

UNIFORM DISTRIBUTIONS ON CURVES AND QUANTIZATION

JOSEPH ROSENBLATT AND MRINAL KANTI ROYCHOWDHURY

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 to make an approximation of a continuous probability distribution by a discrete distribution. It has broad application in signal processing and data compression. In this paper, first we define the uniform distributions on different curves such as a line segment, a circle, and the boundary of an equilateral triangle. Then, we give the exact formulas to determine the optimal sets of n -means and the n th quantization errors for different values of n with respect to the uniform distributions defined on the curves. In each case, we further calculate the quantization dimension and show that it is equal to the dimension of the object; and the quantization coefficient exists as a finite positive number. This supports the well-known result of Bucklew and Wise [2], which says that for a Borel probability measure P with non-vanishing absolutely continuous part the quantization coefficient exists as a finite positive number.

1. Introduction

‘Quantization’ refers to the process of approximating the continuous set of values in the image data with a finite (preferably small) set of values with broad application in engineering and technology (see [4, 7, 12]). Although the work of quantization in engineering science has a long history, rigorous mathematical treatment was given by Graf and Luschgy (see [6]). The quantization for a probability distribution refers to the idea of estimating a given probability by a discrete probability with a given number n of supporting points.

Let P denote a Borel probability measure on \mathbb{R}^d , where $d \geq 1$. For a finite set $\alpha \subset \mathbb{R}^d$, the error $\int \min_{a \in \alpha} \|x - a\|^2 dP(x)$ is often referred to as the *cost* or *distortion error* for α , and is denoted by $V(P; \alpha)$. For any positive integer n , write $V_n := V_n(P) = \inf\{V(P; \alpha) : \alpha \subset \mathbb{R}^d, \text{card}(\alpha) \leq n\}$. Then, V_n is called the n th quantization error for P . We assume that $\int \|x\|^2 dP(x) < \infty$ to make

Received December 28, 2021; Revised June 4, 2022; Accepted July 28, 2022.

2020 *Mathematics Subject Classification.* Primary 60Exx, 94A34.

Key words and phrases. Uniform distribution, optimal quantizers, quantization error, quantization dimension, quantization coefficient.

sure that there is a set α for which the infimum occurs (see [1, 5, 6, 8]). Such a set α for which the infimum occurs and contains no more than n -points is called an *optimal set of n -means* and the elements of an optimal set are called *optimal quantizers*. In some literature it is also referred to as *principal points* (see [9]). The numbers

$$\underline{D}(P) := \liminf_{n \rightarrow \infty} \frac{2 \log n}{-\log V_n(P)}, \text{ and } \overline{D}(P) := \limsup_{n \rightarrow \infty} \frac{2 \log n}{-\log V_n(P)}$$

are, respectively, called the *lower* and *upper quantization dimensions* of the probability measure P . If $\underline{D}(P) = \overline{D}(P)$, the common value is called the *quantization dimension* of P and is denoted by $D(P)$. Quantization dimension measures the speed at which the specified measure of the error tends to zero as n approaches to infinity. For any $s > 0$, the two numbers $\liminf_{n \rightarrow \infty} n^{2/s} V_n(P)$ and $\limsup_{n \rightarrow \infty} n^{2/s} V_n(P)$ are, respectively, called the *s -dimensional lower* and *upper quantization coefficients* for P . If the s -dimensional lower and upper quantization coefficients are equal, we call it the *s -dimensional quantization coefficient* for P . If the s -dimensional lower and the upper quantization coefficients are finite and positive, then s equals the quantization dimension of P (see [6]). It is well-known that for a Borel probability measure P with non-vanishing absolutely continuous part $\lim_{n \rightarrow \infty} n^{2/d} V_n(P)$ is finite and strictly positive (see [2]). This implies that the quantization dimension of a Borel probability measure with non-vanishing absolutely continuous part equals the dimension d of the underlying space. For a finite set $\alpha \subset \mathbb{R}^d$, the *Voronoi region* generated by $a \in \alpha$, denoted by $M(a|\alpha)$, is defined to be the set of all elements in \mathbb{R}^d which are nearest to a . The set $\{M(a|\alpha) : a \in \alpha\}$ is called the *Voronoi diagram* or *Voronoi tessellation* of \mathbb{R}^d with respect to α . The point a is called the centroid of its own Voronoi region if $a = E(X : X \in M(a|\alpha))$, where X is a P -distributed random variable. Let us now state the following proposition (see [4, 6]).

Proposition 1.1. *Let α be an optimal set of n -means, $a \in \alpha$, and $M(a|\alpha)$ be the Voronoi region generated by $a \in \alpha$. Then, for every $a \in \alpha$, (i) $P(M(a|\alpha)) > 0$, (ii) $P(\partial M(a|\alpha)) = 0$, (iii) $a = E(X : X \in M(a|\alpha))$, and (iv) P -almost surely the set $\{M(a|\alpha) : a \in \alpha\}$ forms a Voronoi partition of \mathbb{R}^d .*

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 . Recently, the optimal quantization for uniform distributions on different regions has been investigated by several authors, for example, see [3, 10, 11].

In this paper, there are three subsections. In Subsection 2.1, Subsection 2.2, and Subsection 2.3, first we have defined the uniform distributions on a line segment, on a unit circle, and on the boundary of an equilateral triangle. Then, in Theorem 2.1.1, Theorem 2.2.1, and Theorem 2.3.10, we give the exact formulas to determine the optimal sets of n -means and the corresponding quantization

errors for different values of $n \in \mathbb{N}$. We have further shown that the quantization dimension of the uniform distribution in each case is one, which equals the dimension of the line, circle, and the boundary of the equilateral triangle. Moreover, we have shown that in each case the quantization coefficient exists as a finite positive number. As described in Remark 2.4, once the optimal sets of n -means are known for a line segment, a circle, and an equilateral triangle, by giving an affine transformation, one can easily obtain them for any line segment, any circle, and the boundary of any equilateral triangle. We did not find any literature where the optimal quantization for a probability distributions on a curve has been investigated. Thus, our work in this paper can be considered as a first advance in this direction. Finally, we would like to mention that the technique used in this paper will be of great help to investigate the optimal quantization for any distribution on any curve.

2. Main result

In this section, in three subsections, we give the main results of the paper. Let i and j be the unit vectors in the positive directions of x_1 - and x_2 -axes, respectively. By the position vector \tilde{a} of a point A , it is meant that $\overrightarrow{OA} = \tilde{a}$. In the sequel, we will identify the position vector of a point (a_1, a_2) by $(a_1, a_2) := a_1i + a_2j$, and apologize for any abuse in notation. For any two vectors \vec{u} and \vec{v} , let $\vec{u} \cdot \vec{v}$ denote the dot product between the two vectors \vec{u} and \vec{v} . Then, for any vector \vec{v} , by $(\vec{v})^2$, we mean $(\vec{v})^2 := \vec{v} \cdot \vec{v}$. Thus, $|\vec{v}| := \sqrt{\vec{v} \cdot \vec{v}}$, which is called the length of the vector \vec{v} . For any two position vectors $\tilde{a} := (a_1, a_2)$ and $\tilde{b} := (b_1, b_2)$, we write $\rho(\tilde{a}, \tilde{b}) := \|(a_1, a_2) - (b_1, b_2)\|^2 = (a_1 - b_1)^2 + (a_2 - b_2)^2$, which gives the squared Euclidean distance between the two points (a_1, a_2) and (b_1, b_2) . 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 and Q with position vectors \tilde{p} and \tilde{q} belong to an optimal set of n -means for some positive integer n , and let D with position vector \tilde{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 $(\tilde{p} - \tilde{d})^2 = (\tilde{q} - \tilde{d})^2$, i.e., $\rho(\tilde{p}, \tilde{d}) = \rho(\tilde{q}, \tilde{d})$. In the sequel, such an equation will be referred to as a *canonical equation*. In the paper, there are some decimal numbers; they are rational approximations of some real numbers.

2.1. Optimal quantization for a uniform distribution on a line segment

Without any loss of generality we can assume the line segment as a closed interval $[a, b]$, where $0 < a < b < +\infty$. Let P be the uniform distribution defined on the closed interval $[a, b]$. Let the probability density function for P

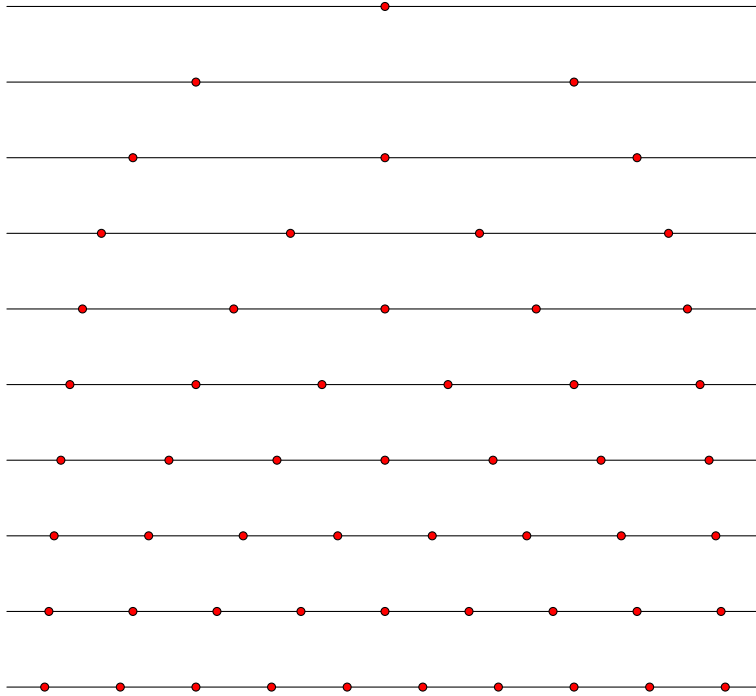


FIGURE 1. Optimal configuration of n means with respect to a uniform distribution on a line segment for $1 \leq n \leq 10$.

is given by f . Then, we have

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b], \\ 0 & \text{otherwise,} \end{cases}$$

implying $dP(x) = P(dx) = f(x)dx = \frac{1}{b-a}dx$.

The following theorem gives the optimal sets of n -means for all $n \in \mathbb{N}$.

Theorem 2.1.1. *Let P be the uniform distribution on the closed interval $[a, b]$. Then, the optimal set n -means is given by $\alpha_n := \{a + \frac{2^{i-1}}{2^n}(b-a) : 1 \leq i \leq n\}$, and the corresponding quantization error is $V_n := V_n(P) = \frac{(a-b)^2}{12n^2}$.*

Proof. Let $\alpha_n := \{a_1, a_2, \dots, a_n\}$ be an optimal set of n -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 $a < a_1 < a_2 < \dots < a_n < b$. Then, by Proposition 1.1, we have

$$(1) \quad a_1 = E(X : X \in [a, \frac{1}{2}(a_1 + a_2)]),$$

$$(2) \quad a_i = E(X : X \in [\frac{a_{i-1} + a_i}{2}, \frac{a_i + a_{i+1}}{2}]) \text{ for } 2 \leq i \leq n - 1,$$

$$(3) \quad a_n = E(X : X \in [\frac{a_{n-1} + a_n}{2}, b]).$$

By (1), we have

$$a_1 = \frac{\int_a^{\frac{1}{2}(a_1+a_2)} x dP}{\int_a^{\frac{1}{2}(a_1+a_2)} dP} = \frac{\int_a^{\frac{1}{2}(a_1+a_2)} x f(x) dx}{\int_a^{\frac{1}{2}(a_1+a_2)} f(x) dx} = \frac{1}{4} (2a + a_1 + a_2)$$

implying

$$(4) \quad 2(a_1 - a) = a_2 - a_1.$$

Similarly, by (2) and (3), we have

$$(5) \quad a_{i-1} - a_i = a_{i+1} - a_i \text{ for } 2 \leq i \leq n - 1, \text{ and } a_n - a_{n-1} = 2(b - a_n).$$

Combining the equations in (4) and (5), we have

$$(6) \quad 2(a_1 - a) = a_i - a_{i-1} = 2(b - a_n) = k \text{ (say) for } 2 \leq i \leq n$$

implying

$$(a_1 - a) + \sum_{i=2}^n (a_i - a_{i-1}) + (b - a_n) = \frac{k}{2} + (n - 1)k + \frac{k}{2}, \text{ i.e., } b - a = nk.$$

Thus, putting $k = \frac{b-a}{n}$, by (6), we have

$$a_1 = a + \frac{1}{2n}(b-a), a_2 = a + \frac{3}{2n}(b-a), a_3 = a + \frac{5}{2n}(b-a), \dots, a_n = a + \frac{2n-1}{2n}(b-a)$$

yielding the fact that $\alpha_n := \{a + \frac{2i-1}{2n}(b-a) : 1 \leq i \leq n\}$ forms an optimal set of n -means for P (see Figure 1). Notice that the probability density function is constant, and the Voronoi regions of the points a_i for $1 \leq i \leq n$ are of equal lengths. This yields the fact that the distortion errors due to each a_i are equal. Hence, the n th quantization error $V_n := V_n(P)$ is given by

$$V_n = \int \min_{a \in \alpha_n} (x - a)^2 dP = \frac{n}{b-a} \int_a^{\frac{1}{2}(a_1+a_2)} (x - a_1)^2 dx = \frac{(a-b)^2}{12n^2}.$$

Thus, the proof of the theorem is complete. □

Remark 2.1.2. By Theorem 2.1.1, we see that if V_n is the quantization error of n -means for the uniform distribution on $[a, b]$, then $V_n = \frac{(a-b)^2}{12n^2}$. Thus,

$$D(P) = \lim_{n \rightarrow \infty} \frac{2 \log n}{-\log V_n} = \lim_{n \rightarrow \infty} \frac{2 \log n}{-\log \frac{(a-b)^2}{12n^2}} = 1,$$

which equals the dimension of the line segment. Moreover, we have $\lim_{n \rightarrow \infty} n^2 V_n = \frac{(a-b)^2}{12}$, which is a finite positive number.

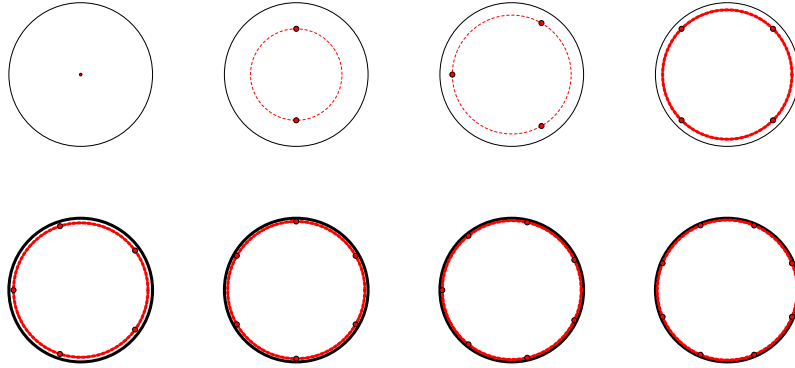


FIGURE 2. Optimal configuration of n means with respect to a uniform distribution on a circle for $1 \leq n \leq 8$.

2.2. Optimal quantization for a uniform distribution on a unit circle

Let L be the unit circle given by the parametric equations: $L := \{(x_1, x_2) : x_1 = \cos \theta, x_2 = \sin \theta \text{ for } 0 \leq \theta \leq 2\pi\}$. Let the positive direction of the x_1 -axis cut the circle at the point A , i.e., A is represented by the parametric value $\theta = 0$. Let s be the distance of a point on L along the arc starting from the point A in the counterclockwise direction. Then,

$$ds = \sqrt{\left(\frac{dx_1}{d\theta}\right)^2 + \left(\frac{dx_2}{d\theta}\right)^2} d\theta = d\theta.$$

Then, the probability density function (pdf) $f(x_1, x_2)$ for P is given by

$$f(x_1, x_2) = \begin{cases} \frac{1}{2\pi} & \text{if } (x_1, x_2) \in L, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, we have $dP(s) = P(ds) = f(x_1, x_2)ds = \frac{1}{2\pi}d\theta$. Moreover, we know that if $\hat{\theta}$ radians is the central angle subtended by an arc of length S of the unit circle, then $S = \hat{\theta}$, and

$$P(S) = \int_S dP(s) = \frac{1}{2\pi} \int_S d\theta = \frac{\hat{\theta}}{2\pi}.$$

The following theorem gives the optimal sets of n -means and the n th quantization errors for all $n \in \mathbb{N}$.

Theorem 2.2.1. *Let α_n be an optimal set of n -means for the uniform distribution P on the unit circle $x_1^2 + x_2^2 = 1$ for $n \in \mathbb{N}$. Then,*

$$\alpha_n := \left\{ \frac{n}{\pi} \left(\sin\left(\frac{\pi}{n}\right) \cdot \cos\left((2j-1)\frac{\pi}{n}\right), \sin\left(\frac{\pi}{n}\right) \cdot \sin\left((2j-1)\frac{\pi}{n}\right) \right) : j = 1, 2, \dots, n \right\}$$

forms an optimal set of n -means, and the corresponding quantization error is given by $1 - \frac{n^2}{\pi^2} \sin^2\left(\frac{\pi}{n}\right)$.

Proof. Let $\alpha_n := \{a_1, a_2, \dots, a_n\}$ be an optimal set of n -means for P for $n \in \mathbb{N}$. Let $\theta_1 - \theta_0, \theta_2 - \theta_1, \dots, \theta_n - \theta_{n-1}$ be the angles in radians subtended by the Voronoi regions of the points a_1, a_2, \dots, a_n , and let the corresponding lengths of the arcs be S_1, S_2, \dots, S_n , respectively, where $\theta_0 = 0$ and $\theta_n = 2\pi$. Then, $S_k = \theta_k - \theta_{k-1}$ for $1 \leq k \leq n$. Thus, for $1 \leq k \leq n$, we have

$$\begin{aligned} a_k &= E(X : X \in S_k) = \frac{1}{P(S_k)} \int_{S_k} (i \cos \theta + j \sin \theta) dP \\ &= \frac{2\pi}{\theta_k - \theta_{k-1}} \int_{\theta_{k-1}}^{\theta_k} \frac{1}{2\pi} (i \cos \theta + j \sin \theta) d\theta, \end{aligned}$$

which after simplification implies

$$(7) \quad a_k = \frac{1}{\theta_k - \theta_{k-1}} (\sin \theta_k - \sin \theta_{k-1}, \cos \theta_{k-1} - \cos \theta_k).$$

The n th quantization error is given by

$$\begin{aligned} V_n &= \int \min_{a \in \alpha_n} \|(\cos \theta, \sin \theta) - a\|^2 dP \\ &= \frac{1}{2\pi} \sum_{k=1}^n \int_{\theta_{k-1}}^{\theta_k} \|(\cos \theta, \sin \theta) - a_k\|^2 d\theta \\ &= \frac{1}{2\pi} \sum_{k=1}^n \int_{\theta_{k-1}}^{\theta_k} \left(\left(\cos \theta - \frac{\sin \theta_k - \sin \theta_{k-1}}{\theta_k - \theta_{k-1}} \right)^2 + \left(\sin \theta - \frac{\cos \theta_{k-1} - \cos \theta_k}{\theta_k - \theta_{k-1}} \right)^2 \right) d\theta \\ &= \frac{1}{2\pi} \sum_{k=1}^n \left(\theta_k - \theta_{k-1} + 2 \frac{(\cos(\theta_{k-1} - \theta_k) - 1)}{(\theta_k - \theta_{k-1})} \right) \\ &= \frac{1}{2\pi} \sum_{k=1}^n \left(\theta_k - \theta_{k-1} - 4 \frac{\sin^2 \frac{\theta_k - \theta_{k-1}}{2}}{\theta_k - \theta_{k-1}} \right) \\ &= 1 - \frac{2}{\pi} \sum_{k=1}^n \frac{\sin^2 \frac{\theta_k - \theta_{k-1}}{2}}{\theta_k - \theta_{k-1}}. \end{aligned}$$

The distortion error V_n being optimal, we must have $\frac{\partial V_n}{\partial \theta_k} = 0$ for all $k = 1, 2, \dots, n - 1$, yielding the fact that

$$-\frac{\sin^2 \frac{\theta_k - \theta_{k-1}}{2}}{(\theta_k - \theta_{k-1})^2} + \frac{\sin(\theta_k - \theta_{k-1})}{(\theta_k - \theta_{k-1})} + \frac{\sin^2 \frac{\theta_{k+1} - \theta_k}{2}}{(\theta_{k+1} - \theta_k)^2} - \frac{\sin(\theta_{k+1} - \theta_k)}{(\theta_{k+1} - \theta_k)} = 0$$

implying

$$(8) \quad \frac{\sin^2 \frac{\theta_k - \theta_{k-1}}{2}}{(\theta_k - \theta_{k-1})^2} - \frac{\sin(\theta_k - \theta_{k-1})}{(\theta_k - \theta_{k-1})} = \frac{\sin^2 \frac{\theta_{k+1} - \theta_k}{2}}{(\theta_{k+1} - \theta_k)^2} - \frac{\sin(\theta_{k+1} - \theta_k)}{(\theta_{k+1} - \theta_k)}$$

for all $k = 1, 2, \dots, n - 1$. From the above recurrence relation, we have

$$(9) \quad \frac{\sin^2 \frac{\theta_1 - \theta_0}{2}}{(\theta_1 - \theta_0)^2} - \frac{\sin(\theta_1 - \theta_0)}{(\theta_1 - \theta_0)} = \frac{\sin^2 \frac{\theta_k - \theta_{k-1}}{2}}{(\theta_k - \theta_{k-1})^2} - \frac{\sin(\theta_k - \theta_{k-1})}{(\theta_k - \theta_{k-1})}$$

for $k = 2, 3, \dots, n$. This yields the fact that

$$\theta_1 - \theta_0 = \theta_2 - \theta_1 = \theta_3 - \theta_2 = \dots = \theta_n - \theta_{n-1} = \frac{2\pi}{n}.$$

Thus, we have $\theta_k = \frac{2\pi k}{n}$ for $k = 1, 2, \dots, n$. Hence, by (7), we deduce that if $\alpha_n := \{a_1, a_2, \dots, a_n\}$ is an optimal set of n -means, then $a_k = \frac{n}{\pi} \left(\sin(\frac{\pi}{n}) \cdot \cos((2k-1)\frac{\pi}{n}), \sin(\frac{\pi}{n}) \cdot \sin((2k-1)\frac{\pi}{n}) \right)$ for $k = 1, 2, \dots, n$ (see Figure 2). Notice that in the optimal set of n -means, the distortion errors contributed by each point in their own Voronoi regions are equal. Thus, the quantization error for n -means is given by

$$\begin{aligned} V_n &= n \left(\text{quanization error due to the point } a_1 \text{ in the opitmal set of } n\text{-means} \right) \\ &= n \int_0^{\frac{2\pi}{n}} \left\| (\cos \theta, \sin \theta) - \left(\frac{n}{2\pi} \sin \frac{2\pi}{n}, \frac{n}{\pi} \sin^2 \frac{\pi}{n} \right) \right\|^2 dP \\ &= n \int_0^{\frac{2\pi}{n}} \frac{1}{2\pi} \left\| (\cos \theta, \sin \theta) - \left(\frac{n}{2\pi} \sin \frac{2\pi}{n}, \frac{n}{\pi} \sin^2 \frac{\pi}{n} \right) \right\|^2 d\theta, \end{aligned}$$

which after simplification yields $V_n = 1 - \frac{n^2}{\pi^2} \sin^2 \frac{\pi}{n}$. Thus, the proof of the theorem is complete. \square

Remark 2.2.2. By Theorem 2.2.1, we see that if V_n is the quantization error of n -means for the uniform distribution on the circle, then $V_n = 1 - \frac{n^2}{\pi^2} \sin^2 \frac{\pi}{n}$. Thus,

$$D(P) = \lim_{n \rightarrow \infty} \frac{2 \log n}{-\log V_n} = \lim_{n \rightarrow \infty} \frac{2 \log n}{-\log \left(1 - \frac{n^2}{\pi^2} \sin^2 \frac{\pi}{n} \right)} = 1,$$

which equals the dimension of the circle. Moreover, we have $\lim_{n \rightarrow \infty} n^2 V_n = \frac{\pi^2}{3}$, which is a finite positive number.

Proposition 2.2.3. *The points in an optimal set of n -means for the uniform distribution P on the unit circle $x_1^2 + x_2^2 = 1$ form a concentric circle of radius $\frac{n}{\pi} \sin \frac{\pi}{n}$.*

Proof. By Theorem 2.2.1, we see that if α_n is an optimal set of n -means for a positive integer n , then $\alpha_n = \{a_1, a_2, \dots, a_n\}$, where $a_k = \frac{n}{\pi} \left(\sin(\frac{\pi}{n}) \cdot \cos((2k-1)\frac{\pi}{n}), \sin(\frac{\pi}{n}) \cdot \sin((2k-1)\frac{\pi}{n}) \right)$ for $k = 1, 2, \dots, n$. The distance of the points a_k from the center $(0, 0)$ is $\frac{n}{\pi} \sin \frac{\pi}{n}$ yielding the fact that the points a_k form the circle $x_1^2 + x_2^2 = \frac{n^2}{\pi^2} \sin^2 \frac{\pi}{n}$, which is the proposition. \square

Remark 2.2.4. By Proposition 2.2.3, we know that the points in an optimal set of n -means for the uniform distribution P on the unit circle $x_1^2 + x_2^2 = 1$ lie on the circle $x_1^2 + x_2^2 = \frac{n^2}{\pi^2} \sin^2 \frac{\pi}{n}$. Notice that the radius of the circle $\frac{n}{\pi} \sin \frac{\pi}{n}$ is an increasing function of n and approaches to one as $n \rightarrow \infty$, i.e., the concentric circle formed by the points in an optimal set of n -means approaches to the circle $x_1^2 + x_2^2 = 1$ as $n \rightarrow \infty$ (see Figure 2).

2.3. Optimal quantization for a uniform distribution on the boundary of an equilateral triangle

Let P be the uniform distribution defined on the boundary L of the equilateral triangle with vertices $O(0, 0)$, $A(1, 0)$, and $B(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Let s represent the distance of any point on L from the origin tracing along the boundary of the triangle in the counterclockwise direction. Then, the points O, A, B are, respectively, represented by $s = 0, s = 1, s = 2$, and the point with $s = 3$ coincides with O . Then, the probability density function (pdf) $f(s)$ of the uniform distribution P is given by $f(s) := f(x_1, x_2) = \frac{1}{3}$ for all $(x_1, x_2) \in L$, and zero otherwise. Notice that $L = L_1 \cup L_2 \cup L_3$, where

$$\begin{aligned} L_1 &= \{(x_1, x_2) : x_1 = t, x_2 = 0 \text{ for } 0 \leq t \leq 1\}, \\ L_2 &= \{(x_1, x_2) : x_1 = t, x_2 = -\sqrt{3}(t - 1) \text{ for } \frac{1}{2} \leq t \leq 1\}, \\ L_3 &= \{(x_1, x_2) : x_1 = t, x_2 = \sqrt{3}t \text{ for } 0 \leq t \leq \frac{1}{2}\}. \end{aligned}$$

Again, $dP(s) = P(ds) = f(x_1, x_2)ds = \frac{1}{3}ds$. On L_1 , we have

$$ds = \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2} |dt| = |dt|$$

yielding $dP = \frac{1}{3}|dt|$. On L_2 and L_3 , we have $ds = \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2} |dt| = 2|dt|$ yielding $dP = \frac{2}{3}|dt|$.

Let us now prove the following lemma.

Lemma 2.3.1. *Let X be a continuous random variable with uniform distribution taking values on L . Then, $E(X) = (\frac{1}{2}, \frac{\sqrt{3}}{6})$ and $V(X) = \frac{1}{6}$.*

Proof. Recall that by (a, b) it is meant $ai + bj$, where i and j are the two unit vectors in the positive directions of x_1 - and x_2 -axes, respectively. Thus, we have,

$$\begin{aligned} E(X) &= \int_L (x_1 i + x_2 j) dP \\ &= \frac{1}{3} \int_0^1 (t, 0) dt - \frac{2}{3} \int_{\frac{1}{2}}^1 (t, -\sqrt{3}(t - 1)) dt - \frac{2}{3} \int_{\frac{1}{2}}^0 (t, \sqrt{3}t) dt \\ &= \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right), \end{aligned}$$

and

$$\begin{aligned}
 V(X) &= E\|X - E(X)\|^2 \\
 &= \frac{1}{3} \left(\int_0^1 \left((t - \frac{1}{2})^2 + (0 - \frac{\sqrt{3}}{6})^2 \right) dt - 2 \int_{\frac{1}{2}}^1 \left((t - \frac{1}{2})^2 \right. \right. \\
 &\quad \left. \left. + (-\sqrt{3}(t - 1) - \frac{\sqrt{3}}{6})^2 \right) dt - 2 \int_{\frac{1}{2}}^0 \left((t - \frac{1}{2})^2 + (\sqrt{3}t - \frac{\sqrt{3}}{6})^2 \right) dt \right) \\
 &= \frac{1}{6}.
 \end{aligned}$$

Hence, the proof of the lemma is complete. \square

Remark 2.3.2. For any $(a, b) \in \mathbb{R}^2$, we have

$$\begin{aligned}
 E\|X - (a, b)\|^2 &= \int_L \|(x_1, x_2) - (a, b)\|^2 dP \\
 &= \frac{1}{3} \int_L \left((x_1 - a)^2 + (x_2 - b)^2 \right) ds \\
 &= \frac{1}{3} \int_L \left((x_1 - \frac{1}{2})^2 + (x_2 - \frac{\sqrt{3}}{6})^2 \right) ds \\
 &\quad + \frac{1}{3} \left((\frac{1}{2} - a)^2 + (\frac{\sqrt{3}}{6} - b)^2 \right) \\
 &= V(X) + \frac{1}{3} \|(a, b) - (\frac{1}{2}, \frac{\sqrt{3}}{6})\|^2,
 \end{aligned}$$

which is minimum if $(a, b) = (\frac{1}{2}, \frac{\sqrt{3}}{6})$, and the minimum value is $V(X)$. Thus, we see that the optimal set of one-mean is the set $\{(\frac{1}{2}, \frac{\sqrt{3}}{6})\}$, and the corresponding quantization error is the variance $V := V(X)$ of the random variable X .

The following lemma gives the optimal set of two-means.

Lemma 2.3.3. *Let P be the uniform distribution defined on the boundary of the equilateral triangle with vertices $O(0, 0)$, $A(1, 0)$, and $B(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Then, if we divide the equilateral triangle into an isosceles trapezoid and an equilateral triangle in the ratio $\alpha : 1 - \alpha$, where $\alpha = \frac{1}{8}(\sqrt{17} - 1)$, then the conditional expectations of the two regions give an optimal set of two-means. One of the three such sets is given by $\{(0.314187, 0.395954), (0.771396, 0.131985)\}$, and the quantization error for two-means is $V_2 = 0.0994281$.*

Proof. Let the points P and Q with position vectors \tilde{p} and \tilde{q} form an optimal set of two-means for the uniform distribution on L . Let ℓ be the boundary of their Voronoi regions. Let ℓ intersect OA and AB at the two points D and E with position vectors \tilde{d} and \tilde{e} , respectively. Let D and E be given by the parametric values $t = \alpha$ and $t = \beta$. Then, $\tilde{d} = (\alpha, 0)$, and $\tilde{e} = (\beta, -\sqrt{3}(\beta - 1))$.

Let P be in the region which contains the point A , and Q be in the region which contains the point B . Then,

$$\tilde{p} = E(X : X \in DA \cup AE) \quad \text{and} \quad \tilde{q} = E(X : X \in EB \cup BO \cup OD),$$

yielding

$$\begin{aligned} \tilde{p} &= \frac{1}{P(DA \cup AE)} \int_{DA \cup AE} x dP \\ &= \left(\frac{-\frac{\alpha^2}{2} - 2(\frac{\beta^2}{2} - \frac{1}{2}) + \frac{1}{2}}{-\alpha - 2(\beta - 1) + 1}, -\frac{2(-\frac{1}{2}\sqrt{3}\beta^2 + \sqrt{3}\beta - \frac{\sqrt{3}}{2})}{-\alpha - 2(\beta - 1) + 1} \right) \end{aligned}$$

and

$$\begin{aligned} \tilde{q} &= \frac{1}{P(EB \cup BO \cup OD)} \int_{EB \cup BO \cup OD} x dP \\ &= \left(\frac{\frac{\alpha^2}{2} - 2(\frac{1}{8} - \frac{\beta^2}{2}) + \frac{1}{4}}{\alpha - 2(\frac{1}{2} - \beta) + 1}, \frac{\frac{\sqrt{3}}{4} - 2(\frac{\sqrt{3}\beta^2}{2} - \sqrt{3}\beta + \frac{3\sqrt{3}}{8})}{\alpha - 2(\frac{1}{2} - \beta) + 1} \right). \end{aligned}$$

Since D and E lie on ℓ , and the boundary of the Voronoi regions ℓ is the perpendicular bisector of the line segment joining P and Q , we have the canonical equations as $\rho(\tilde{d}, \tilde{p}) = \rho(\tilde{d}, \tilde{q})$ and $\rho(\tilde{e}, \tilde{p}) = \rho(\tilde{e}, \tilde{q})$ which after simplification give the following two equations:

$$\begin{aligned} &2\alpha^5 + \alpha^4(12\beta - 11) + 4\alpha^3(4\beta^2 - 10\beta + 3) + 2\alpha^2(8\beta^3 - 26\beta^2 + 28\beta - 3) \\ &+ 2\alpha(8\beta^3 - 20\beta^2 + 12\beta - 3) - 64\beta^5 + 240\beta^4 - 352\beta^3 + 252\beta^2 - 84\beta + 9 = 0, \end{aligned}$$

and

$$\begin{aligned} &-9 - 5\alpha^4 + 2\alpha^5 + 12\beta + 108\beta^2 - 288\beta^3 + 240\beta^4 - 64\beta^5 + \alpha^3(-8\beta^2 + 4\beta + 4) \\ &+ \alpha^2(-32\beta^3 + 76\beta^2 - 44\beta + 6) - 2\alpha(48\beta^4 - 144\beta^3 + 160\beta^2 - 78\beta + 15) = 0. \end{aligned}$$

Solving the above two equations in α and β , we get the following four sets of solutions:

$$\begin{aligned} &\left\{ \alpha = 0, \beta = \frac{3}{4} \right\}, \left\{ \alpha = \frac{1}{2}, \beta = \frac{1}{2} \right\}, \{ \alpha = 1, \beta = 1 \}, \text{ and} \\ &\left\{ \alpha = \frac{1}{8}(\sqrt{17} - 1), \beta = \frac{\frac{79}{8}(\sqrt{17} - 1) - 51}{65(\sqrt{17} - 1) - 232} \right\} \end{aligned}$$

Thus, we have the following results:

(i) Putting $\{\alpha = 0, \beta = \frac{3}{4}\}$, we obtain $\tilde{p} = (\frac{5}{8}, \frac{1}{8\sqrt{3}})$ and $\tilde{q} = (\frac{3}{8}, \frac{7}{8\sqrt{3}})$. Notice that in this case D coincides with the vertex O , and \tilde{p} and \tilde{q} are symmetric about the line ℓ , i.e., the boundary line ℓ of the two Voronoi regions becomes a median of the triangle passing through the vertex O , and thus, the corresponding distortion error, denoted by $V_{2,(i)}$, is given by

$$V_{2,(i)} = 2 \int_{OA \cup AE} \rho((x_1, x_2), \tilde{p}) dP$$

$$= \frac{2}{3} \left(\int_{\alpha}^1 \rho((t, 0), \tilde{p}) dt - 2 \int_1^{\beta} \rho((t, -\sqrt{3}(t-1)), \tilde{p}) dt \right)$$

yielding $V_{2,(i)} = \frac{5}{48} = 0.104167$.

(ii) Putting $\left\{ \alpha = \frac{1}{2}, \beta = \frac{1}{2} \right\}$, we obtain $\tilde{p} = \left(\frac{3}{4}, \frac{1}{2\sqrt{3}} \right)$ and $\tilde{q} = \left(\frac{1}{4}, \frac{1}{2\sqrt{3}} \right)$. Notice that in this case E coincides with the vertex B , and \tilde{p} and \tilde{q} are symmetric about the line ℓ , i.e., the boundary line ℓ of the two Voronoi regions becomes a median of the triangle passing through the vertex B , and thus, the corresponding distortion error, denoted by $V_{2,(ii)}$, is given by

$$\begin{aligned} V_{2,(ii)} &= 2 \int_{DA \cup AB} \rho((x_1, x_2), \tilde{p}) dP \\ &= \frac{2}{3} \left(\int_{\alpha}^1 \rho((t, 0), \tilde{p}) dt - 2 \int_1^{\beta} \rho((t, -\sqrt{3}(t-1)), \tilde{p}) dt \right) \end{aligned}$$

yielding $V_{2,(ii)} = \frac{5}{48} = 0.104167$.

(iii) Putting $\{\alpha = 1, \beta = 1\}$, we see that the points D and E coincides with A , which is not true.

(iv) Putting $\left\{ \alpha = \frac{1}{8}(\sqrt{17}-1), \beta = \frac{79(\sqrt{17}-1)-51}{65(\sqrt{17}-1)-232} \right\}$, we see that the points D and E , respectively, are given by $\left(\frac{1}{8}(\sqrt{17}-1), 0 \right)$ and

$$\left(\frac{\frac{79}{8}(\sqrt{17}-1)-51}{65(\sqrt{17}-1)-232}, -\sqrt{3} \left(\frac{\frac{79}{8}(\sqrt{17}-1)-51}{65(\sqrt{17}-1)-232} - 1 \right) \right).$$

Thus, in this case we have $OD = BE$ and $DA = AE = ED$ implying the fact that the boundary of the Voronoi regions ℓ divides the equilateral triangle into an isosceles trapezoid and an equilateral triangle in the ratio $\alpha : 1 - \alpha$, where $\alpha = \frac{1}{8}(\sqrt{17}-1)$. The corresponding distortion error $V_{2,(iv)}$ is given by

$$\begin{aligned} V_{2,(iv)} &= \frac{1}{3} \left(\int_{\alpha}^1 \rho((t, 0), \tilde{p}) dt - 2 \int_1^{\beta} \rho((t, -\sqrt{3}(t-1)), \tilde{p}) dt \right. \\ &\quad \left. - 2 \int_{\beta}^{\frac{1}{2}} \rho((t, -\sqrt{3}(t-1)), \tilde{q}) dt - 2 \int_{\frac{1}{2}}^0 \rho((t, \sqrt{3}t), \tilde{q}) dt \right. \\ &\quad \left. + \int_0^{\alpha} \rho((t, 0), \tilde{q}) dt \right) \\ &= \frac{1}{3} (0.0330382 + 0.0330382 + 0.0716907 + 0.0888266 + 0.0716907) \\ &= 0.0994281. \end{aligned}$$

Thus, comparing the distortion errors $V_{2,(i)}$, $V_{2,(ii)}$, and $V_{2,(iv)}$, we can say that when the boundary of the Voronoi regions ℓ divides the equilateral triangle into an isosceles trapezoid and an equilateral triangle in the ratio $\alpha : 1 - \alpha$, where $\alpha = \frac{1}{8}(\sqrt{17}-1)$, then the conditional expectations of the two Voronoi regions give an optimal set of two-means. There are three such sets; one of them is given

by $\{(0.314187, 0.395954), (0.771396, 0.131985)\}$ and the quantization error for two-means is $V_2 = 0.0994281$. Thus, the proof of the lemma is complete. \square

The following lemma gives the optimal set of three-means.

Lemma 2.3.4. *Let P be the uniform distribution defined on the boundary of the equilateral triangle with vertices $O(0, 0)$, $A(1, 0)$, and $B(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Then, the set $\{(\frac{13}{16}, \frac{\sqrt{3}}{16}), (\frac{1}{2}, \frac{3\sqrt{3}}{8}), (\frac{3}{16}, \frac{\sqrt{3}}{16})\}$ forms the optimal set of three-means, and the quantization error for three-means is $V_3 = \frac{7}{192} = 0.0364583$.*

Proof. Let the points P , Q , and R with position vectors \tilde{p} , \tilde{q} , and \tilde{r} form an optimal set of three-means for the uniform distribution on L . Let the boundaries of their Voronoi regions cut the sides OA , AB , and BO at the points D , E , and F with parametric values given by $t = \alpha$, $t = \beta$, and $t = \gamma$, respectively. Let \tilde{d} , \tilde{e} , and \tilde{f} be the position vectors of D , E , and F , respectively. Then,

$$\tilde{d} = (\alpha, 0), \tilde{e} = (\beta, -\sqrt{3}(\beta - 1)), \text{ and } \tilde{f} = (\gamma, \sqrt{3}\gamma).$$

Let $\tilde{p} = E(X : X \in DA \cup AE)$, $\tilde{q} = E(X : X \in EB \cup BF)$, and $\tilde{r} = E(X : X \in FO \cup OD)$. Then, using the definitions of conditional expectations, and after some calculations, we have

$$\begin{aligned} \tilde{p} &= \left(\frac{-\frac{\alpha^2}{2} - 2(\frac{\beta^2}{2} - \frac{1}{2}) + \frac{1}{2}}{-\alpha - 2(\beta - 1) + 1}, -\frac{2(-\frac{1}{2}\sqrt{3}\beta^2 + \sqrt{3}\beta - \frac{\sqrt{3}}{2})}{-\alpha - 2(\beta - 1) + 1} \right), \\ \tilde{q} &= \left(\frac{-2(\frac{1}{8} - \frac{\beta^2}{2}) - 2(\frac{\gamma^2}{2} - \frac{1}{8})}{-2(\frac{1}{2} - \beta) - 2(\gamma - \frac{1}{2})}, \frac{-2(\frac{\sqrt{3}\beta^2}{2} - \sqrt{3}\beta + \frac{3\sqrt{3}}{8}) - 2(\frac{\sqrt{3}\gamma^2}{2} - \frac{\sqrt{3}}{8})}{-2(\frac{1}{2} - \beta) - 2(\gamma - \frac{1}{2})} \right), \text{ and} \\ \tilde{r} &= \left(\frac{\frac{\alpha^2}{2} + \gamma^2}{\alpha + 2\gamma}, \frac{\sqrt{3}\gamma^2}{\alpha + 2\gamma} \right). \end{aligned}$$

Notice that the point D lies on the boundary of the Voronoi regions of P and R implying $\rho(\tilde{d}, \tilde{r}) = \rho(\tilde{d}, \tilde{p})$. Similarly, for the points E and F , we have $\rho(\tilde{e}, \tilde{p}) = \rho(\tilde{e}, \tilde{q})$, and $\rho(\tilde{f}, \tilde{q}) = \rho(\tilde{f}, \tilde{r})$. Thus, we have three equations in α , β , and γ . Solving the three equations, we obtain the following three sets of solutions:

$$\{\alpha = 0, \beta = 1, \gamma = \frac{1}{2}\}, \{\alpha = \frac{1}{2}, \beta = \frac{3}{4}, \gamma = \frac{1}{4}\}, \text{ and } \{\alpha = 1, \beta = \frac{1}{2}, \gamma = 0\}.$$

Thus, we obtain the following three results:

(i) For $\{\alpha = 0, \beta = 1, \gamma = \frac{1}{2}\}$, we have $\tilde{p} = (\frac{1}{2}, 0)$, $\tilde{q} = (\frac{3}{4}, \frac{\sqrt{3}}{4})$, $\tilde{r} = (\frac{1}{4}, \frac{\sqrt{3}}{4})$ with distortion error obtained as $\frac{1}{12} = 0.0833333$.

(ii) For $\{\alpha = \frac{1}{2}, \beta = \frac{3}{4}, \gamma = \frac{1}{4}\}$, we have $\tilde{p} = (\frac{13}{16}, \frac{\sqrt{3}}{16})$, $\tilde{q} = (\frac{1}{2}, \frac{3\sqrt{3}}{8})$, $\tilde{r} = (\frac{3}{16}, \frac{\sqrt{3}}{16})$ with distortion error obtained as $\frac{7}{192} = 0.0364583$.

(iii) For $\{\alpha = 1, \beta = \frac{1}{2}, \gamma = 0\}$, we have $\tilde{p} = (\frac{3}{4}, \frac{\sqrt{3}}{4})$, $\tilde{q} = (\frac{1}{4}, \frac{\sqrt{3}}{4})$, $\tilde{r} = (\frac{1}{2}, 0)$ with distortion error obtained as $\frac{1}{12} = 0.0833333$.

Comparing the distortion errors, we deduce that the set

$$\left\{ \left(\frac{13}{16}, \frac{\sqrt{3}}{16} \right), \left(\frac{1}{2}, \frac{3\sqrt{3}}{8} \right), \left(\frac{3}{16}, \frac{\sqrt{3}}{16} \right) \right\}$$

forms the optimal set of three-means (see Figure 3), and the quantization error for three-means is $V_3 = \frac{7}{192} = 0.0364583$. Thus, the lemma is yielded. \square

The following lemma gives the optimal set of four-means.

Lemma 2.3.5. *Let P be the uniform distribution defined on the boundary of the equilateral triangle with vertices $O(0, 0)$, $A(1, 0)$, and $B(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Then, the set*

$$\{(0.133784, 0.140735), (0.5, 0), (0.866216, 0.140735), (0.5, 0.653763)\}$$

forms an optimal set of four-means, and the corresponding quantization error is $V_4 = 0.028269$.

Proof. Let the points $P, Q, R,$ and S with position vectors $\tilde{p}, \tilde{q}, \tilde{r},$ and \tilde{s} form an optimal set of four-means for the uniform distribution on L . Let the boundaries of their Voronoi regions cut the side OA at two points D and E , and cut AB, BO at the points F, G , respectively. Let the position vectors of D, E, F, G be $\tilde{d}, \tilde{e}, \tilde{f}, \tilde{g}$ with parametric values given by $t = \alpha, \beta, \gamma, \delta$, respectively. Then,

$$\tilde{d} = (\alpha, 0), \tilde{e} = (\beta, 0), \tilde{f} = (\gamma, -\sqrt{3}(\gamma - 1)), \text{ and } \tilde{g} = (\delta, \sqrt{3}\delta).$$

Let $\tilde{p} = E(X : X \in GO \cup OD), \tilde{q} = E(X : X \in DE), \tilde{r} = E(X : X \in EA \cup AF),$ and $\tilde{s} = E(X : X \in FB \cup BG)$. Then, using the definitions of conditional expectations, and after some calculations, we have

$$\begin{aligned} \tilde{p} &= \left(\frac{\alpha^2 + \delta^2}{\alpha + 2\delta}, \frac{\sqrt{3}\delta^2}{\alpha + 2\delta} \right), \tilde{q} = \left(\frac{\alpha + \beta}{2}, 0 \right), \\ \tilde{r} &= \left(\frac{-\frac{\beta^2}{2} - 2(\frac{\gamma^2}{2} - \frac{1}{2}) + \frac{1}{2}}{-\beta - 2(\gamma - 1) + 1}, -\frac{2(-\frac{1}{2}\sqrt{3}\gamma^2 + \sqrt{3}\gamma - \frac{\sqrt{3}}{2})}{-\beta - 2(\gamma - 1) + 1} \right), \text{ and} \\ \tilde{s} &= \left(\frac{-2(\frac{1}{8} - \frac{\gamma^2}{2}) - 2(\frac{\delta^2}{2} - \frac{1}{8})}{-2(\frac{1}{2} - \gamma) - 2(\delta - \frac{1}{2})}, \frac{-2(\frac{\sqrt{3}\gamma^2}{2} - \sqrt{3}\gamma + \frac{3\sqrt{3}}{8}) - 2(\frac{\sqrt{3}\delta^2}{2} - \frac{\sqrt{3}}{8})}{-2(\frac{1}{2} - \gamma) - 2(\delta - \frac{1}{2})} \right). \end{aligned}$$

In this case, we obtain the following four canonical equations in $\alpha, \beta, \gamma,$ and δ :

$$\rho(\tilde{d}, \tilde{p}) = \rho(\tilde{d}, \tilde{q}), \rho(\tilde{e}, \tilde{q}) = \rho(\tilde{e}, \tilde{r}), \rho(\tilde{f}, \tilde{r}) = \rho(\tilde{f}, \tilde{s}), \text{ and } \rho(\tilde{g}, \tilde{s}) = \rho(\tilde{g}, \tilde{p}).$$

Solving the above four equations, we obtain

$$\alpha = 0.28985, \beta = 0.71015, \gamma = 0.7451, \delta = 0.2549$$

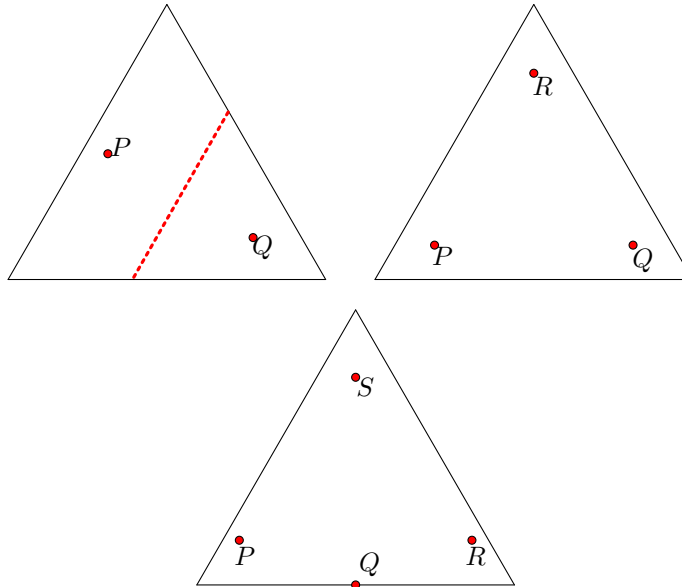


FIGURE 3. Optimal configuration of n means for $2 \leq n \leq 4$.

yielding $\tilde{p} = (0.133784, 0.140735)$, $\tilde{q} = (0.5, 0)$, $\tilde{r} = (0.866216, 0.140735)$, and $\tilde{s} = (0.5, 0.653763)$ (see Figure 3). Notice that the above four points are symmetric about the median passing through the vertex B . Thus, the corresponding quantization error is given by

$$V_4 = \frac{2}{3} \left(\int_{\frac{1}{2}}^{\beta} \rho((t, 0), \tilde{q}) dt + \int_{\beta}^1 \rho((t, 0), \tilde{r}) dt - 2 \int_1^{\gamma} \rho((t, -\sqrt{3}(t-1)), \tilde{r}) dt - 2 \int_{\gamma}^{\frac{1}{2}} \rho((t, -\sqrt{3}(t-1)), \tilde{s}) dt \right) = 0.028269.$$

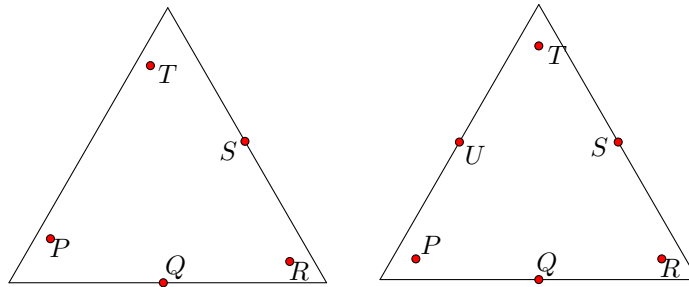
Thus, the lemma is yielded. □

Remark 2.3.6. Due to the symmetry of the equilateral triangle with respect to a rotation of $\frac{2\pi}{3}$, by Lemma 2.3.5, we conclude that there are three optimal sets of four-means (see Figure 3).

Let us now state the following two lemmas which give the optimal sets of five- and six-means. The proofs are similar to the previous lemmas.

Lemma 2.3.7. *Let P be the uniform distribution defined on the boundary of the equilateral triangle with vertices $O(0, 0)$, $A(1, 0)$, and $B(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Then, the set*

$$\{(0.130625, 0.138564), (0.485912, 0), (0.883966, 0.0669921),$$

FIGURE 4. Optimal configuration of n means for $n = 5, 6$.

$$(0.742956, 0.445213), (0.445312, 0.683619)\}$$

forms an optimal set of five-means, and the corresponding quantization error is $V_5 = 0.020525$ (see Figure 4).

Lemma 2.3.8. *Let P be the uniform distribution defined on the boundary of the equilateral triangle with vertices $O(0, 0)$, $A(1, 0)$, and $B(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Then, the set*

$$\{(0.112854, 0.0651563), (0.5, 0), (0.887146, 0.0651563), \\ (0.75, 0.433013), (0.5, 0.735713), (0.25, 0.433013)\}$$

forms an optimal set of six-means, and the corresponding quantization error is $V_6 = 0.0132077$ (see Figure 4).

Remark 2.3.9. By Lemma 2.3.4, Lemma 2.3.5, Lemma 2.3.7, and Lemma 2.3.8, we see that if α_n is an optimal set of n -means for $3 \leq n \leq 6$, then α_n contains three elements in the interior of the triangle each close to one of the vertices, and the remaining elements are on the sides of the triangle. In fact, for $n \geq 6$, if α_n is an optimal set of n -means, then it can be shown that α_n contains three elements in the interior of the triangle each close to one of the vertices, and the remaining elements are on the sides of the triangle. If α_n contains n_1, n_2 , and n_3 elements from the three sides of the triangle, where $n = n_1 + n_2 + n_3 + 3$, it can be shown that $|n_i - n_j| \leq 1$ for $1 \leq i \neq j \leq 3$. If it is not true, then the quantization error can be reduced by redistributing the points until n_1, n_2 , and n_3 satisfy $|n_i - n_j| \leq 1$ for $1 \leq i \neq j \leq 3$, and this contradicts the optimality of the set α_n . Due to technicality, we are not showing the details of the proof in the paper.

The following theorem determines the optimal sets of n -means and the n th quantization errors when $n = 3k + 3$ for some positive integer k . It also helps us to determine the quantization dimension and the quantization coefficient for the uniform distribution defined on the boundary of the equilateral triangle.

Theorem 2.3.10. *Let $n \in \mathbb{N}$ be such that $n \geq 6$, and $n = 3k + 3$ for some positive integer k . Then, the optimal set of n -means for P is given by*

$$\alpha_n = \left\{ \left(\frac{3r}{8}, \frac{\sqrt{3}r}{8} \right), \left(1 - \frac{3r}{8}, \frac{\sqrt{3}r}{8} \right), \left(\frac{1}{2}, -\frac{1}{4}\sqrt{3}(r-2) \right) \right\} \cup \gamma \cup T_1(\gamma) \cup T_2(\gamma),$$

where $\gamma := \{r + \frac{2j-1}{2k}(1-2r) : j = 1, 2, \dots, k\}$, and $r = \frac{8-2\sqrt{7}k}{16-7k^2}$, and T_1, T_2 are two affine transformations on \mathbb{R}^2 such that $T_1(0,0) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$, $T_1(1,0) = (1,0)$, $T_2(0,0) = (0,0)$, and $T_2(1,0) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$. The quantization error for n -means is given by

$$V_n = \frac{7(7k^2 - 8\sqrt{7}k + 16)}{12(16 - 7k^2)^2}.$$

Proof. Let a, b, c be the three points that α_n contains from the interior of the angles $\angle O$, $\angle A$, and $\angle B$, respectively. Again, recall that P is uniform over the boundary of the triangle, and as mentioned in Remark 2.3.9, α_n contains k elements from each side of the triangle yielding the fact that the Voronoi regions of a, b, c will form equilateral triangles P almost surely. Let the lengths of the sides of the equilateral triangles formed by the Voronoi regions of a, b, c equal r . Then,

$$\begin{aligned} a &= \frac{\int_0^r (t, 0) dt - 2 \int_{\frac{r}{2}}^0 (t, \sqrt{3}t) dt}{\int_0^r 1 dt - 2 \int_{\frac{r}{2}}^0 1 dt} = \left(\frac{3r}{8}, \frac{\sqrt{3}r}{8} \right), \\ b &= \frac{\int_{1-r}^1 (t, 0) dt - 2 \int_1^{1-\frac{r}{2}} (t, -\sqrt{3}(t-1)) dt}{\int_{1-r}^1 1 dt - 2 \int_1^{1-\frac{r}{2}} 1 dt} = \left(1 - \frac{3r}{8}, \frac{\sqrt{3}r}{8} \right), \text{ and} \\ c &= \frac{-2 \int_{\frac{1}{2}(1+r)}^{\frac{1}{2}} (t, -\sqrt{3}(t-1)) dt - 2 \int_{\frac{1}{2}}^{\frac{1}{2}(1-r)} (t, \sqrt{3}t) dt}{-2 \int_{\frac{1}{2}(1+r)}^{\frac{1}{2}} 1 dt - 2 \int_{\frac{1}{2}}^{\frac{1}{2}(1-r)} 1 dt} = \left(\frac{1}{2}, -\frac{1}{4}\sqrt{3}(r-2) \right). \end{aligned}$$

Let γ be the set of all the k points that α_n contains from the side OA . Then, the Voronoi regions of the points in γ covers the closed interval $[r, 1-r]$ yielding

$$\gamma = \left\{ \left(r + \frac{2j-1}{2k}(1-2r), 0 \right) : j = 1, 2, \dots, k \right\}.$$

Since T_1 and T_2 are affine transformations with $T_1(OA) = AB$, and $T_2(OA) = OB$, we have the points that α_n contains from AB and OB are, respectively, $T_1(\gamma)$ and $T_2(\gamma)$ yielding

$$\alpha_n = \left\{ \left(\frac{3r}{8}, \frac{\sqrt{3}r}{8} \right), \left(1 - \frac{3r}{8}, \frac{\sqrt{3}r}{8} \right), \left(\frac{1}{2}, -\frac{1}{4}\sqrt{3}(r-2) \right) \right\} \cup \gamma \cup T_1(\gamma) \cup T_2(\gamma).$$

Using the symmetry, the quantization error for n -means is obtained as

$$\begin{aligned} V_n &= 3(\text{quantization error due to the points } a \text{ and the } k \text{ points in } \gamma) \\ &= 3 \left(-\frac{2}{3} \int_{\frac{r}{2}}^0 \rho((t, \sqrt{3}t), a) dt + \frac{1}{3} \int_0^r \rho((t, 0), a) dt \right) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{3}k \int_r^{r+\frac{1-2r}{k}} \rho\left((t,0), \left(r + \frac{1}{2k}(1-2r), 0\right)\right) dt \\
& = \frac{1}{24} \left(7r^3 - \frac{2(2r-1)^3}{k^2} \right).
\end{aligned}$$

Notice that for a given k , the quantization error V_n is a function of r . Solving $\frac{\partial V_n}{\partial r} = 0$, we have $r = \frac{8-2\sqrt{7k}}{16-7k^2}$. Putting $r = \frac{8-2\sqrt{7k}}{16-7k^2}$, we have

$$V_n = \frac{7(7k^2 - 8\sqrt{7k} + 16)}{12(16 - 7k^2)^2}.$$

Thus, the proof the theorem is complete. \square

Remark 2.3.11. When $n = 6$, i.e., when $k = 1$, then Theorem 2.3.10 reduces to Lemma 2.3.8. Using a similar technique as in the proof of Theorem 2.3.10, we can also determine the optimal sets of n -means and the n th quantization errors when $n = 3k + 3 + 1$, or $n = 3k + 3 + 2$ for some positive integer k .

Proposition 2.3.12. *Quantization dimension $D(P)$ of the uniform distribution P defined on the boundary of the equilateral triangle equals the dimension of the boundary of the triangle. Moreover, the quantization coefficient exists as a finite positive number which equals $\frac{3}{4}$.*

Proof. For $n \in \mathbb{N}$, $n \geq 6$, let $\ell(n)$ be the unique positive real number such that $3\ell(n) + 3 \leq n < 3(\ell(n) + 1) + 3$. Then, $V_{3(\ell(n)+1)+3} < V_n \leq V_{3\ell(n)+3}$ implying

$$(10) \quad \frac{2 \log(3\ell(n) + 3)}{-\log V_{3(\ell(n)+1)+3}} < \frac{2 \log n}{-\log V_n} < \frac{2 \log(3(\ell(n) + 1) + 3)}{-\log V_{3\ell(n)+3}}.$$

Notice that

$$\lim_{n \rightarrow \infty} \frac{2 \log(3\ell(n) + 3)}{-\log V_{3(\ell(n)+1)+3}} = \lim_{\ell(n) \rightarrow \infty} \frac{2 \log(3\ell(n) + 3)}{-\log \left(\frac{7(7(\ell(n)+1)^2 - 8\sqrt{7}(\ell(n)+1) + 16)}{12(16 - 7(\ell(n)+1)^2)^2} \right)} = 1,$$

and

$$\lim_{n \rightarrow \infty} \frac{2 \log(3(\ell(n) + 1) + 3)}{-\log V_{3\ell(n)+3}} = \lim_{\ell(n) \rightarrow \infty} \frac{2 \log(3(\ell(n) + 1) + 3)}{-\log \left(\frac{7(7\ell(n)^2 - 8\sqrt{7}\ell(n) + 16)}{12(16 - 7\ell(n)^2)^2} \right)} = 1$$

and hence, by (10), $\lim_{n \rightarrow \infty} \frac{2 \log n}{-\log V_n} = 1$ which is the dimension of the underlying space. Again,

$$(11) \quad (3\ell(n) + 3)^2 V_{3(\ell(n)+1)+3} < n^2 V_n < (3(\ell(n) + 1) + 3)^2 V_{3\ell(n)+3}.$$

We have

$$\begin{aligned}
& \lim_{n \rightarrow \infty} (3\ell(n) + 3)^2 V_{3(\ell(n)+1)+3} \\
& = \lim_{\ell(n) \rightarrow \infty} (3\ell(n) + 3)^2 \frac{7(7(\ell(n) + 1)^2 - 8\sqrt{7}(\ell(n) + 1) + 16)}{12(16 - 7(\ell(n) + 1)^2)^2} = \frac{3}{4},
\end{aligned}$$

and

$$\begin{aligned} & \lim_{n \rightarrow \infty} (3(\ell(n) + 1) + 3)^2 V_{3\ell(n)+3} \\ &= \lim_{\ell(n) \rightarrow \infty} (3(\ell(n) + 1) + 3)^2 \frac{7(7\ell(n)^2 - 8\sqrt{7}\ell(n) + 16)}{12(16 - 7\ell(n)^2)^2} = \frac{3}{4}, \end{aligned}$$

and hence, by (11), we have $\lim_{n \rightarrow \infty} n^2 V_n = \frac{3}{4}$, i.e., the quantization coefficient exists as a finite positive number which equals $\frac{3}{4}$. Thus, the proof of the proposition is complete. \square

Remark 2.3.13. By Remark 2.3.9, we see that an optimal set α_n of n -means for $n \geq 3$ always contains three elements in the interior of the triangle each close to one of the vertices, and the remaining elements are on the sides of the triangle. It can be shown that as n approaches to infinity each of the three elements which are in the interior of the triangle approaches to their closest vertices.

Let us now end the paper with the following remark.

Remark 2.4. We know that an affine transformation T on \mathbb{R}^2 is given by

$$T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix},$$

where $A := \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is an invertible matrix, and $e, f \in \mathbb{R}$. Once the optimal sets of n -means are known for a line segment, circle, or an equilateral triangle, by giving an affine transformation, one can easily obtain them for any line segment, circle, or an equilateral triangle. For example, by Theorem 2.1.1, we see that the optimal set of n -means with respect to the uniform distribution on the line segment $[0, 1]$ is given by $\alpha_n := \{\frac{2i-1}{2n} : 1 \leq i \leq n\}$, and the corresponding quantization error is $V_n := V_n(P) = \frac{1}{12n^2}$. The optimal set of n -means for the uniform distribution on the line segment joining $(0, 0)$ and $(1, \sqrt{3})$ is given by $\alpha_n := \{(\frac{2i-1}{2n}, \frac{\sqrt{3}(2i-1)}{2n}) : 1 \leq i \leq n\}$, and the corresponding quantization error is $V_n := V_n(P) = \frac{1}{3n^2}$.

References

- [1] E. F. Abaya and G. L. Wise, *Some remarks on the existence of optimal quantizers*, *Statist. Probab. Lett.* **2** (1984), no. 6, 349–351. [https://doi.org/10.1016/0167-7152\(84\)90045-2](https://doi.org/10.1016/0167-7152(84)90045-2)
- [2] J. A. Bucklew and G. L. Wise, *Multidimensional asymptotic quantization theory with r th power distortion measures*, *IEEE Trans. Inform. Theory* **28** (1982), no. 2, 239–247. <https://doi.org/10.1109/TIT.1982.1056486>
- [3] C. P. Dettmann and M. K. Roychowdhury, *Quantization for uniform distributions on equilateral triangles*, *Real Anal. Exchange* **42** (2017), no. 1, 149–166. <http://projecteuclid.org/euclid.rae/1490580015>
- [4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic publishers, Boston, 1992.

- [5] R. M. Gray, J. C. Kieffer, and Y. Linde, *Locally optimal block quantizer design*, Inform. and Control **45** (1980), no. 2, 178–198. [https://doi.org/10.1016/S0019-9958\(80\)90313-7](https://doi.org/10.1016/S0019-9958(80)90313-7)
- [6] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics, 1730, Springer-Verlag, Berlin, 2000. <https://doi.org/10.1007/BFb0103945>
- [7] R. M. Gray and D. L. Neuhoff, *Quantization*, IEEE Trans. Inform. Theory **44** (1998), no. 6, 2325–2383. <https://doi.org/10.1109/18.720541>
- [8] A. György and T. Linder, *On the structure of optimal entropy-constrained scalar quantizers*, IEEE Trans. Inform. Theory **48** (2002), no. 2, 416–427. <https://doi.org/10.1109/18.978755>
- [9] S. Matsuura, H. Kurata, and T. Tarpey, *Optimal estimators of principal points for minimizing expected mean squared distance*, J. Statist. Plann. Inference **167** (2015), 102–122. <https://doi.org/10.1016/j.jspi.2015.05.005>
- [10] J. Rosenblatt and M. K. Roychowdhury, *Optimal quantization for piecewise uniform distributions*, Unif. Distrib. Theory **13** (2018), no. 2, 23–55.
- [11] M. K. Roychowdhury, *Optimal quantizers for some absolutely continuous probability measures*, Real Anal. Exchange **43** (2018), no. 1, 105–136. <https://doi.org/10.14321/realanalexch.43.1.0105>
- [12] R. Zam, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*, Cambridge University Press, 2014.

JOSEPH ROSENBLATT
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
1409 W. GREEN STREET
URBANA, ILLINOIS 61801-2907, USA
Email address: rosnbltt@illinois.edu

MRINAL KANTI ROYCHOWDHURY
SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES
UNIVERSITY OF TEXAS RIO GRANDE VALLEY
1201 WEST UNIVERSITY DRIVE
EDINBURG, TX 78539, USA
Email address: mrinal.roychowdhury@utrgv.edu