①

Let A be a finite set in $\mathbb{R}^d$, $|A| = n$.
$d \geq 2$; size of A

For convenience, assume that $A \subset [0,1]^d$,
the unit cube

We shall assume throughout that
$n >$ the amount of the national debt $\cdot d$.

The purpose of this lecture is to study the "dimension" of A using a quantity that comes from geometric measure theory:

$$I_s(A) = n^{-2} \sum_{a \neq a'} |a-a'|^{-s}$$

discrete energy of order $-s$

$0 < s < d$

Let us begin by realizing that we have seen the underlying idea before.

Let $B_d = \underbrace{\{ x \in \mathbb{R}^d : |x| \le 1 \}}_{\text{the unit ball}}$

$x = (x_1, x_2, \dots, x_d)$

$y = (y_1, y_2, \dots, y_d)$

$$\text{Then } \int_{B_d}\int_{B_d} |x-y|^{-s} \, dx \, dy < \infty$$

$$\text{iff} \quad s < d$$

But what if instead of the unit ball we take a lower dimensional object?

Simple example: $d = 2$

$$X = \{ (t, 0) : 0 \le t \le 1 \}$$

What does it mean to integrate over X?

We know how to do this from multi-variable calculus:

We get

$$\int_0^1\int_0^1 |t-t'|^{-s} \, dt\, dt' < \infty$$

iff $s<1$,

so even though $X = \{(t,0) : t \in [0,1]\}$, $\subseteq \mathbb{R}^2$

the exponent restriction is 1-dimensional, not 2-dimensional because X is 1-dimensional!

Let's consider a less transparent example:

$d=2$ $\quad X = \{(\cos\theta, \sin\theta) : \theta \in [0, 2\pi)\}$

The corresponding "energy" quantity is:

$$\iint |x(\theta) - y(\varphi)|^{-s} \, d\theta\, d\varphi$$

$x(\theta) = (\cos\theta, \sin\theta)$ $\quad y(\varphi) = (\cos\varphi, \sin\varphi)$

(4)

We get

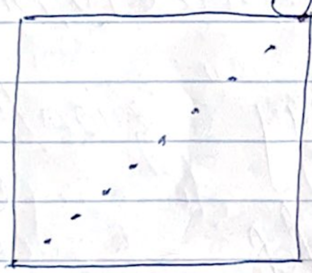
$$\iint \left|2(1-\cos(\theta-\varphi))\right|^{-\frac{s}{2}} d\theta d\varphi$$

$\cos(\theta-\varphi) \sim 1 - \frac{(\theta-\varphi)^2}{2}$, so the integral above is

$$\approx \iint |\theta-\varphi|^{-s} d\theta d\varphi < \infty$$

iff $s<1$, just like the case of the line segment.

With these calculations behind us, let's specialize to the discrete setting.



$$A = \left\{\left(\frac{i}{n}, \frac{i}{n}\right) : 1 \le i \le n\right\}$$

(5)

We have

$$n^{-2}\sum_{i\neq j} dist\left(\left(\frac{i}{n},\frac{i}{n}\right),\left(\frac{j}{n},\frac{j}{n}\right)\right)^{-s}$$

$$\approx n^{-2}\sum_{i\neq j}|i-j|^{-s}\cdot n^{s}\ ; \quad \begin{matrix} 1\leq i\leq n \\ 1\leq j\leq n \end{matrix}$$

Let $u=i-j$, $v=j$

We get

$$n^{-2}\cdot n^{s}\cdot n\sum_{\substack{-n \\ u\neq 0}}^{n}|u|^{-s}$$

$$\approx n^{-1+s}\sum_{n=1}^{n}u^{-s} \sim n^{-s+1}\cdot n^{s-1} \text{ if } s<1$$

$$\text{and } \sim n^{-1+s} \text{ if } s\geq 1.$$

It follows that $I_s(A)\leq C$ independently of n

iff $s\leq 1$.

What happens if $s=1$?

⑥

Let us formalize things a bit.

Definition: Let $\{A_n\}$ be a family of subsets of $[0,1]^d$, $d \geq 2$, such that $|A_n| = \gamma(n)$, increasing integer valued function of n

\# of elements

and $A_n \subset A_{n+1}$ $\forall n$.

We say that $\{A_n\}$ is s-adaptable if

$$I_s(A_n) \leq C$$

independently of $\underline{n}$.

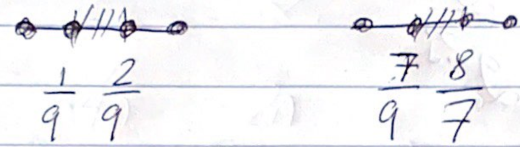
$1 \leq k \leq d$

<u>Exercise:</u> Suppose that $A_n \subset$ k-dimensional manifold. Then $\{A_n\}$ <u>is</u> <u>not</u> s-adaptable if $s > k$.

But there are stranger objects in the world than manifolds!

The classical triadic Cantor set is constructed as follows:

C_2



C_3

and so on!

Let $A_n = C_n \times C_n$ $\quad |A_n| = 2^{n+1} \cdot 2^{n+1} = 2^{2(n+1)}$.

We have

$$2^{-4(n+1)} \sum_{a \neq a'} |a-a'|^{-s}$$

$$\approx 2^{-4n} \sum_{j=0}^{n} \sum_{|a-a'| \approx 3^{-j}} 3^{js}$$

since it makes no sense to consider $|a-a'| < 3^{-n}$!!

(8)

Key question: If $a' \in \mathcal{C}_n \times \mathcal{C}_n$, how many a's are there such that $|a-a'| \approx 3^{-j}$, $0 \leq j \leq n$?

The answer is $\approx \dfrac{2^{2n}}{2^{2j}}$ $2^{2n} \sim$ total size of A_n

It follows that

$$2^{-4n} \sum_{j=0}^{n} 3^{js} \sum_{|a-a'| \approx 3^{-j}} 1$$

$$\approx \cancel{2^{-4n}} \cdot \cancel{2^{2n}} \cdot \cancel{2^{2n}} \cdot \sum_{j=0}^{n} 3^{js} \, 2^{-2j}$$

and the sum converges if

$$s < 2 \cdot \frac{\log(2)}{\log(3)} = s_{critical}$$

This critical value for s gives us a notion of dimension.

Let's amplify some of the technical points that arose above. We used the fact that

$$\sum_{u=1}^{n} u^{-s} \approx n^{-s+1} \quad \text{if} \quad s<1$$

w/ constants independent of n

There are many ways to see this, but here is one that I find particularly instructive.

$$\sum_{u=1}^{n} u^{-s} = \sum_{k=0}^{\log_2(n)} \sum_{u=2^k}^{2^{k+1}-1} u^{-s}$$

small amount of cheating here - please fix it!

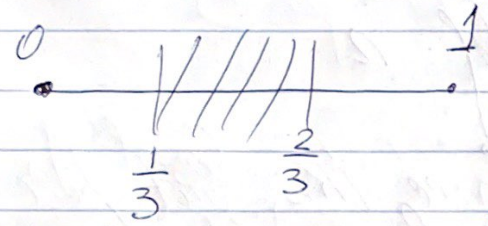
$$\approx \sum_{k=0}^{\log_2(n)} 2^{-ks} \cdot 2^{k}$$

$\hookrightarrow$ converges if $s>1$

yields $\approx \log(n)$ if $s=1$

& n^{-s+1} if $s<1$.

Let's take a more careful look at the Cantor set introduced above.



At the n'th stage, we have 2^n intervals of length 3^{-n}, so the total length of these intervals $\longrightarrow 0$ as $n \longrightarrow \infty$ . . . !

There is a notion of dimension, called the Hausdorff dimension, which can be expressed in this context as the largest s $\ni$

$$\iint |x-y|^{-s} \, d\mu(x) \, d\mu(y)$$

"natural measure on the Cantor set"

Going back to families of discrete sets, we are left w/ some fascinating questions:

Q: Can you construct a family of sets $\{A_n\}$ w/ $A_{n+1} \supset A_n$, so that

$$I_s(A_n) = n^{-2} \sum_{a \neq a'} |a-a'|^{-s}$$

is bounded independently of n

iff $s < s_0$ given any pre-assigned $s_0 \in (0, d)$?

Q: Is it possible to use $I_s(A_n)$, in conjunction with other tools to study clustering and dimension of large point sets?