OPTIMAL QUANTIZATION FOR DISCRETE DISTRIBUTIONS

Russel Cabasag  
_The University of Texas Rio Grande Valley_

Samir Huq  
_The University of Texas Rio Grande Valley_

Eric Mendoza  
_The University of Texas Rio Grande Valley_

Mrinal Kanti Roychowdhury  
_The University of Texas Rio Grande Valley_

Follow this and additional works at: [https://scholarworks.utrgv.edu/mss_fac](https://scholarworks.utrgv.edu/mss_fac)
OPTIMAL QUANTIZATION FOR DISCRETE DISTRIBUTIONS

1RUSSEL CABASAG, 2SAMIR HUQ, 3ERIC MENDOZA, AND 4MRINAL KANTI ROYCHOWDHURY

ABSTRACT. In this paper, we first determine the optimal sets of $n$-means and the $n$th quantization errors for all $1 \leq n \leq 6$ for two nonuniform discrete distributions with support the set \{1, 2, 3, 4, 5, 6\}. Then, for a probability distribution $P$ with support \{\frac{1}{n} : n \in \mathbb{N}\} associated with a mass function $f$, given by $f(x) = \frac{1}{2^k}$ if $x = \frac{1}{k}$ for $k \in \mathbb{N}$, and zero otherwise, we determine the optimal sets of $n$-means and the $n$th quantization errors for all positive integers up to $n = 300$. Further, for a probability distribution $P$ with support the set $\mathbb{N}$ of natural number associated with a mass function $f$, given by $f(x) = \frac{1}{2^k}$ if $x = k$ for $k \in \mathbb{N}$, and zero otherwise, we determine the optimal sets of $n$-means and the $n$th quantization errors for all positive integers $n$. At last we discuss for a discrete distribution, if the optimal sets are given, how to obtain the probability distributions.

1. INTRODUCTION

Quantization is the process of converting a continuous analog signal into a digital signal of $k$ discrete levels, or converting a digital signal of $n$ levels into another digital signal of $k$ levels, where $k < n$. It is essential when analog quantities are represented, processed, stored, or transmitted by a digital system, or when data compression is required. It is a classic and still very active research topic in source coding and information theory. It has broad applications in engineering and technology (see [GG, GN, Z]). For mathematical treatment of quantization one is referred to Graf-Luschgy’s book (see [GL]). Let $\mathbb{R}^d$ denote the $d$-dimensional Euclidean space, $\| \cdot \|$ denote the Euclidean norm on $\mathbb{R}^d$ for any $d \geq 1$, and $n \in \mathbb{N}$. Let $P$ denote a Borel probability measure on $\mathbb{R}^d$. 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 $\alpha$, and is denoted by $V(P; \alpha)$. For any positive integer $n$, write $V_n := V_n(P) = \inf \{ V(P; \alpha) : \alpha \subset \mathbb{R}^d, 1 \leq \text{card}(\alpha) \leq n \}$. Then, $V_n$ is called the $n$th quantization error for $P$. Recently, optimal quantization for different uniform distributions have been investigated by several authors, for example, see [DR, RR, RS].

In this paper, we investigate the optimal quantization for finite, and infinite discrete distributions. In Section 3 we calculate the optimal sets of $n$-means and the $n$th quantization errors for all $1 \leq n \leq 6$ for two nonuniform discrete distributions with support \{1, 2, 3, 4, 5, 6\} associated with two different probability vectors. In Section 4 first, for a probability distribution $P$ with support \{\frac{1}{n} : n \in \mathbb{N}\} associated with a mass function $f$, given by $f(x) = \frac{1}{2^k}$ if $x = \frac{1}{k}$ for $k \in \mathbb{N}$, and zero otherwise, we determine the optimal sets of $n$-means and the $n$th quantization errors for all positive integers up to $n = 300$. Then, for a probability distribution $P$ with support the set $\mathbb{N}$ of natural number associated with a mass function $f$, given by $f(x) = \frac{1}{2^k}$ if $x = k$ for $k \in \mathbb{N}$, and zero otherwise, we determine the optimal sets of $n$-means and the $n$th quantization errors for all positive integers $n$. In Section 5 we discuss for a discrete distribution, if the optimal sets are given, how to obtain the probability distributions.

2. Basic Preliminaries

Given a finite set $\alpha \subset \mathbb{R}^d$, the Voronoi region generated by $a \in \alpha$ is defined by

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

2010 Mathematics Subject Classification. 60Exx, 94A34.

Key words and phrases. Discrete distribution, optimal set of $n$-means, quantization error.
i.e., the Voronoi region generated by \( a \in \alpha \) is the set of all elements in \( \mathbb{R}^d \) which are nearest to \( a \), and the set \( \{ M(a|\alpha) : a \in \alpha \} \) is called the Voronoi diagram or Voronoi tessellation of \( \mathbb{R}^d \) with respect to \( \alpha \).

The following proposition is well-known (see [GG, GL]).

**Proposition 2.1.** Let \( \alpha \) be an optimal set of \( n \)-means for \( P \), and \( a \in \alpha \). Then,

(i) \( P(M(a|\alpha)) > 0 \), (ii) \( P(\partial M(a|\alpha)) = 0 \), (iii) \( a = E(X : X \in M(a|\alpha)) \),

where \( X \) is a random variable with distribution \( P \).

Due to the above proposition, we see that if \( \alpha \) 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 \( a \). In the sequel, we will denote the support of a probability distribution \( P \) by \( \text{supp}(P) \). Let \( P \) be the uniform distribution defined on the set \( \{1, 2, 3, 4, 5, 6\} \). Then, the random variable \( X \) associated with the probability distribution is a discrete random variable with probability mass function \( f \) given by

\[
f(x) = \frac{1}{6}, \text{ for all } x \in \{1, 2, 3, 4, 5, 6\}.
\]

It is not difficult to show that if \( \alpha_n \) is an optimal set of \( n \)-means for \( P \), then

\[
\alpha_1 = \{3.5\}, \quad \alpha_2 = \{2, 5\}, \quad \alpha_3 = \{1.5, 3.5, 5.5\}, \quad \alpha_4 = \{1.5, 3.5, 5, 6\},
\]

\[
\alpha_5 = \{1.5, 3, 4, 5, 6\}, \quad \text{and} \quad \alpha_6 = \text{supp}(P).
\]

**Remark 2.2.** Optimal sets are not unique. For example, in the above, the set \( \alpha_5 \) can be any one of the following sets:

\[
\{1.5, 3, 4, 5, 6\}, \quad \{1, 2.5, 4, 5, 6\}, \quad \{1, 2, 3.5, 5, 6\}, \quad \{1, 2, 3, 4.5, 6\}, \quad \{1, 2, 3, 4, 5.5\}.
\]

In the following sections we give our main results.

3. **Optimal quantization for nonuniform discrete distributions**

In this section, we determine the optimal sets of \( n \)-means for all \( 1 \leq n \leq 6 \) for two nonuniform discrete distributions on the set \( \{1, 2, 3, 4, 5, 6\} \) associated with two different probability vectors. Let \( X \) be the random variable associated with such a distribution. For \( i, j \in \{1, 2, \ldots, 6\} \) with \( i \leq j \), write \( a[i, j] := E(X : X \in \{i, i + 1, \ldots, j\}) \). We give our results in two subsections.

3.1. **Nonuniform distribution associated with the probability vector** \((\frac{1}{2}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27})\).

Let \( P \) be a nonuniform distribution defined on the set \( \{1, 2, 3, 4, 5, 6\} \) with probability mass function \( f \) given by

\[
f(j) = P(X : X = j) = \begin{cases} 
\frac{1}{27} & \text{if } j \in \{1, 2, 3, 4, 5\}, \\
\frac{1}{2} & \text{if } j = 6, \\
0 & \text{otherwise.}
\end{cases}
\]

Notice that \( \text{supp}(P) = \{1, 2, 3, 4, 5, 6\} \). In this subsection, our goal is to calculate the optimal sets \( \alpha_n \) of \( n \)-means and the \( n \)th quantization errors \( V_n \) for all \( n = 1, 2, 3, 4, 5, 6 \). Since

\[
E(X) = \sum_{j=1}^{6} jf(j) = \frac{63}{32},
\]

the optimal set of one-mean is the set \( \{\frac{63}{32}\} \) with quantization error the variance \( V \) of the random variable \( X \), where

\[
V = V_1 = E\|X - E(X)\|^2 = \sum_{j=1}^{6} f(j)(j - \frac{63}{32})^2 = \frac{1695}{1024}.
\]

Moreover, the optimal set \( \alpha_6 \) of six-means is just the support of \( P \), i.e., \( \alpha_6 = \{1, 2, \ldots, 6\} \). In the following propositions, we determine the optimal sets of \( n \)-means for \( 2 \leq n \leq 5 \).
Proposition 3.1.1. The optimal set of two-means is given by \( \{a[1, 2], a[3, 6]\} \) with quantization error \( V_2 = \frac{341}{768} \).

Proof. Notice that \( a[1, 2] = \frac{4}{3} \) and \( a[3, 6] = \frac{31}{8} \). Let us consider the set \( \beta := \{\frac{4}{3}, \frac{31}{8}\} \). Since \( 2 < \frac{1}{2}(\frac{4}{3} + \frac{31}{8}) = 2.60417 < 3 \), the distortion error due to the set \( \beta \) is given by

\[
\sum_{j=1}^{6} f(j) \min_{a \in \beta} (j-a)^2 = \sum_{j=1}^{2} f(j) \left( j - \frac{4}{3} \right)^2 + \sum_{j=3}^{6} f(j) \left( j - \frac{31}{8} \right)^2 = \frac{341}{768}.
\]

Since \( V_2 \) is the quantization error for two-means, we have \( V_2 \leq \frac{341}{768} = 0.44401 \). Let \( \alpha := \{a_1, a_2\} \) be an optimal set of two-means. Without any loss of generality, we can assume that \( 1 \leq a_1 < a_2 \leq 6 \). Notice that the Voronoi region of \( a_1 \) must contain 1. Suppose that the Voronoi region of \( a_1 \) contains only the point 1. Then, the Voronoi region of \( a_2 \) contains all the remaining points, and so

\[ a_2 = a[2, 6] = \frac{47}{16}, \]

implying

\[ V_2 = \sum_{j=2}^{6} f(j) \left( j - \frac{47}{16} \right)^2 = \frac{367}{512} = 0.716797 > V_2, \]

which gives a contradiction. Hence, we can assume that the Voronoi region of \( a_1 \) does not contain 3. Next, suppose that the Voronoi region of \( a_1 \) contains only the point 1. Then, the Voronoi region of \( a_2 \) contains all the remaining points, and so

\[ a_2 = a[2, 6] = \frac{47}{16}, \]

implying

\[ V_2 = \sum_{j=2}^{6} f(j) \left( j - \frac{47}{16} \right)^2 = \frac{367}{512} = 0.716797 > V_2, \]

which yields a contradiction. Hence, we can assume that the Voronoi region of \( a_1 \) contains only the points 1 and 2, and the remaining points are contained in the Voronoi region of \( a_2 \), implying

\[ a_1 = a[1, 2] = \frac{4}{3}, \text{ and } a_2 = a[3, 6] = \frac{31}{8} \]

with quantization error \( V_2 = \frac{341}{768} \). Thus, the proof of the proposition is complete. \( \Box \)

Proposition 3.1.2. The optimal set of three-means is given by \( \{1, a[2, 3], a[4, 6]\} \) with quantization error \( V_3 = \frac{65}{384} \).

Proof. Notice that \( a[2, 3] = \frac{7}{3} \), and \( a[4, 6] = \frac{19}{4} \). The distortion error due to the set \( \beta := \{1, \frac{7}{3}, \frac{19}{4}\} \) is given by

\[
\sum_{j=1}^{6} f(j) \min_{a \in \beta} (j-a)^2 = \sum_{j=2}^{3} f(j) \left( j - \frac{7}{3} \right)^2 + \sum_{j=4}^{6} f(j) \left( j - \frac{19}{4} \right)^2 = \frac{65}{384}.
\]

Since \( V_3 \) is the quantization error for three-means, we have \( V_3 \leq \frac{65}{384} = 0.169271 \). Let \( \alpha := \{a_1, a_2, a_3\} \) be an optimal set of three-means such that \( 1 \leq a_1 < a_2 < a_3 \leq 6 \). Notice that the Voronoi region of \( a_1 \) must contain 1. Suppose that the Voronoi region of \( a_1 \) also contains 3. Then,

\[ V_3 \geq \sum_{j=1}^{3} f(j)(j-a[1, 3])^2 = \frac{13}{28} > V_3, \]

which yields a contradiction. Thus, we can assume that the Voronoi region of \( a_1 \) does not contain 3. Suppose that the Voronoi region of \( a_1 \) contains only the two points 1 and 2. Then, the Voronoi region of \( a_2 \) must contain 3. The following two case can arise:

Case 1. The Voronoi region of \( a_2 \) does not contain 4.
Proposition 3.1.3. The optimal set of four-means is 

\[ V_{a} \] two points 2 and 3, implying the fact that the Voronoi region of 6. Thus, we have 

\[
\left\{ \text{point 1 only, i.e., } a \right\},
\]

which is a contradiction.

Case 2. The Voronoi region of \( a \) contains 4.

Then,

\[
V_{3} \geq \sum_{j=1}^{2} f(j)(j - a_{1,2})^{2} + \sum_{j=4}^{6} f(j)(j - a_{4,6})^{2} = \frac{97}{384} = 0.252604 > V_{3},
\]

which is a contradiction. Hence, we can assume that the Voronoi region of \( a \) does not contain 4. Suppose that the Voronoi region of \( a \) contains only the point 2, i.e., \( a_{1} = 1 \). Then, the Voronoi region of \( a \) must contain 2. Suppose that the Voronoi region of \( a \) also contains 4. Then,

\[
V_{3} \geq \sum_{j=2}^{4} f(j)(j - a_{2,4})^{2} = \frac{13}{56} = 0.232143 > V_{3},
\]

which leads to a contradiction. Hence, by Case 1 and Case 2, we can assume that the Voronoi region of \( a \) contains only the point 1, i.e., \( a_{1} = 1 \). Then, the Voronoi region of \( a \) contains only the point 2. Then, the Voronoi region of \( a \) must contain the remaining points, which yields

\[
V_{3} \geq \sum_{j=3}^{6} f(j)(j - a_{3,6})^{2} = \frac{71}{256} = 0.277344 > V_{3},
\]

which is a contradiction. Hence, we can assume that the Voronoi region of \( a \) contains only the two points 2 and 3, implying the fact that the Voronoi region of \( a \) contains the points 4, 5, and 6. Thus, we have

\[
a_{1} = 1, a_{2} = a_{2,3} = \frac{7}{3}, \text{ and } a_{3} = a_{4,6} = \frac{19}{4},
\]

which yields the proposition. □

Proposition 3.1.3. The optimal set of four-means is \( \{1, 2, a_{3,4}, a_{5,6} \} \) with quantization error \( V_{4} = \frac{11}{192} \).

Proof. The distortion error due to the set \( \beta := \{1, 2, a_{3,4}, a_{5,6} \} \) is given by

\[
\sum_{j=1}^{6} f(j) \min_{a \in \beta} (j - a)^{2} = \sum_{j=3}^{4} f(j)(j - a_{3,4})^{2} + \sum_{j=5}^{6} f(j)(j - a_{5,6})^{2} = \frac{11}{192}.
\]

Since \( V_{4} \) is the quantization error for four-means, we have \( V_{4} \leq \frac{11}{192} = 0.0572917 \). Let \( \alpha := \{a_{1}, a_{2}, a_{3}, a_{4} \} \) be an optimal set of four-means. Without any loss of generality, we can assume that \( 1 \leq a_{1} < a_{2} < a_{3} < a_{4} \leq 6 \). The Voronoi region of \( a_{1} \) must contain 1. Suppose that the Voronoi region of \( a_{1} \) contains 2 as well. Then,

\[
V_{4} \geq \sum_{j=1}^{2} f(j)(j - a_{1,2})^{2} = \frac{1}{6} > V_{4},
\]

which gives a contradiction. Hence, we can assume that the Voronoi region of \( a_{1} \) contains the point 1 only, i.e., \( a_{1} = 1 \). Then, the Voronoi region of \( a_{2} \) must contain 2. Suppose that the Voronoi region of \( a_{2} \) also contains 3. Then,

\[
V_{4} \geq \sum_{j=2}^{3} f(j)(j - a_{2,3})^{2} = \frac{1}{12} = 0.0833333 > V_{4},
\]
which leads to a contradiction. Hence, the Voronoi region of \( a_2 \) does not contain 3, i.e., \( a_2 = 2 \). Then, the Voronoi region of \( a_3 \) must contain 3. Suppose that the Voronoi region of \( a_3 \) contains 5 as well. Then, we have

\[
V_4 \geq \sum_{j=3}^{5} f(j)(j - a[3, 5])^2 = \frac{13}{172} = 0.116071 > V_4,
\]

which yields a contradiction. Thus, we can assume that the Voronoi region of \( a_3 \) does not contain 5. Suppose that the Voronoi region of \( a_3 \) contains 3 only. Then, the Voronoi region of \( a_5 \) contains 4, 5, 6, which implies

\[
V_4 = \sum_{j=4}^{6} f(j)(j - a[4, 6])^2 = \frac{11}{128} = 0.0859375 > V_4,
\]

which gives a contradiction. Hence, the Voronoi region of \( a_3 \) contains 3 and 4, yielding \( a_3 = a[3, 4] \), and \( a_4 = a[5, 6] \). Thus, the optimal set of four-means is \( \{1, 2, a[3, 4], a[5, 6]\} \) with quantization error \( V_4 = \frac{11}{192} \). which is the proposition. 

Using the similar technique as the previous proposition, the following proposition can be proved.

**Proposition 3.1.4.** The optimal set of five-means is \( \{1, 2, 3, 4, a[5, 6]\} \) with quantization error \( V_5 = \frac{1}{64} \).

3.2. Nonuniform distribution associated with a probability vector of the form \((x, (1-x)x, (1-x)^2x, (1-x)^3x, (1-x)^4x, (1-x)^5)\). Let \( P \) be a nonuniform distribution defined on the set \( \{1, 2, 3, 4, 5, 6\} \) with probability mass function \( f \) given by

\[
f(j) = P(X : X = j) = \begin{cases} 
    x & \text{if } j = 1, \\
    (1-x)^{j-1}x & \text{if } j \in \{2, 3, 4, 5\}, \\
    (1-x)^5 & \text{if } j = 6, \\
    0 & \text{otherwise},
\end{cases}
\]

where \( 0 < x < 1 \). Notice that \( \text{supp}(P) = \{1, 2, 3, 4, 5, 6\} \). Fix \( x = \frac{7}{10} \). In this subsection, our goal is to calculate the optimal sets \( \alpha_n \) of \( n \)-means and the \( n \)th quantization errors for all \( n = 1, 2, 3, 4, 5, 6 \) for the given mass function \( f \) with \( x = \frac{7}{10} \). Since

\[
E(X) = \sum_{j=1}^{6} jf(j) = \frac{142753}{100000},
\]

the optimal set of one-mean is the set \( \{\frac{142753}{100000}\} \) with quantization error the variance \( V \) of the random variable \( X \), where

\[
V = V_1 = E\|X - E(X)\|^2 = \sum_{j=1}^{6} f(j)\left(j - \frac{142753}{100000}\right)^2 = \frac{6007880991}{10000000000}.
\]

Moreover, the optimal set \( \alpha_6 \) of six-means is just the support of \( P \), i.e., \( \alpha_6 = \{1, 2, \cdots, 6\} \). In the following propositions, we determine the optimal sets of \( n \)-means for \( 2 \leq n \leq 5 \).

**Proposition 3.2.1.** The optimal set of two-means is given by \( \{1, a[2, 6]\} \) with quantization error \( V_2 = \frac{174296997}{10000000000} \).

**Proof.** The distortion error due to the set \( \beta := \{1, a[2, 6]\} \) is given by

\[
\sum_{2=1}^{6} f(j)\min_{a \in \beta}(j - a)^2 = \sum_{j=2}^{6} f(j)(j - a[2, 6])^2 = \frac{174296997}{10000000000}.
\]
Since $V_2$ is the quantization error for two-means, we have $V_2 \leq \frac{174296997}{1000000000} = 0.174296997$. Let \( \alpha := \{a_1, a_2\} \) be an optimal set of two-means. Without any loss of generality, we can assume that $1 \leq a_1 < a_2 \leq 6$. Notice that the Voronoi region of $a_1$ must contain 1. Suppose that the Voronoi region of $a_1$ contains 3. Then, the point 1, i.e., $1 \leq a_2$ contains 3. Next, suppose that the Voronoi region of $a_1$ contains 2. Then, the Voronoi region of $a_2$ contains all the remaining points, and so

$$ V_2 = \sum_{j=1}^{2} f(j)(j - a[1, 2])^2 + \sum_{j=3}^{6} f(j)(j - a[3, 6])^2 = \frac{272139987}{1300000000} = 0.209338 > V_2, $$

which yields a contradiction. Hence, we can assume that the Voronoi region of $a_1$ does not contain 3. Next, suppose that the Voronoi region of $a_1$ contains 2. Then, the Voronoi region of $a_2$ contains all the remaining points, and so

$$ V_2 = \sum_{j=1}^{3} f(j)(j - a[1, 3])^2 = \frac{4809}{139000} = 0.345971 > V_2, $$

which gives a contradiction. Hence, we can assume that the Voronoi region of $a_1$ contains only the point 1, and the remaining points are contained in the Voronoi region of $a_2$, implying $a_1 = 1$, and $a_2 = a[2, 6]$ with quantization error $V_2 = \frac{174296997}{1000000000}$. Thus, the proof of the proposition is complete.

**Proposition 3.2.2.** The optimal set of three-means is given by $\{1, 2, a[3, 6]\}$ with quantization error $V_3 = \frac{4779999}{100000000}$. 

**Proof.** The distortion error due to the set $\beta := \{1, 2, a[3, 6]\}$ is given by

$$ \sum_{j=3}^{6} f(j) \min_{a \in \beta} (j - a)^2 = \sum_{j=3}^{6} f(j)(j - a[3, 6])^2 = \frac{4779999}{100000000} = 0.04779999. $$

Since $V_3$ is the quantization error for three-means, we have $V_3 \leq 0.04779999$. Let $\alpha := \{a_1, a_2, a_3\}$ be an optimal set of three-means such that $1 \leq a_1 < a_2 < a_3 \leq 6$. Notice that the Voronoi region of $a_1$ must contain 1. Suppose that the Voronoi region of $a_1$ also contains 2. Then,

$$ V_3 = \sum_{j=1}^{2} f(j)(j - a[1, 2])^2 = \frac{21}{130} = 0.161538 > V_3, $$

which yields a contradiction. Thus, we can assume that the Voronoi region of $a_1$ contains only the point 1, i.e., $a_1 = 1$. The Voronoi region of $a_2$ contains 2. Suppose that the Voronoi region of $a_2$ also contains 3. Then,

$$ V_3 = \sum_{j=2}^{3} f(j)(j - a[2, 3])^2 = \frac{63}{1300} = 0.0484615 > V_3, $$

which is a contradiction. Hence, the Voronoi region of $a_2$ contains only the point 2, which yields $a_2 = 2$, and $a_3 = a[3, 6]$, with quantization error $V_3 = \frac{4779999}{100000000}$. Thus, the proof of the proposition is complete.

Following the similar techniques as given in Proposition 3.2.2, we can prove the following two propositions.

**Proposition 3.2.3.** The optimal set of four-means is given by $\{1, 2, 3, a[4, 6]\}$ with quantization error $V_4 = \frac{112833}{100000000}$. 

**Proposition 3.2.4.** The optimal set of five-means is given by $\{1, 2, 3, 4, a[5, 6]\}$ with quantization error $V_5 = \frac{1701}{1000000}$. 

4. Optimal quantization for infinite discrete distributions

In this section, for \( n \in \mathbb{N} \), we investigate the optimal sets of \( n \)-means for two different infinite discrete distributions. We give them in the following two subsections.

4.1. Optimal quantization for an infinite discrete distribution with support \( \{ \frac{1}{n} : n \in \mathbb{N} \} \). Let \( \mathbb{N} := \{1, 2, 3, \cdots \} \) be the set of natural numbers, and let \( P \) be a Borel probability measure on the set \( \{ \frac{1}{n} : n \in \mathbb{N} \} \) with probability mass function \( f \) given by

\[
    f(x) = \begin{cases} \frac{1}{n} & \text{if } x = \frac{1}{k} \text{ for } k \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}
\]

Then, \( P \) is a Borel probability measure on \( \mathbb{R} \), and the support of \( P \) is given by \( \text{supp}(P) = \{ \frac{1}{n} : n \in \mathbb{N} \} \). In this section, our goal is to determine the optimal sets of \( n \)-means and the \( n \)th quantization errors for all positive integers \( n \) for the probability measure \( P \). For \( k, \ell \in \mathbb{N} \), where \( k \leq \ell \), write

\[
    [k, \ell] := \left\{ \frac{1}{n} : n \in \mathbb{N} \text{ and } k \leq n \leq \ell \right\}, \quad \text{and } [k, \infty) := \left\{ \frac{1}{n} : n \in \mathbb{N} \text{ and } n \geq k \right\}.
\]

Further, write

\[
    \text{Av}[k, \ell] := E(\{ X : X \in [k, \ell] \}) = \sum_{n=k}^{\ell} \frac{1}{2n}, \quad \text{Av}[k, \infty) := E(\{ X : X \in [k, \infty) \}) = \sum_{n=k}^{\infty} \frac{1}{2n},
\]

\[
    \text{Er}[k, \ell] := \sum_{n=k}^{\ell} \frac{1}{2n} \left( \frac{1}{n} - \text{Av}[k, \ell] \right)^2, \quad \text{and } \text{Er}[k, \infty] := \sum_{n=k}^{\infty} \frac{1}{2n} \left( \frac{1}{n} - \text{Av}[k, \infty) \right)^2.
\]

Notice that \( E(X) := E(\{ X : X \in \text{supp}(P) \}) = \sum_{n=1}^{\infty} \frac{1}{2n} = \text{Av}[1, \infty) = \log(2), \) and so the optimal set of one-mean is the set \( \{ \log(2) \} \) with quantization error

\[
    V(P) = \sum_{n=1}^{\infty} \frac{1}{2n} \left( \frac{1}{n} - \log(2) \right)^2 = \text{Er}[1, \infty) = \frac{1}{12} \left( \pi^2 - 18 \log^2(2) \right) = 0.101788.
\]

**Proposition 4.1.1.** The set \( \{ \text{Av}[2, \infty), 1 \} \) forms the optimal set of two-means for the probability measure \( P \) with quantization error \( V_2(P) = \text{Er}[2, \infty) = \frac{1}{12} \left( \pi^2 - 12 - 30 \log^2(2) + 24 \log(2) \right) = 0.0076288597. \)

**Proof.** Consider the set \( \beta := \{ \text{Av}[2, \infty), 1 \} \). Since \( \frac{1}{3} < \frac{1}{2} (\text{Av}[2, \infty) + 1) < 1 \), the Voronoi region of 1 contains only the point 1, and the Voronoi region of \( \text{Av}[2, \infty) \) contains the set \( \{ \frac{1}{n} : n \geq 2 \} \). Hence, the distortion error due to the set \( \beta \) is given by

\[
    V(P; \beta) = \text{Er}[2, \infty) = \frac{1}{12} \left( \pi^2 - 12 - 30 \log^2(2) + 24 \log(2) \right) = 0.0076288597.
\]

Since \( V_2(P) \) is the quantization error for two-means, we have \( V_2(P) \leq 0.0076288597. \) Let \( \alpha := \{ a_2, a_1 \} \) be an optimal set of two-means. Due to Proposition 2.1, we can assume that \( 0 \leq a_2 < a_1 \leq 1 \). The Voronoi region of \( a_1 \) must contain 1. Suppose that the Voronoi region of \( a_1 \) also contains \( \frac{1}{2} \). Then,

\[
    V_2(P) \geq \text{Er}[1, 2] = \frac{1}{24} = 0.0416667 > V_2(P),
\]

which leads to a contradiction. Hence, we can assume that the Voronoi region of \( a_1 \) does not contain \( \frac{1}{2} \). Again, by Proposition 2.1, the Voronoi region of \( a_2 \) cannot contain the point 1. Thus, we have \( a_2 = \text{Av}[2, \infty) \), and \( a_1 = 1 \), and the corresponding quantization error is \( V_2(P) = \text{Er}[2, \infty) = 0.0076288597. \) Thus, the proof of the proposition is complete. \( \Box \)

**Proposition 4.1.2.** The set \( \{ \text{Av}[3, \infty), \frac{3}{2}, 1 \} \) forms the optimal set of three-means for the probability measure \( P \) with quantization error \( V_3(P) = \text{Er}[3, \infty) = 0.00116437359. \)
Proceeding as Proposition 4.1.2, we can show that the distortion error due to the set \( \alpha := \{Av[3, \infty), \frac{1}{3}, 1\} \) is given by

\[
V(P; \beta) = Er[3, \infty) = \frac{1}{24} (2\pi^2 - 51 - 108 \log^2(2) + 120 \log(2)) = 0.00116437359.
\]

Since \( V_3(P) \) is the quantization error for three-means, we have \( V_3(P) \leq 0.00116437359 \). Let \( \alpha := \{a_3, a_2, a_1\} \) be an optimal set of three-means such that \( 0 \leq a_3 < a_2 < a_1 \leq 1 \). Proceeding as Proposition 4.1.1, we can show that \( a_1 = 1 \). Suppose that the Voronoi region of \( a_2 \) contains \( \frac{1}{2} \) and \( \frac{1}{3} \). Then,

\[
V_3(P) \geq Er[2, 3] = 0.002314814815 > V_3(P),
\]

which is a contradiction. Hence, the Voronoi region of \( a_2 \) cannot contain \( \frac{1}{3} \). Thus, we have

\[
a_3 = Av[3, \infty), a_2 = \frac{1}{2}, \text{ and } a_1 = 1,
\]

with quantization error \( V_3(P) = Er[3, \infty) = 0.00116437359 \). Thus, the proof of the proposition is complete.

**Proposition 4.1.3.** The set \( \{Av[4, \infty), \frac{1}{3}, \frac{1}{2}, 1\} \) forms the optimal set of four-means for the probability measure \( P \) with quantization error \( V_4(P) = Er[4, \infty) = 0.000241896477 \).

**Proof.** The proof of this proposition is similar to the proof of Proposition 4.1.2.

**Proposition 4.1.4.** The set \( \{Av[5, \infty), \frac{1}{3}, \frac{1}{2}, \frac{1}{3}, 1\} \) forms the optimal set of five-means for the probability measure \( P \) with quantization error \( V_5(P) = Er[5, \infty) = 0.00005991266593 \).

**Proof.** The distortion error due to the set \( \beta := \{Av[5, \infty), \frac{1}{3}, \frac{1}{2}, \frac{1}{3}, 1\} \) is given by

\[
V(P; \beta) = Er[5, \infty) = Er[5, \infty) = 0.00005991266593.
\]

Since \( V_5(P) \) is the quantization error for five-means, we have \( V_5(P) \leq 0.00005991266593 \). Let \( \alpha := \{a_5, a_4, a_3, a_2, a_1\} \) be an optimal set of five-means such that \( 0 \leq a_5 < a_4 < a_3 < a_2 < a_1 \leq 1 \). Proceeding as Proposition 4.1.2, we can show that \( a_1 = 1 \). Suppose that the Voronoi region of \( a_4 \) contains \( \frac{1}{3} \), \( \frac{1}{2} \), and \( \frac{1}{3} \). Then, \( V_5(P) \geq Er[4, 6] = 0.0001116071429 > V_5(P) \), which is a contradiction. Assume that the Voronoi region of \( a_4 \) contains only the points \( \frac{1}{3} \), and \( \frac{1}{2} \). Then, the Voronoi region of \( a_5 \) contains the set \( [6, \infty) \), and so we have

\[
V_5(P) = Er[6, \infty) + Er[4, 5] = 0.00006872664638 > V_5(P),
\]

which leads to a contradiction. Hence, we can assume that the Voronoi region of \( a_4 \) contains only the point \( \frac{1}{3} \). Thus, we have \( a_5 = Av[5, \infty) \), \( a_4 = \frac{1}{3} \), \( a_3 = \frac{1}{3} \), \( a_2 = \frac{1}{2} \), and \( a_1 = 1 \) with quantization error \( V_5(P) = Er[5, \infty) = 0.00005991266593 \). Thus, the proof of the Proposition is complete.

**Proposition 4.1.5.** The set \( \{Av[7, \infty), Av[5, 6], \frac{1}{3}, \frac{1}{2}, \frac{1}{3}, 1\} \) forms the optimal set of six-means for the probability measure \( P \) with quantization error

\[
V_6(P) = Er[7, \infty) + Er[5, 6] = 0.00001658888625.
\]

**Proof.** Notice that \( \frac{1}{3} = 0.142857 < \frac{1}{2}(Av[7, \infty) + Av[5, 6]) = 0.158488 < 0.166667 = \frac{1}{6} \), and \( \frac{1}{6} < \frac{1}{2}(Av[5, 6] + \frac{1}{3}) < \frac{1}{2} \). Hence, the distortion error due to the set \( \beta := \{Av[7, \infty), Av[5, 6], \frac{1}{3}, \frac{1}{2}, \frac{1}{3}, 1\} \) is given by

\[
V(P; \beta) = Er[7, \infty) + Er[5, 6] = 0.00001658888625.
\]

Since \( V_6(P) \) is the distortion error for six-means, we have \( V_6(P) \leq 0.00001658888625 \). Let \( \alpha := \{a_6, a_5, a_4, a_3, a_2, a_1\} \) be an optimal set of six-means such that \( 0 \leq a_6 < a_5 < \cdots < a_1 \leq 1 \). Proceeding in the similar way as in the proof of Proposition 4.1.2, we can show that \( a_3 = \frac{1}{3}, a_2 = \frac{1}{2}, \text{ and } a_1 = 1 \). Proceeding in the similar way as in the proof of Proposition 4.1.4, we can show
that \(a_4 = \frac{1}{4}\). We now show that \(a_5 = \text{Av}[5, 6]\). Notice that the Voronoi region of \(a_5\) must contain \(\frac{1}{5}\). Suppose that the Voronoi region of \(a_5\) contains \(\frac{1}{7}\) and \(\frac{1}{6}\) as well. Then,

\[
V_6(P) \geq Er[5, 7] = 0.00002576328150 > V_6(P),
\]

which leads to a contradiction. Suppose that the Voronoi region of \(a_5\) contains only the point \(\frac{1}{5}\), i.e., \(a_5 = \frac{1}{5}\). Then,

\[
V_6(P) = Er[6, \infty) = 0.00001664331305 > V_6(P),
\]

which yields a contradiction. Hence, we can assume that the Voronoi region of \(a_5\) contains only the two points \(\frac{1}{6}\) and \(\frac{5}{6}\). Thus, we have

\[
a_6 = \text{Av}[7, \infty), a_5 = \text{Av}[5, 6), a_4 = \frac{1}{4}, a_3 = \frac{1}{3}, a_2 = \frac{1}{2}, \text{ and } a_1 = 1,
\]

and the quantization error is \(V_6(P) = Er[7, \infty) + Er[5, 6] = 0.0000165888625\). Thus, the proof of the proposition is complete.

In the following proposition, we calculate the optimal set of \(n\)-means and the \(n\)th quantization error for \(n = 200\).

**Proposition 4.1.6.** The set \(\{\text{Av}[301, \infty), \text{Av}[299, 300], \frac{1}{298}, \frac{1}{297}, \ldots, \frac{1}{5}, \frac{1}{2}, 1\}\) forms the optimal set of 300-means for the probability measure \(P\) with quantization error \(V_{300}(P) = Er[301, \infty) + Er[299, 300] = 1.564317642582409606174128 \times 10^{-100}\).

**Proof.** Notice that \(\frac{1}{301} = 0.003332259136 < \frac{1}{2}(\text{Av}[301, \infty) + \text{Av}[299, 300]) = 0.003326047849 < 0.003333333333 = \frac{1}{300}\), and \(\frac{2}{299} = 0.003334481605 < \frac{1}{2}(\text{Av}[299, 300] + \frac{1}{299}) = 0.003348235106 < 0.003355704698 = \frac{1}{298}\). Hence, the distortion error due to the set

\[
\beta := \{\text{Av}[301, \infty), \text{Av}[299, 300], \frac{1}{198}, \frac{1}{197}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\}
\]

is given by

\[
V(P; \beta) = Er[301, \infty) + Er[299, 300] = 1.564317642582409606174128 \times 10^{-100}.
\]

Since \(V_{300}(P)\) is the distortion error for 300-means, we have

\[
V_{300}(P) \leq 1.564317642582409606174128 \times 10^{-100}.
\]

Let \(\alpha := \{a_{300}, a_{299}, \ldots, a_3, a_2, a_1\}\) be an optimal set of 300-means such that \(0 \leq a_{300} < a_{299} < \cdots < a_1 \leq 1\). Proceeding in the similar way as in the proof of Proposition 4.1.2, we can show that \(a_{297} = \frac{1}{297}, a_{296} = \frac{1}{296}, \ldots, a_3 = \frac{3}{2}, a_2 = \frac{1}{2}, \) and \(a_1 = 1\). Proceeding in the similar way as in the proof of Proposition 4.1.4, we can show that \(a_{298} = \frac{1}{298}\). We now show that \(a_{298} = \text{Av}[299, 300]\). The Voronoi region of \(a_{299}\) must contain \(\frac{1}{299}\). Suppose that the Voronoi region of \(a_{299}\) contains \(\frac{1}{7}\) for \(i = 299, 300, 301, 302\). Then,

\[
V_{300}(P) \geq Er[299, 302] = 1.953916208081177222202350 \times 10^{-100} > V_{300}(P),
\]

which leads to a contradiction. Assume that the Voronoi region of \(a_{299}\) contains only the points \(\frac{1}{7}\) for \(i = 299, 300, 301\). Then,

\[
V_{300}(P) = Er[302, \infty) + Er[299, 301] = 1.69852119259119376459397 \times 10^{-100} > V_{300}(P),
\]

which yields a contradiction. Assume that the Voronoi region of \(a_{299}\) contains only the point \(\frac{1}{299}\). Then,

\[
V_{300}(P) = Er[300, \infty) = 2.345910694878821203973953 \times 10^{-100} > V_{300}(P),
\]

which gives a contradiction. Hence, we can assume that the Voronoi region of \(a_{299}\) contains only the two points \(\frac{1}{299}\) and \(\frac{1}{300}\). Thus, we have

\[
a_{300} = \text{Av}[301, \infty), a_{299} = \text{Av}[299, 300], a_{298} = \frac{1}{298}, \ldots, a_4 = \frac{1}{4}, a_3 = \frac{1}{3}, a_2 = \frac{1}{2}, \text{ and } a_1 = 1,
\]
and the quantization error is given by
\[ V_{300}(P) = Er[301, \infty) + Er[299, 300] = 1.564317642582409606174128 \times 10^{-100}. \]

Thus, the proof of the proposition is complete. \( \square \)

We now give the following theorem.

**Theorem 4.1.7.** For any positive integer \( n \), the sets \{Av[n, \infty), \frac{1}{n-1}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\}, where \( 1 \leq n \leq 5 \), form the optimal sets of \( n \)-means for the probability measure \( P \) with quantization errors \( V_n(P) := Er[n, \infty) \). For the positive integers \( n \), where \( 6 \leq n \leq 300 \), the sets \{Av[n+1, \infty), Av[n-1, n], \frac{1}{n-2}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\} form the optimal sets of \( n \)-means for the probability measure \( P \) with quantization errors \( V_n(P) = Er[n+1, \infty) + Er[n-1, n] \).

**Proof.** Due to Proposition 4.1.1 through Proposition 4.1.6, it follows that for \( 1 \leq n \leq 5 \), the sets \{Av[n, \infty), \frac{1}{n-1}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\} form the optimal sets of \( n \)-means for the probability measure \( P \) with quantization errors \( V_n(P) = Er[n, \infty) \). Proceeding in the similar way as Proposition 4.1.5 and Proposition 4.1.6, we can show that for \( 6 \leq n \leq 300 \), the sets \{Av[n+1, \infty), Av[n-1, n], \frac{1}{n-2}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\} form the optimal sets of \( n \)-means for the probability measure \( P \) with quantization errors \( V_n(P) = Er[n+1, \infty) + Er[n-1, n] \).

Thus, we complete the proof of the theorem. \( \square \)

We now give the following remark.

**Remark 4.1.8.** Proceeding in the similar way, as given in the proof of Theorem 4.1.7, it can be shown that the set \{Av[n, \infty), \frac{1}{n-1}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\} also gives an optimal set of \( n \)-means for \( n = 301 \). It is still not known whether the sets \{Av[n+1, \infty), Av[n-1, n], \frac{1}{n-2}, \ldots, \frac{1}{3}, \frac{1}{2}, 1\} give the optimal sets of \( n \)-means for all positive integers \( n \geq 6 \). If not, then the least upper bound of \( n \in \mathbb{N} \) for which such sets give the optimal sets of \( n \)-means for the probability measure \( P \) is not known yet.

4.2. Optimal quantization for an infinite discrete distribution with support \{\( n : n \in \mathbb{N} \}\}. Let \( \mathbb{N} := \{1, 2, 3, \ldots\} \) be the set of natural numbers, and let \( P \) be a Borel probability measure on the set \{\( n : n \in \mathbb{N} \}\) with probability density function \( f \) given by
\[ f(x) = \begin{cases} \frac{1}{2^n} & \text{if } x = n \text{ for } n \in \mathbb{N}, \\ 0 & \text{otherwise}. \end{cases} \]

Then, \( P \) is a Borel probability measure on \( \mathbb{R} \), and the support of \( P \) is the set \( \mathbb{N} \) of natural numbers. In this section, our goal is to determine the optimal sets of \( n \)-means and the \( n \)th quantization errors for all positive integers \( n \) for the probability measure \( P \). For \( k, \ell \in \mathbb{N} \), where \( k \leq \ell \), write
\[ \{k, \ell\} := \{n : n \in \mathbb{N} \text{ and } k \leq n \leq \ell\}, \quad \text{and } \{k, \infty\} := \{n : n \in \mathbb{N} \text{ and } n \geq k\}. \]

Further, write
\[ Av[k, \ell] := E\left(X : X \in [k, \ell]\right) = \frac{\sum_{n=k}^{\ell} \frac{n}{2^n}}{\sum_{n=k}^{\ell} 1/2^n}, \quad Av[k, \infty) := E\left(X : X \in [k, \infty)\right) = \frac{\sum_{n=k}^{\infty} \frac{n}{2^n}}{\sum_{n=k}^{\infty} 1/2^n}, \]
\[ Er[k, \ell] := \sum_{n=k}^{\ell} \frac{1}{2^n}(n - Av[k, \ell])^2, \quad \text{and } Er[k, \infty) := \sum_{n=k}^{\infty} \frac{1}{2^n}(n - Av[k, \infty])^2. \]
Notice that \( E(P) := E(X : X \in \text{supp}(P)) = \sum_{n=1}^{\infty} \frac{1}{2^n} = Av[1, \infty) = 2 \), and so the optimal set of one-mean is the set \{2\} with quantization error

\[
V(P) = \sum_{n=1}^{\infty} \frac{1}{2^n} (n - 2)^2 = Er[1, \infty) = 2.
\]

**Proposition 4.2.1.** The optimal set of two-means is given by \{Av[1,2], Av[3,\infty)\} with quantization error \( V_2 = \frac{2}{3} \).

*Proof.* We see that \( Av[1,2] = \frac{1}{3} \), and \( Av[3,\infty) = 4 \). Since \( \frac{1}{3} < \frac{1}{2}(\frac{4}{3} + 4) < 4 \), the distortion error due to the set \( \beta := \{\frac{1}{3}, 4\} \) is given by

\[
V(P; \beta) = Er[1,2] + Er[3,\infty) = \frac{2}{3}.
\]

Since \( V_2 \) is the quantization error for two-means, we have \( V_2 \leq \frac{2}{3} \). Notice that the Voronoi region of \( a_1 \) must contain 1. Suppose that the Voronoi region of \( a_1 \) contains only the set \{1,2,3,4\}. Then,

\[
V_2 \geq 4 \sum_{j=1}^{4} \frac{1}{2^2} (j - Av[1,4])^2 = Er[1,4] = \frac{97}{120} = 0.808333 > V_2,
\]

which yields a contradiction. Hence, we can assume that the Voronoi region of \( a_2 \) contains only the set \{1,2,3\}, and so the Voronoi region of \( a_2 \) contains the set \( \{n : n \geq 4\} \). Then, we have

\[
V_2 = Er[1,3] + Er[4,\infty) = \frac{5}{7} = 0.714286 > V_2,
\]

which is a contradiction. Next, suppose that the Voronoi region of \( a_1 \) contains only the element 1, and so the Voronoi region of \( a_2 \) contains the set \( \{n : n \geq 2\} \). Then, we have

\[
V_2 = Er[2,\infty) = 1 > V_2,
\]

which leads to a contradiction. Hence, we can assume that the Voronoi region of \( a_1 \) contains the set \{1,2\}, and so the Voronoi region of \( a_2 \) contains \{3,4,5,\ldots\}, yielding \( a_1 = Av[1,2] \), and \( a_2 = Av[3,\infty) \), and the corresponding quantization error is \( V_2 = \frac{2}{3} \). Thus, the proof of the proposition is complete. \( \Box \)

**Proposition 4.2.2.** The sets \{1, Av[2,3], Av[4,\infty)\}, and \{Av[1,2], Av[3,4], Av[5,\infty)\} form two optimal sets of three-means with quantization error \( V_3 = \frac{1}{3} \).

*Proof.* The distortion error due to set \( \beta := \{1, Av[2,3], Av[4,\infty)\} \) is given by

\[
V(P; \beta) = Er[2,3] + Er[4,\infty) = \frac{1}{3}.
\]

Notice that the distortion error due to the set \{Av[1,2], Av[3,4], Av[5,\infty)\} is also \( \frac{1}{3} \). Since \( V_3 \) is the quantization error for three-means, we have \( V_3 \leq \frac{1}{3} \). Let \( \alpha := \{a_1, a_2, a_3\} \) be an optimal set of three-means, where \( 1 \leq a_1 < a_2 < a_3 < \infty \). Suppose that the Voronoi region of \( a_1 \) contains the set \{1,2,3\}. Then,

\[
V_3 \geq 3 \sum_{j=1}^{3} \frac{1}{2^3} (j - Av[1,3])^2 = \frac{13}{28} > \frac{1}{3} > V_3,
\]

which leads to a contradiction. Hence, we can assume that the Voronoi region of \( a_1 \) contains either the set \{1\}, or the set \{1,2\}. Consider the following two cases:

**Case 1.** The Voronoi region of \( a_1 \) contains only the set \{1\}.

In this case, the Voronoi region of \( a_2 \) must contain the element 2. Suppose that the Voronoi region of \( a_2 \) contains the set \{2,3,4,5\}. Then,

\[
V_3 \geq 5 \sum_{j=2}^{5} \frac{1}{2^5} (j - Av[2,5])^2 = \frac{97}{240} = 0.404167 > V_3,
\]
which yields a contradiction. Assume that the Voronoi region of \(a_2\) contains only the set \(\{2, 3, 4\}\), and so the Voronoi region of \(a_3\) contains the set \(\{n: n \geq 5\}\). Then, the distortion error is

\[
V_3 = Er[2, 4] + Er[5, \infty] = \frac{5}{14} = 0.357143 > V_3,
\]
which gives a contradiction. Next, assume that the Voronoi region of \(a_2\) contains only the element 2, and so the Voronoi region of \(a_3\) contains the set \(\{n: n \geq 3\}\). Then, the distortion error is

\[
V_3 = Er[3, \infty] = \frac{1}{2} > V_3,
\]
which is a contradiction. Hence, in this case, we can conclude that the Voronoi region of \(a_2\) contains only the set \(\{2, 3\}\), yielding \(a_1 = 1, a_2 = Av[2, 3]\), and \(a_3 = Av[4, \infty]\) with quantization error \(V_3 = \frac{1}{3}\).

Case 2. The Voronoi region of \(a_1\) contains only the set \(\{1, 2\}\).

In this case, the Voronoi region of \(a_2\) must contain the element 3. Suppose that the Voronoi region of \(a_2\) contains the set \(\{3, 4, 5, 6\}\). Then,

\[
V_3 \geq \sum_{j=1}^{2} \frac{1}{2^j} (j - Av[1, 2])^2 + \sum_{j=3}^{6} \frac{1}{2^j} (j - Av[3, 6])^2 = \frac{59}{160} = 0.36875 > V_3,
\]
which yields a contradiction. Assume that the Voronoi region of \(a_2\) contains only the set \(\{3, 4, 5\}\), and so the Voronoi region of \(a_3\) contains the set \(\{n: n \geq 6\}\). Then, the distortion error is

\[
V_3 = Er[1, 2] + Er[3, 5] + Er[6, \infty] = \frac{29}{84} = 0.345238 > V_3,
\]
which gives a contradiction. Next, assume that the Voronoi region of \(a_2\) contains only the element 3, and so the Voronoi region of \(a_3\) contains the set \(\{n: n \geq 4\}\). Then, the distortion error is

\[
V_3 = Er[1, 2] + Er[4, \infty] = \frac{5}{12} = 0.416667 > V_3,
\]
which yields a contradiction. Hence, in this case, we can conclude that the Voronoi region of \(a_2\) contains only the set \(\{3, 4\}\), yielding \(a_1 = Av[1, 2], a_2 = Av[3, 4]\), and \(a_3 = Av[5, \infty]\) with quantization error \(V_3 = \frac{1}{3}\).

By Case 1 and Case 2, the proof of the proposition is complete. \(\square\)

We need the following lemma.

Lemma 4.2.3. Let \(n \geq 4\), and let \(\alpha_n\) be an optimal set of \(n\)-means. Then, \(\alpha_n\) must contain the set \(\{1, 2, \ldots, (n-3)\}\).

Proof. The distortion error due to the set \(\beta := \{1, 2, \ldots, (n-3), (n-2), Av[n-1, n], Av[n+1, \infty]\}\) is given by

\[
V(P; \beta) = Er[n-1, n] + Er[n+1, \infty] = \frac{2^{3-n}}{3}.
\]

Since \(V_n\) is the quantization error for \(n\)-means, we have \(V_n \leq \frac{2^{3-n}}{3}\). Let \(\alpha_n := \{a_1, a_2, \ldots, a_n\}\) be an optimal set of \(n\)-means such that \(1 \leq a_1 < a_2 < \cdots < a_n < \infty\). We show that \(a_1 = 1, a_2 = 2, \ldots, a_{n-3} = n - 3\). Notice that the Voronoi region of \(a_1\) must contain the element 1. Suppose that the Voronoi region of \(a_1\) also contains the element 2. Then,

\[
V_n > \sum_{j=1}^{2} \frac{1}{2^j} (j - Av[1, 2])^2 = \frac{1}{6} \geq \frac{2^{3-n}}{3} \geq V_n,
\]
which is a contradiction. Hence, we can conclude that the Voronoi region of \(a_1\) contains only the element 1, yielding \(a_1 = 1\). Thus, we can deduce that there exists a positive integer \(k\), where \(1 \leq k < n - 3\), such that \(a_1 = 1, a_2 = 2, \cdots, a_k = k\). We now show that \(a_{k+1} = k + 1\). Notice
that the Voronoi region of $a_{k+1}$ must contain $k + 1$. Suppose that the Voronoi region of $a_{k+1}$ also contains the element $k + 2$. Then, as $k < n - 3$, we have

$$V_n > \sum_{j=k+1}^{k+2} \frac{1}{2^j} (j - \text{Av}[k+1, k+2])^2 = \frac{2^{-k-1}}{3} \geq \frac{2^{3-n}}{3} \geq V_n,$$

which is a contradiction. Hence, we can conclude that the Voronoi region of $a_{k+1}$ contains only the element $k + 1$. Thus, by the Principle of Mathematical Induction, we deduce that $a_1 = 1, a_2 = 2, \cdots, a_{n-3} = n - 3$. Thus, the proof of the lemma is complete. \hfill \Box

**Theorem 4.2.4.** Let $n \geq 4$, and let $\alpha_n$ be an optimal set of $n$-means. Then, either $\alpha_n = \{1, 2, 3, \cdots, n-3, n-2, \text{Av}[n-1, n], \text{Av}[n+1, \infty]\}$, or $\alpha_n = \{1, 2, 3, \cdots, n-3, \text{Av}[n-2, n-1], \text{Av}[n, n+1], \text{Av}[n+2, \infty]\}$ with quantization error $V_n = \frac{2^{3-n}}{3}$.

**Proof.** As shown in the proof of Lemma 4.2.3, we have $V_n \leq \frac{2^{3-n}}{3}$. Let $\alpha_n := \{a_1, a_2, \cdots, a_n\}$ be an optimal set of $n$-means such that $1 \leq a_1 < a_2 < \cdots < a_n < \infty$. By Lemma 4.2.3, we have $a_1 = 1, a_2 = 2, \cdots, a_{n-3} = n - 3$. Recall that $n \geq 4$. Suppose that the Voronoi region of $a_{n-2}$ contains the set $\{n-2, n-1, n\}$. Then,

$$V_n \geq \sum_{j=n-2}^{n} \frac{1}{2^j} (j - \text{Av}[n-2, n])^2 = \frac{13}{7} 2^{1-n} \geq \frac{2^{3-n}}{3} \geq V_n,$$

which leads to a contradiction. Hence, we can assume that the Voronoi region of $a_{n-2}$ contains either the set $\{n-2\}$, or the set $\{n-2, n-1\}$. Consider the following two cases:

**Case 1.** The Voronoi region of $a_{n-2}$ contains only the set $\{n-2\}$.

Proceeding along the similar lines as Case 1 in the proof of Proposition 4.2.2, we can show that the Voronoi region of $a_{n-1}$ contains only the set $\{n-1, n\}$, yielding $a_{n-2} = n-2, a_{n-1} = \text{Av}[n-1, n]$, and $a_n = \text{Av}[n+1, \infty]$ with quantization error $V_n = \frac{2^{3-n}}{3}$.

**Case 2.** The Voronoi region of $a_{n-2}$ contains only the set $\{n-2, n-1\}$.

Proceeding along the similar lines as Case 2 in the proof of Proposition 4.2.2, we can show that the Voronoi region of $a_{n-1}$ contains only the set $\{n, n+1\}$, yielding $a_{n-2} = \text{Av}[n-2, n-1]$, $a_{n-1} = \text{Av}[n, n+1]$, and $a_n = \text{Av}[n+2, \infty]$ with quantization error $V_n = \frac{2^{3-n}}{3}$.

By Case 1 and Case 2, the proof of the theorem is complete. \hfill \Box

5. Probability distributions when the optimal sets are given

Let $P$ be a discrete probability measure on $\mathbb{R}$ with support a finite or an infinite set $\{1, 2, 3, \cdots\}$. Let $(p_1, p_2, p_3, \cdots)$ be a probability vector associated with $\{1, 2, 3, \cdots\}$ such that the probability mass function $f$ of $P$ is given by $f(k) = p_k$ if $k \in \{1, 2, 3, \cdots\}$, and zero otherwise. For $k, \ell \in \{1, 2, 3, \cdots\}$ with $k \leq \ell$, write

$$[k, \ell] := \{n : k \leq n \leq \ell\}, \text{ and } [k, \infty) := \{k, k+1, \cdots\}.$$

For a random variable $X$ with distribution $P$, let $\text{Av}[k, \ell]$ represent the conditional expectation of $X$ given that $X$ takes values on the set $\{k, k+1, k+2, \cdots, \ell\}$, i.e.,

$$\text{Av}[k, \ell] = E(X : X \in [k, \ell]),$$

where $k, \ell \in \{1, 2, 3, \cdots\}$ with $k \leq \ell$. On the other hand, by $\text{Av}[k, \infty)$ it is meant $\text{Av}[k, \infty) = E(X : X \in [k, \infty))$, where $k \in \{1, 2, 3, \cdots\}$. Let $\alpha_n$ be an optimal set of $n$-means for $P$, where $n \in \mathbb{N}$. In this section, our goal is to find a set of probability vectors $(p_1, p_2, p_3, \cdots)$ such that for all $n \in \mathbb{N}$, the optimal sets of $n$-means are given by $\alpha_n = \{1, 2, 3, \cdots, n-1, \text{Av}[n, \infty)\}$.

Consider the following two cases:

**Case 1.** $\{1, 2, 3, \cdots\}$ is a finite set.

In this case, there exists a positive integer $m$, such that the support of $P$ is given by $\{1, 2, 3, \cdots, m\}$. Notice that for any $k \in \{1, 2, \cdots, m\}$, in this case by $[k, \infty)$ it meant the set $[k, m]$. If $m = 1$, then $\alpha_1 = \{1\}$; and if $m = 2$, then $\alpha_1 = \{\text{Av}[1, \infty)\}$, and $\alpha_2 = \{1, \text{Av}[2, \infty)\}$,
i.e., there is nothing to prove. So, we can assume that \( m \geq 3 \). Define the probability vector \((p_1, p_2, \cdots, p_m)\) as follows:

\[
 p_j = \begin{cases} 
 x & \text{if } j = 1, \\
 (1 - x)^{j-1}x & \text{if } 2 \leq j \leq m - 1, \\
 (1 - x)^{j-1} & \text{if } j = m.
\end{cases}
\]

(1)

For the sets \( \alpha_n \) to form the optimal sets of \( n \)-means for all \( 1 \leq n \leq m \), we must have

\[
 (n - 1) \leq \frac{1}{2} (n - 1 + \text{Av}(n, \infty)) \leq n
\]

(2)

for \( 2 \leq n \leq m \). The set of values of \( x \) obtained by solving the above inequalities does not guarantee that the sets \( \alpha_n \) for \( 1 \leq n \leq m \) will form the optimal set of \( n \)-means. Thus, we need further investigation. Due to symmetry in the construction of the probability vectors, we can say that \( \alpha_n \) for \( 1 \leq n \leq m \) will form the optimal sets of \( n \)-means if the following condition is also true:

\[
 V(P; \{1, \text{Av}[2, \infty]\}) \leq V(P; \{\text{Av}[1, 2], \text{Av}[3, \infty]\}).
\]

Thus, we conjecture that the values of \( x \), for which the inequalities given by (2) and (3) are true, form the set of probability vectors \((p_1, p_2, p_3, \cdots, p_m)\), given by (1), for which the sets \( \alpha_n \) for \( 1 \leq n \leq m \) form the optimal sets of \( n \)-means. By several examples, we verified that the conjecture is true, also see Example 5.1 and Example 5.2.

Case 2. \( \{1, 2, 3, \cdots\} \) is an infinite set.

Define the probability vector \((p_1, p_2, p_3, \cdots)\) as follows:

\[
 p_j = \begin{cases} 
 x & \text{if } j = 1, \\
 (1 - x)^{j-1}x & \text{if } 2 \leq j.
\end{cases}
\]

(4)

For the sets \( \alpha_n \) to form optimal sets of \( n \)-means for all \( 1 \leq n \), we must have

\[
 (n - 1) \leq \frac{1}{2} (n - 1 + a(n)) \leq n
\]

(5)

for \( 2 \leq n \). The set of values of \( x \) obtained by solving the above inequalities does not guarantee that \( \alpha_n \) for \( 1 \leq n \) will form an optimal set of \( n \)-means. Thus, we need further investigation. Due to symmetry in the construction of the probability vectors, we can say that the sets \( \alpha_n \) for \( 1 \leq n \) will form the optimal sets of \( n \)-means if the following inequality is also true:

\[
 V(P; \{1, \text{Av}[2, \infty]\}) \leq V(P; \{\text{Av}[1, 2], \text{Av}[3, \infty]\}).
\]

After some calculation, we see that there exists a real number \( y \), the ten-digit rational approximation of which is 0.6666666667, such that the inequalities given by (5) and (6) are satisfied if \( y \leq x < 1 \). Thus, we conjecture that the sets \( \alpha_n \) for \( 1 \leq n \) will form the optimal sets of \( n \)-means if the probability vector \((p_1, p_2, p_3, \cdots)\) is given by (1) for 0.6666666667 \( \leq x < 1 \). By several examples, we verified that the conjecture is true.

**Example 5.1.** Let \( m = 6 \) in Case 1. Then, for \( 0 < x < 1 \) we have

\[
 p_1 = x, \ p_2 = (1 - x)x, \ p_3 = (1 - x)^2x, \ p_4 = (1 - x)^3x, \ p_5 = (1 - x)^4x, \text{ and } p_6 = (1 - x)^5.
\]

After solving the inequalities given by (2), we have 0.4812099363 < \( x < 1 \). Again, solving the inequality (3), we have 0.6628057756 \( \leq x < 1 \). Notice that 0.4812099363 and 0.6628057756 are the ten-digit rational approximations of two real numbers. Thus, the inequalities given by (2) and (3) are true if 0.6628057756 \( \leq x < 1 \). Hence, a set of probability vectors \((p_1, p_2, \cdots, p_6)\) for which the given sets \( \alpha_n \) form the optimal sets of \( n \) means for \( 1 \leq n \leq 6 \) is given by

\[
 \{(x, (1 - x)x, (1 - x)^2x, (1 - x)^3x, (1 - x)^4x, (1 - x)^5) : 0.6628057756 \leq x < 1\}.
\]
Example 5.2. Let $m = 7$ in Case 1. Then, proceeding as Example 5.1 we see that (2) and (3) are true if $0.6654212000 \leq x < 1$. Hence, a set of probability vectors $(p_1, p_2, \ldots, p_7)$ for which the given sets $\alpha_n$ form the optimal sets of $n$ means for $1 \leq n \leq 7$ is given by

$$\{x, (1-x)x, (1-x)^2x, (1-x)^3x, (1-x)^4x, (1-x)^5x, (1-x)^6\}$$

where $0.6654212000 \leq x < 1$.

References

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

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

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

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

[R] M.K. Roychowdhury, *Optimal quantizers for some absolutely continuous probability measures*, Real Analysis Exchange, Vol. 43(1), 2017, pp. 105-136.

[RR] J. Rosenblatt and M.K. Roychowdhury, *Optimal quantization for piecewise uniform distributions*, Uniform Distribution Theory 13 (2018), no. 2, 23-55.

[RS] M.K. Roychowdhury, and Wasiela Salinas, *Quantization for a mixture of uniform distributions associated with probability vectors*, Uniform Distribution Theory 15 (2020), no. 1, 105-142.

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

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

E-mail address: \{1russellloyd.cabasag01, 2samir.huq01, 3eric.mendoza01, 4mrinal.roychowdhury\}@utrgv.edu