Rényi Entropy and Quantization for Densities

Christoph Bunte and Amos Lapidoth
ETH Zurich
Switzerland
Email: {bunte,lapidoth}@isi.ee.ethz.ch

***Abstract*—A random variable Z taking value in a finite, nonatomic measure space $(X, \mathcal{M}, \mu)$ and whose distribution is absolutely continuous with respect to μ is to be described using N labels. We seek the labeling that minimizes the ρ-th moment of the μ-volume of the set of points in X that have the same label as Z. The large-N asymptotics of this minimum are expressed in terms of the Rényi entropy of order $1/(1+\rho)$.**

I. Introduction and Results

Consider a random variable (RV) Z taking value in a measure space $(X, \mathcal{M}, \mu)$, where the measure μ is finite (i.e., $\mu(X) < \infty$) and nonatomic [1]. Assume that Z has a density f with respect to μ, i.e.,

$$\Pr(Z \in A) = \int_A f \, d\mu, \quad \text{for all } A \in \mathcal{M}. \tag{1}$$

For example, X could be a finite interval, μ the Lebesgue measure, and Z a continuous RV with density f.

We wish to describe Z using a quantizer, i.e., a measurable mapping of the form

$$\varphi\colon X \to \{1, \ldots, N\}, \tag{2}$$

where N is a given positive integer.

The description of Z is $\varphi(Z)$, and the cell containing Z comprises the elements of X that have the same description as Z, i.e., $\varphi^{-1}(\varphi(Z))$. The μ-volume of this cell is thus $\mu(\varphi^{-1}(\varphi(Z)))$, and we seek a quantizer φ that minimizes its ρ-th moment

$$\mathrm{E}\big[\mu\big(\varphi^{-1}(\varphi(Z))\big)^{\rho}\big] = \int_X \mu\big(\varphi^{-1}(\varphi(z))\big)^{\rho} f(z) \, d\mu(z), \tag{3}$$

where ρ is a given positive number. (Replacing $\varphi^{-1}(\varphi(Z))$ with $\varphi^{-1}(\varphi(Z)) \cap \operatorname{supp}(f)$, where $\operatorname{supp}(f) = \{z \in X : f(z) > 0\}$, makes little difference; see the remark at the end of this section.)

The problem of encoding tasks studied in [2] corresponds to X being finite and μ being the counting measure (and hence not nonatomic). The asymptotics we study here are, however, different.

We derive upper and lower bounds on the minimum ρ-th moment of $\mu(\varphi^{-1}(\varphi(Z)))$ and study its large-N asymptotics. The Rényi entropy of Z of order $1/(1+\rho)$ will be key. Recall that the Rényi entropy of Z of order α is defined (for positive α other than one) as

$$h_\alpha(Z) = \frac{1}{1-\alpha} \log \int_X f(z)^{\alpha} \, d\mu(z). \tag{4}$$

(Throughout $\log$ denotes the natural logarithm.)

A trivial upper bound can be obtained by partitioning X into N subsets (cells) of μ-volume $\mu(X)/N$ each[1], labeling them with the numbers $\{1, \ldots, N\}$, and setting $\varphi(z) = n$ if z is a member of the n-th subset. This quantizer achieves

$$\mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^{\rho}] = \frac{\mu(X)^{\rho}}{N^{\rho}}, \tag{5}$$

and the right-hand side (RHS) thus upper-bounds the minimum.

Intuition suggests that (5) is the best we can do when Z is uniformly distributed over X (and Theorem I.1 ahead proves this). For other distributions, however, we expect to do better (and Theorem I.2 ahead shows that we can).

Let $\mathcal{F}(X, N)$ denote the set of all measurable mappings from X into $\{1, \ldots, N\}$. What we have shown so far is

$$\min_{\varphi \in \mathcal{F}(X,N)} \mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^{\rho}]^{1/\rho} \leq \frac{\mu(X)}{N}. \tag{6}$$

(Writing $\min$ instead of $\inf$ is justified; see Section III.)

Our first result is the following lower bound.

Theorem I.1. *For every $N \in \mathbb{N}$ and every $\rho > 0$,*

$$\min_{\varphi \in \mathcal{F}(X,N)} \mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^{\rho}]^{1/\rho} \geq \frac{e^{h_{\frac{1}{1+\rho}}(Z)}}{N}. \tag{7}$$

We note that this result combined with (6) implies that

$$h_{\frac{1}{1+\rho}}(Z) \leq \log \mu(X), \tag{8}$$

i.e., that among all distributions on X that are absolutely continuous with respect to μ, the uniform distribution maximizes the Rényi entropy of any order in $(0, 1)$. (See [4] on the maximization of Rényi entropy subject to constraints.)

Proof of Theorem I.1: We use Hölder's inequality (see, e.g., [1, Thm. 5.1.2]): If $p, q > 1$ and $1/p + 1/q = 1$, then for any measurable, nonnegative functions g and h,

$$\int_X g(z)h(z)d\mu(z) \leq \Big(\int_X g(z)^p d\mu(z)\Big)^{1/p} \Big(\int_X h(z)^q d\mu(z)\Big)^{1/q}. \tag{9}$$

If the rightmost integral is positive and finite, we can rearrange to obtain

$$\int_X g(z)^p d\mu(z) \geq \Big(\int_X g(z)h(z)d\mu(z)\Big)^{p} \Big(\int_X h(z)^q d\mu(z)\Big)^{-p/q}. \tag{10}$$

[1] This is possible because μ is nonatomic [3].

The choice

$$p = 1 + \rho, \tag{11a}$$
$$q = (1+\rho)/\rho, \tag{11b}$$
$$g(z) = f(z)^{\frac{1}{1+\rho}} \mu(\varphi^{-1}(\varphi(z)))^{\frac{\rho}{1+\rho}}, \tag{11c}$$
$$h(z) = \mu(\varphi^{-1}(\varphi(z)))^{-\frac{\rho}{1+\rho}}, \tag{11d}$$

gives

$$\begin{aligned} &\int_X \mu\big(\varphi^{-1}(\varphi(z))\big)^\rho f(z) d\mu(z) \\ &\geq \Big(\int_X f(z)^{\frac{1}{1+\rho}} d\mu(z)\Big)^{1+\rho} \Big(\int_X \frac{1}{\mu\big(\varphi^{-1}(\varphi(z))\big)} d\mu(z)\Big)^{-\rho}. \end{aligned} \tag{12}$$

Observe further that

$$\begin{aligned} \int_X \frac{1}{\mu\big(\varphi^{-1}(\varphi(z))\big)} d\mu(z) &= \sum_{n=1}^N \int_{\varphi^{-1}(n)} \frac{1}{\mu\big(\varphi^{-1}(n)\big)} d\mu(z) \\ &= \sum_{n:\mu(\varphi^{-1}(n))>0} 1 \\ &\leq N. \end{aligned} \tag{13}$$

Combining (13) and (12) completes the proof. ∎

It follows from the conditions for equality in Hölder's inequality that (12) holds with equality if, and only if, $\mu(\varphi^{-1}(\varphi(z)))$ is proportional to $f(z)^{-1/(1+\rho)}$ μ-almost everywhere (μ-a.e.). In particular, this is possible only if f is equal μ-a.e. to a function taking on a finite number of different values.

As our next result shows, the lower bound in Theorem I.1 is asymptotically achievable when N is large. The result is reminiscent of the asymptotic result of [5, Thm. 6.2], which concerns the ρ-th moment of the quantization error with respect to a norm on $\mathbb{R}^d$. In our setting, however, X need not be a subset of $\mathbb{R}^d$, and it need not be endowed with a norm or a metric.

Theorem I.2. *For any $\rho > 0$,*

$$\lim_{N\to\infty} N \min_{\varphi \in \mathcal{F}(X,N)} \mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^\rho]^{1/\rho} = e^{h_{\frac{1}{1+\rho}}(Z)}. \tag{14}$$

In view of Theorem I.1, the limit $N \to \infty$ may be replaced with the infimum over $N \in \mathbb{N}$. The proof is in Section II.

Note that, in general, it is necessary to let $N \to \infty$ to achieve the lower bound in Theorem I.1. In fact, the convergence may be quite slow, as the following example illustrates: Let $X = [0,1]$, $\mathcal{M}$ the Borel sets, μ the Lebesgue measure, and $f(x) = 3x^2$. It can be shown [6] that

$$\min_{\varphi \in \mathcal{F}(X,N)} \mathrm{E}\big[\mu(\varphi^{-1}(\varphi(Z)))\big] \geq \frac{e^{h_{1/2}(Z)}}{N} + \frac{1}{4N^3}. \tag{15}$$

Also note that Theorem I.2 combined with the fact that $(\mathrm{E}[Y^\rho])^{1/\rho}$ is nondecreasing in ρ for any nonnegative RV Y (see, e.g., [7, p.193]) implies the well-known result [8] that $h_{\frac{1}{1+\rho}}(Z)$ is nondecreasing in $\rho > 0$.

We can extend Theorem I.2 to σ-finite nonatomic measure spaces as follows.

Corollary I.3. *Let the RV Z take value in a σ-finite, nonatomic measure space $(X, \mathcal{M}, \mu)$ and have a density f with respect to μ. Let $(V_k)_{k\geq 1}$ be an increasing sequence in $\mathcal{M}$ such that $\mu(V_k) < \infty$ and $\bigcup_{k=1}^\infty V_k = X$. Let Z_k be a RV on V_k with density $f_k(z) = f(z)/\int_{V_k} f d\mu$. Then*

$$\lim_{k\to\infty} \lim_{N\to\infty} \min_{\varphi\in\mathcal{F}(V_k,N)} N\, \mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^\rho]^{1/\rho} = e^{h_{\frac{1}{1+\rho}}(Z)}. \tag{16}$$

Proof: In view of Theorem I.2 we only need to show that $\lim_{k\to\infty} h_{\frac{1}{1+\rho}}(Z_k) = h_{\frac{1}{1+\rho}}(Z)$. This is straightforward:

$$\begin{aligned} &h_{\frac{1}{1+\rho}}(Z_k) \\ &= \frac{1+\rho}{\rho} \log \int_{V_k} f(z)^{\frac{1}{1+\rho}} d\mu(z) - \frac{1}{\rho} \log \int_{V_k} f(z) d\mu(z) \\ &\to \frac{1+\rho}{\rho} \log \int_X f(z)^{\frac{1}{1+\rho}} d\mu(z), \quad (k\to\infty), \end{aligned} \tag{17}$$

where (17) follows from the Monotone Convergence Theorem. ∎

We conclude this section with a remark: The reader may object to minimizing the ρ-th moment of $\mu(\varphi^{-1}(\varphi(Z)))$ for the following reason. Conditioned on the observation $\varphi(Z) = n$, with probability 1 Z lies in the set

$$\varphi^{-1}(n) \cap \mathrm{supp}(f), \tag{18}$$

where $\mathrm{supp}(f) = \{z \in X : f(z) > 0\}$, and this set may have strictly smaller μ-volume than $\varphi^{-1}(n)$.

While this is a valid point, replacing everywhere $\varphi^{-1}(\varphi(Z))$ by $\varphi^{-1}(\varphi(Z)) \cap \mathrm{supp}(f)$ will not change our results fundamentally. Indeed, Theorem I.1 remains true, as can be seen by restricting the domain of integration to $X \cap \mathrm{supp}(f)$ everywhere in the proof. And this implies that also Theorem I.2 continues to hold because the achievability only becomes easier.

II. Proof of Theorem I.2

In this section we present a proof of Theorem I.2. One direction—the "converse" part—is an immediate consequence of Theorem I.1. To prove the other direction—the "direct" part—we shall first prove an analogous result for nonnegative simple functions that do not necessarily integrate to 1 (see (20) ahead). We then approximate arbitrary densities by sequences of simple functions.

Suppose then that f is a measurable, nonnegative, simple function, i.e.,

$$f(z) = \sum_{i=1}^k a_i\, \chi_{\mathcal{D}_i}(z), \quad z \in X, \tag{19}$$

where $\chi_{\mathcal{D}_i}$ is the indicator function of the set $\mathcal{D}_i$, where the a_i's are nonnegative numbers, and where the sets $\mathcal{D}_i \in \mathcal{M}$ form a partition of X, i.e., $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset$ whenever $i \neq j$ and $\bigcup_{i=1}^k \mathcal{D}_i = X$. (Remember that f need not integrate to 1.)

We construct $\varphi \in \mathcal{F}(X,N)$ such that

$$\lim_{N\to\infty} N\Big(\int_X f(z)\mu(\varphi^{-1}(\varphi(z)))^\rho d\mu(z)\Big)^{1/\rho} = \Big(\int_X f(z)^{\frac{1}{1+\rho}} d\mu(z)\Big)^{\frac{1+\rho}{\rho}}. \quad (20)$$

To this end, let

$$\mathcal{I} = \{i \in \{1,\dots,k\} : a_i\,\mu(\mathcal{D}_i) > 0\}. \quad (21)$$

We may assume that $\mathcal{I}$ is not empty, for otherwise $f = 0$ μ-a.e., and (20) clearly holds.

For each $i \in \mathcal{I}$ partition the set $\mathcal{D}_i$ into

$$\big\lfloor (N-1)a_i^{\frac{1}{1+\rho}}\gamma^{-1}\mu(\mathcal{D}_i)\big\rfloor \quad (22)$$

subsets of equal measure[2], where

$$\gamma = \sum_{i\in\mathcal{I}} \mu(\mathcal{D}_i)a_i^{\frac{1}{1+\rho}}, \quad (23)$$

and N is assumed large enough so that (22) be at least 2 for every $i \in \mathcal{I}$.

The choice (22) with γ as in (23) guarantees that the total number of subsets does not exceed $N-1$. Indeed,

$$\sum_{i\in\mathcal{I}} \big\lfloor (N-1)a_i^{\frac{1}{1+\rho}}\gamma^{-1}\mu(\mathcal{D}_i)\big\rfloor \le \sum_{i\in\mathcal{I}} (N-1)a_i^{\frac{1}{1+\rho}}\gamma^{-1}\mu(\mathcal{D}_i) = N-1. \quad (24)$$

Consequently, we can label the different subsets with the integers 1 through $N-1$ (or fewer) and set $\varphi(z) = n$ if z is in the n-th subset. This defines $\varphi(z)$ for all $z \in \bigcup_{i\in\mathcal{I}}\mathcal{D}_i$. For all $z \in \bigcup_{i\notin\mathcal{I}}\mathcal{D}_i$ we set $\varphi(z) = N$. Then $\varphi \in \mathcal{F}(X,N)$. Moreover, if $z \in \mathcal{D}_i$ and $i \in \mathcal{I}$,

$$\mu(\varphi^{-1}(\varphi(z))) = \frac{\mu(\mathcal{D}_i)}{\big\lfloor (N-1)a_i^{\frac{1}{1+\rho}}\gamma^{-1}\mu(\mathcal{D}_i)\big\rfloor}. \quad (25)$$

Using the inequality $\lfloor \xi \rfloor > \xi - 1$, we thus have

$$\begin{aligned}
&\int_X f(z)\mu(\varphi^{-1}(\varphi(z)))^\rho d\mu(z) \\
&= \sum_{i\in\mathcal{I}} \int_{\mathcal{D}_i} a_i\,\mu(\varphi^{-1}(\varphi(z)))^\rho d\mu(z) \\
&< \sum_{i\in\mathcal{I}} \mu(\mathcal{D}_i)a_i\left(\frac{\mu(\mathcal{D}_i)}{(N-1)a_i^{\frac{1}{1+\rho}}\gamma^{-1}\mu(\mathcal{D}_i)-1}\right)^\rho \\
&= \frac{1}{N^\rho}\sum_{i\in\mathcal{I}} \mu(\mathcal{D}_i)a_i\left(\frac{\mu(\mathcal{D}_i)}{\frac{N-1}{N}a_i^{\frac{1}{1+\rho}}\gamma^{-1}\mu(\mathcal{D}_i)-\frac{1}{N}}\right)^\rho. \quad (26)
\end{aligned}$$

It follows from (26) and (23) that

$$\begin{aligned}
&\overline{\lim_{N\to\infty}} N\Big(\int_X f(z)\mu(\varphi^{-1}(\varphi(z)))^\rho d\mu(z)\Big)^{1/\rho} \\
&\le \Big(\sum_{i\in\mathcal{I}}^{n} \mu(\mathcal{D}_i)a_i^{\frac{1}{1+\rho}}\Big)^{\frac{1+\rho}{\rho}} \\
&= \Big(\int_X f(z)^{\frac{1}{1+\rho}} d\mu(z)\Big)^{\frac{1+\rho}{\rho}}. \quad (27)
\end{aligned}$$

This, in fact, shows that (20) holds because (7) is true not just for densities (and hence expectations), but more generally for any measurable nonnegative f:

$$\Big(\int_X f(z)\mu(\varphi^{-1}(\varphi(z)))^\rho d\mu(z)\Big)^{1/\rho} \ge \frac{1}{N}\Big(\int_X f(z)^{\frac{1}{1+\rho}} d\mu(z)\Big)^{\frac{1+\rho}{\rho}}. \quad (28)$$

That this is true can be gleaned from the proof of Theorem I.1 (see (12) and (13)).

Suppose now that f is an arbitrary density on X, i.e., f is measurable, nonnegative, and $\int_X f d\mu = 1$. We may assume that $h_{\frac{1}{1+\rho}}(f) < \infty$, for otherwise there is nothing to prove.

We begin by constructing a sequence of measurable, nonnegative, simple functions converging to f μ-a.e. on X. To this end, for every positive integer k we define the sets

$$\mathcal{E}_{k,i} = \{z \in X : (i-1)2^{-k} \le f(z) < i2^{-k}\} \quad (29)$$

for $i \in \{1,\dots,2^{2k}\}$, and

$$\mathcal{E}_k = \{z \in X : f(z) \ge 2^k\}. \quad (30)$$

We then define

$$f_k(z) = \sum_{i=1}^{2^{2k}} a_i\,\chi_{\mathcal{E}_{k,i}}(z), \quad z \in X, \quad (31)$$

where

$$a_i = \begin{cases} \mu(\mathcal{E}_{k,i})^{-1}\int_{\mathcal{E}_{k,i}} f d\mu, & \text{if } i \in \mathcal{I}_k, \\ 0 & \text{otherwise,} \end{cases} \quad (32)$$

and $\mathcal{I}_k = \big\{i \in \{1,\dots,2^{2k}\} : \mu(\mathcal{E}_{k,i}) > 0\big\}$. Let $\mathcal{E} = \bigcup_{k=1}^\infty \bigcup_{i\notin\mathcal{I}_k} \mathcal{E}_{k,i}$ and note that $\mu(\mathcal{E}) = 0$. By construction of f_k,

$$|f_k(z) - f(z)| \le 2^{-k}, \quad \text{if } z \in X\setminus\mathcal{E} \text{ and } f(z) < 2^k.$$

Thus, $f_k \to f$ μ-a.e. on X. Moreover, since $f_k(z) \le f(z) + 2^{-k}$ for all $z \in X$, the Dominated Convergence Theorem yields

$$\Big(\int_X f_k(z)^{\frac{1}{1+\rho}} d\mu(z)\Big)^{\frac{1+\rho}{\rho}} \to e^{h_{\frac{1}{1+\rho}}(f)}, \quad (k\to\infty). \quad (33)$$

Indeed, if $\int_X f(z)^{\frac{1}{1+\rho}} d\mu(z) < \infty$, as is implied by our assumption that $h_{\frac{1}{1+\rho}}(f) < \infty$, then we also have $\int_X (f(z)+\delta)^{\frac{1}{1+\rho}} d\mu(z) < \infty$ for any $\delta > 0$ because

$$\begin{aligned}
(f(z)+\delta)^{\frac{1}{1+\rho}} &\le \big(2\max\{f(z),\delta\}\big)^{\frac{1}{1+\rho}} \\
&= 2^{\frac{1}{1+\rho}}\max\{f(z)^{\frac{1}{1+\rho}}, \delta^{\frac{1}{1+\rho}}\} \\
&\le 2^{\frac{1}{1+\rho}}\big(f(z)^{\frac{1}{1+\rho}} + \delta^{\frac{1}{1+\rho}}\big). \quad (34)
\end{aligned}$$

Next, we fix some $\alpha \in (0,1)$ and construct $\varphi_k \in \mathcal{F}(X,\lfloor\alpha N\rfloor)$ for f_k as we did for nonnegative simple functions above (with the $\mathcal{D}_i$'s replaced by the $\mathcal{E}_{k,i}$'s and $\mathcal{E}_k$, and with N replaced by $\lfloor\alpha N\rfloor$). Then by (20)

$$\lim_{N\to\infty} \lfloor\alpha N\rfloor\Big(\int_X f_k(z)\mu(\varphi_k^{-1}(\varphi_k(z)))^\rho d\mu(z)\Big)^{1/\rho} = \Big(\int_X f_k(z)^{\frac{1}{1+\rho}} d\mu(z)\Big)^{\frac{1+\rho}{\rho}}. \quad (35)$$

[2]Here we are using again that μ is nonatomic [3].

Next, we construct $\tilde{\varphi}_k$ by setting $\tilde{\varphi}_k(z) = \varphi_k(z)$ for all $z \in X \setminus \mathcal{E}_k$. To define $\tilde{\varphi}_k$ on $\mathcal{E}_k$, we partition $\mathcal{E}_k$ into $\lfloor(1-\alpha)N\rfloor$ subsets of equal μ-volume, label them with the numbers $\lfloor \alpha N\rfloor + 1$ through N and set $\tilde{\varphi}_k(z) = n$ if z is in the n-th subset of $\mathcal{E}_k$. Then $\tilde{\varphi}_k \in \mathcal{F}(X, N)$ and

$$
\begin{aligned}
&\int_X f(z)\mu(\tilde{\varphi}_k^{-1}(\tilde{\varphi}_k(z)))^\rho d\mu(z) \\
&= \int_{X\setminus\mathcal{E}_k} f(z)\mu(\tilde{\varphi}_k^{-1}(\varphi_k(z)))^\rho d\mu(z) + \frac{\Pr(Z\in\mathcal{E}_k)\mu(\mathcal{E}_k)^\rho}{\lfloor(1-\alpha)N\rfloor^\rho} && (36)\\
&\le \int_{X\setminus\mathcal{E}_k} f(z)\mu(\varphi_k^{-1}(\varphi_k(z)))^\rho d\mu(z) + \frac{\mu(\mathcal{E}_k)^\rho}{\lfloor(1-\alpha)N\rfloor^\rho} && (37)\\
&= \int_{X\setminus\mathcal{E}_k} f_k(z)\mu(\varphi_k^{-1}(\varphi_k(z)))^\rho d\mu(z) + \frac{\mu(\mathcal{E}_k)^\rho}{\lfloor(1-\alpha)N\rfloor^\rho} && (38)\\
&\le \int_X f_k(z)\mu(\varphi_k^{-1}(\varphi_k(z)))^\rho d\mu(z) + \frac{\mu(\mathcal{E}_k)^\rho}{\lfloor(1-\alpha)N\rfloor^\rho} && (39)
\end{aligned}
$$

where (36) follows from the construction of $\tilde{\varphi}_k$; where (37) follows because $\tilde{\varphi}_k^{-1}(\varphi_k(z)) = \varphi_k^{-1}(\varphi_k(z))\setminus\mathcal{E}_k$ for $z\in X\setminus\mathcal{E}_k$ and because $\Pr(Z\in\mathcal{E}_k)\le 1$; where (38) follows because $\mu(\varphi_k^{-1}(\varphi_k(z)))$ is constant on each $\mathcal{E}_{i,k}$ and $\int_{\mathcal{E}_{i,k}} f d\mu = \int_{\mathcal{E}_{i,k}} f_k d\mu$; and where (39) follows because the integrand is nonnegative so enlarging the domain of integration can only increase the value of the integral. It follows that

$$
\begin{aligned}
&N^\rho\, \mathrm{E}[\mu(\tilde{\varphi}_k^{-1}(\tilde{\varphi}_k(Z)))^\rho] \\
&\le \frac{N^\rho}{\lfloor\alpha N\rfloor^\rho}\lfloor\alpha N\rfloor^\rho \int_X f_k(z)\mu(\varphi_k^{-1}(\varphi_k(z)))^\rho d\mu(z) \\
&\quad + \frac{N^\rho}{\lfloor(1-\alpha)N\rfloor^\rho}\mu(\mathcal{E}_k)^\rho. && (40)
\end{aligned}
$$

Hence, by (35),

$$
\begin{aligned}
&\overline{\lim_{N\to\infty}}\, N \min_{\varphi\in\mathcal{F}(X,N)} \mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^\rho]^{1/\rho} \\
&\le \left(\alpha^{-\rho}\left(\int_X f_k(z)^{\frac{1}{1+\rho}} d\mu(z)\right)^{1+\rho} + (1-\alpha)^{-\rho}\mu(\mathcal{E}_k)^\rho\right)^{1/\rho}. && (41)
\end{aligned}
$$

Recalling (33) and noting that $\lim_{k\to\infty}\mu(\mathcal{E}_k) = 0$ (see (30)), the proof is completed by first letting $k\to\infty$ and then $\alpha\to 1$. ■

III. Existence of an Optimal Quantizer

Here we prove that the infimum

$$
\inf_{\varphi\in\mathcal{F}(X,N)} \mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^\rho] \tag{42}
$$

is attained. The case $N = 1$ is trivial, so we assume that $N \ge 2$. Let $(\varphi_k)_{k\ge1}$ be a sequence in $\mathcal{F}(X, N)$ such that

$$
\lim_{k\to\infty} \mathrm{E}[\mu(\varphi_k^{-1}(\varphi_k(Z)))^\rho] = \inf_{\varphi\in\mathcal{F}(X,N)} \mathrm{E}[\mu(\varphi^{-1}(\varphi(Z)))^\rho]. \tag{43}
$$

By passing to a subsequence, we may assume that the sequence of N-tuples

$$
\big(\mu(\varphi_k^{-1}(1)), \ldots, \mu(\varphi_k^{-1}(N))\big) \tag{44}
$$

converges to a limit $(a_1, \ldots, a_N)$, where the a_n's are nonnegative and sum up to $\mu(X)$. By relabeling if necessary, we may further assume that

$$
a_1 \le a_2 \le \cdots \le a_N. \tag{45}
$$

To keep the presentation concise, we focus on the case where $a_1 > 0$. The case where $a_1 = \ldots = a_j = 0$ for some $j \in \{1, \ldots, N-1\}$ requires only minor modifications.

We construct an optimal $\varphi \in \mathcal{F}(X, N)$ as follows. Define the nonincreasing thresholds

$$
t_n = \inf\{t \ge 0 : \mu(\{f \ge t\}) \le a_1 + \ldots + a_n\}, \tag{46}
$$

for $1 \le n \le N-1$, where we use the shorthand $\{f \ge t\} = \{z \in X : f(z) \ge t\}$.

Using the continuity of measure, it is straightforward to verify that $\mu(\{f > t_1\}) \le a_1$ and $\mu(\{f \ge t_1\}) \ge a_1$. Thus, since μ is nonatomic [3], there is a measurable $A_1 \subseteq \{f = t_1\}$ (possibly the empty set) such that $\mu(\{f > t_1\}\cup A_1) = a_1$. We set

$$
\varphi^{-1}(1) = \{f > t_1\} \cup A_1. \tag{47}
$$

To construct $\varphi^{-1}(n)$ for $2 \le n \le N-1$, assume that $\varphi^{-1}(1), \ldots, \varphi^{-1}(n-1)$ and $A_1, \ldots, A_{n-1}$ have been defined. If $t_n = t_{n-1}$, then we can find a measurable $A_n \subseteq \{f = t_n\} \setminus \bigcup_{i=1}^{n-1} A_i$ such that $\mu(A_n) = a_n$. In this case, we set $\varphi^{-1}(n) = A_n$. If $t_n < t_{n-1}$, then we can find a measurable $A_n \subseteq \{f = t_n\}$ such that

$$
\mu\Big(A_n \cup \Big(\{f > t_n\} \setminus \bigcup_{i=1}^{n-1}\varphi^{-1}(i)\Big)\Big) = a_n. \tag{48}
$$

In this case, we set $\varphi^{-1}(n) = A_n \cup (\{f > t_n\}\setminus\bigcup_{i=1}^{n-1}\varphi^{-1}(i))$. Finally, we set $\varphi^{-1}(N) = X \setminus \bigcup_{i=1}^{N-1}\varphi^{-1}(i)$.

The $\varphi \in \mathcal{F}(X, N)$ thus obtained has the properties:

$$
\big(\mu(\varphi^{-1}(1)), \ldots, \mu(\varphi^{-1}(N))\big) = (a_1, \ldots, a_N) \tag{49a}
$$

and

$$
\big(\varphi(z) \le \varphi(z')\big) \implies \big(f(z) \ge f(z')\big). \tag{49b}
$$

As we next show, these properties are the key to optimality.

Select $\delta > 0$ and k sufficiently large so that

$$
|\mu(\varphi_k^{-1}(n))^\rho - a_n^\rho| < \delta, \quad n \in \{1, \ldots, N\}. \tag{50}
$$

We then have

$$
\begin{aligned}
&\int_X f(z)\mu(\varphi_k^{-1}(\varphi_k(z)))^\rho d\mu(z) \\
&\quad = \sum_{n=1}^N \Pr\big(Z \in \varphi_k^{-1}(n)\big)\mu(\varphi_k^{-1}(n))^\rho \\
&\quad \ge \sum_{n=1}^N \Pr\big(Z \in \varphi_k^{-1}(n)\big)a_n^\rho - \delta. && (51)
\end{aligned}
$$

To complete the proof we use a variant of the summation-by-parts identity: If $b_n, \tilde{b}_n, c_n$ are real numbers and $\sum_{n=1}^N b_n =$

$\sum_{n=1}^{N} \tilde{b}_n$, then

$$\sum_{n=1}^{N} c_n b_n - \sum_{n=1}^{N} c_n \tilde{b}_n = \sum_{n=1}^{N-1} (c_n - c_{n+1})\big((b_1 + \ldots + b_n) - (\tilde{b}_1 + \ldots + \tilde{b}_n)\big). \tag{52}$$

Set $b_n = \Pr(Z \in \varphi_k^{-1}(n))$, $\tilde{b}_n = \Pr(Z \in \varphi^{-1}(n))$, and observe that for any $n \in \{1, \ldots, N-1\}$,

$$\begin{aligned}(b_1 + \ldots + b_n) - (\tilde{b}_1 + \ldots + \tilde{b}_n) &= \Pr\Big(Z \in \bigcup_{i=1}^{n} \varphi_k^{-1}(i) \setminus \bigcup_{i=1}^{n} \varphi^{-1}(i)\Big) \\ &\quad - \Pr\Big(Z \in \bigcup_{i=1}^{n} \varphi^{-1}(i) \setminus \bigcup_{i=1}^{n} \varphi_k^{-1}(i)\Big).\end{aligned} \tag{53}$$

Since $\bigcup_{i=1}^{n} \varphi^{-1}(i) = \{f > t_n\} \cup A_n$, where $A_n \subseteq \{f = t_n\}$, it follows that the first probability on the RHS of (53) is upper-bounded by

$$t_n\, \mu\Big(\bigcup_{i=1}^{n} \varphi_k^{-1}(i) \setminus \bigcup_{i=1}^{n} \varphi^{-1}(i)\Big), \tag{54}$$

and the second probability is lower-bounded by

$$t_n\, \mu\Big(\bigcup_{i=1}^{n} \varphi^{-1}(i) \setminus \bigcup_{i=1}^{n} \varphi_k^{-1}(i)\Big). \tag{55}$$

Thus,

$$\begin{aligned}(b_1 + \ldots + b_n) - (\tilde{b}_1 + \ldots + \tilde{b}_n) &\leq t_n\Big(\mu\Big(\bigcup_{i=1}^{n} \varphi_k^{-1}(i)\Big) - \mu\Big(\bigcup_{i=1}^{n} \varphi^{-1}(i)\Big)\Big) \\ &= t_n\Big(\sum_{i=1}^{n} \mu(\varphi_k^{-1}(i)) - \sum_{i=1}^{n} a_i\Big) \\ &= \varepsilon_k^{(n)},\end{aligned} \tag{56}$$

where $\varepsilon_k^{(n)} \to 0$ as $k \to \infty$. Setting $\varepsilon_k = \max_n \varepsilon_k^{(n)}$, recalling that the a_n's are nondecreasing, and using (52) with $c_n = a_n^{\rho}$, we thus obtain

$$\begin{aligned}&\sum_{n=1}^{N} \Pr\big(Z \in \varphi_k^{-1}(n)\big) a_n^{\rho} \\ &\quad \geq \sum_{n=1}^{N} \Pr\big(Z \in \varphi^{-1}(n)\big) a_n^{\rho} - \varepsilon_k \sum_{n=1}^{N-1} (a_{n+1}^{\rho} - a_n^{\rho}) \\ &\quad = \int_X f(z) \mu\big(\varphi^{-1}(\varphi(z))\big)^{\rho} d\mu(z) - \varepsilon_k (a_N^{\rho} - a_1^{\rho}) \\ &\quad \geq \int_X f(z) \mu\big(\varphi^{-1}(\varphi(z))\big)^{\rho} d\mu(z) - \varepsilon_k\, \mu(X)^{\rho},\end{aligned} \tag{57}$$

where (57) follows because $a_N \leq \mu(X)$. Combining (57) and (51), and letting first $k \to \infty$ and then $\delta \to 0$ shows that

$$\int_X f(z) \mu\big(\varphi^{-1}(\varphi(z))\big)^{\rho} d\mu(z) \leq \lim_{k \to \infty} \int_X f(z) \mu(\varphi_k^{-1}(\varphi_k(z)))^{\rho} d\mu(z). \tag{58}$$

In view of (43) this completes the proof. ∎

References

[1] R. M. Dudley, *Real Analysis and Probability*. Cambridge University Press, 2002.

[2] C. Bunte and A. Lapidoth, “Encoding tasks and Rényi entropy,” *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5065–5076, Sept 2014.

[3] D. A. Ross, “An elementary proof of Lyapunov’s theorem,” *American Mathematical Monthly*, vol. 112, no. 7, pp. 651–653, 2005.

[4] C. Bunte and A. Lapidoth, “Maximizing differential Rényi entropy,” submitted to the 2014 IEEE 28th Convention of Electrical and Electronic Engineers in Israel (IEEEI).

[5] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Springer, 2000.

[6] C. Bunte and A. Lapidoth, manuscript in preparation.

[7] A. N. Shiryaev, *Probability*, 2nd ed. Springer, 1995.

[8] L. Wang and M. Madiman, “Beyond the entropy power inequality, via rearrangements,” *arXiv preprint arXiv:1307.6018*, 2013.