Quantization of self-similar probability measures and optimal quantizers

by

Doğan Çömez

Department of Mathematics, North Dakota State University

Outline:

1. Five-minute Crash Course in Probability
2. Probability Measures and Distributions
3. Voronoi Partitions
4. Centers and Moments of Probability Distributions
5. Quantization
6. Some Recent Results

Throughout we will deal with $\mathbb{R}^n$ with a norm $\| \ \|$. In particular $\| \ \|_r$ will be the r-norm for some $1 \leq r < \infty$. Some of the information below are adopted from *Foundations of Quantization for Probability Distributions*, by S. Graf and H. Luschgy, Lecture Notes in Mathematics, # 1730, **Springer**, 2000.

1. Five-minute Crash Course in Probability.

$(\Omega, \mathcal{F}, P)$: a probability space.
A measurable function $X : \Omega \to \mathbb{R}$ is called a *random variable*.
The *expectation* (or *mean*) of X is defined as $E(X) := \int_\Omega X dP$.

(i) $E(X^k)$: *k-th moment* of X; and $E(|X|^k)$: *k-th absolute moment* of X.
(ii) $E[(X - E(X))^k]$: *k-th central moment* of X; and $E[|X - E(X)|^k]$: *k-th absolute central moment* of X.

Note: If $k = 1$, $E[(X - E(X))^k] = 0$. When $k = 2$, $E[(X - E(X))^2]$ is called the *variance* of X.

If X is a random variable, then it induces a probability measure on Borel subsets of $\mathbb{R}$ by $P_X := P \circ X^{-1}$. (Sometimes P_X is called a *law*.)

The *distribution (function) of a r.v.* X is a function $F : \mathbb{R} \to \mathbb{R}$ such that

$$F(x) = P_X(-\infty, x] = P(\{\omega : X(\omega) \leq x\}).$$

Observe that $F(b) - F(a) = P_X(a, b]$.
Fact. A function $F : \mathbb{R} \to \mathbb{R}$ is a probability distribution if and only if it's non-decreasing, continuous from right (i.e., $\lim_{x \to a^+} F(x) = F(a)$) and satisfies

$$\lim_{x \to -\infty} F(x) = 0, \text{ and } \lim_{x \to \infty} F(x) = 1.$$

If F is a function satisfying these properties, there is a unique r.v. X whose distribution is F.

All these are closely connected with Lebesgue-Stieltjes measures and their distributions. Recall that a measure $\mu : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ is called a *Lebesgue-Stieltjes measure* if $\mu(I) < \infty$ for any bounded interval I. In general, a function $F : \mathbb{R} \to \mathbb{R}$ which is non-decreasing and continuous from right is called a *distribution* function. Via

$$\mu(a, b] = F(b) - F(a),$$

there is a 1-1 correspondence between Lebesgue-Stieltjes measures and distribution functions.

A non negative Borel-measurable function f on $\mathbb{R}$ is called a *density* of a r.v. X if

$$F(x) = \int_{(\infty,x]} f(t)dt,$$

where F is the distribution function of X. Hence,

$$P_X(A) = \int_A f(x)dx, \quad A \in \mathcal{B}(\mathbb{R}).$$

That is, f is the Radon-Nikodym derivative of P_X w.r.t. Lebesgue measure.

Random vectors. A measurable function A measurable function $X : \Omega \to \mathbb{R}^n$ is called an *n-dimensional random variable* or a *random vector*. If Π_i is the projection on the i-th coordinate, then $X_i = \Pi_i \circ X$ is a r.v. for each $1 \leq i \leq n$.

Note: $X = (X_1, X_2, \ldots, X_n)$ is Borel measurable if and only if each X_i is Borel measurable. If X is a random vector, then

(i) associated law on Borel subsets of $\mathbb{R}$ is defined the same way by $P_X(B) = P \circ X^{-1}(B), \quad B \in \mathcal{B}(\mathbb{R})$.

(ii) associated distribution function (called *joint distribution*)is defined by

$$F(x) = P_X(-\infty, x] = P(\{\omega : X_i(\omega) \leq x, \ 1 \leq i \leq n\}), \quad x \in \mathbb{R}^n.$$

2. Probability Measures and Distributions.

It is not unusual in probability theory to talk about properties/behavior of random variable X with a distribution function F without any reference to the underlying probability space. This is due to the fact that F determines P_X, which in turn, determines all events involving X (i.e., the σ-algebra needed to study X). So, all one needs to check is that there is a probability space on which X is well-defined. Hence, distribution functions play a crucial role in probability spaces.

There exist only three types of pure probability distributions: discrete, absolutely continuous, and singularly continuous (w..r.t Lebesgue measure).

A r.v. X is called *discrete* if the set of values $\{x_n\}$ of it is countable. In this case the distribution function is a step function F with a discontinuity at each x_n (with magnitude $p_n = P_X(\{x_n\})$. F is also called *discrete distribution.*

A r.v. X is called *absolutely continuous* if $P_X << m$, where m is the Lebesgue measure. Equivalently, the associated distribution function F is absolutely continuous on $\mathbb{R}$.

A r.v. X is called *continuous* if its distribution function is continuous on $\mathbb{R}$.

Densities of (continuous) probability distributions.

1. Uniform density on $[a, b]$:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x < b, \\ 0 & \text{otherwise.} \end{cases}$$

2. Exponential density:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 \ \text{ if } \ x < 0, \ \text{ where } \ \lambda > 0 \end{cases}$$

3. Two sided exponential density:

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}, \quad \text{where} \quad \lambda > 0.$$

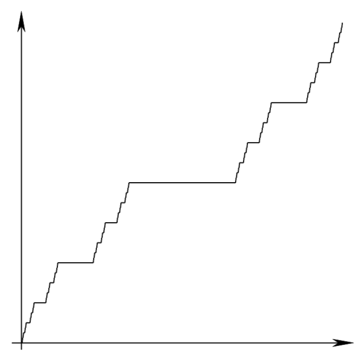
4. Normal density:
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \sigma > 0,\ m \in \mathbb{R}.$$

5. Cauchy density:
$$f(x) = \frac{\theta}{\pi(x^2+\theta^2)}, \quad \theta > 0.$$

All these densities induce absolutely continuous (hence, continuous) distributions.

Recall that given two measures μ and ν, by the Lebesgue decomposition theorem $\nu = \nu_a + \nu_s$, where $\nu_a << \mu$ and $\nu_s \perp \mu$. Then the Radon-Nikodym derivative $\frac{d\nu}{d\mu}$ is defined to be $\frac{d\nu_a}{d\mu}$. If μ is Lebesgue measure, then there is a further decomposition of the singular part of ν into *atomic* and a *singular continuous* part.

The prime example of a singular continuous measure is defined by the distribution function
$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ C(x) & \text{if } 0 \le x < 1, \\ 1 & \text{if } \quad x \ge 1, \end{cases}$$
where $C(x)$ is the *Cantor-Lebesgue function* (aka, *Devil's staircase*) on $[0, 1]$:



Cantor-Lebesgue function $C(x)$

Observe that

(i) $F'(x) = 0$ a.e.,

(ii) F is continuous but not absolutely continuous,

Hence, the measure ν defined by F is singular w.r.t. Lebesgue measure. It follows from the construction that support of ν is the Cantor set (an uncountable set).

F. Riesz constructed a purely singular continuous measure with support all of $[0, 1]$!

There is another convenient method of constructing continuous measures that are not necessarily absolutely continuous. This method involves IFS type fractals which are attractors of some special maps on $\mathbb{R}^n$.

A map $S : \mathbb{R}^n \to \mathbb{R}^n$ is called a *similarity transformation* if $\exists$ a constant $0 < s < 1$ (called *contracting factor*) such that
$$\|S(x) - S(y)\| = s\|x - y\|, \quad \forall\ x, y \in \mathbb{R}^n.$$
Let $\{S_1, S_2, \ldots, S_N\}$ be similarity transformations, then there is a unique nonempty compact set $A \subset \mathbb{R}^n$ with
$$A = \cup_{k=1}^{N} S_k(A). \quad \text{[Hutchinson, 1981]}$$

The set A is called the *attractor* or *invariant set* of $\{S_1, S_2, \ldots, S_N\}$. Furthermore, if $(p_1, p_2, \ldots, p_N)$ is a probability vector (i.e., $p_i \geq 0$ and $\sum_{i=1}^N p_i = 1$), then there is a unique probability measure P on $\mathcal{B}(\mathbb{R}^{\ltimes})$ with

$$P = \sum_{i=1}^{N} p_i (P \circ S_i^{-1}). \quad \text{[Hutchinson, 1981]}$$

If $p_i > 0$ for all $1 \leq i \leq N$, then the support of P is the attractor A. This measure P is called the *self-similar measure* associated with $\{S_1, S_2, \ldots, S_N\}$.

In the case that the collection $\{S_1, S_2, \ldots, S_N\}$ of similarities satisfy some additional conditions (such as *open set condition*, or *strong separation condition*), then the associated attractor is a fractal with n-dimensional Lebesgue measure 0. Hence, P is singular w.r.t. Lebesgue measure.

3. Voronoi Partitions.

Generally speaking, a Voronoi diagram is the partitioning of a (plane) region with n points (called *generating points*) into subregions (convex polygons) such that each subregion contains exactly one generating point and every point in a given subregion is closer to its generating point than to any other.

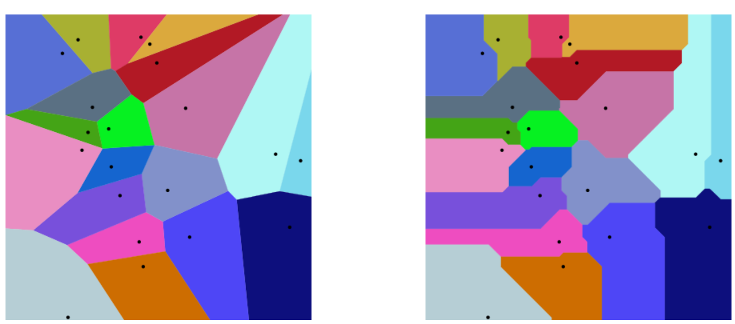
Let $\alpha \subset \mathbb{R}^d$ be a (finite or locally finite) nonempty subset. The *Voronoi region generated by* $a \in \alpha$ is the set defined by

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

i.e., the Voronoi region generated by $a \in \alpha$ is the set of all points in $\mathbb{R}^d$ which are closest to $a \in \alpha$. The set α is called the *generator* and the collection $\{W(a|\alpha) : a \in \alpha\}$ is called the *Voronoi diagram* or *Voronoi tessellation* of $\mathbb{R}^d$ with respect to α.

Let P be a Borel probability measure on $\mathbb{R}^d$. A Borel measurable partition $\{A_a : a \in \alpha\}$ of $\mathbb{R}^d$ is called a *Voronoi partition* of $\mathbb{R}^d$ with respect to α (and P) if

$$A_a \subset W(a|\alpha) \ (P\text{-a.e.}) \text{ for every } a \in \alpha.$$



Voronoi diagrams w.r.t Euclidean metric d_2 (L) and w.r.t. Taxi-cab metric d_1 (R)

Remarks. The following are some properties of Voronoi regions:

1. Voronoi regions depend on the metric (norm); hence so does the associated Voronoi diagram.
2. A Voronoi diagram is a (locally) finite covering for $\mathbb{R}^d$; hence, the number of sets in $\{W(a|\alpha) : a \in \alpha\}$ intersecting any bounded subset of $\mathbb{R}^d$ is finite.
3. If $W_0(a|\alpha) = \{x \in \mathbb{R}^d : \|x - a\| < \min_{b \in \alpha, b \neq a} \|x - b\|\}$, then $W_0(a|\alpha) \subset A_a$ for all $a \in \alpha$.
4. Let $\beta = \{b \in \alpha : W(b|\alpha) \cap W(a|\alpha) \neq \emptyset$, then $W(a|\beta) = W(a|\alpha)$.

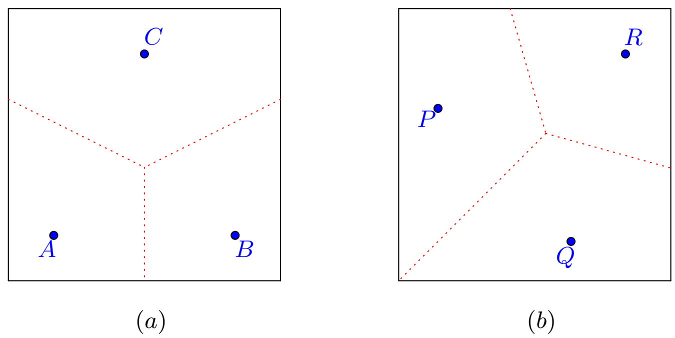


FIGURE 1. (a) Voronoi diagram generated by $\alpha = \{A, B, C\}$; (b) Voronoi diagram generated by $\beta = \{P, Q, R\}$ of the square in $\| \|_2$-norm.

5. The sets $W(a|\alpha)$ are all star-shaped relative to their generator a. Furthermore, for all $a \in \alpha$, $a \in int(W(a|\alpha))$.
6. If the norm $\| \|$ is strictly convex, then $int(W(a|\alpha) = W_0(a|\alpha)$ and $W(a|\alpha) = \overline{W_0(a|\alpha)}$.
7. If $d = 2$ or the norm $\| \|$ is strictly convex, then $m(\partial W(a|\alpha)) = 0$, where m is the d-dimensional Lebesgue measure.
8. In Euclidean spaces Voronoi regions are always convex. The converse statement is also valid; namely, if $W(a|\alpha)$ is convex for all finite $\alpha \subset \mathbb{R}^d$, then the underlying norm $\| \|$ of $\mathbb{R}^d$ is Euclidean. Hence, convexity of Voronoi regions is a characterization of Euclidean norms! This is a classical result due to H. Mann (1935).

Let $B \subset \mathbb{R}^d$ be a Borel set and μ be a Borel measure on $\mathbb{R}^d$. A *μ-tessellation of B* is a countable covering $\{C_n\}_{n\in\mathbb{N}}$ of B by Borel sets $C_n \subset B$ such that $\mu(C_n \cap C_m) = 0$ for $m \neq n$. In general, the Voronoi diagram of a set $\alpha \subset \mathbb{R}^d$ need not be a tessellation of $\mathbb{R}^d$; however, if $\| \|$ is a strictly convex, then it is a tessellation.

4. Centers and Moments of Probability Distributions.

Let $X = (X_1, X_2, \dots, X_d)$ be a $\mathbb{R}^d$-valued random variable with distribution P such that $E\|X\|^r < \infty, \quad 1 \le r < \infty$.

- A point $a \in \mathbb{R}^d$ is called a *center of P (of order r)* if $E\|X - a\|^r = \inf_{b\in\mathbb{R}^d} E\|X - b\|^r$. In general, centers of P need not be unique. The set of all centers of P (of order r) is denoted by $C_r(P)$.
- The value $V_r(P) = \inf_{a\in\mathbb{R}^d} E\|X - a\|^r$, where $a \in C_r(P)$, is called the *r-th moment of P about the center.*

Note. If $\| \|$ is strictly convex and $r > 1$, then $|C_r(P)| = 1$.

A map $T : \mathbb{R}^d \to \mathbb{R}^d$ is called a *similarity transformation* if there exists $c \in (0, \infty)$ such that $\|Tx - Ty\| = c\|x - y\|$. (The real number c is called as the *scaling number*.) The following are known: Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a similarity with scaling factor c.

1. If α is the set of generators of a Voronoi diagram, then the set $T(\alpha) = \{Ta : a \in \alpha\}$ is (locally) finite.
2. $W(Ta|T(\alpha)) = TW(a|\alpha)$.
3. $C_r(T(P)) = TC_r(P)$, and $V_r(TP) = c^r V_r(P)$.

5. Quantization - Set up.

The term quantization has its origin in the theory of signal processing. It is used as a mean to process discretising signals. As a mathematical concept, quantization for probability distributions is the process of approximating a d-dimensional probability distribution P by a discrete probability with a given number of supporting points. In other words, it concerns with the best approximation of a random vector X with distribution P by a random vector Y with range consisting of finitely many points. It follows that this problem concerns determining an appropriate partitioning of the underlying space.

Consequently, two main goals of the theory of quantization are: (1) to find the exact configuration of n-optimal sets that yield to "good" approximation that would allow one to quantize the given probability to within an allowable margin of error, and (2) estimate the rate at which some specified measure of error goes to 0 as $n \to \infty$.

Let X be a $\mathbb{R}^d$-valued random variable with distribution P such that $E\|X\|^r < \infty$. Assume that the set $\alpha = \{a : a \in \mathbb{R}^d\}$ of centers for the quantizing measure of P be given. If $A_\alpha = \{A_a : a \in \alpha\}$ is the Voronoi partition of $\mathbb{R}^d$ w.r.t α, then the quantized version of X is $f(X)$, where $f = \sum_{a\in\alpha} a\chi_{A_a}$. So, the goal is to find n-optimal sets A_α (of centers) such that X and $f(X)$ are within an allowable margin of error in an appropriate metric on the space of probability measures on $\mathbb{R}^d$.

There are many well known such metrics: Levy metric, Prokhorov metric, Ky Fan metric, etc. The most suitable metric for our purposes is **Kantorovich L_r-metric.** For any two Borel probability measures P_1 and P_2 on $\mathbb{R}^d$ with $\int \|x\|^r dP_i < \infty$, the Kantorovich metric is defined as

$$\rho_r(P_1, P_2) = \inf_\mu [\int \|x-y\|^r d\mu(x,y)]^{\frac{1}{r}},$$

where infimum is taken over Borel probability measures μ on $\mathbb{R}^{2d}$ with marginals P_1 and P_2.

Let X be a $\mathbb{R}^d$-valued random variable with distribution P such that $E\|X\|^r < \infty$. Given a positive integer n, let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a measurable (partitioning) function such that $|f(\mathbb{R}^d)| \le n$. We call the collection

$$\mathcal{F}_n = \{f : \mathbb{R}^d \to \mathbb{R}^d : \ f \text{ measurable}, |f(\mathbb{R}^d)| \le n\}$$

as *the set of n-quantizers.* Observe that, for all $f \in \mathcal{F}_n, \ f(X)$ is a quantized version of X with range consisting of at most n points.

Fact.1 If P^f is the image measure of P under f, then

$$\inf_{f\in\mathcal{F}_n} E\|X - f(X)\|^r = \inf_{f\in\mathcal{F}_n} \rho_r^r(P, P^f).$$

Hence, the quantity $\inf_{f\in\mathcal{F}_n} E\|X - f(X)\|^r$ provides the best approximation sought after. Consequently, the *n-th quantization error for P of order r* is defined by

$$V_{n,r}(P) = V_{n,r}(X) = \inf_{f\in\mathcal{F}_n} E\|X - f(X)\|^r.$$

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

$$V_{n,r}(P) = E\|X - f(X)\|^r;$$

i.e., the infimum is attained by $f(X)$. For a fixed n, searching for an n-optimal quantizer for P (of order r) is equivalent to the n-centers problem.

Fact 2. $V_{n,r}(P) = \inf_{\alpha\subset\mathbb{R}^d, \ |\alpha|\le n} E(\min_{a\in\alpha} \|X - a\|^r)$.

A (finite) set $\alpha \subset \mathbb{R}^d, \ |\alpha| \le n$, is an *$n$-optimal set of centers for P of order r* if

$$V_{n,r}(P) = E(\min_{a\in\alpha} \|X - a\|^r).$$

Fact 3. If $f \in \mathcal{F}_n$ is an n-optimal quantizer for P (of order r), then $f(\mathbb{R}^d)$ is an n-optimal set of centers for P (of order r).

There is the following converse statement of Fact 2:

Fact 2′. If $\alpha \subset \mathbb{R}^d$ is n-optimal set of centers for P and $\{A_a : a \in \alpha\}$ is a Voronoi partition of $\mathbb{R}^d$ with respect to α, then $f = \sum_{a\in\alpha} a\chi_{A_a}$ is an n-optimal quantizer for P.

For the rest of this section we assume that $n \geq 2$ (to avoid triviality) and X is a $\mathbb{R}^d$-valued r.v. with distribution P with $|supp(P)| \geq 2$ and $E\|X\|^r < \infty$ for some $1 \leq r < \infty$. P need not be absolutely continuous w.r.t. m.

Let $n \geq 2$ be fixed and let $C_{n,r}(P) = C_{n,r}(X)$ denote the family of all n-optimal set of centers for P of order r. The following statement provides a necessary conditions for optimality.

Theorem 1. Let $\alpha \in C_{n,r}(P)$.

a) Let $r > 1$ or $P(\alpha) = 0$. If the underlying norm is strictly convex, then the Voronoi diagram of α is a P-tessellation of $\mathbb{R}^d$.
b) Let $\{A_a : a \in \alpha\}$ be a Voronoi partition of $\mathbb{R}^d$ w.r.t. α and P. Then $|\alpha| = n$ and $P(A_a) > 0$ for all $a \in \alpha$. In particular, $P(W(a|\alpha)) > 0$.

Notice that P in the previous Theorem need not be absolutely continuous w.r.t. m. The assertion (a) of the Theorem is not valid in general. (see: Example 4.5 on p:40 in [GL]).

Let $S_{n,r}(P)$ be the family of all sets $\alpha \subset \mathbb{R}^d$ with $|\alpha| = n$ satisfying the condition $P(W(a|\alpha)) > 0$. It is clear that $C_{n,r}(P) \subset S_{n,r}(P)$.

Corollary. Let $f \in \mathcal{F}_n$ be an n-optimal quantizer for P of order r and let $\alpha = f(\mathbb{R}^d)$. Then

a) $a \in S_{n,r}(P)$,
b) $\{\{f = a\} : a \in \alpha\}$ is a Voronoi partition of $\mathbb{R}^d$ w.r.t. α and P,
c) $P(\{f = a\}) > 0$, and
d) $a \in C_{n,r}(P(\centerdot|\{f = a\}))$ for every $a \in \alpha$.

The sets $C_{n,r}(X)$ and $S_{n,r}(X)$ have the following same equivariance property.

Fact 4. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a similarity with scaling factor $c > 0$. Then

a) $C_{n,r}(TX) = TC_{n,r}(X)$,
b) $S_{n,r}(TX) = TS_{n,r}(X)$, and
c) $V_{n,r}(TX) = c^r V_{n,r}(X)$.

The n-th quanatization error functional has the following properties:

Fact 5. Let $P = \sum_{i=1}^m s_i P_i$, where $s_i \geq 1$ with $\sum_{i=1}^m s_i = 1$, and $\int \|x\|^r dP_i(x) < \infty$.

a) $V_{n,r}(P) \geq \sum_{i=1}^m s_i V_{n,r}(P_i)$. (Concavity)
b) If $n_i \in \mathbb{N}$ with $\sum_{i=1}^m n_i \leq n$, $\alpha_i \in C_{n,r}(P_i)$, and $\alpha = \cup\alpha_i$, then

$$V_{n,r}(P) \leq \int \min_{a\in\alpha} \|x - a\|^r dP(x) \leq \sum_{i=1}^m s_i V_{n,r}(P_i).$$

The next statement ensures the existence of n-optimal quantizers.

Theorem 2. Let X is a $\mathbb{R}^d$-valued r.v. with distribution P with $E\|X\|^r < \infty$ for some $1 \leq r < \infty$. Then

a) $C_{n,r}(P) \neq \emptyset$ and $\cup\{\alpha : \alpha \in C_{n,r}(P)\}$ is a bounded subset of $\mathbb{R}^d$.
b) $V_{n,r}(P) < V_{n-1,r}(P)$.

Remark. There may be more than one n-optimal set of centers for P, even in the case that P is absolutely continuous w.r.t. m.

Examples. 1. Let P be the uniform distribution on $[0,1]^2 \subset \mathbb{R}^2$. For any $k > 1$, let $A_1, A_2, \ldots, A_n,\ n = 2^k$, be the translations of $[0, \frac{1}{k}]^2$. If a_i is the mid-point of A_1, then $\alpha = \{a_i : 1 \le i \le 2^k\}$ is n-optimal set of centers for P, and

$$E\|X - f_n(X)\|^r = n^{-\frac{r}{2}} \int_{[-\frac{1}{2},\frac{1}{2}]^2} \|x\|^r dx, \quad \text{where} \quad f_n = \sum_{i=1}^{n} a_i \chi_{A_i}.$$

For instance, if $\|\ \|$ is the sup-norm, then $V_{n,r} = n^{-\frac{r}{2}} \frac{2}{(2+r)2^r}$, and α is optimal.

2. Let P be the normal distribution $N(0,1)$ on $(\mathbb{R}^2, d_2)$ and $r = 2$.

(i) When $n = 2$, $\ C_{2,2} = \{\alpha = \{-a, a\} : a \in \mathbb{R}^2, \|a\|_2 = E|X| = \sqrt{\frac{2}{\pi}}\}.$, whereas $V_{2,2}(X) = 2 - \frac{2}{\pi}$. This is optimal.

(ii) When $n = 3$, then one can exhibit more than one set of 3-centers. These are

$$\alpha_1 = \{(-c, 0), (0,0), (c, 0)\}, \text{ where } c \approx 1.224, \text{ with } V_{3,2} \approx 1.190, \text{ and}$$

$$\alpha_2 = \{(0,b), (\frac{\sqrt{3}b}{2}, -\frac{b}{2}), (-\frac{\sqrt{3}b}{2}, -\frac{b}{2}), \quad \text{where } b \approx 1.036, \text{ with } V_{3,2} \approx 1.036.$$

Hence α_2 is optimal.

In one dimensional setting, under a rather relaxed condition, one can ensure uniqueness of optimality. A probability distribution is called strongly unimodal if $P = hm$, where h is a measurable function with $I = \{h > 0\}$ is an open interval and $\log h$ is concave on I.

Theorem 3. (Kieffer, 1983) If P is strongly unimodal, then $|C_{n,r}(P)| = 1$ for every $n \in \mathbb{N},\ 1 \le r < \infty$.

Examples. 1. Uniform distribution $N(0,1)$ on $[0,1]$ is strongly unimodal. For $n = 2^k$, $\alpha = \{\frac{2i-1}{2^k} : i = 1, 2, \ldots, n\}$ is the set of n-optimal centers of order r, for all $r \ge 1$, and $V_{n,r}(P) = \frac{1}{n^r(1+r)2^r}$.

2. Exponential distribution P_e is also strongly unimodal. $P = \lambda h$, where $h(x) = \frac{1}{c} e^{-\frac{x}{c}} \chi_{(0,\infty)}$, with $c = \frac{1}{\log 2}$. Then $\{a_1, a_2, \ldots, a_n\}$, where $a_i = \ln(\frac{\sqrt{n^2+n}}{n+1-i})$, $\ 1 \le i \le n$, is the unique optimal set of n-centers for P of order 1. In this case, $V_{n,1} = \ln(1 + \frac{1}{n})$.

As the following example shows, if P is not strongly unimodal, then the assertion of Theorem.3 may not hold.

Example. Let $P = \lambda h$, be on $\mathbb{R}$, where

$$h(x) = \begin{cases} -\dfrac{|x|}{3} + \dfrac{5}{12} & \text{if } |x| < 1 \\ \dfrac{7 - |x|}{72} & \text{if } 1 \le x < 7, \\ 0 \ \text{ if } & x \ge 7. \end{cases}$$

P is symmetric and unimodal, but not strongly unimodal (since $\log h$ is not concave). For $n = r = 2$, the sets

$$\alpha_1 = \{-1, 3\},\ \alpha_2 = \{-31\}, \text{ and } \alpha_3 = \{-\frac{61}{36}, \frac{61}{36}\}$$

are set of 2-centers for P of order 2. Since

$$E(\min_{a\in\alpha_1}|X-a|^2) = E(\min_{a\in\alpha_2}|X-a|^2) = \frac{47}{18} = 2.611\ldots \text{ with } V_{2,2} = \frac{47}{18};$$
$$E(\min_{a\in\alpha_3}|X-a|^2) = \frac{355}{129} = 2.739\ldots \quad \text{with } V_{2,2} = \frac{53}{18}.$$

Hence α_1 and α_2 are 2-optimal, whereas α_3 is not 2-optimal. Hence, $|C_{2,2}(P)| = 2$.

There is a well known iterative algorithm to construct an n-optimal centers for P of order r, which is originally due to H. Steinhaus (1956) and rediscovered by S.P. Lloyd (1957 but published in 1982).

Having the main results outlined above, we will summarize the (practical) optimality conditions and describe Lloyd algorithm. The necessary conditions for a set α with $|\alpha| = n$ to be optimal for a probability P are: (1) the Voronoi diagram generated by α all have positive P measure, and (2) each $a \in \alpha$ is optimal (i.e., center) for its Voronoi region. The Lloyd algorithm is based on these conditions and proceeds as follows: First choose a set of size n and determine the Voronoi diagram generated by this set, and then replace this set with a new set whose elements are the centers of these Voronoi regions. Continue this process, always using the new set to generate new Voronoi diagrams and then replacing this set with new centers, until the average error has reached to within some predetermined accuracy. If P is strongly unimodal, then the Lloyd algorithm yields to an optimal set. Without unimodality, the Lloyd algorithm yields at best a set which gives the local minimum.

4. Asymptotic Quantization (non-singular case).

In this section we will study the asymptotic behavior of $V_{n,r}(P)$ as $n \to \infty$, mostly in the case that P is nonsingular w.r.t m. Throughout this section $P = P_a + P_s$ is the Lebesgue decomposition, where P_a is the absolutely continuous part and P_s is the singular part of P.

Fact 6. If $E\|X\|^r < \infty$, then $\lim_{n\to\infty} V_{n,r}(P) = 0$.

Remark. Fact.6 reflects that $\cup_{n=1}^{\infty}\mathcal{F}_n$ is a dense subset of all probability distributions on $\mathbb{R}^d$.

Let $A \subset \mathbb{R}^d$ be a bounded Borel set with $m(A) > 0$ and U be the uniform distribution on A. Define

$$M_{n,r}(A) = \frac{V_{n,r}(U(A))}{m(A)^{r/d}},$$

the normalized n-th quantization error for U of order r. Also define

$$Q_r(A) := \inf_{n\geq 1} n^{r/d} M_{n,r}(A), \text{ where } A \subset \mathbb{R}^d \text{ Borel set.}$$

With the *convention* that, for $0 < p < \infty$, $\|f\|_p = [\int |f|^p d\mu]^{1/p}$, we have

Theorem 4. Let $E\|X\|^{r+\delta} < \infty$ for some $\delta > 0$. Then

a) $Q_r([0,1]^d) > 0$, and
b) $\lim_{n\to\infty} n^{r/d} V_{n,r}(P) = Q_r([0,1]^d)\|\frac{dP_a}{dm}\|_{d/d+r}$.

Remarks. 1. If P is a singular distribution, then (b) above only yields $V_{n,r}(P) = o(n^{-\frac{r}{d}})$.
2. The moment condition $E\|X\|^{r+\delta} < \infty$, for some $\delta > 0$, ensures that the limit exists. Without the moment condition we have

$$\liminf_{n\to\infty} n^{r/d} V_{n,r}(P) \geq Q_r([0,1]^d)\|\frac{dP_a}{dm}\|_{d/d+r}.$$

When $P_a \neq 0$ it follows from (a) and (b) of the Theorem that $0 < Q_r([0,1]^d)\|\frac{dP_a}{dm}\|_{d/d+r} < \infty$. The number

$$Q_r(P) = Q_r(X) = Q_r([0,1]^d)\|\frac{dP_a}{dm}\|_{d/d+r}$$

is called the *r-th quantization coefficient* of the probability P on $\mathbb{R}^d$. Notice that $Q_r(P)$ depends on the underlying norm. Also, the scaling property of $V_{n,r}(P)$ carries over to $Q_r(P)$: $Q_r(TX) = c^r Q_r(X)$.

Examples. 1. If P is the uniform distribution $U([a,b])$, then $Q_r(P) = \frac{1}{1+r}(\frac{b-a}{2})^r$.

2. If P is the normal distribution $N(0,1)$, then $Q_r(P) = (\frac{\pi}{2})^{\frac{r}{2}}(1+r)^{\frac{r-1}{2}}$.

3. If P is the exponential distribution$E(\lambda)$, then $Q_r(P) = (\frac{\lambda(1+r)}{2})^r$.

Let $\Gamma = \{\alpha_n\}_{n\geq 1}$ be a sequence where $\alpha_n \subset \mathbb{R}^d$ with $|\alpha_n| \leq n$. The sequence Γ is called *asymptotically n-optimal set of centers for P of order r* if

$$\lim_{n\to\infty} n^{r/d} E(\min_{a\in\alpha_n} \|X-a\|^r) = Q_r(P),$$

provided that $P_a \neq 0$ and $E\|X\|^{r+\delta} < \infty$ for some $\delta > 0$.

Observe that if $\{\alpha_n\}_{n\geq 1}$ is asymptotically n-optimal set of centers for P of order r and $\{A_a : a \in \alpha_n\}$ is a Voronoi partition of $\mathbb{R}^d$ w.r.t. α_n, then $(f_n)_n$ with $f_n = \sum_{a\in\alpha_n} a\chi_{A_a} \in \mathcal{F}_n$ is an *asymptotically n-optimal quantizer of order r*, that is,

$$\lim_{n\to\infty} n^{r/d} E\|X - f_n(X)\|^r) = Q_r(P).$$

Under rather mild conditions, asymptotically n-optimal set of centers for P of order r exists. (see: [GL] Section 7).

For any set $\alpha \subset \mathbb{R}^d$, let $d(x,\alpha) = \inf_{a\in\alpha} \|x-a\|$, the distance of x to α. Now, for any Borel probability measure P on $\mathbb{R}^d$ define

$$e_{n,r}(P) = \inf_{\alpha\subset\mathbb{R}^d,\ |\alpha|\leq n} \|d(x,\alpha)\|_r.$$

From Fact.1, it follows that, for $1 \leq r < \infty$,

$$e_{n,p}(P) = V_{n,r}(P)^{1/r}.$$

When $r = \infty$ (the worst case scenario), the definition becomes

$$e_{n,\infty}(P) = \inf\{\sup_{x\in supp(P)} d(x,\alpha) : \alpha \subset \mathbb{R}^d,\ |\alpha| \leq n\}.$$

Furthermore, if $supp(P)$ is compact, $e_{n,\infty}(P) < \infty$. Now, for any nonempty compact set $A \subset \mathbb{R}^d$, we define $e_{n,r}(A) = e_{n,r}(P)$, where P is a probability with $A = supp(P)$. The number $e_{n,\infty}(A)$ is called an *n-th covering radius for A*, since searching for $\alpha \in C_{n,\infty}$ is equivalent to the problem of finding the most economical covering of A by at most n-balls of equal radius.

Fact.7 Let $n \in \mathbb{N}$.

a) If $1 \leq r \leq s \leq \infty$, then $e_{n,r}(P) \leq e_{n,s}(P)$.
b) If $supp(P)$ is compact, then $\lim_{r\to\infty} e_{n,r}(P) = e_{n,\infty}(P)$.
c) If $A \subset \mathbb{R}^d$ is a nonempty compact set, then $e_{n,\infty}(A) = \inf_{|\alpha|\leq n} d_H(\alpha, A)$, where $d_H(\centerdot,\centerdot)$ is the Hausdorff metric.
d) If $A \subset \mathbb{R}^d$ is a nonempty compact set with $m(A) = 0$, then $e_{n,\infty}(A) = o(n^{1/d})$.
e) If X is a $\mathbb{R}^d$-valued r.v. with distribution P and $supp(P)$ is compact, then

$$e_{n,\infty}(A) = \inf_{f\in\mathcal{F}_n} \|X - f(X)\|_\infty.$$

5. Asymptotic Quantization for Singular Probability Distributions.

In the previous section we have seen that, if the absolutely continuous (w.r.t. m) part of the propability does not vanish, the asymptotics of the minimal error tend to behave "nicely", and certain limits are known to exist. However, such results are either not known or much difficult to obtain in the strictly singular case.

Simplest, and absolutely uninteresting, singular measure is the dirac delta distribution. It is the Lebesgue-Stieltjes measure induced by the Heaviside step function $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geq 0$. This measure has support consisting of only the point 0 (hence, such measures are called *pure point* measures).

Interesting singular measures are those with larger supports. One such measure is the constructed using Cantor-Lebesgue function (also known as the Devil's Staircase). Let $D : [0,1] \to [0,1]$ be the Cantor-Lebesgue function, which is a non-decreasing continuous function with $D(0) = 0,\ D(1) = 1$, and has derivative 0 almost everywhere on $[0,1]$ w.r.t. Lebesgue measure (since Lebesgue measure of the Cantor set is 0). Let μ_D be the measure on given by the Lebesgue-Stieltjes measure induced by the function D. Then the Radon-Nikodym derivative $\frac{d\mu_D}{dm} = 0$, yet μ_D is certainly not 0 since $\mu_D([0;1]) = 1$. At the same time, since D is continuous, μ_D contains no atoms. Therefore, μ_D is a purely singular continuous measure.

The previous example is not really that different from the atomic case except the support of the singular part of the measure is an uncountable fractal set rather than a countable collection of points. To see strange Singular measures can be very "strange"; for instance, consider the measure μ_F given by F. Riesz. It is constructed similarly to μ_D, where the function D is replaced by an increasing function $F : [0,1] \to [0,1]$ with $F(x) \neq 0$ for all $x \in [0,1]$ such that $\frac{dF}{dm} = 0$ m-a.e. Hence μ_F has the Radon-Nikodym derivative $\frac{d\mu_F}{dm} = 0$, yet μ_F has no atoms. Hence μ_F is purely singular continuous measure with support $[0,1]$. (See the article *Singular Continuous Measures* by Michael Pejic, and/or the article *On fine structure of singularly continuous probability measures and random variables with independent Q-symbols* by S. Albeverio, et al.)

Another class of singular measures is obtained via self-similar sets (fractals). If $\{A; f_1, \dots, f_n\}$ is an IFS with attractor A, where each f_i is a contracting similarity on $\mathbb{R}^d$, and $(p_1, \dots, p_n)$ is a probability vector, then there is a unique probability measure P on $\mathbb{R}^d$ with

$$P = \sum_{i=1}^{n} p_i (P \circ f_i^{-1}). \quad \text{[Hutchinson, 1981]}$$

Then this measure P is singular w.r.t. the Lebesgue measure m on $\mathbb{R}^d$. Hence, with this method, one can construct many non-trivial singular measures.

In this section we will study the asymptotic behavior of $e_{n,r}(P)$ as $n \to \infty$ for continuous singular distributions P w.r.t m, where

$$e_{n,p}(P) = V_{n,r}(P)^{1/r},\ 1 \leq r < \infty,\ \text{and}$$
$$e_{n,\infty}(P) = \inf\{ \sup_{x \in supp(P)} d(x,\alpha) : \alpha \subset \mathbb{R}^d,\ |\alpha| \leq n \}.$$

Such singular measures may have have discrete support or continuous; of course, the interesting cases are those with non-discrete support. Since, for singular measures, we only know that $e_{n,p}(P) = o(n^{1/d})$, one needs a more refined tool to study the behaviour of $e_{n,p}$. It turns out that there is a very close relationship between dimension theory and quantization. This, besides leading to the notion of "quantization dimension", also provides a convenient way to investigate asymptotic behavior of $e_{n,r}(P)$ for any probability measure P.

Given a Borel probability measure P on $\mathbb{R}^d$, and $r \in [1, \infty]$, the *lower quantization dimension of P of order r* is defined as

$$\underline{D_r} := \underline{D_r}(P) = \liminf_{n\to\infty} \frac{\log n}{-\log e_{n,r}}, \text{ and}$$

the *upper quantization dimension of P of order r* is defined as

$$\overline{D_r} := \underline{D_r}(P) = \limsup_{n\to\infty} \frac{\log n}{-\log e_{n,r}}.$$

If $\underline{D_r} = \overline{D_r} = D_r = D_r(P)$ (i.e., when the limit exists), then $D_r = \lim_{n\to\infty} \frac{\log n}{-\log e_{n,r}}$ is called the *quantization dimension of P of order r*. Observe that D_r does not depend on the underlying metric (norm), but depends only on the support of P. Also, if D_r exists, we have

$$\log e_{n,r} \sim \log(\frac{1}{n})^{1/D_r}.$$

This can be interpreted that quantization dimension measures the asymptotic rate at which $e_{n,r}$ goes to zero.

Before proceeding further, let's have a short sojourn to the realm of Dimension Theory.

Covering Dimension: A space has Covering (Lebesgue) dimension m if for every open cover of that space has a refinement such that every point is contained in at most m+1 elements (of the refinement).

Box Dimension: $dim_B(A) := \lim_{\epsilon\to 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)}$, where $N(\epsilon)$ is the number of boxes of side-length ϵ required to cover A.

Packing Dimension: The value at which $\mathcal{P}_d(A) = \lim_{\delta\to 0} \inf_{\{U_i\}} \sum_i (diam U_i)^d$ transitions from ∞ to 0, where $\{U_i\}$ is a countable pairwise disjoint closed balls with centers in A and radius δ. **Notation:** $\dim_P(A)$.

Hausdorff Dimension: The value at which $\mathcal{H}_d(A) = \lim_{\epsilon\to 0} \inf_{\{U_i\}} \sum_i (diam U_i)^d$ transitions from ∞ to 0. **Notation:** $\dim_H(A)$.

We also have the concept of dimension of a (probability) measure μ which is defined as

$$\dim_{\cdot}(\mu) := \lim_{\delta\to 0} \inf\{\dim_{\cdot}(A) : \mu(A) = 1-\delta\}.$$

It turns out that these various dimensions are interrelated in some fashion:

(i) In general, we have

$$\dim_H(A) \le \dim_P(A) \le \overline{\dim_B}(A), \text{ and } \dim_H(A) \le \underline{\dim_B}(A) \le \overline{\dim_B}(A).$$

(ii) For self similar sets satisfying Open Set Condition, all these are equal (and is also called similarity dimension).

(iii) It is possible that $\dim_{\cdot}(\mu) < dim_{\cdot}(K)$, where K is the support of μ.

Fact 8. Let P be a Borel probability measure and $r \in [1, \infty]$.

a) If $0 \le t < \overline{D_r} < s$, then $\limsup_{n\to\infty} n e_{n,r}^t = \infty$ and $\lim_{n\to\infty} n e_{n,r}^s = 0$.

b) If $0 \le t < \underline{D_r} < s$, then $\lim_{n\to\infty} n e_{n,r}^t = \infty$ and $\liminf_{n\to\infty} n e_{n,r}^s = 0$.

Corollary. Let P be a Borel probability measure and $r \in [1, \infty]$.

a) If $1 \le r \le s \le \infty$, then $\underline{D_r} \le \underline{D_s}$ and $\overline{D_r} \le \overline{D_s}$.

b) If $D \in (0, \infty)$ is such that $0 < \liminf_{n\to\infty} n e_{n,r}^D \le \limsup_{n\to\infty} n e_{n,r}^D < \infty$, then $D_r = D$.

c) If $r \in [1, \infty)$ and $E(\|X\|^{r+\delta}) < \infty$ for some $\delta > 0$, then $\overline{D_r} \le d$. If $P_a \ne 0$, then $\overline{D_r} = d$.

d) $\overline{D_\infty} \leq d$. If $m(supp(P) > 0$, then $\overline{D_\infty} = d$.

There is an interesting relationships of the quantization dimension to other types of (fractal) dimensions can be summarized as follows.

Fact 9. Let $K \subset \mathbb{R}^d$ be compact and P be a Borel probability measure with $supp(P) = K$. Then

$$\dim_H(K) \leq \underline{D_\infty}(P), \text{ and, } \forall r \geq 1, \dim_H(P) \leq \underline{D_r}(P).$$

Theorem 5. Let $K \subset \mathbb{R}^d$ be compact. Then

$$\overline{\dim}_B(K) = \overline{D_\infty}(K), \text{ and } \underline{\dim}_B(K) = \underline{D_\infty}(K).$$

Corollary. If $\dim_B(K)$ exists for $K \subset \mathbb{R}^d$ compact, then $D_\infty(K)$ also exists and equals the box dimension.

Theorem 6. (Pötzelberger) Let P be a Borel probability measure with compact support $K \subset \mathbb{R}^d$. Then, for $1 \leq r \leq s \leq \infty$,

a) $\overline{D_r}(P) \leq \overline{D_s}(P) \leq \overline{D_\infty}(P) = \overline{D_\infty}(K) = \overline{\dim}_B(K)$, and
b) $\underline{D_r}(P) \leq \underline{D_s}(P) \leq \underline{D_\infty}(P) = \underline{D_\infty}(K) = \underline{\dim}_B(K)$.

6. Self-similar Sets and Measures.

The problem of determining quantization dimension function for a general measure is open; however, for some classes of measures some positive results have been obtained. One such case is the class of self-similar sets and self-similar measures.

Let $S_i : \mathbb{R}^d \to \mathbb{R}^d,\ 1 \leq i \leq N$, be a collection of contractive similarity transformations with corresponding scaling factors $0 < s_i < 1$. Then there exists a unique nonempty compact subset $A \subset \mathbb{R}^d$ with

$$A = \cup_{i=1}^N S_i(A).$$

The set A is called the *attractor* of the collection $\mathbf{S} = \{S_i\}$; it is also called as the *self-similar set associated to* $\{S_i\}$. Hutchinson (1981) showed that there exists a unique real number $D \geq 0$ such that $\sum_{i=1}^N s_i^D = 1$. D is called the *similarity dimension* of $\mathbf{S}$ (or, of A) and is equal to Hausdorff dimension of A. Hutchinson also showed that for a probability vector $(p_1, p_2, \ldots, p_N)$, there is a unique probability measure P on $\mathbb{R}^d$ such that

$$P = \sum_{i=1}^N p_i (P \circ S_i^{-1}).$$

If $p_i > 0$ for all $1 \leq i \leq N$, then $supp(P) = A$. The measure P is called the *self-similar measure* associated to $(\mathbf{S}, \{p_i\}_{i=1}^N)$.

If $\mathbf{S}$ satisfies the open set condition P is the self-similar measure corresponding to $(\mathbf{S}; \{s_i^D\}_{i=1}^N)$, then P is the normalized D-dimensional Hausdorff measure restricted to the attractor A; that is

- $0 < \mathcal{H}^D(A) < \infty$, and
- $P = \frac{1}{\mathcal{H}^D(A)}\ \mathcal{H}^D_{|A}$.

The following are some results known about quantization of self-similar sets A and measures P associated to $(\mathbf{S}, \{p_i\}_{i=1}^N)$.

Fact 10. Ler $r \in [1, \infty)$ be fixed.

a) There exists unique real number $\kappa \in (0, \infty)$ such that $\sum_{i=1}^N (p_i s_i^r)^{\frac{\kappa}{\kappa+r}} = 1$.
b) If κ satisfy (a), then $\limsup_{n\to\infty} n e_{n,r}^\kappa < \infty$; in particular, $\overline{D_r}(P) \leq \kappa$.

c) If D is the similarity dimension of $\mathbf{S}$, then $\limsup_{n\to\infty} ne_{n,\infty}^D < \infty$; in particular, $\overline{D_\infty}(P) \leq D$.
d) If $\mathbf{S}$ satisfy strong separation condition and κ satisfy (a), then $\liminf_{n\to\infty} ne_{n,r}^\kappa > 0$; in particular, $\underline{D_r}(P) \geq \kappa$.
e) If $\mathbf{S}$ satisfy strong separation condition and D is the similarity dimension of $\mathbf{S}$, then $\liminf_{n\to\infty} ne_{n,\infty}^D > 0$; in particular, $\underline{D_\infty}(P) \geq D$.

Theorem 7. Let $r \in [1,\infty)$, κ satisfy (a) in the fact above, and $\mathbf{S}$ satisfy strong separation condition. Then $0 < \liminf_{n\to\infty} ne_{n,r}^\kappa \leq \limsup_{n\to\infty} ne_{n,r}^\kappa < \infty$. In particular, $D_r(P)$ exists and equals κ.

Theorem 8. Let $r \in [1,\infty]$, κ satisfy (a) in the fact above, and $\mathbf{S}$ satisfy strong separation condition.

a) If $p_i = s_i^D,\ 1 \leq i \leq N$, then $D_r(P) = D$.
b) If $(p_1,\dots,p_N) \neq (s_1^D,\dots,s_N^D)$, then $D_r(P)$ exists and

$$D_r(P) = \begin{cases} D, & r = \infty, \\ \kappa, & r < \infty. \end{cases}$$

Moreover, $q < r \Rightarrow D_q(P) < D_r(P)$, and $\lim_{r\to\infty} D_r(P) = D$.

As observed above, for many self-similar probabilities, the inequality

$$0 < \liminf_{n\to\infty} ne_{n,r}^{D_r} \leq \limsup_{n\to\infty} ne_{n,r}^{D_r} < \infty$$

holds for $r \in [1,\infty]$. One naturally asks existence of conditions under which the limit exists. In general, existence of the *quantization coefficient*, the $\frac{r}{D_r}$-th of this limit, is the question. It is known that only under some strict conditions this limit exists (see Theorem 14.12 of [GL]), without which the limit fails to exist.

Cantor Distribution.

Above, we have seen that for self-similar distributions one can obtain more features of the associated quantization. In the very special case of self-similar set/measure of Cantor distribution, there is more. Namely, besides the properties known for general self-similar distributions, one can also describe all optimal sets of n-centers, quantization errors $V_{n,r}(P)$, and all limit points of the sequence $(n^{2/D}V_{n,r}(P))_n$ for $r = 2$. Also, one can show that the quantization coefficient $Q_2(P)$ does not exist.

Recall the Cantor set: Let $S_1, S_2 : \mathbb{R} \to \mathbb{R}$, where $S_1(x) = \frac{1}{3}x$, and $S_1(x) = \frac{1}{3}x = \frac{2}{3}$. The attractor of $\mathbf{S} = \{S_1,\ S_2\}$ is the standard Cantor set $\mathcal{C} \subset [0,1]$. It is known that the similarity dimension of $\mathcal{C}$ is $D = \frac{\log 2}{\log 3}$. Consider the probability vector $(\frac{1}{2},\frac{1}{2})$, and let P be the self similar probability associated to $(\mathbf{S};\frac{1}{2},\frac{1}{2})$. Since $\mathbf{S}$ satisfies strong separation condition, P is the D-dimensional Hausdorff measure on $\mathcal{C}$; hence, it is called as the *Cantor distribution.* Observe that $(s_1^D, s_2^D) = (\frac{1}{2},\frac{1}{2})$, by Theorem 8 above, D is the quantization dimension of P of order r for all $r \in [1,\infty]$.

For any $\sigma \in \{0,1\}^*$ (the set of all words in $\{1,2\}$) let $a_\sigma = S_\sigma(\frac{1}{2})$. Let $l(n) = [\log_2 n],\ n \geq 1$. If $I \subset \{1,2\}^{l(n)}$ with $|I| = n - 2^{l(n)}$, let

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

Also, let $f : [1,2] \to \mathbb{R}$ be defined by $f(x) = \frac{1}{72}x^{2/D}(17-8x)$. Observe that $V_2(P) = \frac{1}{8}$.

Theorem 9. [GL2] Let P be the Cantor distribution and let D, $l(n)$, $\alpha_n(I)$, and f be as defined above.

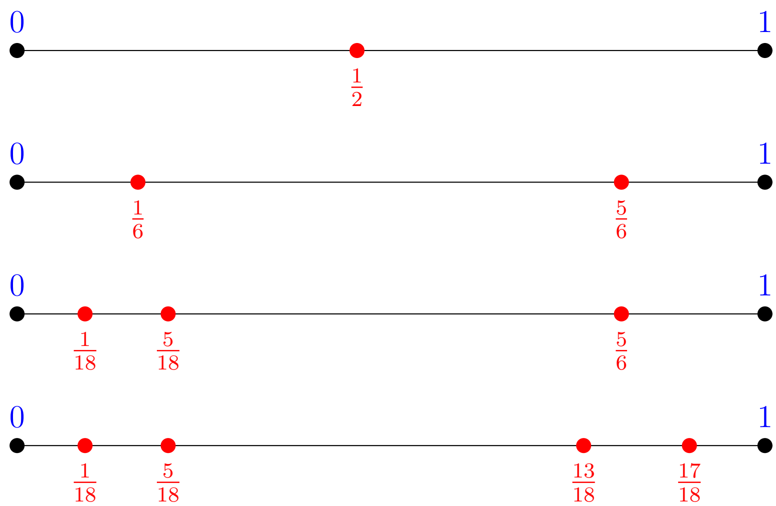
a) For every $n \geq 1$ TFAE:
(i) α is an n-optimal set of centers of order 2 for P.
(ii) $\exists\ I \subset \{1,2\}^{l(n)}$ with $\alpha = \alpha_n(I)$.
b) For every $n \geq 1$,

$$V_{n,2}(P) = \frac{1}{18^{l(n)}}\ \frac{1}{8}\ [2^{l(n)+1} - n + \frac{1}{9}(n - 2^{l(n)})].$$

c) The set of accumulation points of $(n^{2/D} V_{n,2}(P))_n$ is the interval

$$[\frac{1}{8}, f(\frac{17}{8+4D})] = [0.125, 0.2589\dots].$$

The picture of n-optimal sets, $n = 1, 2, 3$ and 4, for the cantor distribution are below.



Some known results (quantization dimension)

Since the early 1980's, beginning with Zador, many mathematicians, notable Graf & Luscghy, have obtained numerous properties of quantization. The field is relatively new, and there are numerous questions we have no answers for.

Recall that if μ has non-vanishing absolutely continuous part, then $D_r(\mu)$ exists and $D_r = d$. If μ is singular, then $D_r(\mu)$ may or may not exist. We don't know the complete picture yet; however there are some results in some cases.

If μ has bounded support K, then

$$\underline{D}_2 \in [\dim_H(\mu), \underline{\dim}_B(\mu)] \quad \text{and} \quad \overline{D}_2 \in [\dim_P(\mu), \overline{\dim}_B(\mu)].$$

This result appears in the papers of Pötzelberger, M. Dai and Z. Liu, among others. Furthermore,

If $supp(\mu) = K$ is compact,

$$\dim_H(\mu) \leq \overline{D}_r(\mu) \leq \dim_B(K), \quad \text{and} \quad \underline{D}_r(\mu) \leq \dim_B(K).$$

Theorem [Graf-Luschgy, 2000] Let μ be a self-similar measure generated by an IFS satisfying strong separation condition and κ satisfy the equation $\sum_{i=1}^N (p_i s_i^r)^{\frac{\kappa}{\kappa+r}} = 1$. Then $0 < \liminf_{n\to\infty} n V_{n,r}^{\kappa/r} \leq \limsup_{n\to\infty} n V_{n,r}^{\kappa/r} < \infty$. In particular, $D_r(P)$ exists and equals κ (similarity dimension).

Some known results (quantization coefficient)

In general, the existence of the quantization coefficient for singular measures is not known. However, under some *very stringent* conditions the quantization coefficient exists (Graf-Luschgy, 2000).

- **Lindsay and Mauldin** Investigated D_r for a measure μ associated with a conformal finite IFS with OSC. In particular, they proved that the upper quantization coefficient for μ associated such and IFS is finite.
- In a series of articles **Roychowdhury** proved the existence of D_r for probability measures μ with bounded distortion (2009), Moran measures (2010, 2011), recurrent self-similar measures (2011), self-similar measures generated by infinite IFS (2011), for Gibbs-like measures (2013); he also obtained estimates of D_r for condensation systems (2012).
- **Roychowdhury (2011) and Zhu (2008)** independently proved that the lower quantization coefficient for μ associated with a conformal finite IFS with OSC is positive. This was an open question for sometime.
- **Roychowdhury (2014)** also studied quantization coefficient for ergodic measures on symbolic spaces.
- Also, **Mihailescu** & **Roychowdhury (2015)** worked on quantization coefficients for infinite IFS.

Some known results (optimal sets)

Study of optimal sets is relatively recent. This is mostly due to the fact that it is a difficult problem to deal with. The optimal sets of n-means for many probability measures is not known. The problem is studied only in the case of some absolutely continuous measures and some special cases of singular measures.

- **Graf** & **Luschgy (1997)** Obtained complete characterization of the optimal sets for the (classical) Cantor distribution. A major accomplishment as well as an inspiration and guide for many others.
- **Dettmann** & **Roychowdhury (2017):** For uniform distributions on equilateral triangles.
- **Roychowdhury:** For distributions generated by dyadic Cantor sets, infinite similitudes on $\mathbb{R}$, etc.
- **Çömez** & **Roychowdhury (2016, 2017):** For probability distributions on Sierpinski carpets, for probability distributions generated by infinite affine transformations, and for probability distributions on R-triangles.

For example, Sierpinski Carpet $\mathcal{S}$ given by, as an IFS, the similarity mapping $S_1(x_1,x_2) = \frac{1}{3}(x_1,x_2)$, $S_2(x_1,x_2) = \frac{1}{3}(x_1,x_2) + (\frac{2}{3},0)$, $S_3(x_1,x_2) = \frac{1}{3}(x_1,x_2) + (0,\frac{2}{3})$, and $S_4(x_1,x_2) = \frac{1}{3}(x_1,x_2) + (\frac{2}{3},\frac{2}{3})$, $(x_1,x_2) \in \mathbb{R}^2$, satisfies SSC. For the associated self-similar Borel probability measure $P_{\mathcal{S}}$ on $\mathbb{R}^2$ the optimal sets of n-means, as in the cantor distribution, (roughly) followthe action of the maps S_ω, $\omega \in \{1,2,3,4\}^k$.

Theorem. [Ç & Roychowhury, 2016] Let $n \geq 4$ and $1 \leq m \leq 3$. If $n = m4^{\ell(n)} + k$, where k is a positive integer such that $0 \leq k < 4^{\ell(n)}$ for some positive integer $\ell(n)$, and $t \subset I^{\ell(n)}$ with card $(t) = k$, then,

$$\alpha_n(t) = \{S_\sigma(\alpha_m) : \sigma \in I^{\ell(n)} \setminus t\} \cup \{S_\sigma(\alpha_{m+1}) : \sigma \in t\}$$

is an optimal set of n-means. The number of such sets is $(2^{m-1})^{4^{\ell(n)}-k} \cdot {}^{4^{\ell(n)}}C_k \cdot 2^{mk}$ if $m = 1,2$, and $(2^{m-1})^{4^{\ell(n)}-k} \cdot {}^{4^{\ell(n)}}C_k$ if $m = 3$. The corresponding quantization error is given by

$$V_n = \sum_{\sigma \in I^{\ell(n)} \setminus t} \int_{J_\sigma} \min_{a \in S_\sigma(\alpha_m)} \|x-a\|^2 dP + \sum_{\sigma \in t} \int_{J_\sigma} \min_{a \in S_\sigma(\alpha_{m+1})} \|x-a\|^2 dP,$$

where ${}^uC_v = \binom{u}{v}$, the binomial coefficients.

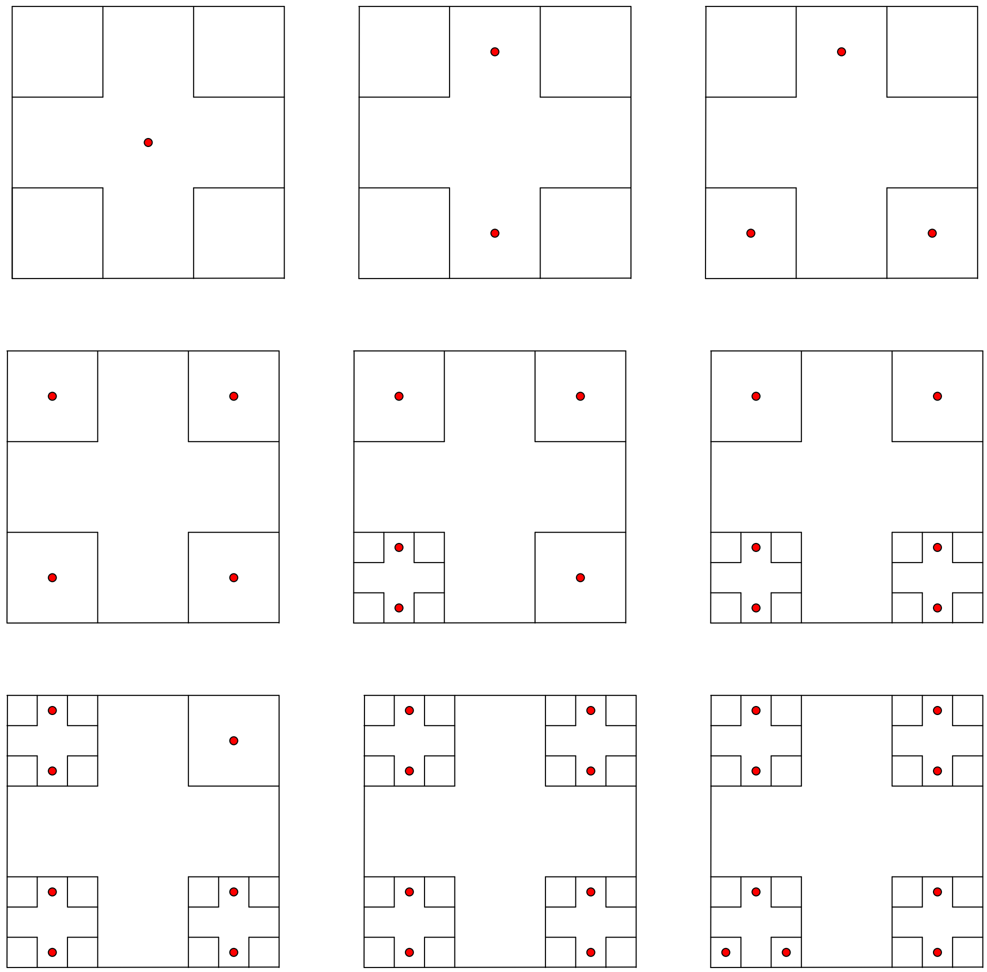


FIGURE 2. Optimal configuration of n points, $1 \leq n \leq 9$ for Sierpinski Carpet with SSC

Beyond the cases mentioned above, the problem of determining n-optimal sets is completely open for self-similar measures with overlap.

7. Condensation Systems

Condensation measures and systems

Let $F = \{S_j\}_{j=1}^N$ be an IFS, $p = (p_0; p_1, p_2, \cdots, p_N)$ a probability vector and ν be a Borel probability measure on $\mathbb{R}^d$ with compact support E. Then, there exists a unique Borel probability measure μ on $\mathbb{R}^d$ with (compact) support K_E such that

$$\mu = p_0\nu + \sum_{j=1}^N p_j \mu \circ S_j^{-1}, \text{ where } K_E = \cup_{j=1}^N S_j(K_E) \cup E.$$

The measure μ is the *condensation measure* (or *inhomogeneous self-similar measure*) of the *condensation system* (F, p, ν).

Consider the condensation system $(\{S_j\}_{j=1}^2, (p_j)_{j=0}^2, \nu)$, where:

$$S_1(x) = \frac{1}{5}x, \ S_2(x) = \frac{1}{5}x + \frac{4}{5}; T_1(x) = \frac{1}{3}x + \frac{4}{15}, \ T_2(x) = \frac{1}{3}x + \frac{2}{5}$$
$$\nu = \frac{1}{2}\nu \circ T_1^{-1} + \frac{1}{2}\nu \circ T_2^{-1}, \text{ and } p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}).$$

Then condensation measure P associated with this system and its support K satisfy

$$(*)\quad P=\frac{1}{3}(P\circ S_1^{-1})+\frac{1}{3}(P\circ S_1^{-1})+\frac{1}{3}\nu,\ \text{and}\ K=S_1(K)\cup S_2(K)\cup C,$$

where C (the dyadic Cantor set) is the support of ν.

Properties of the Condensation measure and the optimal set of 1-mean

1. Iterating (*), for $\omega=\omega_1\omega_2\cdots\omega_n\in\{1,2\}^n$,

$$P=\frac{1}{3^n}\sum_{|\omega|=n}P\circ S_\omega^{-1}+\sum_{k=0}^{n-1}\frac{1}{3^{k+1}}(\sum_{|\omega|=k}\nu\circ S_\omega^{-1}),\ \text{and}\ \nu=\frac{1}{2^k}\sum_{\omega\in I^k}\nu\circ T_\omega^{-1}.$$

2. P **is 'symmetric' about** $\frac{1}{2}$: if two intervals of equal lengths are equidistant from the point $\frac{1}{2}$ then they have the same P-measure.
3. Expected value of $\nu:=E(\nu)=\frac{1}{2}$ and the variance of $Var(\nu):=W=\frac{1}{200}$.
4. Expected value of $P:=E(P)=\frac{1}{2}$ and the variance of $Var(P):=V=\frac{65}{584}$.

$$\forall\ x_0\in\mathbb{R},\ \int(x-x_0)^2dP=(x_0-\frac{1}{2})^2+V.$$

Thus, the optimal set of one-mean for P consists of the expected value $E(P)=\frac{1}{2}$; the quantization error is $V=\frac{65}{584}$.

Optimal sets for $n\geq 2$

The optimal set α_n of n-means and associated quantization errors V_n for $n=2,3,4$:

$\alpha_2=\{a_1,a_2\}$, where $a_1=\frac{19}{90}$, $a_2=\frac{71}{90}$; $V_2=\frac{32929}{1182600}$.
$\alpha_3=\{a_1,a_2,a_3\}$, where $a_1=S_1(\frac{1}{2})=\frac{1}{10}$, $a_2=\frac{1}{2}$, $a_3=S_2(\frac{1}{2})=\frac{9}{10}$; $V_3=\frac{203}{43800}$.
$\alpha_4=\{S_1(\frac{1}{2}),T_1(\frac{1}{2}),T_2(\frac{1}{2}),S_2(\frac{1}{2})\}$; $V_4=\frac{1243}{394200}$.

Description of optimal set of n-means for $n\geq 5$ requires some technical tools.
Define the sequences $\{a(n)\}_{n\geq 1}$ and $\{F(n)\}_{n\geq 1}$ by

$$a(n)=\frac{1}{4}\left(6n+(-1)^{n+1}-7\right),\ \text{and}$$
$$F(n)=\begin{cases}2^n(n+1) & \text{if } 1\leq n\leq 4,\\ 5\cdot 2^n+2^{n-\frac{7}{4}}\sum_{k=5}^{n}2^{\frac{k}{2}+\frac{(-1)^{k+1}}{4}} & \text{if } n\geq 5,\end{cases}$$

Observe that

$$a(n): 0,1,3,4,6,7,9,10,12,13,15,16,18,19,21,22,24,25,27,28,30,\dots,$$

$$F(n): 4,12,32,80,224,576,1664,4352,12800,33792,100352,266240,\dots.$$

Now, define

$S(a(n))=\alpha_{2^{a(n)}}$, $\quad S(a(n-1))=\cup_{\omega\in I}S_\omega(\alpha_{2^{a(n-1)}})$,
$S(a(n-2))=\cup_{\omega\in I^2}S_\omega(\alpha_{2^{a(n-2)}})$,
$\dots\dots\dots\dots,$
$S(a(4))=\cup_{\omega\in I^{n-4}}S_\omega(\alpha_{2^{a(4)}})$, $\quad S(a(3))=\cup_{\omega\in I^{n-3}}S_\omega(\alpha_{2^{a(3)}})$,
$S(2)=\cup_{\omega\in I^{n-2}}S_\omega(\alpha_{2^2})=S(2)$, $\quad S(1)=\cup_{\omega\in I^{n-1}}S_\omega(\alpha_2)=S(1)$, and
$S(0)=\{S_\omega(\frac{1}{2}):\omega\in I^n\}$, where $I=\{1,2\}$.

Set

$$\alpha_{F(n)} := \begin{cases} S(1) \cup S(0) & \text{if } n = 1, \\ S(2) \cup S(1) \cup S(0) & \text{if } n = 2, \\ \Big(\bigcup_{\ell=3}^{n} S(a(\ell)) \Big) \cup S(2) \cup S(1) \cup S(0) & \text{if } n \geq 3. \end{cases}$$

Theorem (Çömez & Roychowdhury, 2017) For any $n \geq 1$, the set $\alpha_{F(n)}(P)$ is an optimal set of $F(n)$-means with quantization error given by

$$V_{F(n)} = \begin{cases} \frac{1}{3.9} W + \frac{2}{75} V & \text{if } n = 1, \\ \frac{1}{3.9^2} W + \frac{2}{75} \frac{1}{3.9} W + \left(\frac{2}{75}\right)^2 V & \text{if } n = 2, \\ \sum_{\ell=3}^{n} \left(\frac{2}{75}\right)^{n-\ell} \frac{W}{3.9^{a(\ell)}} + \left(\frac{2}{75}\right)^{n-2} \frac{W}{3.9^2} + \left(\frac{2}{75}\right)^{n-1} \frac{W}{3.9} + \left(\frac{2}{75}\right)^{n} V, & n \geq 3. \end{cases}$$

($V = \frac{65}{584}$, $W = \frac{1}{200}$, variance of P and ν, respectively.)

Quantization dimension and coefficients

Theorem (ÇR, 2017) Let P be the condensation measure associated with the self-similar measure ν. Then,

- $\lim_{n \to \infty} \frac{2 \log n}{\log V_n(P)} = \beta$, i.e., the quantization dimension $D(P)$ of the measure P exists and equals $\beta = \frac{\log 2}{\log 3}$.
- β-dimensional quantization coefficient for the condensation measure P does not exist, and the β-dimensional lower and upper quantization coefficients for P are finite and positive.

Proposition (ÇR, 2017) Let $D(P)$ be the quantization dimension of the condensation measure P. Then, $D(P) = \max\{k, D(\nu)\}$, where k is the unique number such that

$$\left(\frac{1}{3}\left(\frac{1}{5}\right)^2\right)^{\frac{k}{2+k}} + \left(\frac{1}{3}\left(\frac{1}{5}\right)^2\right)^{\frac{k}{2+k}} = 1.$$

References

AD R. B. Ash and C. A. Doléans-Dade, **Probability and Measure Theory**, Harcourt/Academic Press, San Diego, 2000.

GL S. Graf and H. Luschgy, **Foundations of Quantization for Probability Distributions**, Lecture Notes in Mathematics 1730, Springer, Berlin, 2000.

GL2 S. Graf and H. Luschgy, *The Quantization of the Cantor Distribution*, Math. Nachr., 183 (1997), 113-133.

H J. Hutchinson, *Fractals and self-similarity*, Indiana Univ. J., 30 (1981), 713-747.