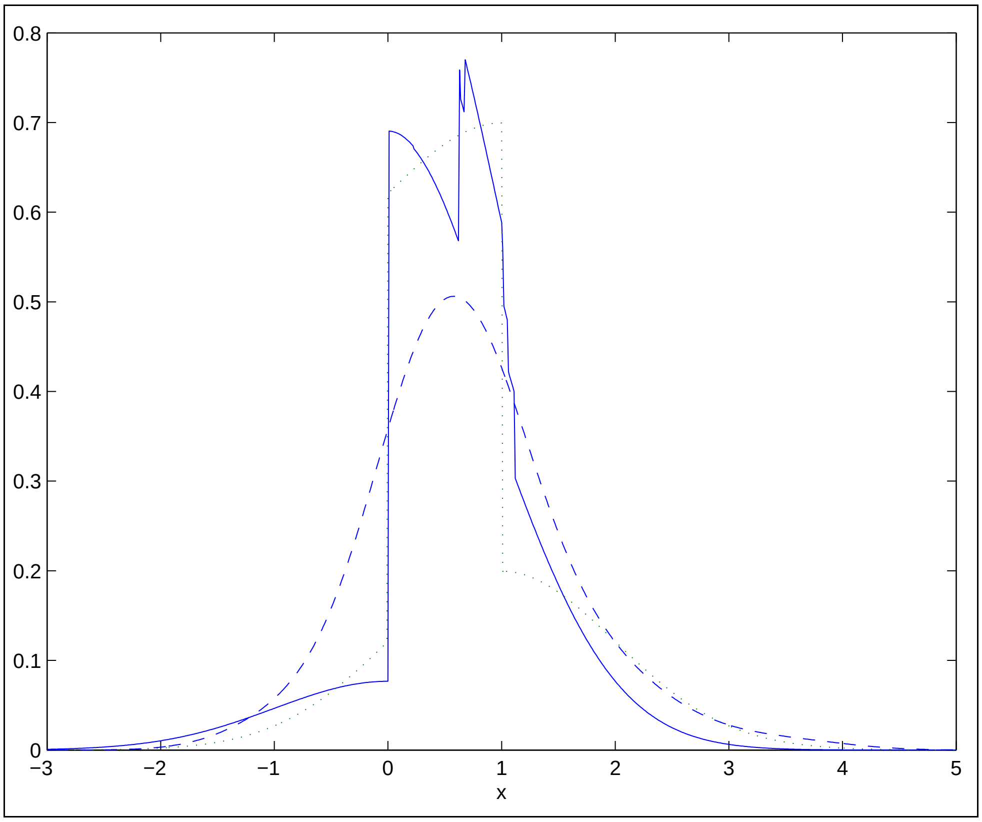

Figure 7. Continuous line: *iterated modified histogram.* Dashed line: *kernel estimate.* Dotted line: *true density to be estimated.* Criterion: Kullback-Leibler error.

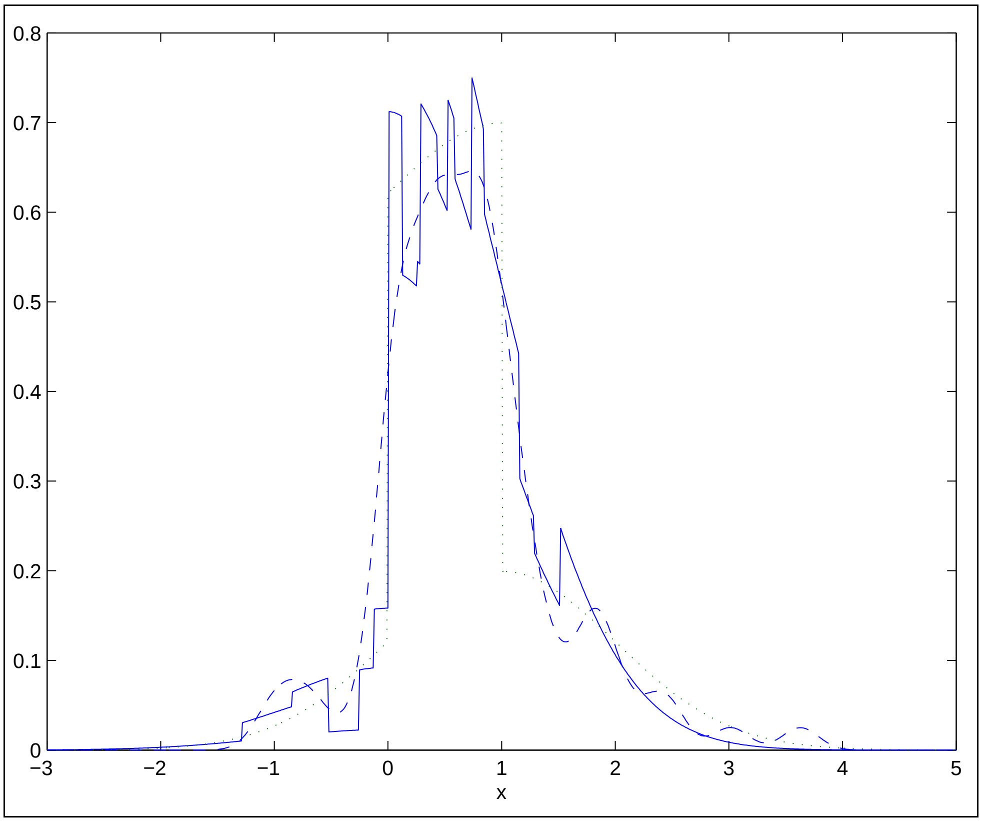

Figure 8. CONTINUOUS LINE: *iterated modified histogram.* DASHED LINE: *kernel estimate.* DOTTED LINE: *true density to be estimated.* CRITERION: L_1 error.

4 Proof of Theorem 1

We start the proof with two general lemmas that will be useful in the sequel. In a sake of clarity, we leave out the notation "$p \geq 0$" in the sequences when no confusion is possible.

Lemma 1 *For every i in $\{1, \ldots, m_n - 1\}$ and every $p \geq 0$:*

- *If $\mu_n(]-\infty, q_i^p]) > ih_n$, then $q_i^{p+1} < q_i^p$.*
- *If $\mu_n(]-\infty, q_i^p]) = ih_n$, then $q_i^{p+1} = q_i^p$.*
- *If $\mu_n(]-\infty, q_i^p]) < ih_n$, then $q_i^{p+1} > q_i^p$.*

Proof. Denote by $\mathbf{1}_A$ the characteristic function of a set A and observe that for every i in $\{1, \ldots, m_n - 1\}$ and every $p \geq 0$,

$$\int_{-\infty}^{q_i^p} B^{p+1} g \, \mathrm{d}\lambda = \int_{-\infty}^{q_i^p} \left[\sum_{j=0}^{i-1} \alpha_{j+1}^p \mathbf{1}_{]q_j^p, q_{j+1}^p]} \, B^p g \right] \mathrm{d}\lambda$$
$$= \sum_{j=0}^{i-1} \frac{n\mu_n(]q_j^p, q_{j+1}^p]) + 1}{nh_n + 1} \int_{q_j^p}^{q_{j+1}^p} B^p g \, \mathrm{d}\lambda,$$

with the convention $q_0^p = -\infty$. Since

$$\int_{q_j^p}^{q_{j+1}^p} B^p g \, \mathrm{d}\lambda = h_n,$$

this leads to

$$\begin{aligned} \int_{-\infty}^{q_i^p} B^{p+1} g \, \mathrm{d}\lambda &= h_n \sum_{j=0}^{i-1} \frac{n\mu_n(]q_j^p, q_{j+1}^p]) + 1}{nh_n + 1} \\ &= h_n \, \frac{n\mu_n(]-\infty, q_i^p]) + i}{nh_n + 1}. \end{aligned}$$

The result is then a straightforward consequence of the following equivalences:

- $q_i^{p+1} < q_i^p \iff \int_{-\infty}^{q_i^p} B^{p+1} g \, \mathrm{d}\lambda > ih_n.$
- $q_i^{p+1} = q_i^p \iff \int_{-\infty}^{q_i^p} B^{p+1} g \, \mathrm{d}\lambda = ih_n.$
- $q_i^{p+1} > q_i^p \iff \int_{-\infty}^{q_i^p} B^{p+1} g \, \mathrm{d}\lambda < ih_n.$

Lemma 2 *Assume that m_n divides n. Then, for every i in $\{1, \ldots, m_n - 1\}$ and every $p \geq 0$:*

- $\mu_n(]-\infty, q_i^p]) > ih_n \iff q_i^p \geq X_{(\frac{in}{m_n}+1)}.$
- $\mu_n(]-\infty, q_i^p]) = ih_n \iff q_i^p \in [X_{(\frac{in}{m_n})}, X_{(\frac{in}{m_n}+1)}[.$
- $\mu_n(]-\infty, q_i^p]) < ih_n \iff q_i^p < X_{(\frac{in}{m_n})}.$

Proof. Clear from the definition of the empirical measure.

Let us now examine the possible behaviors of the sequences of quantiles $\{q_i^p\}$, $i = 1, \ldots, m_n - 1$. Throughout, we will denote by $\{G_p\}$ the sequence of distribution functions associated with the sequence of densities $\{B^p g\}$. We know by Helly's selection theorem (Billingsley [11]) that there exists a subsequence $\{G_{\varphi(p)}\}$ of $\{G_p\}$ and a distribution function (possibly defective) G_∞ such that

$$\lim_{p\to\infty} G_{\varphi(p)}(x) = G_\infty(x) \tag{2}$$

at any continuity point x of G_∞.

Remark 1 *Using Helly's selection theorem, one can show that the sequences* $\{q_i^p\}$ *(*$i = 1, \ldots, m_n - 1$*) are lower bounded.*

Let us first study the possible dynamics of the first sequence $\{q_1^p\}$. We claim that four dynamics have to be considered and we take the liberty, for clarity sake, to point out these dynamics with the abbreviations $\mathbf{R(1)}$ ($\mathbf{R}$ means "Right", $\mathbf{1}$ stands for the first quantile), $\mathbf{I(1)}$ ($\mathbf{I}$ means "Inside"), $\mathbf{L(1)}$ ($\mathbf{L}$ means "Left") and $\mathbf{O(1)}$ ($\mathbf{O}$ means "Oscillation").

- First dynamics: $\mathbf{R(1)}$. Description: for large enough p, say $p \geq p_0$, the sequence $\{q_1^p\}$ remains in the interval $[X_{(\frac{n}{m_n}+1)}, \infty[$.

  Assume that $\mathbf{R(1)}$ holds and pick any x_0 in the interval $]X_{(\frac{n}{m_n})}, X_{(\frac{n}{m_n}+1)}[$ such that x_0 is a continuity point of G_∞. Observe then that for $p \geq p_0$, x_0 belongs to the interval $]-\infty, q_1^p[$ and deduce that for $p > p_0$

  $$G_p(x_0) \geq \delta(p_0) \left[\prod_{j=p_0}^{p-1} \alpha_1^j\right] \times \int_{X_{(\frac{n}{m_n})}}^{x_0} g \, \mathrm{d}\lambda, \tag{3}$$

  where $\delta(p_0)$ is a positive real number. Moreover, from Lemma 2, for $p \geq p_0$,

  $$\alpha_1^p \geq \frac{nh_n + 2}{nh_n + 1}.$$

  This leads, thanks to equations (2) and (3), to

  $$\int_{X_{(\frac{n}{m_n})}}^{x_0} g \, \mathrm{d}\lambda = 0,$$

  i.e.,

  $$G(x_0) = G(X_{(\frac{n}{m_n})}).$$

  We conclude that the dynamics $\mathbf{R(1)}$ is *a.s.* not possible.

- Second dynamics: $\mathbf{I(1)}$. Description: there exists p_0 in $\mathbb{N}$ such that $q_1^{p_0}$ belongs to the interval $[X_{(\frac{n}{m_n})}, X_{(\frac{n}{m_n}+1)}[$.

  In that case, it is deduced from Lemma 1 and Lemma 2 that $q_1^p = q_1^{p_0}$ and $\alpha_1^p = 1$ for all $p \geq p_0$: the quantile is stationary and the first coefficient remains equal to 1.

- THIRD DYNAMICS: $\mathbf{L(1)}$. DESCRIPTION: for large enough p, the sequence $\{q_1^p\}$ remains in the interval $]-\infty, X_{(\frac{n}{m_n})}[$.

  Note that when $\mathbf{L(1)}$ holds, the sequence $\{q_1^p\}$ becomes increasing (Lemma 1 and Lemma 2) and therefore is convergent.

- FOURTH DYNAMICS: $\mathbf{O(1)}$. DESCRIPTION: the sequence $\{q_1^p\}$ takes infinitely many values in both intervals $]-\infty, X_{(\frac{n}{m_n})}[$ and $[X_{(\frac{n}{m_n}+1)}, \infty[$.

  In other words, the quantile q_1^p is oscillating around the interval $[X_{(\frac{n}{m_n})}, X_{(\frac{n}{m_n}+1)}[$.

Let us now turn to the analysis of the dynamics of the second sequence $\{q_2^p\}$. Again, four behaviors are *a priori* allowed and we denote them $\mathbf{R(2)}$, $\mathbf{I(2)}$, $\mathbf{L(2)}$ and $\mathbf{O(2)}$.

- FIRST DYNAMICS: $\mathbf{R(2)}$. DESCRIPTION: for large enough p, the sequence $\{q_2^p\}$ remains in the interval $[X_{(\frac{2n}{m_n}+1)}, \infty[$.

  Let us show that the dynamics $\mathbf{R(2)}$ is *a.s.* not possible. The key argument is that $\mathbf{R(2)}$ is *a.s.* incompatible with any of the dynamics $\mathbf{I(1)}$, $\mathbf{L(1)}$ or $\mathbf{O(1)}$. To see that $\mathbf{R(2)}$ is *a.s.* incompatible with $\mathbf{I(1)}$, assume that $\mathbf{I(1)}$ and $\mathbf{R(2)}$ simultaneously hold, and choose any x_0 in the interval $]X_{(\frac{n}{m_n}+1)}, X_{(\frac{2n}{m_n}+1)}[$ such that

  $$\lim_{p\to\infty} G_{\varphi(p)}(x_0) = G_\infty(x_0). \tag{4}$$

  It is clear that for large enough p, say $p \geq p_0$, $]X_{(\frac{n}{m_n}+1)}, X_{(\frac{2n}{m_n}+1)}[\subset]q_1^p, q_2^p[$ and that x_0 belongs thus to the interval $]q_1^p, q_2^p[$. Consequently, for $p > p_0$,

  $$G_p(x_0) \geq \delta(p_0)\left[\prod_{j=p_0}^{p-1} \alpha_2^j\right] \times \int_{X_{(\frac{n}{m_n}+1)}}^{x_0} g \,\mathrm{d}\lambda, \tag{5}$$

  where $\delta(p_0)$ is a positive real number. Observing that for $p \geq p_0$

  $$\alpha_2^p = \frac{n\mu_n(]q_1^p, q_2^p]) + 1}{nh_n + 1} \geq \frac{nh_n + 2}{nh_n + 1},$$

  leads, in accordance with equations (4) and (5), to

  $$\int_{X_{(\frac{n}{m_n}+1)}}^{x_0} g \,\mathrm{d}\lambda = 0,$$

  i.e.,

  $$G(x_0) = G(X_{(\frac{n}{m_n}+1)}).$$

  We concude that the dynamics $\mathbf{R(2)}$ is *a.s.* incompatible with the dynamics $\mathbf{I(1)}$. Now, concisely:

– To see that $\mathbf{R(2)}$ is *a.s.* incompatible with $\mathbf{L(1)}$, assume that $\mathbf{L(1)}$ and $\mathbf{R(2)}$ simultaneously hold, pick any x_0 in $]X_{(\frac{n}{m_n})}, X_{(\frac{2n}{m_n}+1)}[$ such that equation (4) holds and observe that x_0 belongs for large enough p to the interval $]q_1^p, q_2^p[$ with

$$\alpha_2^p \geq \frac{nh_n + 3}{nh_n + 1}.$$

This leads to

$$\int_{X_{(\frac{n}{m_n})}}^{x_0} g \, \mathrm{d}\lambda = 0.$$

– To show that $\mathbf{R(2)}$ is *a.s.* incompatible with $\mathbf{O(1)}$, assume that both $\mathbf{O(1)}$ and $\mathbf{R(2)}$ hold, choose x_0 in $]X_{(\frac{n}{m_n})}, X_{(\frac{n}{m_n}+1)}[$ such that equation (4) holds and denote by p_0 an integer such that q_2^p belongs to $[X_{(\frac{2n}{m_n}+1)}, \infty[$ for $p \geq p_0$. Then, for $p > p_0$,

$$G_p(x_0) \geq \delta(p_0) \left[\prod_{G(p)} \alpha_2^j \prod_{D(p)} \alpha_1^j \right] \times \int_{X_{(\frac{n}{m_n})}}^{x_0} g \, \mathrm{d}\lambda,$$

where $\delta(p_0)$ is a positive real number and where

$$G(p) = \left\{ j \in \{p_0, \ldots, p-1\} : q_1^j < X_{(\frac{n}{m_n})} \right\}$$

and

$$D(p) = \left\{ j \in \{p_0, \ldots, p-1\} : q_1^j \geq X_{(\frac{n}{m_n}+1)} \right\}.$$

But

$$\alpha_1^j \geq \frac{nh_n + 2}{nh_n + 1} \quad \text{whenever } j \in D(p),$$

and

$$\alpha_2^j \geq \frac{nh_n + 3}{nh_n + 1} \quad \text{whenever } j \in G(p).$$

This implies

$$\int_{X_{(\frac{n}{m_n})}}^{x_0} g \, \mathrm{d}\lambda = 0.$$

- Second dynamics: $\mathbf{I(2)}$. Description: for $p = p_0$, the quantile q_2^p enters the interval $[X_{(\frac{2n}{m_n})}, X_{(\frac{2n}{m_n}+1)}[$.

  $\mathbf{I(2)}$ implies that the second quantile becomes stationary (Lemma 1 and Lemma 2). Arguments similar to those developed for $\mathbf{R(2)}$ show that this dynamics is *a.s.* incompatible with $\mathbf{L(1)}$ and $\mathbf{O(1)}$.

- Third dynamics: $\mathbf{L(2)}$. Description: for large enough p, $q_2^p < X_{(\frac{2n}{m_n})}$.

  According to Lemma 1 and Lemma 2, the sequence $\{q_2^p\}$ becomes increasing and therefore is convergent.

- Fourth dynamics: $\mathbf{O(2)}$. Description: the second quantile q_2^p is oscillating around the interval $[X_{(\frac{2n}{m_n})}, X_{(\frac{2n}{m_n}+1)}[$.

  Under the hypothesis (HL2):

  - $\mathbf{O(2)}$ is *a.s.* incompatible with $\mathbf{L(1)}$ because the dynamics $\mathbf{L(1)}$–$\mathbf{O(2)}$ leads to

  $$\alpha_2^p \geq \begin{cases} 1 & \text{if} \quad q_2^p \in \,]X_{(\frac{2n}{m_n}-1)}, X_{(\frac{2n}{m_n})}[\\ \dfrac{nh_n+3}{nh_n+1} & \text{if} \quad q_2^p \in [X_{(\frac{2n}{m_n}+1)}, \infty[\,, \end{cases}$$

    for large enough p. Pick any x_0 in $]X_{(\frac{n}{m_n})}, X_{(\frac{2n}{m_n}-1)}[$ such that equation (4) holds to obtain

  $$\int_{X_{(\frac{n}{m_n})}}^{x_0} g \,\mathrm{d}\lambda = 0.$$

  - $\mathbf{O(2)}$ is *a.s.* incompatible with $\mathbf{O(1)}$ because, for large enough p, the dynamics $\mathbf{O(1)}$–$\mathbf{O(2)}$ leads to

  $$\alpha_1^p \geq \frac{nh_n+2}{nh_n+1} \quad \text{whenever } q_1^p \geq X_{(\frac{n}{m_n}+1)},$$

    and

  $$\alpha_2^p \geq \begin{cases} 1 & \text{if} \quad q_2^p \in \,]X_{(\frac{2n}{m_n}-1)}, X_{(\frac{2n}{m_n})}[\\ \dfrac{nh_n+3}{nh_n+1} & \text{if} \quad q_2^p \in [X_{(\frac{2n}{m_n}+1)}, \infty[\,, \end{cases}$$

  whenever $q_1^p < X_{(\frac{n}{m_n})}$. Choose now x_0 in $]X_{(\frac{n}{m_n})}, X_{(\frac{n}{m_n}+1)}[$ such that equation (4) holds to get

  $$\int_{X_{(\frac{n}{m_n})}}^{x_0} g \,\mathrm{d}\lambda = 0.$$

To fully understand the behavior of the quantile sequences, it is still necessary to analyze the possible dynamics $\mathbf{R(3)}$, $\mathbf{I(3)}$, $\mathbf{L(3)}$ and $\mathbf{O(3)}$ of the third sequence $\{q_3^p\}$. Classical arguments show that:

- The dynamics $\mathbf{R(3)}$ is *a.s.* incompatible with $\mathbf{I(2)}$, $\mathbf{L(2)}$ and, under (HL2), with $\mathbf{O(2)}$. Thus, under (HL2), the dynamics $\mathbf{R(3)}$ is *a.s.* not allowed.

- The dynamics $\mathbf{I(3)}$ is *a.s.* incompatible with $\mathbf{L(2)}$ and, under (HL2), with $\mathbf{O(2)}$.
- If we assume that $q_3^p > X_{(\frac{3n}{m_n}-1)}$ for large enough p (hypothesis (HL3)), then the dynamics $\mathbf{O(3)}$ is *a.s.* incompatible with $\mathbf{L(2)}$ and, under (HL2), with $\mathbf{O(2)}$.

Repeating the arguments that were presented for the cases $i = 1$, $i = 2$ and $i = 3$, we obtain, under the general hypothesis (H)=(HL2)+...+(HL$m_n - 1$), that

(i) For every i in $\{1, \ldots, m_n - 1\}$, the dynamics $\mathbf{R(i)}$ is *a.s.* not possible.

(ii) If $m_n \geq 3$, for every i in $\{1, \ldots, m_n - 2\}$, the dynamics $\mathbf{I(i+1)}$ is *a.s.* incompatible with the dynamics $\mathbf{L(i)}$ and $\mathbf{O(i)}$.

(iii) If $m_n \geq 3$, for every i in $\{1, \ldots, m_n - 2\}$, the dynamics $\mathbf{O(i+1)}$ is *a.s.* incompatible with the dynamics $\mathbf{L(i)}$ and $\mathbf{O(i)}$.

Finally, let us make three remarks:

(iv) We know that the dynamics $\mathbf{R(1)}$ is *a.s.* not allowed. Considering the densities $\tilde{f}$ and $\tilde{g}$ defined by $\tilde{f}(x) = f(-x)$ and $\tilde{g}(x) = g(-x)$, a symmetry argument shows that the dynamics $\mathbf{L(m_n - 1)}$ is *a.s.* not allowed.

(v) If $m_n = 2$, the dynamics $\mathbf{O(1)}$ is *a.s.* not allowed. To see that, choose any x_0 in the interval $]X_{(\frac{n}{2})}, X_{(\frac{n}{2}+1)}[$ such that equation (4) holds and observe that

$$\alpha_1^p \geq \frac{nh_n + 2}{nh_n + 1} \quad \text{whenever } q_1^p \geq X_{(\frac{n}{2}+1)},$$

and

$$\alpha_2^p \geq \frac{nh_n + 2}{nh_n + 1} \quad \text{whenever } q_1^p < X_{(\frac{n}{2})}.$$

This leads to

$$\int_{X_{(\frac{n}{m_n})}}^{x_0} g \, \mathrm{d}\lambda = 0.$$

(vi) Assume that $m_n \geq 3$. We know that the dynamics $\mathbf{O(1)}$–$\mathbf{I(2)}$ is *a.s.* not allowed. A symmetry argument shows that the dynamics $\mathbf{I(m_n - 2)}$–$\mathbf{O(m_n - 1)}$ is also *a.s.* not allowed.

We can now conclude:

- If $m_n = 2$, the proof of Theorem 1 is clear from points (i), (iv) and (v).

- If $m_n \geq 3$, we conclude from (i), (ii), (iii) and (iv) that the dynamics $\mathbf{I(1)}, \ldots, \mathbf{I(m_n - 2)}$ is *a.s.* the only one possible for the sequences $\{q_i^p\}$, $i = 1, \ldots, m_n - 2$. From points (iv) and (vi), we finally deduce that the dynamics $\mathbf{I(m_n - 1)}$ is *a.s.* the only one possible for the last sequence $\{q_{m_n-1}^p\}$.

This achieves the proof of Theorem 1.

Lemma 3 *Let i in $\{1, \ldots, m_n - 1\}$, let $p \geq 0$, and assume that m_n is not a divisor of in. Then*

- $\mu_n(]-\infty, q_i^p]) > ih_n \iff q_i^p \geq X_{([\frac{in}{m_n}]+1)}$.
- $\mu_n(]-\infty, q_i^p]) < ih_n \iff q_i^p < X_{([\frac{in}{m_n}]+1)}$.

Figure 5 illustrates the dynamics of the sequences of quantiles when $n = 10$ and $m_n = 4$. Arguments similar to those given in the case $n/m_n \in \mathbb{N}$ show that the dynamics $\mathbf{O(1)}$–$\mathbf{I(2)}$–$\mathbf{O(3)}$ is *a.s.* the only one possible under the hypothesis (H) and the additional assumption that, for large enough p, the values of the sequence $\{q_1^p\}$ falling to the right of $X_{(3)}$ are actually "close" to this observation (assumption denoted (HR1)). The consistency of the sequences $\{q_1^p\}$ and $\{q_3^p\}$ towards the observations $X_{(3)}$ and $X_{(8)}$ seems not so easy to prove (it was however observed in many simulations, see Biau [10]). On the other hand, it is seen that

$$\lim_{p\to\infty} B^p g(X_{(3)}) = \infty \quad \text{and} \quad \lim_{p\to\infty} B^p g(X_{(8)}) = \infty,$$

i.e., the iterated densities will present "peaks" at $X_{(3)}$ and $X_{(8)}$. Note also that the second quantile is *a.s.* stationary for large enough p and, under the additional assumptions (HL1) and (HR3) (with obvious notations), that the three coefficients take *a.s.* their values in the set $\{3/3.5; 4/3.5\}$. It is tempting to conjecture that the following dynamics is *a.s.* the only one possible:

$$\mathbf{T(1)}, \ldots, \mathbf{T(m_n - 1)}$$

with

$$\begin{cases} \mathbf{T(i)} = \mathbf{I(i)} & \text{if} \quad in/m_n \in \mathbb{N} \\ \mathbf{T(i)} = \mathbf{O(i)} & \text{if} \quad in/m_n \notin \mathbb{N}. \end{cases}$$

Unfortunately, it is to be feared that more intricate situations may occur. For example, when $n = 10$ and $m_n = 3$, the dynamics $\mathbf{O(1)}$–$\mathbf{R(2)}$ is theoretically *a priori* allowed (see Figure 6). No doubt that there is here room for future research.

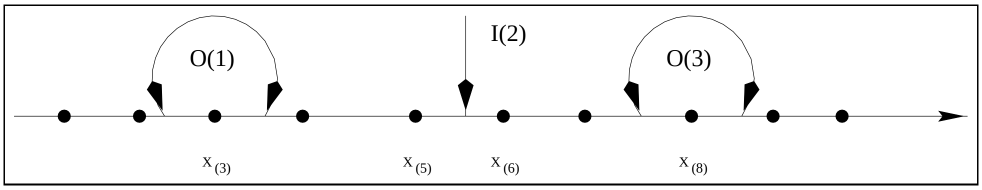

Figure 5. *Illustration of the almost sure dynamics of the sequences* $\{q_i^p\}$ *(*$i = 1, 2, 3$*) when* $n = 10$ *and* $m_n = 4$*.*

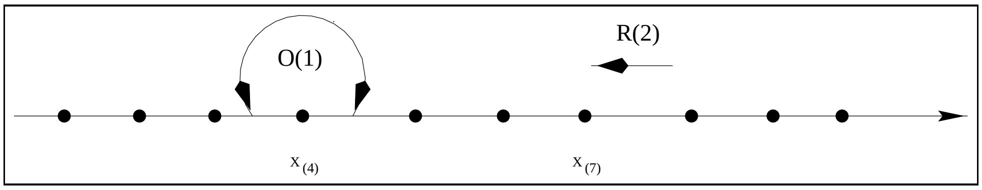

Figure 6. *Illustration of a theoretically possible dynamics for the sequences* $\{q_i^p\}$ *(*$i = 1, 2$*) when* $n = 10$ *and* $m_n = 3$*.*

References

[1] T. W. Anderson, "Some nonparametric multivariate procedures based on statistically equivalent blocks," in *Multivariate Analysis*, New York: Academic, 1966, pp. 5–27.

[2] A. R. Barron, "The convergence in information of probability density estimators," in *Proceedings of the International Symposium of IEEE on Information Theory*, Kobe, Japan, June 19–24, 1988.

[3] A. R. Barron, L. Györfi, and E. van der Meulen, "Distribution estimation consistent in total variation and in two types of information divergence," *IEEE Transactions on Information Theory*, vol. 38, pp. 1437–1454, 1992.

[4] J. Beirlant, A. Berlinet, G. Biau, and I. Vajda, "Divergence-type errors of smooth Barron-type density estimators," *Test*, vol. 11, pp. 191-217, 2002.

[5] A. Berlinet and E. Brunel, "Choix optimal du nombre de classes pour l'estimateur de Barron de la densité," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 331, pp. 713–716, 2000.

[6] A. Berlinet and E. Brunel, "Kullback-Leibler optimization of density estimates," in *Statistics 2001,* Y. Chaubey Ed., Imperial College Press, London, 2002.

[7] A. Berlinet and E. Brunel, "Optimal bin width choice in Kullback-Leibler divergence for modified histograms," Submitted, 2002.

[8] A. Berlinet, L. Györfi, and E. van der Meulen, "Asymptotic normality of relative entropy in multivariate density estimation," *Publications de l'Institut de Statistique de l'Université de Paris*, vol. 41, pp. 3–27, 1997.

[9] A. Berlinet, I. Vajda, and E. van der Meulen, "About the asymptotic accuracy of Barron density estimates," *IEEE Transactions on Information Theory*, vol. 44, pp. 999–1009, 1998.

[10] G. Biau, *Méthodes Itératives en Estimation Fonctionnelle et Systèmes Dynamiques.* PhD Thesis, University Montpellier II, 2000.

[11] P. Billingsley, *Convergence of Probability Measures.* New York: Wiley, 1968.

[12] D. Bosq and J. P. Lecoutre, *Théorie de l'Estimation Fonctionnelle.* Paris: Economica, 1987.

[13] L. Devroye, *A Course in Density Estimation.* Boston: Birkhäuser, 1987.

[14] L. Györfi, F. Liese, I. Vajda, and E. van der Meulen, "Distribution estimates consistent in χ^2-divergence," *Statistics*, vol. 32, pp. 31–57, 1998.

[15] L. Györfi and E. van der Meulen, "Density-free convergence properties of various estimators of entropy," *Computational Statistics and Data Analysis,* vol. 5, pp. 425–436, 1987.

[16] L. Györfi and E. van der Meulen, "There is no density estimate consistent in information divergence for all densities," in *Transactions of the Twelfth Conference on Information Theory, Statistical Decision Functions and Random Processes,* pp. 88–90, 1994.

[17] F. Liese and I. Vajda, *Convex Statistical Distance.* Leipzig: Teubner–Verlag, 1987.

[18] J. W. Tukey, "Nonparametric estimation II. Statistically equivalent blocks and tolerance regions. The continuous case," *The Annals of Mathematical Statistics*, vol. 18, pp. 529–539, 1947.