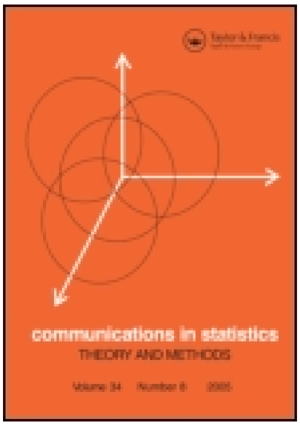

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

Local Distortion and μ-Mass of the Cells of One Dimensional Asymptotically Optimal Quantizers

Sylvain Delattre [a] , Jean-Claude Fort [b d] & Gilles Pagès [c]

[a] Laboratoire de Probabilités et Modèles Aléatoires, Département de Mathématiques, Université Paris 7, Paris, France

[b] Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse and SAMOS, Université Paris I, Paris, France

[c] Laboratoire de Probabilités et Modèles Aléatoires, Université Paris 6, Paris, France

[d] Laboratoire de Statistique et Probabilités, UMR C5583 Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse Cedex 4, France

Version of record first published: 02 Sep 2006.

To cite this article: Sylvain Delattre , Jean-Claude Fort & Gilles Pagès (2004): Local Distortion and μ-Mass of the Cells of One Dimensional Asymptotically Optimal Quantizers, Communications in Statistics - Theory and Methods, 33:5, 1087-1117

To link to this article: http://dx.doi.org/10.1081/STA-120029827

COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 33, No. 5, pp. 1087–1117, 2004

Local Distortion and μ-Mass of the Cells of One Dimensional Asymptotically Optimal Quantizers

Sylvain Delattre,[1] Jean-Claude Fort,[2,*] and Gilles Pagès[3]

[1]Laboratoire de Probabilités et Modèles Aléatoires, Département de Mathématiques, Université Paris 7, Paris, France
[2]Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse and SAMOS, Université Paris I, Paris, France
[3]Laboratoire de Probabilités et Modèles Aléatoires, Université Paris 6, Paris, France

ABSTRACT

We consider one dimensional probability distributions μ having a continuous and positive probability density function. We find the asymptotic of the size and the mass of the Voronoi cells and we prove that the local distortion associated with stationary or optimal quantizers is asymptotically uniform. Numerical simulations and computations illustrate the theoretical results and lead to the design of some good-fit test for the stationary equilibria.

*Correspondence: Jean-Claude Fort, Laboratoire de Statistique et Probabilités, UMR C5583 Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse Cedex 4, France; E-mail: fortjc@cict.fr.

DOI: 10.1081/STA-120029827 0361-0926 (Print); 1532-415X (Online)

Key Words: Vector quantization; Empirical measure; Weak convergence; Local distortion.

MSC Classification (2000): 68P30; 94A29; 62E17.

INTRODUCTION

The quantization of probability distributions appeared in the early 40's and it has now become an important field of Information Theory. It has been conceived in order to drastically cut down the storage of signal data to be analyzed. Nowadays image processing is a typical field of application. The aim of the method is to replace an actual data by a prototype chosen in a small set of prototypes.

The simplest modeling framework is to assume that the data are a sample of i.i.d. μ-distributed random variables. Then the information is entirely summed up in μ and the best choice is when the prototypes minimize some error criterion. The most encountered criterion, called *quadratic distortion*, is the μ-expectation of the square distance between actual data and their corresponding prototypes.

For a survey on mathematical aspects of quantization, one may consult Gersho and Gray (1982) or, more recently Graf and Luschgy (2000) whereas, for more applied aspects in the fields of information theory and signal processing, the book Gersho and Gray (1992) is more appropriate.

More recently, quantization drew the attention of some applied probabilists interested in numerical methods for multidimensional nonlinear problems. Several algorithms based on quantization of probability distributions have been designed recently. Let us mention the pricing of multi-asset American options (see Bally et al., 2002), the discretization of (possibly Reflected) Backward Stochastic Differential Equation (see Bally and Pagès, 2003; Bally et al., 2001), the numerical solving of multidimensional Stochastic Control problems or nonlinear filtering (see Pagès et al., 2003). In all these methods, quantization is used as an efficient and robust tool to compute conditional expectations. The common feature of these new applications of quantization is that they need to compute some very accurate estimates of the desired optimal quantizers. In a previous paper (Fort and Pagès, 2002) we gave semi-closed formulae for quantizers of some particular one dimensional probability distributions. For more general probability distributions finding out some testing procedure of this accuracy was one important motivation for this work as explained below. Thus, the case of Gaussian distributions is especially important for applications (see Pagès and Printems, 2003 where this test is already used to validate the accuracy of an optimal quantization).

From a mathematical viewpoint, the optimal (μ, r)-quantization $(r > 0)$ is the global minimization of the *distortion* function $D_n^{\mu,r}$ with respect to the distribution, μ that is

$$(x_1, \ldots, x_n) \longmapsto D_n^{\mu,r}(x_1, \ldots, x_n) := \int_{\mathbb{R}^d} \min_{1 \le i \le n} |x_i - \xi|^r \mu(d\xi) \qquad (1)$$

(μ is a probability distribution on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ with a moment of order r). So that the resulting error using x_i instead of ξ is measured by the *loss function* $V(\xi) := |\xi|^r$.

When μ has a moment of order r there exists at least a global minimizer $x^{(n)}$ of the distortion function $D_n^{\mu,r}$ defined in (1) (see Abaya and Wise, 1992; Pagès, 1997 or Graf and Luschgy, 2000). Most investigations on optimal quantization deal with the asymptotics of $x^{(n)}$ and that of $D_n^{\mu,r}(x^{(n)}) = \min_{(\mathbb{R}^d)^n} D_n^{\mu,r}$ as the number n of prototypes goes to $+\infty$.

Thus, once established, which is easy, that the minimum distortion converges to 0, the question was to determine the rate of convergence. The main result in that direction, usually known as Zador's Theorem, is due to several authors and was finally completed by Graf and Luschgy 2000 (Bucklew and Wise, 1982; Zador, 1982; see also Benett, 1948 and Graf and Luschgy, 2000).

Let λ_d denote the Lebesgue measure on $\mathbb{R}^d$. If $\mu(d\xi) := f(\xi)\,\lambda_d(d\xi) + \mu_d(d\xi)$, $\mu_d \perp \lambda_d$ and $\int_{\mathbb{R}^d} |\xi|^{r+\eta}\mu(d\xi) < +\infty$ $(\eta > 0)$, then

$$\lim_n \left(n^{\frac{r}{d}} \min_{(\mathbb{R}^d)^n} D_n^{\mu,r}\right) = J_{r,d}\left(\int_{\mathbb{R}^d} f(\xi)^{\frac{d}{d+r}}\,d\xi\right)^{1+\frac{r}{d}} < +\infty. \qquad (2)$$

The constant $J_{r,d}$ is the limiting constant obtained for the uniform distribution over $[0,1]^d$ (or over any Borel set of Lebesgue measure 1).

From a statistical viewpoint, it may look natural to investigate the (standard) empirical measures $\mu^n_{x^{(n)}} := 1/n\sum_{i=1}^n \delta_{x_i^{(n)}}$, $n \ge 1$, related to a sequence of optimal (or locally optimal) quantizers $x^{(n)}$, in order to "reconstruct" the original distribution μ (when absolutely continuous). In 1975, McClure obtained (McClure, 1975) the somewhat unexpected and disappointing result: when the probability density function (p.d.f.) has a compact support,

$$\mu^n_{x^{(n)}} \text{ weakly converges toward } \mu^\infty := \frac{f^{\frac{d}{d+r}}(\xi)}{\int_{\mathbb{R}^d} f^{\frac{d}{d+r}}(u)du}\,d\xi. \qquad (3)$$

It has been extended to general p.d.f.'s $f \in L^{r+\eta}(\mathbb{R}^d)$, $\eta > 0$ in Graf and Luschgy (2000).

This shows that the standard empirical measure is not the right approach to reconstruct the original distribution μ. Actually, weak convergence obtained in (3) turns out to be consistent with the above Zador Theorem (which is actually the key argument of the proof).

The correct method to reconstruct the distribution μ is the following: one considers *a projection following the closest neighbor rule*, denoted $\pi^{(n)}$. Then one considers the image $\widetilde{\mu}^n_{x^{(n)}}$ of μ by this projection, namely

$$\widetilde{\mu}^n_{x^{(n)}} := \mu \circ (\pi^{(n)})^{-1} = \sum_{i=1}^{n} \mu(\{\pi^{(n)} = x_i^{(n)}\})\, \delta_{x_i^{(n)}}, \quad n \geq 1. \tag{4}$$

Furthermore (see Proposition 1 below) $\widetilde{\mu}^n_{x^{(n)}}$ converges weakly toward μ along Lipschitz continuous functions at the same rate as the minimum distortion goes to zero when the size n of the quantizer goes to ∞. So, this rate is ruled by the above Zador Theorem.

Combining these two results leads to infer the behavior of the μ-mass of the cells and of the local distortion. In fact Gersho (1979) conjectured that the generic situation is that the local distortion is asymptotically uniformly distributed. So, a slightly more general conjecture is that, for distributions μ having a positive p.d.f. f

$$\begin{cases} \mu(C_i(x^{(n)})) = \dfrac{1}{n} f^{\frac{r}{r+d}}(x_i^{(n)}) \left(\displaystyle\int_{\mathbb{R}^d} f^{\frac{d}{r+d}}(\xi) d\xi \right) + o(1/n), \\ \displaystyle\int_{C_i(x^{(n)})} |x_i^{(n)} - \xi|^r \mu(d\xi) = \dfrac{D_n^{\mu,r}(x^{(n)})}{n} + o(1/n) \quad \text{as } n \to +\infty, \end{cases} \tag{5}$$

uniformly as $x_i^{(n)}$ lies in a fixed compact. The main result of this paper is to prove this conjecture in one dimension. This result has a direct application in building faster algorithms to search optimal quantizers (see for instance Patane and Russo, 2001). It can also be used to test the accuracy of approximation of a stationary quantizer (see Sec. 6).

The paper is organized as follows. In Sec. 1, we give some definitions and basic properties. In Sec. 2, we recall the main general results about the asymptotics of the minimal distortion and of the optimal quantizers. In Sec. 3, we give a general preliminary result about the asymptotic behavior of the local distortion and the Voronoi cells. The main results are in Sec. 4, where we obtain precise estimates for the local distortion and the μ-mass of the Voronoi cells in the one dimensional case. In Sec. 6,

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

we show on simple examples how to use these results to check the accuracy of a proxy of a stationary quantizer.

Notations.

- $\Longrightarrow$ will denote the weak convergence of finite measures on $\mathbb{R}^d$.
- The letter C will denote a positive real constant that may change from line to line.
- λ_d will denote the Lebesgue measure on $\mathbb{R}^d$ (and λ the Lebesgue measure on $\mathbb{R}$).
- Let $S_n := \{z \in (\mathbb{R}^d)^n / z_i \neq z_j,\ i \neq j\}$ be the set of n-tuples having pairwise distinct components.
- The support of a probability measure μ will be denoted $\operatorname{supp}\mu$.

1. DEFINITIONS AND BASIC PROPERTIES

1.1. Distortion and Optimal Quantization

Let μ be a probability distribution on $(\mathscr{B}(\mathbb{R}^d), \mathbb{R}^d)$.

Definition 1. Let $r > 0$ be a positive real number. Assume that the distribution μ has a r-moment. The (μ, r)-distortion function is defined, for every n-tuple $x := (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$, by

$$D_n^{\mu,r}(x) := \int_{\mathbb{R}^d} \min_{1\le i\le n} |x_i - \xi|^r \mu(d\xi) = \mathbb{E}\Big(\min_{1\le i\le n} |x_i - X|^r \Big)$$
where $X \overset{\mathscr{L}}{\sim} \mu$.

and $|\cdot|$ denotes the canonical Euclidean norm on $\mathbb{R}^d$. The superscript μ will often be dropped.

The distortion function is clearly symmetric, so, if $d = 1$, one may restrict without loss of generality, to n-tuples $x \in \mathbb{R}^n$ such that $x_1 \le x_2 \le \cdots \le x_{n-1} \le x_n$.

In most applications, provided that μ has a second moment, optimal quantization is usually processed with respect to the *quadratic distortion* $D_n^{\mu,2}$.

The r-distortion function is continuous; in fact, when $r \ge 1$, $\sqrt[r]{D_n^{\mu,r}}$ is Lipschitz continuous and r-Hölder when $0 < r < 1$. One shows that this function *always reaches a minimum* (at least one, see, among others, Abaya and Wise, 1992, when $r = 2$, Pagès, 1997, when $r \neq 2$ or Graf and Luschgy, 2000).

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Optimal r-quantization consists in looking for a n-tuple that achieves the minimum of the function $D_n^{\mu,r}$. Such n-tuples have the following elementary properties (see, e.g., Abaya and Wise, 1992; Pagès, 1997 or Graf and Luschgy, 2000):

P1. If μ has an infinite support, then all the possible minima of $D_n^{\mu,r}$ are contained in S_n i.e., have *pairwise distinct* components. Furthermore, if

$$\operatorname{supp}\mu \text{ is a convex set} \tag{6}$$

then

- all the components of *any* minimum of $D_n^{\mu,r}$ lie in the interior $\overset{\circ}{\operatorname{supp}}\,\mu$ of $\operatorname{supp}\mu$,
- any $\operatorname{supp}\mu$-valued *local* minimum of $D_n^{\mu,r}$ actually lies in $S_n \cap \overset{\circ}{\operatorname{supp}}\,\mu$.

P2. Let $x \in S_n$. One defines a *Voronoi tessellation* $(C_i(x))_{1\le i\le n}$ of x as any Borel *partition* of $\mathbb{R}^d$ satisfying

$$C_i(x) \subset \{\xi \in \mathbb{R}^d / |x_i - \xi| = \min_{1\le j\le n} |x_j - \xi|\}, \quad 1 \le i \le n. \tag{7}$$

Then, one may *localize* $D_n^{\mu,r}(x)$ along the Voronoi cells $C_i(x)$, $i = 1, \ldots, n$, as follows

$$D_n^{\mu,r}(x) = \sum_{i=1}^{n} \int_{C_i(x)} |x_i - \xi|^r \mu(d\xi).$$

The boundary of the cell $C_i(x)$ with centroid x_i is contained in the union of finitely many (median) hyperplanes. It does not depend upon the specific choice of the Voronoi tessellation of x. Hence, if the distribution μ is *strongly continuous*, that is $\mu(H) = 0$ for any hyperplane H of $\mathbb{R}^d$, then the Voronoi tessellation of x is μ-essentially unique. In 1-dimension strong continuity of μ amounts to *continuity* i.e., $\mu(\{\xi\}) = 0$ for every $\xi \in \mathbb{R}$.

P3. (a) Assume that μ is strongly continuous and $r \ge 1$. Then (see, e.g., Pagès, 1997) $D_n^{\mu,r}$ is differentiable on S_n and

$$\forall x = (x_1, \ldots, x_n) \in S_n,$$
$$\nabla D_n^{\mu,r}(x) = r\left(\int_{C_i(x)} |x_i - \xi|^{r-1} \frac{x_i - \xi}{|x_i - \xi|} \mu(d\xi)\right)_{1\le i\le n} \tag{8}$$

with the convention $0/|0| = 0$.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

(b) Let $r \in (0,1)$. If there is some $c \in (1-r,1)$ such that, for every compact subset K of $\mathbb{R}^d$,

$$\sup_{x\in K}\int_{\mathbb{R}^d}\frac{\mu(d\xi)}{|\xi-x|^c}<+\infty \tag{9}$$

then, item (a) holds true for $D_n^{\mu,r}$ (see the appendix for a proof). This assumption is satisfied, e.g., by *absolutely continuous distributions* μ whose p.d.f. is *upper bounded on any compact set of* $\mathbb{R}^d$.

Combining **P1** and **P3** shows that the gradient in Eq. (8) vanishes at least

- at any minimum of the r-distortion,
- at any local minimum if Assumption (6) holds (provided that μ satisfies (9) when $0<r<1$).

This leads to the following natural definition.

Definition 2. A n-tuple $x \in S_n$ is a r-stationary quantizer (or stationary code book) if:

$$\nabla D_n^{\mu,r}(x)=0. \tag{10}$$

So, a locally optimal r-quantizer is always r-stationary when $r\geq 1$ (and when $0<r<1$ under appropriate assumptions on μ).

In 1-dimension, one may restrict to quantizers lying in the increasing simplex

$$\mathscr{F}_n^{\mu,+}:=\{x\in(m,M)^n/m<x_1<x_2<\cdots<x_k<\cdots<x_{n-1}<x_n<M\}.$$

where (m,M) denotes the interior of the (convex hull) of $\mathrm{supp}(\mu)$. Let $x\in\mathscr{F}_n^{\mu,+}$; the Voronoi tessellation $(C_i(x))_{1\leq i\leq n}$ of x can be described by introducing the boundary points $\widetilde{x}_1,\ldots,\widetilde{x}_{n+1}$:

$$\widetilde{x}_1:=m,\quad \widetilde{x}_i:=\frac{x_i+x_{i-1}}{2},\quad 2\leq i\leq n,\ \ \widetilde{x}_{n+1}:=M.$$

Then $C_i(x)=(\widetilde{x}_i,\widetilde{x}_{i+1})$, $1\leq i\leq n$. In turn, the differential of $D_n^{\mu,r}$ on $(m,M)^n\cap S_n$ reads: for every $x\in\mathscr{F}_n^{\mu,+}$,

$$\forall i\in\{1,\ldots,n\},\quad \frac{\partial D_n^{\mu,r}}{\partial x_i}(x)=r\int_{\tilde{x}_i}^{\tilde{x}_{i+1}}\mathrm{sign}(x_i-\xi)|\xi-x_i|^{r-1}\mu(d\xi).$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Consequently, a quantizer $x \in S_n \cap (m, M)^n$ is *stationary* iff

$$\forall i \in \{1, \ldots, n\}, \quad \int_{\tilde{x}_i}^{\tilde{x}_{i+1}} \operatorname{sign}(x_i - \xi)|x_i - \xi|^{r-1} \mu(d\xi) = 0. \tag{11}$$

This system of equations is one key to prove convergence of our main result (5).

2. WEAK ASYMPTOTICS OF THE OPTIMAL QUANTIZERS

The results contained in this section are classical. For proofs we refer to Graf and Luschgy (2000).

For the uniform distribution $\mu = U([0,1])$ it is easily seen that $x_i^{(n)} := ((2i-1)/2n)_{1 \le i \le n}$ is the unique stationary – hence optimal – r-quantizer of μ for every $r > 0$. Its Voronoi tessellation is $\left(\left[\frac{i-1}{n}, \frac{i}{n}\right]\right)_{1 \le i \le n}$ and its r-distortion is equal to $\min_{[0,1]^n} D_n^{U([0,1]),r} = 1/(2^r(r+1))$. This fits with the fact that, for that very distribution, the standard empirical measure minimizes the Kolmogorov–Smirnov distance to $U([0,1])$ among all the measure weighting at most n points. But, for a plain distribution μ, there is no reason why all the Voronoi cells of a minimum $x^{(n)}$ of the r-distortion should have the same μ-weight equal to $1/n$. So, it seems more natural to weight every Dirac mass $\delta_{x_i^{(n)}}$ with the μ-measure of its cell i.e., $\mu(C_i(x^{(n)}))$ as it appears in (4). This leads to set, for every n-tuple $x \in (\mathbb{R}^d)^n$:

$$\widetilde{\mu}_x^n := \sum_{1 \le i \le n} \mu(C_i(x)) \delta_{x_i}.$$

One purpose of Quantization Theory is to study the weak asymptotics of a sequence of (r-stationary) quantizers $x^{(n)}$ as $n \to +\infty$ using either the standard empirical measures $\mu_{x^{(n)}}^n := 1/n \sum_{i=1}^n \delta_{x_i^{(n)}}$ or their weighted counterpart $\widetilde{\mu}_x^n$ (weighted empirical measures).

2.1. Weak Convergence of Weighted Empirical Measures $\widetilde{\mu}_x^n$ Toward μ

The result is stated in the following proposition (see Pagès 1997).

Proposition 1. *Let $r > 0$. For every $n \in \mathbb{N}^*$, let $x^{(n)} \in S_n$. Then,*

$$\text{if } 0 < r \le 1, \quad \sup\left\{\left|\int_{\mathbb{R}^d} f\, d\widetilde{\mu}_x^n - \int_{\mathbb{R}^d} f\, d\mu\right|,\ f : \mathbb{R}^d \to \mathbb{R},\ r\text{-Hölder},\ [f]_r \le 1\right\} = D_n^{\mu,r}(x^{(n)}), \tag{12}$$

$$\textit{if } r \geq 1, \quad \sup\left\{\left|\int_{\mathbb{R}^d} f\, d\widetilde{\mu}^n_x - \int_{\mathbb{R}^d} f\, d\mu\right|, f : \mathbb{R}^d \to \mathbb{R}, \text{ Lipschitz}, [f]_{Lip} \leq 1\right\} \leq (D_n^{\mu,r}(x^{(n)}))^{\frac{1}{r}}, \tag{13}$$

so that

$$\widetilde{\mu}^n_{x^{(n)}} \overset{(\mathbb{R}^d)}{\Longrightarrow} \mu \quad \text{as soon as } \lim_n D_n^{\mu,r}(x_n^{(n)}) = 0. \tag{14}$$

(The equality follows by considering the function $f(\xi) := \min_{1\leq i\leq n} |\xi - x_i|^r$.) Furthermore, the rate of convergence of $\min_{(\mathbb{R}^d)^n} D_n^{\mu,r}$ is ruled by Zador's Theorem.

Theorem 1 (Bucklew and Wise, 1982; Graf and Luschgy, 2000; Zador, 1982). *Let $r > 0$. Assume that $\int_{\mathbb{R}^d} |\xi|^{r+\eta}\mu(d\xi) < +\infty$ for some $\eta > 0$. Set $f := d\mu/d\lambda_d$ (which can be possibly 0). Then*

$$\lim_n \left(n^{\frac{r}{d}} \min_{(\mathbb{R}^d)^n} D_n^{\mu,r}\right) = J_{r,d}\ \|f\|_{\frac{d}{d+r}} < +\infty \tag{15}$$

where $\|g\|_p := \left(\int_{\mathbb{R}^d} |g(\xi)|^p d\xi\right)^{1/p} \leq +\infty$, for every $p \in (0, +\infty)$.

The positive real constant $J_{r,d}$ corresponds to the uniform distribution on $[0,1]^d$. One knows that $J_{r,1} = 1/(2^r(r+1))$, $J_{2,2} = 5/(18\sqrt{3})$; the true value of $J_{r,d}$ is unknown for $d \geq 3$ although one knows that $J_{r,d} = (d/2\pi e)^{r/2} + o(d)$ as $d \to +\infty$ (see Graf and Luschgy, 2000).

Remark. If $\mu = U([0,1])$, the limit holds as an equality i.e., $n^r \min_{(\mathbb{R})^n} D_n^{U([0,1]),r} = J_{r,1}$.

2.2. Weak Convergence of the Standard Empirical Measures

Although misleading, the efforts to carry out the reconstruction of μ lead to study the weak convergence of the standard empirical measures $\mu^n_{x^{(n)}} := 1/n\sum_{k=1}^n \delta_{x_k^{(n)}}$. Next result confirms that it is not the right object to reconstruct μ as it could be suspected from Proposition 1.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Definition 3. A sequence $(x^{(n)})_{n\geq 1}$ of quantizers is asymptotically optimal if

$$\lim_n \left(n^{\frac{r}{d}} D_n^{\mu,r}(x^{(n)}) \right) = J_{r,d} \|f\|_{\frac{d}{d+r}}$$

i.e., if it achieves the true rate of convergence in Zador's Theorem.

Theorem 2 (Graf and Luschgy, 2000; McClure, 1975)**.** *Let μ be an absolutely continuous distribution on $\mathbb{R}^d$, with p.d.f. f and let $r > 0$. Assume that $\int_{\mathbb{R}^d} |\xi|^{r+\eta} \mu(d\xi) < +\infty$ for some $\eta > 0$. Let $(x^{(n)})_{n\geq 1}$ be a sequence of asymptotically optimal r-quantizers. Then*

$$\mu^n_{x^{(n)}} := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i^{(n)}} \overset{(\mathbb{R})}{\Longrightarrow} \frac{1}{c_{f,\frac{d}{r+d}}} f^{\frac{d}{r+d}}(\xi) d\xi \quad \textit{with}$$

$$c_{f,\alpha} := \int_{\mathbb{R}^d} f^{\alpha}(\xi)\, d\xi, \quad \alpha > 0. \tag{16}$$

3. FIRST STRONG RESULTS ON THE μ-MASS OF THE CELLS

Proposition 2. *For every $n \in \mathbb{N}^*$, let $x^{(n)} := (x_i^{(n)})_{1\leq i\leq n} \in (\mathrm{supp}(\mu))^n$ be a quantizer.*

(a) *Assume that $\widetilde{\mu}^n_{x^{(n)}} \overset{(\mathbb{R}^d)}{\Longrightarrow} \mu$. Then, for every compact set K of $\mathbb{R}^d$,*

$$\lim_{n\to\infty} \sup_{\{i/x_i^{(n)}\}} \mu(C_i(x^{(n)})) = 0.$$

(b) *Assume that $\lim_{n\to\infty} D_n^{\mu,r}(x^{(n)}) = 0$ for some $r > 0$. Then, $\widetilde{\mu}^n_{x^{(n)}} \overset{(\mathbb{R}^d)}{\Longrightarrow} \mu$ and, furthermore,*

$$\lim_{n\to\infty} \sup_{\{1\leq i\leq n\}} \mu(C_i(x^{(n)})) = 0.$$

Remark. In Proposition 2 no stationarity assumption is made on $x^{(n)}$ (the result is trivial if $x^{(n)}$ is optimal). Of course for numerical applications, one never reaches true optimal quantizers but only some approximations of stationary quantizers via numerical algorithms. The above result still holds in this more realistic setting.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Proof. (a) Let K be a compact set. If the result does not hold, there is some positive ε_0, and a sequence $i_{\phi(n)}$ such that:

$$\forall k \geq 1, \quad \mu(C_{i_{\phi(n)}}(x^{(\phi(n))})) \geq \varepsilon_0.$$

We may assume (up to an extraction) that the sequence $x^{(n)}_{i_{\phi(n)}}$ converges toward x_∞ in K. Let $\alpha > 0$ and g_α be a bounded continuous function such that:

$$g_\alpha(\xi) = 1, \text{ if } |\xi - x_\infty| \leq \alpha \quad \text{and} \quad g_\alpha(\xi) = 0, \text{ if } |\xi - x_\infty| \geq 2\alpha.$$

Then for large enough n , $\int g_\alpha \, d\widetilde{\mu}^{\phi(n)}_{x^{(\phi(n))}} \geq g_\alpha(x^{(\phi(n))}_{i_{\phi(n)}})\mu(C_{i_{\phi(n)}}(x^{(\phi(n))})) \geq \varepsilon_0$.
This implies $\mu(g_\alpha) \geq \varepsilon_0$ for every $\alpha > 0$. This contradicts $\lim_{\alpha \to 0} \mu(g_\alpha) = 0$.

(b) Let $r > 0$, such that $\lim_n D^{\mu,1}_n(x^{(n)}) = 0$. If $r \geq 1$, the monotony of the L^r-norm implies that $D^{\mu,1}_n(x^{(n)}) \leq \sqrt[r]{D^{\mu,r}_n(x^{(n)})}$ goes to 0 as well. So, one may assume without loss of generality that $0 < r \leq 1$. Then, it follows from (12) and (13) that for any r-Holder continuous function g,

$$\left| \int g \, d\widetilde{\mu}^n_{x^{(n)}} - \int g \, d\mu \right| \leq [g]_r D^{\mu,1}_n(x^{(n)}).$$

Now we apply this to the r-Holder function $g(\xi) := |\xi|^r h_K(\xi)$, where h_K is a $[0,1]$-valued Lipschitz continuous function such that $h(\xi) = 1$ if $|\xi| \geq K+1$, $h(\xi) := 0$ if $|\xi| \leq K$. We immediately obtain:

$$\begin{aligned} \overline{\lim_n} \sum_{\{i\,/\,|x^{(n)}_i| \geq K+1\}} |x^{(n)}_i|^r \mu(C_i(x^{(n)})) &\leq \int |\xi|^r h_K(\xi)\mu(d\xi) \\ &\leq \int_{|\xi| \geq K\}} |\xi|^r \, \mu(d\xi). \end{aligned}$$

One concludes using (a) and $\lim_{K \to +\infty} \int_{\{|\xi| \geq K\}} |\xi| \, \mu(d\xi) = 0$. □

4. ASYMPTOTIC LOCAL DISTORTION IN THE ONE DIMENSIONAL CASE

We now assume $d=1$ and that $x^{(n)}$ is a sequence of stationary quantizers for the r-distortion, with $r>0$. Notice that uniqueness is not guaranteed. It is a nontrivial property since it is known that many distributions have several stationary r-quantizers (see, e.g., Pagès, 1997). The classical result about uniqueness is due to several authors (Fleisher, Lloyd, Sharma), finally completed by Trushkin (1982) and Kieffer (1982). It can be established by three different methods (see Kieffer, 1982; Lamberton and Pagès, 1996 or Cohort, 1997; Trushkin, 1982).

Theorem 3 (Trushkin and Kieffer). *Let $r \geq 1$. If the p.d.f. f is log-concave, then*

$$\operatorname{argmin} D_n^{\mu,r} = \operatorname{argmin\,loc} D_n^{\mu,r} = \{\nabla D_n^{\mu,r} = 0\} = \{x^{(n)}\}.$$

Remark. There is no real doubt that this theorem holds true for $0 < r < 1$, although it is not clearly stated in the literature. Thus, methods based on the gradient like Lamberton and Pagès (1996) or Cohort (1997) seem to admit a straightforward extension.

4.1. Stationary Asymptotically Optimal Quantizers

We will rely on Eq. (11) although we do not assume uniqueness. It is natural to introduce the stepwise function associated to the μ-mass of the cells renormalized by a factor n:

$$\varphi_n(\xi) := n \sum_{i=1}^{n} \mathbf{1}_{C_i(x^{(n)})}(\xi)\mu(C_i(x^{(n)})).$$

Theorem 4. *Let μ be a distribution on the real line with a connected support. Assume that μ has a p.d.f. f which is positive and Lipschitz continuous on every compact set of the interior (m, M) of* supp(μ). *Let $(x^{(n)})_{n\geq 1}$ be a sequence of stationary r-quantizers, $r > 0$.*

(a) *The sequence $(\varphi_n)_{n\geq 1}$ is relatively compact for the topology of the uniform convergence on compact sets of (m, M).*
(b) *If, furthermore, $(x^{(n)})_{n\geq 1}$ is asymptotically optimal the sequence φ_n converges uniformly on compacts sets of (m, M) toward $c_{f,1/(r+1)} f^{\frac{r}{r+1}}$ i.e., for every $[a, b] \subset (m, M)$,*

$$\sup_{\{i\,/\,x_i^{(n)} \in [a,b]\}} \left| n\,\mu\big(C_i(x^{(n)})\big) - c_{f,1/(r+1)} f^{\frac{r}{r+1}}\big(x_i^{(n)}\big) \right| \xrightarrow{n\to+\infty} 0. \qquad (17)$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

The local distortion is asymptotically uniformly distributed i.e., for every for every $[a,b] \subset (m, M)$,

$$\sup_{\{i\,/\,x_i^{(n)}\in[a,b]\}} \left| n^{r+1} \int_{C_i(x^{(n)})} |x_i^{(n)} - \xi|^r \mu(d\xi) - J_{r,1} \|f\|_{1/(r+1)} \right| \xrightarrow{n\to+\infty} 0. \tag{18}$$

(c) *Moreover, if μ has a compact support $[m, M]$ and f is Lipschitz continuous on the whole interval $[m, M]$, then all the above convergences hold uniformly on $[m, M]$.*

Proof of the Theorem 4. (a) Once set $\Delta x_i^{(n)} = x_i^{(n)} - x_{i-1}^{(n)}$, Eq. (11) reads:

$$\int_{-\frac{\Delta x_i^{(n)}}{2}}^{\frac{\Delta x_{i+1}^{(n)}}{2}} v^{r-1} \operatorname{sign}(v)\, f(x_i^{(n)} + v) dv = 0, \quad 1 \le i \le n.$$

Introducing $f(x_i^{(n)})$ we obtain:

$$\int_{-\frac{\Delta x_i^{(n)}}{2}}^{0} (-v)^{r-1} \big(f(x_i^{(n)} + v) - f(x_i^{(n)})\big) dv + f(x_i^{(n)}) \frac{(\Delta x_i^{(n)})^r}{r2^r}$$
$$= \int_0^{\frac{\Delta x_{i+1}^{(n)}}{2}} v^{r-1} \left(f(x_i^{(n)} + v) - f(x_i^{(n)})\right) dv + f(x_i^{(n)}) \frac{(\Delta x_{i+1}^{(n)})^r}{r2^r}.$$

Putting $v := (\Delta x_i^{(n)}/2)u$ and $v := (\Delta x_{i+1}^{(n)}/2)u$, respectively, we have, since $f(x_i^{(n)}) > 0$,

$$(\Delta x_i^{(n)})^r \left(1 + r \int_0^1 u^{r-1} \frac{f(x_i^{(n)} - \frac{\Delta x_i^{(n)}}{2} u) - f(x_i^{(n)})}{f(x_i^{(n)})} du \right)$$
$$= (\Delta x_{i+1}^{(n)})^r \left(1 + r \int_0^1 u^{r-1} \frac{f(x_i^{(n)} + \frac{\Delta x_{i+1}^{(n)}}{2} u) - f(x_i^{(n)})}{f(x_i^{(n)})} du \right).$$

Setting $H(x,y) := \int_0^1 r u^{r-1} \dfrac{f(x + u\frac{y-x}{2}) - f(x)}{f(x)} du$ finally leads to

$$\frac{(\Delta x_{i+1}^{(n)})^r}{(\Delta x_i^{(n)})^r} = \frac{1 + H(x_i^{(n)}, x_{i+1}^{(n)})}{1 + H(x_i^{(n)}, x_{i-1}^{(n)})} \tag{19}$$

Let $[a,b]\subset(m,M)$ and let $L_f^{a,b}$ denote the Lipschitz coefficient of f on $[a,b]$.

$$\forall\xi,\ \xi'\in[a,b],\quad |H(\xi,\xi')|\le L_f^{a,b}\frac{r}{r+1}\frac{|\xi-\xi'|}{f(\xi)}\le C\,|\xi-\xi'|$$

since f is bounded away from 0 on $[a,b]$. Consequently:

$$|H(x_i^{(n)},x_{i-1}^{(n)})|\le C\,|x_i^{(n)}-x_{i-1}^{(n)}|\quad\text{and}$$
$$|H(x_i^{(n)},x_{i+1}^{(n)})|\le C\,|x_i^{(n)}-x_{i+1}^{(n)}|,$$

whenever $x_{i\pm1}^{(n)}$ lie in $[a,b]$.

It follows from Proposition 2 that $\max_{\{i/a\le x_{i\pm1}^{(n)}\le b\}}\max(\Delta x_i^{(n)},\Delta x_{i+1}^{(n)})$ goes to 0, so we can estimate the right hand of (19):

$$\frac{1+H(x_i^{(n)},x_{i-1}^{(n)})}{1+H(x_i^{(n)},x_{i+1}^{(n)})}=\exp(H(x_i^{(n)},x_{i-1}^{(n)})-H(x_i^{(n)},x_{i+1}^{(n)}))$$
$$+\,O(\max(\Delta x_i^{(n)},\Delta x_{i+1}^{(n)})^2)$$

where

$$\begin{aligned}&\max_{\{i/a\le x_{i\pm1}^{(n)}\le b\}}|O(\max(\Delta x_i^{(n)},\Delta x_{i+1}^{(n)})^2)|\\&\quad\le C\max_{\{i/a\le x_{i\pm1}^{(n)}\le b\}}\max(\Delta x_i^{(n)},\Delta x_{i+1}^{(n)})^2\overset{n\to\infty}{\longrightarrow}0\end{aligned}.\qquad(20)$$

Now, we have $\sum_{\{i/a\le x_{i\pm1}^{(n)}\le b\}}\Delta x_i^{(n)}\le b-a$, so that:

$$\sum_{\{i/a\le x_{i\pm1}^{(n)}\le b\}}(\Delta x_i^{(n)})^2\le(b-a)\sup_{\{i/a\le x_{i\pm1}^{(n)}\le b\}}(\Delta x_i^{(n)})=(b-a)o(1).\qquad(21)$$

Now set $\alpha_k:=\exp(H(x_{i+k}^{(n)},x_{i+k+1}^{(n)})-H(x_{i+k}^{(n)},x_{i+k-1}^{(n)}))$ and $\delta_k:=O(\max\,(\Delta x_{i+k+1}^{(n)},\Delta x_{i+k}^{(n)})^r)$. Note that, as long as $x_i^{(n)}$ and $x_{i+p}^{(n)}\in[a,b]$, $\min_{1\le k\le p-1}\alpha_k$ is bounded away from 0 independently of n by a real constant $\underline{\alpha}>0$ since the function H is bounded away from $-\infty$ on $[a,b]^2$. Hence

$$\begin{aligned}\left(\frac{\Delta x_{i+p}^{(n)}}{\Delta x_{i+1}^{(n)}}\right)^r&\le\prod_{k=1}^{p-1}(\alpha_k+|\delta_k|)=\left(\prod_{k=1}^{p-1}\alpha_k\right)\prod_{k=1}^{p-1}\left(1+\frac{|\delta_k|}{\alpha_k}\right)\\&\le\left(\prod_{k=1}^{p-1}\alpha_k\right)\exp\left(\sum_{k=1}^{p-1}\frac{|\delta_k|}{\underline{\alpha}}\right)\le\left(\prod_{k=1}^{p-1}\alpha_k\right)e^{(b-a)\,o(1)}.\end{aligned}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

One obtains a lower bound the same way round. Assume that n is large enough so that, $\max\{|\delta_k|,\ a \le x_{k\pm 1}^{(n)} \le b\} \le \underline{\alpha}/2$. Using that $\ln(1-u) \ge -2u$ if $0 \le u \le 1/2$,

$$\left(\frac{\Delta x_{i+p}^{(n)}}{\Delta x_{i+1}^{(n)}}\right)^r \ge \left(\prod_{k=1}^{p-1}\alpha_k\right)\prod_{k=1}^{p-1}\left(1-\frac{|\delta_k|}{\underline{\alpha}}\right)$$
$$\ge \left(\prod_{k=1}^{p-1}\alpha_k\right)\exp\left(-2\sum_{k=1}^{p-1}\frac{|\delta_k|}{\underline{\alpha}}\right) \ge \left(\prod_{k=1}^{p-1}\alpha_k\right)e^{(b-a)\,o(1)}.$$

Thus it yields that for large enough n and every $x_{i+p}^{(n)},\ x_i^{(n)} \in [a,b]$:

$$\exp\left(\sum_{k=1}^{p-1}\ln(\alpha_k)-(b-a)|o(1)|\right)$$
$$\le \left(\frac{\Delta x_{i+p}^{(n)}}{\Delta x_{i+1}^{(n)}}\right)^r \le \exp\left(\sum_{k=1}^{p-1}\ln(\alpha_k)+(b-a)|o(1)|\right). \tag{22}$$

On the other hand, using once again that f is bounded away from 0 on $[a,b]$,

$$\left|\sum_{k=1}^{p-1}\ln(\alpha_k)\right| = \left|\sum_{k=1}^{p-1} r\int_0^1 u^{r-1}\frac{f\left(x_{i+k}^{(n)}-\frac{\Delta x_{i+k}^{(n)}}{2}u\right)-f\left(x_{i+k}^{(n)}+\frac{\Delta x_{i+k+1}^{(n)}}{2}u\right)}{f(x_{i+k}^{(n)})}du\right|$$
$$\le L_f^{a,b}\frac{r}{r+1}\sum_{k=1}^{p-1}\left|\frac{\Delta x_{i+k+1}^{(n)}-\Delta x_{i+k}^{(n)}}{2f(x_{i+k}^{(n)})}\right| \le C(b-a). \tag{23}$$

Combined with Inequality (22), this provides

$$\max_{\{i/a\le x_{i-1}^{(n)}\le x_i^{(n)}\le b\}}\Delta x_i^{(n)} \le C \min_{\{i/a\le x_{i-1}^{(n)}\le x_i^{(n)}\le b\}}\Delta x_i^{(n)}.$$

It follows that

$$\overline{\lim_n}\left(n\max_{\{i/a\le x_{i-1}^{(n)}\le x_i^{(n)}\le b\}}\Delta x_i^{(n)}\right) < C(b-a). \tag{24}$$

Straightforward computations show that, whenever $x_{i\pm1}^{(n)} \in [a,\ b]$,

$$\left|\frac{\mu(C_i(x^{(n)}))}{(\Delta x_{i+1}^{(n)}+\Delta x_i^{(n)})f(x_i^{(n)})/2}-1\right| \le \frac{L_f^{a,b}}{f(x_i^{(n)})}(\Delta x_{i+1}^{(n)}+\Delta x_i^{(n)}).$$

Hence, using that $f > \varepsilon_{a,b} > 0$ on $[a,b]$, one derives

$$\lim_n \max_{\{i/x^{(n)}_{i\pm1}\in[a,b]\}} \left| \frac{\mu(C_i(x^{(n)}))}{(\Delta x^{(n)}_{i+1} + \Delta x^{(n)}_i) f(x^{(n)}_i)/2} - 1 \right| = 0. \tag{25}$$

Combining (24) and (25) yields

$$\overline{\lim_n} \left(n \max_{\{i/x^{(n)}_{i\pm1}\in[a,b]\}} \mu(C_i(x^{(n)})) \right) \le C(b-a) \sup_{[a,b]} f < +\infty. \tag{26}$$

Now, let $u, v \in [a,b]$, $u \le v$. Inequality (22) also yields for large enough n:

$$\frac{\max_{\{i/x^{(n)}_{i\pm1}\in[a,b]\}} \Delta x^{(n)}_i}{\min_{\{i/x^{(n)}_{i\pm1}\in[a,b]\}} \Delta x^{(n)}_i} \le \exp(C_{a,b}(v-u)). \tag{27}$$

It yields, still using (25), that for large enough n,

$$\frac{\max_{\{i/x^{(n)}_{i\pm1}\in[u,v]\}} \mu(C_i(x^{(n)}))}{\min_{\{i/x^{(n)}_{i\pm1}\in[u,v]\}} \mu(C_i(x^{(n)}))} \le \exp(C_{a,b}(v-u)) \frac{\sup_{[a,b]} f}{\inf_{[a,b]} f}. \tag{28}$$

Let a', b' such that $a < a' < b' < b$.

On the one hand, it follows from (26) that $\sup_n \|\varphi_n\|_\infty < +\infty$ ($\|\varphi_n\|_\infty$ is computed over $[a,b]$). On the other hand, it follows from (28) that, for large enough n, for every $u, v \in [a', b']$, $u \le v$,

$$\frac{\max(\varphi_n(u), \varphi_n(v))}{\min(\varphi_n(u), \varphi_n(v))} \le C'_{a,b} e^{C_{a,b}(v-u)}.$$

for some positive real constants $C_{a,b}$ and $C'_{a,b}$, so that

$$|\varphi_n(v) - \varphi_n(u)| \le C'_{a,b}(e^{C_{a,b}(v-u)} - 1)\|\varphi_n\|_\infty$$

which shows the equi-continuity property of the φ_n's on $[a', b']$.

(b) Since $x^{(n)}$ is asymptotically optimal, we know from Theorem 2 that

$$\mu^n_{x^{(n)}} \overset{(\mathbb{R})}{\Longrightarrow} \frac{f^{\frac{1}{1+r}}(\xi)}{c_{f,1/(1+r)}} d\xi.$$

It follows from Proposition 1 that $\widetilde{\mu}^n_{x^{(n)}} \overset{(\mathbb{R})}{\Longrightarrow} \mu$. But $\varphi_n.\mu_n^{x^{(n)}} = \widetilde{\mu}^n_{x^{(n)}}$, so that for any interval I, $\int_I \varphi_n(\xi)d\mu_n^{x^{(n)}}(\xi)$ converges to $\mu(I)$ as n goes to infinity. Hence any limiting function φ of the sequence $(\varphi_n)_{n\geq 1}$ satisfies

$$\int_a^b \varphi(\xi) f^{\frac{1}{1+r}}(\xi)\frac{d\xi}{c_{f,1/(1+r)}} = \int_a^b f(\xi)d\xi.$$

Consequently $\varphi = c_{f,1/(1+r)} f^{\frac{r}{1+r}}$, which shows the first convergence.

Let us pass now to the local distortion. First, one defines for every $n \geq 1$ two functions denoted f_n and $\widetilde{\varphi}_n$ by setting

$$\forall u \in C_i(x^{(n)}) \quad f_n(\xi) := f(x_i^{(n)}) \quad \text{and} \quad \widetilde{\varphi}_n(\xi) := n\,\Delta x_i^{(n)}.$$

Note that $f_n \to f$ uniformly on compact sets. It easily derives from the proof of item (a) that $\widetilde{\varphi}_n \to \varphi/f$ (use (25) and the fact that $f(x_i^{(n)})/f(x_{i+1}^{(n)})$ goes to 1 as $n\to\infty$ uniformly on compacts sets). Easy computations using that f is $L_f^{a,b}$-Lipschitz continuous yield

$$\begin{aligned}
&\left| n^{r+1}\int_{C_i(x^{(n)})} |x_i^{(n)} - \xi|^r (f(\xi) - f(x_i^{(n)}))d\xi \right| \\
&\quad \leq \frac{L^f_{a,b}}{(r+2)2^{r+2}} n^{r+1}\left((\Delta x_i^{(n)})^{r+2} + (\Delta x_{i+1}^{(n)})^{r+2}\right) \\
&\quad \leq \frac{L^f_{a,b}}{(r+2)2^{r+2}} \frac{1}{n}\left(\widetilde{\varphi}_n(x_i^{(n)})^{r+2} + \widetilde{\varphi}_n(x_{i+1}^{(n)})^{r+2}\right)
\end{aligned}$$

whereas

$$\begin{aligned}
&n^{r+1} f(x_i^{(n)}) \int_{C_i(x^{(n)})} |x_i^{(n)} - \xi|^r \, d\xi \\
&\quad = n^{r+1} f(x_i^{(n)}) J_{r,1} \frac{(\Delta x_i^{(n)})^{r+1} + (\Delta x_{i+1}^{(n)})^{r+1}}{2} \\
&\quad = J_{r,1} f(x_i^{(n)}) \frac{\widetilde{\varphi}_n(x_i^{(n)})^{r+1} + \widetilde{\varphi}_n(x_{i+1}^{(n)})^{r+1}}{2}.
\end{aligned}$$

Now $f_n\,\widetilde{\varphi}_n^{r+1} \overset{U_K}{\longrightarrow} f \times (\varphi/f)^{r+1} = c_{f,\frac{1}{r+1}}^{-(r+1)} = \|f\|_{\frac{1}{r+1}}$, so that

$$\lim_n \max_{\{i/x_{i\pm1}^{(n)}\in[a,b]\}} \left| n^{r+1}\int_{C_i(x^{(n)})} |x_i^{(n)} - \xi|^r \mu(d\xi) - J_{r,1}\|f\|_{\frac{1}{r+1}} \right| = 0.$$

(c) is obvious. □

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Remark. Equation (18) allows to retrieve Zador's Theorem since it shows that the asymptotic local distortion is constant and thus it gives an equivalent for $D_n^{\mu,r}(x^{(n)})$.

Our result deals with asymptotically optimal stationary quantizers. These stationary quantizers are generally the ones that are reachable via numerical methods. More precisely when no uniqueness is guaranteed the algorithms designed for finding good quantizers are based on the stationary Eq. (11). When we restrict to a sequence of true optimal quantizers it is possible to relax the Lipschitz condition, using another method based on the optimality property. This is the purpose of the next subsection.

4.2. The Case of Optimal Quantizers

In this section we assume that $(x^{(n)}, n \geq 1)$ is a sequence of optimal quantizers. Of course it is a more restrictive hypothesis than in the previous section and the consequence is that we can relax the Lipschitz condition for f to continuity. Our result reads:

Theorem 5. *Assume that μ has a p.d.f., denoted f, continuous and positive on every compact subset of the interior of* $\text{supp}(\mu)$. *Then for all compact interval $[a,b]$ of the interior of* $\text{supp}(\mu)$:

$$\lim_{n\to\infty} \sup\left\{\frac{\mu(C_i(x^{(n)}))f^{\frac{r}{r+1}}(x_i^n)}{\mu(C_j(x^{(n)}))f^{\frac{r}{r+1}}(x_j^n)}\ 1 \leq i, j \leq n,\, a \leq x_i^{(n)}, x_j^{(n)} \leq b\right\} = 1.$$

Hence, statements (b) *and* (c) *of Theorem* 4 *hold.*

The proof is the combination of three lemmas.

Lemma 1. *Let $\gamma > 1$ and $\Phi_\gamma(x) := x^\gamma - 1 + \gamma(1-x)$. Φ_γ is increasing on $[1,\infty)$, decreasing on $[0,1]$ and vanishes at $x = 1$. Let X be a nonnegative random variable such that $\mathbb{E}\, X^\gamma < +\infty$ with $\mathbb{E}\, X > 0$. Then,*

$$\mathbb{E}\Phi_\gamma\left(\frac{X}{\mathbb{E}X}\right) = \left(\frac{\mathbb{E}X^\gamma}{(\mathbb{E}X)^\gamma} - 1\right).$$

At this stage, we need to introduce some (temporary) notations:

$$\lambda_i := \lambda(C_i(x^{(n)})),$$
$$M_n(i,j) = M_n(j,i) := \max\left\{f(x), x \in \bigcup_{k=i}^{j} C_k(x^{(n)})\right\},$$
$$m_n(i,j) = m_n(j,i) := \min\left\{f(x), x \in \bigcup_{k=i}^{j} C_k(x^{(n)})\right\}.$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Lemma 2. *Let p and i two positive integers, with* $1 \le i \le i+p-1 \le n$. *Then:*

$$\frac{1}{p}\sum_{k=0}^{p-1}\Phi_{r+1}\left(\frac{\lambda_{i+k}}{\frac{1}{p}\sum_{l=0}^{p-1}\lambda_{i+l}}\right) \le \frac{M_n(i,i+p-1)}{m_n(i,i+p-1)}-1.$$

Lemma 3. *Let* p, q, i, j *four positive integers satisfying* $1 \le i < j \le n$ *and* $i+p-1 < j-q+1$. *Then*

$$p\left(m_n(i,i+p-1)-\frac{p^r}{q^r}M_n(i,i+p-1)\right)\left(\frac{1}{p}\sum_{k=0}^{p-1}\lambda_{i+k}\right)^{r+1}$$

$$\le q\left(\frac{q^r}{p^r}M_n(j-q+1,j)-m_n(j-q+1,j)\right)\left(\frac{1}{q}\sum_{k=0}^{q-1}\lambda_{j-k}\right)^{r+1}.$$

This inequality holds also exchanging i, p, $i+p-1$, $i+k$ *with* j, q, $j-q+1$, $j-k$.

Proof of Theorem 5. From Lemma 2 we have for $1 \le i < j \le n$:

$$\sup_{i\le k\le j}\Phi_{r+1}\left(\frac{\lambda_k}{\frac{1}{j-i+1}\sum_{l=i}^{j}\lambda_l}\right) \le (j-i+1)\left(\frac{M_n(i,j)}{m_n(i,j)}-1\right). \qquad (29)$$

Let $[a,b]$ be a compact interval of the interior of $\mathrm{supp}(\mu)$. It is clear that $\lim_{n\to\infty}\sup_{x_i^{(n)}\in[a,b]}\lambda_i = 0$ since

$$\frac{\inf_{x\in[a,b]}f(x)}{2^r(r+1)}\ \lambda_i^{r+1} \le \int_{C_i(x^{(n)})}|\xi-x_i^{(n)}|^r f(\xi)d\xi \le D_n^{\mu,r}(x^{(n)}) = \min_{\mathbb{R}^n} D_n^{\mu,r}.$$

Consequently, inequality (29), the uniform continuity of f on $[a,b]$ and the properties of Φ_{r+1} imply that

$$\text{for all } m\ge 1,\quad \lim_{n\to\infty}\ \sup_{\substack{1\le j-i\le m\\ a\le x_i^{(n)}\le x_j^{(n)}\le b}}\ \sup_{i\le k\le j}\left|\frac{\lambda_k}{\frac{1}{j-i+1}\sum_{l=i}^{j}\lambda_l}-1\right| = 0. \qquad (30)$$

On the other hand, as a consequence of Lemma 3 we have the following inequalities. Let $1 \le i < j \le n$ and $p + q \le j - i + 1$, then if $\frac{p}{q} < \left(\frac{m_n(i, i+p-1)}{M_n(i, i+p-1)}\right)^{\frac{1}{r}}$:

$$\frac{\frac{1}{p}\sum_{k=0}^{p-1}\lambda_{i+k}}{\frac{1}{q}\sum_{k=0}^{q-1}\lambda_{j-k}} \le \frac{q}{p}\left(\frac{M_n(j-q+1, j) - \frac{p^r}{q^r} m_n(j-q+1, j)}{m_n(i, i+p-1) - \frac{p^r}{q^r} M_n(i, i+p-1)}\right)^{\frac{1}{r+1}} \tag{31}$$

Under the same assumption the inequality holds exchanging $i,\ p,\ i+p-1$ with $j,\ q,\ j-q+1$. It follows from Inequality (31) with $p = m,\ q = m+1$ and from the uniform continuity of f that, for large enough m:

$$\lim_{n\to\infty} \sup_{\substack{(i,j),\, j-i \ge 2m \\ a \le x_i^{(n)}, x_j^{(n)} \le b}} \frac{f(x_i^{(n)})^{\frac{1}{r+1}} \frac{1}{m}\sum_{k=0}^{m-1}\lambda_{i+k}}{f(x_j^{(n)})^{\frac{1}{r+1}} \frac{1}{m+1}\sum_{k=0}^{m}\lambda_{j-k}} \le 1 + \frac{1}{m}.$$

Exchanging i, p and j, q in (31) yields:

$$\lim_{n\to\infty} \sup_{\substack{(i,j),\, j-i \ge 2m \\ a \le x_i^{(n)}, x_j^{(n)} \le b}} \frac{f(x_j^{(n)})^{\frac{1}{r+1}} \frac{1}{m}\sum_{k=0}^{m-1}\lambda_{j-k}}{f(x_i^{(n)})^{\frac{1}{r+1}} \frac{1}{m+1}\sum_{k=0}^{m}\lambda_{i+k}} \le 1 + \frac{1}{m}.$$

Using these last two inequalities together with (30), the proof is easily completed. □

Proof of Lemma 1. The properties of Φ_γ follow from the strict convexity of $x \mapsto x^\gamma$. □

Proof of Lemma 2. Let $[\alpha^{(n)}, \beta^{(n)}] := \bigcup_{k=0}^{p-1} C_{i+k}(x^{(n)})$. Consider a quantizer x given by: $x_{i+k} = \alpha^{(n)} + (\beta^{(n)} - \alpha^{(n)})(k/p + 1/2)$, $0 \le k \le p-1$ and $x_\ell = x_\ell^{(n)}$ otherwise. Since $x^{(n)}$ is optimal it follows that:

$$\sum_{k=0}^{p-1}\int_{C_{i+k}(x^{(n)})} |\xi - x_{i+k}^{(n)}|^r f(\xi)d\xi \le \sum_{k=0}^{p-1}\int_{C_{i+k}(x)} |\xi - x_{i+k}|^r f(\xi)d\xi,$$
$$\le \frac{M_n(i, i+p-1)}{2^r(r+1)p^r}(\beta^{(n)} - \alpha^{(n)})^{r+1}.$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Of course we also have:

$$\sum_{k=0}^{p-1} \frac{m_n(i, i+p-1)}{2^r(r+1)} \ \lambda_{i+k}^{r+1} \le \sum_{k=0}^{p-1} \int_{C_{i+k}(x^{(n)})} |\xi - x_{i+k}^{(n)}|^r f(\xi) d\xi. \tag{32}$$

Combining these two inequalities yields:

$$\frac{1}{p} \sum_{k=0}^{p-1} \lambda_{i+k}^{r+1} \le \frac{M_n(i, i+p-1)}{m_n(i, i+p-1)} \left(\frac{1}{p} \sum_{k=0}^{p-1} \lambda_{i+k} \right)^{r+1}.$$

Now, one considers a random variable X satisfying $\mathbb{P}(X = \lambda_{i+k}) = 1/p, 0 \le k \le p-1$. The above inequality reads

$$\frac{\mathbb{E} X^{r+1}}{(\mathbb{E} X)^{r+1}} \le \frac{M_n(i, i+p-1)}{m_n(i, i+p-1)}.$$

Applying Lemma 1 completes the proof. □

Proof of lemma 3. Let $[\alpha^{(n)}, \beta^{(n)}] = \bigcup_{k=0}^{p-1} C_{i+k}(x^{(n)})$, $[\alpha'^{(n)}, \beta'^{(n)}] = \bigcup_{k=0}^{q-1} C_{j-k}(x^{(n)})$, and let x be the "quantizer" defined by:

$$x_{i+k} := \alpha^{(n)} + (\beta^{(n)} - \alpha^{(n)})(k/q + 1/2), \quad 0 \le k \le q-1,$$
$$x_{j-k} := (\beta'^{(n)} - (\beta'^{(n)} - \alpha'^{(n)})(k/p + 1/2), \quad 0 \le k \le p-1$$
$$x_\ell := x_\ell^{(n)} \text{ otherwise.}$$

On one hand, one has

$$\sum_{k=0}^{p-1} \int_{C_{i+k}(x^{(n)})} |\xi - x_{i+k}^{(n)}|^r f(\xi) \, d\xi + \sum_{k=0}^{q-1} \int_{C_{j-k}(x^{(n)})} |\xi - x_{j-k}^{(n)}|^r f(\xi) d\xi$$
$$\le \frac{M_n(i, i+p-1)}{2^r(r+1)q^r} (\beta^{(n)} - \alpha^{(n)})^{r+1} + \frac{M_n(j-q+1, j)}{2^r(r+1)p^r} (\beta'^{(n)} - \alpha'^{(n)})^{r+1}.$$

On the other hand (32) and Jensen inequality yield

$$\sum_{k=0}^{p-1} \int_{C_{i+k}(x^{(n)})} |\xi - x_{i+k}^{(n)}|^r f(\xi) \, du$$
$$\ge \frac{m_n(i, i+p-1)}{2^r(r+1)} \sum_{k=0}^{p-1} \lambda_{i+k}^{r+1} \ge \frac{m_n(i, i+p-1)}{2^r(r+1)} p \left(\frac{1}{p} \sum_{k=0}^{p-1} \lambda_{i+k} \right)^{r+1}$$

and

$$\sum_{k=0}^{q-1}\int_{C_{j-k}(x^{(n)})}|\xi - x_{i+k}^{(n)}|^r f(\xi)du$$
$$\geq \frac{m_n(j,j-q+1)}{2^r(r+1)}q\left(\frac{1}{q}\sum_{k=0}^{q-1}\lambda_{j-k}\right)^{r+1}.$$

Gathering the last three inequalities finally leads to

$$\frac{p^{r+1}}{q^r}M_n(i,i+p-1)\left(\frac{1}{p}\sum_{k=0}^{p-1}\lambda_{i+k}\right)^{r+1}+\frac{q^{r+1}}{p^r}M_n(j,j-q+1)\left(\frac{1}{q}\sum_{k=0}^{q-1}\lambda_{j-k}\right)^{r+1}$$
$$\geq p\,m_n(i,i+p-1)\left(\frac{1}{p}\sum_{k=0}^{p-1}\lambda_{i+k}\right)^{r+1}+q\,m_n(j,j-q+1)\left(\frac{1}{q}\sum_{k=0}^{q-1}\lambda_{j-k}\right)^{r+1},$$

which is the announced result. □

5. APPLICATION TO GOOD-FIT TEST

In one dimension there are many recursive methods to compute stationary or – preferably – optimal quantizers. All of them are looking for a zero of the distortion gradient. As soon as the p.d.f. is numerically accessible, deterministic methods have no serious stochastic rival. There are essentially two kinds of methods, the first kind, which consists in searching a local minimum of the distortion using a dynamical system approach is the so-called Lloyd's Method I (see Kieffer, 1982), the second one, less specific is based on gradient descents. In one dimension, these methods are also the key tool to solve the "uniqueness problem" which says that when the p.d.f. is log-concave, then the distortion has exactly one stationary/optimal quantizer (see Trushkin, 1982 or Kieffer, 1982).

These deterministic methods admit some straightforward multi-dimensional extensions. However, these extensions are essentially intractable because they require at each step to compute several integrals with respect to the distribution μ itself. It is possible to overcome this difficulty when the distribution μ can be (easily) simulated: one may then devise some stochastic counterparts based either on standard Monte Carlo simulations or on recursive stochastic approximation algorithms (stochastic gradient descent). These procedures simply need to process massive computer simulation of μ-distributed numbers. Their rate of convergence is dimension-free: it behaves in all dimensions like $\sqrt{\mathit{Size\ of\ the\ simulated\ sample}}$. In fact, such stochastic procedures (like the CLVQ, see Pagès, 1997) turn out to be the basic tool in all numerical methods based on quantization (like multidimensional numerical integration in Pagès (1997) or conditional expectation computation in Bally et al. (2001)).

Whatever are the methods, when using quantization for numerical purpose one has to pay attention to several important requirements:

- A much greater accuracy is needed than that in its original purpose like coding or data compression.
- The stationarity of the quantizer becomes a crucial point, often more than the true optimality (see Pagès, 1997 or Bally et al., 2002): the equilibrium equation (8) makes some error terms vanish.
- The weights $\mu(C_i(x))$, $1 \le i \le n$, of a n-tuple x and its distortion have to be computed with the same accuracy as x itself.

This led us to look for some numerical test able to state if a n-tuple is close or not to a stationary r-quantizer. We will assume from now on that our r-quantizer x and its companion parameters have been obtained as the result of any numerical optimization. What follows can be implemented for any value of r although it has been tested for quadratic quantizers. At this stage the results of Theorem 2 suggest two possible tests according to what is known about the p.d.f. of μ.

The p.d.f. is known: If the p.d.f. is known, one may fit the graph $\varphi_n \equiv x_i \mapsto (n\widehat{\mu}_i)^{\frac{r+1}{r}}$ to that of $x_i \mapsto (c_{f,\frac{1}{r+1}})^{1+\frac{1}{r}} f(x_i)$ where $\widehat{\mu}_i$ is an estimate of $\mu(C_i(x))$. Of course, the normalizing constant $c_{f,\frac{1}{r+1}}$ is usually unknown. If f is computable one can think to numerically compute $c_{f,\frac{1}{r+1}}$ since it is an integral. If it is not the case one can simply choose the constant c that minimizes the quadratic error $\sum_i (\varphi_n(x_i) - cf(x_i))^2$. If the resulting error is small, this will mean that the quantizer x is numerically stationary and that $c^{\frac{r}{r+1}}$ is *a posteriori* a satisfactory proxy of $c_{f,\frac{1}{r+1}}$.

The p.d.f. is not known: In all usual procedures to compute a stationary/optimal quantizer, it is possible to store an approximation of the local distortion "*local distortion*(i)" at every component x_i of a quantizer x. This can be done without any significant further cost. Thus, it is possible to test – graphically or numerically – the flatness of the function defined by the ratio

$$x_i \longmapsto \frac{\textit{local distortion}(i)}{\textit{global distortion}}$$

on compact sets when the p.d.f. f does not vanish on the interior of its support: the flatter this function is, the more accurate the computed proxy is.

An alternative approach that we did not develop here could be as follows: one computes the Lebesgue measure $\widehat{\lambda}_i \approx \lambda(C_i(x))$ of the Voronoi cells processing *a posteriori* specific Monte Carlo simulation; then one approximates the (unknown) p.d.f. f at x_i by setting $f(x_i) \approx \widehat{f}_i := \widehat{\mu}_i/\widehat{\lambda}_i$. Finally one implements the same procedure as above using $\widehat{f}_i$.

Here we present three examples in 1-dimension: a compactly supported and positive $p.d.f$, the Gaussian and the Arcsine distributions. As a "counterexample" we inspect the case of a p.d.f. vanishing at one point. We compute the quantizers by a Newton–Raphson algorithm involving the gradient and the hessian of the (quadratic) distortion (see Fort and Pagès (1995) for some closed form for the Hessian). It took less that 11 iterations. In the power case ("counterexample") the Hessian was too unstable so we iterate many times a regular gradient descent with variable step sizes. The companion parameters (local and global distortions, μ-mass of the cells) have been computed using analytic formulae.

Our purpose here is to check whether the uniformity of the local distortion is a good criterion for the optimality of a quantizer (size $n = 100$) or not.

In each figure below we show the quantizer values (o), the theoretical value of the p.d.f. at the quantizer values and the local distortion (suitably scaled to appear on the graph).

It turns out that for compactly supported non vanishing p.d.f. (possibly unbounded) the local distortion of the computed quantizer (size 100) does look almost uniform (Fig. 1 p.d.f. is $f(\xi) = 6/7(1+\xi-\xi^2)$, Fig. 2 p.d.f. is $f(\xi) = \text{Arcsine}(\xi)$). In the non-compactly supported case (Fig. 3 Gaussian p.d.f.), the local distortion is still uniform excepted near the edges. What happens for a p.d.f. vanishing at a given point x_0 (Fig. 4 power p.d.f.) is more expected since all the arguments in the proofs fall down in that case.

Furthermore, we verified through Fig. 5 and 6 that the μ-mass of the cell to the power $\frac{3}{2}$, suitably normalized, is a good estimator of the p.d.f. (we did not superimpose the two curves to keep them readable). As a by-product of Theorem 4, we estimated $c_{f,\frac{1}{3}}$ for the Gaussian p.d.f. by 3.164 which is quite accurate when compared to the true value 3.196.

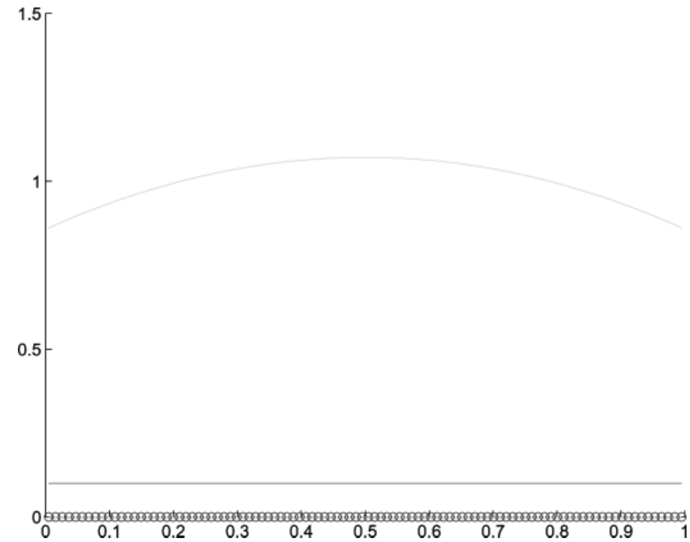

Figure 1. Compactly supported positive distribution: the quantizers (o), the p.d.f. and the local distortion. (*View this art in color in www.dekker.com.*)

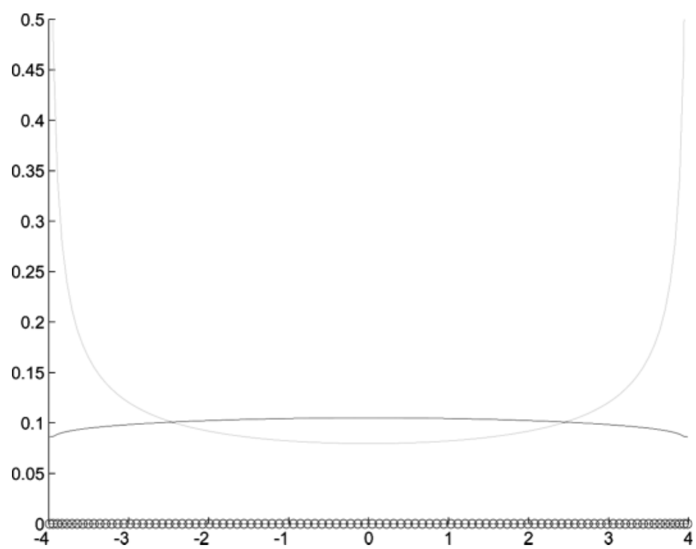

Figure 2. Arcsine: the quantizers (o), the p.d.f. and the local distortion. (*View this art in color in www.dekker.com.*)

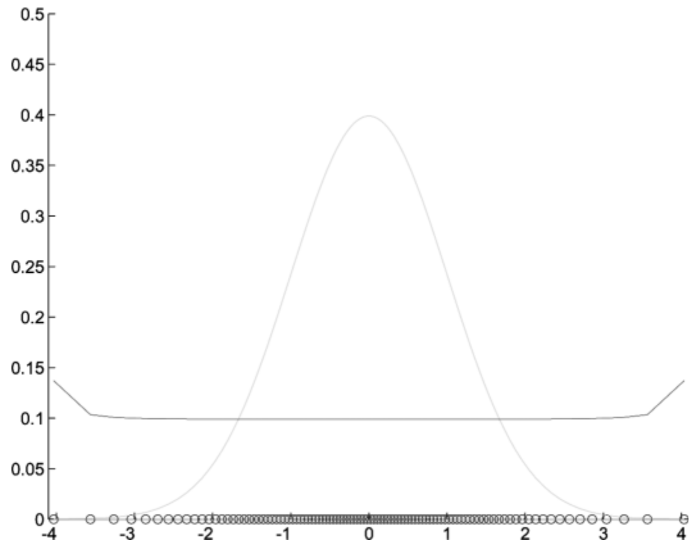

Figure 3. Gaussian distribution: the quantizers (*o*), the p.d.f. and the local distortion. (*View this art in color in www.dekker.com.*)

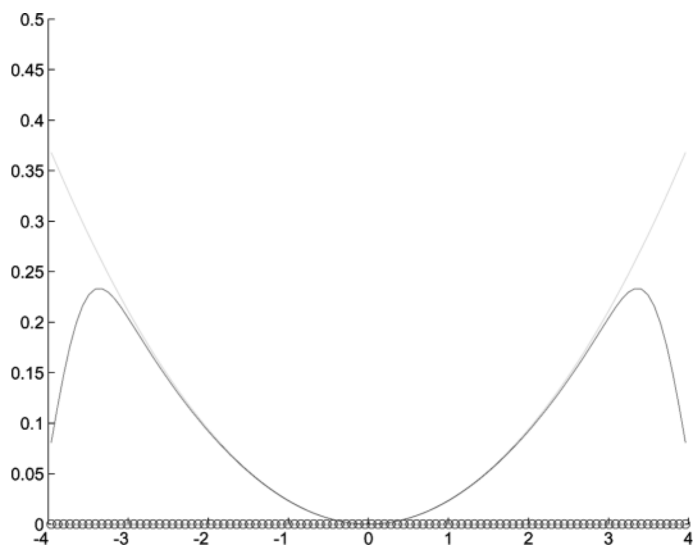

Figure 4. Power distribution: the quantizers (*o*), the p.d.f. and the local distortion. (*View this art in color in www.dekker.com.*)

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

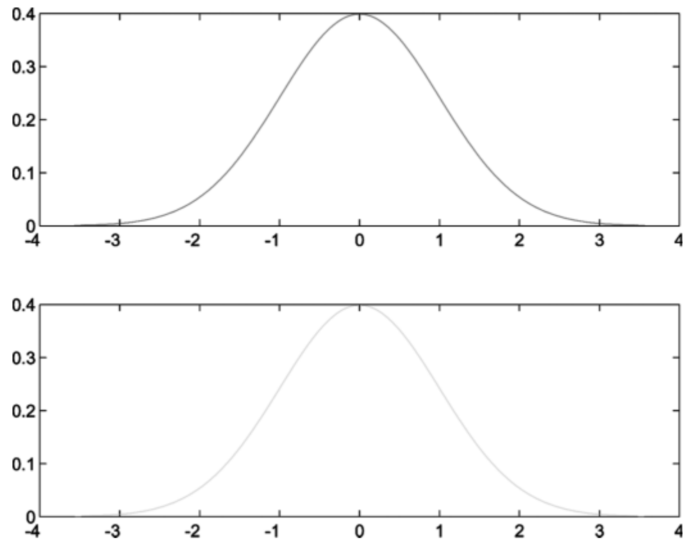

Figure 5. The Gaussian p.d.f. (bottom) and its estimation via the μ-mass of the cells (top). (*View this art in color in www.dekker.com.*)

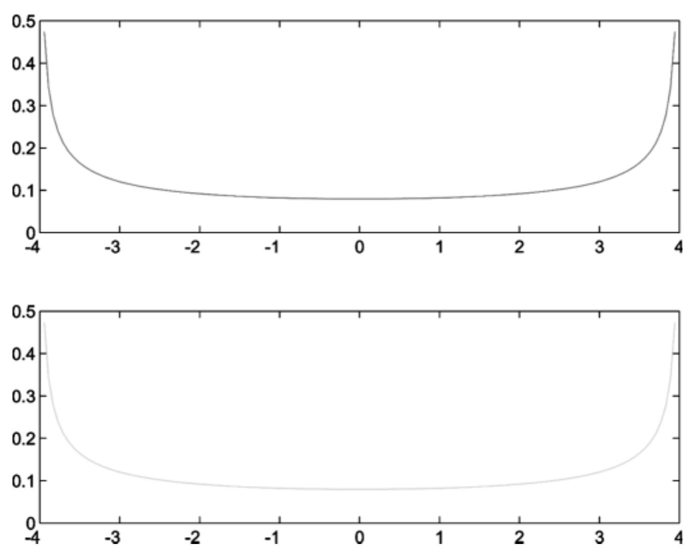

Figure 6. The Arcsine p.d.f. (bottom) and its estimation via the μ-mass of the cells (top). (*View this art in color in www.dekker.com.*)

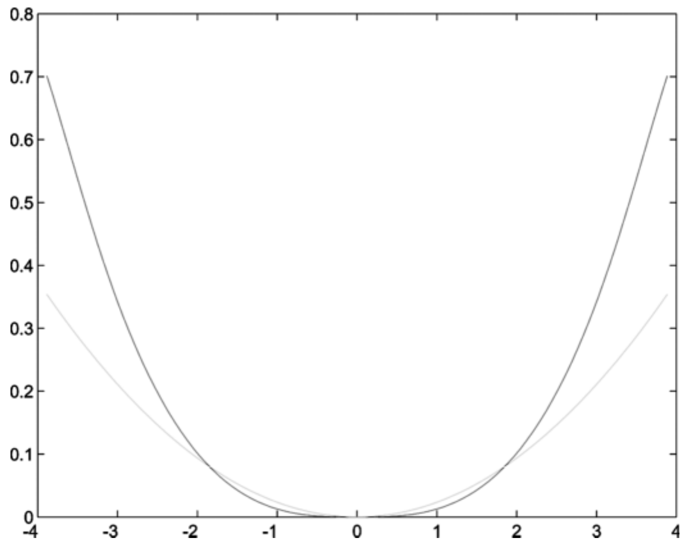

Figure 7. The power p.d.f. and its estimation (bold) via the μ-mass of the cells. (*View this art in color in www.dekker.com.*)

Again what happens for a p.d.f. vanishing at a given point x_0 is clearly different: the estimated density via the Theorem 4 is far from the true one (Fig. 7 power p.d.f. and its estimate $(c_{f,\frac{1}{3}} n\,\mu(C_i(x^{(n)}))^{\frac{2}{3}})$.

6. CONCLUSION

This paper is a first 1-dimensional result to prove that optimal or asymptotically optimal quantizers generate an asymptotically uniform local distortion. One may reasonably guess that the converse is true, which would make the uniformity of the local distortion a good optimality criterion for a quantizer. This is important for numerical applications of quantization (see Pagès and Printems, 2003) which require some very accurate estimates of the optimal or at least locally optimal quantizers.

From a technical viewpoint it seems that our methods of proof do not extend to higher dimensions: a general multidimensional method still needs to be found.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

APPENDIX: DIFFERENTIABILITY OF THE r-DISTORTION, $0 < r < 1$

Let $x \in (\mathbb{R}^d)^n$ and let $h_i := (0, \ldots, 0, h, 0 \ldots, 0)$ (the non zero component is the ith). Set, for every $\xi \neq x_i$,

$$R_i(x, h, \xi) := \min_{j \neq i}(|x_j - \xi|^r \wedge |x_i + h_i - \xi|^r) - \min_j |x_j - \xi|^r - r|x_i - \xi|^{r-1}\left(\frac{x_i - \xi}{|x_i - \xi|} \mid h\right)\mathbf{1}_{C_i(x)}(\xi)$$

where $(\xi \mid \xi')$ denotes the canonical inner product on $\mathbb{R}^d$. One checks as for the standard case $r \geq 1$ that, for every $\xi \in \overset{\circ}{C}_i(x)\backslash\{x_i\}$, hence $\mu(d\xi)$-*a.s.*,

$$\lim_{h \to 0} \frac{R_i(x, h, \xi)}{|h|} = 0.$$

On the other hand, using the two standard inequalities $|\min_j a_j - \min_j b_j| \leq \max_j |a_i - b_j|$ and $|u^r - v^r| \leq |u - v|(u^{r-1} + v^{r-1})$, one derives that, for every $\xi \neq x_i$,

$$\begin{aligned}|R_i(x, h, u)| &\leq |h|\left(\min_{j \neq i}(|x_j - \xi| \wedge |x_i + h - \xi|)^{r-1} + \min_j |x_j - \xi|^{r-1}\right)\\ &\leq |h|\Big(|x_i + h - \xi|^{r-1} + |x_i - \xi|^{r-1}\Big).\end{aligned}$$

Let $\rho := \frac{c}{1-r} > 1$. One has

$$\left(\frac{|R_i(x, h, u)|}{|h|}\right)^\rho \leq 2^{\rho - 1}(|x_i + h - \xi|^{-c} + |x_i - \xi|^{-c}).$$

$$\sup_{|h| \leq 1} \int \left(\frac{|R_i(x, h, \xi)|}{|h|}\right)^\rho \mu(d\xi) \leq 2^{\rho - 1} \sup_{u \in B(x_i, 1)} \int \frac{\mu(d\xi)}{|u - \xi|^c} < +\infty.$$

One derives the existence of $\partial D_n^{\mu,r}/\partial x_i(x)$ using the uniform integrability convergence theorem. The continuity of these partial derivatives follows by similar arguments and this completes the proof. □

REFERENCES

Abaya, E., Wise, G. (1992). On the existence of optimal quantizer. *IEEE Trans. Inf. Theory* 38(2):937–946.

Bally, V., Pagès, G. (2003). A quantization algorithm for solving discrete time multidimensional optimal stopping problems. *Bernoulli* 9(6):1003–1049.

Bally, V., Pagès, G., Printems, J. (2001). A stochastic quantization method for nonlinear problems. *Monte Carlo Methods Appl.* 7(1):21–34.

Bally, V., Pagès, G, Printems, J. (2002). A quantization tree method for pricing and hedging multidimensional American options. Pre-print LPMA-753, Univ. Paris 6, to appear in *Mathematical Finance*.

Benett, N. R. (1948). Spectra of quantized signals. *Bell Syst. Tech. J.* 27:446–472.

Bucklew, J., Wise, G. (1982). Multidimensional asymptotic quantization theory with *rth* power distortion measures. *IEEE Trans. Inf. Theory, Special issue on Quantization* 28(2):239–247.

Cohort, P. (1997). A Geometrical Approach to Uniqueness of a Locally Optimal Quantizer. Technical Report, LPMA, Univ. Paris 6 (France).

Fort, J. C., Pagès, G. (1995). On the *a.s.* convergence of the Kohonen algorithm with a general neighborhood function. *Ann. Appl. Prob.* 5(4):1177–1216.

Fort, J. C., Pagès, G. (2002). Asymptotics of optimal quantizers for some scalar distributions. *J. Comput. Appl. Math.* 146:253–275.

Gersho, A. (1979). Asymptotically optimal block quantization. *IEEE Trans. Inf. Theory* 25(4):373–380.

Gersho, A., Gray, R. eds. (1982). *IEEE Trans. Inf. Theory, Special Issue on Quantization* 28.

Gersho, A., Gray, R. (1992). *Vector Quantization and Signal Compression*. Boston: MA Kluwer, 732 pp.

Graf, S., Luschgy, H. (2000). *Foundations of Quantization for Probability Distributions*. Lecture Notes In Mathematics, 1730, Berlin: Springer Verlag, 230 pp.

Kieffer, J. (1982). Exponential rate of convergence for the Lloyd's method I. *IEEE Trans. Inf. Theory Special Issue Quantization* 28(2):205–210.

Lamberton, D., Pagès, G. (1996). On the critical points of the 1-dimensional competitive learning vector quantization algorithm.

In: De Facto editor. Proceedings of the ESANN'96, Brussels, Belgium.

McClure, D. E. (1975). Nonlinear segmented function approximation and analysis of line patterns. *Quart. Appl. Math.* 33:1–37.

Pagès, G. (1997). A space vector quantization method for numerical integration. *J. Comput. Appl. Math.* 89:1–38.

Pagès, G., Printems, J. (2003). Optimal quadratic quantization for numerics: The Gaussian case. *Monte Carlo Methods Appl.* 9(2):135–166.

Pagès, G., Pham, H., Printems, J. (2003). Optimal quantization methods and applications to numerical problems in finance. Pre-print LPMA-813, Univ. Paris 6/7 (France), to appear in *Handbook of Numerical Methods in Finance*, Birkhauser.

Patane, G., Russo, M. (2001). The enhanced LBG algorithm. *Neural Networks* 14(9).

Shirayev, A. N. (1984). *Probability, Graduate Text in Mathematics*. New-York: Springer, 577 pp.

Trushkin, A. (1982). Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions. *IEEE Trans. Inf. Theory, Special Issue on Quantization* 28(2): 187–198.

Zador, P. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inf. Theory, Special Issue on Quantization* 28(2):139–148.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016