*The Annals of Applied Probability*
2004, Vol. 14, No. 1, 118–143


# LIMIT THEOREMS FOR RANDOM NORMALIZED DISTORTION

BY PIERRE COHORT

*INRIA–CERMICS*

We present some convergence results about the distortion $\mathcal{D}^{\nu}_{\mu,n,r}$ related to the Voronoï vector quantization of a $\mu$-distributed random variable using $n$ i.i.d. $\nu$-distributed codes. A weak law of large numbers for $n^{r/d}\mathcal{D}^{\nu}_{\mu,n,r}$ is derived essentially under a $\mu$-integrability condition on a negative power of a $\delta$-lower Radon–Nikodym derivative of $\nu$. Assuming in addition that the probability measure $\mu$ has a bounded $\varepsilon$-potential, we obtain a strong law of large numbers for $n^{r/d}\mathcal{D}^{\nu}_{\mu,n,r}$. In particular, we show that the random distortion and the optimal distortion vanish almost surely at the same rate. In the one-dimensional setting ($d = 1$), we derive a central limit theorem for $n^r\mathcal{D}^{\nu}_{\mu,n,r}$. The related limiting variance is explicitly computed.

**Introduction.** Quantization is a classical discretization procedure introduced in 1948 to solve some signal processing problems. This procedure consists of approximating an $\mathbb{R}^d$-valued random variable $X$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ by a finite range random variable $f_n(X) : (\Omega, \mathcal{A}) \to \{y_1, \dots, y_n\}$ where $\{y_1, \dots, y_n\} \subset \mathbb{R}^d$ and where $f_n : \mathbb{R}^d \to \{y_1, \dots, y_n\}$ is a Borel map. The above approximation induces a discretization error usually modeled by $\mathbb{E}(\|X - f_n(X)\|^r)$ for some $r > 0$, provided that $X \in L^r(\Omega, \mathcal{A}, \mathbb{P})$. The preceding mean error is called the *distortion* and $f_n$ is called an *$n$-quantizer*. The set of $n$-quantizers will be denoted $\mathcal{Q}_n$. Using the quantization procedure requires knowledge of a good quantizer along with the related distortion. So it is useful to estimate (at least for large $n$) the optimal distortion

$$\inf_{f_n \in \mathcal{Q}_n} \mathbb{E}\big(\|X - f_n(X)\|^r\big) \tag{1}$$

and to study the $n$-quantizers inducing a distortion close to (1). In this setting, we can confine our attention to the set $\mathcal{V}_n$ of the quantizers taking values according to the nearest neighbor rule: $f_n(x) := y_{i(x)}$ where $i(x)$ is the smallest index satisfying $\|x - y_{i(x)}\| = \min_{1 \le j \le n} \|x - y_j\|$. Indeed, one obtains easily that

$$\inf_{f_n \in \mathcal{V}_n} \mathbb{E}\big(\|X - f_n(X)\|^r\big) = \inf_{f_n \in \mathcal{Q}_n} \mathbb{E}\big(\|X - f_n(X)\|^r\big).$$

The quantizers that belong to $\mathcal{V}_n$ are called *Voronoï $n$-quantizers*. The name Voronoï comes from the fact that the closures of the constancy sets of $f_n$ are the so-called Voronoï cells of $\{y_1, \dots, y_n\}$, defined by the polyhedras,

$$C_i(y_1, \dots, y_n) := \Big\{z \in \mathbb{R}^d;\ \|x - y_i\| = \min_{1 \le j \le n} \|x - y_j\|\Big\}, \qquad 1 \le i \le n. \tag{2}$$

Received May 2000; revised December 2001.
*AMS 2000 subject classifications.* Primary 60F25, 60F15, 60F05; secondary 94A29.
*Key words and phrases.* Quantization, distortion, law of large numbers, central limit theorem.

When restricted to $\mathcal{V}_n$, the distortion writes

$$\mathbb{E}(\|X - f_n(X)\|^r) = \int_{\mathbb{R}^d} \min_{1 \le i \le n} \|u - y_j\|^r \mu(du),$$

where $\mu$ is the law of $X$ and where $\{y_1, \ldots, y_n\} = f_n(\mathbb{R}^d)$. So the distortion is usually considered as a function on $(\mathbb{R}^d)^n$ and the set $\mathcal{V}_n$ is usually identified to $(\mathbb{R}^d)^n$ [at $f_n \in \mathcal{V}_n$ then $n!$ $n$-tuples correspond in $(\mathbb{R}^d)^n$]. The distortion will be denoted $\mathcal{D}_{\mu,n,r}(y_1, \ldots, y_n)$. The coordinates of a Voronoï quantizer $(y_1, \ldots, y_n)$ are called the codes.

This paper is devoted to the study of the asymptotics (as $n \to +\infty$) of the distortion when the codes are random, drawn from an i.i.d. sequence $Y_1, \ldots, Y_n$. There is indeed a lack of probabilistic results about the related random distortion since only the mean $\mathbb{E}(\mathcal{D}_{\mu,n,r}(Y_1, \ldots, Y_n))$ has been investigated in the past by Zador [21], who obtained the convergence of $n^{r/d}\mathbb{E}(\mathcal{D}_{\mu,n,r}(Y_1, \ldots, Y_n))$ (see also [10]).

Our goal is then to get more information about the asymptotic behavior of $\mathcal{D}_{\mu,n,r}(Y_1, \ldots, Y_n)$, in particular to see more precisely how the random distortion differs from the optimal one (1). To this end, we will show in a quite general setting the following new results.

*Laws of large numbers* (*Theorems* 1 *and* 2).

$$n^{r/d}\mathcal{D}_{\mu,n,r}(Y_1, \ldots, Y_n) \xrightarrow{L^p} l,$$

$$n^{r/d}\mathcal{D}_{\mu,n,r}(Y_1, \ldots, Y_n) \xrightarrow{a.s.\ and\ L^2} l,$$

where $l$ (defined in Theorem 1) is deterministic, depending on $r$, $d$, $\mu$ and law$(Y_1)$.

*Central limit theorem* (*for* $d = 1$) (*Theorem* 3).

$$n^{1/2}\big(n^r \mathcal{D}_{\mu,n,r}(Y_1, \ldots, Y_n) - l\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{f,g}^2),$$

where $\sigma_{f,g}$ is defined in Theorem 3.

Before stating these results, it will be helpful to give some background on quantization theory, including the basics of optimal quantization.

The optimizing problem related to $\mathcal{D}_{\mu,n,r}$ is known as the *optimal quantization* problem of the probability measure $\mu$. This question arises in various fields of applied mathematics (information theory, statistical clustering, stochastic algorithm theory, etc.) and has been extensively investigated during the past fifty years. Recently, Graf and Luschgy [10] have completed a comprehensive book containing a rigorous mathematical treatment of the classical theory along with some investigation on new topics such as optimal quantization for continuous singular probability measures. The survey by Gray and Neuhoff [11] provides a detailed account of the information-theoretic aspects of quantization (coding problems and links with Shannon's theory) along with a historical review. The deterministic optimizing algorithms of $\mathcal{D}_{\mu,n,r}$ are described in [11] and the stochastic optimizing algorithms are investigated in [1, 2, 7, 16]. Let us recall the following basic facts.

*Existence of an optimal quantizer.* For every probability measure $\mu$ on $\mathbb{R}^d$ and every $r > 0$, there exists a quantizer $(y_1^*, \ldots, y_n^*) \in \mathrm{conv}(\mathrm{supp}(\mu))^n$ such that

$$\mathcal{D}_{\mu,n,r}(y_1^*, \ldots, y_n^*) = \inf_{(y_1,\ldots,y_n)\in(\mathbb{R}^d)^n} \mathcal{D}_{\mu,n,r}(y_1, \ldots, y_n) =: \mathcal{D}^*_{\mu,n,r}.$$

Such a quantizer is called an optimal quantizer and $\mathcal{D}^*_{\mu,n,r}$ is called the optimal distortion.

*Asymptotics of the optimal distortion.* If there exists $\varepsilon > 0$ such that $\int_{\mathbb{R}^d} \|u\|^{r+\varepsilon} \mu(du) < +\infty$. Then (Bucklew and Wise's theorem [5]; see [10] for a correct proof)

$$n^{r/d} \mathcal{D}^*_{\mu,n,r} \to J_{r,d} \|f\|_{d/(d+r)},$$

where $f$ is the density function of the absolutely continuous part of $\mu$, where $\|f\|_{d/(d+r)} = (\int f^{d/(d+r)})^{(d+r)/d}$ and where $J_{r,d}$ is some constant depending only on $r$ and $d$ (see Gersho's conjecture [8] for a geometrical interpretation of $J_{r,d}$). Note that from the moment assumption and Hölder's inequality, one has $\|f\|_{d/(d+r)} < +\infty$.

*Asymptotic structure of optimal quantizers.* Some features of the asymptotic structure of optimal quantizers can be derived or reasonably conjectured. For instance, Graf and Luschgy [10] have shown that the codes of an optimal sequence $(y_1^*, \ldots, y_n^*)_{n\geq 1}$ fulfill ($\stackrel{d}{\to}$ denotes narrow convergence)

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{y_i^*} \stackrel{d}{\to} \frac{f^{d/(d+r)}}{\int f^{d/(d+r)}}\lambda \tag{3}$$

as soon as $\mu$ is absolutely continuous and fulfills $\int \|u\|^{r+\varepsilon}\mu(du) < +\infty$. In [8], Gersho conjectured the following asymptotic geometrical regularity of Voronoï cells $C_i(y_1^*, \ldots, y_n^*)$ [see (2)]: As $n \to +\infty$, the cell $C_i(y_1^*, \ldots, y_n^*)$ becomes congruent to a polyhedron $P^*$ satisfying

$$\frac{1}{\lambda(P^*)^{1+r/d}} \int_{P^*} \|u - c(P^*)\|^r \, du = \inf_{P\in\mathcal{P}} \frac{1}{\lambda(P)^{1+r/d}} \int_P \|u - c(P)\|^r \, du, \tag{4}$$

where $\mathcal{P}$ is the set of polyhedras generating a tesselation $\{P_i\}_{i\geq 1}$ which is Voronoï with respect to the $r$-inertia centroids of the $P_i$'s. Moreover, the constant $J_{r,d}$ in the Bucklew and Wise theorem equals the right-hand term in (4).

*Optimization algorithms.* The optimal quantizers are in general unknown. So one has to use some deterministic [14, 19, 20] or stochastic algorithms [16] that make a local optimization of $\mathcal{D}_{\mu,n,r}$ and provide numerical approximations of some locally optimal quantizers. For instance, the CLVQ algorithm reads as follows: assume that $\operatorname{supp}(\mu)$ is a convex set. Let $(y_1^0, \dots, y_n^0) \in (\mathbb{R}^d)^n$, let $(Y_k)_{k\geq 1}$ be a $\mu$-distributed i.i.d. sequence and let $(\varepsilon_k)_{k\geq 1}$ be a real sequence in $(0, 1)$. Define then $(y^k \in (\mathbb{R}^d)^n)_{k\geq 0}$ by

$$(5)\qquad y_i^{k+1} := \begin{cases} y_i^k - \varepsilon_{k+1}(y_i^k - Y_{k+1}), & \text{if } Y_{k+1} \in C_i(y^k), \\ y_i^k, & \text{otherwise}, \end{cases} \qquad 1 \leq i \leq n,$$

where $C_i(y^k)$ is the $i$th Voronoï cell of $y^k$. If $\varepsilon_k \equiv \varepsilon$, the sequence (5) is known as the constant gain CLVQ algorithm. It is, under some assumptions on $\mu$, a Doeblin recurrent Markov chain whose invariant probability measure "concentrates" on the set $\{\nabla \mathcal{D}_{\mu,n,2} = 0\}$ as $\varepsilon \to 0$ (see [3]). If $\varepsilon_k \downarrow 0$, the sequence (5) is known as a decreasing gain CLVQ algorithm and, under some assumptions on $\mu$ and on $(\varepsilon_k)_{k\geq 1}$, is almost surely converging toward an element of $\operatorname{Argminloc}(\mathcal{D}_{\mu,n,2})$ (see [7]).

In practice, some problems arise when using the above results. For instance, even if the optimization algorithm does not get trapped into some suboptimal local optima, the resulting distortion is difficult to estimate from Bucklew and Wise's theorem even for large $n$ since the constant $J_{r,d}$ is in general unknown as soon as $d > 1$. So in order to give some upper bounds on $J_{r,d}$, numerous sequences $(y_1^n, \dots, y_n^n)_{n\geq 1}$ of deterministic suboptimal quantizers have been investigated (for instance, the *lattice* quantizers, [10, 11]). From some geometrical considerations, these sequences are shown to provide the convergence

$$n^{r/d} \mathcal{D}_{\lambda_{[0,1]^d},n,r}(y_1^n, \dots, y_n^n) \to W_{r,d},$$

where $W_{r,d}$ is a known constant depending on $(y_1^n, \dots, y_n^n)_{n\geq 1}$ and then yields the upper bound $J_{r,d} \leq W_{r,d}$ (the introduction of such suboptimal sequences also originated in the fact that they allow the reduction of coding and algorithmic complexity problems in some applications; see [11], Section II.E, page 2338 and Section V, page 2361).

Another kind of quantizer has been introduced: a random quantizer (see [10, 21]). Let $\nu$ be an absolutely continuous probability measure on $\mathbb{R}^d$ such that $\operatorname{supp}(\nu) \supset \operatorname{supp}(\mu)$ and let $(Y_i)_{i\geq 1}$ be an i.i.d. sample from $\nu$. One can consider a sample path $(Y_1(\omega), \dots, Y_n(\omega))$ as a Voronoï quantizer. At this time, random quantizers have been used only to get an upper bound on $J_{r,d}$ and only the mean distortion $\mathbb{E}(\mathcal{D}_{\mu,n,r}(Y_1, \dots, Y_n))$ has been studied. In [21], Zador showed (unfortunately under quite unrealistic assumptions) that

$$(6)\qquad \mathbb{E}\big(n^{r/d} \mathcal{D}_{\mu,n,r}(Y_1, \dots, Y_n)\big) \to B_d^{-r/d} \Gamma\left(1 + \frac{r}{d}\right) \int_{\mathbb{R}^d} \frac{\mu(du)}{g(u)^{r/d}},$$

where $B_d$ is the volume of the $d$-dimensional unit ball and $g$ is the density function of $\nu$. Using the inequality $\mathcal{D}^*_{\mu,n,r} \le \mathbb{E}(\mathcal{D}_{\mu,n,r}(Y_1,\dots,Y_n))$ and letting $\mu = \nu := \lambda_{[0,1]}$, the author obtained the bound $J_{r,d} \le B_d^{-r/d}\Gamma(1+r/d)$. More recently, Graf and Luschgy [10] derived rigorously convergence (6) when $\nu$ is the uniform distribution on some compact set of $\mathbb{R}^d$ satisfying a uniform repartition mass principle (e.g., convex).

Here we note that in some situations it may be interesting to have a better knowledge of the random distortion $\mathcal{D}_{\mu,n,r}(Y_1,\dots,Y_n)$. In particular, we have in mind the initializing problem of the competitive learning vector quantization algorithm (5). A good initialization of (5) requires at least that $(y_1^0,\dots,y_n^0) \in \mathrm{supp}(\mu)^n$. But, the probability measure $\mu$ [and in particular $\mathrm{supp}(\mu)$] is in general known only through the data $(Y_k)_{k\ge1}$. So an a priori choice of $(y_1^0,\dots,y_n^0)$, not depending on the data, could lead to $y_i^0 \notin \mathrm{supp}(\mu)$ for some $i$. Such $y_i^0$ could then be frozen since an excessively large $d(y_i^0, \mathrm{supp}(\mu))$ implies $Y_{k+1} \notin C_i(y_1^k,\dots,y_n^k)$ for every $k$ and then $y_i^k = y_i^0$ for every $k$.

Subsequently, it has been natural to set $(y_1^0,\dots,y_n^0) = (Y_1,\dots,Y_n)$ and then run the algorithm with the remaining data. The above random self-initialization ensures that $(y_1^0,\dots,y_n^0) \in \mathrm{supp}(\mu)^n$, but one can ask if it provides a good initial distortion $\mathcal{D}_{\mu,n,2}(Y_1,\dots,Y_n)$. To answer this question, the only knowledge of $\mathbb{E}(\mathcal{D}_{\mu,n,2}(Y_1,\dots,Y_n))$ is not very satisfactory and we need some additional features of the (specifically pathwise) behavior of $\mathcal{D}_{\mu,n,2}(Y_1,\dots,Y_n)$.

Returning to the general case, the fact that $\mathrm{supp}(\nu) \supset \mathrm{supp}(\mu)$ is equivalent to $\mathcal{D}_{\mu,n,r}(Y_1,\dots,Y_n) \overset{\text{a.s.}}{\to} 0$. The goal of this paper is to investigate the $L^p$ and the almost sure rate in the preceding convergence and, in the case $d=1$, to derive a central limit theorem for $n^r\mathcal{D}_{\mu,n,r}(Y_1,\dots,Y_n)$. Using a method based on the estimation of the integral moments and of the variance of the random distortion, we derive an $L^p$ and a strong law of large numbers for $n^{r/d}\mathcal{D}_{\mu,n,r}(Y_1,\dots,Y_n)$ (Section 1) which answer our question in a quite general setting and yield the result that the random and the optimal distortion vanish almost surely at the same rate (with different constants). When $d=1$, the CLT is derived from an extension of the Pyke–Hall method for spacing statistics (see [12]). In Sections 2 and 3, we give some comments on these results. The proofs are derived in Section 4.

**1. Results.** For notational convenience, we set $\mathcal{D}^{\nu}_{\mu,n,r} := \mathcal{D}_{\mu,n,r}(Y_1,\dots,Y_n)$ where $\nu$ is the common law of the i.i.d. random variables $Y_1,\dots,Y_n$. The random distortion normalized at the optimal rate $n^{r/d}$ will be called the random normalized distortion. The main object used to derive our results is the function

$$\mathbb{R}^d \times \mathbb{R}_+^* \to \mathbb{R}_+,$$

$$(u,\delta) \mapsto g_\delta(u) := \inf_{s\in(0,\delta]} \frac{\nu(B(u,s))}{\lambda(B(u,s))},$$

where $\lambda$ is the Lebesgue measure on $\mathbb{R}^d$ and where $B(u,s)$ is the open ball with center $u \in \mathbb{R}^d$ and radius $s \geq 0$.

For a fixed $\delta > 0$, we will call $u \mapsto g_\delta(u)$ the $\delta$-lower Radon–Nikodym derivative of the probability measure $\nu$. This terminology is justified by the fact that if $\delta$ is small, one can hope to have $g_\delta \simeq d\nu/d\lambda$ under some mild regularity assumption on $g := d\nu/d\lambda$.

In the following sections, $B_d$ denotes $\lambda(B(0,1))$ and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$.

1.1. *$L^p$ convergence.* Our first result is an $L^p$ law of large numbers for the random normalized distortion.

THEOREM 1. *Let $d \in \mathbb{N}^*$, $p \in \mathbb{N} \setminus \{0,1\}$ and $r \in \mathbb{R}_+^*$. Let $\mu$ and $\nu$ be some absolutely continuous probability measures on $\mathbb{R}^d$ with* $\operatorname{supp}(\mu) \subset \operatorname{supp}(\nu)$. *Let $g$ be the density function of $\nu$. Assume that*:

T1.1. *There exists $\rho > 0$ such that* $\int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\|u\|+\rho}(u)^{pr/d}} < +\infty$.

T1.2. $\int_{\mathbb{R}^d} \|u\|^{pr}(\mu+\nu)(du) < +\infty$.

*Then*

$$n^{r/d}\mathcal{D}_{\mu,n,r}^{\nu} \xrightarrow{L^p} B_d^{-r/d}\Gamma\left(1+\frac{r}{d}\right)\int_{\mathbb{R}^d}\frac{\mu(du)}{g(u)^{r/d}}.$$

If the quantized probability measure $\mu$ has a greater moment order than the quantizing probability measure $\nu$, we can derive the $L^p$ convergence under a more tractable version of assumption T1.1.

COROLLARY 1. *Let $d \in \mathbb{N}^*$, $p \in \mathbb{N} \setminus \{0,1\}$ and $r \in \mathbb{R}_+^*$. Let $\mu$ and $\nu$ be some absolutely continuous probability measures on $\mathbb{R}^d$ with* $\operatorname{supp}(\mu) \subset \operatorname{supp}(\nu)$. *Let $g$ be the density function of $\nu$. Assume that there exists some $\eta \in (0,1)$ such that*:

C1.1. *There exists $\delta > 0$ such that* $\int_{\mathbb{R}^d} \frac{\mu(du)}{g_\delta(u)^{pr/(d(1-\eta))}} < +\infty$.

C1.2. $\int_{\mathbb{R}^d} \|u\|^{pr/\eta}\mu(du) < +\infty$; $\int_{\mathbb{R}^d} \|u\|^{pr}\nu(du) < +\infty$.

*Then assumptions* T1.1 *and* T1.2 *hold*; *in particular,*

$$n^{r/d}\mathcal{D}_{\mu,n,r}^{\nu} \xrightarrow{L^p} B_d^{-r/d}\Gamma\left(1+\frac{r}{d}\right)\int_{\mathbb{R}^d}\frac{\mu(du)}{g(u)^{r/d}}.$$

1.2. *Almost sure convergence.* Here, we provide a strong law of large numbers for the random normalized distortion. The result is obtained under a mild strengthening of the assumptions ensuring the $L^2$ convergence along with the existence of a bounded $\varepsilon$-potential for the quantized probability measure.

THEOREM 2. *Let* $d \in \mathbb{N}^*$ *and* $r \in \mathbb{R}_+^*$. *Let* $\mu$ *and* $\nu$ *be some absolutely continuous probability measures on* $\mathbb{R}^d$ *with* $\mathrm{supp}(\mu) \subset \mathrm{supp}(\nu)$. *Let* $g$ *be the density function of* $\nu$. *Assume that there exists* $\varepsilon > 0$ *such that*:

T2.1. *The function* $c \mapsto \int_{\mathbb{R}^d} \frac{\mu(du)}{\|u-c\|^\varepsilon}$ *is bounded on* $\mathbb{R}^d$.
T2.2. *There exists* $\rho > 0$ *such that* $\int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\|u\|+\rho}(u)^{(2r+\varepsilon)/d}} < +\infty$.
T2.3. $\int_{\mathbb{R}^d} \|u\|^{2r+\varepsilon}(\mu+\nu)(du) < +\infty$.

*Then*

$$n^{r/d}\mathcal{D}_{\mu,n,r}^{\nu} \xrightarrow{a.s.\text{ and }L^2} B_d^{-r/d}\Gamma\left(1+\frac{r}{d}\right)\int_{\mathbb{R}^d}\frac{\mu(du)}{g(u)^{r/d}}.$$

As for the $L^p$ law, if the quantized probability measure $\mu$ has a greater moment order than the quantizing probability measure $\nu$, we can improve T2.2.

COROLLARY 2. *Let* $d \in \mathbb{N}^*$ *and* $r \in \mathbb{R}_+^*$. *Let* $\mu$ *and* $\nu$ *be some absolutely continuous probability measures on* $\mathbb{R}^d$. *Let* $g$ *be the density function of* $\nu$. *Assume that there exist* $\eta \in (0, 1)$ *and* $\varepsilon > 0$ *such that*:

C2.1. *The function* $c \mapsto \int_{\mathbb{R}^d} \frac{\mu(du)}{\|u-c\|^\varepsilon}$ *is bounded on* $\mathbb{R}^d$.
C2.2. *There exists* $\delta > 0$ *such that* $\int_{\mathbb{R}^d} \frac{\mu(du)}{g_\delta(u)^{(2r+\varepsilon)/(d(1-\eta))}} < +\infty$.
C2.3. $\int_{\mathbb{R}^d} \|u\|^{(2r+\varepsilon)/\eta}\mu(du) < +\infty$; $\int_{\mathbb{R}^d} \|u\|^{2r+\varepsilon}\nu(du) < +\infty$.

*Then assumptions* T2.1, T2.2 *and* T2.3 *hold*; *in particular*,

$$n^{r/d}\mathcal{D}_{\mu,n,r}^{\nu} \xrightarrow{a.s.\text{ and }L^2} B_d^{-r/d}\Gamma\left(1+\frac{r}{d}\right)\int_{\mathbb{R}^d}\frac{\mu(du)}{g(u)^{r/d}}.$$

1.3. *Central limit theorem.* In the one-dimensional setting ($d = 1$), we show that the rate of convergence in Theorem 2 is governed by the following central limit theorem. We write $\xrightarrow{\mathcal{L}}$ to denote the usual convergence in law for $\mathbb{R}$-valued random variables.

THEOREM 3. *Let* $r > 1$. *Let* $\mu$ *and* $\nu$ *be some absolutely continuous probability measures on* $[0, 1]$ *with respective density functions* $f$ *and* $g$. *Assume that* $f$ *and* $g$ *are continuously differentiable on* $[0, 1]$ *and bounded away from zero. Then*

$$n^{1/2}\left(n^r\mathcal{D}_{\mu,n,r}^{\nu} - \frac{1}{2^r}\Gamma(1+r)\int_0^1\frac{f(u)}{g(u)^r}\,du\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{f,g}^2),$$

*where*

$$\sigma_{f,g}^2 := \frac{1}{2^{2r}(1+r)^2}\int_0^1 dz\, g(z)B(z)$$

*with*

$$\begin{aligned}
B(z) &:= \big((r+1)\Gamma(r+2)\big)^2\bigg(\bigg(\frac{A(z)}{1-G(z)}\bigg)^2 + 2\frac{A(z)f(z)}{g(z)^{1+r}(1-G(z))}\bigg)\\
&\qquad + \big(\Gamma(2r+3)-\Gamma(r+2)^2\big)\bigg(\frac{f(z)}{g(z)^{1+r}}\bigg)^2,\\
A(z) &:= \int_z^1 dy\bigg[\frac{f'(y)(1-G(y))}{(r+1)g(y)^{1+r}} - \frac{f(y)}{g(y)^r}\bigg(1+\frac{(1-G(y))g'(y)}{g(y)^2}\bigg)\bigg],\\
G(z) &:= \int_0^z g(t)\,dt.
\end{aligned}$$

**2. Comments.** *$L^p$ convergence.* Theorem 1 and Corollary 1 are derived in a rather general setting. They involve only some moment conditions on $\mu$ and $\nu$ and, roughly speaking, a $\mu$-integrability condition on a negative power of the density $g$. Indeed, from Lemma 8 (see Section 4.7) the value of $\delta$ in assumption C1.1 can be chosen small enough so that one can hope to obtain $g_\delta \simeq g$. Moreover, under some mild additional regularity property of $\nu$ (e.g., $\nu(B(u,s)) \geq c(g(u)\wedge 1)s^d$ for $s \in (0,\delta]$), the condition C1.1 becomes

$$\int_{\mathbb{R}^d} \frac{\mu(du)}{g(u)^{pr/(d(1-\eta))}} < +\infty. \tag{7}$$

Assumption (7) seems satisfactory since it is close (at least for $\eta$ close to 0) to the assumption $\int \mu/g^{r/d} < +\infty$ which ensures the existence of the limiting constant in Theorem 1. Assumption T1.1 is less intuitive than (7), but, if $\eta$ is close to 1, it would be better to use T1.1 instead of (7).

*Almost sure convergence.* Theorem 2 shows that the optimal distortion $\mathcal{D}^*_{\mu,n,r}$ and the random distortion $\mathcal{D}^\nu_{\mu,n,r}$ vanish almost surely at the same rate. Assumptions T2.2 and T2.3 are similar to T1.1 and T1.2 with $p=2$. Hence, this result is obtained from the $L^2$ convergence essentially under the additional assumption of a bounded $\varepsilon$-potential for the quantized measure $\mu$, where $\varepsilon$ can be chosen small enough. Therefore, T2.1 is not very stringent; for instance, every bounded probability density function satisfies T2.1.

A practical consequence of Theorem 2 is that one can hope to obtain a good initial distortion in the CLVQ algorithm, at least for large $n$. Nevertheless, we point out that the corresponding deterministic limiting constant $\lim n^{r/d}\mathcal{D}^\mu_{\mu,n,r}$ is not the best achievable by random quantization. Indeed, from the reverse Hölder's inequality, one can show that the infimum $\inf_g \int g^{-r/d}\, d\mu$ is attained for the probability density function

$$g = g_{\text{opt}} := \frac{f^{d/(d+r)}}{\int f^{d/(d+r)}}$$

(see [21]; note that from the assumption $\int \|u\|^{r+\varepsilon} \mu(du) < +\infty$ and from Hölder's inequality, one has $\int f^{d/(d+r)} < +\infty$). The corresponding limiting constant is then $B_d^{-r/d}\Gamma(1 + r/d)\|f\|_{d/(d+r)}$. Hence, from (3) and from the Bucklew and Wise theorem, the asymptotic suboptimality of the "random optimal" quantization is only due to the geometric unstability of the related Voronoï tesselation. The self-quantization procedure induces a second source of suboptimality, due to the suboptimal $f$-repartition of the codes.

*Central limit theorem.* We proved the CLT in a more stringent setting for a technical reason: assuming $d = 1$ along with some regularity on $f$ and $g$ allowed us to use and extend some spacing statistics techniques, namely, the Pyke–Hall method (see Section 4.6.1). The limiting variance $\sigma^2_{f,g}$ appears as a generalization of Hall's, which equals $2^{2r}(r + 1)^2\sigma^2_{1,g}$.

In the following section, we show that the assumptions of Theorems 1 and 2 satisfy two interesting robustness properties.

**3. Two robustness properties for T1 and T2.** Here we investigate the robustness of the assumptions of Theorems 1 and 2 under the convex combination and the tensor product of the codes laws.

3.1. *Convex combination.* Let $\mu$, $\nu_1$, $\nu_2$ be some absolutely continuous probability measures on $\mathbb{R}^d$ such that $\text{supp}(\nu_1) \cup \text{supp}(\nu_2) \supset \text{supp}(\mu)$ and let $\mu_i := \mu_{|\text{supp}(\nu_i)}$, $i = 1, 2$.

We check that for every $\alpha \in (0, 1)$, the couple $(\mu, \alpha\nu_1 + (1-\alpha)\nu_2)$ satisfies T1.1 (resp., T1.2, T2.1, T2.2, T2.3) as soon as $(\mu_1, \nu_1)$ and $(\mu_2, \nu_2)$ satisfy T1.1 (resp., T1.2, T2.1, T2.2, T2.3).

ASSUMPTION T1.1. Assume that $(\mu_1, \nu_1)$, $(\mu_2, \nu_2)$ satisfy T1.1.
Let

$$g_\delta(u) := \inf_{s\in(0,\delta]} \frac{\lambda(B(0, s))}{\alpha\nu_1 + (1-\alpha)\nu_2(B(0, s))},$$

$$g_\delta^{(i)}(u) := \inf_{s\in(0,\delta]} \frac{\lambda(B(0, s))}{\nu_i(B(0, s))}, \qquad i = 1, 2,$$

and let $\rho > 0$. Since $\mathrm{supp}(\nu_1) \cup \mathrm{supp}(\nu_2) \supset \mathrm{supp}(\mu)$, one has $\mu \le \mu_1 + \mu_2$ and then

$$\begin{aligned}
&\int_{\mathbb{R}^d} \biggl( \sup_{s\in(0,\|u\|+\rho)} \frac{\lambda(B(u,s))}{\alpha\nu_1 + (1-\alpha)\nu_2(B(u,s))} \biggr)^{pr/d} \mu(du) \\
&\quad \le \int_{\mathbb{R}^d} \biggl( \sup_{s\in(0,\|u\|+\rho)} \frac{\lambda(B(u,s))}{\alpha\nu_1 + (1-\alpha)\nu_2(B(u,s))} \biggr)^{pr/d} (\mu_1+\mu_2)(du) \\
&\quad = \frac{1}{\alpha^{pr/d}} \int_{\mathbb{R}^d} \frac{\mu_1(du)}{g^{(1)}_{\|u\|+\rho}(u)^{pr/d}} + \frac{1}{(1-\alpha)^{pr/d}} \int_{\mathbb{R}^d} \frac{\mu_2(du)}{g^{(2)}_{\|u\|+\rho}(u)^{pr/d}} \\
&\quad < +\infty
\end{aligned}$$

from Lemma 8 and the fact that $(\mu_1,\nu_1)$ and $(\mu_2,\nu_2)$ satisfy T1.1.

Assumption T2.2 is similar to Assumption T1.1 and Assumption T1.2 are straightforward. As a consequence, one obtains the following robustness property for the random quantization: If $(\mu_1,\nu_1)$ and $(\mu_2,\nu_2)$ satisfy T1.1 and T1.2 (which implies that $n^{r/d}\mathcal{D}^{\nu_i}_{\mu_i,n,r} \to_{L^p} l_{r,d,\mu_i,\nu_i}$), then, $n^{r/d}\mathcal{D}^{\alpha\nu_1+(1-\alpha)\nu_2}_{\mu,n,r} \to_{L^p} l_{r,d,\mu,\alpha\nu_1+(1-\alpha)\nu_2}$. A similar result holds for the almost sure convergence.

3.2. *Tensor product.* Let $d_1, d_2 \in \mathbb{N}^*$ such that $d_1 + d_2 = d$ and let $\nu_1$ (resp., $\nu_2$, $\mu$) be an absolutely continuous probability measure on $\mathbb{R}^{d_1}$ (resp., $\mathbb{R}^{d_1}$, $\mathbb{R}^d$). Let $\mu_1$ (resp., $\mu_2$) be the margin distributions of $\mu$ over $\mathbb{R}^{d_1}$ (resp., $\mathbb{R}^{d_2}$). We derive that $(\mu, \nu_1\otimes\nu_2)$ satisfies the assumptions of Theorem 1 (resp., Theorem 2) as soon as $(\mu_1,\nu_1)$ and $(\mu_2,\nu_2)$ satisfy the assumptions of Theorem 1 (resp., Theorem 2).

ASSUMPTIONS T1.1 and T1.2. Let $C(u,r)$ denote the open cube with center $u \in \mathbb{R}^d$ and half-side $r \ge 0$. First, one has

$$\begin{aligned}
g^{-1}_{\|u\|+\rho}(u) &\le \sup_{s\in(0,\|u\|+\rho]} \frac{\lambda(C(u,s))}{\nu(C(u,d^{-1/2}s))} \\
&\le d^{d/2} \prod_{i=1,2} \sup_{s\in(0,\|u\|+\rho]} \frac{\lambda_i(C^{(i)}(u^{(i)},d^{-1/2}s))}{\nu_i(C^{(i)}(u^{(i)},d^{-1/2}s))},
\end{aligned} \tag{8}$$

where $(u^{(1)},u^{(2)}) \in \mathbb{R}^{d_1}\times\mathbb{R}^{d_2} = u \in \mathbb{R}^d$ and $C^{(1)}\times C^{(2)} = C$. Second, from Hölder's inequality,

$$\begin{aligned}
&\int_{\mathbb{R}^d} \biggl( \prod_{i=1,2} \sup_{s\in(0,\|u\|+\rho]} \frac{\lambda_i(C^{(i)}(u^{(i)},d^{-1/2}s))}{\nu_i(C^{(i)}(u^{(i)},d^{-1/2}s))} \biggr)^{pr/d} \mu(du) \\
&\quad \le \prod_{i=1,2} \biggl( \int_{\mathbb{R}^d} \biggl( \sup_{s\in(0,\|u\|+\rho]} \frac{\lambda_i(C^{(i)}(u^{(i)},d^{-1/2}s))}{\nu_i(C^{(i)}(u^{(i)},d^{-1/2}s))} \biggr)^{pr/d_i} \mu(du) \biggr)^{d_i/d}.
\end{aligned}$$

However,

$$\sup_{s\in(0,\|u\|+\rho]} \frac{\lambda_i(C^{(i)}(u^{(i)}, d^{-1/2}s))}{\nu_i(C^{(i)}(u^{(i)}, d^{-1/2}s)}$$

$$(9) \qquad \le \sup_{s\in(0,d^{1/2}(\|u^{(i)}\|+\rho)]} \frac{\lambda_i(C^{(i)}(u^{(i)}, d^{-1/2}s))}{\nu_i(C^{(i)}(u^{(i)}, d^{-1/2}s))} + \frac{\lambda_i(C^{(i)}(u^{(i)}, d^{-1/2}(\|u\|+\rho)))}{\nu_i(C^{(i)}(u^{(i)}, (\|u^{(i)}\|+\rho)))}.$$

Therefore, from Lemma 3 (see Section 4.7) there exists $\xi_{\nu_i} > 0$ such that

$$(10) \qquad \sup_{s\in(0,\|u\|+\rho]} \frac{\lambda_i(C^{(i)}(u^{(i)}, d^{-1/2}s))}{\nu_i(C^{(i)}(u^{(i)}, d^{-1/2}s))} \le d_i^{d_i/2} \sup_{s\in(0,\|u^{(i)}\|+\rho]} \frac{\lambda_i(B^{(i)}(u^{(i)}, s))}{\nu_i(B^{(i)}(u^{(i)}, s))} + \frac{\lambda_i(B^{(i)}(u^{(i)}, (\|u\|+\rho)))}{\xi_{\nu_i}}.$$

From (8), (9) and (10), one finally gets that

$$\int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\|u\|+\rho}(u)^{pr/d}} \le M \prod_{i=1,2} \left[ \int_{\mathbb{R}^{d_i}} \frac{\mu_i(du^{(i)})}{g^{(i)}_{\|u^{(i)}\|+\rho}(u)^{pr/d_i}} + \int_{\mathbb{R}^d} (\|u\|+\rho)^{pr} \mu(du) \right]^{d_i/d} < +\infty.$$

Hence $(\mu, \nu_1 \otimes \nu_2)$ satisfies Assumption T1.1. The case of Assumption T1.2 is straightforward and Assumptions T2.1, T2.2 and T2.3 are similar to Assumptions T1.1 and T1.2.

As a consequence, one obtains the following robustness property for the random quantization: if $(\mu_1, \nu_1)$ and $(\mu_2, \nu_2)$ satisfy T1.1 and T1.2 (which implies that $n^{r/d_i} \mathcal{D}^{\nu_i}_{\mu_i,n,r} \to_{L^p} l_{r,d_i,\mu_i,\nu_i}$) then, $n^{r/d} \mathcal{D}^{\nu_1\otimes\nu_2}_{\mu,n,r} \to_{L^p} l_{r,d,\mu,\nu_1\otimes\nu_2}$. A similar result holds for the almost sure convergence.

## 4. Proofs.

4.1. *Notation.* In the following items, $f$ denotes a nonnegative Borel function on $(\mathbb{R}^d)^p \times \mathbb{R}_+^p$ and $(\mathbf{u}, \mathbf{v})$ denotes $(u_1, \dots, u_p, v_1, \dots, v_p) \in (\mathbb{R}^d)^p \times \mathbb{R}_+^p$.

1. $$I_{\mu,p}[f(\mathbf{u},\mathbf{v})] := \int_{(\mathbb{R}^d)^p} \mu(du_1)\cdots\mu(du_p) \int_{\mathbb{R}_+^p} dv_1 \cdots dv_p \, f(\mathbf{u},\mathbf{v}).$$

2. $$\Upsilon_{n,p}(\mathbf{u},\mathbf{v}) := \left(1 - \nu\left(\bigcup_{j=1}^{p} B\left(u_j, \frac{v_j^{1/r}}{n^{1/d}}\right)\right)\right)^n.$$

3. $$l_{r,d,\mu,\nu} := \int_{\mathbb{R}^d} \mu(du) \int_{\mathbb{R}_+} dv\, \exp(-B_d g(u) v^{d/r}) = B_d^{-r/d} \Gamma\left(1 + \frac{r}{d}\right) \int_{\mathbb{R}^d} \frac{\mu(du)}{g(u)^{r/d}}.$$

4. $$A_{n,\rho}^{(p)} := \left\{(\mathbf{u},\mathbf{v}) \in (\mathbb{R}^d)^p \times \mathbb{R}_+^p;\ \max_{1\le i\le p} \frac{v_i^{1/r}}{n^{1/d}} \le \max_{1\le i\le p} \|u_i\| + \rho\right\} \subset (\mathbb{R}^d)^p \times \mathbb{R}_+^p.$$

The convergence in probability (resp., the mean, the variance) with respect to $\mathbb{P}$ will be denoted $\stackrel{\mathbb{P}}{\to}$, (resp., $\mathbb{E}$, $\mathbb{V}$).

4.2. *Proof of Theorem* 1. One has (Lemma 1)

$$\mathbb{E}(n^{r/d} \mathcal{D}_{\mu,n,r}^{\nu})^p = I_{\mu,p}[\mathbb{1}_{A_{n,\rho}^{(p)}} \Upsilon_{n,p}] + I_{\mu,p}[\mathbb{1}_{\complement A_{n,\rho}^{(p)}} \Upsilon_{n,p}]. \tag{11}$$

However, (Lemma 2), for every $(\mathbf{u},\mathbf{v}) \in (\mathbb{R}^d)^p \times \mathbb{R}_+^p$,

$$\mathbb{1}_{A_{n,\rho}^{(p)}} \Upsilon_{n,p}(\mathbf{u},\mathbf{v}) \le \sum_{j=1}^{p} \exp\left(-B_d g_{\max_{1\le i\le p}\|u_i\|+\rho}(u_j) \max_{1\le i\le p} v_i^{d/r}\right) \tag{12}$$

and (Lemma 4)

$$I_{\mu,p}\left[\sum_{j=1}^{p} \exp\left(-B_d g_{\max_{1\le i\le p}\|u_i\|+\rho}(u_j) \max_{1\le i\le p} v_i^{d/r}\right)\right] < +\infty. \tag{13}$$

Since $A_{n,\rho}^{(p)} \uparrow (\mathbb{R}^d)^p \times \mathbb{R}_+^p$ as $n \uparrow +\infty$, Lemma 5 yields

$$\mathbb{1}_{A_{n,\rho}^{(p)}} \Upsilon_{n,p} \xrightarrow{\mu^{\otimes p}\otimes\lambda^{\otimes p}\text{-a.e.}} \exp\left(-B_d \sum_{j=1}^{p} g(u_j) v_j^{d/r}\right). \tag{14}$$

So (12), (13), (14) and the Lebesgue dominated convergence theorem yield

$$I_{\mu,p}[\mathbb{1}_{A_{n,\rho}^{(p)}} \Upsilon_{n,p}] \to I_{\mu,p}\left[\exp\left(-B_d \sum_{j=1}^{p} g(u_j) v_j^{d/r}\right)\right] = l_{r,d,\mu,\nu}^p. \tag{15}$$

However (Lemma 6),

$$I_{\mu,p}[\mathbb{1}_{\complement A_{n,\rho}^{(p)}} \Upsilon_{n,p}] \to 0. \tag{16}$$

Therefore (15), (16) and (11) yield

$$\mathbb{E}(n^{r/d} \mathcal{D}_{\mu,n,r}^{\nu})^p \to l_{r,d,\mu,\nu}^p. \tag{17}$$

Now, observe that Assumptions T2.1 and T2.2 hold for every $q \in \{1, \dots, p\}$ and the fact that $p \geq 2$ has not been used to derive (17). As a consequence, (17) holds for every $1 \leq q \leq p$, $\mathbb{E}(n^{r/d}\mathcal{D}^{\nu}_{\mu,n,r})^q \to l^q_{r,d,\mu,\nu}$ and, in particular, $n^{r/d}\mathcal{D}^{\nu}_{\mu,n,r} \to_{\mathbb{P}} l_{r,d,\mu,\nu}$. Finally, Lemma 7 yields $n^{r/d}\mathcal{D}^{\nu}_{\mu,n,r} \to_{L^p} l_{r,d,\mu,\nu}$ which completes the proof.

4.3. *Proof of Corollary* 1. We have to show that C1.1 and C1.2 imply T2.1 and T2.2.

First, C1.2 straightforwardly implies T1.2.

Second, one has for every $u \in \operatorname{supp}(g)$,

$$g_{\|u\|+\rho}(u)^{-1} \leq g_\rho(u)^{-1} + \lambda\big(B(u, \|u\| + \rho)\big)/\nu\big(B(u, \rho)\big).$$

Hence,

$$\int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\|u\|+\rho}(u)^{pr/d}} \leq \int_{\mathbb{R}^d} \frac{\mu(du)}{g_\rho(u)^{pr/d}} + \int_{\mathbb{R}^d} \mu(du)\bigg(\frac{\lambda(B(u, \|u\| + \rho))}{\nu(B(u, \rho))}\bigg)^{pr/d}.$$

So, from Hölder's inequality [recall that $\eta \in (0, 1)$], $\int_{\mathbb{R}^d} g_{\|u\|+\rho}(u)^{-pr/d}\mu(du)$ is less than

$$\int_{\mathbb{R}^d} \frac{\mu(du)}{g_\rho(u)^{pr/d}} + \bigg(\int_{\mathbb{R}^d} \mu(du)\big(\lambda\big(B(u, \|u\| + \rho)\big)\big)^{pr/d\eta}\bigg)^{\eta}\bigg(\int_{\mathbb{R}^d} \frac{\mu(du)}{\nu(B(u, \rho))^{pr/(d(1-\eta))}}\bigg)^{1-\eta}.$$

Therefore,

$$\int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\|u\|+\rho}(u)^{pr/d}} \leq \int_{\mathbb{R}^d} \frac{\mu(du)}{g_\rho(u)^{pr/d}} + \frac{B_d^{pr/d}}{\rho^{pr}}\bigg(\int_{\mathbb{R}^d} \mu(du)(\|u\| + \rho)^{pr/\eta}\bigg)^{\eta}\bigg(\int_{\mathbb{R}^d} \frac{\mu(du)}{g_\rho(u)^{pr/(d(1-\eta))}}\bigg)^{1-\eta}.$$

Assumptions C1.1 and C1.2 along with Lemma 8 then ensure that $\int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\|u\|+\rho}(u)^{pr/d}} < +\infty$. Lemma 8 completes the proof.

4.4. *Proof of Theorem* 2. First, we show that the assumptions of Lemma 9 are fulfilled with $\mathcal{E}_n := A^{(2)}_{n,\rho}$ and

$$\psi(\mathbf{u}, \mathbf{v}) := \sum_{j=1}^{2} \exp\big(-B_d g_{\|u_1\| \vee \|u_2\|+\rho}(u_j) v_1^{d/r} \vee v_2^{d/r}\big).$$

ASSUMPTION L9.1. Lemma 2 with $p = 2$ yields for every $n \in \mathbb{N}^*$ and every $(\mathbf{u}, \mathbf{v}) \in (\mathbb{R}^d)^2 \times \mathbb{R}_+^2$, $\mathbb{1}_{A_{n,\rho}^{(2)}} \Upsilon_{n,2} \leq \psi(\mathbf{u}, \mathbf{v})$. Assumption L9.1 then holds.

ASSUMPTION L9.2. From Lemma 4, we already know that $I_{\mu,2}[\psi] < +\infty$. So, by Lemma 11, one obtains $I_{\mu,2}[(1 + \delta_\varepsilon)\psi] < +\infty$. Assumption L9.2 then holds.

ASSUMPTION L9.3. From Lemma 12, one has $I_{\mu,2}[\mathbb{1}_{\complement A_{n,\rho}^{(2)}} \delta_\varepsilon \Upsilon_{n,2}] \to 0$. In addition, since $A_{n,\rho}^{(2)} \uparrow (\mathbb{R}^d)^2 \times \mathbb{R}_+^2$, one has $I_{\mu,2}[\mathbb{1}_{\complement A_{n,\rho}^{(2)}} \Upsilon_{n,2}] \to 0$. Hence, $I_{\mu,2}[\mathbb{1}_{\complement A_{n,\rho}^{(2)}} (1 + \delta_\varepsilon)\Upsilon_{n,2}] \to 0$. In particular, Assumption L9.3 holds.

Then we get from Lemma 9 that for every $k \geq d/\varepsilon$,

$$n^{2kr/d} \mathcal{D}_{\mu,n^{2k},r}^{\nu} - \mathbb{E}\big(n^{2kr/d} \mathcal{D}_{\mu,n^{2k},r}^{\nu}\big) \overset{\text{a.s.}}{\to} 0. \tag{18}$$

However, Assumptions T2.2 and T2.3 imply Assumptions T1.1 and T1.2 (with $p = 2$). Hence, from Theorem 1, one has $n^{r/d} \mathcal{D}_{\mu,n,r}^{\nu} \to_{L^2} l_{r,d,\mu,\nu}$. In particular, $\mathbb{E}(n^{2kr/d} \mathcal{D}_{\mu,n^{2k},r}^{\nu}) \to l_{r,d,\mu,\nu}$ and from (18), one obtains $n^{2kr/d} \mathcal{D}_{\mu,n^{2k},r}^{\nu} \overset{\text{a.s.}}{\to} l_{r,d,\mu,\nu}$. Finally Lemma 10 applied with $a_n := \mathcal{D}_{\mu,n,r}^{\nu}$ completes the proof.

4.5. *Proof of Corollary* 2. As in the proof of Corollary 1, one can show, using Hölder's inequality, that Assumptions C2.2 and C2.3 entail Assumptions T2.2 and T2.3. The details are then omitted.

4.6. *Proof of Theorem* 3. In this section, $\mathcal{M}$ (resp., $\mathcal{P}$) denotes the set of the absolutely continuous positive finite (resp., probability) measures having a density function $g$ such that $g \in \mathcal{C}^1([0, 1])$ and $\inf g > 0$.

4.6.1. *The Pyke–Hall method.* Let $(Z_{1,n}, \ldots, Z_{n,n})$ be the order statistics of $(Y_1, \ldots, Y_n)$. Then the normalized uniform distortion $n^r \mathcal{D}_{\lambda_{[0,1]},n,r}^{\nu}$ can be approximated by Kimball's type spacing statistics

$$S_n^{(1)} := \frac{1}{n} \sum_{k=1}^{n-1} \big(n(Z_{k+1,n} - Z_{k,n})\big)^{r+1},$$

where $(Z_{1,n}, \ldots, Z_{n,n})$ is the order statistics of $(Y_1, \ldots, Y_n)$. Indeed, one can prove that (e.g., if $\nu \in \mathcal{P}$)

$$n^{1/2}\left(n^r \mathcal{D}_{\lambda_{[0,1]},n,r}^{\nu} - \frac{1}{2^r(r+1)} S_n^{(1)}\right) \overset{\mathbb{P}}{\to} 0. \tag{19}$$

So, CLT for $n^r \mathcal{D}_{\lambda_{[0,1]},n,r}^{\nu}$ reduces to CLT for $S_n^{(1)}$.

In [18], Pyke introduced a method based on Rényi's version of order statistics (see the Appendix for the related notation) to derive some limit theorems for $S_n^{(1)}$. This method has been extended to general statistics of $m$-spacings by Hall [12]. For 1-spacings, Hall's CLT reads as follows.

THEOREM 4. *With the preceding notation*,

$$n^{1/2}\Big(S_n^{(1)} - \int_0^1 dx\, g(x)^2 \int_0^1 dz\, z^{r+1}\exp\big(-zg(x)\big)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2).$$

(See [12] for the explicit computation of $\sigma^2$.)

The key to the proof of Theorem 4 is first to assume that $(Z_{1,n},\dots,Z_{n,n})$ is Rényi's version of order statistics (see the Appendix), that is,

$$Z_{k,n} = K_g\Bigg(\sum_{j=1}^{k}\frac{X_j}{n-j+1}\Bigg), \tag{20}$$

where $(X_j)_{j\geq 0}$ is an i.i.d. exponential sequence with parameter 1, where $K_g := G^{-1}(1-\exp(\cdot))$ and where $G$ is the distribution function of $\nu$. Second, the key approximation is as follows.

PROPOSITION 1. *Let $r>1$; let $\nu\in\mathcal{P}$ with density function $g$. Then*

$$n^{1/2}\big((S_n^{(1)} - S) - (S_n^{(2)} - \mathbb{E}(S_n^{(2)}))\big) \xrightarrow{\mathbb{P}} 0,$$

*where*

$$\begin{aligned} S_n^{(1)} &:= \frac{1}{n}\sum_{k=1}^{n-1}\big(n(Z_{k+1,n}-Z_{k,n})\big)^{r+1},\\ S &:= \int_0^{+\infty} dz\,\exp(-z)\int_0^1 dx\,\Big(\frac{z}{1-x}K_g'\big(-\log(1-x)\big)\Big)^{r+1} \end{aligned} \tag{21}$$

*and*

$$\begin{aligned} S_n^{(2)} = \frac{1}{n}\sum_{k=1}^{n-1}&\frac{n}{n-k}\Bigg(\sum_{j=1}^{k}\frac{X_j-1}{n-j+1}\Bigg)K_g''\Big(-\log\Big(1-\frac{k}{n}\Big)\Big)\\ &\times\Gamma(r+2)(r+1)\Big(\frac{n}{n-k}K_g'\Big(-\log\Big(1-\frac{k}{n}\Big)\Big)\Big)^r\\ &+\Big(\frac{n}{n-k}X_{k+1}K_g'\Big(-\log\Big(1-\frac{k}{n}\Big)\Big)\Big)^{r+1}. \end{aligned} \tag{22}$$

Reordering the terms in (22) according to the $X_j$'s, one obtains a sum of independent random variables to which the Lindeberg central limit theorem applies. The related CLT for $S_n^{(2)}$ together with Proposition 1 finally yield Theorem 4.

Our goal is to extend Proposition 1 in order to deal with the nonuniform distortion $n^r\mathcal{D}_{\mu,n,r}^{\nu}$.

4.6.2. *Proof* (*of Theorem* 3). We observe that if $d = 1$, the distortion $\mathcal{D}^{\nu}_{\mu,n,r}$ is a function of $(Z_{1,n}, \dots, Z_{n,n})$. So, in order to derive a CLT for $n^r \mathcal{D}^{\nu}_{\mu,n,r}$, we can deal with a specific version of order statistics $(Z_{1,n}, \dots, Z_{n,n})$; in particular, we assume without loss of generality that $(Z_{1,n}, \dots, Z_{n,n})$ is Rényi's version of order statistics defined by (20).

For $1 \le k \le n-1$, let $\zeta_{k,n} \in [Z_{k,n}, Z_{k+1,n}]$. Then (Lemma 13)

$$n^{1/2}\left(n^r \mathcal{D}^{\nu}_{\mu,n,r} - \frac{1}{2^r(r+1)} W_n\right) \xrightarrow{\mathbb{P}} 0. \tag{23}$$

Hence, the CLT for $n^r \mathcal{D}^{\nu}_{\mu,n,r}$ reduces to the CLT for $W_n$.

The next step is to apply Lemma 14 to $W_n$.

From the mean value theorem, there exists for $1 \le k \le n-1$ a $[0,1]$-valued random variable $\theta_{k,n}$ such that

$$\frac{X_{k+1}}{n-k} K_g'\left(\sum_{j=1}^{k} \frac{X_j}{n-j+1} + \theta_{k,n}\frac{X_{k+1}}{n-k}\right) = Z_{k+1,n} - Z_{k,n}.$$

Since the function $K_g$ is increasing, one has

$$K_g\left(\chi_{k,n} + \theta_{k,n}\frac{X_{k+1}}{n-k}\right) \in [Z_{k,n}, Z_{k+1,n}] \qquad \text{where } \chi_{k,n} := \sum_{j=1}^{k} \frac{X_j}{n-j+1}.$$

Therefore, we can set $\zeta_{k,n} := K_g(\chi_{k,n} + \theta_{k,n}X_{k+1}/(n-k))$. Subsequently, $W_n$ writes

$$\frac{1}{n}\sum_{k=1}^{n-1}\left(\frac{n}{n-k}X_{k+1}\left((f\circ K_g)^{1/(1+r)}\frac{\exp(-\cdot)}{g\circ K_g}\right)\left(\chi_{k,n} + \theta_{k,n}\frac{X_{k+1}}{n-k}\right)\right)^{r+1}.$$

However, since $f, g \in \mathcal{P}$, one has $gf^{-1/(1+r)} \in \mathcal{M}$. Therefore, in order to apply Lemma 14 we can set $\varphi := gf^{-1/(1+r)}$ and $\psi := g$. With this choice,

$$(f\circ K_g)^{1/(1+r)}\frac{\exp(-\cdot)}{g\circ K_g} = H_{gf^{-1/(1+r)},g}$$

and the sum $S_n^{(1)}$ defined in Lemma 14 equals $W_n$. Consequently, Lemmas 14, 15 and 13 yield

$$n^{1/2}\left(n^r \mathcal{D}^{\nu}_{\mu,n,r} - \frac{1}{2^r(r+1)}S\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{2^{2r}(r+1)^2}\mathfrak{S}^2_{gf^{-1/(1+r)},g}\right)$$

and an elementary computation shows that

$$\frac{1}{2^{2r}(r+1)^2}\mathfrak{S}^2_{gf^{-1/(1+r)},g} = \sigma^2_{f,g} \quad \text{and} \quad \frac{1}{2^r}\Gamma(1+r)\int_0^1 \frac{f(u)}{g(u)^r}\,du = \frac{1}{2^r(r+1)}S,$$

which completes the proof of Theorem 3.

4.7. *Lemmas.*

4.7.1. *Lemmas for Theorems* 1 *and* 2.

LEMMA 1. *Let $\mu, \nu$ be two probability measures on $\mathbb{R}^d$. Then, for every $n \in \mathbb{N}^*$ and every $p \in \mathbb{N}^*$, one has $\mathbb{E}(n^{r/d}\mathcal{D}^{\nu}_{\mu,n,r})^p = I_{\mu,p}[\Upsilon_{n,p}]$.*

PROOF. One easily writes that

$$\mathbb{E}(n^{r/d}\mathcal{D}^{\nu}_{\mu,n,r})^p = \mathbb{E}\biggl(\int_{(\mathbb{R}^d)^p} \prod_{j=1}^{p} n^{r/d} \min_{1\le i\le n} \|Y_i - u_j\|^r \mu(du_1)\cdots\mu(du_p)\biggr)$$

$$= I_{\mu,p}\biggl[\mathbb{P}\biggl(\bigcap_{i=1}^{n}\bigcap_{j=1}^{p}\{n^{r/d}\|Y_i - u_j\|^r \ge v_j\}\biggr)\biggr].$$

Using that the $Y_i$'s are i.i.d. finally yields the needed result. □

LEMMA 2. *Let $d \in \mathbb{N}^*$, $p \in \mathbb{N}^*$, $r > 0$ and $\rho > 0$. For every $\delta > 0$ and every $u \in \mathbb{R}^d$, set $g_\delta(u) := \inf_{s\in(0,\delta]} \frac{\nu(B(u,s))}{\lambda(B(u,s))}$. Then for every $n \in \mathbb{N}^*$ and every $(\mathbf{u}, \mathbf{v}) \in (\mathbb{R}^d)^p \times \mathbb{R}^p_+$,*

$$\mathbb{1}_{A^{(p)}_{n,\rho}} \Upsilon_{n,p}(\mathbf{u}, \mathbf{v}) \le \sum_{j=1}^{p} \exp\Bigl(-B_d\, g_{\max_{1\le i\le p}\|u_i\|+\rho}(u_j) \max_{1\le i\le p} v_i^{d/r}\Bigr).$$

PROOF. One has

$$\Upsilon_{n,p}(\mathbf{u}, \mathbf{v}) \le \sum_{j=1}^{p} \Upsilon_{n,1}\Bigl(u_j, \max_{1\le i\le p} v_i^{d/r}\Bigr). \tag{24}$$

Let $j \in \{1, \dots, p\}$. One gets

$$\Upsilon_{n,1}\Bigl(u_j, \max_{1\le i\le p} v_i^{d/r}\Bigr) \le \exp\biggl(-B_d \frac{\nu(B(u_j, \max_{1\le i\le p} v_i^{1/r}/n^{1/d}))}{\lambda(B(u_j, \max_{1\le i\le p} v_i^{1/r}/n^{1/d}))} \max_{1\le i\le p} v_i^{d/r}\biggr).$$

By the definition of the set $A^{(p)}_{n,\rho}$, one finally obtains

$$\begin{aligned} &\mathbb{1}_{A^{(p)}_{n,\rho}} \Upsilon_{n,1}\Bigl(u_j, \max_{1\le i\le p} v_i^{d/r}\Bigr) \\ &\quad \le \exp\biggl(-B_d \inf_{s\in(0,\max_{1\le i\le p}\|u_i\|+\rho]} \frac{\nu(B(u_j, s))}{\lambda(B(u_j, s))} \max_{1\le i\le p} v_i^{d/r}\biggr). \end{aligned} \tag{25}$$

Summing (25) over $j$ and using (24) completes the proof. □

LEMMA 3. *Let $\nu$ be a probability measure on $\mathbb{R}^d$ and let $\rho > 0$. Then*

$$\inf_{u \in \mathrm{supp}(\nu)} \nu\big(B(u, \|u\| + \rho)\big) > 0.$$

PROOF. Assume that there is a sequence $(u_n)_{n \geq 0}$ in $\mathrm{supp}(\nu)$ such that $\nu(B(u_n, \|u_n\| + \rho)) \to 0$. Seeing that the function $u \mapsto \nu(B(u, \|u\| + \rho))$ is lower semicontinuous and does not vanish on every compact subset of $\mathrm{supp}(\nu)$, one can assume w.l.g. that $\|u - u_n\| < \|u_n\|$ for every $n$ for some $u \in \mathrm{supp}(\nu)$. Then one has $B(u, \rho) \subset \liminf B(u_n, \|u_n\| + \rho)$, and, from Fatou's lemma,

$$\liminf \nu\big(B(u_n, \|u_n\| + \rho)\big) \geq \nu\big(\liminf B(u_n, \|u_n\| + \rho)\big) \geq \nu\big(B(u, \rho)\big) > 0$$

giving a contradiction. □

LEMMA 4. *Under the assumptions of Theorem* 1, *one has*

$$I_{\mu,p}\left[\sum_{j=1}^{p} \exp\left(-B_d\, g_{\max_{1 \leq i \leq p} \|u_i\| + \rho}(u_j) \max_{1 \leq i \leq p} v_i^{d/r}\right)\right] < +\infty.$$

PROOF.

$$\begin{aligned}
&I_{\mu,p}\left[\sum_{j=1}^{p} \exp\left(-B_d\, g_{\max_{1 \leq i \leq p} \|u_i\| + \rho}(u_j) \max_{1 \leq i \leq p} v_i^{d/r}\right)\right] \\
&\quad \leq M_1 \sum_{j=1}^{p} \int_{(\mathbb{R}^d)^p} \mu(du_1) \cdots \mu(du_p) g_{\max_{1 \leq i \leq p} \|u_i\| + \rho}(u_j)^{-pr/d} \\
&\quad \leq M_1 \sum_{j=1}^{p} \sum_{i=1}^{p} \int_{(\mathbb{R}^d)^2} \mu(du_j) \mu(du_i) g_{\|u_i\| + \rho}(u_j)^{-pr/d} \\
&\quad \leq M_2 \int_{(\mathbb{R}^d)^2} \mu(du) \mu(dw) \left(g_{\|u\| + \rho}(u)^{-1} + \frac{\lambda(B(u, \|w\| + \rho))}{\nu(B(u, \|u\| + \rho))}\right)^{pr/d} \\
&\quad \leq M_3 \int_{\mathbb{R}^d} \mu(du) \big(g_{\|u\| + \rho}(u)^{-pr/d} + (\|u\| + \rho)^{pr}\big) \qquad \text{(Lemma 3)}
\end{aligned}$$

and the preceding term is finite. □

LEMMA 5. *Let $d \in \mathbb{N}^*$, $p \in \mathbb{N}^*$ and $r \in \mathbb{R}_+^*$. Let $\mu$ and $\nu$ be some absolutely continuous probability measures on $\mathbb{R}^d$. Then*

$$\Upsilon_{n,p} \to \exp\left(-B_d \sum_{j=1}^{p} g(u_j) v_j^{d/r}\right), \qquad \mu^{\otimes p} \otimes \lambda^{\otimes p}\text{-}a.e. \tag{26}$$

The proof is sraightforward.

LEMMA 6. *Let $d \in \mathbb{N}^*$, $p \in \mathbb{N}^*$, $r > 0$ and $\rho > 0$. Assume that $\int_{\mathbb{R}^d} \|u\|^{pr}(\mu + \nu)(du) < +\infty$.*

*Then $I_{\mu,p}[\mathbb{1}_{\complement A_{n,\rho}^{(p)}} \Upsilon_{n,p}] \to 0$.*

PROOF. The change of variable $\mathbf{w} := \mathbf{v}/n^{r/d}$ yields

$$I_{\mu,p}[\mathbb{1}_{\complement A_{n,\rho}^{(p)}} \Upsilon_{n,p}] = I_{\mu,p}[\mathbb{1}_{B_\rho} n^{pr/d} \Theta_{p,r}^n], \tag{27}$$

where

$$B_\rho := \left\{ (\mathbf{u}, \mathbf{w}) \in \operatorname{supp}(\mu)^p \times \mathbb{R}_+^p;\ \max_{1 \le i \le p} w_i^{1/r} > \max_{1 \le i \le p} \|u_i\| + \rho \right\}$$

and where $\Theta_{p,r}(\mathbf{u}, \mathbf{w})$ is the function defined on $(\mathbb{R}^d)^p \times \mathbb{R}_+^p$ by

$$\Theta_{p,r}(\mathbf{u}, \mathbf{w}) := 1 - \nu\left( \bigcup_{j=1}^p B(u_j, w_j^{1/r}) \right).$$

On the set $B_\rho$, one has

$$\bigcup_{j=1}^p B(u_j, w_j^{1/r}) \supset B\left( u_{j_0}, \max_{1 \le j \le p} \|u_j\| + \rho \right) \supset B(u_{j_0}, \|u_{j_0}\| + \rho),$$

where the index $j_0$ satisfies $w_{j_0} = \max_{1 \le j \le p} w_j$.

Since $\operatorname{supp}(\mu) \subset \operatorname{supp}(\nu)$, Lemma 3 then yields the existence of $\xi_\nu > 0$ such that for every $(\mathbf{u}, \mathbf{w})$, $\mathbb{1}_{B_\rho} \Theta_{p,r}(\mathbf{u}, \mathbf{w}) \le 1 - \xi_\nu$. Consequently,

$$\mathbb{1}_{B_\rho} n^{pr/d} \Theta_{p,r}^n(\mathbf{u}, \mathbf{w}) \to 0 \qquad \text{as } n \uparrow +\infty. \tag{28}$$

In addition, one can easily show that the sequence $(\mathbb{1}_{B_\rho} n^{pr/d} \Theta_{p,r}^n)_{n \ge 1}$ is decreasing for $n \ge -pr/(d \log(1 - \xi_\nu))$. Subsequently, since $\Theta_{p,r} \le 1$, one obtains

$$\mathbb{1}_{B_\rho} n^{pr/d} \Theta_{p,r}^n \le \left( -\frac{pr}{d \log(1 - \xi_\nu)} \vee 1 \right)^{pr/d} \Theta_{p,r} \tag{29}$$

for every $n \ge -\frac{pr}{d \log(1-\xi_\nu)} \vee 1$.

Now, from $\int_{\mathbb{R}^d} \|u\|^{pr}(\mu + \nu)(du) < +\infty$, one easily obtains $I_{\mu,p}[\Theta_{p,r}] < +\infty$.

Then from (28), (29) and from the Lebesgue dominated convergence theorem, one finally gets that $I_{\mu,p}[\mathbb{1}_{B_\rho} n^{pr/d} \Theta_{p,r}^n] \to 0$ which, from (27), completes the proof. □

LEMMA 7. *Let $p \ge 1$ and $(X_n)_{n \ge 0}$, $X$ be random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $X_n \to_{\mathbb{P}} X$ and $\mathbb{E}|X_n|^p \to \mathbb{E}|X|^p$. Then $X_n \to_{L^p} X$.*

The proof follows easily from Fatou's lemma.

LEMMA 8. *Let $\varphi:\mathbb{R}^d \to \mathbb{R}_+$ and $\alpha > 0$. Then*

$$\left(\exists\,\rho > 0 \text{ such that } \int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\varphi(u)+\rho}(u)^\alpha} < +\infty\right) \iff \left(\forall\,\rho > 0, \int_{\mathbb{R}^d} \frac{\mu(du)}{g_{\varphi(u)+\rho}(u)^\alpha} < +\infty\right).$$

PROOF. This follows from the fact that for every $\rho, \rho' > 0$, $u \in \mathbb{R}^d$,

$$g_{\varphi(u)+\rho}(u)^{-1} \le \left(1 + \left(\frac{\varphi(u)+\rho}{\varphi(u)+\rho'}\right)^d\right) g_{\varphi(u)+\rho'}(u)^{-1}. \qquad \square$$

In the following lemmas, we use the following notation: for $u_1, u_2 \in \mathbb{R}^d$ ($u_1 \neq u_2$) and $v_1, v_2 \ge 0$,

$$\delta_\varepsilon(u_1, u_2, v_1, v_2) := \left(\frac{v_1^{1/r} + v_2^{1/r}}{\|u_1 - u_2\|}\right)^\varepsilon.$$

LEMMA 9. *Let $d \in \mathbb{N}^*$ and $r \in \mathbb{R}_+^*$. Let $(\mathcal{E}_n)_{n\ge1}$ be a sequence of Borel sets of $(\mathbb{R}^d)^2 \times \mathbb{R}_+^2$ and let $\mu$ and $\nu$ be some absolutely continuous probability measures on $\mathbb{R}^d$. Assume that there exists a function $\psi : (\mathbb{R}^d)^2 \times \mathbb{R}_+^2 \mapsto \mathbb{R}_+$ and $\varepsilon > 0$ such that*:

L9.1. $\mathbb{1}_{\mathcal{E}_n} \Upsilon_{n,2}(\mathbf{u}, \mathbf{v}) \le \psi(\mathbf{u}, \mathbf{v})$ $\mu \otimes \mu \otimes \lambda \otimes \lambda$*-a.e. for every* $n \in \mathbb{N}^*$.
L9.2. $I_{\mu,2}[(1 + \delta_\varepsilon)\psi] < +\infty$.
L9.3. $\limsup I_{\mu,2}[\mathbb{1}_{\complement\mathcal{E}_n}(1 + \delta_\varepsilon)\Upsilon_{n,2}] < +\infty$.

*Then, for every integer $k \ge d/\varepsilon$,*

$$n^{2kr/d} \mathcal{D}^\nu_{\mu,n^{2k},r} - \mathbb{E}\big(n^{2kr/d} \mathcal{D}^\nu_{\mu,n^{2k},r}\big) \overset{a.s.}{\to} 0. \tag{30}$$

PROOF. Let $k \in \mathbb{N}$ such that $k \ge d/\varepsilon$. To obtain (30), we show that

$$\sum_{n\ge1} \mathbb{V}\big(n^{2rk/d} \mathcal{D}^\nu_{\mu,n^{2k},r}\big) < +\infty. \tag{31}$$

To this end, write

$$\begin{aligned} n\mathbb{V}&\big(n^{rk/d} \mathcal{D}^\nu_{\mu,n^k,r}\big) \\ &= n\big(I_{\mu,2}[\Upsilon_{n^k,2}] - (I_{\mu,1}[\Upsilon_{n^k,1}])^2\big) \\ &= nI_{\mu,2}\big[\Upsilon_{n^k,2}(u_1,u_2,v_1,v_2) - \Upsilon_{n^k,1}(u_1,v_1)\Upsilon_{n^k,1}(u_2,v_2)\big]. \end{aligned}$$

Let $D_{d,2} := \{\mathbf{u} \in \mathbb{R}^d \times \mathbb{R}^d;\ u_1 = u_2\}$. Assume that $(\mathbf{u}, \mathbf{v}) \in \complement D_{d,2} \times \mathbb{R}_+^2$ and that $n > \delta_{d/k}(\mathbf{u}, \mathbf{v})$.

We have $B(u_1, v_1^{1/r}/n^{k/d}) \cap B(u_2, v_2^{1/r}/n^{k/d}) = \varnothing$ and then

$$\Upsilon_{n^k,2}(u_1, u_2, v_1, v_2) = \left(1 - \nu\left(B\left(u_1, \frac{v_1^{1/r}}{n^{k/d}}\right)\right) - \nu\left(B\left(u_2, \frac{v_2^{1/r}}{n^{k/d}}\right)\right)\right)^{n^k}.$$

Therefore, $\Upsilon_{n^k,2}(u_1, u_2, v_1, v_2) - \Upsilon_{n^k,1}(u_1, v_1)\Upsilon_{n^k,1}(u_2, v_2) \le 0$, which entails that

$$\begin{aligned} &n(\Upsilon_{n^k,2}(u_1, u_2, v_1, v_2) - \Upsilon_{n^k,1}(u_1, v_1)\Upsilon_{n^k,1}(u_2, v_2)) \\ &\quad \le \delta_{d/k}\Upsilon_{n^k,2}(u_1, u_2, v_1, v_2). \end{aligned} \tag{32}$$

If on the contrary, $n \le \delta_{d/k}(u_1, u_2, v_1, v_2)$, then the inequality (32) straightforwardly holds. Since $\mu \otimes \mu(D_{d,2}) = 0$, one then obtains

$$\begin{aligned} n\mathbb{V}(n^{rk/d}\mathcal{D}^{\nu}_{\mu,n^k,r}) &\le I_{\mu,2}[\delta_{d/k}\Upsilon_{n^k,2}] \\ &\le I_{\mu,2}[\mathbb{1}_{\mathcal{E}_{n^k}}\delta_{d/k}\Upsilon_{n^k,2}] + I_{\mu,2}[\mathbb{1}_{\complement\mathcal{E}_{n^k}}\delta_{d/k}\Upsilon_{n^k,2}]. \end{aligned}$$

Assumption L9.1 then yields

$$n\mathbb{V}(n^{rk/d}\mathcal{D}^{\nu}_{\mu,n^k,r}) \le I_{\mu,2}[\delta_{d/k}\psi] + I_{\mu,2}[\mathbb{1}_{\complement\mathcal{E}_{n^k}}\delta_{d/k}\Upsilon_{n^k,2}]. \tag{33}$$

Therefore, for sufficiently large $n$,

$$n\mathbb{V}(n^{rk/d}\mathcal{D}^{\nu}_{\mu,n^k,r}) \le I_{\mu,2}[\delta_{d/k}\psi] + \limsup I_{\mu,2}[\mathbb{1}_{\complement\mathcal{E}_{n^k}}\delta_{d/k}\Upsilon_{n^k,2}] + 1. \tag{34}$$

However, since $\varepsilon \ge d/k$, Assumptions L9.2 and L9.3 easily yield that $I_{\mu,2}[\delta_{d/k}\psi] < +\infty$ and that $\limsup I_{\mu,2}[\mathbb{1}_{\complement\mathcal{E}_{n^k}}\delta_{d/k}\Upsilon_{n^k,2}] < +\infty$.

Then, from (34), one obtains $\limsup n\mathbb{V}(n^{rk/d}\mathcal{D}^{\nu}_{\mu,n^k,r}) < +\infty$ which implies (31) and completes the proof. □

LEMMA 10. *Let $\alpha > 0$, $l \in \mathbb{R}$, $k \in \mathbb{N}^*$ and let $(a_n)_{n>0}$ be a decreasing sequence of real numbers such that $n^{k\alpha}a_{n^k} \underset{n\to+\infty}{\longrightarrow} l$. Then $n^{\alpha}a_n \underset{n\to+\infty}{\longrightarrow} l$.*

The proof is straightforward.

LEMMA 11. *Under the assumptions of Theorem 2, one has*

$$I_{\mu,2}\left[\delta_{\varepsilon}(u_1, u_2, v_1, v_2)\sum_{j=1}^{2}\exp\bigl(-B_d g_{\|u_1\|\vee\|u_2\|+\rho}(u_j)v_1^{d/r} \vee v_2^{d/r}\bigr)\right] < +\infty.$$

For the proof, the only difference with Lemma 4 is the additional term $\delta_{\varepsilon}(u_1, u_2, v_1, v_2)$, which is easily handled by the existence of a bounded $\varepsilon$-potential for $\mu$.

LEMMA 12. *Let $d \in \mathbb{N}^*$, $r > 0$, $\varepsilon > 0$, $\rho > 0$. Assume that*:

L12.1. *The function $c \to \int_{\mathbb{R}^d} \frac{\mu(du)}{\|u-c\|^{\varepsilon}}$ is bounded over $\mathbb{R}^d$.*
L12.2. $\int_{\mathbb{R}^d \|u\|^{2r+\varepsilon}} (\mu + \nu)(du) < +\infty$.

*Then*

$$I_{\mu,2}\big[\mathbb{1}_{\complement A^{(2)}_{n,\rho}} \delta_\varepsilon \Upsilon_{n,2}\big] \to 0.$$

The proof is similar to the proof of Lemma 6.

4.7.2. *Lemmas for Theorem* 3.

LEMMA 13. *Let $r > 1$, let $\mu, \nu \in \mathcal{P}$ and let $f$ denote the density function of the probability measure $\mu$. For every $n \ge 1$ and every $1 \le k \le n-1$, let $\zeta_{k,n}$ be a random variable taking values in $[Z_{k,n}, Z_{k+1,n}]$. Then*

$$n^{1/2}\left(n^r \mathcal{D}^{\nu}_{\mu,n,r} - \frac{1}{2^r(r+1)} \frac{1}{n} \sum_{k=1}^{n-1} f(\zeta_{k,n})\big(n(Z_{k+1,n} - Z_{k,n})\big)^{r+1}\right) \xrightarrow{\mathbb{P}} 0.$$

The proof follows immediately from the fact that $f$ is Lipschitz and from the elementary extreme values and spacings theory.

The following lemma is the extension of the Pyke–Hall method used in Theorem 3. We use the notation

$$\Sigma_{k,n} := \sum_{j=1}^{k} 1/(n-j+1),$$

$$\chi_{k,n} := \sum_{j=1}^{k} X_j/(n-j+1),$$

$$\lambda_{k,n} := -\log(1-k/n).$$

LEMMA 14. *Let $r > 1$ and let $\nu \in \mathcal{P}$ with density function $g$. For every $n \ge 1$ and every $1 \le k \le n-1$, let $\theta_{k,n}$ be the $[0,1]$-valued random variable satisfying*

$$\frac{X_{k+1}}{n-k} K'_g\left(\chi_{k,n} + \theta_{k,n}\frac{X_{k+1}}{n-k}\right) = Z_{k+1,n} - Z_{k,n}. \tag{35}$$

*Let $\varphi \in \mathcal{M}$, let $\psi \in \mathcal{P}$ and define*

$$H_{\varphi,\psi} := \frac{\exp(-\cdot)}{\varphi \circ K_\psi}.$$

*Then*

$$n^{1/2}\big(\big(S_n^{(1)} - S\big) - \big(S_n^{(2)} - \mathbb{E}\big(S_n^{(2)}\big)\big)\big) \xrightarrow{\mathbb{P}} 0, \tag{36}$$

*where*

$$S = \int_0^{+\infty} dz\, \exp(-z) \int_0^1 dx \left( \frac{z}{1-x} H_{\varphi,\psi}(-\log(1-x)) \right)^{r+1}, \tag{37}$$

$$S_n^{(1)} := \frac{1}{n} \sum_{k=1}^{n-1} \left( \frac{n}{n-k} X_{k+1} H_{\varphi,\psi}\left( \chi_{k,n} + \theta_{k,n} \frac{X_{k+1}}{n-k} \right) \right)^{r+1} \tag{38}$$

*and*

$$\begin{aligned} S_n^{(2)} = \frac{1}{n} \sum_{k=1}^{n-1} & \frac{n}{n-k} (\chi_{k,n} - \Sigma_{k,n}) H'_{\varphi,\psi}(\lambda_{k,n}) \\ & \times \Gamma(r+2)(r+1) \left( \frac{n}{n-k} H_{\varphi,\psi}(\lambda_{k,n}) \right)^r \\ & + \left( \frac{n}{n-k} X_{k+1} H_{\varphi,\psi}(\lambda_{k,n}) \right)^{r+1}. \end{aligned} \tag{39}$$

PROOF. A careful reading of Hall's proof of Proposition 1 (derived in [12] in a more complex technical setting) reveals that the only properties of $K'_g = \exp(-\cdot)/g \circ K_g$ required to prove this result are those of the functions belonging to the set $\{\exp(-\cdot)/\varphi \circ K_\psi, \varphi \in \mathcal{M}, \psi \in \mathcal{P}\}$ (in particular, the fact that $K'_g$ is the derivative of $K_g$ is not used).

However, the statement of Lemma 14 is precisely obtained by replacing mutatis mutandis the function $K'_g$ in Proposition 1 by a function $H_{\varphi,\psi}$. So the proof of Lemma 14 is very similar to the proof of Proposition 1 and we give only a brief sketch of it.

The function $H_{\varphi,\psi}$ is continuously differentiable on $\mathbb{R}_+$. So applying twice the mean value theorem, one can check that

$$S_n^{(1)} = \frac{1}{n} \sum_{k=1}^{n-1} A_{n,k}^{(1)} + \frac{1}{n} \sum_{k=1}^{n-1} A_{n,k}^{(2)},$$

where

$$A_{n,k}^{(1)} := \left( \frac{n}{n-k} X_{k+1} H_{\varphi,\psi}(\Sigma_{k,n}) \right)^{r+1},$$

$$A_{n,k}^{(2)} := (r+1) \frac{n}{n-k} X_{k+1} \Delta_{n,k} \left( \frac{n}{n-k} X_{k+1} [H_{\varphi,\psi}(\Sigma_{k,n}) + \xi_{n,k} \Delta_{n,k}] \right)^r$$

with

$$\begin{aligned} \Delta_{n,k} := & \left( \chi_{k,n} - \Sigma_{k,n} + \theta_{n,k} \frac{X_{k+1}}{n-k} \right) \\ & \times H'_{\varphi,\psi}\left( \Sigma_{k,n} + \eta_{n,k} \chi_{k,n} - \Sigma_{k,n} + \eta_{n,k} \theta_{n,k} \frac{X_{k+1}}{n-k} \right) \end{aligned}$$

and where $\eta_{n,k}$ and $\xi_{n,k}$ are some $[0,1]$-valued random variable [defined on $(\Omega, \mathcal{A}, \mathbb{P})$].

Then one can prove the following approximations:

$$\begin{aligned} n^{-1/2}\sum_{k=1}^{n-1} A_n^{(2)} - \Big[ \frac{n}{n-k} X_{k+1}(\chi_{k,n} - \Sigma_{k,n}) H'_{\varphi,\psi}(\lambda_{k,n}) \\ \times (r+1)\Big(\frac{n}{n-k} X_{k+1} H_{\varphi,\psi}(\lambda_{k,n})\Big)^r \Big] \xrightarrow{\mathbb{P}} 0, \end{aligned} \tag{40}$$

$$\begin{aligned} n^{-1/2}\sum_{k=1}^{n-1} (\chi_{k,n} - \Sigma_{k,n}) H'_{\varphi,\psi}(\lambda_{k,n}) \\ \times \Big[ \frac{n}{n-k} X_{k+1}\Big(\frac{n}{n-k} X_{k+1} H_{\varphi,\psi}(\lambda_{k,n})\Big)^r \\ - \mathbb{E}\Big(\frac{n}{n-k} X_{k+1}\Big(\frac{n}{n-k} X_{k+1} H_{\varphi,\psi}(\lambda_{k,n})\Big)^r\Big)\Big] \xrightarrow{\mathbb{P}} 0, \end{aligned} \tag{41}$$

$$n^{-1/2}\sum_{k=1}^{n-1} A_{n,k}^{(1)} - \Big(\frac{n}{n-k} X_{k+1} H_{\varphi,\psi}(\lambda_{k,n})\Big)^{r+1} \xrightarrow{\mathbb{P}} 0, \tag{42}$$

$$n^{1/2}\bigg(S - \frac{1}{n}\sum_{k=1}^{n-1} \mathbb{E}\Big(\Big(\frac{n}{n-k} X_{k+1} H_{\varphi,\psi}(\lambda_{k,n})\Big)^{r+1}\Big)\bigg) \to 0. \tag{43}$$

Finally, combining (40)–(43), one obtains

$$n^{1/2}\big((S_n^{(1)} - S) - (S_n^{(2)} - \mathbb{E}(S_n^{(2)}))\big) \xrightarrow{\mathbb{P}} 0,$$

which completes the proof. $\square$

LEMMA 15. *Under the notation and assumptions of Lemma* 14,

$$n^{1/2}\big(S_n^{(2)} - \mathbb{E}(S_n^{(2)})\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathfrak{S}^2_{\varphi,\psi}),$$

*where*

$$\begin{aligned} \mathfrak{S}^2_{\varphi,\psi} = \int_0^1 dx \Big(\frac{1}{1-x}\Big)^2 \Big(\int_x^1 \Phi(t)\,dt\Big)^2 + \big(\Gamma(2r+3) - \Gamma(r+2)^2\big)\Psi(x)^2 \\ + \frac{2}{1-x}\big(\Gamma(r+3) - \Gamma(r+2)\big)\Psi(x) \int_x^1 \Phi(t)\,dt \end{aligned}$$

*with*

$$\Phi(t) := \frac{1}{1-t} H'_{\varphi,\psi}\big(-\log(1-t)\big)\Gamma(r+2)(r+1)\Big(\frac{1}{1-t} H_{\varphi,\psi}\big(-\log(1-t)\big)\Big)^r$$

*and*

$$\Psi(t) := \left(\frac{1}{1-t} H_{\varphi,\psi}(-\log(1-t))\right)^{r+1}.$$

For the proof, reordering the summation in $S_n^{(2)}$ and Lindeberg's theorem for triangular arrays (see, e.g., [13]) yield the result.

## APPENDIX

**Rényi's version of order statistics.** Let $\nu$ be a probability measure on $\mathbb{R}$, let $(Y_i)_{1\le i\le n}$ be a $\nu$-distributed i.i.d. sample and let $(\widehat{Y}_{k,n})_{1\le k\le n}$ be the related order statistics. Rényi's version of order statistics is a version of the vector $(\widehat{Y}_{k,n})_{1\le k\le n}$ constructed from an i.i.d. standard exponential sequence $(X_j)_{1\le j\le n}$. More precisely, letting $(\widehat{X}_{k,n})_{1\le k\le n}$ be the order statistics of $(X_j)_{1\le j\le n}$, it is classical that

$$(\widehat{X}_j)_{1\le j\le n} \stackrel{\mathcal{L}}{=} \left(\sum_{j=1}^{k} \frac{X_j}{n-j+1}\right)_{1\le k\le n}.$$

Then one obtains (Rényi's version of order statistics)

$$G^{-1}\left(1-\exp\left(\sum_{j=1}^{k}\frac{X_j}{n-j+1}\right)\right)_{1\le k\le n} \stackrel{\mathcal{L}}{=} (\widehat{Y}_{k,n})_{1\le k\le n}, \tag{44}$$

where $G$ denotes the distribution function of the probability measure $\nu$. In this paper, for notational convenience, we set $K_g := G^{-1}(1-\exp(\cdot))$.

**Acknowledgments.** This work is part of the author's Ph.D. dissertation at the University of Paris 6, written under the supervision of Professor G. Pagès whose guidance and suggestions are gratefully acknowledged and appreciated. Thanks are also due to the referees for useful suggestions and corrections.

## REFERENCES

[1] BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1990). *Adaptative Algorithms and Stochastic Approximations.* Springer, Berlin.

[2] BOUTON, C. and PAGÈS, G. (1994). Convergence in distribution of the one-dimensional Kohonen algorithms when the stimuli are not uniform. *Adv. in Appl. Probab.* **26** 80–103.

[3] BOUTON, C. and PAGÈS, G. (1997). About the multidimensional competitive learning vector quantization algorithm with constant gain. *Ann. Appl. Probab.* **7** 679–710.

[4] BUCKLEW, J. A. (1981). Companding and random quantization in several dimensions. *IEEE Trans. Inform. Theory* **27** 207–211.

[5] BUCKLEW, J. A. and WISE, G. (1982). Multidimensional asymptotic quantization theory with $r$th power distortion measures. *IEEE Trans. Inform. Theory* **28** 239–247.

[6] CONWAY, J. H. and SLOANE, J. A. (1982). Voronoï regions of lattices, second moments of polytopes and quantization. *IEEE Trans. Inform. Theory* **28** 211–226.

[7] FORT, J. C. and PAGÈS, G. (1995). About the a.s. convergence of the Kohonen algorithm with a general neighborhood function. *Ann. Appl. Probab.* **5** 1177–1216.

[8] GERSHO, A. (1979). Asymptotically optimal block quantization. *IEEE Trans. Inform. Theory* **25** 373–380.

[9] GERSHO, A. and GRAY, R. M. (1992). *Vector Quantization and Signal Compression*, 7th ed. Kluwer, Dordrecht.

[10] GRAF, S. and LUSCHGY, H. (1999). *Foundations of Quantization for Probability Distributions*. Springer, Berlin.

[11] GRAY, R. M. and NEUHOFF, D. L. (1998). Quantization. *IEEE Trans. Inform. Theory* **44** 2325–2383.

[12] HALL, P. (1984). Limit theorems for sums of general functions of $m$-spacing. *Math. Proc. Cambridge Philos. Soc.* **96** 517–532.

[13] HALL, P. and HEYDE, C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

[14] LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inform. Theory* **28** 129–137.

[15] NA, S. and NEUHOFF, D. L. (1995). Bennett's integral for vector quantizer. *IEEE Trans. Inform. Theory* **41** 886–899.

[16] PAGÈS, G. (1997). A space vector quantization for numerical integration. *J. Appl. Comput. Math.* **89** 1–38.

[17] PAGÈS, G., BALLY, V. and PRINTEMS, J. (2001). A stochastic quantization method for nonlinear problems. *Monte Carlo Methods Appl.* **7** 21–34.

[18] PYKE, R. (1965). Spacings. *J. Roy. Statist. Soc. Ser. B* **27** 395–449.

[19] SABIN, M. J. and GRAY, R. M. (1986). Global convergence and empirical consistency of the generalized Lloyd algorithm. *IEEE Trans. Inform. Theory* **32** 148–155.

[20] TRUSHKIN, A. V. (1993). On the design of an optimal quantizer. *IEEE Trans. Inform. Theory* **39** 1180–1194.

[21] ZADOR, P. L. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory* **28** 139–149.

ENPC-CERMICS
6-8 AVENUE BLAISE PASCAL
CITE DESCARTES-CHAMPS SUR MARNE
F-77455 MARNE LA VALLEE
FRANCE
E-MAIL: cohort@cermics.enpc.fr