Optimal Quantization for Mixtures of Two Uniform Distributions

Eduardo Orozco
The University of Texas Rio Grande Valley

OPTIMAL QUANTIZATION FOR MIXTURES OF TWO UNIFORM DISTRIBUTIONS

A Thesis

by

EDUARDO OROZCO

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2021

Major Subject: Mathematics

OPTIMAL QUANTIZATION FOR MIXTURES OF TWO UNIFORM DISTRIBUTIONS

A Thesis
by
EDUARDO OROZCO

COMMITTEE MEMBERS

Dr. Mrinal Kanti Roychowdhury
Chair of Committee

Dr. Santanu Chakraborty
Committee Member

Dr. Hansapani Rodrigo
Committee Member

Dr. Josef Sifuentes
Committee Member

May 2021

ABSTRACT

Orozco, Eduardo, Optimal quantization for mixtures of two uniform distributions. Master of Science (MS), May, 2021, 17 pp., 1 table, 24 references.

The basic goal of quantization for probability distribution is to reduce the number of values, which is typically uncountable, describing a probability distribution to some finite set and thus approximation of a continuous probability distribution by a discrete distribution. Mixtures of probability distributions, also known as mixed distributions, are an exciting new area for optimal quantization. In this thesis, for a mixed distribution we determine the optimal sets of n-means and the nth quantization errors for all positive integers n.

DEDICATION

To my parents, brother, and sister.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee chair, Dr. Mrinal Kanti Roychowdhury for his mentorship. He has provided persistent counsel, without which this thesis would not be possible.

I am also extremely thankful for my committee members Dr. Santanu Chakraborty, Dr. Hansapani Rodrigo, and Dr. Josef Sifuentes for all their feedback which was invaluable to the completion of this thesis.

I am grateful to all the teachers whose guidance has built me up into the student I am now. I would also like to thank my parents for all their encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Continuous-valued signals can take any real value either in the entire range of real numbers or in a range limited by some system constraints. In either of the two cases, an uncountably infinite set of values is required to represent the signal values. If a signal has to be processed or stored digitally, each of its values must be representable by a finite number of bits. Thus, all values together have to form a finite countable set. A signal consisting only of such discrete values is said to be a quantized signal. The process of transformation of a continuous-valued signal into a discrete-valued one is called 'quantization'. 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 [GL1]). Let $\mathbb{R}^d$ denote the d-dimensional Euclidean space equipped with a metric $\|\cdot\|$ compatible with the Euclidean topology. Let P be a Borel probability measure on $\mathbb{R}^d$ and α be a finite subset of $\mathbb{R}^d$. Then, $\int \min_{a\in\alpha} \|x-a\|^2 dP(x)$ is often referred to as the *cost,* or *distortion error* for α with respect to the probability measure P, and is denoted by $V(P;\alpha)$. Write $\mathscr{D}_n := \{\alpha \subset \mathbb{R}^d : 1 \leq \text{card}(\alpha) \leq n\}$. Then, $\inf\{V(P;\alpha) : \alpha \in \mathscr{D}_n\}$ is called the *nth quantization error* for the probability measure P, and is denoted by $V_n := V_n(P)$. A set α for which the infimum occurs and contains no more than n points is called an *optimal set of n-means*. Since $\int \|x\|^2 dP(x) < \infty$ such a set α always exists (see [AW, GKL, GL1, GL2]). Furthermore, whenever the support of the probability measure P is an infinite set, then an optimal set of n-means for P contains exactly n elements (see Theorem 4.12 in [GL1]). For some recent work in this direction one can see [CR, DR1, DR2, GL3, HMRT, L1, PRRSS, R1, R2, R3, R4, R5, RR1].

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

Proposition I.1. *Let α 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 $M(a|\alpha)$ is the Voronoi region of $a \in \alpha$, i.e., $M(a|\alpha)$ is the set of all elements x in $\mathbb{R}^d$ which are closest to a among all the elements in α.

Proposition I.1 says that if α is an optimal set and $a \in \alpha$, then a is the *conditional expectation* of the random variable X given that X takes values in the Voronoi region of a. The following theorem is known.

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

Mixed distributions are an exciting new area for optimal quantization. For any two Borel probability measures P_1 and P_2, and $p \in (0,1)$, if $P := pP_1 + (1-p)P_2$, then the probability measure P is called the *mixture* or the *mixed distribution* generated by the probability measures (P_1, P_2) associated with the probability vector $(p, 1-p)$. This kind of problems has rigorous applications in many areas including signal processing. For example, while driving long distances, sometimes we experience cellular signals getting cut off. This happens because of being far away from the tower, or there is no tower nearby to catch the signal. In optimal quantization for mixed distributions, one of our goals is to find the exact locations of the towers by giving different weights, also called importance, to different portions of a path. Interested readers can also see the paper [R6].

Let P_1 and P_2 be two uniform distributions on the two disconnected line segments $J_1 := [0, \frac{1}{3}]$ and $J_2 := [\frac{2}{3}, 1]$ of equal lengths, and P be the mixed distribution generated by (P_1, P_2) associated with a probability vector $(p, 1-p)$. Take $p = \frac{1}{3}$, i.e., for the mixed distribution $P = \frac{1}{3}P_1 + \frac{2}{3}P_2$, in this thesis, we determine the optimal sets of n-means and the nth quantization errors for all $n \in \mathbb{N}$. In this regard, we would like to mention that if α_n is an optimal set of n-means for all $n \in \mathbb{N}$ and if $p = \frac{1}{3}$, then $\alpha_n \cap J_1 \neq \emptyset$ and $\alpha_n \cap J_2 \neq \emptyset$. But it is not true for all probability vectors $(p, 1-p)$, where $0 < p < \frac{1}{2}$, see Remark II.13 and Proposition II.12. One of the main significance of such a result is that the technique utilized in this thesis can be useful to find the optimal sets of n-means and the nth quantization errors for the mixed distributions on any two disconnected line segments for all $n \in \mathbb{N}$. At the end of the thesis, in a remark, we also mentioned about some open questions.

CHAPTER II

OPTIMAL QUANTIZATION FOR THE MIXTURE OF TWO UNIFORM DISTRIBUTIONS ON TWO DISCONNECTED LINE SEGMENTS

Let P_1 and P_2 be uniform distributions, respectively, on the intervals given by

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

Let f_1 and f_2 be their respective density functions. Then, $f_1(x) = 3$ if $x \in [0, \frac{1}{3}]$, and zero otherwise; and $f_2(x) = 3$ if $x \in [\frac{2}{3}, 1]$, and zero otherwise. The underlying mixed distribution considered is given by $P := pP_1 + (1-p)P_2$, where $p = \frac{1}{3}$. By $E(X)$ we mean the expectation of a random variable X with distribution P, and $V(X)$ represents the variance of P. By $\alpha_n(\mu)$, we denote an optimal set of n-means with respect to a probability distribution μ, and $V_n(\mu)$ represents the corresponding quantization error for n-means. If μ is the mixed distribution P, sometimes we denote them by α_n instead of $\alpha_n(P)$, and the corresponding quantization error by V_n instead of $V_n(P)$.

Lemma II.1. *Let P be the mixed distribution defined by $P = pP_1 + (1-p)P_2$. Then, $E(X) = \frac{1}{6}(5-4p)$, and $V(X) = \frac{1}{108}(-48p^2 + 48p + 1)$.*

Proof. We have

$$E(X) = \int x dP = p \int x d(P_1(x)) + (1-p) \int x d(P_2(x)) = p \int_{J_1} 3x\, dx + (1-p) \int_{J_2} 3x\, dx$$

yielding $E(X) = \frac{1}{6}(5-4p)$, and

$$V(X) = \int (x - E(X))^2 dP = p \int (x - E(X))^2 d(P_1(x)) + (1-p) \int (x - E(X))^2 d(P_2(x)),$$

implying $V(X)=\frac{1}{108}(-48p^2+48p+1)$, and thus, the lemma is yielded. □

Remark II.2. The optimal set of one-mean is the set $\{\frac{1}{6}(5-4p)\}$, and the corresponding quantization error is the variance $V:=V(X)$ of a random variable with distribution $P:=pP_1+(1-p)P_2$. Recall that in our case, $p=\frac{1}{3}$, and then $E(X)=\frac{11}{18}$ and $V(X)=\frac{35}{324}$.

Proposition II.3. *For $n\geq 2$ let α_n be an optimal set of n-means for P. Then, $\alpha_n\cap J_1\neq\emptyset$ and $\alpha_n\cap J_2\neq\emptyset$.*

Proof. The distortion error due to the set $\beta:=\{\frac{1}{6},\frac{5}{6}\}$ is given by

$$\int\min_{a\in\beta}(x-a)^2dP=p\int(x-\frac{1}{6})^2dP_1+(1-p)\int(x-\frac{5}{6})^2dP_2=\frac{1}{108}.$$

Let $\alpha_n:=\{a_1,a_2,\cdots,a_n\}$ be an optimal set of n-means for P such that $a_1<a_2<\cdots<a_n$ for $n\geq 2$. Since V_n is the quantization error for n-means for $n\geq 2$, we have $V_n\leq V_2\leq\frac{1}{108}$. Suppose that $\alpha_n\cap J_1=\emptyset$, which yields $\frac{1}{3}<a_1$. Then, we have

$$V_2\geq\int_{J_1}(x-\frac{1}{3})^2dP=\frac{1}{81}>\frac{1}{108}\geq V_2,$$

which leads to a contradiction. Next, suppose that $\alpha_n\cap J_2=\emptyset$. Then, $a_n<\frac{2}{3}$, which yields

$$V_2\geq\int_{J_2}(x-\frac{2}{3})^2dP=\frac{2}{81}>\frac{1}{108}\geq V_2,$$

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

Corollary II.4. Proposition II.3 implies that the set $\beta:=\{\frac{1}{6},\frac{5}{6}\}$ forms an optimal set of two means with quantization error $V_2=\frac{1}{108}$.

Proposition II.5. *For $n\geq 2$ let α_n be an optimal set of n-means for P. If α_n contains a point from the open interval $(\frac{1}{3},\frac{2}{3})$, then we must have card$(\alpha_n\cap(\frac{1}{3},\frac{2}{3}))=1$, where card$(A)$ of a set A means the cardinality of the set A.*

Proof. By Proposition II.3, we see that α_2 does not contain any point from the open interval $(\frac{1}{3}, \frac{2}{3})$, and if α_3 contains a point from the open interval, it cannot contain more than one point from the open interval. Thus, the proposition is true for $n = 2, 3$. We now prove that the proposition is true for $n \geq 4$. For $n \geq 4$, let $\alpha_n := \{a_1, a_2, \cdots, a_n\}$ be an optimal set of n-means for P such that $a_1 < a_2 < \cdots < a_n$. Since the points in an optimal set are the conditional expectations in their own Voronoi regions, we must have $0 < a_1 < a_2 < a_3 < \cdots < a_n < 1$. Let k be the largest positive integer such that $a_k \leq \frac{1}{3}$. Suppose that α_n contains a point from the open interval $(\frac{1}{3}, \frac{2}{3})$. For the sake of contradiction assume that $\text{card}(\alpha_n \cap (\frac{1}{3}, \frac{2}{3})) = m$, where $m \geq 2$. Then,

$$\frac{1}{3} < a_{k+1} < a_{k+2} < \cdots < a_{k+m} < \frac{2}{3}.$$

The following two cases can arise:

Case 1. $m \geq 3$.

In this case, we have $P(M(a_{k+j}|\alpha_n)) = 0$ for $2 \leq j \leq m-1$, which because of Proposition I.1, leads to a contradiction.

Case 2. $m = 2$.

In this case, we have $a_k \leq \frac{1}{3} < a_{k+1} < a_{k+2} < \frac{2}{3} \leq a_{k+3}$. Indeed, due to Proposition I.1, we have

$$a_k < \frac{1}{2}(a_k + a_{k+1}) < \frac{1}{3} < a_{k+1} < \frac{1}{2}(a_{k+1} + a_{k+2}) < a_{k+2} < \frac{2}{3} < \frac{1}{2}(a_{k+2} + a_{k+3}) < a_{k+3}.$$

Now notice that the total error contributed by the two points a_{k+1} and a_{k+2} is given by

$$\int_{[\frac{1}{2}(a_k + a_{k+1}), \frac{1}{3}]} (x - a_{k+1})^2 dP + \int_{[\frac{2}{3}, \frac{1}{2}(a_{k+2} + a_{k+3})]} (x - a_{k+2})^2 dP,$$

which can be strictly reduced if we replace a_{k+1} by $\frac{1}{3}$, and a_{k+2} by $\frac{2}{3}$, which is a contradiction, as we assumed that α_n is an optimal set of n-means with $\frac{1}{3} < a_{k+1} < a_{k+2} < \frac{2}{3}$.

By Case 1 and Case 2, we can deduce that $m \leq 1$, i.e., if α_n contains a point from the open interval $(\frac{1}{3}, \frac{2}{3})$, then we must have $\text{card}(\alpha_n \cap (\frac{1}{3}, \frac{2}{3})) = 1$, which completes the proof of the proposition. $\square$

Lemma II.6. *The set* $\{\frac{1}{6}, \frac{3}{4}, \frac{11}{12}\}$ *forms an optimal set of three-means with quantization error* $V_3 = \frac{1}{216}$.

Proof. Consider the set of three points β such that $\beta := \{\frac{1}{6}, \frac{3}{4}, \frac{11}{12}\}$. The distortion error due to the set β is given by

$$\int \min_{a\in\beta}(x-a)^2 dP = \frac{1}{3}\int_{J_1}(x-\frac{1}{6})^2 dP_1 + 2\cdot\frac{2}{3}\int_{[\frac{2}{3},\frac{5}{6}]}(x-\frac{3}{4})^2 dP_2 = \frac{1}{216}.$$

Since V_3 is the quantization error for three-means, we have $V_3 \leq \frac{1}{216}$. Let $\alpha := \{a_1, a_2, a_3\}$ be an optimal set of three-means. Since the points in an optimal set are the conditional expectations in their own Voronoi regions, without any loss of generality, we can assume that $0 < a_1 < a_2 < a_3 < 1$. By Proposition II.3, we have $a_1 < \frac{1}{3}$, and $\frac{2}{3} < a_3$. We now show that $\frac{2}{3} < a_2$. Suppose that $a_2 \leq \frac{1}{3}$. Then, notice that the Voronoi region of a_2 does not contain any point from J_2. If it does, then we must have $\frac{1}{2}(a_2 + a_3) > \frac{2}{3}$ implying $a_3 > \frac{4}{3} - a_2 \geq \frac{4}{3} - \frac{1}{3} = 1$, which leads to a contradiction as we know $a_3 < 1$. Thus, if $a_2 \leq \frac{1}{3}$, then $a_3 = E(X : X \in J_2) = \frac{5}{6}$, and so

$$V_3 \geq \int_{J_2}(x-\frac{5}{6})^2 dP = \frac{2}{3}\int_{J_2}(x-\frac{5}{6})^2 dP_2 = \frac{1}{162} > \frac{1}{216} \geq V_3,$$

which leads to a contradiction. Hence, we can assume that $\frac{1}{3} < a_2$. Next, suppose that $\frac{1}{3} < a_2 < \frac{2}{3}$. Then, the following two cases can arise:

Case 1. $\frac{1}{3} < a_2 \leq \frac{1}{2}$.

Then, the Voronoi region of a_2 must contain points from J_2, i.e., $\frac{1}{2}(a_2 + a_3) \geq \frac{2}{3}$ implying $a_3 \geq \frac{4}{3} - a_2 \geq \frac{4}{3} - \frac{1}{2} = \frac{5}{6}$, otherwise the quantization error can be strictly reduced by moving the point a_2 to $\frac{1}{3}$. We see that

$$\begin{aligned}
&\int_{J_2} \min_{a \in \{a_2, a_3\}} (x-a)^2 dP \geq \int_{J_2} \min_{a \in \{\frac{1}{2}, a_3\}} (x-a)^2 dP \\
&= \frac{2}{3}\Big(\int_{[\frac{2}{3}, \frac{1}{2}(\frac{1}{2}+a_3)]} (x-\frac{1}{2})^2 dP_2 + \int_{[\frac{1}{2}(\frac{1}{2}+a_3), 1]} (x-a_3)^2 dP_2\Big) \\
&= -\frac{a_3^3}{2} + \frac{7a_3^2}{4} - \frac{15a_3}{8} + \frac{833}{1296}
\end{aligned}$$

which is minimum when $a_3 = \frac{5}{6}$, and the minimum value is $\frac{1}{162}$. Thus, we have

$$V_3 \geq \frac{1}{162} > \frac{1}{216} \geq V_3,$$

which leads to a contradiction.

Case 2. $\frac{1}{2} \leq a_2 < \frac{2}{3}$.

Then, we must have $\frac{1}{2}(a_1 + \frac{1}{2}) \leq \frac{1}{3}$ implying $a_1 \leq \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$. Notice that

$$\begin{aligned}
V_3 &= \int_{J_1} \min_{a \in \{a_1, a_2\}} (x-a)^2 dP + \int_{J_2} \min_{a \in \{a_2, a_3\}} (x-a)^2 dP \\
&\geq \frac{1}{3} \int_{J_1} \min_{a \in \{a_1, \frac{1}{2}\}} (x-a)^2 dP_1 + \frac{2}{3} \int_{J_2} \min_{b \in \{\frac{2}{3}, a_3\}} (x-b)^2 dP \\
&= \frac{a_1^3}{4} + \frac{a_1^2}{8} - \frac{a_1}{16} + \frac{23}{2592} - \frac{a_3^3}{2} + \frac{5a_3^2}{3} - \frac{16a_3}{9} + \frac{50}{81}.
\end{aligned}$$

Under the condition $a_1 < \frac{1}{6}$, and $\frac{1}{2} \leq a_2 < \frac{2}{3} < a_3 < 1$, we see that the minimum value of the following expression

$$\frac{a_1^3}{4} + \frac{a_1^2}{8} - \frac{a_1}{16} + \frac{23}{2592} - \frac{a_3^3}{2} + \frac{5a_3^2}{3} - \frac{16a_3}{9} + \frac{50}{81}$$

is $\frac{17}{2916}$, which occurs when $a_1 = \frac{1}{6}$ and $a_3 = \frac{8}{9}$. Thus, in this case, we have $V_3 \geq \frac{17}{2916} \geq \frac{1}{216} > V_3$, which is a contradiction.

Hence, by Case 1 and Case 2, we can conclude that $\frac{2}{3} < a_2$, i.e., $a_1 < \frac{1}{3}$ and $\frac{2}{3} < a_2 < a_3 < 1$. Since the Voronoi region of a_1 does not contain any point from J_2 and the Voronoi region of a_2

does not contain any point from J_1, by Theorem I.2, we have $a_1 = \frac{1}{6}$, $a_2 = \frac{3}{4}$, and $a_3 = \frac{11}{12}$ with the quantization error for three means is given by $V_3 = \frac{1}{216}$. Thus, the proof of the lemma is complete. □

Remark II.7. Proceeding similarly as in Lemma II.6, we can show that the set $\{\frac{1}{12}, \frac{1}{4}, \frac{3}{4}, \frac{11}{12}\}$ forms an optimal set of four-means with quantization error $V_4 = \frac{1}{432}$.

Proposition II.8. *For $n \geq 2$ let α_n be an optimal set of n-means for P. Then, α_n does not contain any point from the open interval $(\frac{1}{3}, \frac{2}{3})$. Moreover, the Voronoi region of any point in $\alpha_n \cap J_1$ does not contain any point from J_2, and the Voronoi region of any point in $\alpha_n \cap J_2$ does not contain any point from J_1.*

Proof. By Corollary II.4, Lemma II.6, and Remark II.7, we can conclude that the proposition is true for $2 \leq n \leq 4$. We now prove the proposition for $n \geq 5$. Let $\alpha_n := \{a_1, a_2, \cdots, a_n\}$ be an optimal set of n-means for $n \geq 5$ such that $0 < a_1 < a_2 < \cdots < a_n < 1$. First, we show that α_n does not contain any point from the open interval $(\frac{1}{3}, \frac{2}{3})$. Using Theorem I.2, we see that the distortion error due to the set $\beta := \{\frac{1}{12}, \frac{1}{4}, \frac{13}{18}, \frac{5}{6}, \frac{17}{18}\}$ is given by

$$\int \min_{a\in\beta}(x-a)^2 dP = \frac{17}{11664}.$$

Since V_n is the quantization error for n-means, where $n \geq 5$, we have

$$V_n \leq V_5 \leq \frac{17}{11664}.$$

By Proposition II.3, we have $\alpha_n \cap J_1 \neq \emptyset$, and $\alpha_n \cap J_2 \neq \emptyset$. Let k be the largest positive integer such that $a_k \leq \frac{1}{3}$. For the sake of contradiction, assume that α_n contains a point from the open interval $(\frac{1}{3}, \frac{2}{3})$. Then, by Proposition II.5, we must have $a_{k+1} \in (\frac{1}{3}, \frac{2}{3})$, and $\frac{2}{3} \leq a_{k+2}$. The following two cases can arise:

Case 1. $\frac{1}{3} < a_{k+1} \leq \frac{1}{2}$.

Then, the Voronoi region of a_{k+1} must contain points from J_2, i.e., $\frac{1}{2}(a_{k+1} + a_{k+2}) \geq \frac{2}{3}$

implying $a_{k+2} \geq \frac{4}{3} - a_{k+1} \geq \frac{4}{3} - \frac{1}{2} = \frac{5}{6}$, otherwise the quantization error can be strictly reduced by moving the point a_{k+1} to $\frac{1}{3}$. Then,

$$V_n \geq \int_{[\frac{2}{3},\frac{5}{6}]} (x - \frac{5}{6})^2 dP = \frac{1}{324} > \frac{17}{11664} \geq V_n,$$

which is a contradiction.

Case 2. $\frac{1}{2} \leq a_{k+1} < \frac{2}{3}$.

Then, we must have $\frac{1}{2}(a_k + a_{k+1}) \leq \frac{1}{3}$ implying $a_k \leq \frac{2}{3} - a_{k+1} \leq \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$, and so

$$V_n \geq \int_{[\frac{1}{6},\frac{1}{3}]} (x - \frac{1}{6})^2 dP = \frac{1}{648} > \frac{17}{11664} \geq V_n,$$

which leads to a contradiction.

By Case 1 and Case 2, we conclude that α_n does not contain any point from the open interval $(\frac{1}{3}, \frac{2}{3})$. Thus, $\frac{2}{3} \leq a_{k+1}$. To complete the proof, assume that the Voronoi region of a_k contains points from J_2. Then, $\frac{1}{2}(a_k + a_{k+1}) > \frac{2}{3}$ implying $a_{k+1} > \frac{4}{3} - a_k \geq \frac{4}{3} - \frac{1}{3} = 1$, which is a contradiction. Similarly, we can show that if the Voronoi region of a_{k+1} contains points from J_1, then a contradiction arises. Thus, the proof of the proposition is complete. □

We are now ready to prove the following theorem.

Theorem II.9. *Let α_n be an optimal set of n-means for P for $n \geq 2$. Let card$(\alpha_n \cap J_1) = k$. Then, α_n contains k elements from J_1, and $(n-k)$ elements from J_2, and $\alpha_n(P) = \alpha_k(P_1) \cup \alpha_{n-k}(P_2)$ with quantization error*

$$V_n(P) = \frac{1}{324}\Big(\frac{1}{k^2} + \frac{2}{(n-k)^2}\Big).$$

Proof. By Proposition II.3, we have $\alpha_n \cap J_1 \neq \emptyset$ and $\alpha_n \cap J_2 \neq \emptyset$. Thus, there exist two positive integers n_1 and n_2 such that card$(\alpha_n \cap J_1) = n_1$, and card$(\alpha_n \cap J_2) = n_2$. Since by Proposition II.8, α_n does not contain any point from the open interval $(\frac{1}{3}, \frac{2}{3})$, we have $n = n_1 + n_2$. Hence, by taking $n_1 = k$, we see that α_n contains k elements from J_1, and $(n-k)$ elements from J_2. Again, by Proposition II.8, we know that the Voronoi region of any point in $\alpha_n \cap J_1$ does not contain any point from J_2, and the Voronoi region of any point from $\alpha_n \cap J_2$ does not contain any point from J_1. This implies the fact that $\alpha_n(P) = \alpha_k(P_1) \cup \alpha_{n-k}(P_2)$, and the corresponding quantization error is given by

$$V_n(P) = \frac{1}{3}V_k(P_1) + \frac{2}{3}V_{n-k}(P_2) = \frac{1}{324}\Big(\frac{1}{k^2} + \frac{2}{(n-k)^2}\Big).$$

Thus, the proof of the theorem is complete. □

Remark II.10. Let k be the positive integer as stated in Theorem II.9. By Theorem I.2, $\alpha_k(P_1)$ and $\alpha_{n-k}(P_2)$ are known. Thus, once k is known, we can easily determine the optimal sets of n-means and the nth quantization errors for all $n \in \mathbb{N}$ with $n \geq 2$. For $n \geq 2$, consider the real valued function

$$F(n,x) := \frac{1}{324}\Big(\frac{1}{x^2} + \frac{2}{(n-x)^2}\Big)$$

defined in the domain $1 \leq x \leq n-1$. Notice that $F(n,x)$ is concave upward, and so $F(n,x)$ attains its minimum at a unique x in the interval $[1, n-1]$. Thus, we can say that for a given positive integer $n \geq 2$, there exists a unique positive integer $k := k(n)$, depending on n, for which $F(n,x)$ is minimum if x ranges over the positive integers in the interval $[1, n-1]$. In other words, $k := k(n)$ is the positive integer such that $1 \leq k \leq n-1$, and satisfies:

$$F(n,k) = \min\{F(n,j) : j \in \mathbb{N}, 1 \leq j \leq n-1\}. \tag{2.1}$$

Using the expression (2.1), for any positive integer $n \geq 2$ we can easily determine $k(n)$. In Table 2.1, we give the values of $k(n)$ for $8 \leq n \leq 79$.

Remark II.11. In the statement of the following proposition, the two decimal numbers are the rational approximations of two real numbers that minimize the expression (2.2).

Proposition II.12. *For the mixed distribution $P := \frac{1}{100}P_1 + \frac{99}{100}P_2$ the optimal set of two means is given by $\{0.731517, 0.910506\}$ with quantization error $V_2 = \frac{1314-53\sqrt{53}}{163350}$.*

Proof. Since $\frac{2}{3} < \frac{1}{2}(0.731517 + 0.910506) = 0.821012$, the distortion error due to the set $\beta :=$

$\{0.731517, 0.910506\}$ is given by

$$\int_{J_1}(x-0.731517)^2dP+\int_{[\frac{2}{3},0.821012]}(x-0.731517)^2dP+\int_{[0.821012,1]}(x-0.910506)^2dP=0.005682.$$

Let $\alpha := \{a_1, a_2\}$ be an optimal set of two-means with $0 < a_1 < a_2 < 1$. Since V_2 is the quantization error for two-means, we have $V_2 \leq 0.005682$. If $a_2 < \frac{2}{3}$, then

$$V_2 \geq \int_{J_2}(x-\frac{2}{3})^2dP = \frac{22}{621} > V_2,$$

which leads to a contradiction. Hence, $\frac{2}{3} < a_2$. Assume that $a_1 \leq \frac{1}{2}$. Notice that $E(X : X \in J_2) = \frac{5}{6}$, and $\frac{1}{2}(\frac{1}{2}+\frac{5}{6}) = \frac{2}{3}$, and so by Proposition I.1, we can assume that $a_1 = E(X : X \in J_1) = \frac{1}{6}$, and $a_2 = E(X : X \in J_2) = \frac{5}{6}$ yielding

$$V_2 = \int_{J_1}(x-\frac{1}{6})^2dP+\int_{J_2}(x-\frac{5}{6})^2dP = \frac{1}{108} > V_2,$$

which is a contradiction. Hence, we can assume that $\frac{1}{2} < a_1$. Then, the Voronoi region of a_1 must contain points from J_2, i.e., $\frac{2}{3} < \frac{1}{2}(a_1+a_2)$. Thus, the distortion error is given by

$$\frac{1}{100}\int_0^{\frac{1}{3}}3(x-a_1)^2\,dx+\frac{99}{100}\Big(\int_{\frac{2}{3}}^{\frac{a_1+a_2}{2}}3(x-a_1)^2\,dx+\int_{\frac{a_1+a_2}{2}}^{1}3(x-a_2)^2\,dx\Big),$$

which upon simplification yields

$$\frac{1}{10800}(8019a_1^3+27a_1^2(297a_2-788)-9a_1\left(891a_2^2-1580\right)-8019a_2^3+32076a_2^2-32076a_2+7528), \tag{2.2}$$

the minimum value of which occurs at $a_1 = 0.731517$ and $a_2 = 0.910506$, and the minimum value is $\frac{1314-53\sqrt{53}}{163350}$. Thus, the set $\{0.731517, 0.910506\}$ forms an optimal set of two-means for P with quantization error $V_2 = \frac{1314-53\sqrt{53}}{163350}$, which is the proposition. □

Remark II.13. In Proposition II.3, we have shown that if α_n is an optimal set of n-means for the mixed distribution $P := \frac{1}{3}P_1 + \frac{2}{3}P_2$, then, $\alpha_n \cap J_1 \neq \emptyset$ and $\alpha_n \cap J_2 \neq \emptyset$ for all $n \geq 2$. Proposition II.12 shows that the set $\{0.731517, 0.910506\}$ forms an optimal set of two-means for the mixed distribution $P := \frac{1}{100}P_1 + \frac{99}{100}P_2$. Notice that here both the points 0.731517 and 0.910506 lie in the interval J_2. Thus, we see that the condition $\alpha_n \cap J_1 \neq \emptyset$ and $\alpha_n \cap J_2 \neq \emptyset$ are not true for all mixed distributions $P := pP_1 + (1-p)P_2$ with $0 < p < \frac{1}{2}$ and $n \geq 2$.

Let us now end the discussion with some open questions given in the following remark.

Remark II.14. By Table 2.1, we observe that there are some positive integers n for which

$$\text{either } k(n+9) = k(n+10) = k(n+11), \text{ or } k(n+7) = k(n+8) = k(n+9); \tag{2.3}$$

and for all $n \in \mathbb{N}$ with $n \geq 2$, either $k(n) = k(n-1)$, or $k(n) = k(n+1)$. It is still not known whether there is a subsequence (n_j) of positive integers for which the expression (2.3) is true. It is also not known whether there is a closed form of the sequence $(k(n))$, which will help us to determine $k(n)$ for any positive integer $n \geq 2$, without the help of expression given in (2.1).

n	$k(n)$	n	$k(n)$	n	$k(n)$	n	$k(n)$	n	$k(n)$	n	$k(n)$
8	4	20	9	32	14	44	19	56	25	68	30
9	4	21	9	33	15	45	20	57	25	69	31
10	4	22	10	34	15	46	20	58	26	70	31
11	5	23	10	35	15	47	21	59	26	71	31
12	5	24	11	36	16	48	21	60	27	72	32
13	6	25	11	37	16	49	22	61	27	73	32
14	6	26	12	38	17	50	22	62	27	74	33
15	7	27	12	39	17	51	23	63	28	75	33
16	7	28	12	40	18	52	23	64	28	76	34
17	8	29	13	41	18	53	23	65	29	77	34
18	8	30	13	42	19	54	24	66	29	78	35
19	8	31	14	43	19	55	24	67	30	79	35

Table 2.1: Values of $k(n)$ for $8 \leq n \leq 79$.

BIBLIOGRAPHY

[AW] E.F. Abaya and G.L. Wise, *Some remarks on the existence of optimal quantizers*, Statistics & Probability Letters, Volume 2, Issue 6, December 1984, Pages 349-351.

[BW] J.A. Bucklew and G.L. Wise, *Multidimensional asymptotic quantization theory with rth power distortion measures*, IEEE Transactions on Information Theory, 1982, Vol. 28 Issue 2, 239-247.

[CR] D. Comez and M.K. Roychowdhury, *Quantization for uniform distributions on stretched Sierpinski triangles*, Monatshefte für Mathematik, Volume 190, Issue 1, 79-100 (2019).

[DR1] C.P. Dettmann and M.K. Roychowdhury, *Quantization for uniform distributions on equilateral triangles*, Real Analysis Exchange, Vol. 42(1), 2017, pp. 149-166.

[DR2] C.P. Dettmann and M.K. Roychowdhury, *An algorithm to compute CVTs for finitely generated Cantor distributions*, Southeast Asian Bulletin of Mathematics (2021) 45: 173-188.

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

[GKL] R.M. Gray, J.C. Kieffer and Y. Linde, *Locally optimal block quantizer design*, Information and Control, 45 (1980), pp. 178-198.

[GL1] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics 1730, Springer, Berlin, 2000.

[GL2] A. György and T. Linder, *On the structure of optimal entropy-constrained scalar quantizers*, IEEE transactions on information theory, vol. 48, no. 2, February 2002.

[GL3] S. Graf and H. Luschgy, *The Quantization of the Cantor Distribution*, Math. Nachr., 183, 113-133 (1997).

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

[HMRT] J. Hansen, I. Marquez, M.K. Roychowdhury, and E. Torres, *Quantization coefficients for uniform distributions on the boundaries of regular polygons*, Statistics & Probability Letters, Volume 173, June 2021, 109060.

[L] L.J. Lindsay, *Quantization dimension for probability distributions*, PhD dissertation, 2001, University of North Texas, Texas, USA.

[L1] L. Roychowdhury, *Optimal quantization for nonuniform Cantor distributions*, Journal of Interdisciplinary Mathematics, Vol 22 (2019), pp. 1325-1348.

[PRRSS] G. Pena, H. Rodrigo, M.K. Roychowdhury, J. Sifuentes, and E. Suazo, *Quantization for uniform distributions on hexagonal, semicircular, and elliptical curves*, Journal of Optimization Theory and Applications, (2021) 188: 113-142.

[R1] M.K. Roychowdhury, *Quantization and centroidal Voronoi tessellations for probability measures on dyadic Cantor sets*, Journal of Fractal Geometry, 4 (2017), 127-146.

[R2] M.K. Roychowdhury, *Optimal quantizers for some absolutely continuous probability measures*, Real Analysis Exchange, Vol. 43(1), 2017, pp. 105-136.

[R3] M.K. Roychowdhury, *Optimal quantization for the Cantor distribution generated by infinite similitudes*, Israel Journal of Mathematics 231 (2019), 437-466.

[R4] M.K. Roychowdhury, *Least upper bound of the exact formula for optimal quantization of some uniform Cantor distributions*, Discrete and Continuous Dynamical Systems- Series A, Volume 38, Number 9, September 2018, pp. 4555-4570.

[R5] M.K. Roychowdhury, *Center of mass and the optimal quantizers for some continuous and discrete uniform distributions*, Journal of Interdisciplinary Mathematics, Vol. 22 (2019), No. 4, pp. 451-471.

[R6] M.K. Roychowdhury, *Optimal quantization for mixed distributions*, to appear, Real Analysis Exchange.

[RR1] J. Rosenblatt and M.K. Roychowdhury, *Optimal quantization for piecewise uniform distributions*, Uniform Distribution Theory 13 (2018), no. 2, 23-55.

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

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

BIOGRAPHICAL SKETCH

Eduardo Orozco grew up in San Perlita and graduated from Rio Hondo High School in Texas. He received a Master of Science degree in mathematics from the University of Texas Rio Grande Valley, May 2021, where he also earned a Bachelor of Science degree in mathematics, May 2018. During 2019-2020 he worked as a basketball official. Eduardo's email address is eorozc549@gmail.com.