

University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

---

Mathematical and Statistical Sciences Faculty  
Publications and Presentations

College of Sciences

---

2020

## Optimal Quantization Via Dynamics

Joseph Rosenblatt

Mrinal Kanti Roychowdhury

*The University of Texas Rio Grande Valley*

Follow this and additional works at: [https://scholarworks.utrgv.edu/mss\\_fac](https://scholarworks.utrgv.edu/mss_fac)



Part of the [Mathematics Commons](#)

---

### Recommended Citation

Rosenblatt, Joseph and Roychowdhury, Mrinal Kanti, "Optimal Quantization Via Dynamics" (2020).

*Mathematical and Statistical Sciences Faculty Publications and Presentations*. 22.

[https://scholarworks.utrgv.edu/mss\\_fac/22](https://scholarworks.utrgv.edu/mss_fac/22)

This Article is brought to you for free and open access by the College of Sciences at ScholarWorks @ UTRGV. It has been accepted for inclusion in Mathematical and Statistical Sciences Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact [justin.white@utrgv.edu](mailto:justin.white@utrgv.edu), [william.flores01@utrgv.edu](mailto:william.flores01@utrgv.edu).

To appear, Dynamical Systems

## OPTIMAL QUANTIZATION VIA DYNAMICS

JOSEPH ROSENBLATT AND MRINAL KANTI ROYCHOWDHURY

**ABSTRACT.** Quantization for probability distributions refers broadly to estimating a given probability measure by a discrete probability measure supported by a finite number of points. We consider general geometric approaches to quantization using stationary processes arising in dynamical systems, followed by a discussion of the special cases of stationary processes: random processes and Diophantine processes. We are interested in how close stationary process can be to giving optimal  $n$ -means and  $n^{\text{th}}$  optimal mean distortion errors. We also consider different ways of measuring the degree of approximation by quantization, and their advantages and disadvantages in these different contexts.

### 1. INTRODUCTION

**1.1. The General Setting.** Let  $\mathbb{R}^d$  denote the  $d$ -dimensional Euclidean space with the Euclidean metric  $\|\cdot\|$ . Let  $P$  be a Borel probability measure on  $\mathbb{R}^d$ . Let  $V_n = V_n(P)$  be

$$V_n(P) = \inf_{\alpha \in \mathcal{D}_n} \int \min_{a \in \alpha} \|x - a\| dP(x),$$

where  $\mathcal{D}_n := \{\alpha \subset \mathbb{R}^d : 1 \leq \text{card}(\alpha) \leq n\}$ . We assume that  $\int \|x\| dP < \infty$  to make sure that there is a set  $\alpha_n^*$  for which the infimum occurs (see [AW, GKL, GL, GL2]). Here  $V_n(P)$  is the *optimal mean distortion error*. The set  $\alpha_n^*$  for which we get the optimal mean distortion error is called an *optimal set of  $n$ -means* and the elements of an optimal set are called *optimal quantizers*. For some mathematical background on quantization, see the text by Graf and Luschgy [GL]. For some recent work on optimal quantizers, see [CR1, CR2, DR, GL3, R1, R2, R3, R4, RR1]. Note that it is traditional to use  $\|x - a\|^2$  in the definition above, but we do not do this here because we want to keep as simple as possible the connection between the integral itself and various ways of approximating the integrand.

Given a fixed  $\alpha_n \in \mathcal{D}_n$ , we call  $d_n^{\alpha_n}(x) = \min_{a \in \alpha_n} \|x - a\|$  the *point distortion error*, and we call the mean of the point distortion error,  $V_n^{\alpha_n} = \int d_n^{\alpha_n}(x) dP(x) = \int \min_{a \in \alpha_n} \|x - a\| dP(x)$ , the *mean distortion error*. We consider various methods of generating *model*  $\alpha_n$ . We are particularly interested in estimates for the rate of convergence to zero, as  $n$  goes to infinity, of the distortion errors of model sequences that have been generated by some dynamical stochastic process. If we know both a) the rate of convergence to zero of the distortion errors in the optimal case, and b) in a given model we have good estimates for the rate of convergence to zero of the distortion errors, then we will have some understanding of how far from optimality these models are.

Clearly, if we had a uniform estimate for the rate of convergence to zero of the point distortion error, then by Lebesgue's Bounded Convergence Theorem we would obtain a rate estimate for the mean distortion error. Indeed, given  $(\rho_n)$ , even if we just had, for a.e.  $x$

---

*Date:* January 24, 2020.

*2010 Mathematics Subject Classification.* 94A34, 37A45, 11J71, 60B05.

*Key words and phrases.* Optimal quantizers, quantization error, dynamical systems, Diophantine approximation, and independent random variables.

with respect to  $P$ , an estimate  $\min_{a \in \alpha_n} \|x - a\| = O(\rho_n)$  as  $n \rightarrow \infty$ , we would know that also  $\int \min_{a \in \alpha_n} \|x - a\| dP(x) = O(\rho_n)$  as  $n \rightarrow \infty$ . However, estimates for the mean distortion error will generally give us limited insight into the best estimates for the point distortion error that would hold for a.e.  $x$ .

In the case that the support  $K$  of  $P$  is compact, there is a geometric approach for getting a uniform estimate for the point distortion error. Take any  $\alpha_n \in \mathcal{D}_n$ . Think of it as generated by some process that we believe has a distortion error that is close to the optimal distortion error. Let  $r_n^{\alpha_n}$  be the minimum radius  $r$  such that the balls of radius  $r$  with centers at  $a \in \alpha_n$  cover  $K$ . We call  $r_n^{\alpha_n}$  the *geometric distortion error*. This distance provides a good geometric proxy for the distortion errors. Indeed, it is clear that uniformly the point distortion error for  $\alpha_n$  is not larger than  $r_n^{\alpha_n}$  and hence the minimal mean distortion error  $\int \min_{a \in \alpha_n^*} \|x - a\| dP(x)$  is no larger than  $r_n^{\alpha_n}$ . We will see that many stochastic processes can give good estimates for  $r_n^{\alpha_n}$  with a suitable choice of  $\alpha_n$ , and hence using these processes not only would we have an estimate for the point distortion error, but we also would have an estimate for the mean distortion error.

**1.2. Using Stochastic Processes.** Here is the full range of how one might consider distortion errors generated by some sequence of stochastic processes  $\mathcal{T} = (\mathcal{T}_k : k \geq 1)$ . We assume that the processes are defined on an underlying state space  $(\Omega, \beta_p, p)$  which might be a probability space, but generally would be a  $\sigma$ -finite measure space. For  $\omega \in \Omega$ , the range values of  $\mathcal{T}_k(\omega)$  are in another measure space  $(Y, \beta_P, P)$  which is also naturally a metric space with metric  $\delta_Y$ . Elements  $\omega \in \Omega$  are the parameters and the values  $(\mathcal{T}_k(\omega) : k \geq 1)$  the quantizers associated with the parameter  $\omega$ .

Given  $\omega \in \Omega$  and  $y \in Y$ , we form the point distortion errors  $d_n^{\mathcal{T}}(\omega, y) = \min_{1 \leq k \leq n} \delta_Y(y, \mathcal{T}_k(\omega))$ .

The mean distortion error  $I_n^{\mathcal{T}}(\omega) = \int d_n^{\mathcal{T}}(\omega, y) dP(y)$  is itself a stochastic process implicitly dependent on the choice of  $\omega$ . For a full analysis, we want to calculate the distribution function and all the moments of  $d_n^{\mathcal{T}}(\omega, y)$ . So in particular, we would want to know the distributional behavior and moments of  $I_n^{\mathcal{T}}(\omega)$ . It would be ideal to have the distribution functions and moments of these quantities explicitly in terms of any additional parameters defining the stochastic process  $(\mathcal{T}_k : k \geq 1)$ . But we may have to accept just good estimates for sizes of these as a function of  $n$ . In the case when the range space  $Y$  is actually a Euclidean space  $\mathbb{R}^l$  with the usual Euclidean metric for  $\delta$ , then we would also want to know how these quantities depend on the dimension  $l$ .

As indicated above, we can always try to overestimate these quantities using the geometric distortion error. It turns out that sometimes the best we can do is to overestimate the geometric distortion error using the discrepancy of the stochastic processes  $(\mathcal{T}_k : k \geq 1)$ . This discrepancy is the main tool in Monte-Carlo integration for the stochastic process and it plays an important role in quantization too. In this case, the geometric distortion would be calculated in terms of  $\delta_Y$  and would be a function primarily of  $\omega$  and  $n$ . The discrepancy would also depend on  $\omega$  and  $n$ , but would be calculated with respect to a test family of sets, for example the  $l$ -dimensional bounded rectangles in the case that  $Y = \mathbb{R}^l$ .

In summary, this is the basic program that we use. We choose a sequence  $\alpha = (\alpha(k) : k \geq 1)$  by via some stochastic process that is explicit, even though it might not give the optimal quantizers. We then compute an estimate for the mean distortion error, the point distortion error, or the geometric distortion error using the  $n$  points  $\alpha_n = (\alpha(1), \dots, \alpha(n))$ . Say we have

done this for the geometric distortion error. We know that

$$V_n(P) \leq \int \min_{1 \leq k \leq n} \|x - \alpha(k)\| dP(x) \leq r_n^{\alpha_n}.$$

If we also have  $r_n^{\alpha_n} \leq KV_n(P)$ , then we have a good estimate for the optimal mean distortion error using the geometric distortion error. The best outcome would be if in fact

$$V_n(P) \sim \int \min_{1 \leq k \leq n} \|x - \alpha(k)\| dP(x).$$

Then  $\alpha_n = (\alpha(1), \dots, \alpha(n))$  would be what we call an *asymptotically optimal set of  $n$ -means*. An even better situation would be if in addition we had  $r_n^{\alpha_n} \sim V_n(P)$ . However, this last estimate, indeed both of these, seem to rarely be the case.

**1.3. Outline.** The comments above explain why we focus on results for stochastic processes that give estimates of the form  $r_n^{\alpha_n} = O(\rho_n)$  with  $\rho_n \rightarrow 0$  as  $n \rightarrow \infty$ . We are most interested in explicit sequences that give asymptotically optimal  $n$ -means, for given probability measures, but these are difficult to obtain. So we are also interested in methods that would seem likely to give good estimates of this type, and how close the error rate is to that of the asymptotically optimal  $n$ -means.

In this article, we first consider general dynamical models in Section 2 for which there usually is correlation of the outputs. We prove a Baire category result to show that there is no global distortion rate for this large class. Then we discuss what is available for the more special case of random models in Section 3 with uncorrelated variables. Following this, we look at another special class of dynamical systems, the Diophantine models in Section 4. For these there is again correlation, but now number theory plays a critical role. Each of these approaches has advantages over the other. They also have advantages over carrying out the detailed, hard work needed to construct explicit optimal  $n$ -means. The trade-off is that one generally obtains only at best an estimate that is on the order of the optimal results.

## 2. DYNAMICAL MODELS

Suppose  $\tau$  is an invertible, Lebesgue measure  $m$  preserving transformation of  $[0, 1]$ . Sometimes we may want to consider maps acting on probability spaces with a more complicated geometry, but for many of the points we want to make in this article using  $[0, 1]$  as the underlying space is perfect for giving the basic results.

Suppose  $\tau$  is ergodic. Then for a.e.  $x \in [0, 1]$ , the orbit  $\{\tau^k(x) : k \geq 1\}$  is dense in  $[0, 1]$ . For any  $y \in [0, 1]$ , consider the approximation  $d_n^\tau(x, y) = \min_{1 \leq k \leq n} |\tau^k(x) - y|$ . We know that for every  $y$  and a.e.  $x$ , we have  $d_n^\tau(x, y) \rightarrow 0$  as  $n \rightarrow \infty$ . We would like to know estimates for the rate that this tends to zero. Some ways this could be considered are these, with each condition being weaker than the previous one: a) what rate does the  $\max_{y \in [0, 1]} d_n^\tau(x, y)$  tend to zero for a.e.  $x$ , b) what rates does  $d_n^\tau(x, y)$  tend to zero for every (or just a.e.)  $y$  and a.e.  $x$ , or c) what rate does  $\int_0^1 d_n^\tau(x, y) dm(y)$  tend to zero for a.e.  $x$ ? The ergodicity of  $\tau$  guarantees that  $M_n^\tau = \max_{y \in [0, 1]} d_n^\tau(x, y)$  tends to zero for a.e.  $x$ , and so each of these measures do converge to zero. We are asking for types of quantization error rate given by the  $n$  means  $\{\tau(x), \dots, \tau^n(x)\}$ .

We do want to keep in mind that there is another quantization error that is stronger than all of the ones above: a gap measurement. For each  $n$ , take  $(x_k : 0 \leq k \leq n + 1)$  to be

$(\tau^k(x) : 1 \leq k \leq n) \cup \{0, 1\}$  in increasing order i.e.  $0 = x_0 \leq x_1 \leq \dots \leq x_n \leq x_{n+1} = 1$ . Let the gap measurement be  $g_n^\tau(x) = \max\{x_{k+1} - x_k : 0 \leq k \leq n\}$ . Knowing the rate this goes to zero would be the best we could expect to do since  $g_n^\tau(x) \geq 2 \max_{y \in [0,1]} d_n^\tau(x, y)$ . But also, the geometric distortion error is the same thing since  $2r_n^\tau(x) = g_n^\tau(x)$ .

It is clear that the same question as asked above can be considered for any minimal dynamical system. That is, take a compact metric space  $(X, d_X)$  and a minimal homeomorphism  $\tau$  of  $X$ . Let  $d_n^\tau(x, y) = \min_{1 \leq k \leq n} d_X(\tau^k(x), y)$ . What can be said about how fast this tends to zero for each  $y$  or maximized over  $y \in X$ ?

**Remark 2.1.** We expect that the questions above for ergodic dynamical systems might be best asked in the original natural geometric structure on which the map  $\tau$  is defined. For example, take an ergodic automorphism  $\tau$  of the two torus  $\mathbb{T}^2$ . Replace the distance function of absolute value by the two dimensional Euclidean distance inherited on  $\mathbb{T}^2$  from the natural map  $\pi : [0, 1]^2 \rightarrow \mathbb{T}^2$ . See Remark 2.13 for more details.

We could also be considering a differentiable mapping  $\tau$  on a manifold  $M$  that is ergodic with respect to a natural probability measure on  $M$ . These types of geometric structures have been studied extensively from the viewpoint of generalizing the classical discrepancy calculations that we commented on earlier. But there seems to be not as much known about the quantization approximation itself.

It is evident also from a number of cases that a perhaps better measure of the quantization error would be  $I_n^\tau = \int_0^1 d_n^\tau(x, y) dm(y)$ , taken for any  $x \in [0, 1]$ . Clearly,  $I_n^\tau(x) \leq M_n^\tau(x)$ . See in particular this issue as discussed in Section 3 and Section 4.

**2.2. Rates are Always Non-generic.** The above discussion show us that the following Baire category result gives us some important information about the limits of geometric distortion errors for the class of stationary dynamical systems. We take  $\mathcal{M}$  to be the group of invertible, measure-preserving transformations of the probability space  $([0, 1], m)$ . We consider the weak topology on  $\mathcal{M}$  i.e. the strong operator topology on the group of continuous linear operators  $T^\tau(f) = f \circ \tau$ , with  $f \in L^2([0, 1])$ . A standard fact is that the weak topology on  $\mathcal{M}$  is given by a complete pseudo-metric. In particular, it is a *Baire space* i.e. any intersection of a countable set of open, dense sets is also dense. We say that a set  $\mathcal{G}$  in a Baire space is residual if it contains such a dense  $G_\delta$  set, and a set  $\mathcal{F}$  is meager if it is a complement of a residual set.

**Theorem 2.3.** *Take a fixed  $\delta > 0$  and  $\rho_n \rightarrow 0$ ,  $\rho_n > 0$  for all  $n$ . Consider the set  $\mathcal{F}$  of maps  $\tau \in \mathcal{M}$  such that on a set of  $x \in [0, 1]$  with measure at least  $\delta$ , we have  $I_n^\tau(x) \leq K\rho_n$  for some constant  $K$  and for large enough  $n$ , both depending on  $x$ . Then this is a meager set in the weak topology; actually its complement is a dense  $G_\delta$  set.*

*Proof.* We write  $\mathcal{F}$  as

$$\bigcup_{K=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\tau : m\{x : I_n^\tau(x) \leq K\rho_n\} \geq \delta\}.$$

Let  $\mathcal{B}_N = \bigcap_{n=N}^{\infty} \{\tau : m\{x : I_n^\tau(x) \leq K\rho_n\} \geq \delta\}$ . We can see this is closed in the weak topology determined by  $\Delta$  if we show each of the sets  $\{\tau : m\{x : I_n^\tau(x) \leq K\rho_n\} \geq \delta\}$  is closed in this topology. We can see this last fact is true as follows. Take any sequence  $(\tau_s)$  in  $\{\tau : m\{x : I_n^\tau(x) \leq K\rho_n\} \geq \delta\}$ . Suppose  $(\tau_s)$  converges to  $\sigma \in \mathcal{M}$  in the metric  $\Delta$ . This

means that for all  $k \geq 1$ ,  $\|f \circ \tau_s^k - f \circ \sigma^k\|_1 \rightarrow 0$  as  $s \rightarrow \infty$  for every  $f \in L^1(X)$ . So for all  $k \geq 1$  and  $x$ ,  $\int |x - \tau_s^k(y)| dm(y) \rightarrow \int |x - \tau^k(y)| dm(y)$  as  $s \rightarrow \infty$ . Hence, for all  $n$  and  $x$ ,  $I_n^{\tau_s}(x) = \int \min_{1 \leq k \leq n} |x - \tau_s^k(y)| dm(y)$  converges to  $I_n^\tau(x) = \int \min_{1 \leq k \leq n} |x - \tau^k(y)| dm(y)$  as  $s \rightarrow \infty$ .

Thus, because  $m\{x : I_n^{\tau_s}(x) \leq K\rho_n\} \geq \delta$  for all  $s$ , we also have  $m\{x : I_n^\tau(x) \leq K\rho_n\} \geq \delta$ .

This shows that  $\mathcal{F}$  is an  $F_\sigma$  set in the weak topology. So to prove our result, we need only to show that  $\mathcal{B}_N$  contains no interior. Suppose on the contrary that  $\tau \in W \subset B_N$  with  $W$  being a weak open set. Then by standard methods one can show there is a cyclic transformation  $\tau_0 \in W$ . Even more, we can have a partition  $\{E_j : j = 1, \dots, J\}$  of  $X$  with each set  $m(E_j) = 1/J$  and such that if  $\tau_0$  is any measure preserving map that sends the sets  $E_j$ , sending  $E_j$  to  $E_{j+1}$  for all  $j$ , and  $E_J$  to  $E_1$ , then  $\tau_0 \in W$ . Now (if necessary) we modify  $\tau_0$  (but keeping it a cyclic permutation of the sets  $E_j$  as above). First, take intervals  $S_j$  of the same small, positive length and let  $D_j = E_j \setminus S_j$  and  $C_j = E_j \cap D_j$ . We now need to also adjust the lengths of  $S_j$ , keeping them of some small, positive length (possibly different) so that the sets  $D_j$  all have the same measure. We now modify  $\tau_0$  so that it actually cyclically permutes the sets  $D_j$  in the sense above, and as a result the sets  $C_j$  too, with the same permutation of the index  $j \rightarrow j+1, 1 \leq j < J$  and  $J \rightarrow 1$ . These adjustments are made so that  $\tau_0$  is still in  $W$ . Considering the center half of the intervals  $S_j$ , we see that for all  $x \in \bigcup_{j=1}^J D_j$ , we have for all  $m$ ,  $I_m^{\tau_0}(x) \geq \gamma$  for some, possibly quite small, value  $\gamma > 0$ . Now, as part of these adjustments, we can arrange that  $\bigcup_{j=1}^J D_j$  has measure at least  $1 - \frac{\delta}{2}$ . This combination of estimates guarantees that for all  $m$ , there exists  $x$  such that  $\gamma \leq I_m^{\tau_0}(x) \leq K\rho_m$ . But since  $\rho_m \rightarrow 0$  as  $m \rightarrow \infty$ , we cannot have  $\gamma \leq K\rho_m$  for all  $m$ . The conclusion is that  $B_N$  cannot have interior in the weak topology.  $\square$

**Remark 2.4.** This result means that there is no rate for  $I_n^\tau$ , no matter how slow, for the quantization error rate which would apply to a Baire category large set of maps  $\mathcal{M}$ . Since the ergodic mappings are themselves a dense  $G_\delta$  set, for any specific quantization error rate as above, there would be (many) ergodic mappings that do not satisfy this condition. This means the same thing would be true when attempting to get a general rate result for  $d_n^\tau(x)$ . Moreover, it means that there cannot be a pointwise rate either i.e. some rate  $(\rho_n)$  such that for all  $\tau \in \mathcal{M}$ ,  $I_n^\tau(x) \leq \rho_n$  for large enough  $n$  depending on  $x$ , which holds for all  $x$  in some set of non-zero measure.

**Remark 2.5.** There has recently been a similar result proved by A. Junqueira [J] in the topological setting. It seems likely that just assuming a one point rate as in [J] will not be adequate in the measure-theoretic category because maps can be changed on a null set without changing their measure-theoretic dynamics.

We state this corollary of Theorem 2.3 to emphasize what has been proved.

**Corollary 2.6.** Take a fixed  $\delta > 0$  and  $\rho_n \rightarrow 0, \rho_n > 0$  for all  $n$ . Consider the set  $\mathcal{F}$  of maps  $\tau \in \mathcal{M}$  such that on a set of  $x \in [0, 1]$  with measure at least  $\delta$ , we have the maximum gap  $g_n^\tau(x) \leq K\rho_n$  for some constant  $K$  and for all large enough  $n$ , depending on  $x$ . Then this is a meager set in the weak topology; actually its complement is a dense  $G_\delta$  set.

This is not to say that there cannot be a good rate result for a large set of ergodic maps. Indeed, take any ergodic rotation  $\tau_\alpha$  given by  $\alpha$  whose terms in the simple continued fraction

for  $\alpha$  are bounded. Then the result in Graham and Van Lint [GVL], and also in Schoissengeier [Sch] shows that  $\rho_n = \frac{1}{n}$  does work for the quantization error rate.

It is not clear how pervasive this rate result can be since conjugating these examples does not necessarily preserve the gap structure. However, we can certainly get some rate for a dense class. Just take a countable dense set of ergodic mappings and use diagonalization to obtain a sufficiently slow rate. This gives,

**Proposition 2.7.** *There is a countable set  $\mathcal{S}$  of ergodic mappings in  $\mathcal{M}$  which is dense in weak topology and a decreasing sequence  $\rho_n \rightarrow 0$  as  $n \rightarrow \infty$  such that  $r_n^\tau(x) = O(\rho_n)$  for all  $\tau \in \mathcal{S}$  and a.e.  $x$ .*

**Remark 2.8.** It would be very useful to know that via cutting and stacking one can construct an ergodic, rank one map  $\tau$  such that  $2nd_n^\tau(x) \rightarrow 1$  as  $n \rightarrow \infty$  for a.e.  $x$ . It is not clear what rate works for maps like Chacon's map, and it is not likely to have this property. But another such construction, with many divisions and some spacers at each inductive step should give at least an ergodic mapping, and one with this asymptotically optimal point distortion rate.

**2.9. Discrepancy.** The quantization process is closely related to discrepancy estimates. See Kuipers and Niederreiter [KN], especially the chapter notes, for a wealth of background information and references on discrepancy. Another more recent reference that is also excellent is Drmota and Tichy [DT]. We again take our interval modulo one, but we suppress this in the notation for simplicity.

**Definition 2.10.** Given a sequence  $\alpha = (\alpha(k) : k \geq 1)$  in  $[0, 1]$ , the discrepancy

$$D_n(\alpha) = \sup\left\{\left|\frac{1}{n} \sum_{k=1}^n 1_{[x,y]}(\alpha(k)) - (y-x)\right| : 0 \leq x < y \leq 1\right\}.$$

The discrepancy

$$D_n^*(\alpha) = \sup\left\{\left|\frac{1}{n} \sum_{k=1}^n 1_{[0,y]}(\alpha(k)) - y\right| : 0 < y \leq 1\right\}.$$

It is easy to see that  $D_n^* \leq D_n \leq 2D_n^*$ .

Now if  $D_n < \delta$ , then for any interval  $I$  of length  $\delta$ , there must be some  $\alpha_k \in I$  with  $k \leq n$ . So  $\min_{1 \leq k \leq n} |x - \alpha(k)| \leq \delta$ . Hence, we have the following useful basic estimate:

**Lemma 2.11.** *For any  $x \in [0, 1]$ , we have  $\min_{1 \leq k \leq n} |x - \alpha(k)| \leq D_n(\alpha)$ .*

In particular, we can consider the discrepancy  $D_n(O_\tau(x))$  where  $O_\tau(x)$  is the forward time sequence  $(\tau^k(x) : k \geq 1)$  for an ergodic map  $\tau$ . As a consequence of the estimate in Lemma 2.11 and the non-generic behavior in Theorem 2.3, we have this Baire category result.

**Corollary 2.12.** Take a fixed  $\delta > 0$  and  $\rho_n \rightarrow 0$ ,  $\rho_n > 0$  for all  $n$ . Consider the set  $\mathcal{F}$  of maps  $\tau \in \mathcal{M}$  such that on a set of  $x \in [0, 1]$  with measure at least  $\delta$ , we have  $D_n(O_\tau(x)) \leq K\rho_n$  for some constant  $K$  and for large enough  $n$ , both depending on  $x$ . Then this is a meager set in the weak topology on  $\mathcal{M}$ .

**Remark 2.13.** Here is an example of related dynamical rate questions that occur when the underlying measure space is changed for which the only result on geometric distortion error rates that we know is through discrepancy estimates. Take an ergodic automorphism  $A$  of

the two torus  $\mathbb{T}^2$ . Let  $d_n^A(x, y) = \min_{1 \leq k \leq n} d_{\mathbb{T}^2}(y, A^k x)$ , where  $d_{\mathbb{T}^2}$  is the natural Euclidean metric distance (taken mod one actually). Let  $I_n^A(x) = \int_{\mathbb{T}^2} d_n^A(x, y) d\lambda_{\mathbb{T}^2}(y)$ . What can be said about how fast this goes to zero for a.e.  $x$ ? The estimate to prove or deny here would be that  $I_n^A = O(1/n)$  since the ergodic map  $A$  is Bernoulli, and hence reasonably quickly mixing. If instead, we consider  $d_n^A(x) = \max_{y \in \mathbb{T}^2} d_n^A(x, y)$ , what is the rate that this overestimate for  $I_n^A$  goes to zero for a.e.  $x$ ? What we are most interested in is if these rates are really faster than what would come out of using overestimates given by the natural two dimensional discrepancy

$$D_n^A = \sup_{r < s, u < v} \left| \frac{1}{n} \sum_{k=1}^n 1_{[r,s] \times [u,v]}(T^k x) - (s-r)(v-u) \right|.$$

See Losert, Nowak, and Tichy [LNT], and Nowak and Tichy [NT], and citations in these articles, for the overestimates on discrepancy that are available here, and in higher dimensions. At least when one eigenvalue of  $A$  has modulus larger than 1, we have  $D_n^A = O(\ln^5(n)/\sqrt{n})$ .

**2.14. Shrinking Targets.** Consider an ergodic mapping of  $[0, 1]$  and a fixed sequence  $(\rho_n)$  decreasing to zero. A *geometric shrinking target* result would give information about the Lebesgue measure of the set of points  $x$  such that we have for all  $y$  (or perhaps only a.e.  $y$ ), infinitely often  $\tau^n(x) \in [y - \rho_n, y + \rho_n]$ . Typically we would want to know that either this holds for a.e.  $x$ , or in contrast to this it might hold only on a null set of  $x$ . The first case is called a *visible geometric shrinking target*, and the second case is called an *invisible geometric shrinking target*. These extremes are the best types of geometric shrinking target properties.

It turns out that the geometric distortion error  $r_n^\tau(x)$  gives an estimate for a shrinking target rate. Indeed, using Boshernitzan's Theorem [B], we prove this theorem in [RR2].

**Proposition 2.15.** *Suppose we have a sequence  $(\rho_n)$  such that for a.e.  $x$ , we have the geometric distortion error  $r_n^\tau(x) \leq \rho_n$  for large enough  $n$  depending on  $x$ . Then we have the geometric shrinking target behavior that for a.e.  $x$ , we have for all  $y$ , infinitely often  $\tau^n(x) \in [y - \rho_n, y + \rho_n]$ .*

It follows that there is always some shrinking target rate.

**Proposition 2.16.** *For every ergodic mapping  $\tau$ , there is a sequence  $(\rho_n)$  tending to zero so that for all  $y \in [0, 1]$ , the sequence of intervals  $(B_{\rho_n}(y) : n \geq 1)$  is an a.e. visible geometric shrinking target with respect to  $\tau$ .*

**Remark 2.17.** Shrinking target theorems for general dynamical systems are studied extensively in the Rosenblatt and Roychowdhury [RR2], where there also are background and references on this topic. In the end, the best values of  $r_n^\tau$  are not known in general. Indeed, we see in this article sometimes we have to use the (generally) much larger discrepancy of the sequence  $(\tau^k(x) : k \geq 1)$  to get an upper bound on the geometric distortion error. Similarly, there is generally a loss of speed when using the geometric distortion error to derive a geometric shrinking target results. This is what allows for difference in Baire category results when considering suitable rates in shrinking target theorems as opposed to geometric distortion error rates. In particular, in [RR2], it is shown there that shrinking target properties are sufficiently flexible so the behavior is actually generic in many non-trivial cases, unlike with quantization error rates where the behavior is only for a first category class of maps, as noted in Theorem 2.3 and Corollary 2.12.

## 3. IID MODELS

Consider a method of randomly generating  $n$ -means for uniform measure on the interval  $[0, 1]$  modulo one. We take  $\alpha = (\alpha(k) : k \geq 1)$  to be IID random variables with uniform distribution. We actually are taking  $\alpha(k, \omega)$  with  $\omega \in \Omega$  as the model underlying probability space  $(\Omega, P)$ , but we will suppress the dependence on  $\omega$  if it will not create confusion.

Suppose we want an estimate for  $d_n(\{\alpha_1, \dots, \alpha_n\})$ . The simplest approach would be to estimate how many terms  $(\alpha(1), \dots, \alpha(n))$  are needed so that each interval  $I_j = [j/M, (j+1)/M], j = 0, \dots, M-1$  contains at least one point, with high probability. This will guarantee that our quantization error  $\int_0^1 \min_{1 \leq k \leq n} |x - \alpha(k)| dx$  is no larger than  $M \int_0^{1/M} x dx = O(1/M)$ .

**Proposition 3.1.** *With probability  $1 - 1/\ln(n)$ , the point distortion error  $d_n(\{\alpha_1, \dots, \alpha_n\}) = O(\ln(n)/n)$ .*

*Proof.* It is easiest to consider the probability of the complementary case: no term  $\alpha(k), k = 1, \dots, n$  is in some  $I_j$ . This probability is  $(1 - \frac{1}{M})^n$  for each such  $j$ . So an estimate for the entire scope of the possibility is  $M(1 - \frac{1}{M})^n$ . Taking  $M = n/\ln(n)$  as a real variable would give for large  $n$ ,  $M(1 - \frac{1}{M})^n \sim 1/\ln(n)$ . Hence, with probability  $1 - 1/\ln(n)$ , each  $I_j$  contains some  $\alpha(k), 1 \leq k \leq n$ . This gives the estimate  $1/M = \ln(n)/n$  for the quantization error with this probability.  $\square$

**Remark 3.2.** Proposition 3.1 only gives convergence in measure as  $n$  goes to  $\infty$ , but a simple increase in the size of  $M$  can guarantee an almost sure result. Note: instead of the optimal distortion error of  $C/n$ , this approach is giving a somewhat worse estimate of  $C \ln(n)/n$ .

We can get more information from a distributional calculation. Consider the probability  $P(\{\omega : n \min_{1 \leq k \leq n} |x - \alpha(k, \omega)| \geq t\})$ . It is easy to see that this is  $(1 - \frac{2t}{n})^n$ . So scaling of the distortion error by  $n$  results in convergence in distribution to the distribution function  $d(t) = 1 - e^{-2t}, t \geq 0$ , one can also compute expectations, and other moments. For example, we have the following result.

**Proposition 3.3.** *The probabilistic mean of the point distortion error in the IID case satisfies*

$$\int_{\Omega} \min_{1 \leq k \leq n} |x - \alpha(k, \omega)| dP(\omega) = O(1/n).$$

*Proof.* One computes

$$\begin{aligned} & \int_{\Omega} n \min_{1 \leq k \leq n} |x - \alpha(k, \omega)| dP(\omega) \\ &= \int_0^{\infty} P(\{\omega : n \min_{1 \leq k \leq n} |x - \alpha(k, \omega)| \geq t\}) dt = \int_0^{n/2} (1 - 2t/n)^n dt = \frac{n}{2(n+1)}. \end{aligned}$$

$\square$

**Remark 3.4.** Going further than this distributional convergence is not going to be possible because of the Hewitt-Savage Theorem [HS]. It shows that if this sequence converges a.e. or even just in measure, then the limit function would be a constant. The distributional convergence shows that this is not possible.

Suppose we want to improve our estimate on the point distortion error. As remarked in Section 2.9, the quantization process is closely related to discrepancy estimates. Here is how this works out in the case of IID sequences. We use the following result of K-L Chung [KLC] which is the best possible discrepancy result for IID sequences.

**Theorem 3.5.** *For a.e.  $\omega$ ,*

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{2n} D_n^*(\boldsymbol{\alpha}(\omega))}{\sqrt{\ln \ln(n)}} = 1.$$

Combining this with Lemma 2.11 gives this upper bound on the distortion error.

**Corollary 3.6.** *For a.e.  $\omega$ ,  $\min_{1 \leq k \leq n} |x - \alpha(k)(\omega)| = O(\sqrt{\ln \ln(n)}/\sqrt{n})$ .*

However, the actual point distortion error rate in the IID case is much faster than what this discrepancy estimate gives. For simplicity of notation, for the maximum gap measure in this case, write  $g_n(\omega)$  instead of  $g_n^\alpha$  where  $\alpha = \{\alpha(1, \omega), \dots, \alpha(n, \omega)\}$ . Here is the result of Levy [L].

**Proposition 3.7.** *For an IID uniform sequence in  $[0, 1]$ ,  $ng_n(\omega) = O(\ln(n))$ .*

What we are considering here is the first step in the classical problem of non-parametric statistics: gap measurements. Beyond the largest gaps, one can also look for the next largest gap, the next, and so on, and the sizes and distribution of the values of these successive gaps. This type of order statistics of uniformly distributed IID processes in  $[0, 1]$  is now very well understood. For improvement on Levy's Theorem and more on order statistics for uniform random variables see the articles by L. Devroye [LD1, LD2] and P. Deheuvels [PD1, PD2, PD3]. Indeed, Levy's Theorem can be made more specific. For example, in L. Devroye [LD2], the following result is shown.

**Proposition 3.8.** *Almost surely in  $\omega$*

$$\liminf_{n \rightarrow \infty} (ng_n(\omega) - \ln(n) + \ln \ln \ln(n)) = -\ln(2)$$

and

$$\limsup_{n \rightarrow \infty} (ng_n(\omega) - \ln(n))/2 \ln \ln(n) = 1.$$

There is one important improvement that can be made; see Cohort [PC]. If we also integrate with respect to  $x$  as above, then there is a.s. convergence to a computable constant. That is, we switch from a point distortion error to the mean distortion error for the IID model. Then we have the following.

**Proposition 3.9.** *There is a non-zero constant  $C$  such that for a.e.  $\omega$ ,*

$$\int_0^1 n \min_{1 \leq k \leq n} |x - \alpha(k, \omega)| dx$$

converges to  $C$  as  $n \rightarrow \infty$ .

**Remark 3.10.** Proposition 3.9 is proved by calculating variances and using an infinite series method. Cohort [PC] actually carries these calculations out in greater generality than our one dimensional setting. This article contains other interesting results related to a.s. convergence of the random proxy for optimal  $n$ -means and conclusions that follow about the asymptotic optimality of the random  $n$ -means.

**Remark 3.11.** a) The most important point that the result in Cohort [PC] gives us is that the mean distortion error is generally much smaller than the point distortion error, or a geometric distortion error like the maximum gap measurement. These in turn are much smaller than what would be given by using just the discrepancy of the random sequence. But then it is well known that  $\limsup_{n \rightarrow \infty} nD_n = \infty$  for every sequence. So it is not surprising that the same thing is true in this IID model for  $nr_n(\omega)$ , although we will see in Section 4 that this type of divergence does not hold in some interesting deterministic settings. In any case, the probability and statistics literature give good results for the IID model for the exact rate at worst that  $nr_n(\omega)$  tends to  $\infty$ . Calculating this rate for higher dimensional cubes, and with respect to other common distributions for the IID sequence besides the uniform distribution, would be very worthwhile.

b) By the results in Section 2.14, specifically Proposition 2.15, by the above results, we would get a geometric shrinking target rate of  $\ln(n)/n$  for the random sequence of centers. However, there is a much better result available as given in Shepp [Sh]. See Rosenblatt and Roychowdhury [RR2] for more details.

If the measure  $P$  that we are quantizing is not uniform, then to get a good quantization, we need to adjust the placement of the random variables  $(\alpha(k) : k \geq 1)$ . The obvious approach is to just take  $\alpha(k)$  to be IID with distribution given by the fixed probability measure  $P$ . Then we would have the empirical measures  $\frac{1}{n} \sum_{k=1}^n \delta_{\alpha(k)}$  converging weakly to  $P$ . The result of Theorem 7.5 in Graf and Luschgy [GL] shows that our random empirical measure would not be asymptotically optimal except in the case of uniform measure. See also the discussion following Theorem 7.5 in [GL].

#### 4. DIOPHANTINE MODELS

**4.1. The Weyl sequences.** Now consider a method of generating good  $n$ -means that relies on some classical number theory: a Diophantine model. We take  $\alpha(k, \theta) = \{k\theta\}$  for all  $k \geq 1$ . Here  $\theta$  is some irrational number and  $\{t\}$  denotes the fraction in  $[0, 1)$  such that  $t = \{t\} + k$  for some integer  $k$ .

We know that  $\alpha(\theta) = (\alpha(k, \theta) : k \geq 1)$  is uniformly distributed on  $[0, 1]$  and moreover there is a classical estimate in Khinchin [AK] for the discrepancy  $D_n(\alpha(\theta))$  that holds for a.e.  $\theta$ . This estimate comes from metric facts about continued fractions and Diophantine approximation. This discrepancy estimate is in some sense parallel to the iterated logarithm method of Chung. See Kuipers and Niederreiter [KN] for some discussion of this theorem of Khinchin.

**Theorem 4.2.** *For any non-decreasing  $g$  such that  $\sum_{n=1}^{\infty} \frac{1}{g(n)} < \infty$ , for a.e.  $\theta$ , one has for the sequence  $\alpha(\theta) = (k\theta \bmod 1 : k \geq 1)$*

$$nD_n(\alpha(\theta)) = O(\ln(n)g(\ln \ln(n))).$$

**Remark 4.3.** Curiously enough, the discrepancy estimate here is better than the discrepancy estimate available in the IID case.

From Theorem 4.2 and Lemma 2.11, we clearly have this parallel to Proposition 3.1.

**Proposition 4.4.** *If  $\delta > 0$ , for a.e.  $\theta$ , the point distortion error  $d_n(\alpha(\theta)) = O(\ln^{(1+\delta)}(n)/n)$ .*

**Remark 4.5.** The estimate in Proposition 4.4 is not as good as the optimal one that would be  $C/n$ . Actually, this estimate is not even as good as the one from the IID model, specifically Proposition 3.1. However, this Diophantine approach has the virtue of being somewhat more explicit than the IID approach.

The overestimate above from Khinchin's Discrepancy Theorem is most likely too large to give a good rate for the distortion error in the Diophantine model. For example, see the results in Graham and Van Lint [GVL]. Their results on gaps tells us some very interesting facts. Let  $d_n^\theta(y) = \min_{1 \leq k \leq n} |y - \{k\theta\}|$  and let  $M_n^\theta = \max_{y \in [0,1]} d_n^\theta(y)$ .

**Proposition 4.6.** *There is an absolute constant  $\gamma > 0$  such that for any irrational  $\theta$ , we have*

$$\liminf_{n \rightarrow \infty} nM_n^\theta + \gamma \leq \limsup_{n \rightarrow \infty} nM_n^\theta.$$

This means that the quantization error rate obtained from  $(\{k\theta\} : k \geq 1)$  can never be optimal in the strict sense, but only optimal by a bounded proportion of the optimal error rate. These examples also demonstrate another sure way that discrepancy is too large. But interestingly enough, if  $\theta$  has bounded terms in its simple, continued fraction expansion, then  $\sup_{n \geq 1} nM_n^\theta < \infty$ . Indeed, this is characteristic of such irrational numbers.

**Remark 4.7.** There does not seem to be a result in the literature of the following type: there is a sequence  $(h_n^* : n \geq 1)$  with  $h_n^*$  increasing to  $\infty$ , such that for Lebesgue a.e.  $\theta$ , we have  $nM_n^\theta = O(h_n^*)$ , and for any other sequence  $(h_n)$  with this property,  $\sup_{n \geq 1} h_n^*/h_n$  is bounded.

Note: the discrepancy estimate we have used shows that for any  $\delta > 0$ ,  $nM_n^\theta = O(\ln^{1+\delta}(n))$ . So  $\ln^{1+\delta}(n)$  a candidate for  $h_n^*$ . However, we do not believe that this rate is optimal and we are still seeking the optimal quantization error rate  $h_n^*$  for this model.

Generally, we have for some  $C > 0$ ,  $nD_n \geq C \ln(n)$  infinitely often; see Kuipers and Niederreiter [KN], Theorem 2.2 with  $k = 1$ , while  $nd_n^\theta$  is bounded when  $\theta$  has bounded terms in its continued fraction expansion. Note: for these same  $\theta$ , the overestimate for  $nD_n$  is  $C \ln(n)$ ; see Kuipers and Niederreiter [KN], Theorem 3.4.

**Remark 4.8.** It is possible that the Diophantine results here can be improved by a couple of other different approaches. One approach would be to take a specific very good value of  $\theta$ , actually the Golden Mean ratio. See Motta, Shipman, and Springer [MSS] where optimal transitivity is studied to limit the gaps in the sequence. Another approach would be to use bounded remainder sets so that the discrepancy error can be perhaps better controlled. See both Haynes, Kelly, and Koivusalo [HK1] and Haynes and Koivusalo [HK2].

**Remark 4.9.** Another interesting direction to pursue here is to take as our quantization sequences such as  $\{2^k\theta\}$ , or more generally  $\{n_k\theta\}$  for some increasing sequence  $(n_k)$ . We do not see right now any clear way to get gap measurements for a point distortion error rate in these cases. However, there is considerable literature on discrepancy in these cases. For example, see Philipp [P] for the case of powers of 2, that in some ways can be considered as a Diophantine version of the IID model, as can be the case using any sequence  $(n_k)$  which is lacunary. This analogy might suggest that the point distortion error and the mean distortion error are like the ones for the IID uniformly distributed case. For example, is it true that the maximal gap  $g_n$  in  $(\{2^k\theta\} : k \geq 1)$  is on the order of  $\ln(n)/n$ ? Also, see articles by Berkes, Fukuyama, and Nishimura [BFN], Aistleitner and Fukuyama [AF], and Aistleitner

and Larcher [AL1, AL2] where very interesting results on what types of discrepancy estimates are available for the sequences  $(\{n_k\theta\} : k \geq 1)$ . They show that any rate (up to standard restrictions) is possible for a suitable sequence  $(n_k)$ .

**4.10. The Farey points.** One additional parallel here is that one can replace the Diophantine model by (in some sense) its close relative: Farey fractions. It is more natural here to consider each set  $\mathcal{F}_n$  of Farey fractions as a potential good quantization for uniform measure. The Farey fractions  $\mathcal{F}_n$  of order  $n$  are all rational fractions  $p/q$  with  $q \leq n$  and  $\gcd(p, q) = 1$ . The cardinality of  $\mathcal{F}_n$  is  $\frac{3n^2}{\pi^2} + (n \ln(n))$ . As in Kargaev and Zhigljavsky [KZ1], let  $d_n^{\mathcal{F}}(x) = \min_{p/q \in \mathcal{F}_n} |x - \frac{p}{q}|$ . They give the following mean and point distortion asymptotics.

**Proposition 4.11.** *Moreover, the mean distortion error*

$$I_n^{\mathcal{F}} = \int_0^1 d_n^{\mathcal{F}}(x) dx = \frac{3 \ln(n)}{\pi^2 n^2} + O\left(\frac{1}{n^2}\right)$$

as  $n \rightarrow \infty$ .

**Proposition 4.12.** *For any  $\epsilon > 0$  and for a.e.  $x$ :*

$$\begin{aligned} \lim_{n \rightarrow \infty} n^2 d_n^{\mathcal{F}} \ln^{1+\epsilon}(n) &= \infty \\ \liminf_{n \rightarrow \infty} n^2 d_n^{\mathcal{F}} \ln(n) &= 0 \\ \limsup_{n \rightarrow \infty} n^2 d_n^{\mathcal{F}} / \ln(n) &= \infty \\ \lim_{n \rightarrow \infty} n^2 d_n^{\mathcal{F}} / \ln^{1+\epsilon}(n) &= 0. \end{aligned}$$

**Remark 4.13.** These very good estimates suggest that in the Diophantine case we might have similar results. One reason for expecting this is that the arguments for the approximation by Farey fractions often take advantage of properties of Farey fractions that also give insights into the Three Gaps Theorem for the Weyl sequence in the Diophantine model, see Polanco, Schultz, and Zaharescu [PSZ]. For a recent article about the Three Gap Theorem itself, see T. Van Ravenstein [VR]. What is thus suggested is that we can perhaps expect the gaps in  $(\{n\theta\})$  to be on the order of  $\ln(n)/n$  for a.e.  $\theta$ . But the precise over and under-estimates in terms of limit supremum and limit infimum may cause some variation in this. But also, perhaps for a.e.  $\theta$ , there is some constant  $C(\theta)$  such that

$$\int_0^1 \min_{1 \leq k \leq n} |x - \{k\theta\}| dx \sim \frac{C \ln(n)}{n}$$

as  $n \rightarrow \infty$ .

**Remark 4.14.** We should observe that there is a classical discrepancy estimate when using the Farey fractions  $\mathcal{F}_n$ . The first result in this direction appears in Niederreiter [N1]; he shows that using the  $N_n$  points in  $\mathcal{F}_n$ , one has for the discrepancy  $\frac{c_1}{\sqrt{N_n}} \leq D_{N_n}^{\mathcal{F}} \leq \frac{c_2}{\sqrt{N_n}}$  for some constants  $c_1$  and  $c_2$ . Following this, Dress [FD] gave the very nice explicit value:  $D_{N_n}^{\mathcal{F}} = \frac{1}{n}$ . Given that  $N_n \sim \frac{3n^2}{\pi^2}$ , this explains the result that Niederreiter had obtained earlier on the discrepancy of the Farey points.

The asymptotic deviation (in a suitable measure) of the Farey fractions from uniformly placed points is equivalent to the Riemann Hypothesis. The classical result is this. Take the Farey fractions  $\mathcal{F}_n$  and let  $N_n$  be the cardinality of  $\mathcal{F}_n$ . Let  $(f_k^n : 1 \leq k \leq N_n)$  be  $\mathcal{F}_n$  written in increasing order. Consider  $\Delta_n(s) = \sum_{k=1}^n |x_k - \frac{k}{n}|^s$  for some fixed  $s$ . Franel [F] proved that the Riemann Hypothesis is equivalent to  $\Delta_n(2) = O(n^r)$  for all  $r > -1$ . In particular, this would show that  $\Delta_n(2) = O(1/\sqrt{n})$ , and hence this sum deviation tends to zero as  $n \rightarrow \infty$ . Also, Landau [Landau] showed that the Riemann Hypothesis is equivalent to  $\Delta_n(1) = O(n^r)$  for all  $r > 1/2$ .

This way of measuring the offset from a regular distribution represented by uniformly spaced points suggests another way in which we can test our sequence  $\alpha$  for the distance from optimal  $n$ -means. We take any sequence  $\alpha = (\alpha(k) : k \geq 1)$  generated by random models in Section 2, Diophantine models in this section, or dynamical methods as in Section 4. We take  $x_1, \dots, x_n$  to be  $\alpha(k), k = 1, \dots, n$  written in increasing order. Then consider the difference  $\Delta_n^\alpha(s) = \sum_{k=1}^n |x_k - \frac{k}{n}|^s$  for some fixed  $s$ , e.g.  $s = 1$  or  $s = 2$ . What is the best over-estimate  $\rho(n)$  in each of these cases so that  $\Delta_n^\alpha(s) = O(\rho(n))$  for all  $n \geq 1$ ? Indeed, when do we have  $\Delta_n^\alpha(s) \rightarrow 0$  as  $n \rightarrow \infty$ ? This is at least conjecturally the case for the Farey fraction estimate in place of the Weyl sequence. What happens in the Diophantine model? What happens with a random model or a dynamical systems model? We do not know the answers here yet, but something can be said in the Diophantine case. Given  $\theta$ , take  $(x_k^{(\theta, n)} : 1 \leq k \leq n)$  to be  $\{\{k\theta\} : 1 \leq k \leq n\}$  written in increasing order. Let  $\Delta_n^\theta(s) = \sum_{k=1}^{N_n} |x_k^{(\theta, N_n)} - f_n^k|^s$ . What is the best rate control for this in terms of  $n$ , given fixed  $\theta$ ? It is not hard to see using the Three Gaps Theorem that in this case  $\Delta_n^\theta(s) = O(n)$ . So to get smaller sizes for  $\Delta_n^M(s)$ , we will have to switch to a random model or dynamical model  $\mathcal{M}$  that is not as rigid as the Diophantine model.

**Remark 4.15.** Given two sequences  $\alpha_1$  and  $\alpha_2$ , it is not clear what geometric or measure-theoretic property is equivalent to a rate control on  $\Delta_n^{(\alpha_1, \alpha_2)}(s) = \sum_{k=1}^n |\alpha_1(k) - \alpha_2(k)|^s$ .

**Remark 4.16.** It is not too difficult to prove a Baire category result of this type. Consider a Lebesgue measure-preserving, ergodic invertible map  $\tau$  of  $[0, 1]$ . Given  $y \in [0, 1]$ , let  $(x_k^{(\tau, y)} : 1 \leq k \leq n)$  be  $(\tau^k(y) : 1 \leq k \leq n)$  in increasing order. For  $p, 0 < p < \infty$ , let the deviation  $\Delta_n^{(\tau, y)}(s) = \sum_{k=1}^n |x_k^{(\tau, y)} - \frac{k}{n}|^s$ . Given any rate  $\rho_n > 0$  such that  $\rho_n/n \rightarrow 0$  as  $n \rightarrow \infty$ , and any  $\epsilon > 0$ , the set  $\mathcal{B}$  consisting of mappings  $\tau$  such that  $\Delta_n^{(\tau, y)}(s) = O(\rho_n)$  for all  $y \in B$  with  $m(B) \geq \epsilon$ , is first category in the ergodic mappings with the usual symmetric pseudo-metric. Hence, the generic mapping yields no rate result at all. It is not actually clear at this time if there are ergodic mappings for which there are rate results. The obvious candidate would be a rotation by an angle whose continued fraction expansion has bounded elements, like the Golden Mean ratio.

## REFERENCES

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

- [AF] C. Aistleitner and K. Fukuyama, *Extremal discrepancy behavior of lacunary sequences*, Monatshefte für Math 177 no. 2 (2015) 167-184.
- [AL1] C. Aistleitner and G. Larcher, *On sequences with prescribed metric discrepancy behavior*, Monatshefte für Math 181 no. 3 (2016) 507-514.
- [AL2] C. Aistleitner and G. Larcher, *Metric results on the discrepancy of sequences  $(a_n\alpha)_{n\geq 1}$  modulo one for integer sequences  $(a_n)_{n\geq 1}$  of polynomial growth*, Mathematika 62 no. 2 (2016) 478-491.
- [BN] O. Barndorff-Nielsen, *On the rate of growth of the partial maxima of a sequence of independent identically distributed random variables*, Math Scandinavica 9 (1961) 383-394.
- [BFN] I. Berkes, K. Fukuyama, and T. Nishimura, *A metric discrepancy result with given speed*, Acta Math Hungarica 151 no. 1 (2017) 199-216.
- [B] M. Boshernitzan, *Quantitative recurrence results*, Inventiones Math 113 (1993) 617-631.
- [BC] M. Boshernitzan and J. Chaika, *Diophantine properties of IETs and general systems: quantitative proximality and connectivity*, Inventiones Math 192 (2013) 375-412.
- [C] J. Chaika, *Shrinking targets for IETs: extending a theorem of Kurzweil*, Geometric Functional Analysis 21(2011) 1020-1042.
- [CR1] D. Çömez and M. K. Roychowdhury, *Quantization for uniform distributions on stretched Sierpiński triangles*, Monatshefte für Mathematik, Volume 190, Issue 1, 79-100 (2019).
- [CR2] D. Çömez and M. K. Roychowdhury, *Quantization for uniform distributions of Cantor dusts on  $\mathbb{R}^2$* , Topology Proceedings, Volume 56 (2020), Pages 195-218.
- [KLC] Kai Lai Chung, *An estimate concerning the Kolmogoroff limits distribution*, Transactions of the AMS 67 (1949) 36-50.
- [PC] P. Cohort, *Limit theorems for random normalized distortion*, Annals Applied Probability, 14, no. 1 (2004) 118-143.
- [PD3] P. Deheuvels, *Strong bounds for multidimensional spacings*, Z. Wahr. Verw. Gebiete 64 (1983) 411-424.
- [PD2] P. Deheuvels, *Upper bounds for  $k$ th maximal spacings*, Z. Wahrsch. Verw. Gebiete 62 no. 4 (1983) 465-474.
- [PD1] P. Deheuvels, *Strong limiting bounds for maximal uniform spacings*, Annals of Probability 10 no. 4 (1982) 1058-1065.
- [DR] C. P. Dettmann and M. K. Roychowdhury, *Quantization for uniform distributions on equilateral triangles*, Real Analysis Exchange, Vol. 42(1) (2017) 149-166.
- [LD2] L. Devroye, *A log log law for maximal uniform spacing*, Annals of Probability 10 no. 3 (1982) 863-868.
- [LD1] L. Devroye, *Laws of the iterated logarithm for order statistics of uniform spacings*, Annals of Probability 9 no. 5 (1981) 860-867.
- [FD] F. Dress, *Discrépance des suites de Farey*, Journal de Théorie des Nombres de Bordeaux, 11 no. 2 (1999) 345-367.
- [DT] M. Drmota and R. F. Tichy, *Sequences, Discrepancies, and Applications*, Lecture Notes in Math no. 1651, Springer-Verlag, Berlin-Heidelberg, 1997.
- [VD] V. Drobot, *On dispersion and Markov constants*, Acta Math Hungarica 47 (1-2) (1986) 89-93.
- [F] J. Franel, *Les suites de Farey et le problème des nombres premiers*, Göttinger Nachrichten (1924) 198-201.
- [FY] K. Fukuyama and M. Yamashita, *Metric discrepancy results for geometric progressions with large ratios*, Monatshefte für Math 180 no. 4 (2016) 731-742.
- [GG] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, Kluwer Academy Publishing, Boston, 1992.
- [GL] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics 1730, Springer, Berlin, 2000.
- [GL3] S. Graf and H. Luschgy, *The Quantization of the Cantor Distribution*, Math. Nachr., 183 (1997), pp. 113-133.
- [GVL] R. Graham and J. H. Van Lint, *On the distribution of  $n\theta$  modulo 1*, Canadian Journal Math, 20 (1968) 1020-1024.
- [GKL] R. M. Gray, J. C. Kieffer, and Y. Linder, *Locally optimal block quantizer design*, Information and Control, 45 (1980) 178-198.
- [GN] R. Gray and D. Neuhoff, *Quantization*, IEEE Trans. Inform. Theory, 44 (1998) 2325-2383.

- [GL2] A. György and T. Linder, *On the structure of optimal entropy-constrained scalar quantizers*, IEEE transactions on information theory, 48, no. 2 (February 2002) 416-427.
- [HK1] A. Haynes, M. Kelly, and H. Koivusalo, *Constructing bounded remainder sets and cut-and-project sets which are bounded distance to lattices*, II. Indagationes Math (N.S.) 28, no. 1 (2017) 138-144.
- [HK2] A. Haynes and H. Koivusalo, *Constructing bounded remainder sets and cut-and-project sets which are bounded distance to lattices*, Israel Journal Math. 212, no. 1 (2016) 189-201.
- [HS] E. Hewitt and L. Savage, *Symmetric measures on Cartesian products*, Transactions American Math Society 80 (1955) 470-501.
- [J] A. Junqueira, *Quantitative recurrence for generic homeomorphisms*, Proceedings of the AMS, 145 no. 11 (2017) 4751-4761.
- [KZ2] P. Kargaev and A. Zhigljavsky, *Asymptotic distribution of the distance function to the Farey points*, Journal of Number Theory 65 (1997) 130-149.
- [KZ1] P. Kargaev and A. Zhigljavsky, *Approximation of real numbers by rationals: some metric theorems*, Journal of Number Theory 61 (1996) 209-225.
- [HK] H. Kesten, *Some probabilistic theorems on Diophantine approximation*, Transactions of the AMS 103 (1962) 189-217.
- [AK] A. Khinchin, *Einige Sätze über Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen*, Math Annalen 92 (1924) 115-125.
- [K2] D. H. Kim, *Refined shrinking target property for rotations*, Nonlinearity 27 (2014) 1985-1997.
- [K1] D. H. Kim, *The shrinking target property of irrational rotations*, Nonlinearity 20 (2007) 1637-1643.
- [KN] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley and Sons, New York-London-Sydney, 1974.
- [Landau] E. Landau, *Bemerkungen zu der vorstehenden Abhandlung von Herrn Franel*, Göttinger Nachrichten (1924) 202-206.
- [L] P. Levy, *Sur la division d'un segment par des points choisis au hasard*, Comptes Rendus (1939) 147-149.
- [LNT] V. Losert, W-G. Nowak, and R. F. Tichy, *On the asymptotic distribution of the powers of  $s \times s$ -matrices*, Compositio Math 45, no 2 (1982) 273-291.
- [MSS] F. Motta, P. Shipman, and B. Springer, *Optimally topologically transitive orbits in discrete dynamical systems*, American Math Monthly 123, no. 2 (2016) 115-135.
- [N2] H. Niederreiter, *On a measure of denseness for sequences*, Topics in Classical Number Theory, Vol. I, II (Budapest, 1981), 1163-1208, Colloquium Math Soc. János Bolyai, 34, North-Holland, Amsterdam, 1984.
- [N1] H. Niederreiter, *The distribution of Farey points*, Math Annalen 201 (1973) 341-345.
- [NT] W-G. Nowak and R. F. Tichy, *An improved estimate on the distribution mod 1 of powers of real matrices*, Compositio Math 49, no. 2 (1983) 283-289.
- [P] G. Pagès, *A space quantization method for numerical integration*, Journal of Computational and Applied Math 89 (1997) 1-38.
- [W1] W. Philipp, *Empirical distribution functions and uniform distribution mod 1*, in *Diophantine Approximation and its Applications*, pp. 211-234, Academic Press, New York, 1973.
- [W2] W. Philipp, *Limit theorems for lacunary series and uniform distribution mod 1*, Acta Arithmetica 26 (1975) 241-251.
- [PSZ] G. Polanco, D. Schultz, and A. Zaharescu, *Continuous distributions arising from the Three Gap Theorem*, International Journal of Number Theory, 12 no. 7 (2016) 1743-1764.
- [RR1] J. Rosenblatt and M. K. Roychowdhury, *Optimal quantization for piecewise uniform distributions*, Uniform Distribution Theory 13 (2018), no. 2, 23-55.
- [RR2] J. Rosenblatt and M. K. Roychowdhury, *Shrinking Targets in Dynamical Systems*, Journal of Mathematical Analysis and Applications, Volume 485, Issue 1, 1 May 2020, 123769
- [R1] M. K. Roychowdhury, *Quantization and centroidal Voronoi tessellations for probability measures on dyadic Cantor sets*, Journal of Fractal Geometry, 4 (2017), 127-146.
- [R2] M. K. Roychowdhury, *Optimal quantizers for some absolutely continuous probability measures*, Real Analysis Exchange, Vol. 43(1), 2017, pp. 105-136.
- [R3] M. K. Roychowdhury, *Optimal quantization for the Cantor distribution generated by infinite similitudes*, Israel Journal of Mathematics 231 (2019), 437-466.

- [R4] M. K. Roychowdhury, *Least upper bound of the exact formula for optimal quantization of some uniform Cantor distributions*, Discrete and Continuous Dynamical Systems- Series A, Volume 38, Number 9, September 2018, pp. 4555-4570.
- [Sch] J. Schoissengeier, *On the longest gaps in the sequence  $(n\alpha) \bmod 1$* , Österreichisch-Ungarisch-Slowakisches Kolloquium über Zahlentheorie (Mariatrost, 1992), 155-166, Grazer Math Ber., 318, Karl-Franzens-Universität Graz, Graz, 1993.
- [Sh] L. Shepp, *Covering the circle with random arcs*, Israel Journal Math 11 (1972) 328345.
- [VR] T. Van Ravenstein, *The Three Gap Theorem (Steinhaus Conjecture)*, Journal Australian Math Society (Series A) 45 (1988) 360-370.
- [ZA] A. Zhigljavsky and I. Aliev, *Weyl sequences: asymptotic distributions of the partition lengths*, Acta Arithmetica 88 no. 4 (1999) 351-361.

DEPARTMENT OF MATHEMATICAL SCIENCES, INDIANA UNIVERSITY-PURDUE UNIVERSITY INDIANAPOLIS,  
402 N. BLACKFORD STREET, INDIANAPOLIS, IN 46202-3217, USA.

*E-mail address:* rosnbltt@illinois.edu

SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES, UNIVERSITY OF TEXAS RIO GRANDE VALLEY,  
1201 WEST UNIVERSITY DRIVE, EDINBURG, TX 78539-2999, USA.

*E-mail address:* mrinal.roychowdhury@utrgv.edu