7-23-2020

Quantization for a Mixture of Uniform Distributions Associated with Probability Vectors

Mrinal Kanti Roychowdhury
The University of Texas Rio Grande Valley

Wasiela Salinas
The University of Texas Rio Grande Valley

Recommended Citation

Roychowdhury, Mrinal Kanti and Salinas, Wasiela. "Quantization for a Mixture of Uniform Distributions Associated with Probability Vectors" Uniform distribution theory, vol.15, no.1, 2020, pp.105-142. https://doi.org/10.2478/udt-2020-0006 sciendo

uniform distribution theory

DOI: 10.2478/udt-2020–0006
Unif. Distrib. Theory **15** (2020), no.1, 105–142

QUANTIZATION FOR A MIXTURE OF UNIFORM DISTRIBUTIONS ASSOCIATED WITH PROBABILITY VECTORS

Mrinal Kanti Roychowdhury — Wasiela Salinas

School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, USA

ABSTRACT. The basic goal of quantization for probability distribution is to reduce the number of values, which is typically uncountable, describing a probability distribution to some finite set and thus approximation of a continuous probability distribution by a discrete distribution. Mixtures of probability distributions, also known as mixed distributions, are an exciting new area for optimal quantization. In this paper, we investigate the optimal quantization for three different mixed distributions generated by uniform distributions associated with probability vectors.

Communicated by Manfred Kühleitner

1. Introduction

Continuous-valued signals can take any real value either in the entire range of real numbers or in a range limited by some system constraints. In either of the two cases, an uncountably infinite set of values is required to represent the signal values. If a signal has to be processed or stored digitally, each of its values must be representable by a finite number of bits. Thus, all values together have to form a finite countable set. A signal consisting only of such discrete values is said to be quantized. The process of transformation of a continuous-valued signal into a discrete-valued one is called 'quantization'. It has broad application in engineering and technology (see [GG, GN, Z]). For mathematical treatment

2010 Mathematics Subject Classification: 60Exx, 94A34.

Keywords: Mixed distribution, uniform distribution, optimal sets, quantization error, quantization dimension, quantization coefficient.

of quantization one is referred to Graf-Luschgy's book (see [GL1]). Let $\mathbb{R}^d$ denote the d-dimensional Euclidean space equipped with the Euclidean norm $\|\cdot\|$, and let P be a Borel probability measure on $\mathbb{R}^d$. Then, the nth *quantization error* for P, with respect to the squared Euclidean distance, is defined by

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

represents the distortion error for P due to the set α. 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 known that for a continuous Borel probability measure an optimal set of n-means always has exactly n-elements (see [GL1]). Optimal sets of n-means for different probability distributions were calculated by several authors, for example, one can see [CR, DR1, DR2, GL2, RR1, L1, R1–R5]. The number

$$\lim_{n\to\infty} \frac{2\log n}{-\log V_n(P)},$$

if it exists, is called the *quantization dimension* of the probability measure P, and is denoted by $D(P)$; on the other hand, for any $s \in (0,+\infty)$, 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 (see [GL1, P]).

Let us now state the following proposition (see [GG, GL1]):

Proposition 1.1. *Let α be an optimal set of n-means for P, and $a \in \alpha$. Then,* (i) $P\big(M(a|\alpha)\big) > 0$, (ii) $P\big(\partial M(a|\alpha)\big) = 0$, (iii) $a = E\big(X : X \in M(a|\alpha)\big)$, *where $M(a|\alpha)$ is the Voronoi region of $a \in \alpha$, i.e., $M(a|\alpha)$ is the set of all elements x in $\mathbb{R}^d$ which are closest to a among all the elements in α.*

Proposition 1.1 says that if α is an optimal set and $a \in \alpha$, then a is the *conditional expectation* of the random variable X given that X takes values in the Voronoi region of a. The following theorem is known.

Theorem 1.2 (see [RR2]). *Let P be a uniform distribution on the closed interval $[a,b]$. Then, the optimal set n-means is given by $\alpha_n := \{a + \frac{2i-1}{2n}(b-a) : 1 \le i \le n\}$, and the corresponding quantization error is $V_n := V_n(P) = \frac{(a-b)^2}{12n^2}$.*

Theorem 1.3. *Let α_n be an optimal set of n-means for a uniform distribution on the unit circular arc S given by*

$$S := \{(\cos\theta, \sin\theta) : \alpha \le \theta \le \beta\}, \quad \textit{where} \quad 0 \le \alpha < \beta \le 2\pi.$$

Then,

$$\alpha_n := \Big\{ \frac{2n}{\beta-\alpha}\sin\Big(\frac{\beta-\alpha}{2n}\Big)\Big(\cos\Big(\alpha+(2j-1)\frac{\beta-\alpha}{2n}\Big), \sin\Big(\alpha+(2j-1)\frac{\beta-\alpha}{2n}\Big)\Big) : j=1,2,\dots,n\Big\}$$

forms an optimal set of n-means, and the corresponding quantization error is given by

$$V_n = \frac{(\alpha-\beta)^2 - 2n^2 + 2n^2\cos\frac{\alpha-\beta}{n}}{(\alpha-\beta)^2}.$$

Proof. Notice that S is an arc of the unit circle $x_1^2+x_2^2=1$ which subtends a central angle of $\beta-\alpha$ radian, and the probability distribution is uniform on S. Hence, the density function is given by $f(x_1,x_2)=\frac{1}{\beta-\alpha}$ if $(x_1,x_2)\in S$, and zero, otherwise. Thus, the proof follows in the similar way as the proof in the similar theorem in [RR2]. □

Mixed distributions are an exciting new area for optimal quantization. For any two Borel probability measures P_1 and P_2, and $p\in(0,1)$, if $P := pP_1+(1-p)P_2$, then the probability measure P is called the *mixture* or the *mixed distribution* generated by the probability measures (P_1,P_2) associated with the probability vector $(p,1-p)$. Such kind of problems has rigorous applications in many areas including signal processing. For example, while driving long distances, we have seen sometimes cellular signals get cut off. This happens because of being far away from the tower, or there is no tower nearby to catch the signal. In optimal quantization for mixed distributions one of our goals is to find the exact locations of the towers by giving different weights, also called importance, to different portions of a path.

The following theorem about the quantization dimension for the mixed distributions is well-known. For some more details please see [L, Theorem 2.1].

Theorem 1.4. *Let P_1 and P_2 be any two Borel probability measures on $\mathbb{R}^d$ such that both $D(P_1)$ and $D(P_2)$ exist. If $P=pP_1+(1-p)P_2$, where $0<p<1$, then $D(P)=\max\{D(P_1),D(P_2)\}$.*

In this paper, in Section 2, we have considered a mixed distribution generated by two uniform distributions on a circle and on one of its diameters associated with the probability vector $(\frac{1}{2},\frac{1}{2})$. For this mixed distribution, in Theorem 2.10, we have explicitly determined the optimal sets of n-means and the nth quantization errors for all positive integers $n\geq 2$. In Proposition 2.12, we have proved that the quantization dimension $D(P)$ of the mixed distribution is one, which supports Theorem 1.4 because $D(P_1) = D(P_2) = 1$, and the quantization coefficient exists as a finite positive number which equals $\frac{3}{8}\left(4+\pi^2\right)$. Optimal sets of n-means and the nth quantization errors are calculated, in Section 3, for the mixture of two uniform distributions on two disconnected line segments $[0, \frac{1}{2}]$ and $[\frac{3}{4}, 1]$ associated with the probability vector $(\frac{3}{4}, \frac{1}{4})$, and in Section 4, for the mixture of two uniform distributions on two connected line segments $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ associated with the probability vector $(\frac{3}{4}, \frac{1}{4})$. We would like to mention that in these two sections, to determine the optimal sets of n-means and the nth quantization errors for the mixed distributions we need to take the help of two different sequences $\{a(n)\}_{n=1}^{\infty}$ given by Definition 3.8, and Definition 4.6. If the probability vector $(\frac{3}{4}, \frac{1}{4})$ is replaced by some other probability vector $(p, 1-p)$, where $0 < p < 1$, what will be the two such sequences are not known yet. In fact, optimal sets of n-means and the nth quantization errors are not known yet for a more general mixed distribution.

2. Quantization for a mixed distribution on the circles including a diameter

Let i and j be the unit vectors in the positive directions of the x_1- and x_2-axes, respectively. By the position vector a of a point A, it is meant that $\overrightarrow{OA} = a$. We will identify the position vector of a point (a_1, a_2) by $(a_1, a_2) := a_1 i + a_2 j$, and apologize for any abuse in notation. For any two position vectors $a := (a_1, a_2)$ and $b := (b_1, b_2)$, we write $\rho(a, b) := \|(a_1, b_1) - (a_2, b_2)\|^2 = (a_1 - a_2)^2 + (b_1 - b_2)^2$, which gives the squared Euclidean distance between the two points (a_1, a_2) and (b_1, b_2). Let P and Q belong to an optimal set of n-means for some positive integer n, and let D be a point on the boundary of the Voronoi regions of the points P and Q. Since the boundary of the Voronoi regions of any two points is the perpendicular bisector of the line segment joining the points, we have $|\overrightarrow{DP}| = |\overrightarrow{DQ}|$, i.e., $(\overrightarrow{DP})^2 = (\overrightarrow{DQ})^2$ implying $(p-d)^2 = (q-d)^2$, i.e., $\rho(d, p) - \rho(d, q) = 0$. We call such an equation a *canonical equation*. By $E(X)$ and $V := V(X)$, we represent the expectation and the variance of a random variable X with respect to the probability distribution under consideration.

Let P_1 be the uniform distribution defined on the circle $x_1^2 + x_2^2 = 1$ with center $O(0, 0)$, and P_2 be the uniform distribution on one of its diameters. Let us denote the diameter by L_1 and the circle by L_2. Without any loss of generality, we can assume that the diameter is horizontal, i.e., the diameter is represented by

$$L_1 := \{(x_1, 0) : -1 \le x_1 \le 1\}$$

which intersects the circle at the two points $A(-1,0)$ and $B(0,1)$. Let L be the path formed by the circle and the diameter AB. Thus, we have

$$L = L_1 \cup L_2,$$

where

$$L_1 = \{(t,0) : -1 \leq t \leq 1\}, \quad \text{and} \quad L_2 = \{(\cos\theta, \sin\theta) : 0 \leq \theta \leq 2\pi\}.$$

Let s represent the distance of any point on L from the origin tracing along the boundary L in the positive direction of the x_1-axis, and in the counterclockwise direction. Thus, $s = 1$ represents the point $B(1,0)$, $s = 1 + \frac{\pi}{2}$ represents the point $(0,-1)$, and so on. Take the mixed distribution P as

$$P := \frac{1}{2}P_1 + \frac{1}{2}P_2,$$

i.e., P is generated by (P_1, P_2) associated with the probability vector $(\frac{1}{2}, \frac{1}{2})$. For this mixed distribution P in this section, we determine the optimal sets of n-means and the nth quantization errors for all $n \in \mathbb{N}$. The probability density function (pdf) $f(x_1, x_2)$ for the mixed distribution P is given by

$$f(x_1, x_2) = \begin{cases} \frac{1}{4} & \text{if } (x_1, x_2) \in L_1, \\ \frac{1}{4\pi} & \text{if } (x_1, x_2) \in L_2. \end{cases}$$

On L_1 we have $\mathrm{d}s = \sqrt{(\frac{\mathrm{d}x_1}{\mathrm{d}t})^2 + (\frac{\mathrm{d}x_2}{\mathrm{d}t})^2}\,\mathrm{d}t = \mathrm{d}t$ yielding $\mathrm{d}P(s) = P(\mathrm{d}s) = f(x_1, x_2)\,\mathrm{d}s = \frac{1}{4}\,\mathrm{d}t$. Similarly, on L_2, we have $\mathrm{d}s = \mathrm{d}\theta$ yielding $\mathrm{d}P(s) = P(\mathrm{d}s) = f(x_1, x_2)\,\mathrm{d}s = \frac{1}{4\pi}\,\mathrm{d}\theta$.

Lemma 2.1. *Let X be a continuous random variable with mixed distribution taking values on L. Then,*

$$E(X) = (0,0) \quad \textit{and} \quad V := V(X) = \frac{2}{3}.$$

Proof. We have,

$$E(X) = \int_L (x_1 i + x_2 j)\,\mathrm{d}P = \frac{1}{4}\int_{L_1}(t,0)\,\mathrm{d}t + \frac{1}{4\pi}\int_{L_2}(\cos\theta, \sin\theta)\,\mathrm{d}\theta = (0,0).$$

To calculate the variance, we know that $V(X) = E\|X - E(X)\|^2$, which implies

$$V(X) = \frac{1}{4}\int_{L_1}\rho\big((t,0),(0,0)\big)\,\mathrm{d}t + \frac{1}{4\pi}\int_{L_2}\rho\big((\cos\theta, \sin\theta),(0,0)\big)\,\mathrm{d}\theta = \frac{2}{3}.$$

Thus, the lemma is yielded. □

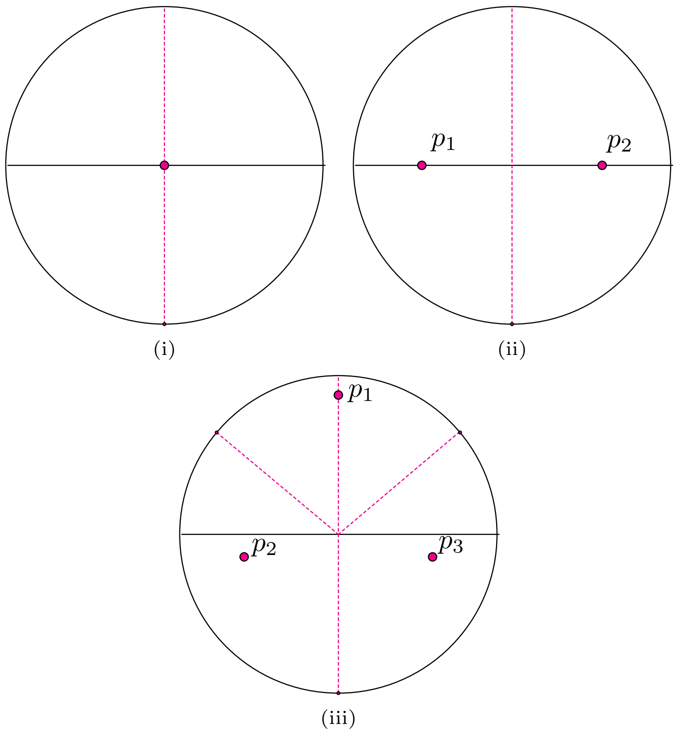

FIGURE 1.

REMARK 2.2. Using the standard theory of probability, for any $(a,b) \in \mathbb{R}^2$, we have

$$E\|X-(a,b)\|^2 = \int_L \|(x_1,x_2)-(a,b)\|^2 \, \mathrm{d}P = V(X) + \|(a,b)-(0,0)\|^2,$$

which is minimum if $(a,b)=(0,0)$, and the minimum value is $V(X)$. Thus, we see that the optimal set of one-mean is the set $\{(0,0)\}$, and the corresponding quantization error is the variance $V := V(X)$ of the random variable X (see Figure 1 (i)).

PROPOSITION 2.3. *The set* $\left\{\left(-\frac{1}{4}-\frac{1}{\pi},0\right),\left(\frac{1}{4}+\frac{1}{\pi},0\right)\right\}$ *forms the optimal set of two-means, and the corresponding quantization error is given by* $V_2 = 0.343691$.

Proof. Since P is a mixed distribution giving the equal weights to both the component probabilities P_1 and P_2, and the path L is symmetric with respect to the x_2-axis, without going into much calculation, we can assume that the boundary of the Voronoi regions of the two points in an optimal set of two-means lies along the x_2-axis. Thus, the optimal set of two-means is given by $\{p_1, p_2\}$ (see Figure 1 (ii)), where

$$\begin{aligned} p_1 &= E\big(X : X \in \overline{AO} \,\cup \quad \text{(left half of the circle)}\big) \\ &= \frac{\frac{1}{4}\int_{-1}^{0}(x,0)\,\mathrm{d}x + \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}}(\cos\theta,\sin\theta)\,\mathrm{d}\theta}{\frac{1}{4}\int_{-1}^{0}\,\mathrm{d}x + \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}}\,\mathrm{d}\theta} \\ &= \Big(-\frac{1}{4}-\frac{1}{\pi},0\Big), \end{aligned}$$

and similarly, $p_2 = \big(\frac{1}{4}+\frac{1}{\pi},0\big)$. The quantization error for two-means is given by

$$V_2 = 2\bigg(\frac{1}{4}\int_{-1}^{0}\rho\big((x,0),p_1\big)\,\mathrm{d}x + \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}}\rho\big((\cos\theta,\sin\theta),p_1\big)\,\mathrm{d}\theta\bigg) = 0.343691.$$

Thus, the proposition is yielded. □

The following proposition gives the optimal set of three-means (see Figure 1 (iii)). The proof follows in the similar way as Proposition 2.5 which is given later.

Proposition 2.4. *The set*

$$\{(0,0.877439),(-0.593906,-0.14179),(0.593906,-0.14179)\}$$

forms an optimal set of three-means, and the corresponding quantization error is given by $V_3 = 0.2386$.

Proposition 2.5. *The set*

$$\{(0,0.90407),(-0.633881,0),(0,-0.90407),(0.633881,0)\}$$

forms an optimal set of four-means, and the corresponding quantization error is given by $V_4 = 0.163013$.

Proof. Let $\alpha := \{p_1, p_2, p_3, p_4\}$ be an optimal set of four-means. The following cases can arise:

Case 1. α contains one point from L_1, the Voronoi region of which does not contain any point from L_2.

In this case, we can assume that p_1, p_2, p_3, p_4 can be located as shown in Figure 2 (i). Let the boundary of the Voronoi regions of p_1 and p_2 intersect L_2 at the point d_1 given by the parametric value $\theta = \pi - b$, where $0 < b < \frac{\pi}{2}$, and the boundary of the Voronoi regions of p_2 and p_3 intersect L_1 at the point d_2 given by $x_1 = -a$, where $0 < a < 1$. Thus, due to symmetry, we have

$$p_1 = \frac{\int_b^{\pi-b}(\cos\theta, \sin\theta)\ \mathrm{d}\theta}{\int_b^{\pi-b}\ \mathrm{d}\theta} = \Big(0, \frac{2\cos b}{\pi - 2b}\Big),$$

$$p_2 = \frac{\frac{1}{4}\int_{-1}^{-a}(x,0)\,\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\frac{3\pi}{2}}(\cos\theta, \sin\theta)\ \mathrm{d}\theta}{\frac{1}{4}\int_{-1}^{-a}\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\frac{3\pi}{2}}\ \mathrm{d}\theta}$$

$$= \Big(\frac{-\pi a^2 + 2\sin b + \pi + 2}{\pi(2a-3) - 2b}, -\frac{2\cos b}{-2\pi a + 2b + 3\pi}\Big),$$

$$p_3 = (0,0), \quad d_1 = (-\cos b, \sin b), \text{ and } d_2 = (-a, 0).$$

Thus, solving the canonical equations $\rho(d_1, p_1) - \rho(d_1, p_2) = 0$, and $\rho(d_2, p_2) - \rho(d_2, p_3) = 0$, we have $a = 0.377997$, $b = 0.678642$. Hence, putting the values of a and b we have, $p_1 = (0, 0.872524)$, $p_2 = (-0.707525, -0.185184)$, and $p_3 = (0,0)$, and so, due to symmetry, $p_4 = (0.707525, -0.185184)$. The corresponding distortion error is given by

$$V(P;\alpha) = \frac{1}{4\pi}\int_b^{\pi-b}\rho\big((\cos\theta, \sin\theta), p_2\big)\ \mathrm{d}\theta + 2\Big(\frac{1}{4}\int_{-1}^{-a}\rho((x,0), p_2)\,\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\frac{3\pi}{2}}\rho\big((\cos\theta, \sin\theta), p_2\big)\ \mathrm{d}\theta\Big) + \frac{1}{4}\int_{-a}^{a}\rho((x,0), p_3)\,\mathrm{d}x = 0.21596.$$

Case 2. α does not contain any point from L_1, the Voronoi region of which does not contain any point from L_2.

In this case, we can assume that p_1, p_2, p_3, p_4 can be located as shown in Figure 2 (ii). Let the boundary of the Voronoi regions of p_1 and p_2 intersect L_2 at the point d_1 given by the parametric value $\theta = \pi - b$, where $0 < b < \frac{\pi}{2}$. Thus, due to symmetry, we have

$$p_1 = \frac{\int_b^{\pi-b}(\cos\theta, \sin\theta)\ \mathrm{d}\theta}{\int_b^{\pi-b}\ \mathrm{d}\theta} = \Big(0, \frac{2\cos b}{\pi - 2b}\Big),$$

$$p_2 = \frac{\frac{1}{4}\int_{-1}^{0}(x,0)\,\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\pi+b}(\cos\theta, \sin\theta)\ \mathrm{d}\theta}{\frac{1}{4}\int_{-1}^{0}\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\pi+b}\ \mathrm{d}\theta} = \Big(-\frac{4\sin b + \pi}{4b + 2\pi}, 0\Big),$$

and

$$d_1 = (-\cos b, \sin b).$$

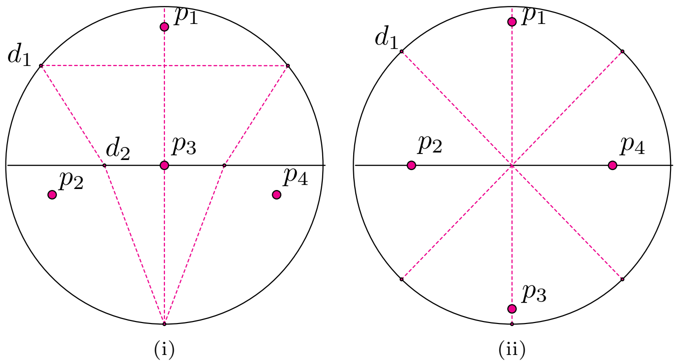

FIGURE 2.

Thus, solving the canonical equations $\rho(d_1, p_1) - \rho(d_1, p_2) = 0$, we have $b = 0.800791$. Hence, putting the values of b, we have, $p_1 = (0, 0.90407)$, $p_2 = (-0.633881, 0)$, and so, due to symmetry, we have $p_3 = (0.633881, 0)$, and $p_4 = (0, -0.90407)$. The corresponding distortion error is given by

$$V(P; \alpha) = 2\Big(\frac{1}{4\pi}\int_b^{\pi-b} \rho\big((\cos\theta, \sin\theta), p_2\big)\, d\theta + \frac{1}{4}\int_{-1}^{0} \rho\big((x, 0), p_2\big)\, dx + \frac{1}{4\pi}\int_{\pi-b}^{\pi+b} \rho\big((\cos\theta, \sin\theta), p_2\big)\, d\theta\Big) = 0.163013.$$

Comparing Case 1 and Case 2, we see that if α contains only one point from L_1, the Voronoi regions of which does not contain any point from L_2, then the distortion error is larger than the distortion error obtained in Case 2. Similarly, we can show that if α contains more than one point from L_1, the Voronoi regions of which do not contain any point from L_2, then the distortion error is larger than the distortion error obtained in Case 2. Considering all the above cases, we see that the distortion error in Case 2 is the smallest. Hence, the points in α obtained in Case 2 form an optimal set of four-means, and the corresponding quantization error is given by $V_4 = 0.163013$. Thus, the proof of the proposition is complete. □

PROPOSITION 2.6. *An optimal set of five-means is given by*

$$\{(0, 0.903584), (-0.788308, 0), (0, 0), (0, -0.903584), (0.788308, 0)\}$$

and the corresponding quantization error is $V_5 = 0.119779$.

P r o o f. Let $\alpha := \{p_1, p_2, p_3, p_4, p_5\}$ be an optimal set of five-means. The following cases can arise:

Case 1. α contains two points from L_1, the Voronoi regions of which do not contain any point from L_2.

In this case, we can assume that $p_1, p_2, \ldots, p_5$ can be located as shown in Figure 3 (i). Let the boundary of the Voronoi regions of p_1 and p_2 intersect L_2 at the point d_1 given by the parametric value $\theta = \pi - b$, where $0 < b < \frac{\pi}{2}$, and the boundary of the Voronoi regions of p_2 and p_3 intersect L_1 at the point d_2 given by $x_1 = -a$, where $0 < a < 1$. Thus, due to symmetry, we have

$$p_1 = \frac{\int_b^{\pi-b}(\cos\theta, \sin\theta)\ \mathrm{d}\theta}{\int_b^{\pi-b}\ \mathrm{d}\theta} = \Big(0, \frac{2\cos b}{\pi - 2b}\Big),$$

$$p_2 = \frac{\frac{1}{4}\int_{-1}^{-a}(x,0)\,\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\frac{3\pi}{2}}(\cos\theta, \sin\theta)\ \mathrm{d}\theta}{\frac{1}{4}\int_{-1}^{-a}\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\frac{3\pi}{2}}\mathrm{d}\theta}$$

$$= \Big(\frac{-\pi a^2 + 2\sin b + \pi + 2}{\pi(2a-3) - 2b}, -\frac{2\cos b}{-2\pi a + 2b + 3\pi}\Big),$$

$$p_3 = (-\frac{a}{2}, 0), \quad d_1 = (-\cos b, \sin b), \quad \textit{and} \quad d_2 = (-a, 0).$$

Thus, solving the canonical equations $\rho(d_1, p_1) - \rho(d_1, p_2) = 0$, and $\rho(d_2, p_2) - \rho(d_2, p_3) = 0$, we have $a = 0.567815$, $b = 0.656426$. Hence, putting the values of a and b we have, $p_1 = (0, 0.866365)$, $p_2 = (-0.74607, -0.220972)$, and $p_3 = (-0.283907, 0)$, and so, due to symmetry, $p_4 = (0.283907, 0)$, and $p_5 = (0.74607, -0.220972)$. The corresponding distortion error is given by

$$V(P;\alpha) = \frac{1}{4\pi}\int_b^{\pi-b}\rho\big((\cos\theta,\sin\theta),p_2\big)\ \mathrm{d}\theta + 2\Big(\frac{1}{4}\int_{-1}^{-a}\rho((x,0),p_2)\,\mathrm{d}x + \frac{1}{4\pi}\int_{\pi-b}^{\frac{3\pi}{2}}\rho\big((\cos\theta,\sin\theta),p_2\big)\ \mathrm{d}\theta + \frac{1}{4}\int_{-a}^{0}\rho\big((x,0),p_3\big)\,\mathrm{d}x\Big) = 0.18911.$$

Case 2. α contains only one point from L_1, the Voronoi region of which does not contain any point from L_2.

In this case, we can assume that $p_1, p_2, \ldots, p_5$ can be located as shown in Figure 3 (ii). Let the boundary of the Voronoi regions of p_1 and p_2 intersect L_2 at the point d_1 given by the parametric value $\theta = \pi - b$, where $0 < b < \frac{\pi}{2}$, the boundary of the Vonoroi regions of p_2 and p_3 intersect L_1 at the point d_2

given by $x_1 = -a$, where $0 < a < 1$. Thus, due to symmetry, we have

$$p_1 = \frac{\int_b^{\pi-b} (\cos\theta, \sin\theta)\ d\theta}{\int_b^{\pi-b}\ d\theta} = \Big(0, \frac{2\cos b}{\pi - 2b}\Big),$$

$$p_2 = \frac{\frac{1}{4}\int_{-1}^{-a}(x,0)\,dx + \frac{1}{4\pi}\int_{\pi-b}^{\pi+b}(\cos\theta,\sin\theta)\ d\theta}{\frac{1}{4}\int_{-1}^{-a}\,dx + \frac{1}{4\pi}\int_{\pi-b}^{\pi+b}\ d\theta} = \Big(-\frac{-\pi a^2 + 4\sin b + \pi}{-2\pi a + 4b + 2\pi}, 0\Big),$$

$$p_3 = (0,0), \quad d_1 = (-\cos b, \sin b), \quad d_2 = (-a, 0).$$

Thus, solving the canonical equations $\rho\,(d_1, p_1) - \rho\,(d_1, p_2) = 0$, $\rho\,(d_2, p_2) - \rho\,(d_2, p_3) = 0$, we have $a = 0.394154$, and $b = 0.798783$. Hence, putting the values of a, and b, we have, $p_1 = (0, 0.903584)$, $p_2 = (-0.788308, 0)$, and $p_3 = (0,0)$, and so, due to symmetry, $p_4 = (0, -0.903584)$, and $p_5 = (0.788308, 0)$. The corresponding distortion error is given by

$$V(P;\alpha) = 2\Big(\frac{1}{4\pi}\int_b^{\pi-b}\rho\big((\cos\theta,\sin\theta),p_2\big)\ d\theta + \frac{1}{4}\int_{-1}^{-a}\rho\big((x,0),p_2\big)\,dx + \frac{1}{4\pi}\int_{\pi-b}^{\pi+b}\rho\big((\cos\theta,\sin\theta),p_2\big)\ d\theta\Big) + \frac{1}{4}\int_{-a}^{a}\rho\big((x,0),p_3\big)\,dx = 0.119779.$$

Case 3. α does not contain any point from L_1, the Voronoi region of which does not contain any point from L_2.

In this case, we can assume that $p_1, p_2, \ldots, p_5$ can be located as shown in Figure 3 (iii). Let the boundary of the Voronoi regions of p_1 and p_2 intersect L_2 at the point d_1 given by the parametric value $\theta = \pi - b$, where $0 < b < \frac{\pi}{2}$, and the boundary of the Voronoi regions of p_2 and p_3 intersect L_2 as the point d_2 given by the parametric value $\theta = \pi + c$, where $0 < c < \frac{\pi}{2}$. Thus, due to symmetry, we have

$$p_1 = \frac{\int_{\frac{\pi}{2}}^{\pi-b}(\cos\theta,\sin\theta)\ d\theta}{\int_{\frac{\pi}{2}}^{\pi-b}\ d\theta} = \Big(\frac{2(\sin b - 1)}{\pi - 2b}, \frac{2\cos b}{\pi - 2b}\Big),$$

$$p_2 = \frac{\frac{1}{4}\int_{-1}^{0}(x,0)\,dx + \frac{1}{4\pi}\int_{\pi-b}^{\pi+c}(\cos\theta,\sin\theta)\ d\theta}{\frac{1}{4}\int_{-1}^{0}\,dx + \frac{1}{4\pi}\int_{\pi-b}^{\pi+c}\ d\theta} = \Big(-\frac{2\sin b + 2\sin c + \pi}{2(b + c + \pi)}, \frac{\cos c - \cos b}{b + c + \pi}\Big),$$

$$p_3 = \frac{\int_{\pi+c}^{2\pi-c}(\cos\theta, \sin\theta)\ \mathrm{d}\theta}{\int_{\pi+c}^{2\pi-c}\ \mathrm{d}\theta} = \Big(0, -\frac{2\cos c}{\pi-2c}\Big),$$

$$d_1 = (-\cos b, \sin b),\ \text{and}\ d_2 = (-\cos c, -\sin c).$$

Thus, solving the canonical equations $\rho(d_1, p_1) - \rho(d_1, p_2) = 0$, and $\rho(d_2, p_2) - \rho(d_2, p_3) = 0$, we have $b = 0.426473$, and $c = 0.837847$. Hence, putting the values of b, and c, we have, $p_1 = (-0.512388, 0.795606)$, $p_2 = (-0.619091, -0.0547824)$, $p_3 = (0, -0.912839)$, and so, due to symmetry, $p_4 = (0.619091, -0.0547824)$, and $p_5 = (0.512388, 0.795606)$. The corresponding distortion error is given by

$$\begin{aligned} &V(P;\alpha) \\ &= 2\Big(\frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\pi-b}\rho\big((\cos\theta,\sin\theta),p_2\big)\ \mathrm{d}\theta\Big) + \frac{1}{4}\int_{-1}^{0}\rho\big((x,0),p_2\big)\,\mathrm{d}x \\ &\quad + \frac{1}{4\pi}\int_{\pi-b}^{\pi+c}\rho\big((\cos\theta,\sin\theta),p_2\big)\ \mathrm{d}\theta\Big) + \frac{1}{4\pi}\int_{\pi+c}^{2\pi-c}\rho\big((\cos\theta,\sin\theta),p_3\big)\ \mathrm{d}\theta \\ &= 0.1355. \end{aligned}$$

Comparing Case 1 and Case 2, we see that if α contains two points from L_1, the Voronoi regions of which do not contain any point from L_2, then the distortion error is larger than the distortion error obtained in Case 2. Similarly, we can show that if α contains more than two points from L_1, the Voronoi regions of which do not contain any point from L_2, then the distortion error is larger than the distortion error obtained in Case 2. Comparing Case 2 and Case 3, we see that Case 3 can not happen as the distortion error is larger in Case 3. Considering all the above cases, we see that the distortion error in Case 2 is the smallest. Hence, the points in α obtained in Case 2 form an optimal set of five-means, and the corresponding quantization error is given by $V_5 = 0.119779$. Thus, the proof of the proposition is complete. □

Proposition 2.7. *An optimal set of six-means is*

$$\{(-0.497577, 0.809422), (-0.786245, -0.0706781), (0,0), (0, -0.913921),\\ (0.786245, -0.0706781), (0.497577, 0.809422)\}$$

and the corresponding quantization error for six-means is given by $V_6 = 0.093342$.

Proof. Let $\alpha := \{p_1, p_2, p_3, p_4, p_5, p_6\}$ be an optimal set of six-means. As in Proposition 2.6, here also we consider three different cases as shown in Figure 4.

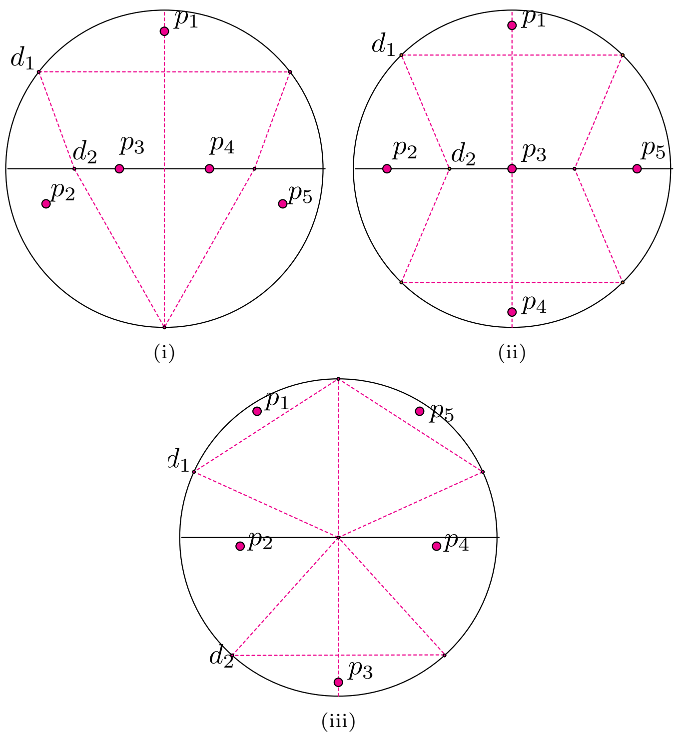

FIGURE 3.

In each case, we calculate the distortion errors. Then, comparing the distortion errors, we see that the points given by the proposition give the smallest distortion error for six points, and hence they form an optimal set of six-means, which is shown by Figure 4 (ii). Thus, the proof of the proposition is deduced. ☐

Proceeding in the similar way as Proposition 2.6 and Proposition 2.7, we can deduce that the following proposition is also true.

PROPOSITION 2.8. *Let α_n be an optimal set of n-means, and let V_n be the corresponding quantization error. Then,*

$$\alpha_7 = \{(-0.476891, 0.827476), (-0.788772, 0), (0, 0), (-0.476891, -0.827476), (0.476891, -0.827476), (0.788772, 0), (0.476891, 0.827476)\}$$

with $V_7 = 0.070674$, see Figure 5 (i)*;*

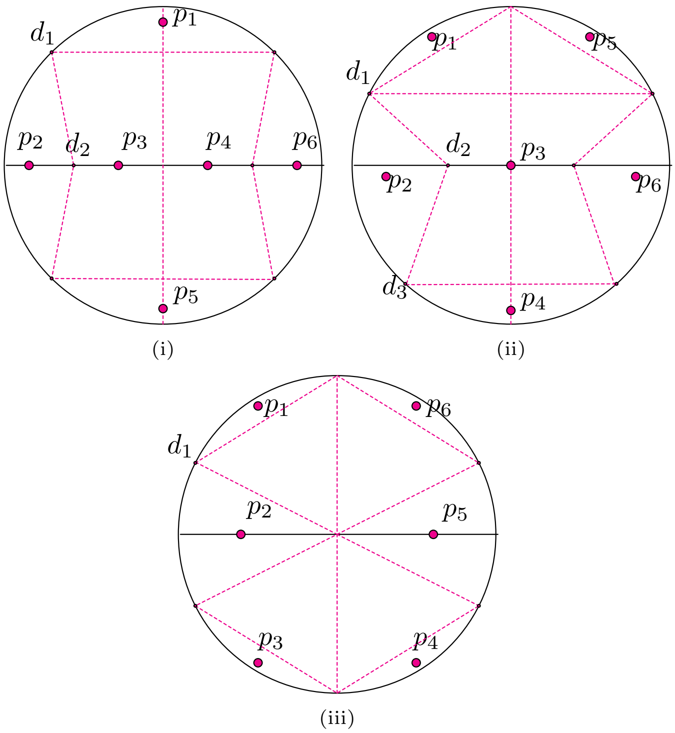

FIGURE 4.

$$
\begin{aligned}
\alpha_8 = \{&(-0.475258, 0.828843), (-0.860649, 0),\\
&(-0.286883, 0), (-0.475258, -0.828843), (0.475258, -0.828843),\\
&(0.860649, 0), (0.286883, 0), (0.475258, 0.828843)\}\\
&\text{with } V_8 = 0.0577852, \text{ see Figure 5 (ii)};
\end{aligned}
$$

$$
\begin{aligned}
\alpha_9 = \{&-0.463928, 0.838108), (-0.857223, 0.0396484), (-0.286659, 0),\\
&(-0.704114, -0.671446), (0, -0.972943), (0.704114, -0.671446),\\
&(0.286659, 0), (0.857223, 0.0396484), (0.463928, 0.838108)\}\\
&\text{with } V_9 = 0.04803, \text{ see Figure 5 (iii)};
\end{aligned}
$$

$$
\begin{aligned}
\alpha_{10} = \{&(0, 0.974386), (-0.690161, 0.687826), (-0.854308, 0), (-0.284769, 0),\\
&(-0.690161, -0.687826), (0, -0.974386), (0.690161, -0.687826),\\
&(0.854308, 0), (0.284769, 0), (0.690161, 0.687826)\},\\
&\text{with } V_{10} = 0.039046, \text{ see Figure 5 (iv)}.
\end{aligned}
$$

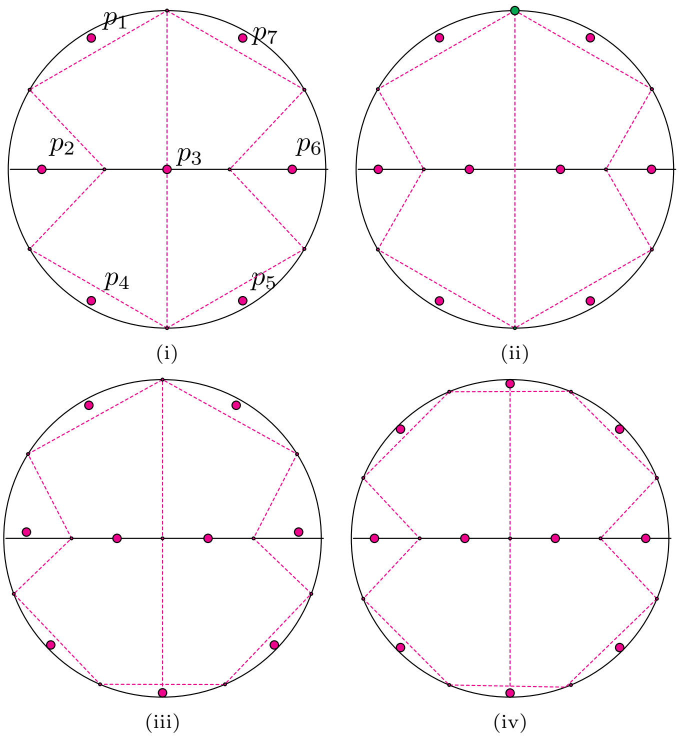

FIGURE 5.

The following proposition plays an important role in the paper.

PROPOSITION 2.9. *Let α_n be an optimal set of n-means for P, and $n \geq 5$. Then, α_n contains at least one point from L_1, the Voronoi region of which does not contain any point from L_2; and at least one point from L_2, the Voronoi region of which does not contain any point from L_1.*

Proof. Let V_n denote the nth quantization error for any positive integer n. By the previous propositions, the lemma is true for $5 \leq n \leq 10$. Let $n \geq 11$. Then, $V_n \leq V_{11} < V_{10} = 0.039046$. For the sake of contradiction, assume that for $n \geq 11$, the set α_n does not contain any point from L_1, the Voronoi region of which does not contain any point from L_2. Then,

$$V_n > \int_{L_1} \min_{a \in \{(-\frac{1}{2},0),(0,\frac{1}{2})\}} \rho\big((x,0),a\big)\, \mathrm{d}P$$
$$= \frac{1}{4}\int_{-1}^{0} \rho\Big((t,0),\Big(-\frac{1}{2},0\Big)\Big)\, \mathrm{d}t + \frac{1}{4}\int_{0}^{1} \rho\Big((t,0),\Big(\frac{1}{2},0\Big)\Big)\, \mathrm{d}t = \frac{1}{24},$$

implying $V_n > \frac{1}{24} = 0.0416667 > V_{10}$, which leads to a contradiction. Hence, α_n contains at least one point from L_1, the Voronoi region of which does not contain any point from L_2. Similarly, we can prove the other part of the proposition. Thus, the proof of the proposition is complete. □

We now state and prove the following theorem, which is the main theorem of this section. Notice that we are saying the theorem as the main theorem of this section, because as mentioned in Remark 2.11, this theorem helps us to calculate all the optimal sets of n-means, and so, the nth quantization errors for all $n \geq 5$ for the mixed distribution P.

Theorem 2.10. *Let $n \geq 5$ be a positive integer, and let α_n be an optimal set of n-means for P. Let $3k+2 \leq n \leq 3k+4$ for some positive integer k. Then, α_n contains k elements from L_1, the Voronoi regions of which do not contain any point from L_2.*

Proof. By Proposition 2.9, for $n \geq 5$, the set α_n always contains points from L_1, the Voronoi regions of which do not contain any point from L_2, and points from L_2, the Voronoi regions of which do not contain any point from L_1. Since the Voronoi region of a point in an optimal set covers maximum area within a shortest distance P-almost surely, the set α_n, given in the theorem, must contain the two points, the Voronoi regions of which contain points from both L_1 and L_2, in other words, the Voronoi regions of these two points contain points around the two intersections of L_1 and L_2. Each of the remaining $n-2$ points occurs due to the uniform distribution on L_1, or L_2, the Voronoi region of which contains points only from L_1, or from L_2, respectively.

Let $n = n_1 + n_2 + k + 2$ be such that α_n contains k elements from L_1, the Voronoi regions of which do not contain any point from L_2; n_1 elements from above the x_1-axis, the Voronoi regions of which do not contain any point from L_1, and n_2 elements from below the x_1-axis, the Voronoi regions of which do not contain any point from L_1. Then, there exist three real numbers a, b, and c, where $-1 < a < 1$, $0 < b < \frac{\pi}{2}$, and $0 < c < \frac{\pi}{2}$, such that the following occur:

(i) The k elements that α_n contains from L_1 occur due to the uniform distribution on $[-a, a]$, and as mentioned in Theorem 1.2, are given by the set

$$\Big\{-a + \frac{2i-1}{k}a : 1 \leq i \leq k\Big\},$$

with distortion error given by

$$k\big(\text{distortion error due to the point } -a+\tfrac{a}{k} \text{ in the interval } \big[-a, -a+\tfrac{2a}{k}\big]\big)$$

$$= \frac{k}{4}\int_{-a}^{-a+\frac{2a}{k}} \Big(t - \Big(-a + \frac{a}{k}\Big)\Big)^2 \,\mathrm{d}t = \frac{a^3}{6k^2}.$$

(ii) The n_1 elements that α_n contains from above the x_1-axis, the Voronoi regions of which do not contain any point from L_1, occur due to the uniform distribution on the circular arc $\{(\cos\theta, \sin\theta) : b \leq \theta \leq \pi - b\}$, and by Theorem 1.3, are given by the set

$$\Big\{\frac{2n_1}{\pi-2b}\sin\frac{\pi-2b}{2n_1}\Big(\cos\Big(b+(2j-1)\frac{\pi-2b}{2n_1}\Big), \sin\Big(b+(2j-1)\frac{\pi-2b}{2n_1}\Big)\Big) : 1 \leq j \leq n_1\Big\},$$

with distortion error

$$n_1\Big(\frac{1}{4\pi}\int_b^{b+\frac{\pi-2b}{n_1}}\rho\Big((\cos\theta,\sin\theta), \frac{2n_1}{\pi-2b}\sin\Big(\frac{\pi-2b}{2n_1}\Big)\Big(\cos\Big(b+\frac{\pi-2b}{2n_1}\Big), \sin\Big(b+\frac{\pi-2b}{2n_1}\Big)\Big)\Big)\,d\theta\Big)$$
$$= \frac{(\pi-2b)^2 - 2n_1^2 + 2n_1^2\cos(\frac{2b-\pi}{n_1})}{4\pi(\pi-2b)},$$

and we denote it by D_{n_1}.

(iii) The n_2 elements that α_n contains from below the x_1-axis, the Voronoi regions of which do not contain any point from L_1, occur due to the uniform distribution on the circular arc $\{(\cos\theta, \sin\theta) : \pi + c \leq \theta \leq 2\pi - c\}$, and by Theorem 1.3, are given by the set

$$\Big\{\frac{2n_2}{\pi-2c}\sin\frac{\pi-2c}{2n_2}\Big(\cos\Big(\pi+c+(2j-1)\frac{\pi-2c}{2n_2}\Big), \sin\Big(\pi+c+(2j-1)\frac{\pi-2c}{2n_2}\Big)\Big) : 1 \leq j \leq n_2\Big\},$$

with distortion error

$$n_2\Big(\frac{1}{4\pi}\int_{\pi+c}^{\pi+c+\frac{\pi-2c}{n_2}}\rho\Big((\cos\theta,\sin\theta), \frac{2n_2}{\pi-2c}\sin\Big(\frac{\pi-2c}{2n_2}\Big)\Big(\cos\Big(\pi+c+\frac{\pi-2c}{2n_2}\Big), \sin\Big(\pi+c+\frac{\pi-2c}{2n_2}\Big)\Big)\Big)\,d\theta\Big)$$
$$= \frac{(\pi-2c)^2 - 2n_2^2 + 2n_2^2\cos(\frac{2c-\pi}{n_2})}{4\pi(\pi-2c)},$$

and we denote it by D_{n_2}.

(iv) The two points in α_n, the Voronoi regions of which contain points from both L_1 and L_2, are given by the set $\{(-r,s),(r,s)\}$, where

$$\begin{aligned}(-r,s)&=\frac{\frac14\int_{-1}^{-a}(t,0)\ \mathrm{d}t+\frac{1}{4\pi}\int_{\pi-b}^{\pi+c}(\cos\theta,\sin\theta)\,\mathrm{d}\theta}{\frac14\int_{-1}^{-a}\ \mathrm{d}t+\frac{1}{4\pi}\int_{\pi-b}^{\pi+c}\ \mathrm{d}\theta}\\
&=\Big(-\frac{-\pi a^2+2\sin b+2\sin c+\pi}{2(-\pi a+b+c+\pi)},\frac{\cos c-\cos b}{-\pi a+b+c+\pi}\Big),\end{aligned}$$

i.e.,

$$r=\frac{-\pi a^2+2\sin b+2\sin c+\pi}{2(-\pi a+b+c+\pi)},\quad\text{and}\quad s=\frac{\cos c-\cos b}{-\pi a+b+c+\pi},$$

and the distortion error for both the two points is given by

$$\begin{aligned}&2\Big(\frac14\int_{-1}^{-a}\rho\big((t,0),(-r,s)\big)\ \mathrm{d}t+\frac{1}{4\pi}\int_{\pi-b}^{\pi+c}\rho\big((\cos\theta,\sin\theta),(-r,s)\big)\ \mathrm{d}\theta\Big)\\
&=\frac{1}{24\pi(-\pi a+b+c+\pi)}\Big(\pi^2a^4-4\pi a^3b-4\pi a^3c-4\pi^2a^3\\
&\quad+12\pi(a^2-1)\sin b+12\pi a^2\sin c+6\pi^2a^2\\
&\quad-12\pi ab-12\pi ac-4\pi^2a+12b^2+24bc\\
&\quad+24\cos(b+c)+16\pi b+12c^2+16\pi c-12\pi\sin c+\pi^2-24\Big),\end{aligned}$$

and we denote it by $D(a,b,c)$.

Let $V(n_1,n_2,k)$ denote the distortion error due to the all above n_1+n_2+k+2 elements in α_n. Then, we have

$$V(n_1,n_2,k)=\frac{a^3}{6k^2}+D_{n_1}+D_{n_2}+D(a,b,c).\tag{1}$$

Let n_1,n_2, and k be fixed. Then, using the partial derivatives we can obtain the following equations

$$\frac{\partial}{\partial a}\big(V(n_1,n_2,k)\big)=0,\quad\frac{\partial}{\partial b}\big(V(n_1,n_2,k)\big)=0,\quad\text{and}\quad\frac{\partial}{\partial c}\big(V(n_1,n_2,k)\big)=0.\tag{2}$$

For a given set of values of n_1,n_2, and k, solving the equations in (2), we can obtain the values of a,b,c. Putting the values of a,b,c in (1), we can obtain the distortion error for the given set of values of n_1,n_2,k.

Now, to prove the theorem we use induction on k. If $k=1$, and $k=2$, the theorem is true due to the previous propositions. Let us assume that the theorem is true for $k=m$, i.e., when $3m+2\le n\le 3m+4$. We now prove that the theorem is true for $3(m+1)+2\le n\le 3(m+1)+4$. By the assumption, the theorem is true for $n=3m+4$, i.e., the set α_{3m+4} contains

m points from L_1, the Voronoi regions of which do not contain any point from L_2, and $(2m+2)$ points occur due to the uniform distribution on L_2, the Voronoi region of which do not contain any point from L_1. Again, due to the mixed distribution with equal weights to the component probabilities, and symmetry of the circle with respect to the x_1-axis, we can can assume that α_n contains $m+1$ elements from above, and $m+1$ elements from below. Now, to calculate α_{n+1}, we need to add one extra point either to L_1, or L_2 in an optimal way, i.e., the Voronoi regions of the new point will contain only the points from L_1, or from L_2, and the overall distortion error due to $n+1$ points becomes smallest. First suppose that the extra point is added to L_1, the Voronoi region of which does not contain any point from L_2. As described above using (1), we calculate the distortion error $V(m+1, m+1, m+1)$. Next, suppose that the extra point is added to L_2, the Voronoi region of which does not contain any point from L_1, and using (1), we calculate the distortion error $V(m+2, m+1, m)$, or $V(m+1, m+2, m)$. We see that the distortion error $V(m+1, m+1, m+1)$ is the smallest, which implies the fact that α_{n+1} contains $m+1$ points from L_1. Once, α_{n+1} is known, similarly we can obtain α_{n+2}, and α_{n+3} with distortion errors, respectively, $V(m+1, m, m+1)$ and $V(m+1, m+1, m+1)$. Thus, we see that each of α_{n+1}, α_{n+2}, and α_{n+3} contains $m+1$ points from L_1, the Voronoi regions of which do not contain any point from L_2. Notice that $n+1 = 3(m+1)+2$, $n+2 = 3(m+1)+3$, and $n+1 = 3(m+1)+4$, i.e., for the positive integer n satisfying $3(m+1)+2 \leq n \leq 3(m+1)+4$, the set α_n contains $m+1$ elements from L_1, the Voronoi regions of which do not contain any point from L_2. Thus, the theorem is true for $k = m+1$ if it is true for $k = m$. Hence, by the principle of mathematical induction, the theorem is true for all positive integers k, and thus, the proof of the theorem is complete. ☐

Remark 2.11. For $n \geq 5$, let $3k+2 \leq n \leq 3k+4$ for some positive integer k. Then, by Theorem 2.10, we can say that if $n-k-2$ is an even number, then an optimal set of n-means contains $\frac{1}{2}(n-k-2)$ elements from either side of the x_1-axis, the Voronoi regions of which do not contain any point from L_1; and if $n-k-2$ is an odd number, then an optimal set of n-means contains $\frac{1}{2}\lfloor n-k-2 \rfloor$ elements from one side of the x_1-axis, and $\frac{1}{2}\lfloor n-k-2 \rfloor + 1$ elements from the other side of the x_1-axis, the Voronoi regions of which do not contain any point from L_1. Thus, by Theorem 2.10, using Theorem 1.2, and Theorem 1.3, we can easily determine the optimal sets of n-means and the nth quantization errors for all $n \geq 5$.

The following proposition gives the quantization dimension and the quantization coefficient for the mixed distribution.

Proposition 2.12. *Quantization dimension $D(P)$ of the mixed distribution P is one, which is the dimension of the underlying space, and the quantization coefficient exists as a finite positive number which equals $\frac{3}{8}\left(4+\pi^2\right)$.*

Proof. By Remark 2.11, we see that if n is of the form $n = 3k + 2$ for some positive integer k, then α_n contains k elements from L_1, the Voronoi regions of which do not contain any point from L_2, and k elements from the above, and k elements from below the x_1-axis, the Voronoi region of which do not contain any point from L_1. For $n \in \mathbb{N}$, $n \geq 5$, let $\ell(n)$ be the unique positive integer such that $3\ell(n) + 2 \leq n < 3(\ell(n) + 1) + 2$. Then, $V_{3(\ell(n)+1)+2} < V_n \leq V_{3\ell(n)+2}$ implying

$$\frac{2\log(3\ell(n)+2)}{-\log V_{3(\ell(n)+1)+2}} < \frac{2\log n}{-\log V_n} < \frac{2\log(3(\ell(n)+1)+2)}{-\log V_{3\ell(n)+2}}. \tag{3}$$

Notice that if $n \to \infty$, then $\ell(n) \to \infty$. Moreover, if $n \to \infty$, they by (1) and (2), we can see that $a \to 1$, $b \to 0$, and $c \to 0$. Assume that n is sufficiently large, in other words, assume that $\ell(n)$ is sufficiently large, and then as $a \to 1$, $b \to 0$, and $c \to 0$, by (1) we have $D(a, b, c) \to 0$, implying

$$V_{3\ell(n)+2} = V\big(\ell(n), \ell(n), \ell(n)\big) = \frac{-6\ell(n)^4 + 6\ell(n)^4 \cos\frac{\pi}{\ell(n)} + 3\pi^2\ell(n)^2 + \pi^2}{6\pi^2\ell(n)^2},$$

yielding

$$\lim_{n\to\infty} \frac{2\log(3\ell(n)+2)}{-\log V_{3(\ell(n)+1)+3}}$$

$$= \lim_{\ell(n)\to\infty} \frac{2\log(3\ell(n)+2)}{-\log\left(\frac{-6(\ell(n)+1)^4+3\pi^2(\ell(n)+1)^2+6(\ell(n)+1)^4\cos(\frac{\pi}{\ell(n)+1})+\pi^2}{6\pi^2(\ell(n)+1)^2}\right)} = 1,$$

and

$$\lim_{n\to\infty} \frac{2\log(3(\ell(n)+1)+2)}{-\log V_{3\ell(n)+2}}$$

$$= \lim_{\ell(n)\to\infty} \frac{2\log(3(\ell(n)+1)+2)}{-\log\left(\frac{-6\ell(n)^4+6\ell(n)^4\cos(\frac{\pi}{\ell(n)})+3\pi^2\ell(n)^2+\pi^2}{6\pi^2\ell(n)^2}\right)} = 1$$

and hence, by (3),

$$\lim_{n\to\infty} \frac{2\log n}{-\log V_n} = 1,$$

which is the dimension of the underlying space.

Again,

$$\big(3\ell(n)+2\big)^2 V_{3(\ell(n)+1)+2} < n^2 V_n < \Big(3\big(\ell(n)+1\big)+2\Big)^2 V_{3\ell(n)+2}. \tag{4}$$

We have

$$\begin{aligned}
&\lim_{n\to\infty}\big(3\ell(n)+2\big)^2 V_{3(\ell(n)+1)+2}\\
&=\lim_{\ell(n)\to\infty}\big(3\ell(n)+2\big)^2\\
&\quad\frac{-6(\ell(n)+1)^4+3\pi^2(\ell(n)+1)^2+6(\ell(n)+1)^4\cos(\frac{\pi}{\ell(n)+1})+\pi^2}{6\pi^2(\ell(n)+1)^2}\\
&=\frac{3}{8}\left(4+\pi^2\right),
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{n\to\infty}\Big(3\big(\ell(n)+1\big)+2\Big)^2 V_{3\ell(n)+2}\\
&=\lim_{\ell(n)\to\infty}\Big(3\big(\ell(n)+1\big)+2\Big)^2\frac{-6\ell(n)^4+6\ell(n)^4\cos(\frac{\pi}{\ell(n)})+3\pi^2\ell(n)^2+\pi^2}{6\pi^2\ell(n)^2}\\
&=\frac{3}{8}\left(4+\pi^2\right),
\end{aligned}$$

and hence, by (4) we have

$$\lim_{n\to\infty} n^2V_n=\frac{3}{8}\left(4+\pi^2\right),$$

i.e., the quantization coefficient exists as a finite positive number which equals $=\frac{3}{8}\left(4+\pi^2\right)$. Thus, the proof of the proposition is complete. □

3. Optimal quantization for the mixture of two uniform distributions on two disconnected line segments

Let P_1 and P_2 be two uniform distributions, respectively, on the intervals $[0,\frac{1}{2}]$ and $[\frac{3}{4},1]$. Write

$$J_1:=\Big[0,\frac{1}{2}\Big], \quad\text{and}\quad J_2:=\Big[\frac{3}{4},1\Big].$$

Let f_1 and f_2 be their respective density functions. Then, $f_1(x)=2$ if $x\in[0,\frac{1}{2}]$, and zero, otherwise; and $f_2(x)=4$ if $x\in[\frac{3}{4},1]$, and zero, otherwise. Let $P:=\frac{3}{4}P_1+\frac{1}{4}P_2$. In the sequel, for the mixed distribution P, we determine the optimal sets of n-means and the nth quantization errors for all positive integers n. By $E(P)$ and $V(P)$, we mean the expectation and the variance of a random variable with distribution P. By $\alpha_n(\mu)$, we denote an optimal set of n-means with respect to a probability distribution μ, and $V_n(\mu)$ represents the corresponding quantization error for n-means. If μ is the mixed distribution P, in the sequel, we sometimes denote it by α_n instead of $\alpha_n(P)$, and the corresponding quantization error is denoted by V_n instead of $V_n(P)$.

Lemma 3.1. *Let P be the mixed distribution defined by $P = \frac{3}{4}P_1 + \frac{1}{4}P_2$. Then,*

$$E(P) = \frac{13}{32}, \quad \textit{and} \quad V(P) = \frac{277}{3072}.$$

Proof. We have

$$E(P) = \int x\,\mathrm{d}P = \frac{3}{4}\int x\,\mathrm{d}\big(P_1(x)\big)+\frac{1}{4}\int x\ \mathrm{d}\big(P_2(x)\big) = \frac{3}{4}\int_0^{\frac{1}{2}} 2x\,\mathrm{d}x+\frac{1}{4}\int_{\frac{3}{4}}^1 4x\,\mathrm{d}x$$

yielding $E(P) = \frac{13}{32}$, and

$$V(P)=\int\big(x-E(P)\big)^2\,\mathrm{d}P=\frac{3}{4}\int\big(x-E(P)\big)^2\ \mathrm{d}\big(P_1(x)\big)+\frac{1}{4}\int\big(x-E(P)\big)^2\ \mathrm{d}\big(P_2(x)\big),$$

implying $V(P) = \frac{277}{3072}$, and thus, the lemma is yielded. □

Remark 3.2. The optimal set of one-mean is the set $\{\frac{13}{32}\}$, and the corresponding quantization error is the variance $V := V(P)$ of a random variable with distribution P.

Lemma 3.3. *The set $\alpha := \{\frac{1}{4}, \frac{7}{8}\}$ is an optimal set of two-means, and the corresponding quantization error is given by $V_2 = \frac{13}{768}$.*

Proof. Consider the set of two points β given by $\beta := \{\frac{1}{4}, \frac{7}{8}\}$. The distortion error due to the set β is given by

$$\begin{aligned}\int\min_{a\in\beta}(x-a)^2\,\mathrm{d}P &= \int_{J_1}\Big(x-\frac{1}{4}\Big)^2\,\mathrm{d}P + \int_{J_2}\Big(x-\frac{7}{8}\Big)^2\,\mathrm{d}P\\ &= \frac{3}{4}\int_0^{\frac{1}{2}}2\Big(x-\frac{1}{4}\Big)^2\,\mathrm{d}x + \frac{1}{4}\int_{\frac{3}{4}}^1 4\Big(x-\frac{7}{8}\Big)^2\,\mathrm{d}x\\ &= \frac{13}{768} = 0.0169271.\end{aligned}$$

Since V_2 is the quantization error for two-means, we have $V_2 \le 0.0169271$. Let $\alpha := \{a_1, a_2\}$ 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 $0 < a_1 < a_2 < 1$. We now show that the

Voronoi region of a_1 does not contain any point from J_2, and the Voronoi region of a_2 does not contain any point from J_1. Suppose that $\frac{13}{40} \leq a_1$. Then,

$$V_2 > \int_{[0, \frac{13}{40}]} \Big(x - \frac{13}{40}\Big)^2 \, dP = \frac{2197}{128000} = 0.0171641 > V_2,$$

which is a contradiction, and so, we can assume that $a_1 < \frac{13}{40} < \frac{1}{2}$.
Since $a_1 < \frac{13}{40}$, the Voronoi region of a_1 does not contain any points from J_2. If it contains points from J_2, then $\frac{1}{2}(a_1 + a_2) > \frac{3}{4}$, implying $a_2 > \frac{3}{2} - a_1 \geq \frac{3}{2} - \frac{13}{40} = \frac{47}{40} > 1$, which is a contradiction. Hence, we can assume that

$$a_1 \leq E(X : X \in J_1) = \frac{1}{4}, \quad \text{and} \quad a_2 \leq E(X : X \in J_2) = \frac{7}{8}. \tag{5}$$

Suppose that $a_2 < \frac{5}{8}$. Then,

$$V_2 > \frac{1}{4} \int_{\frac{3}{4}}^{1} 4\Big(x - \frac{5}{8}\Big)^2 \, dx = \frac{13}{768} = 0.0169271 \geq V_2,$$

which leads to a contradiction. So, we can assume that $\frac{5}{8} \leq a_2$. Thus, by (5), we have $\frac{5}{8} \leq a_2 \leq \frac{7}{8}$. Assume that $\frac{5}{8} \leq a_2 \leq \frac{3}{4}$. Since $a_1 \leq \frac{1}{4}$, the following cases can arise:

Case 1. $\frac{1}{8} \leq a_1 \leq \frac{1}{4}$.

Then, notice that $\frac{13}{32} < \frac{1}{2}(\frac{1}{4} + \frac{5}{8}) = \frac{7}{16} < \frac{1}{2}$, and so,

$$\int_{[0, \frac{13}{32}]} \min_{a \in \{a_1, a_2\}} (x - a)^2 \, dP = \frac{13\left(3072a_1^2 - 1248a_1 + 169\right)}{65536},$$

the minimum value of which is $\frac{2197}{262144}$, and it occurs when $a_1 = \frac{13}{64}$.
Notice that for $a_1 = \frac{13}{64}$, we have

$$\frac{13}{32} = 0.40625 < \frac{1}{2}\Big(\frac{13}{64} + \frac{5}{8}\Big) = 0.414063.$$

Thus, we have

$$V_2 \geq \frac{2197}{262144} + \frac{3}{4} \int_{\frac{13}{32}}^{\frac{7}{16}} 2\Big(x - \frac{1}{4}\Big)^2 \, dx + \frac{3}{4} \int_{\frac{7}{16}}^{\frac{1}{2}} 2\Big(x - \frac{5}{8}\Big)^2 \, dx + \frac{1}{4} \int_{\frac{3}{4}}^{1} 4\Big(x - \frac{3}{4}\Big)^2 \, dx = \frac{13603}{786432},$$

yielding $V_2 \geq 0.0172971 > V_2$, which is a contradiction.

Case 2. $a_1 < \frac{1}{8}$.

Then, $\frac{1}{2}(\frac{1}{8} + \frac{5}{8}) = \frac{3}{8} < \frac{1}{2}$, and so

$$V_3 \geq \frac{3}{4}\int_{\frac{1}{8}}^{\frac{3}{8}} 2\Big(x-\frac{1}{8}\Big)^2\,\mathrm{d}x + \frac{3}{4}\int_{\frac{3}{8}}^{\frac{1}{2}} 2\Big(x-\frac{5}{8}\Big)^2\,\mathrm{d}x + \frac{1}{4}\int_{\frac{3}{4}}^{1} 4\Big(x-\frac{3}{4}\Big)^2\,\mathrm{d}x = \frac{61}{3072} = 0.0198568 > V_3,$$

which leads to a contradiction.

Hence, by Case 1 and Case 2, we can conclude that $\frac{3}{4} \leq a_2 \leq \frac{7}{8}$. Suppose that $\frac{3}{4} \leq a_2 \leq \frac{13}{16}$. Then, the Voronoi region of a_2 must contain points prom J_1 implying $\frac{1}{2}(a_1 + a_2) < \frac{1}{2}$, which yields $a_1 < 1 - a_2 \leq 1 - \frac{3}{4} = \frac{1}{4}$. Again,

$$\int_{J_1} (x-a_1)^2\,\mathrm{d}P = \frac{1}{16}(12a^2 - 6a + 1),$$

the minimum value of which is $\frac{1}{64}$ when $a_1 = \frac{1}{4}$. Thus, we have

$$V_2 \geq \int_{J_1}\Big(x-\frac{1}{4}\Big)^2\,\mathrm{d}P + \int_{J_2}\Big(x-\frac{13}{16}\Big)^2\,\mathrm{d}P = \frac{55}{3072} = 0.0179036 > V_2,$$

which gives a contradiction. Hence, we can assume that $\frac{13}{16} < a_2 \leq \frac{7}{8}$. Suppose that the Voronoi region of a_2 contains points from J_1, i.e., $\frac{1}{2}(a_1+a_2) < \frac{1}{2}$. Then, $a_1 < 1 - a_2 \leq 1 - \frac{13}{16} = \frac{3}{16}$. Notice that

$$\int_{J_1} (x-a_1)^2\,\mathrm{d}P = \frac{1}{16}(12a_1^2 - 6a_1 + 1),$$

the minimum value of which is $\frac{19}{1024}$ when $a_1 = \frac{3}{16}$. Thus, we have $V_2 \geq \frac{19}{1024} = 0.0185547 > V_2$, which is a contradiction. Thus, we can assume that the Voronoi region of a_2 does not contain any point from J_1. Previously, we have proved that the Voronoi region of a_1 does not contain any point from J_2. Hence, we have

$$a_1 = E(X : X \in J_1) = \frac{1}{4}, \quad \text{and} \quad a_2 = E(X : X \in J_2) = \frac{7}{8},$$

and the corresponding quantization error for two-means is given by $V_2 = \frac{13}{768}$. □

Lemma 3.4. *The set* $\left\{\frac{1}{8}, \frac{3}{8}, \frac{7}{8}\right\}$ *forms an optimal set of three-means with quantization error* $V_3 = \frac{1}{192}$.

Proof. Consider the set of three points β, such that $\beta := \left\{\frac{1}{8}, \frac{3}{8}, \frac{7}{8}\right\}$. The distortion error due to the set β is given by

$$\int \min_{a\in\beta}(x-a)^2\,\mathrm{d}P = 2\cdot\frac{3}{4}\int_0^{\frac{1}{4}} 2\Big(x-\frac{1}{8}\Big)^2\,\mathrm{d}x + \frac{1}{4}\int_{\frac{3}{4}}^{1} 4\Big(x-\frac{7}{8}\Big)^2\,\mathrm{d}x = \frac{1}{192}.$$

Since V_3 is the quantization error for three-means, $V_3 \leq \frac{1}{192} = 0.00520833$. Let $\alpha := \{a_1, a_2, a_3\}$ be an optimal set of three-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 $0 < a_1 < a_2 < a_3 < 1$. We now show that $a_2 < \frac{1}{2}$, and $\frac{3}{4} < a_3$. If $a_3 < \frac{3}{4}$, then

$$V_3 > \int_{J_2} \Big(x - \frac{3}{4}\Big)^2 \, \mathrm{d}P = \frac{1}{4} \int_{\frac{3}{4}}^{1} 4\Big(x - \frac{3}{4}\Big)^2 \, \mathrm{d}x = \frac{1}{192} = 0.00520833 \geq V_3,$$

which leads to a contradiction. Hence, we can assume that $\frac{3}{4} < a_3$. Next, we show that $a_2 < \frac{1}{2}$. Suppose that $\frac{1}{2} \leq a_2$. Then,

$$\begin{aligned} &\int_{J_1} \min_{a \in \{a_1, \frac{1}{2}\}} (x-a)^2 \, \mathrm{d}P \\ &= \frac{3}{4} \int_0^{\frac{1}{2}(a_1 + \frac{1}{2})} 2(x - a_1)^2 \, \mathrm{d}x + \frac{3}{4} \int_{\frac{1}{2}(a_1+\frac{1}{2})}^{\frac{1}{2}} 2\Big(x - \frac{1}{2}\Big)^2 \, \mathrm{d}x \\ &= \frac{1}{64}(24a_1^3 + 12a_1^2 - 6a_1 + 1), \end{aligned}$$

the minimum value of which is $\frac{1}{144}$, and it occurs when $a_1 = \frac{1}{6}$. Thus, in this case, we see that $V_3 \geq \frac{1}{144} = 0.00694444 > V_3$, which leads to a contradiction. Hence, we can assume that $0 < a_1 < a_2 < \frac{1}{2}$. Suppose that the Voronoi region of a_2 contains points from J_2. Then, $\frac{1}{2}(a_2 + a_3) > \frac{3}{4}$ implying $a_3 > \frac{3}{2} - a_1 \geq \frac{3}{2} - \frac{1}{2} = 1$, which is a contradiction, as $a_3 < 1$. Thus, we see that the Voronoi region of a_2 does not contain any point from J_2. Suppose that the Voronoi region of a_3 contains points from J_1. Then, $\frac{1}{2}(a_2 + a_3) < \frac{1}{2}$ implying $a_2 < 1 - a_3 \leq 1 - \frac{3}{4} = \frac{1}{4}$, and so

$$V_3 > \frac{3}{4} \int_{\frac{1}{4}}^{\frac{1}{2}} 2\Big(x - \frac{1}{4}\Big)^2 \, \mathrm{d}x = \frac{1}{128} = 0.0078125 > V_3,$$

which is a contradiction. So, we can assume that the Voronoi region of a_3 does not contain any point from J_1. Thus, by Theorem 1.2, we can conclude that $a_1 = \frac{1}{8}$, $a_2 = \frac{3}{8}$, and $a_3 = \frac{7}{8}$, and

$$V_3 = \int \min_{a \in \alpha} (x-a)^2 \, \mathrm{d}P = \frac{1}{192},$$

which completes the proof of the lemma. □

Remark 3.5. By Lemma 3.3, and Lemma 3.4, we see that $\alpha_2 = \alpha_1(P_1) \cup \alpha_1(P_2)$, and $\alpha_3 = \alpha_2(P_1) \cup \alpha_1(P_2)$. Using the similar technique, we can show that $\alpha_4 = \alpha_3(P_1) \cup \alpha_1(P_2)$, $\alpha_5 = \alpha_3(P_1) \cup \alpha_2(P_2)$, $\alpha_6 = \alpha_4(P_1) \cup \alpha_2(P_2)$, $\alpha_7 = \alpha_5(P_1) \cup \alpha_2(P_2)$, $\alpha_8 = \alpha_6(P_1) \cup \alpha_2(P_2)$, and $\alpha_9 = \alpha_6(P_1) \cup \alpha_3(P_2)$.

We now prove the following propositions.

Proposition 3.6. *Let α_n be an optimal set of n-means for P for $n \geq 2$. Then, the set α_n does not contain any point from the open interval $\left(\frac{1}{2}, \frac{3}{4}\right)$.*

Proof. By Remark 3.5, the proposition is true for $2 \leq n \leq 9$. We now prove that the proposition is true for any positive integer $n \geq 10$. Take any $n \geq 10$. Since $\alpha_9 = \alpha_6(P_1) \cup \alpha_3(P_2)$, and the Voronoi region of any point in $\alpha_9 \cap J_1$ does not contain any point from J_2, and the Voronoi region of any point in $\alpha_9 \cap J_2$ does not contain any point from J_1, we have

$$V_9 = \frac{3}{4}V_6(P_1) + \frac{1}{4}V_3(P_2) = \frac{1}{1728} = 0.000578704.$$

Since V_n is the quantization error for n-means for $n \geq 10$, we have $V_n \leq V_9 = 0.000578704$. Let $\alpha_n := \{a_1, a_2, \ldots, a_n\}$ be an optimal set of n-means for P such that $a_1 < a_2 < \cdots < a_n$. Let $j = \max\{i : a_i \leq \frac{1}{2}\}$. Then, $a_j \leq \frac{1}{2} < a_{j+1}$. The proposition will be proved if we can show that $a_{j+1} \in J_2$. For the sake of contradiction, assume that $a_{j+1} \in \left(\frac{1}{2}, \frac{3}{4}\right)$. Then, the following two cases can arise:

Case 1. $\frac{1}{2} < a_{j+1} \leq \frac{5}{8}$.

In this case, the Voronoi region of a_{j+1} must contain points from J_2, otherwise, the quantization error can be strictly reduced my moving the point a_{j+1} to $\frac{1}{2}$. Thus, $\frac{1}{2}(a_{j+1} + a_{j+2}) > \frac{3}{4}$ implying $a_{j+2} > \frac{3}{2} - a_{j+1} \geq \frac{3}{2} - \frac{5}{8} = \frac{7}{8}$, which yields the fact that

$$V_n \geq \int_{[\frac{3}{4}, \frac{7}{8}]} \left(x - \frac{7}{8}\right)^2 \mathrm{d}P = \frac{1}{4} \int_{\frac{3}{4}}^{\frac{7}{8}} 4\left(x - \frac{7}{8}\right)^2 \mathrm{d}x = 0.000651042 > V_n,$$

which leads to a contradiction.

Case 2. $\frac{5}{8} \leq a_{j+1} < \frac{3}{4}$.

In this case, we have $\frac{1}{2}(a_j + a_{j+1}) < \frac{1}{2}$ implying $a_j < 1 - a_{j+1} \leq 1 - \frac{5}{8} = \frac{3}{8}$, which yields the fact that

$$V_n \geq \int_{[\frac{3}{8}, \frac{1}{2}]} \left(x - \frac{3}{8}\right)^2 \mathrm{d}P = \frac{3}{4} \int_{\frac{3}{8}}^{\frac{1}{2}} 2\left(x - \frac{3}{8}\right)^2 \mathrm{d}x = 0.000976563 > V_n,$$

which is a contradiction.

In light of the above two cases, we can conclude that $a_{j+1} \notin \left(\frac{1}{2}, \frac{3}{4}\right)$. Hence, $\frac{3}{4} < a_{j+2}$, i.e., $a_{j+2} \in J_2$. Thus, the proof of the proposition is complete. □

Proposition 3.7. *Let α_n be an optimal set of n-means for P for $n \geq 2$. Then, for $n \geq 2$, $\alpha_n \cap J_1 \neq \emptyset$, and $\alpha_n \cap J_2 \neq \emptyset$. Moreover, for $n \geq 2$, any point in $\alpha_n \cap J_1$ does not contain any point from J_2, and any point in $\alpha_n \cap J_2$ does not contain any point from J_1,*

Proof. As shown in the proof of Lemma 3.3, and Lemma 3.4, we see that the proposition is true for $n = 2, 3$. By Lemma 3.4, we know $V_3 = \frac{1}{192} = 0.00520833$. We now prove the proposition for $n \geq 4$. Let $n \geq 4$. Since V_n is the quantization error for n-means for $n \geq 4$, we have $V_n \leq V_3 = 0.00520833$. Let $\alpha_n := \{a_1, a_2, \ldots, a_n\}$ be an optimal set of n-means for P such that $a_1 < a_2 < \cdots < a_n$. If $\alpha_n \cap J_2 = \emptyset$, then

$$V_n > \frac{1}{4}\int_{\frac{3}{4}}^{1} 4\Big(x - \frac{3}{4}\Big)^2 \, dx = 0.00520833,$$

which is a contradiction as $V_n \leq 0.00520833$. On the other hand, if $\alpha_n \cap J_1 = \emptyset$, then

$$V_n > \frac{3}{4}\int_0^{\frac{1}{2}} 2\Big(x - \frac{1}{4}\Big)^2 \, dx = \frac{1}{64} = 0.015625 > V_n,$$

which leads to a contradiction. Hence,

$$\alpha_n \cap J_1 \neq \emptyset, \quad \text{and} \quad \alpha_n \cap J_2 \neq \emptyset.$$

Let $j = \max\{i : a_i \leq \frac{1}{2}\}$. Then, $a_j \leq \frac{1}{2}$, and due to Proposition 3.6, we have $\frac{3}{4} \leq a_{j+1}$. If the Voronoi region of a_j contains points from J_2, then $\frac{1}{2}(a_j + a_{j+1}) > \frac{3}{4}$ implying $a_{j+1} > \frac{3}{2} - a_j \geq \frac{3}{2} - \frac{1}{2} = 1$, which is a contradiction. If the Voronoi region of a_{j+1} contains points from J_1, then $\frac{1}{2}(a_j + a_{j+1}) < \frac{1}{2}$ implying $a_j < 1 - a_{j+1} \leq 1 - \frac{3}{4} = \frac{1}{4}$. Then,

$$V_n \geq \int_{[\frac{1}{4}, \frac{1}{2}]} \Big(x - \frac{1}{4}\Big)^2 dP = \frac{3}{4}\int_{\frac{1}{4}}^{\frac{1}{2}} 2\Big(x - \frac{1}{4}\Big)^2 \, dx = \frac{1}{128}$$

yielding $V_n \geq 0.0078125 > V_n$, which leads to a contradiction. Thus, the proof of the proposition is complete. □

Definition 3.8. For $n \in \mathbb{N}$, and $n \geq 2$, define the function $a(n)$ as follows:

$$a(n) = \min\{k \in \mathbb{N} : H(n, k) > 0\},$$

where $H(n, k) = \frac{1}{n^3} - \sum_{i=k}^{\infty} \frac{1}{(i+1)^4}$.

Remark 3.9. Notice that $\sum_{i=k}^{\infty} \frac{1}{(i+1)^4}$ is a decreasing function of $k \in \mathbb{N}$, and so for a given $n \geq 2$, $H(n, k)$ is an increasing function of k, and thus the function $a(n)$ is well defined. Moreover, $\{\frac{1}{n^3}\}_{n \geq 2}$ is a decreasing sequence, and so, the sequence $\{a(n)\}_{n=2}^{\infty}$ is an increasing sequence. In fact,

$$\{a(n)\}_{n=2}^{\infty} = \{1, 2, 3, 3, 4, 5, 6, 6, 7, 8, 8, 9, 10, 10, \\ 11, 12, 12, 13, 14, 15, 15, 16, 17, 17, 18, \dots\}.$$

By $\lfloor x \rfloor$ it is meant the greatest integer not exceeding x. To find the value of $a(n)$ for any positive integer n, one can start checking by putting $k = \lfloor \frac{2n}{3} \rfloor$ in the function $H(n, k)$. If $H(n, k) > 0$, then find $H(n, k-1), H(n, k-2), \dots$ until one obtains some positive integer m, such that $H(n, m) > 0$, and $H(n, m-1) < 0$, and then $a(n) = m$. If $H(n, k) < 0$, then find $H(n, k+1), H(n, k+2), \dots$ until one obtains some positive integer m, such that $H(n, m) > 0$, and $H(n, m-1) < 0$, and then $a(n) = m$.

Remark 3.10. For $n \geq 2$ let α_n be an optimal set of n-means for P. Due to Proposition 3.6 and Proposition 3.7, we can conclude that if α_n contains k elements from J_1, then α_n contains $n-k$ elements from J_2. Thus, we have

$$\begin{aligned} V_n := V_n(P) &= \int \min_{a \in \alpha_n} (x-a)^2 \, \mathrm{d}P \\ &= \frac{3}{4} \int \min_{a \in \alpha_n \cap J_1} (x-a)^2 \, \mathrm{d}P_1 + \frac{1}{4} \int \min_{a \in \alpha_n \cap J_2} (x-a)^2 \, \mathrm{d}P_2, \end{aligned}$$

yielding

$$V_n(P) = \frac{3}{4} V_k(P_1) + \frac{1}{4} V_{n-k}(P_2).$$

Let us now give the following theorem, which gives the optimal sets of n-means and the nth quantization errors for the mixed distribution P for all positive integers $n \geq 2$.

Theorem 3.11. *For $n \geq 2$, let α_n be an optimal set of n-means for P. Then, α_n contains $a(n)$ elements from J_1, i.e.,*

$$\alpha_n(P) = \alpha_{a(n)}(P_1) \cup \alpha_{n-a(n)}(P_2), \ \textit{and} \ V_n(P) = \frac{3}{4} V_{a(n)}(P_1) + \frac{1}{4} V_{n-a(n)}(P_2).$$

Proof. Assume that α_n contains k elements from J_1. Let $V(k, n-k)$ is the corresponding distortion error. Then, as mentioned in Remark 3.10, we have

$$V(k, n-k) = \frac{3}{4} V_k(P_1) + \frac{1}{4} V_{n-k}(P_2).$$

Notice that if our assumption is correct, then we must have $V_n = V(k, n-k)$.

Let us now run the following algorithm:

(i) Write $k := \lfloor \frac{2n}{3} \rfloor$.

(ii) If $V\big(k-1, n-(k-1)\big) < V(k, n-k)$ replace k by $k-1$ and return, else go to step (iii).

(iii) If $V\big(k+1, n-(k+1)\big) < V(k, n-k)$ replace k by $k+1$ and return, else step (iv).

(iv) End.

After running the above algorithm, we see that $k = a(n)$, i.e., our assumption is correct. Thus, the proof of the theorem is complete. □

Remark 3.12. If $n = 14$, then $k = \lfloor \frac{28}{3} \rfloor = 9$. By running the algorithm as mentioned in the theorem, we obtain $k = 10$. Moreover, notice that $a(14) = 10$, i.e., α_{14} contains $a(14)$ elements from J_1, and $n - a(14)$ elements from J_2, i.e., $\alpha_{14} = \alpha_{a(14)}(P_1) \cup \alpha_{14-a(14)}(P_2)$. If $n = 100$, then $k = \lfloor \frac{200}{3} \rfloor = 66$. By running the algorithm as mentioned in the theorem, we obtain $k = 69$. Moreover, we have $a(100) = 69$, i.e., α_{100} contains $a(100)$ elements from J_1, and $n - a(100)$ elements from J_2, i.e., $\alpha_{100} = \alpha_{a(100)}(P_1) \cup \alpha_{100-a(100)}(P_2)$.

4. Optimal quantization for the mixture of two uniform distributions on two connected line segments

Let P_1 and P_2 be two uniform distributions, respectively, on the intervals $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$. Write

$$J_1 := \left[0, \frac{1}{2}\right], \quad \text{and} \quad J_2 := \left[\frac{1}{2}, 1\right].$$

Let f_1 and f_2 be their respective density functions. Then, $f_1(x) = 2$ if $x \in [0, \frac{1}{2}]$, and zero, otherwise; and $f_2(x) = 2$ if $x \in [\frac{1}{2}, 1]$, and zero, otherwise. Let $P := \frac{3}{4}P_1 + \frac{1}{4}P_2$. For such a mixed distribution, in this section, we investigate the optimal sets of n-means and the nth quantization errors for all $n \in \mathbb{N}$. Notice that the density function of the mixed distribution P can be written as follows:

$$f(x) = \begin{cases} \frac{3}{2} & \text{if } x \in J_1, \\ \frac{1}{2} & \text{if } x \in J_2, \\ 0, & \text{otherwise.} \end{cases}$$

Let us now prove the following lemma.

Lemma 4.1. *Let P be the mixed distribution defined by $P = \frac{3}{4}P_1 + \frac{1}{4}P_2$. Then, $E(P) = \frac{3}{8}$, and $V(P) = \frac{13}{192}$.*

Proof. We have

$$E(P)=\int x\,\mathrm{d}P = \frac{3}{4}\int x\,\mathrm{d}\big(P_1(x)\big)+\frac{1}{4}\int x\,\mathrm{d}\big(P_2(x)\big) = \frac{3}{4}\int_0^{\frac{1}{2}} 2x\,\mathrm{d}x+\frac{1}{4}\int_{\frac{1}{2}}^1 2x\,\mathrm{d}x$$

yielding $E(P) = \frac{3}{8}$, and

$$V(P)=\int \big(x-E(P)\big)^2\,\mathrm{d}P=\frac{3}{4}\int\big(x-E(P)\big)^2\,\mathrm{d}\big(P_1(x)\big)+\frac{1}{4}\int\big(x-E(P)\big)^2\,\mathrm{d}\big(P_2(x)\big),$$

implying $V(P) = \frac{13}{192}$, and thus, the lemma is yielded. □

Remark 4.2. The optimal set of one-mean is the set $\{\frac{3}{8}\}$, and the corresponding quantization error is the variance $V := V(P)$ of a random variable with distribution P.

Proposition 4.3. *For $n \geq 2$, let α_n be an optimal set of n-means. Then, $\alpha_n \cap J_1 \neq \emptyset$, and $\alpha_n \cap J_2 \neq \emptyset$.*

Proof. Consider the set of two points $\beta := \big\{\frac{1}{4}, \frac{3}{4}\big\}$. The distortion error due to the set β is given by

$$\begin{aligned}\int \min_{b\in\beta}(x-b)^2\,\mathrm{d}P &= \int_{J_1}\Big(x-\frac{1}{4}\Big)^2\,\mathrm{d}P + \int_{J_2}\Big(x-\frac{3}{4}\Big)^2\,\mathrm{d}P\\ &= \frac{3}{4}\int_0^{\frac{1}{2}} 2\Big(x-\frac{1}{4}\Big)^2\,\mathrm{d}x + \frac{1}{4}\int_{\frac{1}{2}}^1 2\Big(x-\frac{3}{4}\Big)^2\,\mathrm{d}x = \frac{1}{48}.\end{aligned}$$

Since V_n is the quantization error for two-means, and $n \geq 2$, we have $V_n \leq V_2 \leq \frac{1}{48} = 0.0208333$. For the sake of contradiction assume that $\alpha_n \cap J_2 = \emptyset$. Then,

$$V_n > \int_{J_2}\Big(x-\frac{1}{2}\Big)^2\,\mathrm{d}P = \frac{1}{4}\int_{\frac{1}{2}}^1 2\Big(x-\frac{1}{2}\Big)^2\,\mathrm{d}x = \frac{1}{48} \geq V_n,$$

which is a contradiction. Hence, we can assume that $\alpha \cap J_2 \neq \emptyset$. Similarly, we can show that $\alpha_n \cap J_1 \neq \emptyset$. Thus, the proof of the proposition is complete. □

Lemma 4.4. *The set $\{\frac{1}{4}, \frac{3}{4}\}$ forms an optimal set of two-means with quantization error $V_2 = \frac{1}{48}$.*

Proof. Let $\alpha := \{a_1, a_2\}$ be an optimal set of two-means such that $0 < a_1 < a_2 < 1$. By Proposition 4.3, we have $a_1 < \frac{1}{2} < a_2$. The following two cases can arise:

Case 1. $\frac{1}{2} \leq \frac{a_1+a_2}{2}$.

In this case, we have

$$a_1 = \frac{\frac{3}{4}\int_0^{\frac{1}{2}} 2x\,\mathrm{d}x + \frac{1}{4}\int_{\frac{1}{2}}^{\frac{1}{2}(a_1+a_2)} 2x\,\mathrm{d}x}{\frac{3}{4}\int_0^{\frac{1}{2}} 2\,\mathrm{d}x + \frac{1}{4}\int_{\frac{1}{2}}^{\frac{1}{2}(a_1+a_2)} 2\,\mathrm{d}x}, \quad \text{and} \quad a_2 = \frac{1}{2}\Big(\frac{1}{2}\,(a_1+a_2)+1\Big).$$

Solving the above two equations, we have $a_1 = \frac{1}{4}$, and $a_2 = \frac{3}{4}$, with distortion error

$$\begin{aligned} V(P;\alpha) &= \frac{3}{4}\int_0^{\frac{1}{2}} 2(x-a_1)^2\,\mathrm{d}x \\ &\quad + \frac{1}{4}\int_{\frac{1}{2}}^{\frac{1}{2}(a_1+a_2)} 2(x-a_1)^2\,\mathrm{d}x + \frac{1}{4}\int_{\frac{1}{2}(a_1+a_2)}^{1} 2\,(x-a_2)^{\,2}\,\mathrm{d}x = \frac{1}{48}. \end{aligned}$$

Case 2. $\frac{a_1+a_2}{2} < \frac{1}{2}$.

Proceeding in the similar way as Case 1, we obtain two equations, and see that there is no solution in this case.

Considering the above two cases, we see that the set $\{\frac{1}{4}, \frac{3}{4}\}$ forms an optimal set of two-means with quantization error $\frac{1}{48}$, which is the lemma. □

Lemma 4.5. *The set*

$$\Big\{\frac{1}{3}\Big(\frac{1}{8}(21-\sqrt{3})-2\Big),\ \frac{1}{8}(21-\sqrt{3})-2,\ \frac{1}{24}(21-\sqrt{3})\Big\}$$

forms an optimal set of three-means with quantization error $V_3 = 0.00787482$.

Proof. Consider the set of three points $\beta := \{u, v, w\}$, where

$$u = \frac{1}{3}\Big(\frac{1}{8}\Big)(21-\sqrt{3})-2, \quad v = \frac{1}{8}(21-\sqrt{3})-2, \quad \text{and} \quad w = \frac{1}{24}(21-\sqrt{3}).$$

Since $0 < u < v < \frac{1}{2} < \frac{v+w}{2} < w < 1$, the distortion error due to the set β is given by

$$\begin{aligned} V(P;\beta) &= \frac{3}{4}\int_0^{\frac{u+v}{2}} 2(x-u)^2\,\mathrm{d}x + \frac{3}{4}\int_{\frac{u+v}{2}}^{\frac{1}{2}} 2(x-v)^2\,\mathrm{d}x \\ &\quad + \frac{1}{4}\int_{\frac{1}{2}}^{\frac{v+w}{2}} 2(x-v)^2\,\mathrm{d}x + \frac{1}{4}\int_{\frac{v+w}{2}}^{1} 2(x-w)^2\,\mathrm{d}x \end{aligned}$$

yielding $V(P;\beta) = 0.00787482$. Since V_3 is the quantization error for three-means we have $V_3 \leq 0.00787482$. Let $\alpha := \{a, b, c\}$ be an optimal set of three-means. Without any loss of generality we can assume that $0 < a < b < c < 1$. By Proposition 4.3, we know $a < \frac{1}{2} < c$. We now show that $b < \frac{1}{2}$. Suppose that $\frac{9}{16} < b$. Then,

$$\begin{aligned} V_3 &\geq \int_{J_1} \min_{r\in\{a,\frac{9}{16}\}} (x-r)^2 \, \mathrm{d}P \\ &= \frac{3}{4}\int_0^{\frac{1}{2}(a+\frac{9}{16})} 2(x-a)^2 \, \mathrm{d}x + \frac{3}{4}\int_{\frac{1}{2}(a+\frac{9}{16})}^{\frac{1}{2}} 2\Big(x-\frac{9}{16}\Big)^2 \, \mathrm{d}x \\ &= \frac{12288a^3 + 6912a^2 - 3888a + 725}{32768}, \end{aligned}$$

the minimum value of which is 0.00976563 and it occurs when $a = \frac{3}{16}$, and thus, we have $V_3 \geq 0.00976563 > V_3$, which is a contradiction. So, we can assume that $b \leq \frac{9}{16}$. Next, assume that $\frac{1}{2} \leq b \leq \frac{9}{16}$. Notice that then $\frac{9}{16} < c < 1$. Then, as before we have

$$\begin{aligned} V_3 &\geq \int_{J_1} \min_{r\in\{a,\frac{1}{2}\}} (x-r)^2 \, \mathrm{d}P + \int_{\frac{9}{16}}^{1} \min_{s\in\{\frac{9}{16},c\}} (x-r)^2 \, \mathrm{d}P \\ &= \frac{1}{64}(24a^3 + 12a^2 - 6a + 1) + \frac{-12288c^3 + 42240c^2 - 45264c + 15655}{98304}, \end{aligned}$$

the minimum value of which is $\frac{1}{144} + \frac{343}{221184} = 0.00849519$, and it occurs when $a = 0.166667$, and $c = 0.854167$. Thus, we have $V_3 \geq 0.00849519 > V_3$, which is a contradiction. Hence, we can assume that $b < \frac{1}{2}$. Then, the two cases can arise: either $\frac{1}{2}(b+c) < \frac{1}{2}$, or $\frac{1}{2} \leq \frac{1}{2}(b+c)$. Proceeding as in Lemma 4.4, we can see that $\frac{1}{2}(b+c) < \frac{1}{2}$ can not happen. Thus, we have $\frac{1}{2} \leq \frac{1}{2}(b+c)$ implying

$$a = \frac{a+b}{4}, \quad b = \frac{\frac{3}{4}\int_{\frac{a+b}{2}}^{\frac{1}{2}} 2x \, \mathrm{d}x + \frac{1}{4}\int_{\frac{1}{2}}^{\frac{b+c}{2}} 2x \, \mathrm{d}x}{\frac{3}{4}\int_{\frac{a+b}{2}}^{\frac{1}{2}} 2 \, \mathrm{d}x + \frac{1}{4}\int_{\frac{1}{2}}^{\frac{b+c}{2}} 2 \, \mathrm{d}x}, \quad \text{and} \quad c = \frac{\int_{\frac{b+c}{2}}^{1} 2x \, \mathrm{d}x}{\frac{4}{4}\int_{\frac{b+c}{2}}^{1} 2 \, \mathrm{d}x}.$$

Solving the above equations, we have

$$a = \frac{1}{3}\Big(\frac{1}{8}(21-\sqrt{3}) - 2\Big), \quad b = \frac{1}{8}(21-\sqrt{3}) - 2, \quad \text{and} \quad c = \frac{1}{24}(21-\sqrt{3}),$$

and the corresponding quantization error is given by $V_3 = 0.00787482$, and thus, the proof of the lemma is complete. □

Definition 4.6. For $n \in \mathbb{N}$, define the sequence $\{a(n)\}_{n=1}^{\infty}$ as follows:

$$a(n) := \Big\lfloor \frac{5(n+1)}{8} \Big\rfloor,$$

i.e.,

$$\{a(n)\}_{n=1}^{\infty} = \{1, 1, 2, 3, 3, 4, 5, 5, 6, 6, 7, 8, 8, 9, 10, 10, 11, 11, 12, 13, 13, 14, 15, 15, 16, 16, \ldots\}.$$

Let us now state and prove the following two claims.

Claim 4.7. Let $\{a(n)\}$ be the sequence defined by Definition 4.6. Take $n = 8$, and then $a(n) = 5$. Assume that $\alpha_n := \{a_1 < a_2 < a_3 < a_4 < a_5 < b_1 < b_2 < b_3\}$ is an optimal set of eight-means for P. Then, $\frac{1}{2} \leq \frac{1}{2}(a_5 + b_1)$.

P r o o f. For the sake of contradiction, assume that $\frac{1}{2}(a_5 + b_1) < \frac{1}{2}$. Then,

$$a_1 = \frac{1}{2}\Big(0 + \frac{a_1 + a_2}{2}\Big), \quad \text{and} \quad a_2 = \frac{1}{2}\Big(\frac{a_1 + a_2}{2} + \frac{a_2 + a_3}{2}\Big)$$

implying $a_1 = \frac{1}{3}a_2$, and $a_2 = \frac{3}{5}a_3$. Similarly, $a_3 = \frac{5}{7}a_4$, $a_4 = \frac{7}{9}a_5$. Again,

$$b_2 = \frac{1}{2}\Big(\frac{b_1 + b_2}{2} + \frac{b_2 + b_3}{2}\Big), \quad \text{and} \quad b_3 = \frac{1}{2}\Big(\frac{b_2 + b_3}{2} + 1\Big)$$

implying $b_2 = \frac{3}{5}b_1 + \frac{2}{5}$, and $b_3 = \frac{1}{3}b_2 + \frac{2}{3}$. Moreover,

$$a_5 = \frac{1}{2}\Big(\frac{a_4 + a_5}{2} + \frac{a_5 + b_1}{2}\Big) = \frac{1}{2}\Bigg(\frac{\frac{7}{9}a_5 + a_5}{2} + \frac{a_5 + b_1}{2}\Bigg)$$

implying $a_5 = \frac{9}{11}b_1$, and

$$b_1 = E\Big(X : X \in \Big[\frac{a_5 + b_1}{2}, \frac{1}{2}\Big] \cup \Big[\frac{1}{2}, \frac{b_1 + b_2}{2}\Big]\Big)$$

$$= \frac{-6a_5 b_1 - 3a_5^2 - 2b_1^2 + b_2^2 + 2b_1 b_2 + 2}{-12a_5 - 8b_1 + 4b_2 + 8}.$$

Next, putting the values of a_5 and b_2 in the expression of b_1, we have

$$b_1 = \frac{-11128b_1^2 + 1936b_1 + 3267}{14520 - 23320b_1} \quad \textit{yielding} \quad b_1 = \frac{11\left(143 \pm 5i\sqrt{5}\right)}{3048},$$

which is not real. Thus, $\frac{1}{2}(a_5 + b_1) < \frac{1}{2}$ leads to a contradiction. Hence, $\frac{1}{2} \leq \frac{1}{2}(a_5 + b_1)$. □

Claim 4.8. Let $\{a(n)\}$ be the sequence defined by Definition 4.6. Take $n = 9$, and then $a(n) = 6$. Assume that $\alpha_n := \{a_1 < a_2 < a_3 < a_4 < a_5 < a_6 < b_1 < b_2 < b_3\}$ is an optimal set of nine-means for P. Then, $\frac{1}{2} \leq \frac{1}{2}(a_6 + b_1)$.

Proof. For the sake of contradiction, assume that $\frac{1}{2}(a_6+b_1)<\frac{1}{2}$. Then,

$$a_1=\frac{1}{2}\Big(0+\frac{a_1+a_2}{2}\Big), \quad \text{and} \quad a_2=\frac{1}{2}\Big(\frac{a_1+a_2}{2}+\frac{a_2+a_3}{2}\Big)$$

implying $a_1=\frac{1}{3}a_2$, and $a_2=\frac{3}{5}a_3$. Similarly, $a_3=\frac{5}{7}a_4$, $a_4=\frac{7}{9}a_5$, and $a_5=\frac{9}{11}a_6$. Again,

$$b_2=\frac{1}{2}\Big(\frac{b_1+b_2}{2}+\frac{b_2+b_3}{2}\Big), \quad \text{and} \quad b_3=\frac{1}{2}\Big(\frac{b_2+b_3}{2}+1\Big)$$

implying $b_2=\frac{3}{5}b_1+\frac{2}{5}$, and $b_3=\frac{1}{3}b_2+\frac{2}{3}$. Moreover,

$$a_6=\frac{1}{2}\Big(\frac{a_5+a_6}{2}+\frac{a_6+b_1}{2}\Big)=\frac{1}{2}\Big(\frac{\frac{9}{11}a_6+a_6}{2}+\frac{a_6+b_1}{2}\Big) \quad \text{implying} \quad a_6=\frac{11}{13}b_1,$$

and

$$b_1=E\Big(X : X\in\Big[\frac{a_6+b_1}{2},\frac{1}{2}\Big]\cup[\frac{1}{2},\frac{b_1+b_2}{2}]\Big)$$
$$=\frac{-6a_5b_1-3a_5^2-2b_1^2+b_2^2+2b_1b_2+2}{-12a_5-8b_1+4b_2+8}.$$

Next, putting the values of a_5 and b_2 in the expression of b_1, we have

$$b_1=\frac{-16192b_1^2+2704b_1+4563}{20280-33280b_1} \quad \text{yielding} \quad b_1=\frac{13\left(169\pm 5i\sqrt{11}\right)}{4272},$$

which is not real. Thus, $\frac{1}{2}(a_6+b_1)<\frac{1}{2}$ leads to a contradiction. Hence, $\frac{1}{2}\leq\frac{1}{2}(a_6+b_1)$. ☐

Lemma 4.9. *Let α_n be an optimal set of n-means for Pň1, where $n\geq 2$, and $\{a(n)\}$ be the sequence defined by Definition 4.6. Then, card$(\alpha_n\cap J_1)=a(n)$, and card$(\alpha_n\cap J_2)=n-a(n)$.*

Proof. We prove the lemma by induction. By Lemma 4.4 and Lemma 4.5, the lemma is true for $n=2,3$. Assume that that the lemma is true for $n=\ell$, i.e., card$(\alpha_\ell\cap J_1)=a(\ell)$, and card$(\alpha_\ell\cap J_2)=n-a(\ell)$. We need to show that card$(\alpha_{\ell+1}\cap J_1)=a(\ell+1)$. Assume that card$(\alpha_{\ell+1}\cap J_1)=k$, i.e., $\alpha_{\ell+1}$ contains k elements from J_1, and $n-k$ elements from J_2. Let

$$\alpha_{\ell+1}\cap J_1=\{a_1<a_2<\cdots<a_k\}, \quad \text{and} \quad \alpha_{\ell+1}\cap J_2=\{b_1<b_2<\cdots<b_{n-k}\}.$$

Then, either $\frac{1}{2}(a_k+b_1)<\frac{1}{2}$, or $\frac{1}{2}<\frac{1}{2}(a_k+b_1)$. In each case, using the similar techniques as in the proofs of Claim 4.7 and Claim 4.8, if the solution exists, we solve for $a_1,a_2,\ldots,a_k,b_1,\ldots,b_{n-1}$, and find the distortion errors. Notice that at least one solution will exist. Let $V(k,n-k)$ be the minimum of the distortion errors if $\alpha_{\ell+1}$ contains k elements from J_1, and $n-k$ elements from J_2.

Let us now run the following algorithm:

(i) Write $k := a(\ell)$.

(ii) If $V\big(k-1, n-(k-1)\big) < V(k, n-k)$ replace k by $k-1$ and return, else go to step (iii).

(iii) If $V\big(k+1, n-(k+1)\big) < V(k, n-k)$ replace k by $k+1$ and return, else step (iv).

(iv) End.

After running the above algorithm, we see that the value of k obtained equals $a(\ell+1)$, i.e., the lemma is true for $n = \ell+1$ if it is true for $n = \ell$. Hence, by the Induction Principle, we can say that the lemma is true for all positive integers $n \geq 2$, i.e., $\text{card}(\alpha_n \cap J_1) = a(n)$ for any positive integer $n \geq 2$. Since $\text{card}(\alpha_n \cap J_1) + \text{card}(\alpha_n \cap J_2) = n$, we have $\text{card}(\alpha_n \cap J_2) = n - a(n)$. Thus, the proof of the lemma is complete. □

Let us now state and prove the following theorem which is the main theorem in this section.

Theorem 4.10. *Let α_n be an optimal set of n-means for P, where $n \geq 2$, and $\{a(n)\}$ be the sequence defined by Definition 4.6. Write $k := a(n)$, $m := n - a(n)$. Then,*

$$\alpha_n := \{a_1 < a_2 < \cdots < a_k < b_1 < b_2 < \cdots < b_m\},$$

where

$$a_j = \begin{cases} \frac{a_1+a_2}{4} & \text{if} \quad j = 1, \\ \frac{1}{2}\big(\frac{a_{j-1}+a_j}{2} + \frac{a_j+a_{j+1}}{2}\big) & \text{if} \quad 2 \leq j \leq k-1, \\ E\big(X : X \in [\frac{a_{k-1}+a_k}{2}, \frac{1}{2}] \cup [\frac{1}{2}, \frac{a_k+b_1}{2}]\big) & \text{if} \quad j = k, \end{cases}$$

and

$$b_j = \begin{cases} \frac{1}{2}\big(\frac{a_k+b_1}{2} + \frac{b_1+b_2}{2}\big) & \text{if} \quad j = 1, \\ \frac{1}{2}\big(\frac{b_{j-1}+b_j}{2} + \frac{b_j+b_{j+1}}{2}\big) & \text{if} \quad 2 \leq j \leq m-1, \\ \frac{1}{2}\big(\frac{b_{m-1}+b_m}{2} + 1\big) & \text{if} \quad j = m, \end{cases}$$

and the corresponding quantization error is given by

$$\begin{aligned} V_n = \frac{1}{48}\big(&-3b_1^2ma_k + 3b_1ma_k^2 - 3b_1^2a_k + 3b_1a_k^2 - ma_k^3 + 21a_1^3(k-1) + \\ &9a_2a_1^2(k-1) - 9a_2^2a_1(k-1) + 3a_2^3(k-1) - \\ &3a_{k-1}^3 - 14a_k^3 - 9a_{k-1}a_k^2 + 24a_k^2 + 9a_{k-1}^2a_k - 12a_k + \\ &b_2^3m - 3b_1b_2^2m + 3b_1^2b_2m + b_1^3 + 2\big). \end{aligned}$$

Proof. By Lemma 4.9, the optimal set α_n of n-means contains k elements from J_1, and m elements from J_2, where $k = a(n)$ and $m = n - k$. Let $\alpha_n := \{a_1 < a_2 < \cdots < a_k < b_1 < b_2 < \cdots < b_m\}$. Recall Theorem 1.2, and the fact that P_1 is a uniform distribution on $[0, \frac{1}{2}]$, and P_2 is a uniform distribution on $[\frac{1}{2}, 1]$. Thus, we have

$$a_j = \begin{cases} \frac{a_1+a_2}{4} & \text{if} \quad j = 1, \\ \frac{1}{2}\big(\frac{a_{j-1}+a_j}{2} + \frac{a_j+a_{j+1}}{2}\big) & \text{if} \quad 2 \leq j \leq k-1, \end{cases}$$

and

$$b_j = \begin{cases} \frac{1}{2}\big(\frac{b_{j-1}+b_j}{2} + \frac{b_j+b_{j+1}}{2}\big) & \text{if} \quad 2 \leq j \leq m-1, \\ \frac{1}{2}(\frac{b_{m-1}+b_m}{2} + 1) & \text{if} \quad j = m. \end{cases}$$

The following two cases can arise:

Case 1. $\frac{1}{2} \leq \frac{1}{2}(a_k + b_1)$.

In this case, we have

$$a_k = E\big(X : X \in \big[\tfrac{a_{k-1}+a_k}{2}, \tfrac{1}{2}\big] \cup \big[\tfrac{1}{2}, \tfrac{a_k+b_1}{2}\big]\big), \text{ and } b_1 = \tfrac{1}{2}\big(\tfrac{a_k+b_1}{2} + \tfrac{b_1+b_2}{2}\big).$$

Case 2. $\frac{1}{2}(a_k + b_1) < \frac{1}{2}$.

In this case, we have

$$a_k = \tfrac{1}{2}\big(\tfrac{a_{k-1}+a_k}{2} + \tfrac{a_k+b_1}{2}\big), \text{ and } b_1 = E\big(X : X \in \big[\tfrac{a_k+b_1}{2}, \tfrac{1}{2}\big] \cup \big[\tfrac{1}{2}, \tfrac{b_1+b_2}{2}\big]\big).$$

For any given positive integer, using the similar techniques as in the proofs of Claim 4.7 and Claim 4.8, we see that in Case 2, the system of equations to obtain $a_1, a_2, \ldots, a_k, b_1, \ldots, b_m$ does not have any solution. Hence Case 2 cannot happen.

Thus, we have $\frac{1}{2} \leq \frac{1}{2}(a_k + b_1)$, i.e., the system of equations to obtain the elements $a_1, a_2, \ldots, a_k, b_1, \ldots, b_m$ as stated in the theorem are true, and hence, the corresponding quantization error is given by

$$\begin{aligned} V_n &= \frac{3(k-1)}{4} \int_0^{\frac{a_1+a_2}{2}} 2(x-a_1)^2 \mathrm{d}x + \frac{3}{4} \int_{\frac{a_{k-1}+a_k}{2}}^{\frac{1}{2}} 2(x-a_k)^2 \mathrm{d}x \\ &\quad + \frac{1}{4} \int_{\frac{1}{2}}^{\frac{a_k+b_1}{2}} 2(x-a_k)^2 \mathrm{d}x \frac{m}{4} \int_{\frac{a_k+b_1}{2}}^{\frac{b_1+b_2}{2}} 2(x-b_1)^2 \mathrm{d}x \\ &= \frac{1}{48}\Big(- 3b_1^2 m a_k + 3b_1 m a_k^2 - 3b_1^2 a_k + 3b_1 a_k^2 - m a_k^3 + 21 a_1^3 (k-1) \\ &\quad + 9a_2 a_1^2 (k-1) - 9a_2^2 a_1 (k-1) + 3a_2^3 (k-1) - 3a_{k-1}^3 - 14 a_k^3 - 9a_{k-1} a_k^2 \\ &\quad + 24 a_k^2 + 9a_{k-1}^2 a_k - 12 a_k + b_2^3 m - 3b_1 b_2^2 m + 3b_1^2 b_2 m + b_1^3 + 2\Big). \end{aligned}$$

Thus, we complete the proof of the theorem. □

Now, we give the following example.

Example 4.11. Take $n = 16$. Then, $k = a(n) = 10$, and so, $m = 6$. Thus, by Theorem 4.10, we have

$$\begin{aligned}\{a_1 &= 0.0255733, & a_2 &= 0.0767199, & a_3 &= 0.127866, & a_4 &= 0.179013,\\ a_5 &= 0.23016, & a_6 &= 0.281306, & a_7 &= 0.332453, & a_8 &= 0.383599,\\ a_9 &= 0.434746, & a_{10} &= 0.485893, & b_1 &= 0.564986, & b_2 &= 0.644079,\\ b_3 &= 0.723173, & b_4 &= 0.802266, & b_5 &= 0.88136, & b_6 &= 0.960453,\}\end{aligned}$$

and the corresponding quantization error is given by

$$\begin{aligned}V_{16} = \frac{1}{48}\big(&-21a_{10}b_1^2 + 21a_{10}^2 b_1 + 189a_1^3 + 81a_2a_1^2 - 81a_2^2a_1 + \\ &27a_2^3 - 3a_9^3 - 20a_{10}^3 - 9a_9a_{10}^2 + 24a_{10}^2 + 9a_9^2a_{10} - \\ &12a_{10} + b_1^3 + 6b_2^3 - 18b_1b_2^2 + 18b_1^2b_2 + 2\big) = 0.000293827.\end{aligned}$$

REFERENCES

[BW] BUCKLEW, J. A.— WISE, G. L.: *Multidimensional asymptotic quantization theory with rth power distortion measures*, IEEE Trans. Inform. Theory, **28** (1982), no. 2, 239–247.

[CR] ÇÖMEZ, D.—ROYCHOWDHURY, M. K.: *Quantization for uniform distributions on stretched Sierpiński triangles*, Monatsh.Math. **190** (2019), no. 1, 79–100.

[DR1] DETTMANN, C. P.—ROYCHOWDHURY, M. K.: *Quantization for uniform distributions on equilateral triangles*, Real Anal. Exchange, **42** (2017), no. (1), 149–166.

[DR2] DETTMANN, C. P.—ROYCHOWDHURY, M. K.: *An algorithm to compute CVTs for finitely generated Cantor distributions*, Southeast Asian Bull. Math. (to appear).

[GG] GERSHO, A.—GRAY, R. M.: *Vector Quantization and Signal Compression*. Kluwer Academy publishers: Boston, MA, 1992.

[GL1] GRAF, S.—LUSCHGY, H.: *Foundations of Quantization for Probability Distributions*. In: *Lecture Notes in Mathematics Vol. 1730*, Springer-Verlag, Berlin, 2000.

[GL2] GRAF, S.—LUSCHGY, H.: *The Quantization of the Cantor distribution*, Math. Nachr. **183** (1997), 113–133.

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

[L] LINDSAY, L. J.: *Quantization dimension for probability distributions*, Thesis (Ph.D.)–University of North Texas. ProQuest LLC, Ann Arbor, MI, 2001.

[L1] ROYCHOWDHURY, L.: *Optimal quantization for nonuniform Cantor distributions*, J. Interdiscip. Math. **22** (2019), no. 8, 1325–1348.

[P] PÖTZELBERGER, K.: *The quantization dimension of distributions*, Math. Proc. Cambridge Philos. Soc. **131** (2001), no. 3, 507–519.

[R1] ROYCHOWDHURY, M. K.: *Quantization and centroidal Voronoi tessellations for probability measures on dyadic Cantor sets*, J. Fractal Geom. **4** (2017), no. 2, 127–146.

[R2] ROYCHOWDHURY, M. K.: *Optimal quantizers for some absolutely continuous probability measures*, Real Anal. Exchange **43** (2017), no. 1, 105–136.

[R3] ROYCHOWDHURY, M. K.: *Optimal quantization for the Cantor distribution generated by infinite similitudes*, Israel J. Math. **231** (2019), no. 1, 437–466.

[R4] ROYCHOWDHURY, M. K.: *Least upper bound of the exact formula for optimal quantization of some uniform Cantor distributions*, Discrete Contin. Dyn. Syst. Ser. A, **38** (2018), no. 9, 4555–4570.

[R5] ROYCHOWDHURY, M. K.: *Center of mass and the optimal quantizers for some continuous and discrete uniform distributions*, J. Interdiscip. Math. **22** (2019), no. 4, 451–471.

[RR1] ROSENBLATT, J.—ROYCHOWDHURY, M. K.: *Optimal quantization for piecewise uniform distributions*, Unif. Distrib. Theory **13** (2018), no. 2, 23–55.

[RR2] ROSENBLATT, J.—ROYCHOWDHURY, M. K.: *Uniform distributions on curves and quantization*, arXiv:1809.08364 [math.PR].

[Z] ZAMIR, R.: *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge University Press, Cambridge, 2014.

Received October 18, 2019
Accepted March 1, 2020

Mrinal Kanti Roychowdhury
Wasiela Salinas
School of Mathematical and
Statistical Sciences
University of Texas Rio Grande Valley
1201 West University Drive
Edinburg, TX 78539-2999, USA.
E-mail: mrinal.roychowdhury@utrgv.edu
wsalinas47@gmail.com