The quantization of the Cantor distribution

S. Graf, H. Luschgy

1995

Abstract

For a real–valued random variable whose distribution is the classical Cantor probability the n–quantization error and the n–optimal quantization rules are calculated for every natural number n. Moreover the connection between the rate of convergence of the logarithms of the quantization errors for n going to infinity and the Hausdorff dimension of the Cantor set is indicated.

1 Introduction

Let P denote a probability measure on $\mathcal{R}^d$. Fix $n \in \mathcal{N}$ and let $\mathcal{F}_n$ be the set of all n–quantizers, i.e. the set of all Borel measurable maps $f : \mathcal{R}^d \to \mathcal{R}^d$ with $card(f(\mathcal{R}^d)) \le n$. If X is a random variable in $\mathcal{R}^d$ with distribution P then, for each $f \in \mathcal{F}_n$, $f(X)$ gives a quantized version of X. The **quantization error** is defined as

$$V_n = \inf_{f \in \mathcal{F}_n} \int \|x - f(x)\|^2 \, dP(x).$$

A quantizer $f \in \mathcal{F}_n$ is called an **optimal n–quantizer** for P if

$$V_n = \int \|x - f(x)\|^2 \, dP(x).$$

The quantization problem is to find an optimal n–quantizer for P or at least an asymptotically optimal sequence of n–quantizers as $n \to \infty$. This problem arose in the context of coding speech and visual signals effectively (see Abut[1] for an overview of the relevant research articles). Most of the mathematical theory of quantization was developed for probability measures P which are absolutely continuous (with respect to Lebesgue measure) or discrete. In the present article we want to study the quantization problem for the canonical probability measure on the classical Cantor set $C \subseteq \mathcal{R}$ which is a rather simple example of a continuous singular distribution. We will calculate the values V_n and determine the optimal quantizers for the Cantor distribution. For probability measures P on $\mathcal{R}^d$ with non–vanishing absolutely continuous part $\lim_{n\to\infty} n^{\frac{2}{d}} V_n$ is finite and strictly positive (see Bucklew–Wise[2]). For the Cantor measure we show that the sequence $(n^{\frac{2}{\beta}} V_n)_{n\in\mathcal{N}}$ is still bounded and bounded away from 0, where $\beta = \frac{\log 2}{\log 3}$ denotes the Hausdorff dimension of the Cantor set C. But $(n^{\frac{2}{\beta}} V_n)_{n\in\mathcal{N}}$ does not converge: $(n^{\frac{2}{\beta}} V_n)_{n\in\mathcal{N}}$ has a non degenerate interval of accumulation points (see Theorem 6.3). It is an interesting question whether this last result is characteristic of (continuous) singular probability measures.

The considerations described above also lead to a new dimension formula for the Cantor set (see Theorem 6.6), namely $\lim_{n\to\infty} \frac{\log V_n}{\log n} = -\frac{2}{\beta}$.

This is a natural analogue to the result for probabiliy measures with non–vanishing absolutely continuous part where the corresponding formula with β replaced by d follows immediately from $\lim_{n\to\infty} n^{\frac{2}{d}} V_n \in]0, +\infty[$.

2 Some basic facts about quantization

In this section P is an arbitrary probability measure on $\mathcal{R}^d$ and $\| \;\|$ denotes the euclidean norm on $\mathcal{R}^d$.

THEOREM 1 *If* $\int \|x\|^2 \, dP(x) < \infty$ *then*

$$V_n = \inf_{\alpha} \int \min_{a\in\alpha} \|x - a\|^2 \, dP(x),$$

where the infimum is taken over all subsets α of $\mathcal{R}^d$ with card $\alpha \leq n$.

PROOF see, for instance, [3]. Λ

DEFINITION 2 A subset α of $\mathcal{R}^d$ is called an **optimal set of n–means** for P if *card* $\alpha \leq n$ and

$$V_n = \int \min_{a \in \alpha} \|x - a\|^2 \, dP(x).$$

Λ

THEOREM 3 *For every $n \geq 1$, there is an optimal subset of n–means for P.* Λ

PROOF see, for instance, [3]. Λ

DEFINITION 4 Given a finite subset $\alpha = \{a_1, \dots, a_m\} \subset \mathcal{R}^d$ we call a Borel measurable partition $A_1, \dots, A_m$ of $\mathcal{R}^d$ a **Voronoi partition** w.r.t. α if

$$A_i \subset \{x \in \mathcal{R}^d \mid \|x - a_i\| = \min_{a \in \alpha} \|x - a\|\}$$

for every $i \in \{1, \dots, m\}$. Λ

REMARK 5 As a consequence of Proposition 2.1 the problem of finding an optimal n–quantizer for P is equivalent to the problem of finding an optimal subset of n–means for P: If $\alpha = \{a_1, \dots, a_m\}$ is an optimal set of n–means for P and $A_1, \dots, A_m$ is a Voronoi partition w.r.t. α then $f = \sum_{i=1}^{m} a_i 1_{A_i}$ is an optimal n–quantizer. Conversely, if $f \in \mathcal{F}_n$ is an optimal n–quantizer then $\alpha = f(\mathcal{R}^d)$ is an optimal set of n–means for P.
It can be shown that for a continuous probability P an optimal set of n–means always has exactly n elements. For $d = 1$ and a finite set $\alpha = \{a_1, \dots, a_m\} \subset \mathcal{R} = \mathcal{R}^d$ a Voronoi partition $A_1, \dots, A_m$ with respect to α always consists of intervals and half–limes. Thus $\sum_{i=1}^{m} a_i 1_{A_i}$ is a monotone step function. The n–quantization problem on $\mathcal{R}$ is, therefore (up to modifications on P–nullsets), equivalent to finding all monotone step function with at most n values that optimally approximate the identity of $\mathcal{R}$ in $L^2(P)$. $\diamondsuit$

THEOREM 6 *Let α be an optimal set of n–means and, for $a \in \alpha$, define*

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

Then, for every $a \in \alpha$,

(i) $P(M(a)) > 0$,

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

(iii) $a = E(X \mid X \in M(a))$.

Λ

PROOF see [3]. Λ

3 An upper bound for the quantization error

From now on P denotes the Cantor probability measure. The following is a well–known characterization of P (see Hutchinson[5], Theorem 4.4(1) for a generalization to self–similar measures) which provides the key tool for the results in the rest of the paper.

THEOREM 7 *Let $S_1, S_2 : \mathcal{R} \to \mathcal{R}$ be defined by*

$$S_1 x = \frac{1}{3}x \text{ and } S_2 x = \frac{1}{3}x + \frac{2}{3}.$$

Then P is the unique Borel probability on $\mathcal{R}$ such that

$$P = \frac{1}{2} P \circ S_1^{-1} + \frac{1}{2} P \circ S_2^{-1},$$

where $P \circ S_i^{-1}$ denotes the image measure of P with respect to S_i $(i = 1, 2)$. Λ

REMARK 8 Let S_1 and S_2 be as in Theorem 3.1. For $= \sigma \in \{1,2\}^k$ with $\sigma = (\sigma_1, \ldots, \sigma_k)$ set $S_\sigma = S_{\sigma_1} \circ \ldots \circ S_{\sigma_k}$, $J_\sigma = S_\sigma([0,1])$, and a_σ the midpoint of J_σ. The sets $(J_\sigma)_{\sigma \in \{1,2\}^k}$ are just the 2^k intervals of length $\frac{1}{3^k}$ in the k–th level of the classical Cantor construction. The intervals $J_{\sigma * 1}, J_{\sigma * 2}$ into which J_σ is split up at the $(k+1)$–th level are called the children of J_σ. The set $C = \bigcap_{k \in \mathcal{N}} \bigcup_{\sigma \in \{1,2\}^k} J_\sigma$ is the Cantor set and equals the support of the Cantor measure P. $\diamondsuit$

COROLLARY 9 *Let $f : \mathcal{R} \to \mathcal{R}_+$ be Borel measurable and $k \in \mathcal{N}$. Then*

$$\int f dP = \frac{1}{2^k} \sum_{\sigma \in \{1,2\}^k} \int f \circ S_\sigma dP.$$

Λ

PROOF The result follows from Theorem 3.1 by induction on k. Λ

For later use we will need the expectation $E(P)$ and variance $V(P)$ of P.

LEMMA 10

$$E(P) = \frac{1}{2}, \ V(P) = V_1 = \frac{1}{8},$$

and for fixed $x_0 \in \mathcal{R}$,

$$\int (x - x_0)^2 dP(x) = (x_0 - \frac{1}{2})^2 + \frac{1}{8}.$$

Λ

PROOF

(i) It follows from Theorem 3.1 that

$$\begin{aligned} E(P) = \int x dP(x) &= \tfrac{1}{2}\int \tfrac{1}{3}x dP(x) + \tfrac{1}{2}\int \left(\tfrac{1}{3}x + \tfrac{2}{3}\right) dP(x) \\ &= \tfrac{1}{3}E(P) + \tfrac{1}{3}. \end{aligned}$$

This equation implies

$$E(P) = \frac{1}{2}.$$

(ii) First we will calculate $\int x^2 dP(x)$.
Using 3.1 we obtain

$$\begin{aligned} \int x^2 dP(x) &= \tfrac{1}{2}\int \left(\tfrac{1}{3}x\right)^2 dP(x) + \tfrac{1}{2}\int \left(\tfrac{1}{3}x + \tfrac{2}{3}\right)^2 dP(x) \\ &= \tfrac{1}{18}\int x^2 dP(x) + \tfrac{1}{18}\int x^2 P(x) + \tfrac{2}{9}\int x dP(x) + \tfrac{2}{9} \\ &= \tfrac{1}{9}\int x^2 dP(x) + \tfrac{1}{3}. \end{aligned}$$

This equation implies

$$\int x^2 dP(x) = \frac{3}{8}.$$

(iii) Using (i) & (ii) we deduce

$$V(P) = \int x^2 dP(x) - E(P)^2 = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}.$$

(iv) Since $\displaystyle\int (x - x_0)^2 dP(x) = V(P) + (E(P) - x_0)^2$ the last equation in the lemma holds.

Λ

DEFINITION 11 For $n \in \mathcal{N}$ with $n \geq 1$ let $l(n)$ be the unique natural number with $2^{l(n)} \leq n < 2^{l(n)+1}$. For $I \subset \{1,2\}^{l(n)}$ with $card(I) = n - 2^{l(n)}$ let $\alpha_n(I)$ be the set consisting of all midpoints a_σ of intervals J_σ with $\sigma \in \{1,2\}^{l(n)} \setminus I$ and of all midpoints $a_{\sigma*1}, a_{\sigma*2}$ of the children of J_σ with $\sigma \in I$. Formally,

$$\alpha_n(I) = \{a_\sigma \mid \sigma \in \{1,2\}^{l(n)} \setminus I\} \cup \{a_{\sigma*1} \mid \sigma \in I\} \cup \{a_{\sigma*2} \mid \sigma \in I\}.$$

Λ

REMARK 12 It is plausible and will be shown later that the sets $\alpha_n(I)$ are exactly the optimal sets of n–means for P. $\diamondsuit$

THEOREM 13

$$\int \min_{a \in \alpha_n(I)} \|x - a\|^2 dP(x) = \frac{1}{18^{l(n)}} \frac{1}{8} \left(2^{l(n)+1} - n + \frac{1}{9}(n - 2^{l(n)})\right)$$

Λ

PROOF Obviously we have

$$\int \min_{a\in\alpha_n(I)} \|x-a\|^2 dP(x) = \sum_{\sigma\in\{1,2\}^{l(n)}\setminus I} \int_{J_\sigma} \|x-a_\sigma\|^2 dP(x) + \sum_{\sigma\in I}\left[\int_{J_{\sigma*1}} \|x-a_{\sigma*1}\|^2 dP(x) + \int_{J_{\sigma*2}} \|x-a_{\sigma*2}\|^2 dP(x)\right].$$

It follows from Corollary 3.3 and Lemma 3.4 that, for every $k\in\mathcal{N}$ and every $\tau\in\{1,2\}^k$,

$$\begin{aligned}\int_{J_\tau} \|x-a_\tau\|^2 dP(x) &= \tfrac{1}{2^k}\int \|S_\tau(x)-S_\tau(\tfrac{1}{2})\|^2 dP(x)\\ &= \tfrac{1}{2^k 9^k}\int \|x-\tfrac{1}{2}\|^2 dP(x)\\ &= \tfrac{1}{18^k}\tfrac{1}{8}.\end{aligned}$$

Thus we deduce

$$\begin{aligned}\int \min_{a\in\alpha_n(I)} \|x-a\|^2 dP(x) &= \tfrac{1}{8}\cdot\tfrac{1}{18^{l(n)}} card(\{1,2\}^{l(n)}\setminus I)\\ &\qquad + \tfrac{1}{8}\cdot\tfrac{1}{18^{l(n)+1}} 2 card(I)\\ &= \tfrac{1}{8}\cdot\tfrac{1}{18^{l(n)}}(2^{l(n)+1}-n+\tfrac{1}{9}(n-2^{l(n)})).\end{aligned}$$

Λ

COROLLARY 14 *For every $n\in\mathcal{N}$,*

$$V_n \le \frac{1}{8}\cdot\frac{1}{18^{l(n)}}(2^{l(n)+1}-n+\frac{1}{9}(n-2^{l(n)})).$$

Λ

PROOF Since $card(\alpha_n(I)) = n$ it follows from Proposition 2.1 that

$$V_n \le \int \min_{a\in\alpha_n(I)} \|x-a\|^2 dP(x).$$

By Proposition 3.7 this implies the corollary. Λ

REMARK 15 For integers $n = 2^l$ the upper bound takes the form $\frac{1}{8}(\frac{1}{9})^l$ and coincides with the upper bound given by Bucklew and Wise [2], p. 243. $\diamondsuit$

4 On optimal sets of 2– and 3–means

In this section we will determine the opitmal sets of n–means for $n = 2$ and 3. The results and proofs for these two cases are the key for understanding the general case and, moreover, contain the principal part of the proof of the general result. The first step to this proof is the following lemma.

LEMMA 16 *Let $n\ge 2$ and let $\alpha = \{a_1,\dots,a_n\}$ be an optimal set of n–means for P and assume w.l.o.g. $a_1 < a_2 < \dots < a_n$. Then $\alpha\cap[0,\frac{1}{3}[\neq\emptyset$, $\alpha\cap]\frac{2}{3},1]\neq\emptyset$, and, for $j := \max\{i \mid a_i\le\frac{1}{3}\}$,* *(i)* $a_i \geq \frac{2}{3}$ *for all* $i > j+1$.

(ii) If $a_{j+1} \in]\frac{1}{3}, \frac{2}{3}[$ *then* $\frac{1}{2}(a_j + a_{j+1}) < \frac{1}{3}$ *and* $\frac{1}{2}(a_{j+1} + a_{j+2}) > \frac{2}{3}$.

In particular, an optimal set of 2–means does not contain a point from $]\frac{1}{3}, \frac{2}{3}[$.$\Lambda$

PROOF Assume $\frac{1}{3} \leq a_1$. Then Corollary 3.8, Theorem 3.1 and Lemma 3.4 imply

$$
\begin{aligned}
\tfrac{1}{72} \geq V_2 \geq V_n &= \int \min_{a\in\alpha}(x-a)^2\,dP(x) \\
&\geq \int_{[0,\frac{1}{3}]} (x-a_1)^2\,dP(x) \\
&= \tfrac{1}{2}\int (S_1(x)-a_1)^2\,dP(x) \\
&= \tfrac{1}{18}\int (x-3a_1)^2\,dP(x) \\
&= \tfrac{1}{18}\left[(3a_1-\tfrac{1}{2})^2+\tfrac{1}{8}\right] \geq \tfrac{3}{8}\cdot\tfrac{1}{8} > \tfrac{1}{72}.
\end{aligned}
$$

Thus we get a contradiction and $a_1 < \frac{1}{3}$ is proved. In a similar fashion one can show that $a_n > \frac{2}{3}$. By the preceding considerations we get

$$1 \leq j \leq n-1.$$

Set $k = \min\{i \in \{1,\dots,n\} \mid \frac{2}{3} \leq a_i\}$. Then $1 \leq j < k \leq n$. We will show that $k-j \leq 2$. Assume the contrary. Then we have $a_{j+1}, a_{j+2} \in]\frac{1}{3}, \frac{2}{3}[$. By Theorem 2.6 we have $P(M(a_{j+1})) > 0$ and $P(M(a_{j+2})) > 0$. Since $M(a_{j+1}) = [\frac{1}{2}(a_j + a_{j+1}), \frac{1}{2}(a_{j+1} + a_{j+2})]$ and $M(a_{j+2}) = [\frac{1}{2}(a_{j+1} + a_{j+2}), \frac{1}{2}(a_{j+2} + a_{j+3})]$ we, therefore, deduce

$$\frac{1}{2}(a_j + a_{j+1}) < \frac{1}{3} \text{ and } \frac{1}{2}(a_{j+2} + a_{j+3}) > \frac{2}{3}.$$

By moving a_{j+1} to $\frac{1}{3}$ and a_{j+2} to $\frac{2}{3}$ we could strictly reduce the quantization error which is a contradiction to the optimality of α.

Since $k \neq j$ we have $k = j+1$ or $k = j+2$. In any case (i) is proved.

Now suppose that $\frac{1}{3} < a_{j+1} < \frac{2}{3}$. Since $P(M(a_{j+1})) > 0$ we know that $\frac{1}{2}(a_{j+1} + a_j) < \frac{1}{3}$ or $\frac{1}{2}(a_{j+1}+a_{j+2}) > \frac{2}{3}$. Assume that only one of these inequalities holds, say the first one. Then the quantization error can be strictly reduced by moving a_{j+1} to $\frac{1}{3}$, which yields a contradiction to the optimality of α. Similarly we can proceed in the second case. Thus statement (ii) follows. Λ

For the following considerations we will need two more technical lemmas.

LEMMA 17 *Let* $F :]0,1] \to \mathcal{R}$ *be defined by*

$$F(t) = \frac{1}{P([0,t])} \int_{[0,t]} x\,dP(x).$$

Then F is an increasing continuous function with

$$(i) \qquad F(1-\tfrac{1}{3^n}) = 1 - \tfrac{1}{2}\left(\tfrac{1-(\frac{1}{6})^n}{1-(\frac{1}{2})^n}\right)$$

and

$$(ii) \qquad F(\tfrac{1}{3}t) = \tfrac{1}{3}F(t)$$

for all $t \in]0,1]$. Λ

Proof Since P is a continuous measure F is continuous. To show that F is non–decreasing let $t \in]0,1]$ and $h > 0$ with $t+h \leq 1$ be given. Then

$$\begin{aligned} F(t+h)-F(t) &= \tfrac{1}{P([0,t+h])P([0,t])}\Big(P([0,t]) \int\limits_{[0,t+h]} x dP(x) \\ &\qquad\qquad - P([0,t+h]) \int\limits_{[0,t]} x dP(x)\Big) \\ &= \tfrac{1}{P([0,t+h])P([0,t])}\Big(P([0,t]) \int\limits_{]t,t+h]} x dP(x) - P(]t,t+h]) \int\limits_{[0,t]} x dP(x)\Big) \\ &\geq \tfrac{1}{P([0,t+h])P([0,t])}\big(P([0,t])tP(]t,t+h]) - P(]t,t+h])tP([0,t])\big) \\ &= 0. \end{aligned}$$

Thus F is non–decreasing.
To prove (i) note that

$$(1) \qquad P([0,1-\tfrac{1}{3^n}]) = 1 - \tfrac{1}{2^n}$$

and

$$(2) \qquad \int\limits_{[0,1-\frac{1}{3^n}]} x dP(x) = \sum_{k=1}^{n} \int\limits_{[1-\frac{1}{3^{k-1}},1-\frac{1}{3^{k-1}}+\frac{1}{3^k}]} x dP(x)$$

for $n \in \mathcal{N}$ with $n \geq 1$.
Set $s_k := \int\limits_{[1-\frac{1}{3^{k-1}},1-\frac{1}{3^{k-1}}+\frac{1}{3^k}]} x dP(x)$.
Using Theorem 3.1 we obtain

$$(3) \qquad \begin{aligned} s_k &= \tfrac{1}{2}\int \tfrac{1}{3}x 1_{[1-\frac{1}{3^{k-1}},1-\frac{1}{3^{k-1}}+\frac{1}{3^k}]}(\tfrac{1}{3}x)P(x) \\ &\qquad + \tfrac{1}{2}\int (\tfrac{1}{3}x+\tfrac{2}{3}) 1_{[1-\frac{1}{3^{k-1}},1-\frac{1}{3^{k-1}}+\frac{1}{3^k}]}(\tfrac{1}{3}x+\tfrac{2}{3}) dP(x). \end{aligned}$$

For $k=1$ the second integral equals 0 and by Lemma 3.4 we get

$$(4) \qquad s_1 = \tfrac{1}{2}\int \tfrac{1}{3} x dP(x) = \tfrac{1}{12}.$$

For $k \geq 2$ the first integral vanishes and

$$1 - \frac{1}{3^{k-1}} \leq \frac{1}{3}x + \frac{2}{3} \leq 1 - \frac{1}{3^{k-1}} + \frac{1}{3^k}$$

is equivalent to

$$1 - \frac{1}{3^{k-2}} \leq x \leq 1 - \frac{1}{3^{k-2}} + \frac{1}{3^{k-1}},$$

hence

$$s_k = \frac{1}{2} \int_{[1-\frac{1}{3^{k-2}}, 1-\frac{1}{3^{k-2}}+\frac{1}{3^{k-1}}]} \frac{1}{3}x + \frac{2}{3} dP(x). \tag{5}$$

This implies

$$s_k = \frac{1}{6} s_{k-1} + \frac{1}{3} P([1 - \frac{1}{3^{k-2}}, 1 - \frac{1}{3^{k-2}} + \frac{1}{3^{k-1}}]). \tag{6}$$

It is easy to see that

$$P([1 - \frac{1}{3^{k-2}}, 1 - \frac{1}{3^{k-2}} + \frac{1}{3^{k-1}}]) = \frac{1}{2^{k-1}}.$$

Thus (6) turns to

$$s_k = \frac{1}{6} s_{k-1} + \frac{1}{3} \cdot \frac{1}{2^{k-1}}. \tag{7}$$

By induction (4) and (7) lead to

$$s_k = (\tfrac{1}{6})^{k-1} \tfrac{1}{12} + (\tfrac{1}{2})^k (1 - (\tfrac{1}{3})^{k-1}). \tag{8}$$

Using (8) and the definition of s_k the identity (2) turns to

$$\begin{aligned} \int_{[0,1-\frac{1}{3^n}]} x dP(x) &= \sum_{k=1}^{n} s_k \\ &= \sum_{k=1}^{n} \left[(\tfrac{1}{6}^{k-1} \tfrac{1}{12} + (\tfrac{1}{2})^k (1 - (\tfrac{1}{3})^{k-1}) \right] \\ &= \sum_{k=1}^{n} \left[(\tfrac{1}{2})^k - \tfrac{5}{12} (\tfrac{1}{6})^{k-1} \right] \\ &= \tfrac{1}{2} \frac{1-(\frac{1}{2})^n}{1-\frac{1}{2}} - \tfrac{5}{12} \frac{1-(\frac{1}{6})^n}{1-\frac{1}{6}} \\ &= 1 - (\tfrac{1}{2})^n - \tfrac{1}{2}(1 - (\tfrac{1}{6})^n). \end{aligned} \tag{9}$$

Combining (9) with (1) and the definition of F yields

$$F(1 - \frac{1}{3^n}) = 1 - \frac{1}{2} \left(\frac{1 - (\frac{1}{6})^n}{1 - (\frac{1}{2})^n} \right)$$

and (i) is proved.
To prove (ii) note that Theorem 3.1 implies

$$\begin{aligned} P([0, \tfrac{1}{3}t]) &= \tfrac{1}{2} \int 1_{[0,\frac{1}{3}t]} (\tfrac{1}{3}x) dP(x) \\ &= \tfrac{1}{2} P([0, t]) \end{aligned} \tag{10}$$

and

$$\begin{aligned} \int_{[0,\frac{1}{3}t]} x dP(x) &= \tfrac{1}{2} \int 1_{[0,\frac{1}{3}t]} (\tfrac{1}{3}x) \tfrac{1}{3} x dP(x) \\ &= \tfrac{1}{6} \int_{[0,t]} x dP(x). \end{aligned} \tag{11}$$

Using (10) and (11) in the definition of F yields

$$F(\frac{1}{3}t) = \frac{1}{3}F(t)$$

and proves (ii). Λ

LEMMA 18 *Let* $f :]-\frac{1}{2}, \frac{3}{2}] \to \mathcal{R}$ *be defined by*

$$f(t) = F\left(\frac{1}{4} + \frac{1}{2}t\right).$$

Then $f(t) > t$ *for all* $t \in]-\frac{1}{2}, \frac{1}{6} - \frac{2}{81}]$. Λ

PROOF For $t < 0$ the inequality $t < f(t)$ is obvious.
For $0 \leq t < \frac{1}{18}$ Lemma 4.2 implies

$$\begin{aligned} f(t) &= F(\tfrac{1}{4} + \tfrac{1}{2}t) \\ &\geq F(\tfrac{1}{4}) \\ &\geq F(\tfrac{2}{9}) = F(\tfrac{1}{3} - \tfrac{1}{3^2}) = \tfrac{1}{3}F(1 - \tfrac{1}{3}) \\ &= \tfrac{1}{3}\left(1 - \tfrac{1}{2}\left(\tfrac{1-\frac{1}{6}}{1-\frac{1}{2}}\right)\right) = \tfrac{1}{18} > t. \end{aligned}$$

For $\frac{1}{18} \leq t < \frac{19}{162}$ Lemma 4.2 yields

$$f(t) \geq F(\tfrac{1}{4} + \tfrac{1}{2} \cdot \tfrac{1}{18}) = F(\tfrac{5}{18}) \geq F(\tfrac{1}{3} - \tfrac{2}{27}).$$

Using $C \cap]\frac{1}{3} - \frac{2}{27}, \frac{1}{3} - \frac{1}{27}[= \emptyset$ and Lemma 4.2 we get

$$\begin{aligned} F(\tfrac{1}{3} - \tfrac{2}{27}) &= F(\tfrac{1}{3} - \tfrac{1}{27}) = F(\tfrac{1}{3} - \tfrac{1}{3^3}) \\ &= \tfrac{1}{3}F(1 - \tfrac{1}{3^2}) \\ &= \tfrac{19}{162} > t. \end{aligned}$$

For $\frac{19}{162} \leq t < \frac{1}{6} - \frac{2}{3^4}$ Lemma 4.2 implies

$$\begin{aligned} f(t) &\geq F(\tfrac{1}{4} + \tfrac{1}{2}\tfrac{19}{162}) = F(\tfrac{25}{81}) \\ &= F(\tfrac{1}{3} - \tfrac{2}{81}) \\ &= F(\tfrac{1}{3} - \tfrac{1}{3^4}) \\ &= \tfrac{1}{3}F(1 - \tfrac{1}{3^3}) \\ &= \tfrac{163}{1134} > \tfrac{161}{1134} = \tfrac{23}{162} = \tfrac{1}{6} - \tfrac{2}{3^4} > t. \end{aligned}$$

Thus the lemma is proved. Λ

THEOREM 19 *If α is an optimal set of 3–means then $\alpha \cap]\frac{1}{3}, \frac{2}{3}[= \emptyset$.* $\mathit{\Lambda}$

PROOF Using Remark 2.5 we know that $\alpha = \{a_1, a_2, a_3\}$ with $a_1 < a_2 < a_3$. By Lemma 4.1 we have $a_1 \in [0, \frac{1}{3}]$ and $a_3 \in [\frac{2}{3}, 1]$. We will show that $a_2 \notin]\frac{1}{3}, \frac{2}{3}[$. Assume the contrary.
Case 1: $a_2 \in [\frac{1}{2}, \frac{2}{3}[$.
Lemma 4.1 implies

(1) $\frac{1}{2}(a_1 + a_2) < \frac{1}{3}$

hence

(2) $a_1 < \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$.

Since $M(a_1) \cap [0, 1] = [0, \frac{1}{2}(a_1 + a_2)]$ it follows from Theorem 2.6(iii) and Lemma 4.2 that

$$\begin{aligned} a_1 = \frac{1}{P([0, \frac{a_1+a_2}{2}])} \int_{[0, \frac{a_1+a_2}{2}]} x dP(x) &= F(\tfrac{1}{2}(a_1 + a_2)) \\ &\geq F(\tfrac{1}{4} + \tfrac{1}{2}a_1). \\ &= f(a_1). \end{aligned}$$

By Lemma 4.3 this implies

(3) $a_1 \geq \frac{1}{6} - \frac{2}{3^4}$.

Using (1) we deduce

(4) $a_2 < \frac{2}{3} - a_1 \leq \frac{1}{2} + \frac{2}{3^4}$.

By Lemma 4.1 we have $\frac{1}{2}(a_2 + a_3) > \frac{2}{3}$, hence

(5) $a_3 > \frac{4}{3} - a_2 \geq \frac{5}{6} - \frac{2}{3^4}$.

Let $h : [0, 1] \to \mathcal{R}$ be defined by

$$h(x) = \min_{1 \leq i \leq 3} | x - a_i | .$$

For $0 \leq x \leq \frac{1}{9}$ we have

(6) $h(x) =| x - a_1 |\geq \frac{1}{6} - \frac{2}{3^4} - x \geq 0$.

For $\frac{2}{9} \leq x \leq \frac{1}{3}$ we have

(7) $h(x) = \min(| x - a_1 |, | x - a_2 |) \geq x - \frac{1}{6} \geq 0$.

For $\frac{2}{3} \leq x \leq \frac{7}{9}$ we have

(8) $h(x) = \min(| x - a_2 |, | x - a_3 |) \geq \frac{5}{6} - \frac{2}{3^4} - x \geq 0$

because (5) implies $a_3 - x \geq \frac{5}{6} - \frac{2}{3^4} - x \geq 0$ and (4) implies

$$\begin{aligned} x - a_2 &\geq x - (\tfrac{1}{2} + \tfrac{2}{3^4}) \geq \tfrac{2}{3} - (\tfrac{1}{2} + \tfrac{2}{3^4}) \\ &\geq \tfrac{1}{6} - \tfrac{2}{3^4} = \tfrac{5}{6} - \tfrac{2}{3^4} - \tfrac{2}{3} \geq \tfrac{5}{6} - \tfrac{2}{3^4} - x. \end{aligned}$$

Since $\alpha = \{a_1, a_2, a_3\}$ is optimal we have (using (6), (7) and (8) for the last inequality)

$$\begin{aligned}
V_3 &= \int \min_{1\le i\le 3}(x-a_i)^2\,dP(x)\\
&= \int_{[0,1]} h(x)^2\,dP(x)\\
&\ge \int_{[0,\frac{1}{9}]} h(x)^2\,dP(x) + \int_{[\frac{2}{9},\frac{1}{3}]} h(x)^2\,dP(x) + \int_{[\frac{3}{9},\frac{7}{9}]} h(x)^2\,dP(x)\\
&\ge \int_{[0,\frac{1}{9}]} (x-(\tfrac{1}{6}-\tfrac{2}{3^4}))^2\,dP(x) + \int_{[\frac{2}{9},\frac{1}{3}]} (x-\tfrac{1}{6})^2\,dP(x) + \int_{[\frac{2}{3},\frac{7}{9}]} (x-(\tfrac{5}{6}-\tfrac{2}{3^4}))^2\,dP(x).
\end{aligned}$$

Applying Corollary 3.3 to these integrals yields

$$\begin{aligned}
V_3 &\ge \tfrac{1}{2^2}\int(\tfrac{1}{9}x-(\tfrac{1}{6}-\tfrac{2}{3^4}))^2\,dP(x) &&+ \tfrac{1}{2^2}\int(\tfrac{1}{9}x+\tfrac{2}{9}-\tfrac{1}{6})^2\,dP(x)\\
& &&+ \tfrac{1}{2^2}\int(\tfrac{1}{9}x+\tfrac{2}{3}-(\tfrac{5}{6}-\tfrac{2}{3^4}))^2\,dP(x)\\
&= \tfrac{1}{18^2}\int(x-\tfrac{23}{18})^2\,dP(x) &&+ \tfrac{1}{18^2}\int(x+\tfrac{1}{2})^2\,dP(x)\\
& &&+ \tfrac{1}{18^2}\int(x-\tfrac{23}{18})^2\,dP(x).
\end{aligned}$$

We evaluate these integrals by Lemma 3.4 and obtain

$$\begin{aligned}
V_3 &\ge \tfrac{1}{18^2}[(\tfrac{23}{18}-\tfrac{1}{2})^2+\tfrac{1}{8}+(-\tfrac{1}{2}-\tfrac{1}{2})^2+\tfrac{1}{8}+(\tfrac{23}{18}-\tfrac{1}{2})^2+\tfrac{1}{8}]\\
&= \tfrac{1}{18^2}[\tfrac{3}{8}+(\tfrac{7}{9})^2+1+(\tfrac{7}{9})^2] \qquad (9)\\
&= \tfrac{1}{18^2}\tfrac{1675}{648} > \tfrac{5}{2}\cdot\tfrac{1}{18^2}.
\end{aligned}$$

By Corollary 3.8 we have

$$\begin{aligned}
V_3 &\le \tfrac{1}{18}\cdot\tfrac{1}{8}(4-3+\tfrac{1}{9}(3-2))\\
&= \tfrac{1}{18^2}\cdot\tfrac{5}{2}
\end{aligned}$$

so that (9) is a contradiction.

Case 2: $a_2 \in]\frac{1}{3},\frac{1}{2}[$
This leads to a contradiction in a similar way as case 1. Λ

Lemma 20 *Let $n \ge 2$ and let α be an optimal set of n–means with $\alpha \cap]\frac{1}{3},\frac{2}{3}[= \emptyset$. Set $\alpha_1 = \alpha\cap[0,\frac{1}{3}]$, $\alpha_2 = \alpha\cap[\frac{2}{3},1]$, and $j = card(\alpha_1)$. Then $3\alpha_1$ $(= S_1^{-1}(\alpha_1))$ is an optimal set of j–means and $3\alpha_2 - 2$ $(= S_2^{-1}(\alpha_2))$ is an optimal set of $(n-j)$–means. Moreover,*

$$V_n = \frac{1}{18}(V_j + V_{n-j}).$$

Λ

Proof By Lemma 4.1 we have $\alpha_1 \ne \emptyset$ and $\alpha_2 \ne \emptyset$. Since $\alpha = \alpha_1 \cup \alpha_2$ we deduce

$$(1)\quad V_n = \int_{[0,\frac{1}{3}]} \min_{a\in\alpha_1}(x-a)^2\,dP(x) + \int_{[\frac{2}{3},1]} \min_{a\in\alpha_2}(x-a)^2\,dP(x).$$

By the scaling properties of P (Theorem 3.1) the restriction of P to $[0, \frac{1}{3}]$ equals $\frac{1}{2}P \circ S_1^{-1}$ and the restriction of P to $[\frac{2}{3}, 1]$ equals $\frac{1}{2}P \circ S_2^{-1}$. Using the definitions of S_1 and S_2 this implies

(2) $V_n = \frac{1}{18} \int \min_{a \in \alpha_1} (x - 3a)^2 \, dP(x) + \frac{1}{18} \int \min_{a \in \alpha_2} (x - (3a - 2))^2 \, dP(x)$.

If $3\alpha_1$ were not an optimal set of j–means then we could find a set $\beta \subset \mathcal{R}$ with $card(\beta) = j$ and $\int \min_{b \in \beta} (x - b)^2 \, dP(x) < \int \min_{a \in \alpha_1} (x - 3a)^2 \, dP(x)$. But then $(\frac{1}{3}\beta) \cup \alpha_2$ is a set of cardinality n with $\int \min_{a \in \frac{1}{3}\beta \cup \alpha_2} (x - a)^2 \, dP(x) < \int \min_{a \in \alpha} (x - a)^2 \, dP(x)$, which contradicts the optimality of α. Similarly $3\alpha_2 - 2$ can be shown to be an optimal set of $(n - j)$–means. Thus (2) implies $V_n = \frac{1}{18}(V_j + V_{n-j})$. Λ

THEOREM 21 *(i) The only optimal set of* 1*–means is* $\{\frac{1}{2}\}$. *Moreover,* $V_1 = \frac{1}{8}$.

(ii) The only optimal set of 2*–means consists of the midpoints of* $[0, \frac{1}{3}]$ *and* $[\frac{2}{3}, 1]$. *Moreover,* $V_2 = \frac{1}{9} V_1$.

(iii) There are two optimal sets of 3*–means. One consists of the midpoints of* $[0, \frac{1}{9}], [\frac{2}{9}, \frac{1}{3}]$ *and* $[\frac{2}{3}, 1]$, *the other of the midpoints of* $[0, \frac{1}{3}], [\frac{2}{3}, \frac{7}{9}]$, *and* $[\frac{8}{9}, 1]$. *Moreover,* $V_3 = \frac{1}{18}(V_1 + V_2)$.

Λ

PROOF

(i) is true since $\frac{1}{2}$ is the expected value of P (Lemma 3.4).

(ii) An optimal set of 2–means does not contain a point from $]\frac{1}{3}, \frac{2}{3}[$ (Lemma 4.1). Thus Lemma 4.5 together with (i) yields that the set of midpoints of $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ is the only candidate for an optimal set of 2–means. By the existence result for n–means (Proposition 1) it, therefore, is the only set of 2-means. Lemma 4.5 also implies $V_2 = \frac{1}{18}(V_1 + V_1) = \frac{1}{9} V_1$.

(iii) By Proposition 4.4 an optimal set α of 3–means does not contain a point from $]\frac{1}{3}, \frac{2}{3}[$. Set $\alpha_1 = \alpha \cap [0, \frac{1}{3}]$ and $\alpha_2 = \alpha \cap [\frac{2}{3}, 1]$. If $card(\alpha_1) = 2$ then $3\alpha_1$ is an optimal set of 2–means (Lemma 4.5). Hence (ii) implies that α_1 consists of the midpoints of $[0, \frac{1}{9}]$ and $[\frac{2}{9}, \frac{1}{3}]$. Similarly, α_2 consists of the midpoint of $[\frac{2}{3}, 1]$. If $card(\alpha_1) = 1$ then we get the set of midpoints of $[0, \frac{1}{3}], [\frac{2}{3}, \frac{7}{9}]$, and $[\frac{8}{9}, 1]$ for α. By Lemma 4.5 there are no other possibilities for an optimal set of 3–means and $V_3 = \frac{1}{18}(V_1 + V_2)$. Using the last equality and Proposition 3.7 we see that the two sets are indeed optimal sets of 3–means.

Λ

5 Determination of the quantization error and the optimal sets of n–means in the general case.

In this section we will compute V_n and determine all optimal sets of n–means for the Cantor probability P.

THEOREM 22 *Let $n \geq 2$ and let α be an optimal set of n–means. Then $\alpha \cap]\frac{1}{3}, \frac{2}{3}[= \emptyset$.* Λ

PROOF For $n < 4$ this follows from Lemma 4.1 and Proposition 4.4.
For $n \geq 4$ let j be as in Lemma 4.1. Assume $a_{j+1} \in]\frac{1}{3}, \frac{2}{3}[$.
Case 1: $a_{j+1} \geq \frac{1}{2}$
Then

$$V_n \geq \int_{[0,\frac{1}{3}]} \min_{1 \leq i \leq j+1} (x - a_i)^2 \, dP(x). \tag{1}$$

Since $\frac{1}{2}(a_j + a_{j+1}) < \frac{1}{3}$ by Lemma 4.1 we obtain

$$a_j < \frac{2}{3} - a_{j+1} < \frac{1}{6}.$$

For $x \in [\frac{2}{9}, \frac{1}{3}]$ this implies

$$\min_{1 \leq i \leq j+1} (x - a_i)^2 = \min_{i \in \{j, j+1\}} (x - a_i)^2 \geq (x - \tfrac{1}{6})^2 \tag{2}$$

because $x - a_j > x - \frac{1}{6} > 0$ and $\quad a_{j+1} - x > \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \geq x - \frac{1}{6} \geq 0.$
Combining (1) and (2) yields

$$V_n \geq \int_{[\frac{2}{9},\frac{1}{3}]} (x - \tfrac{1}{6})^2 \, dP(x) \tag{3}$$

Applying Corollary 3.3 to the last integral we get

$$\begin{aligned} V_n &\geq \tfrac{1}{4} \int (\tfrac{1}{9}x + \tfrac{2}{9} - \tfrac{1}{6})^2 \, dP(x) \\ &= \tfrac{1}{18^2} \int (x + \tfrac{1}{2})^2 \, dP(x) \\ &= \tfrac{1}{18^2} \tfrac{9}{8}. \end{aligned}$$

By Proposition 3.7 we have the contradiction $\frac{1}{2} \cdot \frac{1}{18^2} \geq V_4 \geq V_n$

Case 2: $a_{j+1} < \frac{1}{2}$
A contradiction can be obtained in a similar way (reflecting the situation at $\frac{1}{2}$). Λ

THEOREM 23 *For $n \in \mathcal{N}$ with $n \geq 1$ let $l(n) \in \mathcal{N}$ satisfy $2^{l(n)} \leq n < 2^{l(n)+1}$. A set $\alpha \subset \mathcal{R}$ is an optimal set of n–means if and only if there exists a subset $I \subset \{1,2\}^{l(n)}$ with $card(I) = n - 2^{l(n)}$ and such that α consists of all midpoints of intervals J_σ with $\sigma \in \{1,2\}^{l(n)} \setminus I$ and all midpoints of the children of J_σ with $\sigma \in I$ (i.e. $\alpha = \alpha_n(I)$). Moreover,*

$$V_n = \frac{1}{18^{l(n)}} \cdot \frac{1}{8}(2^{l(n)+1} - n + \frac{1}{9}(n - 2^{l(n)})).$$

Λ

Proof We will proceed by induction on n. For $n < 4$ the statement of the theorem is proved in Proposition 4.6.

Suppose that the assertion of the theorem holds for all $m < n, n \geq 2$. According to Proposition 5.1 and Lemma 4.5 there exists a $j \in \{1, \ldots, n-1\}$ with

(1) $$V_n = \tfrac{1}{18}(V_j + V_{n-j}).$$

Without loss of generality we may assume $j \geq n - j$.

We will show that

(2) $$2^{l(n)-1} \leq j \leq 2^{l(n)} \text{ and } 2^{l(n)-1} \leq n - j < 2^{l(n)}.$$

Let $k, l \in \mathcal{N}$ be such that

(3) $$2^k \leq j < 2^{k+1} \text{ and } 2^l \leq n - j < 2^{l+1}.$$

Since $j \geq n - j$ we have $j \geq \frac{1}{2}n \geq 2^{l(n)-1}$ and $n - j \leq \frac{1}{2}n < 2^{l(n)}$ hence

(4) $$k \geq l(n) - 1 \text{ and } l \leq l(n) - 1.$$

In particular we have

(5) $$k \geq l.$$

Using Corollary 3.8 and the induction hypothesis, (1) implies

(6) $$\begin{aligned} \tfrac{1}{18^{l(n)}} \cdot \tfrac{1}{8} \quad & (2^{l(n)+1} - n + \tfrac{1}{9}(n - 2^{l(n)})) \\ & \geq \tfrac{1}{18^{k+1}} \cdot \tfrac{1}{8}(2^{k+1} - j + \tfrac{1}{9}(j - 2^k)) \\ & \ + \tfrac{1}{18^{l+1}} \cdot \tfrac{1}{8}(2^{l+1} - (n-j) + \tfrac{1}{9}(n - j - 2^l)) \quad . \end{aligned}$$

By some easy reformulations this leads to

(7) $$2(17 - 8\tfrac{n}{2^{l(n)}}) \geq 9^{l(n)-k-1}(17 - 8\tfrac{j}{2^k}) + 9^{l(n)-l-1}(17 - 8\tfrac{n-j}{2^l}).$$

Since $\frac{n}{2^{l(n)}} \geq 1$, $\frac{j}{2^k} < 2$, and $\frac{n-j}{2^l} < 2$ we deduce

(8) $$18 > 9^{l(n)-k-1} + 9^{l(n)-l-1}.$$

Since $k \leq l(n)$ we deduce

(9) $$l(n) - 2 \leq l.$$

Thus the possible choices for k and l are

a) $k = l(n) - 1$ and $l = l(n) - 2$

b) $k = l(n) - 1$ and $l = l(n) - 1$

c) $k = l(n)$ and $l = l(n) - 1$

d) $k = l(n)$ and $l = l(n) - 2$

For a) we obtain by a reformulation of (7)

$$(n-j) \geq 2^{l(n)-1}$$

hence $l \geq l(n)-1$, a contradiction.
For c) we obtain by a reformulation of (7)

$$j \leq 2^{l(n)}.$$

Since $\quad 2^{l(n)} \leq j < 2^{l(n)+1}$ this yields

$$j = 2^{l(n)}.$$

For d) we obtain by a reformulation of (7)

$$272\tfrac{n-1}{2^{l(n)}} \geq \tfrac{64}{9}\cdot 17 + \tfrac{136}{9}\tfrac{j}{2^{l(n)}} \qquad .$$

Since $1 \leq \frac{j}{2^{l(n)}}$ we deduce

$$n-j \geq 2^{l(n)-1}$$

which is a contradiction, since

$$2^{l(n)-2} \leq n-j < 2^{l(n)-1}.$$

Thus we have proved (2).
¿From (2) we get

$$\begin{aligned} V_n = \tfrac{1}{18}(V_j + V_{n-j}) &= \tfrac{1}{18}(\tfrac{1}{18^{l(n)-1}}\tfrac{1}{8}(2^{l(n)} - j + \tfrac{1}{9}(j - 2^{l(n)-1}) \\ &+ \tfrac{1}{18^{l(n)-1}}\cdot\tfrac{1}{8}(1^{l(n)} - (n-j) + \tfrac{1}{9}(n-j-2^{l(n)-1})) \\ &= \tfrac{1}{18^{l(n)}}\tfrac{1}{8}(2^{l(n)+1} - n + \tfrac{1}{9}(n - 2^{l(n)})). \end{aligned} \tag{10}$$

Now let α be an optimal set of n–means. By Lemma 4.5 $\alpha_1 := S_1^{-1}(\alpha \cap [0, \frac{1}{3}])$ and $\alpha_2 := S_2^{-1}(\alpha \cap [\frac{2}{3}, 1])$ are optimal sets of j – resp. $(n-j)$ – means. According to the induction hypothesis there exists a subset $I_1 \subset \{1,2\}^{l(n)-1}$ with $card(I_1) = j - 2^{l(n)-1}$ and $I_2 \subset \{1,2\}^{l(n)-1}$ with $card(I_2) = n-j-2^{l(n)-1}$ such that $\alpha_1 = \alpha_j(I_1)$ and $\alpha_2 = \alpha_{n-j}(I_2)$. Then it is easy to check that, for $I = \{\tau \in \{1,2\}^{l(n)} : (\tau_1 = 1 \wedge (\tau_2, \ldots, \tau_{l(n)}) \in I_1)$ or $(\tau_1 = 2 \wedge (\tau_2, \ldots, \tau_{l(n)}) \in I_2\}$ we get

$$\begin{aligned} card\ I = card\ I_1 + card\ I_2 &= j - 2^{l(n)-1} + n - j - 2^{l(n)-1} \\ &= n - 2^{l(n)} \end{aligned}$$

and

$$\alpha = \alpha_n(I).$$

That every set α as described in the statement of the theorem is an optimal set of n–means follows from (10) and Propositionn 3.7. Thus Theorem 5.2 is proved. Λ

6 Asymptotics for V_n and the quantization dimension of the Cantor measure

In this section we will study the asymptotics of V_n. In what follows $\beta = \frac{\log\ 2}{\log\ 3}$ is the Hausdorff dimension of the Cantor distribution P which is the same as the Hausdorff dimension of the Cantor set. As pointed out in the introduction we will show that the sequence $(n^{\frac{2}{\beta}} V_n)_{n\in\mathcal{N}}$ does not converge. This contrasts the corresponding result for absolutely continuous distributions, where $\lim_{n\to\infty} n^{\frac{2}{d}} V_n$ exists. But first we will introduce a new concept of dimension for distributions (see also Zador[7]).

DEFINITION 24 The **quantization dimension** of a probability P is defined to be the number

$$\lim_{n\to\infty} -\frac{log\ n}{log(V_n^{1/2})},$$

if the limit exists in $\mathcal{R}$.

We will show that the quantization dimension of P equals β. Λ

LEMMA 25 *Define* $f : [1,2] \to \mathcal{R}$ *by*

$$f(x) = \frac{1}{72} x^{\frac{2}{\beta}} (17 - 8x)$$

Then

(i) $1 \le \frac{17}{8+4\beta} \le 2$

(ii) $f(1) = \frac{1}{8} = f(2), f(\frac{17}{8+4\beta}) \ge f(x)$ *for all* $x \in [1,2]$
and $f([1,2]) = [\frac{1}{8}, f(\frac{17}{8+4\beta})]$.

Λ

PROOF (i) is obviously true since

$$8 + 4\beta \le 12 \le 17$$

and

$$17 \le 2 \cdot (8 + 4\beta).$$

(ii) A direct calculation shows $f(1) = \frac{1}{8} = f(2)$. Now

$$f'(x) = \frac{17}{36\beta} x^{\frac{2}{\beta}-1} - \frac{1}{9}(\frac{2}{\beta} + 1) x^{\frac{2}{\beta}}.$$

Thus

$$f'(1) = \frac{1}{36\beta}(17 - (8 + 4\beta)) = \frac{1}{36\beta}(9 - 4\beta) > 0$$

and

$$\begin{aligned} f'(2) &= \tfrac{17}{36\beta}\tfrac{1}{2}9 - \tfrac{1}{9}(\tfrac{2}{\beta}+1)9 \\ &= \tfrac{153}{72\beta} - \tfrac{2}{\beta} - 1 = \tfrac{153-144-72\beta}{72\beta} < 0. \end{aligned}$$

Moreover, for $x \in [1,2]$,

$$\begin{aligned} f'(x) = 0 &\Leftrightarrow \tfrac{1}{36\beta} x^{\frac{2}{\beta}-1}(17-(8+4\beta)x) = 0 \\ &\Leftrightarrow x = \tfrac{17}{8+4\beta}. \end{aligned}$$

This implies that f has its maximum on $[1,2]$ at the point $\frac{17}{8+4\beta}$ and that $f([1,2])$ is as described in (ii). Λ

THEOREM 26 *The set of accumulation points of the sequence* $\left(\frac{2}{n^\beta} V_n\right)_{n\in\mathcal{N}}$ *equals*

$$[\frac{1}{8}, f(\frac{17}{8+4\beta})],$$

where f is as in Lemma 6.2 Λ

PROOF Let $y \in [\frac{1}{8}, f(\frac{17}{8+4\beta})]$ be given. Then there exists an $x \in [1,2]$ with $y = f(x)$. Set $n_l = [x2^l]$, where $[x2^l]$ denotes the greatest integer less than or equal to $x2^l$. Then $n_l < n_{l+1}$, $l(n_l) = l$, and there exists an $x_l \in [1,2]$ with $n_l = x_l 2^l$. Thus, we obtain

$$\begin{aligned} n_l^{\frac{2}{\beta}} V_{n_l} &= n_l^{\frac{2}{\beta}} \tfrac{1}{8}\tfrac{1}{18^l}(2^{l+1} - n_l + \tfrac{1}{9}(n_l - 2^l)) \\ &= \tfrac{1}{8} x_l^{\frac{2}{\beta}} (2^{\frac{2}{\beta}})^l \tfrac{1}{18^l}(2^{l+1} - x_l 2^l + \tfrac{1}{9}(x_l 2^l - 2^l)) \\ &= \tfrac{1}{8} x_l^{\frac{2}{\beta}} 9^l \tfrac{1}{18^l} 2^l (\tfrac{17}{9} - \tfrac{8}{9} x_l) \\ &= \tfrac{1}{8} x_l^{\frac{2}{\alpha}} (\tfrac{17}{9} - \tfrac{8}{9} x_l) \\ &= f(x_l). \end{aligned}$$

Now

$$x_l 2^l \leq x2^l < x_l 2^l + 1$$

hence

$$x - \frac{1}{2^l} < x_l \leq x$$

and, therefore, $\lim_{l\to\infty} x_l = x$.
Since f is continuous this implies

$$\lim_{l\to\infty} n_l^{\frac{2}{\beta}} V_{n_l} = f(x) = y.$$

Now let y be an accumulation point of the sequence $\left(n^{\frac{2}{\beta}} V_n\right)_{n\in\mathcal{N}}$. Then there exists a subsequence $\left(n_k^{\frac{2}{\beta}} V_{n_k}\right)_{k\in\mathcal{N}}$ with

$$\lim_{k\to\infty} n_k^{\frac{2}{\beta}} V_{n_k} = y.$$

Set $l_k = l(n_k)$ and $x_k = \frac{n_k}{2^{l_k}}$. Then we have $x_k \in [1,2]$ and

$$n_k^{\frac{2}{\beta}} V_{n_k} = f(x_k).$$

Now we choose a convergent subsequence (x_{k_l}) of (x_k) and we obtain

$$y = \lim_{k\to\infty} n_k^{\frac{2}{\beta}} V_{n_k} = \lim_{l\to\infty} f(x_{k_l}) = f(\lim_{l\to\infty} x_{k_l}) \in [\frac{1}{8}, f(\frac{17}{8+4\beta})].$$

Λ

LEMMA 27 *For all $n \in \mathcal{N}$,*

$$\frac{1}{72} \leq n^{\frac{2}{\beta}} V_n \leq \frac{9}{8}.$$

Λ

PROOF For $n \in \mathcal{N}$ we have

$$n^{\frac{2}{\beta}} V_n \geq (2^{l(n)})^{\frac{2}{\beta}} V_{2^{l(n)+1}} = 9^{l(n)} \cdot \frac{1}{8}\frac{1}{18^{l(n)}}\frac{1}{9}2^{l(n)} = \frac{1}{72}$$

and

$$n^{\frac{2}{\beta}} V_n \leq (2^{l(n)+1})^{\frac{2}{\beta}} V_{2^{l(n)}} = 9^{l(n)+1} \cdot \frac{1}{18^{l(n)}}2^{l(n)} = \frac{9}{8}.$$

Λ

REMARK 28 For $k \in \mathcal{N}$, let $\alpha_k = \{S_\sigma(\frac{1}{2}) : \sigma \in \{1,2\}^k\}$ be the (uniquely determined) optimal set of 2^k–means for the Cantor distribution (see Theorem 5.2) and let

$$\begin{aligned} Q_k &= \sum_{a\in\alpha_k} P(\{x \in \mathcal{R} :\mid x-a \mid = \min_{b\in\alpha_k} \mid x-b \mid\}\varepsilon_a \\ &= \tfrac{1}{2^k} \sum_{a\in\alpha_k} \varepsilon_a. \end{aligned}$$

Then $(V_{2^k})^{1/2} = l_2(P, Q_k)$, where l_2 denotes the minimal L^2–metric (see Pollard [6], Graf–Luschgy [4]). Thus we obtain from Theorem 6.3

$$l_2(P, Q_k) = O(3^{-k}) \text{ as } k \to \infty.$$

$\diamondsuit$

THEOREM 29 *The quantization dimension of P equals the Hausdorff dimension β of the Cantor set.* Λ

PROOF By Lemma 6.4 we have

$$\frac{1}{72}n^{-\frac{2}{\beta}} \leq V_n \leq \frac{9}{8}n^{-\frac{2}{\beta}}$$

hence

$$\frac{1}{2}log\frac{1}{72} - \frac{1}{\beta}log\ n \leq log V_n^{1/2} \leq \frac{1}{2}log\frac{9}{8} - \frac{1}{\beta}log\ n.$$

This implies

$$\frac{-log\ n}{\frac{1}{2}log\frac{1}{72}-\frac{1}{\beta}log\ n} \le -\frac{log\ n}{log(V_n^{1/2})} \le \frac{-log\ n}{\frac{1}{2}log\frac{9}{8}-\frac{1}{2}log\ n}.$$

Since

$$\lim_{n\to\infty}\frac{-log\ n}{\frac{1}{2}log\frac{1}{72}-\frac{1}{\beta}log\ n}=\beta$$

and

$$\lim_{n\to\infty}\frac{-log\ n}{\frac{1}{2}log\frac{9}{8}-\frac{1}{\beta}log\ n}=\beta$$

we get

$$\lim_{n\to\infty}-\frac{log\ n}{log(V_n^{1/2})}=\beta$$

and the theorem is proved. Λ

References

[1] H. Abut(ed.), Vector quantization. IEEE Press, New York 1990

[2] J.A. Bucklew - G.L. Wise, Multidimensional asymptotic quantization with rth power distortion measures, IEEE Trans. Inform. Theory 28(1982), 239–247

[3] S. Graf - H. Luschgy, Foundations of quantization for random vectors, Preprint No. 16, Angew. Mathematik und Informatik, Universität Münster 1994

[4] S. Graf - H. Luschgy, Consistent Estimation in the quantization problem for random vectors, in: Transactions of the 12^{th} Prague Conference on Information Theory, Statistical Decission Functions, Random Processes, Acad. Sci. Czech. Republic & Charles Univ. Prague, Prague 1994, pp. 84–87

[5] J. Hutchinson, Fractals and self–similarity. Indiana Univ. J. 30(1981), 713–747

[6] D. Pollard, Quantization and the method of k–means, IEEE Trans. Inform. Theory 28(1982), 199–205

[7] P.L. Zador, Asymptotic quantization error of continuous signals and the quantization dimension, IEEE Trans. Inform. Theory 28(1982), 139–249

S. Graf
Fakultät für Mathematik
und Informatik
Universität Passau
D–94030 Passau

H. Luschgy
Institut für Mathematische
Statistik
Universität Münster
Einsteinstr. 62
D-48149 Münster