10-17-2022

Optimal Quantization for Some Triadic Uniform Cantor Distributions with Exact Bounds

Mrinal Kanti Roychowdhury
The University of Texas Rio Grande Valley

Recommended Citation

Roychowdhury, M.K. Optimal Quantization for Some Triadic Uniform Cantor Distributions with Exact Bounds. Qual. Theory Dyn. Syst. 21, 149 (2022). https://doi.org/10.1007/s12346-022-00678-8

To appear, Qualitative Theory of Dynamical Systems

OPTIMAL QUANTIZATION FOR SOME TRIADIC UNIFORM CANTOR DISTRIBUTIONS WITH EXACT BOUNDS

MRINAL KANTI ROYCHOWDHURY

Abstract. Let $\{S_j : 1 \leq j \leq 3\}$ be a set of three contractive similarity mappings such that $S_j(x) = rx + \frac{j-1}{2}(1-r)$ for all $x \in \mathbb{R}$, and $1 \leq j \leq 3$, where $0 < r < \frac{1}{3}$. Let $P = \sum_{j=1}^{3} \frac{1}{3} P \circ S_j^{-1}$. Then, P is a unique Borel probability measure on $\mathbb{R}$ such that P has support the Cantor set generated by the similarity mappings S_j for $1 \leq j \leq 3$. Let $r_0 = 0.1622776602$, and $r_1 = 0.2317626315$ (which are ten digit rational approximations of two real numbers). In this paper, for $0 < r \leq r_0$, we give a general formula to determine the optimal sets of n-means and the nth quantization errors for the triadic uniform Cantor distribution P for all positive integers $n \geq 2$. Previously, Roychowdhury gave an exact formula to determine the optimal sets of n-means and the nth quantization errors for the standard triadic Cantor distribution, i.e., when $r = \frac{1}{5}$. In this paper, we further show that $r = r_0$ is the greatest lower bound, and $r = r_1$ is the least upper bound of the range of r-values to which Roychowdhury formula extends. In addition, we show that for $0 < r \leq r_1$ the quantization coefficient does not exist though the quantization dimension exists.

1. Introduction

Let P be a Borel probability measure on $\mathbb{R}^d$, where $d \geq 1$. For a finite set $\alpha \subset \mathbb{R}^d$, write

$$V(P;\alpha) = \int \min_{a \in \alpha} \|x-a\|^2 dP(x), \text{ and } V_n := V_n(P) = \inf\Big\{V(P;\alpha) : \alpha \subset \mathbb{R}^d, \text{ card}(\alpha) \leq n\Big\},$$

where $\|\cdot\|$ represents the Euclidean norm on $\mathbb{R}^d$. Then, $V(P;\alpha)$ is called the *cost* or *distortion error* for P with respect to the set α, and V_n is called the nth quantization error for P with respect to the squared Euclidean distance. A set $\alpha \subset \mathbb{R}^d$ is called an *optimal set of n-means* for P if $V_n(P) = V(P;\alpha)$. It is well-known that for a continuous Borel probability measure an optimal set of n-means contains exactly n-elements (see [4]). To see some work in the direction of optimal sets of n-means, one is referred to [2, 5, 16]. For theoretical results in quantization we refer to [4, 6–8, 11], and for its promising application see [12, 13]. For a finite set $\alpha \subset \mathbb{R}^d$ and $a \in \alpha$, by $M(a|\alpha)$ we denote the set of all elements in $\mathbb{R}^d$ which are nearest to a among all the elements in α, i.e.,

$$M(a|\alpha) = \{x \in \mathbb{R}^d : \|x-a\| = \min_{b \in \alpha} \|x-b\|\}.$$

$M(a|\alpha)$ is called the *Voronoi region* generated by $a \in \alpha$. On the other hand, the set $\{M(a|\alpha) : a \in \alpha\}$ is called the *Voronoi diagram* or *Voronoi tessellation* of $\mathbb{R}^d$ with respect to the set α.

Definition 1.1. *A set $\alpha \subset \mathbb{R}^d$ is called a centroidal Voronoi tessellation (CVT) with respect to a probability distribution P on $\mathbb{R}^d$, if it satisfies the following two conditions:*

(i) $P(M(a|\alpha) \cap M(b|\alpha)) = 0$ for $a, b \in \alpha$, and $a \neq b$;

(ii) $E(X : X \in M(a|\alpha)) = a$ for all $a \in \alpha$,

where X is a random variable with distribution P, and $E(X : X \in M(a|\alpha))$ represents the conditional expectation of the random variable X given that X takes values in $M(a|\alpha)$.

A Borel measurable partition $\{A_a : a \in \alpha\}$ is called a *Voronoi partition* of $\mathbb{R}^d$ with respect to the probability distribution P, if P-almost surely $A_a \subset M(a|\alpha)$ for all $a \in \alpha$. Let us now state the following proposition (see [3, 4]).

2010 *Mathematics Subject Classification.* 60Exx, 28A80, 94A34.

Key words and phrases. Cantor set, probability distribution, optimal sets, quantization error, centroidal Voronoi tessellation.

Proposition 1.2. *Let α be an optimal set of n-means, $a \in \alpha$, and $M(a|\alpha)$ be the Voronoi region generated by $a \in \alpha$, i.e., $M(a|\alpha) = \{x \in \mathbb{R}^d : \|x - a\| = \min_{b \in \alpha} \|x - b\|\}$. Then, for every $a \in \alpha$, (i) $P(M(a|\alpha)) > 0$, (ii) $P(\partial M(a|\alpha)) = 0$, (iii) $a = E(X : X \in M(a|\alpha))$.*

The number $D(P) := \lim\limits_{n\to\infty} \frac{2 \log n}{-\log V_n(P)}$, if it exists, is called the *quantization dimension* of the probability measure P. On the other hand, for $s \in (0, +\infty)$, the number $\lim\limits_{n\to\infty} n^{\frac{2}{s}} V_n(P)$, if it exists, is called the s-dimensional *quantization coefficient* for P. To know details about the quantization dimension and the quantization coefficient one is referred to [4].

Let $\{S_j : 1 \leq j \leq 3\}$ be a set of three contractive similarity mappings such that $S_j(x) = rx + \frac{j-1}{2}(1-r)$ for all $x \in \mathbb{R}$, where $0 < r < \frac{1}{3}$ and $1 \leq j \leq 3$. For any positive integer n, if $\sigma := \sigma_1\sigma_2 \cdots \sigma_n \in \{1,2,3\}^n$, then we say that σ is a word of length n. By $\{1,2,3\}^*$, we denote the set of all words including the empty word $\emptyset$. The empty word $\emptyset$ has length zero. For $\sigma := \sigma_1\sigma_2 \cdots \sigma_n \in \{1,2,3\}^n$, by S_σ it is meant that $S_\sigma := S_{\sigma_1} \circ \cdots \circ S_{\sigma_n}$, and by $a(\sigma)$, we mean $a(\sigma) := S_\sigma(\frac{1}{2})$. For the empty word $\emptyset$, by $S_\emptyset$ it is meant the identity mapping on $\mathbb{R}$. For $\sigma := \sigma_1\sigma_2 \cdots \sigma_n \in \{1,2,3\}^n$, set $J_\sigma := S_\sigma([0,1])$. For the empty word $\emptyset$, write $J := J_\emptyset = S_\emptyset([0,1]) = [0,1]$. Then, the set $C := \bigcap_{n\in\mathbb{N}} \bigcup_{\sigma\in\{1,2,3\}^n} J_\sigma$ is known as the *Cantor set* generated by the mappings S_j, and equals the support of the probability measure P given by $P = \sum_{j=1}^3 \frac{1}{3} P \circ S_j^{-1}$. Notice that C satisfies the invariance equality $C = \mathop{\cup}\limits_{j=1}^{3} S_j(C)$ (see [10]). In this paper a Cantor set C, which is generated by a set of three contractive similarity mappings, is called a *triadic Cantor set*, and a probability measure P which has support the triadic Cantor set, is called a *triadic Cantor distribution*. For words $\beta, \gamma, \cdots, \delta$ in $\{1,2,3\}^*$, we write

$$a(\beta, \gamma, \cdots, \delta) := E(X|X \in J_\beta \cup J_\gamma \cup \cdots \cup J_\delta) = \frac{1}{P(J_\beta \cup \cdots \cup J_\delta)} \int_{J_\beta \cup \cdots \cup J_\delta} x dP(x),$$

where X is a random variable with probability distribution P, and $E(X)$ and $V := V(X)$ represent the expectation and the variance of the random variable X. Notice that for any $\omega \in \{1,2,3\}^*$, the similarity mapping S_ω is an injective mapping on $\mathbb{R}$; on the other hand, for any discrete subset A of $\mathbb{R}$, the set $S_\omega(A)$ represents the set of values obtained by applying S_ω to each of the elements in A. Let us now give the following two definitions.

Definition 1.3. *For $n \in \mathbb{N}$ with $n \geq 3$ let $\ell(n)$ be the unique natural number with $3^{\ell(n)} \leq n < 3^{\ell(n)+1}$. Write $\beta_2 := \{a(1), a(2,3)\}$ and $\beta_3 := \{a(1), a(2), a(3)\}$. For $n \geq 3$, define $\beta_n := \beta_n(I)$ as follows:*

$$\beta_n(I) = \left\{ \begin{array}{ll} \{a(\omega) : \omega \in \{1,2,3\}^{\ell(n)} \setminus I\} \bigcup \mathop{\cup}\limits_{\omega \in I} S_\omega(\beta_2) & \text{if } 3^{\ell(n)} \leq n \leq 2 \cdot 3^{\ell(n)}, \\ \{S_\omega(\beta_2) : \omega \in \{1,2,3\}^{\ell(n)} \setminus I\} \bigcup \mathop{\cup}\limits_{\omega \in I} S_\omega(\beta_3) & \text{if } 2 \cdot 3^{\ell(n)} < n < 3^{\ell(n)+1}, \end{array} \right.$$

where $I \subset \{1,2,3\}^{\ell(n)}$ is arbitray with card$(I) = n - 3^{\ell(n)}$ if $3^{\ell(n)} \leq n \leq 2 \cdot 3^{\ell(n)}$; and card$(I) = n - 2 \cdot 3^{\ell(n)}$ if $2 \cdot 3^{\ell(n)} < n < 3^{\ell(n)+1}$.

Definition 1.4. *For $n \in \mathbb{N}$ with $n \geq 3$ let $\ell(n)$ be the unique natural number with $3^{\ell(n)} \leq n < 3^{\ell(n)+1}$. Write $\gamma_2 := \{a(1,21), a(22,23,3)\}$ and $\gamma_3 := \{a(1), a(2), a(3)\}$. For $n \geq 3$, define $\gamma_n := \gamma_n(I)$ as follows:*

$$\gamma_n(I) = \left\{ \begin{array}{ll} \{a(\omega) : \omega \in \{1,2,3\}^{\ell(n)} \setminus I\} \bigcup \mathop{\cup}\limits_{\omega \in I} S_\omega(\gamma_2) & \text{if } 3^{\ell(n)} \leq n \leq 2 \cdot 3^{\ell(n)}, \\ \{S_\omega(\gamma_2) : \omega \in \{1,2,3\}^{\ell(n)} \setminus I\} \bigcup \mathop{\cup}\limits_{\omega \in I} S_\omega(\gamma_3) & \text{if } 2 \cdot 3^{\ell(n)} < n < 3^{\ell(n)+1}, \end{array} \right.$$

where $I \subset \{1,2,3\}^{\ell(n)}$ is arbitrary with card$(I) = n - 3^{\ell(n)}$ if $3^{\ell(n)} \leq n \leq 2 \cdot 3^{\ell(n)}$; and card$(I) = n - 2 \cdot 3^{\ell(n)}$ if $2 \cdot 3^{\ell(n)} < n < 3^{\ell(n)+1}$.

Remark 1.5. In the paper there are several decimal numbers, they are rational approximations of some real numbers up to ten decimal places.

Roychowdhury showed that if $r=\frac{1}{5}$, then the sets γ_n given by Definition 1.3, determine the optimal sets of n-means for all positive integers $n\geq 2$ (see [15]). Proposition 2.5 implies that γ_n forms a CVT if $\frac{1}{79}\left(21-2\sqrt{51}\right)\leq r\leq \frac{1}{41}\left(2\sqrt{31}-1\right)$, i.e., if $0.08502712839\leq r\leq 0.2472080177$. Thus, we see that the range of r values for which the sets γ_n form the optimal sets of n-means is bounded below by $\frac{1}{79}\left(21-2\sqrt{51}\right)$, and bounded above by $\frac{1}{41}\left(2\sqrt{31}-1\right)$. But, the greatest lower bound and the least upper bound of the range of r values for which the sets γ_n form the optimal sets of n-means were not known. In this paper, in Theorem 5.1 we give an answer of it.

Remark 1.6. Notice that if $r=0$, then $S_1(x)=0$, $S_2(x)=\frac{1}{2}$, and $S_3(x)=1$ for all $x\in\mathbb{R}$, and then the probability measure P becomes a discrete uniform distribution with support $\{0,\frac{1}{2},1\}$. Because of that in our study we are assuming that the contractive ratios r are positive.

The arrangement of the paper is as follows: In Section 2, we give the basic preliminaries. In Section 3, we show that the sets β_n form the optimal sets of n-means if $r=\frac{1}{25}$. In Section 4, we prove the following theorem:

Theorem 1.7. *Let $\gamma_n:=\gamma_n(I)$ be the set for arbitrary I as defined by Definition 1.4. Let $r_0,r_1\in(0,\frac{1}{3})$ be the unique real numbers satisfying, respectively, the equations*

$$-\frac{3r^5+15r^4+6r^3-42r^2+31r-13}{240(r+1)}=-\frac{3r^3-3r^2+r-1}{24(r+1)},$$

$$-\frac{3r^5+15r^4+6r^3-42r^2+31r-13}{240(r+1)}=-\frac{3r^7+15r^6+60r^5+66r^4+18r^3-324r^2+283r-121}{2184(r+1)}.$$

Then, $r_0=0.1622776602$, and $r_1=0.2317626315$. Then, for all $n\geq 3$, the sets γ_n form the optimal sets of n-means for $r=r_0$ and $r=r_1$.

In Theorem 5.1, we show that the sets β_n form the optimal sets of n-means if $0<r\leq r_0$, and the sets γ_n form the optimal sets of n-means if $r_0\leq r\leq r_1$. Thus, Theorem 5.1 implies the fact that the greatest lower bound, and the least upper bound of r for which the sets γ_n form the optimal sets of n-means are, respectively, given by $r=r_0$ and $r=r_1$. Notice that for $r=r_0$ both the sets β_n and γ_n form the optimal sets of n-means for P. In addition, in Theorem 5.2, we show that the quantization coefficient for $0<r\leq r_1$ does not exist though the quantization dimension exists.

2. Preliminaries

As defined in the previous section, let S_j for $1\leq j\leq 3$ be the contractive similarity mappings on $\mathbb{R}$ given by $S_j(x)=rx+\frac{j-1}{2}(1-r)$ for all $x\in\mathbb{R}$, and $1\leq j\leq 3$, where $0<r<\frac{1}{3}$. For $\sigma:=\sigma_1\sigma_2\cdots\sigma_k\in\{1,2,3\}^k$ and $\tau:=\tau_1\tau_2\cdots\tau_\ell\in\{1,2,3\}^\ell$, by $\sigma\tau:=\sigma_1\cdots\sigma_k\tau_1\cdots\tau_\ell$ we mean the word obtained from the concatenation of the words σ and τ. For $\sigma=\sigma_1\sigma_2\cdots\sigma_n\in\{1,2,3\}^*$, $n\geq 0$, write $p_\sigma:=\frac{1}{3^n}$ and $s_\sigma:=\frac{1}{r^n}$. Recall that if C is the Cantor set, then $C:=\bigcap_{n\in\mathbb{N}}\bigcup_{\sigma\in\{1,2,3\}^n}J_\sigma$. For $n\geq 1$, the intervals J_σ, where $\sigma\in\{1,2,3\}^n$, are called the *nth level basic intervals* of the Cantor set C.

The following two lemmas are well-known and easy to prove (see [5, 15]).

Lemma 2.1. *Let $f:\mathbb{R}\to\mathbb{R}^+$ be Borel measurable and $k\in\mathbb{N}$, and P be the probability measure on $\mathbb{R}$ given by $P=\sum_{j=1}^3\frac{1}{3}P\circ S_j^{-1}$. Then,*

$$\int f(x)dP(x)=\sum_{\sigma\in\{1,2,3\}^k}\frac{1}{3^k}\int f\circ S_\sigma(x)dP(x).$$

Lemma 2.2. *Let X be a random variable with the probability distribution P. Then,*

$$E(X)=\frac{1}{2}\text{ and }V:=V(X)=\frac{1-r}{6(r+1)},\text{ and }\int(x-x_0)^2dP(x)=V(X)+(x_0-\frac{1}{2})^2,$$

where $x_0\in\mathbb{R}$.

The following corollary is useful to obtain the distortion errors.

Corollary 2.3. Let $\sigma \in \{1,2,3\}^k$ for $k \geq 1$, and $x_0 \in \mathbb{R}$. Then,

$$\int_{J_\sigma} (x-x_0)^2 dP(x) = \frac{1}{3^k}\Big(r^{2k}V + (S_\sigma(\frac{1}{2}) - x_0)^2\Big).$$

Proof. By induction, $P = \frac{1}{3}\sum_{j=1}^3 P \circ S_j^{-1}$ implies $P = \sum_{\sigma \in \{1,2,3\}^k} p_\sigma P \circ S_\sigma^{-1}$. Using this fact, Lemma 2.1 and Lemma 2.2, the proof of the corollary follows. □

Proposition 2.4. *Let $\beta_n(I)$ be the set given by Definition 1.3. Then, $\beta_n(I)$ forms a CVT if $0 < r \leq 2 - \sqrt{3}$, i.e., if $0 < r \leq 0.2679491924$. Moreover, if $3^{\ell(n)} \leq n \leq 2 \cdot 3^{\ell(n)}$, then*

$$V(P, \beta_n(I)) = \frac{1}{3^{\ell(n)}} \cdot r^{2\ell(n)}\Big((2 \cdot 3^{\ell(n)} - n)V + (n - 3^{\ell(n)})V(P;\beta_2)\Big),$$

and if $2 \cdot 3^{\ell(n)} \leq n < 3^{\ell(n)+1}$, then

$$V(P, \beta_n(I)) = \frac{1}{3^{\ell(n)}} \cdot r^{2\ell(n)}\Big((3^{\ell(n)+1} - n)V(P;\beta_2) + (n - 2 \cdot 3^{\ell(n)})V(P;\beta_3)\Big).$$

Proof. By the definition, we have $\beta_2 = \{a(1), a(2,3)\}$ and $\beta_3 = \{a(1), a(2), a(3)\}$. Recall that $\beta_n := \beta_n(I)$ is defined for $n \geq 3$, where $I \subset \{1,2,3\}^{\ell(n)}$ with $\text{card}(I) = n - 3^{\ell(n)}$ if $3^{\ell(n)} \leq n \leq 2 \cdot 3^{\ell(n)}$; and $\text{card}(I) = n - 2 \cdot 3^{\ell(n)}$ if $2 \cdot 3^{\ell(n)} < n < 3^{\ell(n)+1}$. Notice that for $n \geq 3$, if $n \neq 3^{\ell(n)}$ or $n \neq 2 \cdot 3^{\ell(n)}$, the subset I can be chosen more than one way. This leads to the fact that if $n \neq 3^{\ell(n)}$ or $n \neq 2 \cdot 3^{\ell(n)}$, the sets β_n can be chosen multiple ways. Let us take

$$\begin{aligned}
&\beta_4 = \{a(1), a(2), a(31), a(32,33)\} \text{ (by choosing } I = \{3\}),\\
&\beta_5 = \{a(1), a(21), a(22,23), a(31), a(32,33)\} \text{ (by choosing } I = \{2,3\}),\\
&\beta_6 = \{a(11), a(12,13), a(21), a(22,23), a(31), a(32,33)\} \text{ (where } I = \{1,2,3\}),\\
&\beta_7 = \{a(11), a(12), a(13), a(21), a(22,23), a(31), a(32,33)\} \text{ (by choosing } I = \{1\}).
\end{aligned}$$

Since similarity mappings preserve the ratio of the distances of a point from any other two points, $\beta_n(I)$ will form a CVT if we can show that β_2, β_3, β_4, β_5, β_6, β_7 form a CVT. Recall that $a(1) = E(X : X \in J_1)$ and $a(2,3) = E(X : X \in J_2 \cup J_3)$, and also recall the Definition 1.1. Thus, β_2 will form a CVT if

$$P(M(a(1)|\beta_2) \cap M(a(2,3)|\beta_2)) = 0. \tag{1}$$

Since the basic intervals in the first level are $J_1 := [S_1(0), S_1(1)]$, $J_2 := [S_2(0), S_2(1)]$, and $J_3 := [S_3(0), S_3(1)]$, the relation (1) will be true if

$$S_1(1) \leq \frac{1}{2}\,(a(1) + a(2,3)) \leq S_2(0).$$

Similarly, β_3 will form a CVT if $S_i(1) < \frac{1}{2}(a(i) + a(i+1)) < S_{i+1}(0)$ for $i = 1, 2$; β_4 will form a CVT if

$$\begin{aligned}
S_1(1) < \frac{1}{2}(a(1) + a(2)) < S_2(0) < S_2(1) < \frac{1}{2}(a(2) + a(31)) < S_{31}(0) < S_{31}(1)\\
< \frac{1}{2}(a(31) + a(32,33)) < S_{32}(0).
\end{aligned}$$

Similarly, we can obtain the inequalities for which β_5, β_6, and β_7 will form a CVT. Due to similarity, combining all the inequalities, we see that they will be true if the following inequalities are true:

$$\begin{aligned}S_1(1) &\leq \frac{1}{2}\left(a(1)+a(2,3)\right) \leq S_2(0),\\ S_1(1) &\leq \frac{1}{2}\left(a(1)+a(21)\right) \leq S_{21}(0),\\ S_{13}(1) &\leq \frac{1}{2}(a(12,13)+a(21)) \leq S_{21}(0),\\ S_{13}(1) &\leq \frac{1}{2}\left(a(13)+a(21)\right) \leq S_{21}(0).\end{aligned}$$

Upon some simplification, we see that the above inequalities are true if $0 < r \leq 2-\sqrt{3}$, i.e., if $0 < r \leq 0.2679491924$. If $3^{\ell(n)} \leq n \leq 2\cdot 3^{\ell(n)}$, then

$$\begin{aligned}V(P;\beta_n(I)) &= \sum_{\sigma\in\{1,2,3\}^{\ell(n)}\setminus I}\int_{J_\sigma}(x-a(\sigma))^2dP + \sum_{\sigma\in I}\int_{J_\sigma}\min_{a\in S_\sigma(\beta_2)}(x-a)^2dP\\ &= \frac{1}{3^{\ell(n)}}r^{2\ell(n)}\Big(\sum_{\sigma\in\{1,2,3\}^{\ell(n)}\setminus I}V+\sum_{\sigma\in I}V(P;\beta_2)\Big)\\ &= \frac{1}{3^{\ell(n)}}\cdot r^{2\ell(n)}\Big((2\cdot 3^{\ell(n)}-n)V+(n-3^{\ell(n)})V(P;\beta_2)\Big).\end{aligned}$$

Similarly, if $2\cdot 3^{\ell(n)} \leq n < 3^{\ell(n)+1}$, then

$$V(P,\beta_n(I)) = \frac{1}{3^{\ell(n)}}\cdot r^{2\ell(n)}\Big((3^{\ell(n)+1}-n)V(P;\beta_2)+(n-2\cdot 3^{\ell(n)})V(P;\beta_3)\Big).$$

Thus, the proof of the proposition is complete. $\square$

Proposition 2.5. *Let $\gamma_n(I)$ be the set given by Definition 1.4. Then, $\gamma_n(I)$ forms a CVT if $\frac{1}{79}\left(21-2\sqrt{51}\right) \leq r \leq \frac{1}{41}\left(2\sqrt{31}-1\right)$, i.e., if $0.08502712839 \leq r \leq 0.2472080177$. Moreover, if $3^{\ell(n)} \leq n \leq 2\cdot 3^{\ell(n)}$, then*

$$V(P,\gamma_n(I)) = \frac{1}{3^{\ell(n)}}\cdot r^{2\ell(n)}\Big((2\cdot 3^{\ell(n)}-n)V+(n-3^{\ell(n)})V(P;\gamma_2)\Big),$$

and if $2\cdot 3^{\ell(n)} \leq n < 3^{\ell(n)+1}$, then

$$V(P,\gamma_n(I)) = \frac{1}{3^{\ell(n)}}\cdot r^{2\ell(n)}\Big((3^{\ell(n)+1}-n)V(P;\gamma_2)+(n-2\cdot 3^{\ell(n)})V(P;\gamma_3)\Big).$$

Proof. By the definition, we have $\gamma_2 = \{a(1,21), a(22,23,3)\}$ and $\gamma_3 = \{a(1), a(2), a(3)\}$. For $n \geq 3$, if $n \neq 3^{\ell(n)}$ or $n \neq 2\cdot 3^{\ell(n)}$, the subset I can be chosen more than one way. This leads to the fact that if $n \neq 3^{\ell(n)}$ or $n \neq 2\cdot 3^{\ell(n)}$, the sets γ_n can be chosen multiple ways. Proceeding in the similar way, as Proposition 2.4, let us choose

$$\begin{aligned}&\gamma_4 = \{a(1), a(2), a(31,321), a(322,323,33)\}\\ &\gamma_5 = \{a(1), a(21,221), a(222,223,23), a(31,321), a(322,323,33)\}\\ &\gamma_6 = \{a(11,121), a(122,123,13), a(21,221), a(222,223,23), a(31,321), a(322,323,33)\}\\ &\gamma_7 = \{a(11), a(12), a(13), a(21,221), a(222,223,23), a(31,321), a(322,323,33)\}.\end{aligned}$$

Due to the same reasoning as described in the proof of Proposition 2.4, to show $\gamma_n(I)$ forms a CVT, it is enough to prove that the following inequalities are true:

$$\begin{aligned}
S_{21}(1) &\leq \frac{1}{2}\left((a(1,21)+a(22,23,3)\right) \leq S_{22}(0),\\
S_1(1) &\leq \frac{1}{2}\left(a(1)+a(21,221)\right) \leq S_{21}(0),\\
S_{13}(1) &\leq \frac{1}{2}(a(122,123,13)+a(21,221)) \leq S_{21}(0),\\
S_{13}(1) &\leq \frac{1}{2}\left(a(13)+a(21,221)\right) \leq S_{21}(0).
\end{aligned}$$

Upon some simplification, we see that the above inequalities are true if $\frac{1}{79}\left(21-2\sqrt{51}\right) \leq r \leq \frac{1}{41}\left(2\sqrt{31}-1\right)$, i.e., if $0.08502712839 \leq r \leq 0.2472080177$. The rest of the proof follows in the similar way as it is given for $V(P;\beta_n)$ in Proposition 2.4. Thus, the proof of the proposition is complete. $\square$

Definition 2.6. *For $n \in \mathbb{N}$ with $n \geq 3$ let $\ell(n)$ be the unique natural number with $3^{\ell(n)} \leq n < 3^{\ell(n)+1}$. Write $\delta_2 := \{a(1,21,221), a(222,223,23,3)\}$ and $\delta_3 := \{a(1), a(2), a(3)\}$. For $n \geq 3$, define $\delta_n := \delta_n(I)$ as follows:*

$$\delta_n(I) = \left\{\begin{array}{ll} \{a(\omega) : \omega \in \{1,2,3\}^{\ell(n)} \setminus I\} \bigcup \underset{\omega \in I}{\cup} S_\omega(\delta_2) & \text{if } 3^{\ell(n)} \leq n \leq 2\cdot 3^{\ell(n)},\\ \{S_\omega(\delta_2) : \omega \in \{1,2,3\}^{\ell(n)} \setminus I\} \bigcup \underset{\omega \in I}{\cup} S_\omega(\delta_3) & \text{if } 2\cdot 3^{\ell(n)} < n < 3^{\ell(n)+1}, \end{array}\right.$$

where $I \subset \{1,2,3\}^{\ell(n)}$ with $card(I) = n - 3^{\ell(n)}$ if $3^{\ell(n)} \leq n \leq 2\cdot 3^{\ell(n)}$; and $card(I) = n - 2\cdot 3^{\ell(n)}$ if $2\cdot 3^{\ell(n)} < n < 3^{\ell(n)+1}$.

Proposition 2.7. *Let $\delta_n(I)$ be the set given by Definition 2.6. Then, $\delta_n(I)$ forms a CVT if $0.1845020699 \leq r \leq 0.2705731187$. Moreover, if $3^{\ell(n)} \leq n \leq 2\cdot 3^{\ell(n)}$, then*

$$V(P,\delta_n(I)) = \frac{1}{3^{\ell(n)}}\cdot r^{2\ell(n)}\Big((2\cdot 3^{\ell(n)}-n)V + (n-3^{\ell(n)})V(P;\delta_2)\Big),$$

and if $2\cdot 3^{\ell(n)} \leq n < 3^{\ell(n)+1}$, then

$$V(P,\delta_n(I)) = \frac{1}{3^{\ell(n)}}\cdot r^{2\ell(n)}\Big((3^{\ell(n)+1}-n)V(P;\delta_2) + (n-2\cdot 3^{\ell(n)})V(P;\delta_3)\Big).$$

Proof. By the definition, we have $\delta_2 = \{a(1,21,221), a(222,223,23,3)\}$ and $\delta_3 = \{a(1), a(2), a(3)\}$. For $n \geq 3$, if $n \neq 3^{\ell(n)}$ or $n \neq 2\cdot 3^{\ell(n)}$, the subset I can be chosen more than one way. This leads to the fact that if $n \neq 3^{\ell(n)}$ or $n \neq 2\cdot 3^{\ell(n)}$, the sets δ_n can be chosen multiple ways. Proceeding in the similar way, as Proposition 2.4, let us choose

$$\begin{aligned}
\delta_4 &= \{a(1), a(2), a(31,321,3221), a(3222,3223,323,33)\}\\
\delta_5 &= \{a(1), a(21,221,2221), a(2222,2223,223,23),\\
&\qquad a(31,321,3221), a(3222,3223,323,33)\}\\
\delta_6 &= \{a(11,121,1221), a(1222,1223,123,13), a(21,221,2221), a(2222,2223,223,23),\\
&\qquad a(31,321,3221), a(3222,3223,323,33)\}\\
\delta_7 &= \{a(11), a(12), a(13), a(21,221,2221), a(2222,2223,223,23),\\
&\qquad a(31,321,3221), a(3222,3223,323,33)\}.
\end{aligned}$$

Due to the same reasoning as described in the proof of Proposition 2.4, to show $\delta_n(I)$ forms a CVT, it is enough to prove that the following inequalities are true:

$$\begin{aligned} S_{221}(1) &\leq \frac{1}{2}\left(a(1, 21, 221) + a(222, 223, 23, 3)\right) \leq S_{222}(0), \\ S_1(1) &\leq \frac{1}{2}\left(a(1) + a(21, 221, 2221)\right) \leq S_{21}(0), \\ S_{13}(1) &\leq \frac{1}{2}(a(1222, 1223, 123, 13) + a(21, 221, 2221)) \leq S_{21}(0), \\ S_{13}(1) &\leq \frac{1}{2}\left(a(13) + a(21, 221, 2221)\right) \leq S_{21}(0). \end{aligned}$$

The above inequalities are true if $0.1845020699 \leq r \leq 0.2705731187$. The rest of the proof follows in the similar way as it is given for $V(P; \beta_n(I))$ in Proposition 2.4. Thus, the proof of the proposition is complete. $\square$

The following proposition is useful to establish Lemma 3.1, and Lemma 4.1.

Proposition 2.8. *Let $\kappa := \{a_1, a_2\}$, where $a_1 := E(X : X \in [0, \frac{1}{2}])$, and $a_2 := E(X : X \in [\frac{1}{2}, 1])$. Then, $a_1 = \frac{r+1}{6-2r}$, and $a_2 = \frac{5-3r}{6-2r}$, and the corresponding distortion error is given by*

$$V(P; \kappa) = \frac{-7r^3 + 13r^2 - 9r + 3}{6(r-3)^2(r+1)}.$$

Proof. By the hypothesis, we have

$$\begin{aligned} a_1 &= E(X : X \in [0, \frac{1}{2}]) = E\Big(X : X \in J_1 \cup J_{21} \cup J_{221} \cup \cdots\Big), \text{ and} \\ a_2 &= E(X : X \in [\frac{1}{2}, 1]) = E\Big(X : X \in J_3 \cup J_{23} \cup J_{223} \cup \cdots\Big), \end{aligned}$$

yielding

$$a_1 = 2\sum_{n=1}^{\infty} \frac{1}{3^n} \frac{1}{2}(-r^{n-1} + r^n + 1) = \frac{r+1}{6-2r}, \text{ and } a_2 = 2\sum_{n=1}^{\infty} \frac{1}{3^n} \frac{1}{2}(r^{n-1} - r^n + 1) = \frac{5-3r}{6-2r},$$

and the corresponding distortion error is given by

$$V(P; \kappa) = 2\int_{J_1 \cup J_{21} \cup J_{221} \cup J_{2221} \cdots} \Big(x - \frac{r+1}{6-2r}\Big)^2 dP$$

implying

$$V(P; \kappa) = 2\Big(\sum_{n=1}^{\infty} \frac{r^{2n}}{3^n} V + \sum_{n=1}^{\infty} \frac{1}{3^n}\Big(\frac{1}{2}\left(-r^{n-1} + r^n + 1\right) - \frac{r+1}{6-2r}\Big)^2\Big) = \frac{-7r^3 + 13r^2 - 9r + 3}{6(r-3)^2(r+1)}.$$

Thus, the proposition is yielded. $\square$

3. Optimal sets of n-means and the nth quantization errors for $r = \frac{1}{25}$

Let β_n be the set given by Definition 1.3. In this section, we show that for all $n \geq 2$, the sets β_n form the optimal sets of n-means for $r = \frac{1}{25}$. To calculate the distortion errors we will frequently use the formula given by Corollary 2.3. Notice that by Lemma 2.2, in this case, we have $E(X) = \frac{1}{2}$ and $V := V(X) = \frac{1-r}{6(r+1)} = \frac{2}{13}$.

Lemma 3.1. *The set $\beta := \{a(1), a(2, 3)\}$ forms the optimal set of two-means, and the corresponding quantization error is given by $V_2 = \frac{314}{8125} = 0.0386462$.*

Proof. Let $\beta := \{a_1, a_2\}$ be an optimal set of two-means. Since the points in an optimal set are the conditional expectations in their own Voronoi regions, without any loss of generality, we can assume that $0 < a_1 < a_2 < 1$. Let us consider the set $\kappa := \{a(1), a(2,3)\}$. The distortion error due to the set κ is given by

$$V(P;\kappa) = \int_{J_1} (x - a(1))^2 dP + \int_{J_2 \cup J_3} (x - a(2,3))^2 dP = 0.0386462. \tag{2}$$

Since V_2 is the quantization for two-means, we have $V_2 \leq 0.0386462$. Assume that $0.38 < a_1$. Then,

$$V_2 \geq \int_{J_1} (x - 0.38)^2 dP = 0.0432821 > V_2,$$

which is a contradiction. Hence, $a_1 \leq 0.38$. Similarly, $0.62 \leq a_2$. Since $\frac{1}{2}(a_1+a_2) \leq \frac{1}{2}(0.38+1) = 0.69 < S_3(0) = 0.96$, the Voronoi region of a_1 does not contain any point from J_3. Similarly, the Voronoi region of a_2 does not contain any point from J_1. Since the union of the Voronoi regions of a_1 and a_2 covers $J_1 \cup J_2 \cup J_3$, without any loss of generality, we can assume that the Voronoi region of a_2 contains points from J_2, and $\frac{1}{2}(a_1 + a_2) \leq \frac{1}{2}$. If $\frac{1}{2}(a_1 + a_2) = \frac{1}{2}$, then substituting $r = \frac{1}{25}$, by Proposition 2.8, we have

$$V_2 = \frac{866}{17797} = 0.0486599 > V_2,$$

which leads to a contradiction. Hence, we can conclude that $\frac{1}{2}(a_1 + a_2) < \frac{1}{2}$. Using the similar technique as it is given in the proof of Lemma 3.1 in [15], we can show that $S_1(1) \leq \frac{1}{2}(a_1+a_2) \leq S_2(0)$ yielding the fact that $a_1 = a(1)$, $a_2 = a(2,3)$, and $V_2 = \frac{314}{8125} = 0.0386462$. Hence, the proof of the lemma is complete. $\square$

Lemma 3.2. *The set $\beta := \{a(1), a(2), a(3)\}$ forms an optimal set of three-means, and the corresponding quantization error is given by $V_3 = \frac{2}{8125} = 0.000246154$.*

Proof. Consider the set of three points $\kappa := \{a(1), a(2), a(3)\}$. The distortion error due to the set κ is given by

$$V(P;\kappa) = \sum_{j=1}^{3} \int_{J_j} (x - a(j))^2 dP = \frac{2}{8125} = 0.000246154.$$

Since V_3 is the quantization error for three-means, we have $V_3 \leq 0.000246154$. Let $\beta := \{a_1, a_2, a_3\}$, where $0 < a_1 < a_2 < a_3 < 1$, be an optimal set of three-means. If $S_1(1) = \frac{1}{25} < \frac{1}{23} < a_1$, then

$$V_3 \geq \int_{J_1} (x - \frac{1}{23})^2 dP = \frac{13709}{51577500} = 0.000265794 > V_3,$$

which gives a contradiction. Thus, we can assume that $a_1 \leq \frac{1}{23}$. Similarly, $\frac{22}{23} \leq a_3$. Suppose that $\beta \cap J_1 = \emptyset$. Then, due to symmetry, we can assume that $\beta \cap J_3 = \emptyset$, and then

$$V_3 \geq 2 \int_{J_1} (x - a_1)^2 dP = 2 \int_{J_1} (x - S_1(1))^2 dP = \frac{7}{16250} = 0.000430769 > V_3,$$

which leads to a contradiction. So, we can assume that $\beta \cap J_1 \neq \emptyset$, i.e., $a_1 < S_1(1)$. Similarly, $\beta \cap J_3 \neq \emptyset$, i.e., $S_3(0) < a_3$. Now, we show that $\beta \cap J_2 \neq \emptyset$. Suppose that $\beta \cap J_2 = \emptyset$. Then, either $a_2 < \frac{12}{25} = S_2(0)$, or $\frac{13}{25} = S_2(1) < a_2$. First, assume that $a_2 < S_2(0)$. Then, notice that $S_2(1) = \frac{13}{25} < \frac{1}{2}(S_2(0)+S_3(0)) < S_3(0)$ yielding the fact that the Voronoi region of $S_2(0)$ contains J_2. Hence,

$$V_3 \geq \int_{J_2} (x - S_2(0))^2 dP + \int_{J_3} (x - a(3))^2 dP = \frac{29}{97500} = 0.000297436 > V_3,$$

which is a contradiction. Similarly, we can show that a contradiction arises if $\frac{13}{25} = S_2(1) < a_2$. Thus, we can assume that $\beta \cap J_2 \neq \emptyset$. Now, if the Voronoi region of a_1 contains points from J_2, we have $\frac{1}{2}(a_1 + a_2) > \frac{12}{25} = S_2(0)$ implying $a_2 > \frac{24}{25} - a_1 \geq \frac{24}{25} - \frac{1}{25} = \frac{23}{25} > S_2(1)$, which is a contradiction as $\beta \cap J_2 \neq \emptyset$. Hence, we can assume that the Voronoi region of a_1 does not contain any point from J_2, and so from J_3. Similarly, we can show that the Voronoi region of a_2 does not contain any point from J_1 and J_3, and the Voronoi region of a_3 does not contain any point from J_2, and so from J_1. Thus, by Proposition 1.2, we conclude that $a_1 = a(1)$, $a_2 = a(2)$, and $a_3 = a(3)$, and the corresponding quantization error is given by $V_3 = \frac{2}{8125} = 0.000246154$, which is the lemma. □

Proposition 3.3. *Let β_n be an optimal set of n-means for any $n \geq 3$. Then, $\beta_n \cap J_j \neq \emptyset$ for all $1 \leq j \leq 3$, and β_n does not contain any point from the open intervals $(S_1(1), S_2(0))$ and $(S_2(1), S_3(0))$. Moreover, the Voronoi region of any point in $\beta_n \cap J_j$ does not contain any point from J_i, where $1 \leq i \neq j \leq 3$.*

Proof. By Lemma 3.2, the proposition is true for $n = 3$. Let us prove the lemma for $n \geq 4$. Let $\beta_n := \{a_1, a_2, \cdots, a_n\}$ be an optimal set of n-means for $n \geq 4$. Since the points in an optimal set are the conditional expectations in their own Voronoi regions, without any loss of generality, we can assume that $0 < a_1 < a_2 < \cdots < a_n < 1$. Consider the set of four elements $\kappa := S_1(\beta_2) \cup \{a(2), a(3)\}$. Then,

$$V(P; \kappa) = \int_{J_1} \min_{a \in S_1(\beta_2)} (x-a)^2 dP + \int_{J_2} (x-a(2))^2 dP + \int_{J_3} (x-a(3))^2 dP = \frac{938}{5078125} = 0.000184714.$$

Since V_n is the quantization error for n-means for $n \geq 4$, we have $V_n \leq V_4 \leq 0.000184714$. Suppose that $S_1(1) \leq a_1$. Then,

$$V_n \geq \int_{J_1} (x - S_1(1))^2 dP = \frac{7}{32500} = 0.000215385 > V_n,$$

which is a contradiction. So, we can assume that $a_1 < S_1(1)$, i.e., $\beta_n \cap J_1 \neq \emptyset$. Similarly, $\beta_n \cap J_3 \neq \emptyset$. We now show that $\beta_n \cap J_2 \neq \emptyset$. For the sake of contradiction, assume that $\beta_n \cap J_2 = \emptyset$. Let $a_j := \max\{a_i : a_i < S_2(0) \text{ for } 1 \leq i \leq n-1\}$. Then, $a_j < S_2(0)$. As $\beta_n \cap J_2 = \emptyset$, we have $S_2(1) < a_{j+1}$. If $a_j < \frac{1}{2}(S_1(1) + S_2(0)) = \frac{13}{50}$, then as $\frac{1}{2}(a_j + a_{j+1}) < \frac{1}{2}(\frac{13}{50} + S_2(1)) = \frac{39}{100} < \frac{12}{25} = S_2(0)$, we have

$$V_n \geq \int_{J_2} (x - S_2(1))^2 dP = \frac{7}{32500} = 0.000215385 > V_n,$$

which leads to a contradiction. So, we can assume that $\frac{13}{50} \leq a_j < S_2(0)$. Then, by Proposition 1.2, we have $\frac{1}{2}(a_{j-1} + a_j) < \frac{1}{25}$ implying $a_{j-1} < \frac{2}{25} - a_j \leq \frac{2}{25} - \frac{13}{50} = -\frac{9}{50} < 0$, which gives a contradiction as $\beta_n \cap J_1 \neq \emptyset$. Hence, we can conclude that $\beta_n \cap J_2 \neq \emptyset$. Notice that $(S_1(1), S_2(0)) = (\frac{1}{25}, \frac{12}{25})$. Suppose that β_n contains a point from the open interval $(\frac{1}{25}, \frac{12}{25})$. Let $a_j := \max\{a_i : a_i < \frac{1}{25} \text{ for } 1 \leq i \leq n-2\}$. Then, due to Proposition 1.2, $a_{j+1} \in (\frac{1}{25}, \frac{12}{25})$, and $a_{j+2} \in J_2$. The following cases can arise:

Case 1. $\frac{1}{25} < a_{j+1} \leq \frac{13}{50}$.

Then, $\frac{1}{2}(a_{j+1} + a_{j+2}) > \frac{12}{25}$ implying $a_{j+2} > \frac{24}{25} - a_{j+1} \geq \frac{24}{25} - \frac{13}{50} = \frac{35}{50} > S_2(1)$, which leads to a contradiction because $a_{j+2} \in J_2$.

Case 2. $\frac{13}{50} \leq a_{j+1} < \frac{12}{25}$.

Then, $\frac{1}{2}(a_j + a_{j+1}) < \frac{1}{25}$ implying $a_j \leq \frac{2}{25} - a_{j+1} \leq \frac{2}{25} - \frac{13}{50} = -\frac{9}{50}$, which is a contradiction because $a_j > 0$.

Thus, by Case 1 and Case 2, we can conclude that β_n does not contain any point from the open interval $(S_1(1), S_2(0))$. Reflecting the situation with respect to the point $\frac{1}{2}$, we can conclude that β_n does not contain any point from the open interval $(S_2(1), S_3(0))$ as well. To prove the last part of the proposition, we proceed as follows: Let $a_j := \max\{a_i : a_i < \frac{1}{25} \text{ for } 1 \leq i \leq n-2\}$. Then, a_j is the rightmost element in $\beta_n \cap J_1$, and $a_{j+1} \in \beta_n \cap J_2$. Suppose that the Voronoi region of a_j

contains points from J_2. Then, $\frac{1}{2}(a_j+a_{j+1}) > \frac{12}{25}$ implying $a_{j+1} > \frac{24}{25}-a_j \geq \frac{24}{25}-\frac{1}{25}=\frac{23}{25} > S_2(1)$, which yields a contradiction as $a_{j+1} \in J_2$. Thus, the Voronoi region of any point in $\beta_n \cap J_1$ does not contain any point from J_2, and J_3 as well. Similarly, we can prove that the Voronoi region of any point in $\beta_n \cap J_2$ does not contain any point from J_1 and J_3, and the Voronoi region of any point in $\beta_n \cap J_3$ does not contain any point from J_1 and J_2. Thus, the proof of the proposition is complete. □

The following lemma is a modified version of Lemma 4.5 in [5], and the proof follows similarly. One can also see Lemma 3.5 in [15].

Lemma 3.4. *Let $n \geq 3$, and let β_n be an optimal set of n-means such that $\beta_n \cap J_j \neq \emptyset$ for all $1 \leq j \leq 3$, and β_n does not contain any point from the open intervals $(S_1(1), S_2(0))$ and $(S_2(1), S_3(0))$. Further assume that the Voronoi region of any point in $\beta_n \cap J_j$ does not contain any point from J_i, where $1 \leq i \neq j \leq 3$. Set $\kappa_j := \beta_n \cap J_j$, and $n_j :=$ card (κ_j) for $1 \leq j \leq 3$. Then, $S_j^{-1}(\kappa_j)$ is an optimal set of n_j-means, and $V_n = \frac{1}{1875}(V_{n_1}+V_{n_2}+V_{n_3})$.*

Let us now state and prove the following theorem which gives the optimal sets of n-means for all $n \geq 3$, where $r = \frac{1}{25}$.

Theorem 3.5. *Let P be the probability measure on $\mathbb{R}$ with support the Cantor set C generated by the three contractive similarity mappings S_j for $j = 1, 2, 3$. Let $n \in \mathbb{N}$ with $n \geq 3$. Take $r = \frac{1}{25}$. Then, the sets $\beta_n := \beta_n(I)$ given by Definition 1.3 form the optimal sets of n-means for P with the corresponding quantization error $V_n := V(P; \beta_n(I))$, where $V(P; \beta_n(I))$ is given by Proposition 2.4.*

Proof. We will proceed by induction on $\ell(n)$. If $n = 3$, then by Lemma 3.2, the theorem is true. Now, we show that the theorem is true if $n = 4$. Let $\kappa_j := \beta_n \cap J_j$, and $n_j :=$ card (κ_j) for $1 \leq j \leq 3$. Since $S_j^{-1}(\kappa_j)$ is an optimal set of n_j-means for $1 \leq j \leq 3$, and for $n = 4$ the possible choices for the triplet (n_1, n_2, n_3) are $(2, 1, 1)$, $(1, 2, 1)$, and $(1, 1, 2)$, by Proposition 3.3 and Lemma 3.4, the set β_4 forms an optimal set of four-means with quantization error $V(P; \beta_4)$ given by Proposition 2.4. Remember that for a given n, among all the possible choices of the triplets (n_1, n_2, n_3), the triplets (n_1, n_2, n_3) which give the smallest distortion error will give the optimal sets of n-means. Notice that for $n = 5$, the possible choices of the triplets are $(3, 1, 1)$, $(1, 3, 1)$, $(1, 1, 3)$, $(1, 2, 2)$, $(2, 1, 2)$, $(2, 2, 1)$ among which $(1, 2, 2)$, $(2, 1, 2)$, $(2, 2, 1)$ give the smallest distortion error. Hence, the optimal sets of five-means are $\{a(1)\} \cup S_2(\beta_2) \cup S_3(\beta_2)$, $S_1(\beta_2) \cup \{a(2)\} \cup S_3(\beta_2)$, and $S_1(\beta_2) \cup S_2(\beta_2) \cup \{a(3)\}$ which are the sets β_5 given by Definition 1.3. Similarly, we can calculate the optimal sets of six- and seven-means. Thus, the theorem is true for $\ell(n) = 1$. Let us assume that the theorem is true for all $\ell(n) < m$, where $m \in \mathbb{N}$ and $m \geq 2$. We now show that the theorem is true if $\ell(n) = m$. Let us first assume that $3^m \leq n \leq 2 \cdot 3^m$. Let β_n be an optimal set of n-means for P such that $3^m \leq n \leq 2 \cdot 3^m$. Let card $(\beta_n \cap J_j) = n_j$ for $j = 1, 2, 3$, and then by Lemma 3.4, we have

$$V_n = \frac{1}{1875}(V_{n_1}+V_{n_2}+V_{n_3}). \tag{3}$$

Without any loss of generality, we can assume that $n_1 \geq n_2 \geq n_3$. Let $u, v, w \in \mathbb{N}$ be such that

$$3^u \leq n_1 \leq 2 \cdot 3^u, \ 3^v \leq n_2 \leq 2 \cdot 3^v, \text{ and } 3^w \leq n_3 \leq 2 \cdot 3^w. \tag{4}$$

Proceeding in the similar lines as the proof of Theorem 3.6 in [15], we can show that $u = v = w = m-1$. Since by Lemma 3.4, for $S_j^{-1}(\beta_n \cap J_j)$ is an optimal set of n_j means where $3^{m-1} \leq n_j \leq 2 \cdot 3^{m-1}$, we have

$$S_j^{-1}(\beta_n \cap J_j) = \{a(\omega) : \omega \in \{1,2,3\}^{m-1} \setminus I_j\} \cup \left(\cup_{\omega \in I_j} S_\omega(\beta_2)\right),$$

where $I_j \subseteq \{1, 2, 3\}^{m-1}$ with card $(I_j) = n_j - 3^{m-1}$ for $1 \leq j \leq 3$. Hence,

$$\beta_n := \beta_n(I) = \bigcup_{j=1}^{3} S_j^{-1}(\beta_n \cap J_j) = \{a(\omega) : \omega \in \{1,2,3\}^{\ell(n)} \setminus I\} \cup \left(\cup_{\omega \in I} S_\omega(\beta_2)\right),$$

where $I \subseteq \{1, 2, 3\}^m$ with card $(I) = n - 3^m$, is an optimal set of n-means. The corresponding quantization error is

$$V_n = \frac{1}{3^m} r^{2m} \left((2 \cdot 3^m - n)V + (n - 3^m)V_2\right) = V(P; \beta_n(I)),$$

where $V(P; \beta_n(I))$ is given by Proposition 2.4. Thus, the theorem is true if $3^m \leq n \leq 2 \cdot 3^m$. Similarly, we can prove that the theorem is true if $2 \cdot 3^m < n < 3^{m+1}$. Hence, by the induction principle, the proof of the theorem is complete. □

4. Optimal sets of n-means and the nth quantization errors for $r = r_0$ and $r = r_1$

In this section, we give the proof of Theorem 1.7. First, we prove the following two lemmas.

Lemma 4.1. *Let r_0 and r_1 be the real numbers given by Theorem 1.7. Then, the set $\gamma := \{a(1, 21), a(22, 23, 3)\}$ for $r = r_0$ and $r = r_1$ form the optimal sets of two-means, and the corresponding quantization errors are, respectively, given by $V_2 = 0.0324042$, and $V_2 = 0.026897$.*

Proof. First, we prove that γ forms an optimal set of two-means for $r = r_0$. Let $\gamma := \{a_1, a_2\}$ be an optimal set of two-means. Since, the points in an optimal set are the expected values of their own Voronoi regions, without any loss of generality, we can assume that $0 < a_1 < a_2 < 1$. Let us consider the set $\kappa := \{a(1, 21), a(22, 23, 3)\}$. The distortion error due to the set κ is given by

$$V(P; \kappa) = \int_{J_1} (x - a(1, 21))^2 dP + \int_{J_2 \cup J_3} (x - a(22, 23, 3))^2 dP = 0.0324042. \tag{5}$$

Since V_2 is the quantization error for two-means, we have $V_2 \leq 0.0324042$. Assume that $0.39 < a_1$. Then,

$$V_2 \geq \int_{J_1} (x - 0.39)^2 dP = 0.0328529 > V_2,$$

which is a contradiction. Hence, $a_1 \leq 0.39$. Similarly, $0.61 \leq a_2$. Since $\frac{1}{2}(a_1 + a_2) \leq \frac{1}{2}(0.39 + 1) = 0.695 < S_3(0) = 0.837722$, the Voronoi region of a_1 does not contain any point from J_3. Similarly, the Voronoi region of a_2 does not contain any point from J_1. Since the union of the Voronoi regions of a_1 and a_2 covers $J_1 \cup J_2 \cup J_3$, without any loss of generality, we can assume that the Voronoi region of a_2 contains points from J_2, and $\frac{1}{2}(a_1 + a_2) \leq \frac{1}{2}$. If $\frac{1}{2}(a_1 + a_2) = \frac{1}{2}$, then substituting $r = 0.1622776602$, by Proposition 2.8, we have

$$V(P; \kappa) = 0.0329779,$$

which contradicts (5). Hence, we can conclude that $\frac{1}{2}(a_1 + a_2) < \frac{1}{2}$. Using the similar technique as it is given in the proof of Lemma 3.1 in [15], we can show that either $\frac{1}{2}(a_1 + a_2) = \frac{1}{2}(a(1, 21) + a(22, 23, 3)) = 0.466886$, or $\frac{1}{2}(a_1 + a_2) = \frac{1}{2}(a(1) + a(2, 3)) = 0.395285$, i.e., either $S_{21}(1) < \frac{1}{2}(a_1 + a_2) < S_{22}(0)$, or $S_1(1) < \frac{1}{2}(a_1 + a_2) < S_2(0)$. Notice that if $S_{21}(1) < \frac{1}{2}(a_1 + a_2) < S_{22}(0)$, then γ_2, given by Definition 1.4, forms the optimal set of two-means. On the other hand, if $S_1(1) < \frac{1}{2}(a_1 + a_2) < S_2(0)$, then β_2, given by Definition 1.3, forms the optimal set of two-means. In fact, later we will see that $V(P; \gamma_2) = V(P; \beta_2) = 0.0324042$ for $r = 0.1622776602$. Thus, γ_2 forms the optimal set of two-means for $r = r_0$ with quantization error $V_2 = 0.0324042$. Similarly, we can show that γ_2 forms the optimal set of two-means if $r = r_1$ with quantization error $V_2 = 0.026897$. Hence, the lemma is yielded. □

The following lemma is true analogously as Lemma 3.3 in [15].

Lemma 4.2. *The set $\gamma_3 := \{a(1), a(2), a(3)\}$ for $r = r_0$, and $r = r_1$ form the optimal sets of three-means, and the corresponding quantization errors are, respectively, given by $V_3 = 0.00316342$, and $V_3 = 0.00558347$.*

The following proposition is true analogously as Proposition 3.5 in [15].

Proposition 4.3. *Let $n \geq 3$, and let γ_n be an optimal set of n-means for $r = r_0$, and $r = r_1$. Then, $\gamma_n \cap J_j \neq \emptyset$ for all $1 \leq j \leq 3$, and γ_n does not contain any point from the open intervals $(S_1(1), S_2(0))$ and $(S_2(1), S_3(0))$. Moreover, the Voronoi region of any point in $\gamma_n \cap J_j$ does not contain any point from J_i, where $1 \leq i \neq j \leq 3$.*

The following remark is true due to Proposition 4.3.

Remark 4.4. Let $n \geq 3$, and let γ_n be an optimal set of n-means for $r = r_0$, and $r = r_1$. Set $\kappa_j := \gamma_n \cap J_j$, and $n_j :=$ card (κ_j) for $1 \leq j \leq 3$. Then, $S_j^{-1}(\kappa_j)$ is an optimal set of n_j-means, and for $r = r_0$ and $r = r_1$, respectively, we have $V_n = \frac{1}{3}r_0^n(V_{n_1} + V_{n_2} + V_{n_3})$ and $V_n = \frac{1}{3}r_1^n(V_{n_1} + V_{n_2} + V_{n_3})$.

Proof of Theorem 1.7. We proceed to prove it by induction on $\ell(n)$. By Lemma 4.2, we see that the theorem is true for $n = 3$. Proceeding in the similar way, as mentioned in the proof of Theorem 3.5, we can show that for $n = 4, 5, 6, 7$, the sets γ_n form the optimal sets of n-means for $r = r_0$ and $r = r_1$. Thus, the theorem is true if $\ell(n) = 1$. Let us assume that the theorem is true for all $\ell(n) < m$, where $m \in \mathbb{N}$ and $m \geq 2$. We now show that the theorem is true if $\ell(n) = m$. Let us first assume that $3^m \leq n \leq 2 \cdot 3^m$. Let γ_n be an optimal set of n-means for P such that $3^m \leq n \leq 2 \cdot 3^m$. Let card $(\gamma_n \cap J_j) = n_j$ for $j = 1, 2, 3$, and then by Remark 4.4, we have

$$V_n = \frac{1}{3}r_0^n(V_{n_1} + V_{n_2} + V_{n_3}) \text{ for } r = r_0, \text{ and } V_n = \frac{1}{3}r_1^n(V_{n_1} + V_{n_2} + V_{n_3}) \text{ for } r = r_1.$$

The rest of the proof for $r = r_0$ and $r = r_1$ follow in the similar way as the proof of Theorem 3.5. Thus, we complete the proof of the theorem. □

5. Main results

The two theorems in this section, state and prove the main results of the paper.

Theorem 5.1. *Let $r_0, r_1 \in (0, \frac{1}{3})$ be the unique real numbers satisfying, respectively, the equations*

$$-\frac{3r^5 + 15r^4 + 6r^3 - 42r^2 + 31r - 13}{240(r+1)} = -\frac{3r^3 - 3r^2 + r - 1}{24(r+1)},$$
$$-\frac{3r^5 + 15r^4 + 6r^3 - 42r^2 + 31r - 13}{240(r+1)} = -\frac{3r^7 + 15r^6 + 60r^5 + 66r^4 + 18r^3 - 324r^2 + 283r - 121}{2184(r+1)}.$$

Then, $r_0 = 0.1622776602$, and $r_1 = 0.2317626315$. Let the sets β_n and γ_n be, respectively, given by Definition 1.3, and Definition 1.4. Then, β_n form the optimal sets of n-means for $0 < r \leq r_0$, and γ_n forms the optimal sets of n-means for $r_0 \leq r \leq r_1$.

Proof. By Proposition 2.4, Proposition 2.5, and Proposition 2.7, we see that both β_n and γ_n form CVTs if $0.08502712839 \leq r \leq 0.2472080177$; both γ_n and δ_n form CVTs if $0.1845020699 \leq r \leq 0.2472080177$; both β_n and δ_n form CVTs if $0.1845020699 \leq r \leq 0.2679491924$. Again, $V(P; \beta_3) = V(P; \gamma_3) = V(P; \delta_3)$. Thus, for any $3^{\ell(n)} \leq n < 3^{\ell(n)+1}$, from the aforementioned propositions, in the case of $V(P; \beta_n(I))$ and $V(P; \gamma_n(I))$, we see that $V(P; \beta_n(I)) > V(P; \gamma_n(I))$, $V(P; \beta_n(I)) = V(P; \gamma_n(I))$, and $V(P; \beta_n) < V(P; \gamma_n)$ will be true if $V(P; \beta_2) > V(P; \gamma_2)$, $V(P; \beta_2) = V(P; \gamma_2)$, and $V(P; \beta_2) < V(P; \gamma_2)$, respectively. Similarly, it hold in the case of $V(P; \beta_n)$ and $V(P; \delta_n)$, and in the case of $V(P; \gamma_n)$ and $V(P; \delta_n)$. Next, we have

$$V(P; \beta_2) = -\frac{3r^3 - 3r^2 + r - 1}{24(r+1)},$$
$$V(P; \gamma_2) = -\frac{3r^5 + 15r^4 + 6r^3 - 42r^2 + 31r - 13}{240(r+1)},$$
$$V(P; \delta_2) = -\frac{3r^7 + 15r^6 + 60r^5 + 66r^4 + 18r^3 - 324r^2 + 283r - 121}{2184(r+1)}.$$

After some calculation, we observe that $V(P;\beta_2) < V(P;\gamma_2)$ is true if $0.08502712839 \leq r < 0.1622776602$; $V(P;\beta_2) = V(P;\gamma_2)$ if $r = 0.1622776602$, and $V(P;\beta_2) > V(P;\gamma_2)$ if $0.1622776602 < r \leq 0.2472080177$. Again, $V(P;\beta_2) > V(P;\delta_2)$ if $0.1701473031 < r \leq 0.2679491924$ and $V(P;\beta_2) = V(P;\delta_2)$ if $r = 0.1701473031$. Recall that the sets β_n form CVTs if $0 < r \leq 0.2679491924$. Hence, we can say that the sets β_n do not form the optimal sets of n-means if $0.1622776602 < r \leq 0.2679491924$. In Theorem 1.7, we have seen that the sets β_n form the optimal sets of n-means if $r = \frac{1}{25}$. Using the similar technique, we can show that the sets β_n form the optimal sets of n-means if $0 < r \leq \frac{1}{25}$. Since $V(P;\beta_2) = V(P;\gamma_2)$ if $r = r_0$; and by Theorem 1.7, the sets γ_n form the optimal sets of n-means if $r = r_0$, we can say that the sets β_n also form the optimal sets of n-means if $r = r_0$. Again, $V(P;\beta_2)$ is strictly decreasing in the closed interval $[0, r_0]$. Hence, the sets β_n form the optimal sets of n-means for $0 < r \leq r_0$.

To prove the remaining part of the theorem, we see that

(*i*) $V(P;\beta_2) < V(P;\gamma_2)$ if $0.08502712839 \leq r < 0.1622776602$; $V(P;\beta_2) = V(P;\gamma_2)$ if $r = 0.1622776602$, and $V(P;\beta_2) > V(P;\gamma_2)$ if $0.1622776602 < r \leq 0.2472080177$.

(*ii*) $V(P;\delta_2) < V(P;\gamma_2)$ if $0.2317626315 < r \leq 0.2472080177$; $V(P;\delta_2) = V(P;\gamma_2)$ if $r = 0.2317626315$, and $V(P;\delta_2) > V(P;\gamma_2)$ if $0.1845020699 \leq r < 0.2317626315$.

Thus, the sets γ_n do not form the optimal sets of n-means if $0.08502712839 \leq r < 0.1622776602$, or if $0.2317626315 < r \leq 0.2472080177$; in other words, the range of r values for which the sets γ_n form the optimal sets of n-means is bounded below by $r_0 = 0.1622776602$ and bounded above by $r_1 = 0.2317626315$. By Theorem 1.7, we see that the sets γ_n form the optimal sets of n-means if $r = r_0$, and $r = r_1$. Again, $V(P;\gamma_2)$ is strictly decreasing in the closed interval $[r_0, r_1]$. Hence, the precise range of r values for which the sets γ_n form the optimal sets of n-means is given by $r_0 \leq r \leq r_1$. Thus, the proof of the theorem is complete. $\square$

Since the Cantor set C under investigation satisfies the strong separation condition, with each S_j having contracting factor of r, the Hausdorff dimension of the Cantor set is equal to the similarity dimension. Hence, from the equation $3(r)^\beta = 1$, we have $\dim_{\mathrm{H}}(C) = \beta = -\frac{\log 3}{\log r}$. By Theorem 14.17 in [4], the quantization dimension $D(P)$ exists and is equal to β. In Theorem 5.2, we show that β dimensional quantization coefficient for P does not exist.

Theorem 5.2. *The β-dimensional quantization coefficient for $0 < r \leq r_1$ does not exist.*

Proof. We have $3^{\frac{1}{\beta}} = \frac{1}{r}$. Notice that $\left\{\left(3^{\ell(n)}\right)^{\frac{2}{\beta}} V_{3^{\ell(n)}}(P)\right\}$ and $\left\{\left(2\cdot 3^{\ell(n)}\right)^{\frac{2}{\beta}} V_{2\cdot 3^{\ell(n)}}(P)\right\}$ are two different subsequences of the sequence $\left\{n^{\frac{2}{\beta}} V_n(P)\right\}$. First, assume that $0 < r \leq r_0$. Then, by Theorem 5.1, β_n is an optimal set of n-means for $0 < r \leq r_0$. Recall Proposition 2.4. Then, we have

$$\lim_{n\to\infty} \left(3^{\ell(n)}\right)^{\frac{2}{\beta}} V_{3^{\ell(n)}}(P) = \lim_{n\to\infty} \frac{1}{r^{2\ell(n)}} \frac{1}{3^{\ell(n)}} r^{2\ell(n)} 3^{\ell(n)} V = V, \tag{6}$$

and

$$\lim_{n\to\infty} \left(2\cdot 3^{\ell(n)}\right)^{\frac{2}{\beta}} V_{2\cdot 3^{\ell(n)}}(P) = \lim_{n\to\infty} 2^{\frac{2}{\beta}} \frac{1}{r^{2\ell(n)}} \frac{1}{3^{\ell(n)}} r^{2\ell(n)} 3^{\ell(n)} V(P;\beta_2) = 2^{\frac{2}{\beta}} V(P;\beta_2). \tag{7}$$

By (6) and (7), we see that $\left\{n^{\frac{2}{\beta}} V_n(P)\right\}$ has two different subsequences having two different limits, and so $\lim_{n\to\infty} n^{\frac{2}{\beta}} V_n(P)$ does not exist. Due to Theorem 5.1, and Proposition 2.5, similarly, we can show that if $r_0 \leq r \leq r_1$, then $\lim_{n\to\infty} n^{\frac{2}{\beta}} V_n(P)$ does not exist. Thus, we show that the β-dimensional quantization coefficient for $0 < r \leq r_1$ does not exist, which completes the proof of the theorem. $\square$

Acknowledgement. The author would like to express his sincere gratitude to the referees for their valuable comments.

References

[1] Du, Q., Faber, V., Gunzburger, M.: *Centroidal Voronoi Tessellations: Applications and Algorithms.* Siam Rev., 41(4), 637-676 (1999)

[2] Dettmann, C.P. and Roychowdhury, M.K.: *Quantization for uniform distributions on equilateral triangles.* Real Analysis Exchange, 42(1), 149-166 (2017)

[3] Gersho, A. and Gray, R.M.: *Vector quantization and signal compression.* Kluwer Academy publishers, Boston (1992)

[4] Graf, S. and Luschgy, H.: *Foundations of quantization for probability distributions.* Lecture Notes in Mathematics 1730, Springer, Berlin (2000)

[5] Graf, S. and Luschgy, H.: *The quantization of the Cantor distribution.* Math. Nachr., 183, 113-133 (1997)

[6] Graf, S. and Luschgy, H.: *Asymptotics of the quantization error for self-similar probabilities.* Real Anal. Exchange, 26, 795-810 (2001)

[7] Graf, S. and Luschgy, H.: *Quantization for probability measures with respect to the geometric mean error.* Math. Proc. Camb. Phil. Soc., 136, 687-717 (2004)

[8] Graf, S., Luschgy, H. and Pagés, G.: *The local quantization behavior of absolutely continuous probabilities.* Ann. Probab. 40(4), 1795-1828 (2012)

[9] Gray, R. and Neuhoff, D.: *Quantization.* IEEE Trans. Inform. Theory, 44, 2325-2383 (1998)

[10] Hutchinson, J.: *Fractals and self-similarity.* Indiana Univ. J., 30, 713-747 (1981)

[11] Pötzelberger, K.: *The quantization dimension of distributions.* Math. Proc. Camb. Phil. Soc., 131, 507-519 (2001)

[12] Pagès, G.: *A space quantization method for numerical integration.* J. Comput. Appl. Math., 89, 1-38 (1998)

[13] Pagès, G. and Printems, J.: *Functional quantization for numerics with an application to option pricing.* Monte Carlo Methods Appl., 11, 407-446 (2005)

[14] Roychowdhury, M.K.: *Least upper bound of the exact formula for optimal quantization of some uniform Cantor distributions.* Discrete and Continuous Dynamical Systems- Series A, 38(9), 4555-4570 (2018)

[15] Roychowdhury, M.K.: *The quantization of the standard triadic Cantor distribution.* Houston Journal of Mathematics, 46(2), 389-407 (2020)

[16] Rosenblatt, J. and Roychowdhury, M.K.: *Optimal quantization for piecewise uniform distributions.* Uniform Distribution Theory, 13(2), 23-55 (2018)

[17] Zador, P.L.: *Development and Evaluation of Procedures for Quantizing Multivariate Distributions.* Ph.D. Thesis (Stanford University, (1964)

School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, 1201 West University Drive, Edinburg, TX 78539-2999, USA.

Email address: `mrinal.roychowdhury@utrgv.edu`