自然数集上一组离散分布的量化

[1]Juan Gomez, [2]Haily Martinez, [3]Mrinal K. Roychowdhury, [4]Alexis Salazar, 和 [5]Daniel J. Vallez

arXiv:2305.02372v1 [math.PR] 3 May 2023

Abstract. 概率论中的量化方案旨在寻找给定概率分布的最佳近似，即寻找一个支撑集为有限点集的概率分布。本文首先陈述并证明一个定理，然后提出一个猜想，并通过若干例子验证该猜想。假设猜想成立，对于自然数集上一组离散分布，我们计算了所有正整数 n对应的最优 n均值集以及第 n个量化误差。此外，我们还计算了量化维数。

1. Introduction

量化最常见的形式是舍入。其目的是减少表示空间的基数，尤其是在输入数据为实数值时。它在通信、信息论、信号处理和数据压缩中有广泛应用（参见 [GG, GL1, GL2, GN, P, Z1, Z2]）。令 $\mathbb{R}^d$ 表示配备欧几里得范数 $\|\cdot\|$ 的 d 维欧几里得空间，并令 P 为 $\mathbb{R}^d$ 上的博雷尔概率测度。那么，对于 P，相对于欧几里得距离平方，其 n 阶量化误差定义为

$$V_n := V_n(P) = \inf\Big\{V(P;\alpha) : \alpha \subset \mathbb{R}^d, 1 \leq \operatorname{card}(\alpha) \leq n\Big\},$$

其中，$V(P;\alpha) = \int \min_{a\in\alpha} \|x-a\|^2 dP(x)$ 表示集合 α 相对于概率分布 P 所产生的失真误差；对于集合 A，$\operatorname{card}(A)$ 表示集合 A 的基数。使下确界得以取到且至多包含 n 个点的集合 α 称为一个最优 n-均值集，记作 $\alpha_n := \alpha_n(P)$。最优集的元素也称为最优量化器。已知，对于博雷尔概率测度 P，若其支撑集包含无穷多个元素且 $\int \|x\|^2 dP(x)$ 有限，则最优 n-均值集总是存在，并且恰好有 n 个元素[AW, GKL, GL1, GL2]。该数

$$D(P) := \lim_{n\to\infty} \frac{2\log n}{-\log V_n(P)}, \tag{1}$$

如果存在，则称为 P 的量化维数。量化维数衡量指定误差度量随着 n 趋于无穷而趋于零的速度。对于有限集 $\alpha \subset \mathbb{R}^d$和 $a \in \alpha$，我们用 $M(a|\alpha)$ 表示 $\mathbb{R}^d$ 中距离 a 最近的所有元素组成的集合，其中 α 中的元素均参与比较，即 $M(a|\alpha) = \{x \in \mathbb{R}^d : \|x-a\| = \min_{b\in\alpha}\|x-b\|\}$. $M(a|\alpha)$ 称为由 $a \in \alpha$ 生成的沃罗诺伊区域。另一方面，集合 $\{M(a|\alpha) : a \in \alpha\}$ 称为 $\mathbb{R}^d$ 相对于集合 α 的沃罗诺伊图或沃罗诺伊镶嵌。以下命题进一步说明了最优 n-均值集所生成的沃罗诺伊区域（参见 [GG, GL2]）。

命题 1.1. 设 α 是 n-均值的一个最优集，$a \in \alpha$，且 $M(a|\alpha)$ 是由 $a \in \alpha$ 生成的沃罗诺伊区域，即，

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

于是，对于每个 $a \in \alpha$，

(i) $P(M(a|\alpha)) > 0$,

(ii) $P(\partial M(a|\alpha)) = 0$,

2010 年数学主题分类。60Exx，94A34。关键词和短语。离散分布，最优集，量化误差。

(iii) $a = E(X : X \in M(a|\alpha))$，以及 (iv) P-几乎必然地，集合 $\{M(a|\alpha): a \in \alpha\}$ 构成 $\mathbb{R}^d$ 的沃罗诺伊划分。

由上述命题可知，如果 α 是 P 的一个最优 n-均值集，那么每个 $a \in \alpha$ 都是随机变量 X 在给定 X 取值于 a 的沃罗诺伊区域这一条件下的条件期望。有时，我们也将这样的 $a \in \alpha$ 称为其自身沃罗诺伊区域的质心。对此感兴趣的读者可参见 [DFG、DR、R1]。

如果对于所有 $j \in \mathbb{N}$ 和 $j \geq 0$，都有 $0 < p_j < 1$，且满足 $\sum_{j\geq 1} p_j = 1$，则称向量 $(p_1, p_2, p_3, \cdots)$ 为概率分布。请注意，$(p_1, p_2, p_3, \cdots)$ 可以是有限向量，也可以是无限向量；有限向量是指其坐标个数有限的向量，否则称为无限向量。

对于实数集 $\mathbb{R}$ 上的博雷尔概率测度 P，设 U 是满足 $P(U) = 0$ 的 $\mathbb{R}$ 的最大开子集，则 $\mathbb{R} \setminus U$ 称为概率测度 P 的支撑集。例如，掷一次均匀骰子时，P 是实数轴上的一个博雷尔概率测度，其

$$\text{support}(P) = \{1, 2, 3, 4, 5, 6\},$$

相应的概率分布为（$\frac{1}{6}$，$\frac{1}{6}$，$\frac{1}{6}$，$\frac{1}{6}$，$\frac{1}{6}$，$\frac{1}{6}$）。当均匀硬币抛掷两次时，P 是 $\mathbb{R}^2$ 上的一个博雷尔概率测度，其

$$\text{support}(P) = \{(1, 1), (1, 2), (2, 1), (2, 2)\},$$

相应的概率分布为（$\frac{1}{4}$，$\frac{1}{4}$，$\frac{1}{4}$，$\frac{1}{4}$），其中 1 表示“正面”，2 表示“反面”。

定义 1.2。设（p_1，p_2，$\cdots$，p_{k-1}）是集合 $\{\frac{1}{2}, \frac{1}{2^2}, \cdots, \frac{1}{2^{k-1}}\}$ 的一个排列，其中 $k \in \mathbb{N}$：$= \{1, 2, \cdots\}$，且 $k \geq 2$。在实数集 $\mathbb{R}$ 上定义一个概率测度 P，其支撑集为自然数集 $\mathbb{N}$，如下所示：

$$P := \sum_{j=1}^{k-1} p_j \delta_j + \sum_{j=k}^{\infty} \frac{1}{2^j} \delta_j,$$

其中，对于 $x \in \mathbb{R}$，函数 δ_x 表示狄拉克测度，也就是说，对于任意子集 $A \subseteq \mathbb{R}$，我们有 $\delta_x(A) = 1$ 当 $x \in A$ 时成立，否则为零。

现在陈述以下定理和猜想。

定理 1.3. 设 $P := \sum_{j=1}^{k-1} p_j \delta_j + \sum_{j=k}^{\infty} \frac{1}{2^j} \delta_j$ 为定义 1.2 所定义的概率测度。设 $\{a_1, a_2, \cdots, a_{n-3}, a_{n-2}, a_{n-1}, a_n\}$ 为具有 $n \geq k + 2$ 的 n-均值最优集。假设 $a_1 = 1, a_2 = 2, \cdots, a_{n-3} = n - 3$。那么，要么 $a_{n-2} = n - 2$，$a_{n-1} = Av[n - 1, n], a_n = Av[n + 1, \infty)$，要么 $a_{n-2} = Av[n - 2, n - 1], a_{n-1} = Av[n, n + 1], a_n = Av[n + 2, \infty)$，其量化误差为 $V_n = \frac{2^{3-n}}{3}$，其中对于任意 $k, \ell \in \mathbb{N}$，$Av[k, \ell]$ 和 $Av[k, \infty)$ 均在下一节中定义。

例 1.4. 设 $(\frac{1}{2^3}, \frac{1}{2^2}, \frac{1}{2})$ 是集合 $\{\frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}\}$ 的一个排列。记

$$P := \frac{1}{2^3}\delta_1 + \frac{1}{2^2}\delta_2 + \frac{1}{2}\delta_3 + \sum_{j=4}^{\infty} \frac{1}{2^j}\delta_j.$$

那么，P 是 $\mathbb{R}$ 上以集合 $\mathbb{N}$ 为支撑集的博雷尔概率测度。假设 $\{a_1, a_2, \cdots, a_n\}$ 是 $n = 6$ 的一个最优 n 均值集。如果 $a_1 = 1$、$a_2 = 2$ 和 $a_3 = 3$，那么根据定理 1.3，集合 $\{a_4, a_5, a_6\}$ 必须等于集合 $\{4, Av[5, 6], Av[7, \infty)\}$、集合 $\{Av[4, 5], Av[6, 7], Av[8, \infty)\}$，且量化误差为 $V_6 = \frac{1}{24}$。

猜想 1.5. 设 $P := \sum_{j=1}^{k-1} p_j \delta_j + \sum_{j=k}^{\infty} \frac{1}{2^j} \delta_j$ 是由定义 1.2 定义的概率测度。设 $\{a_1, a_2, a_3, \cdots, a_n\}$ 是一个最优 n 均值集，其中 $n \geq k + 2$ 满足 $a_1 < a_2 < \cdots < a_n$。那么，$a_1 = 1$、$a_2 = 2$、$\cdots$、$a_{n-3} = n - 3$。

本文首先给出定理 1.3 的完整证明。然后，我们通过注释 3.2.2 和注释 3.3.2 中提到的两个离散分布验证该猜想。在假设该猜想成立的条件下，我们计算这两个分布的 n-均值的最优集以及第 n 阶量化误差。

所有 $n \in \mathbb{N}$ 的离散分布。一旦知道量化误差，就可以轻松计算量化维数，见命题 3.3.3。此外，在最后一节中，我们给出了命题 5.1。根据该命题，我们推断，如果 $(p_1, p_2, \cdots, p_{k-1})$ 不是集合 $\{\frac{1}{2}, \frac{1}{2^2}, \cdots, \frac{1}{2^{k-1}}\}$ 的排列，其中 $k \in \mathbb{N} := \{1, 2, \cdots\}$，且 $k \geq 2$，那么猜想不成立。目前尚不清楚该猜想的一般性证明。这个问题仍未解决。

2. Preliminaries

令 $\mathbb{N} := \{1, 2, 3, \cdots\}$ 为自然数集。设 $(p_1, p_2, p_3, \cdots)$ 为一个概率分布，其中 $0 < p_j < 1$ 对所有 $j \in \mathbb{N}$ 和 $\sum_{j=1}^{\infty} p_j = 1$ 均成立。设

$$P = \sum_{j=1}^{\infty} p_j \delta_j,$$

其中 δ_j 是定义 1.2 中给出的狄拉克测度。然后，P 是 $\mathbb{R}$ 的实数集合上的离散概率测度，其支撑集是与概率分布（p_1、p_2、p_3、$\cdots$）相关联的自然数集合 $\mathbb{N}$。事实上，如果 X 是与概率测度P相关联的随机变量，且 f 是概率质量函数，那么我们有

$$P(X = j) = f(j) = p_j.$$

定义以下记号：对于 k、$\ell \in \mathbb{N}$，其中 $k \leq \ell$，记作

$$[k, \ell] := \{n : n \in \mathbb{N} \text{ and } k \leq n \leq \ell\}, \text{ and } [k, \infty) := \{n : n \in \mathbb{N} \text{ and } n \geq k\}.$$

此外，记为

$$Av[k, \ell] := E\Big(X : X \in [k, \ell]\Big) = \frac{\sum_{n=k}^{\ell} p_n n}{\sum_{n=k}^{\ell} p_n}, \; Av[k, \infty) := E\Big(X : X \in [k, \infty)\Big) = \frac{\sum_{n=k}^{\infty} p_n n}{\sum_{n=k}^{\infty} p_n},$$

$$Er[k, \ell] := \sum_{n=k}^{\ell} p_n \Big(n - Av[k, \ell]\Big)^2, \text{ and } Er[k, \infty) := \sum_{n=k}^{\infty} p_n \Big(n - Av[k, \infty)\Big)^2.$$

注意，$E(X) := E(X : X \in \text{supp}(P)) = \sum_{n=1}^{\infty} p_n n$，因此单均值的最优集是集合$\{\sum_{n=1}^{\infty} p_n n\}$，其量化误差为

$$V(P) = \sum_{n=1}^{\infty} p_n (n - E(X))^2.$$

在以下各节中，我们将给出本文的主要结果。

3. Proof of Theorem 1.3 and verifications of Conjecture 1.5

在本节的以下各小节中，我们首先证明定理 1.3，然后通过两个不同的例子验证猜想 1.5 成立。接着，我们陈述并证明命题 3.3.3，该命题给出了概率测度 P 的量化维数。

3.1. 定理 1.3 的证明。设 $n \geq k + 2$，其中 $k \geq 2$。由集合β:
$= \{1, 2, \cdots, n-3, n-2, Av[n-1, n], Av[n+1, \infty)\}$引起的失真误差为

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

由于 V_n 是 n-均值的量化误差，因此我们有 $V_n \leq \frac{2^{3-n}}{3}$。设 $\alpha := \{a_1, a_2, a_3, \cdots, a_n\}$ 是 n-均值的最优集，其中 $1 \leq a_1 < a_2 < a_3 < \cdots < a_n < \infty$。假设 $a_1 = 1, a_2 = 2, \cdots, a_{n-3} = n - 3$。那么，$a_{n-2}$的沃罗诺伊区域必定包含元素 $n - 2$。假设 a_{n-2}的沃罗诺伊区域包含集合 $\{n-2, n-1, n\}$。那么，

$$V_n \geq Er[n-2, n] = \frac{13}{7} 2^{1-n} > \frac{2^{3-n}}{3} \geq V_n,$$

这就产生了矛盾。因此，我们可以假设 a_{n-2} 的沃罗诺伊区域仅包含集合$\{n-2\}$ 或集合 $\{n-2, n-1\}$。下面考虑以下两种情况:

情形 1. a_{n-2} 的沃罗诺伊区域仅包含集合 $\{n-2\}$。

那么， a_{n-1} 的沃罗诺伊区域必定包含元素 $n-1$。假设 a_{n-1} 的沃罗诺伊区域包含集合 $\{n-1, n, n+1, n+2\}$。那么,

$$V_n \geq \frac{97}{15}\, 2^{-n-1} > V_n,$$

这就产生了矛盾。假设 a_{n-1} 的沃罗诺伊区域仅包含集合 $\{n-1, n, n+1\}$。那么，由于 a_n 的沃罗诺伊区域包含集合 $\{k:\ k \geq n+2\}$,

$$V_n \geq Er[n-1, n+1] + Er[n+2, \infty) = \frac{5}{7}2^{2-n} > V_n,$$

这导致矛盾。接下来，假设 a_{n-1} 的沃罗诺伊区域仅包含集合$\{n-1\}$。那么， a_n 的沃罗诺伊区域包含集合 $\{k:\ k \geq n\}$，从而得到

$$V_n = Er[n, \infty) = 2^{2-n} > V_n,$$

这就产生了矛盾。因此， a_{n-1} 的沃罗诺伊区域仅包含集合$\{n-1, n\}$，进而， a_n 的沃罗诺伊区域仅包含集合 $\{k:\ k \geq n+1\}$。因此，在这种情况下，我们有 $a_{n-2} = n-2, a_{n-1} = Av[n-1, n], a_n = Av[n+1, \infty)$，量化误差为 $V_n = \frac{2^{3-n}}{3}$.

情形 2。 a_{n-2} 的沃罗诺伊区域仅包含集合 $\{n-2, n-1\}$.

那么， a_{n-1} 的沃罗诺伊区域必须包含元素 n。假设a_{n-1} 的沃罗诺伊区域包含集合 $\{n, n+1, n+2, n+3\}$。那么,

$$V_n \geq Er[n-2, n-1] + Er[n, n+3] = \frac{59}{5}2^{-n-2} > \frac{2^{3-n}}{3} \geq V_n,$$

这会导致矛盾。假设 a_{n-1} 的沃罗诺伊区域仅包含集合 $\{n, n+1, n+2\}$。那么，由于 a_n 的沃罗诺伊区域包含集合 $\{k:\ k \geq n+3\}$,

$$V_n \geq Er[n-2, n-1] + Er[n, n+2] + Er[n+3, \infty) = \frac{29}{21}2^{1-n} > V_n,$$

这会导致矛盾。因此，我们可以假设， a_{n-1} 的沃罗诺伊区域仅包含集合 $\{n\}$或集合 $\{n, n+1\}$。如果 a_{n-1} 的沃罗诺伊区域仅包含集合 $\{n\}$，那么

$$V_n \geq Er[n-2, n-1] + Er[n+1, \infty) = \frac{5}{3}2^{1-n} > V_n,$$

这会导致矛盾。因此，我们可以假设， a_{n-1} 的沃罗诺伊区域仅包含集合 $\{n, n+1\}$，而 a_n 的沃罗诺伊区域仅包含集合 $\{k:\ k \geq n+2\}$ 得到$a_{n-2} = Av[n-2, n-1]$、 $a_{n-1} = Av[n, n+1]$、 $a_n = Av[n+2, \infty)$，其量化误差为 $V_n = \frac{2^{3-n}}{3}$。

情形 1 和情形 2 合起来给出了所有正整数 n对应的 n-均值最优集和第 n个量化误差。因此，定理 定理 1.3 的证明完成。□

3.2. 当 $(p_1, p_2, p_3, p_4, \cdots) = (\frac{1}{2^2}, \frac{1}{2}, \frac{1}{2^3}, \frac{1}{2^4}, \frac{1}{2^5}, \cdots)$ 时对猜想 1.5 的验证。在这种情况下，实数集合 $\mathbb{R}$ 上概率测度 P 的概率质量函数 f 为

$$f(j) = \begin{cases} \frac{1}{2^2} & \text{if } j = 1, \\ \frac{1}{2} & \text{if } j = 2, \\ \frac{1}{2^n} & \text{if } j = n \text{ for } n \in \mathbb{N} \text{ and } n \neq 1, 2, \\ 0 & \text{otherwise.} \end{cases}$$

注意，此处 $k = 3$，且 $(p_1, \cdots, p_{k-1}) = (p_1, p_2) = (\frac{1}{2^2}, \frac{1}{2})$，其中 $k \in \mathbb{N}$ 的定义见定义 1.2。现在我们来证明以下命题。

命题 3.2.1。设 $n \geq 5$，且 α_n 是由概率测度 P 给出的 n-均值的最优集、】【

$$P = \frac{1}{2^2}\delta_1 + \frac{1}{2}\delta_2 + \sum_{j=3}^{\infty} \frac{1}{2^j}\delta_j.$$

那么， α_n必须包含集合 $\{1,2,\cdots,(n-3)\}$。

证明。失真误差 由集合 $\beta:=\{1,2,\cdots,(n-2)\ Av[n-1,n],Av[n+1,\infty)\}$产生，定义为 b y

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

由于 V_n是 n均值的量化误差，我们有 $V_n\leq\frac{2^{3-n}}{3}$。令 $\alpha_n:=\{a_1,a_2,\cdots,a_n\}$为一个 n均值的最优集，使得 $1\leq a_1<a_2<\cdots<a_n<\infty$。我们证明 $a_1=1,a_2=2,\cdots,a_{n-3}=n-3$。我们用归纳法证明这一点。注意， a_1的沃罗诺伊区域必须包含元素 1。假设 a_1的沃罗诺伊区域也包含元素 2。那么，

$$V_n>\sum_{j=1}^{2}f(j)(j-Av[1,2])^2=\frac{1}{6}\geq\frac{2^{3-n}}{3}\geq V_n,$$

这就产生了矛盾。因此，我们可以得出结论： a_1的沃罗诺伊区域只包含元素 1，从而得到 $a_1=1$。由此可推知，存在一个正整数 ℓ，其中 $1\leq\ell<n-3$，使得 $a_1=1,a_2=2,\cdots,a_\ell=\ell$。现在我们来证明 $a_{\ell+1}=\ell+1$。注意， $a_{\ell+1}$的沃罗诺伊区域必定包含 $\ell+1$。假设 $a_{\ell+1}$的沃罗诺伊区域也包含元素 $\ell+2$。于是，我们有

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

这就产生了矛盾。因此，我们可以得出结论： $a_{\ell+1}$的沃罗诺伊区域只包含元素 $\ell+1$，从而得到 $a_{\ell+1}=\ell+1$。注意， $2\leq\ell+1\leq n-3$。因此，根据数学归纳法原理，我们可以推得 $a_1=1$, $a_2=2,\cdots,a_{n-3}=n-3$。至此，命题的证明完成。□

备注 3.2.2. 命题 3.2.1 验证了猜想 猜想 1.5 是正确的 .

3.3. 当 $(p_1,p_2,p_3,p_4,\cdots)=(\frac{1}{2^3},\frac{1}{2^2},\frac{1}{2},\frac{1}{2^4},\frac{1}{2^5},\cdots)$时对猜想 1.5 的验证。在此情形下，实数集 $\mathbb{R}$ 上概率测度 P 的概率质量函数 f 为

$$f(j)=\begin{cases}\frac{1}{2^3} & \text{if } j=1,\\ \frac{1}{2} & \text{if } j=3,\\ \frac{1}{2^n} & \text{if } j=n \text{ for } n\in\mathbb{N} \text{ and } n\neq 1,3,\\ 0 & \text{otherwise.}\end{cases}$$

注意，此处 $k=4$且 $(p_1,\cdots,p_{k-1})=(p_1,p_2,p_3)=(\frac{1}{2^3},\frac{1}{2^2},\frac{1}{2})$，其中 $k\in\mathbb{N}$ 的定义见定义 1.2。

现在证明以下命题。

命题 3.3.1. 设 $n\geq 6$，并设 α_n 是概率测度 P 的一个 n-均值最优集，其定义为

$$P=\frac{1}{2^3}\delta_1+\frac{1}{2^2}\delta_2+\frac{1}{2}\delta_3+\sum_{j=4}^{\infty}\frac{1}{2^j}\delta_j.$$

那么， α_n必须包含集合 $\{1,2,\cdots,(n-3)\}$。

证明。失真误差 由于集合 $\beta:=\{1,2,\cdots,(n-2)\ Av[n-1,n],Av[n+1,\infty)\}$ 由 b 给出。 y

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

由于 V_n 是 n-均值的量化误差，我们有 $V_n\leq\frac{2^{3-n}}{3}$。令 $\alpha_n:=\{a_1,a_2,\cdots,a_n\}$ 为 n-均值的最优集，使得 $1\leq a_1<a_2<\cdots<a_n<\infty$。我们证明 $a_1=1,a_2=2,\cdots,a_{n-3}=n-3$。我们通过归纳法证明。 a_1 的沃罗诺伊区域必须包含元素 1。假设 a_1 的沃罗诺伊区域也包含元素 2。注意，该集合的其余元素

自然数集合包含在具有正失真误差的 a_2、a_3、$\cdots$、a_n 的沃罗诺伊区域并集中，从而得到

$$V_n > \sum_{j=1}^{2} f(j)(j - Av[1,2])^2 = \frac{1}{12} \geq \frac{2^{3-n}}{3} \geq V_n,$$

这与假设矛盾。因此，我们可以得出结论：a_1 的沃罗诺伊区域只包含元素 1，从而得到 $a_1 = 1$。由此可推得，存在一个正整数 ℓ，其中 $1 \leq \ell < n-3$，使得 $a_1 = 1, a_2 = 2, \cdots, a_\ell = \ell$。现在我们证明 $a_{\ell+1} = \ell + 1$。注意，$a_{\ell+1}$ 的沃罗诺伊区域必定包含 $\ell + 1$。假设 $a_{\ell+1}$ 的沃罗诺伊区域也包含元素 $\ell + 2$。接着按照命题 3.2.1 中类似的步骤推理，可以看出这会导致矛盾。因此，我们可以得出结论：$a_{\ell+1}$ 的沃罗诺伊区域只包含元素 $\ell + 1$，从而得到$a_{\ell+1} = \ell + 1$。注意，$2 \leq \ell + 2 \leq n - 3$。因此，根据数学归纳法原理，我们可以推得 $a_1 = 1, a_2 = 2, \cdots, a_{n-3} = n - 3$。至此，命题得证。□

注释 3.3.2. 命题 3.3.1 证明了猜想 1.5 是正确的。

命题 3.3.3. 设 $P := \sum_{j=1}^{k-1} p_j \delta_j + \sum_{j=k}^{\infty} \frac{1}{2^j} \delta_j$ 为定义 1.2 所定义的概率测度。假设猜想 1.5 成立。那么，量化维数 $D(P)$ 存在且等于零。

证明. 根据定理 1.3，在猜想 1.5 成立的假设下，对于概率测度 P 及任意正整数 $n \geq k + 2$，其 n 阶量化误差由定义 1.2 定义为 $V_n(P) = \frac{2^{3-n}}{3}$。因此，利用公式 (1)，可得 $D(P) = 0.$ □

4. Optimal quantization for the two probability distributions described in Section 3

在本节的以下两个小节中，我们在猜想 1.5 成立的假设下，确定第 3.2 小节和第 3.3 小节中给出的两个概率测度的 n-均值最优集以及所有正整数 $n \geq 2$ 对应的 n 次量化误差。

4.1. 当 P 满足 $(p_1, p_2, p_3, p_4, \cdots) = (1\,2^2, 1\,2, 1\,2^3, 1\,2^4, 1\,2^5, \cdots)$ 时的最优量化。

下面给出相应的命题。

命题 4.1.1。二均值最优集合为 $\{Av[1,3], Av[4,\infty)\}$，量化误差为 $V_2 = \frac{17}{28}$。

证明。我们看到 $Av[1,3] = \frac{13}{7}$，且 $Av[4,\infty) = 5$。由于 $3 < \frac{1}{2}(\frac{13}{7} + 5) = \frac{24}{7} < 4$，集合 $\beta := \{\frac{13}{7}, 5\}$ 所导致的失真误差为

$$V(P;\beta) = Er[1,3] + Er[4,\infty) = \frac{17}{28}.$$

由于 V_2 是二均值的量化误差，因此有 $V_2 \leq \frac{17}{28}$。设 $\alpha := \{a_1, a_2\}$ 为一个二均值最优集合，满足 $a_1 < a_2$。由于最优集中的点是其各自沃罗诺伊区域中的条件期望，因此有 $1 \leq a_1 < a_2 < \infty$。注意，$a_1$ 的沃罗诺伊区域必须包含 1。假设 a_1 的沃罗诺伊区域包含集合 $\{1,2,3,4\}$。那么，

$$V_2 \geq \sum_{j=1}^{4} f(j)(j - Av[1,4])^2 = Er[1,4] = \frac{5}{8} > V_2,$$

这会导致矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$ 或 $\{1,2\}$，或者集合 $\{1,2,3\}$。假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$，那么 a_2 的沃罗诺伊区域就包含集合 $\{n\colon n \geq 2\}$。于是，我们有

$$V_2 = Er[2,\infty) = \frac{7}{6} > V_2,$$

这会导致矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1,2\}$，或者集合 $\{1,2,3\}$。假设 a_1 的沃罗诺伊区域仅包含集合 $\{1,2\}$。那么， a_2 的沃罗诺伊区域包含 $\{3,4,5,\cdots\}$，从而得到

$$V_2 = Er[1,2] + Er[3,\infty) = \frac{2}{3} > V_2,$$

这就产生了矛盾。因此，我们可以得出结论： a_1 的沃罗诺伊区域仅包含集合$\{1,2,3\}$，而 a_2 的沃罗诺伊区域包含集合 $\{j:\ j \geq 4\}$ 从而得到

$$a_1 = Av[1,3] \text{ and } a_2 = Av[4,\infty) \text{ with quantization error } V_2 = Er[1,3] + Er[4,\infty) = \frac{17}{28}.$$

至此，命题的证明完成。 □

命题 **4.1.2**。集合 $\{Av[1,2], Av[3,4], Av[5,\infty)\}$ 构成三均值的最优集，其量化误差为 $V_3 = \frac{1}{3}$。

集合 $\beta := \{Av[1,2], Av[3,4], Av[5,\infty)\}$ 所导致的失真误差为

$$V(P;\beta) = Er[1,2] + Er[3,4] + Er[5,\infty) = \frac{1}{3}.$$

由于 V_3 是三均值的量化误差，因此我们有 $V_3 \leq \frac{1}{3}$。设 $\alpha := \{a_1, a_2, a_3\}$ 是三均值的一个最优集。由于最优集中的点是其各自沃罗诺伊区域内的条件期望，因此我们有 $1 \leq a_1 < a_2 < a_3 < \infty$。假设 a_1 的沃罗诺伊区域包含集合 $\{1,2,3\}$。那么，

$$V_3 \geq \sum_{j=1}^{3} f(j)(j - Av[1,3])^2 = Er[1,3] = \frac{5}{14} > \frac{1}{3} \geq V_3,$$

这将导致矛盾。因此，我们可以假设， a_1 的沃罗诺伊区域仅包含集合 $\{1\}$或集合 $\{1,2\}$。为导出矛盾，假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$。那么， a_2 的沃罗诺伊区域必然包含元素 2。假设 a_2 的沃罗诺伊区域包含集合 $\{2,3,4,5\}$。于是，

$$V_3 \geq Er[2,5] = \frac{181}{368} = 0.491848 > V_3,$$

这将产生矛盾。假设 a_2 的沃罗诺伊区域仅包含集合 $\{2,3,4\}$，因此 a_3 的沃罗诺伊区域包含集合 $\{n:\ n \geq 5\}$。那么，失真误差为

$$V_3 = Er[2,4] + Er[5,\infty) = \frac{9}{22} = 0.409091 > V_3,$$

这将导致矛盾。接下来，假设 a_2 的沃罗诺伊区域仅包含集合 $\{2,3\}$，因此 a_3 的沃罗诺伊区域包含集合 $\{n:\ n \geq 4\}$。那么，失真误差为

$$V_3 = Er[2,3] + Er[4,\infty) = \frac{7}{20} > V_3,$$

这导致矛盾。最后，假设 a_2 的沃罗诺伊区域仅包含集合 $\{2\}$，因此 a_3 的沃罗诺伊区域包含集合 $\{n:\ n \geq 3\}$。那么，失真误差为

$$V_3 = Er[3,\infty) = \frac{1}{2} > V_3,$$

这会导致矛盾。因此，我们可以得出结论： a_1 的沃罗诺伊区域只包含集合$\{1,2\}$。那么， a_2 的沃罗诺伊区域必须包含元素 3。假设a_2 的沃罗诺伊区域包含集合 $\{3,4,5,6\}$。那么，

$$V_3 \geq \sum_{j=1}^{2} f(j)(j - Av[1,2])^2 + \sum_{j=3}^{6} f(j)(j - Av[3,6])^2 = Er[1,2] + Er[3,6] = \frac{59}{160} = 0.36875 > V_3,$$

这会导致矛盾。假设 a_2 的沃罗诺伊区域只包含集合 $\{3,4,5\}$，因此 a_3 的沃罗诺伊区域包含集合 $\{n\colon n\geq 6\}$。那么，失真误差为

$$V_3 = Er[1,2] + Er[3,5] + Er[6,\infty) = \frac{29}{84} = 0.345238 > V_3,$$

这就产生了矛盾。接下来，假设 a_2 的沃罗诺伊区域仅包含元素 3，因此 a_3 的沃罗诺伊区域包含集合 $\{n\colon n\geq 4\}$。那么，失真误差为

$$V_3 = Er[1,2] + Er[4,\infty) = \frac{5}{12} = 0.416667 > V_3,$$

这导致矛盾。因此，我们可以得出结论：a_2 的沃罗诺伊区域仅包含集合 $\{3,4\}$，从而得到 $a_1 = Av[1,2]$、$a_2 = Av[3,4]$ 和 $a_3 = Av[5,\infty)$，其量化误差为 $V_3 = \frac{1}{3}$。由此，命题的证明完成。□

命题 4.1.3. 集合 $\{1,2,Av[3,4],Av[5,\infty)\}$ 构成四均值的最优集，量化误差为 $V_4 = \frac{1}{6}$。

证明。由于集合 $\beta := \{1,2,Av[3,4],Av[5,\infty)\}$ 产生的失真误差为

$$V(P;\beta) = Er[3,4] + Er[5,\infty) = \frac{1}{6}.$$

由于 V_4 是四均值的量化误差，因此有 $V_4 \leq \frac{1}{6}$。令 $\alpha := \{a_1,a_2,a_3,a_4\}$ 为四均值的最优集。由于最优集中的点是其各自沃罗诺伊区域中的条件期望，因此有 $1 \leq a_1 < a_2 < a_3 < a_4 < \infty$。显然，$a_1$ 的沃罗诺伊区域包含点 1。假设 a_1 的沃罗诺伊区域包含集合 $\{1,2,3\}$。那么，

$$V_3 \geq \sum_{j=1}^{3} f(j)(j - Av[1,3])^2 = Er[1,3] = \frac{5}{14} > \frac{1}{6} \geq V_4,$$

这会导致矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$，或集合 $\{1,2\}$。假设 a_1 的沃罗诺伊区域仅包含集合 $\{1,2\}$。那么，自然数集合中的其余元素都包含在 a_2、a_3 和 a_4 的沃罗诺伊区域的并集中。注意，点 a_2、a_3 和 a_4 所贡献的总失真误差为正。因此，

$$V_4 > \text{distortion error contributed by the point } a_1 = Er[1,2] = \frac{1}{6} = V_4,$$

这将导致矛盾。因此，a_1 的沃罗诺伊区域不能包含 $\{1,2\}$，即 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$，也就是 $a_1 = 1$。于是，a_2 的沃罗诺伊区域必须包含 2。假设 a_2 的沃罗诺伊区域包含集合 $\{2,3,4\}$。那么，

$$V_4 \geq Er[2,4] = \frac{25}{88} > V_4,$$

这将导致矛盾。因此，我们可以假设 a_2 的沃罗诺伊区域仅包含集合 $\{2\}$或集合 $\{2,3\}$。假设 a_2 的沃罗诺伊区域仅包含集合 $\{2,3\}$。设 a_3 的沃罗诺伊区域包含集合 $\{4,5,6,7\}$。那么，

$$V_4 \geq Er[2,3] + Er[4,7] = \frac{193}{960} > V_4,$$

这将导致矛盾。因此，我们可以假设 a_3 的沃罗诺伊区域仅包含集合 $\{4\}$、$\{4,5\}$或 $\{4,5,6\}$。假设 a_3 的沃罗诺伊区域仅包含集合 $\{4,5,6\}$。那么，a_4 的沃罗诺伊区域包含集合 $\{n : n \geq 7\}$。于是，

$$V_4 = Er[2,3] + Er[4,6] + Er[7,\infty) = \frac{53}{280} > V_4,$$

这构成矛盾。假设 a_3 的沃罗诺伊区域仅包含集合 $\{4,5\}$。那么，a_4 的沃罗诺伊区域包含集合 $\{n : n \geq 6\}$。接着，

$$V_4 = Er[2,3] + Er[4,5] + Er[6,\infty) = \frac{11}{60} > V_4,$$

这会导致矛盾。假设 a_3 的沃罗诺伊区域仅包含集合 $\{4\}$。那么， a_4 的沃罗诺伊区域包含集合 $\{n:$ $n \geq 5\}$。然后，

$$V_4 = Er[2,3] + Er[5,\infty) = \frac{9}{40} > V_4,$$

这导致了矛盾。因此，我们可以看到，如果 a_2 的沃罗诺伊区域仅包含集合 $\{2,3\}$，那么就会产生矛盾。因此，我们可以得出结论， a_2 的沃罗诺伊区域仅包含集合 $\{2\}$，换句话说，我们有 $a_2 = 2$。然后， a_3 的沃罗诺伊区域包含集合 $\{3\}$。假设 a_3 的沃罗诺伊区域包含集合 $\{3,4,5,6\}$，那么与之前一样，我们可以看到会产生矛盾。因此， a_3 的沃罗诺伊区域仅包含集合 $\{3\}$、 $\{3,4\}$ 或集合 $\{3,4,5\}$。注意，如果 a_3 的沃罗诺伊区域仅包含集合 $\{3,4,5\}$，那么 a_4 的沃罗诺伊区域包含集合 $\{n:\ n \geq 6\}$；如果 a_3 的沃罗诺伊区域仅包含集合 $\{3\}$，那么 a_4 的沃罗诺伊区域包含集合 $\{n:\ n \geq 4\}$。在这两种情况下，按照之前的方法继续推导，我们可以看到都会产生矛盾。因此，我们可以得出结论， a_3 的沃罗诺伊区域仅包含集合 $\{3,4\}$。因此， a_4 的沃罗诺伊区域包含 $\{n:\ n \geq 5\}$。由此，我们得到

$$a_1 = 1, a_2 = 2, a_3 = [3,4], \text{ and } a_4 = [5,\infty) \text{ with } V_4 = \frac{1}{6}.$$

因此，该命题的证明完成。 □

命题 4.1.4。集合 $\{1,2,\cdots,n-3, Av[n-2,n-1], Av[n,n+1], Av[n+2,\infty)\}$ 和 $\{1,2,\cdots,n-3, n-2, Av[n-1,n], Av[n+1,\infty)\}$ 构成所有 $n \geq 5$ 的 n 均值最优集，其量化误差为 $V_n = \frac{2^{3-n}}{3}$.

证明。假设猜想 1.5 成立，则该结论由定理 1.3 和猜想 1.5 得出。□

4.2. 当 $(p_1, p_2, p_3, p_4, \cdots) = (\frac{1}{2^3}, \frac{1}{2^2}, \frac{1}{2}, \frac{1}{2^4}, \frac{1}{2^5}, \cdots)$ 时， P 的最优量化。下面给出相关命题。

命题 4.2.1. 二均值最优集合为 $\{Av[1,3], Av[4,\infty)\}$，其量化误差为 $V_2 = \frac{5}{7}$。

证明。我们看到 $Av[1,3] = \frac{17}{7}$，且 $Av[4,\infty) = 5$。由于 $3 < \frac{1}{2}(\frac{17}{7} + 5) = \frac{26}{7} < 4$，集合 $\beta := \{\frac{17}{7}, 5\}$ 所导致的失真误差为

$$V(P;\beta) = Er[1,3] + Er[4,\infty) = \frac{5}{7}.$$

由于 V_2 是二均值的量化误差，我们有 $V_2 \leq \frac{5}{7}$。设 $\alpha := \{a_1, a_2\}$ 是一个二均值最优集合，满足 $a_1 < a_2$。由于最优集中的点是其各自沃罗诺伊区域内的条件期望，我们有 $1 \leq a_1 < a_2 < \infty$。注意， a_1 的沃罗诺伊区域必定包含 1。假设 a_1 的沃罗诺伊区域包含集合 $\{1,2,3,4,5\}$。那么，

$$V_2 \geq Er[1,5] = \frac{393}{496} > V_2,$$

这就导出了矛盾。因此，我们可以得出结论： a_1 的沃罗诺伊区域不包含点 5。假设 a_1 的沃罗诺伊区域仅包含 $\{1,2,3,4\}$，因此 a_2 的沃罗诺伊区域包含集合 $\{n:\ n \geq 5\}$。于是，我们有

$$V_2 = Er[1,4] + Er[5,\infty) = \frac{11}{15} > V_2,$$

这就导致了矛盾。类似地，我们可以证明，如果 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$或集合 $\{1,2\}$，那么就会得到矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1,2,3\}$，而 a_2 的沃罗诺伊区域仅包含集合 $\{n:\ n \geq 4\}$。于是，我们有

$$a_1 = Av[1,3],\ a_2 = Av[4,\infty) \text{ with } V_2 = \frac{5}{7}.$$

至此，该命题的证明完成。 □

命题 4.2.2. 集合 $\{Av[1,2], Av[3,4], Av[5,\infty)\}$ 构成三均值的最优集，其量化误差为 $V_3 = \frac{19}{72}$。

证明。由集合 $\beta := \{Av[1,2], Av[3,4], Av[5,\infty)\}$ 导致的失真误差为

$$V(P;\beta) = Er[1,2] + Er[3,4] + Er[5,\infty) = \frac{19}{72}.$$

由于 V_3 是三均值的量化误差，因此我们有 $V_3 \leq \frac{19}{72} = 0.263889$。令 $\alpha := \{a_1, a_2, a_3\}$ 为三均值的一个最优集。由于最优集中的点是其各自沃罗诺伊区域内的条件期望，因此我们有 $1 \leq a_1 < a_2 < a_3 < \infty$。假设 a_1 的沃罗诺伊区域包含集合 $\{1,2,3\}$。那么，

$$V_3 \geq Er[1,3] = \frac{13}{28} > \frac{19}{72} \geq V_3,$$

这会导致矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$ 或集合 $\{1,2\}$。假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$。在这种情况下，a_2 的沃罗诺伊区域必须包含元素 2。假设 a_2 的沃罗诺伊区域包含集合 $\{2,3,4\}$。那么，

$$V_3 \geq Er[2,4] = \frac{7}{26} = 0.269231 > V_3,$$

这导致矛盾。假设 a_2 的沃罗诺伊区域仅包含集合 $\{2,3\}$，因此 a_3 的沃罗诺伊区域包含集合 $\{n : n \geq 4\}$。那么，失真误差为

$$V_3 = Er[2,3] + Er[4,\infty) = \frac{5}{12} > V_3,$$

这导致矛盾。最后，假设 a_2 的沃罗诺伊区域仅包含集合 $\{2\}$，因此 a_3 的沃罗诺伊区域包含集合 $\{n : n \geq 3\}$。那么，失真误差为

$$V_3 = Er[3,\infty) = \frac{13}{20} > V_3,$$

这导致矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1,2\}$。那么，a_2 的沃罗诺伊区域必须包含集合 $\{3\}$。假设 a_2 的沃罗诺伊区域包含集合 $\{3,4,5,6\}$。那么，

$$V_3 \geq Er[1,2] + Er[3,6] = \frac{151}{416} > V_3,$$

这导致矛盾。假设 a_2 的沃罗诺伊区域仅包含集合 $\{3,4,5\}$，因此 a_3 的沃罗诺伊区域包含集合 $\{n\colon\ n \geq 6\}$。那么，失真误差为

$$V_3 = Er[1,2] + Er[3,5] + Er[6,\infty) = \frac{35}{114} > V_3,$$

这就产生了矛盾。接下来，假设 a_2 的沃罗诺伊区域仅包含元素 3，因此 a_3 的沃罗诺伊区域包含集合 $\{n\colon\ n \geq 4\}$。那么，失真误差为

$$V_3 = Er[1,2] + Er[4,\infty) = \frac{1}{3} > V_3,$$

这就产生了矛盾。因此，我们可以得出结论：a_2 的沃罗诺伊区域仅包含集合 $\{3,4\}$，从而得到 $a_1 = Av[1,2]$、$a_2 = Av[3,4]$ 和 $a_3 = Av[5,\infty)$，其量化误差为 $V_3 = \frac{19}{72}$。□

命题 4.2.3。集合 $\{Av[1,2], 3, Av[4,5], Av[6,\infty)\}$ 构成四均值的最优集，其量化误差为 $V_4 = \frac{1}{6}$。

证明。该 由集合 $\beta := \{Av[1,2], 3, Av[4,5], Av[6,\infty)\}$ 产生的失真误差为 y

$$V(P;\beta) = Er[1,2] + Er[4,5] + Er[6,\infty) = \frac{1}{6}.$$

由于 V_4 是四均值的量化误差，因此有 $V_4 \leq \frac{1}{6} = 0.166667$。令 $\alpha := \{a_1, a_2, a_3, a_4\}$ 为四均值的一个最优集。由于最优集中的点是其

各自沃罗诺伊区域内的条件期望，因此有 $1 \leq a_1 < a_2 < a_3 < a_4 < \infty$。假设 a_1的沃罗诺伊区域包含集合 $\{1, 2, 3\}$。那么，

$$V_4 \geq Er[1, 3] = \frac{13}{28} > V_4,$$

这会导致矛盾。因此，我们可以假设， a_1 的沃罗诺伊区域仅包含集合 $\{1\}$或集合 $\{1, 2\}$。假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$。在这种情况下， a_2 的沃罗诺伊区域必须包含元素 2。假设 a_2 的沃罗诺伊区域包含集合 $\{2, 3, 4\}$。那么，

$$V_4 \geq Er[2, 4] = \frac{7}{26} = 0.269231 > V_4,$$

由此得到矛盾。假设 a_2 的沃罗诺伊区域仅包含集合 $\{2, 3\}$。那么，注意到 a_3 和 a_4 的沃罗诺伊区域包含所有元素 $\{n : n \geq 4\}$。因此， a_3 和 a_4 产生的总失真误差必定为正。这说明

$$V_4 > Er[2, 3] = \frac{1}{6} \geq V_4,$$

这就导致矛盾。假设 a_2 的沃罗诺伊区域仅包含集合 $\{2\}$。那么，和之前一样，我们可以看到这会导致矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1, 2\}$。那么， a_2 的沃罗诺伊区域必须包含 3。如果 a_2 的沃罗诺伊区域包含更多点，运用之前类似的论证可以证明这会导致矛盾。因此，我们可以得出结论： $a_2 = 3$。同样，运用类似的论证可以证明，a_3 的沃罗诺伊区域仅包含集合 $\{4, 5\}$，而 a_4 的沃罗诺伊区域仅包含集合 $\{n : n \geq 6\}$。因此，我们有

$$a_1 = Av[1, 2], a_2 = 3, a_3 = Av[4, 5], \text{ and } a_4 = Av[6, \infty) \text{ with quantization error } V_4 = \frac{1}{6}.$$

因此，该命题的证明完成。 □

命题4.2.4。集合 $\{1, 2, 3, Av[4, 5], Av[6, \infty)\}$ 构成五均值的最优集，其量化误差为 $V_5 = \frac{1}{12}$。

证明。T 由集合 $\beta := \{1, 2, 3, Av[4, 5], Av[6, \infty)\}$ 产生的失真误差为

$$V(P; \beta) = Er[4, 5] + Er[6, \infty) = \frac{1}{12}.$$

由于 V_5 是五均值的量化误差，我们有$V_5 \leq \frac{1}{12} = 0.0833333$。设 $\alpha := \{a_1, a_2, a_3, a_4, a_5\}$是一个满足 $a_1 < a_2 < a_3 < a_4 < a_5$的五均值最优集。由于最优集中的点是其各自沃罗诺伊区域内的条件期望，我们有 $1 \leq a_1 < a_2 < a_3 < a_4 < a_5 < \infty$。显然， a_1 的沃罗诺伊区域包含点1。为导出矛盾，假设 a_1 的沃罗诺伊区域包含集合 $\{1, 2, 3\}$。那么，

$$V_5 \geq Er[1, 3] = \frac{13}{28} > V_5,$$

这与假设矛盾。接下来，假设 a_1 的沃罗诺伊区域仅包含集合 $\{1, 2\}$。那么请注意， a_2、 a_3、 a_4 和 a_5 的沃罗诺伊区域的并集包含所有元素 $\{n : n \geq 3\}$。因此，我们必须有

$$V_5 > Er[1, 2] = \frac{1}{12} \geq V_5,$$

这是一个矛盾。因此，我们可以得出结论： a_1 的沃罗诺伊区域仅包含元素 1，即 $a_1 = 1$。显然， a_2 的沃罗诺伊区域包含元素 2。假设a_2 的沃罗诺伊区域包含集合 $\{2, 3\}$。那么，我们有

$$V_5 \geq Er[2, 3] = \frac{1}{6} > V_5,$$

这会导致矛盾。因此， a_2 的沃罗诺伊区域只包含元素 2，即 $a_2 = 2$。同理，我们可以证明 $a_3 = 3$。其余证明与

命题 4.1.3 中给出的证明思路相同。因此，我们看到 $a_4 = Av[4,5]$ 和 $a_5 = Av[6,\infty)$ 的量化误差为 $V_5 = \frac{1}{12}$。至此，命题得证。□

命题 4.2.5. 集合 $\{1,2,\cdots,n-3, Av[n-2,n-1], Av[n,n+1], Av[n+2,\infty)\}$ 和 $\{1,2,\cdots,n-3, n-2, Av[n-1,n], Av[n+1,\infty)\}$ 对所有满足 $n \geq 6$ 的 n-均值构成最优集，其量化误差为 $V_n = \frac{2^{3-n}}{3}$.

证明. 假设猜想 1.5 成立，则根据定理 1.3 和猜想 1.5 即可得到证明。□

5. Observation and Remarks

在猜想 1.5 中，概率测度 P 定义为 $P := \sum_{j=1}^{k-1} p_j\delta_j + \sum_{j=k}^{\infty} \frac{1}{2^j}\delta_j$，其中 $(p_1, p_2, \cdots, p_{k-1})$ 是集合 $\{\frac{1}{2}, \frac{1}{2^2}, \cdots, \frac{1}{2^{k-1}}\}$ 的一个排列，且 $k \in \mathbb{N}$ 满足 $k \geq 2$。如果 $P := \sum_{j=1}^{k-1} p_j\delta_j + \sum_{j=k}^{\infty} \frac{1}{2^j}\delta_j$，并且 $(p_1, p_2, \cdots, p_{k-1})$ 不是集合 $\{\frac{1}{2}, \frac{1}{2^2}, \cdots, \frac{1}{2^{k-1}}\}$ 的一个排列，那么猜想 1.5 不成立。对此，我们给出以下命题。

命题 5.1. 对于由 P 给出的概率测度 $P: = \frac{149}{200}\delta_1 + \frac{1}{200}\delta_2 + \sum_{j=3}^{\infty} \frac{1}{2^j}\delta_j$ 五均值最优集为

$$\{1, Av[2,3], 4, Av[5,6], Av[7,\infty)\}, \text{ or } \{1, Av[2,3], Av[4,5], Av[6,7], Av[8,\infty)\}$$

其量化误差为 $V_5 = \frac{29}{624}$。

证明。由集合 $\beta: = \{1, Av[2,3], 4, Av[5,6], Av[7,\infty)\}$ 产生的失真误差为

$$V(P;\beta) = Er[2,3] + Er[5,6] + Er[7,\infty) = \frac{29}{624}.$$

由于 V_5 是五均值的量化误差，因此有 $V_5 \leq \frac{29}{624} = 0.0464744$。设 $\alpha: = \{a_1, a_2, a_3, a_4, a_5\}$ 是一个满足 $a_1 < a_2 < a_3 < a_4 < a_5$ 的五均值最优集。由于最优集中的点是其各自沃罗诺伊区域内的条件期望，因此有 $1 \leq a_1 < a_2 < a_3 < a_4 < a_5 < \infty$。显然， a_1 的沃罗诺伊区域包含点 1。为反证起见，假设 a_1 的沃罗诺伊区域包含集合 $\{1,2,3\}$。于是，

$$V_5 \geq Er[1,3] = \frac{7537}{17500} = 0.430686 > V_5,$$

这与假设矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域仅包含集合 $\{1\}$ 或集合 $\{1,2\}$。假设 a_1 的沃罗诺伊区域仅包含集合 $\{1,2\}$。那么， a_2 的沃罗诺伊区域必定包含元素 3。假设 a_2 的沃罗诺伊区域包含集合 $\{3,4\}$。那么，

$$V_5 \geq Er[1,2] + Er[3,4] = \frac{1399}{30000} = 0.0466333 > V_5,$$

这将导致矛盾。因此，我们可以假设 a_2 的沃罗诺伊区域仅包含元素 3，即 $a_2 = 3$。那么， a_3、a_4、a_5 的沃罗诺伊区域的并集包含集合 $\{4,5,6,\cdots\}$，且每个 $j \in \{4,5,6,\cdots\}$ 对应的概率为 $\frac{1}{2^j}$。因此，按照定理 1.3 的证明中所述的类似思路，我们可以证明

$$(2) \qquad \{a_3, a_4, a_5\} \text{ equals the set } \{4, Av[5,6], Av[7,\infty)\}, \text{ or } \{Av[4,5], Av[6,7], Av[8,\infty)\}$$

与量化误差相关的

$$V_5 = Er[1,2] + Er[5,6] + Er[7,\infty) = \frac{1399}{30000} = 0.0466333 > V_5,$$

这会导致矛盾。因此，我们可以假设 a_1 的沃罗诺伊区域只包含元素 1，即 $a_1 = 1$。于是， a_2 的沃罗诺伊区域必须包含 2。假设 a_2 的沃罗诺伊区域包含集合 $\{2,3,4\}$。那么，

$$V_5 \geq Er[2,4] = \frac{31}{616} > V_5,$$

这会导致矛盾。因此， a_2 的沃罗诺伊区域只包含集合 $\{2\}$或 $\{2,3\}$。假设 a_2 的沃罗诺伊区域只包含集合 $\{2\}$，即 $a_2 = 2$。那么，根据 $a_1 = 1$、 $a_2 = 2$，按照定理 1.3 的证明中所述的类似步骤，我们可以证明 $a_3 = 3$、 $a_4 = Av[4,5]$、 $a_5 = Av[6,\infty)$；或者 $a_3 = Av[3,4]$、 $a_4 = Av[5,6]$、 $a_5 = Av[7,\infty)$，其量化误差为 $V_5 = \frac{1}{12} > V_5$，这与假设矛盾。因此，我们可以假设 a_2 的沃罗诺伊区域只包含集合 $\{2,3\}$。同样，按照定理 1.3 的证明中所述的类似步骤，我们可以证明 $\{a_3, a_4, a_5\}$ 等于集合 $\{4, Av[5,6], Av[7,\infty)\}$；或者 $\{Av[4,5], Av[6,7], Av[8,\infty)\}$。因此，我们得出结论：五均值最优集要么是 $\{1, Av[2,3], 4, Av[5,6], Av[7,\infty)\}$，要么是 $\{1, Av[2,3], Av[4,5], Av[6,7], Av[8,\infty)\}$，其量化误差为 $V_5 = \frac{29}{624}$。命题得证。□

注 5.2. 命题 5.1 表明，对于与正整数集 $\mathbb{N}$ 相关联的任意概率分布 $(p_1, p_2, p_3, \cdots)$，猜想 1.5 并不成立。

注 5.3. 第 3.2 小节和第 3.3 小节给出的两个例子验证了猜想 1.5。我们仍无法给出该猜想的一般性证明。进一步探究该猜想的一般性证明将很有价值。

References

[AW] E.F. Abaya 和 G.L. Wise，《关于最优量化器存在性的一些评述》，Statistics & Probability Letters，第 2 卷，第 6 期，1984年12月，第 349–351 页。[DFG] Q. Du、V. Faber 和 M. Gunzburger，《质心 Voronoi 镶嵌：应用与算法》，SIAM Review，第 41 卷，第 4 期（1999），第 637–676 页。[DR] C.P. Dettmann 和 M.K. Roychowdhury，《计算有限生成 Cantor 分布的质心 Voronoi 镶嵌的算法》，Southeast Asian Bulletin of Mathematics（2021）45：173–188。[GG] A. Gersho 和 R.M. Gray，《矢量量化与信号压缩》，Kluwer Academy publishers：波士顿，1992年。[GKL] R.M. Gray、J.C. Kieffer 和 Y. Linde，《局部最优块量化器设计》，Information and Control，45（1980），第 178–198 页。[GL1] A. Gy¨orgy 和 T. Linder，《关于最优熵约束标量量化器的结构》，IEEE transactions on information theory，第 48 卷，第 2 期，2002年2月。[GL2] S. Graf 和 H. Luschgy，《概率分布量化基础》，Lecture Notes in Mathematics 1730，Springer，柏林，2000年。[GN] R.M. Gray 和 D.L. Neuhoff，《量化》，IEEE Transactions on Information Theory，第 44 卷，第 6 期，1998年10月，第 2325–2383 页。[P] D. Pollard，《量化与 k-均值法》，IEEE Transactions on Information Theory，28（1982），199–205。[R1] M.K. Roychowdhury，《二进 Cantor 集上概率测度的量化与质心 Voronoi 镶嵌》，Journal of Fractal Geometry，4（2017），127–146。[Z1] P.L. Zador，《连续信号的渐近量化误差与量化维数》，IEEE Trans-actions on I nformation Theory，28（1982），139–149。[Z2] R. Zam，《信号与网络的格编码：量化、调制与多用户信息论的结构化编码方法》，剑桥大学出版社，2014年。

School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, 1201 West University Drive, Edinburg, TX 78539-2999, USA.电子邮箱地址： {[1]juan.gomez15, [2]haily.martinez01, [3]mrinal.roychowdhury}@utrgv.edu 电子邮箱地址： {[4]alexis.salazar01, [5]daniel.vallez01}@utrgv.edu