36710 - 36752

ADVANCED PROBABILITY OVERVIEW

FALL 2020

LECTURE 28: WED, DEC 9, 2020

CONCENTRATION OF MEASURE

- HIGH-DIMENSIONAL PROBABILITY, BY R. VERSHYNIN
- THE CONCENTRATION OF MEASURE PHENOMENON BY M. LEDOUX
- ELEMENTARY INTRODUCTION TO MODERN CONVEX GEOMETRY, BY K. BALL

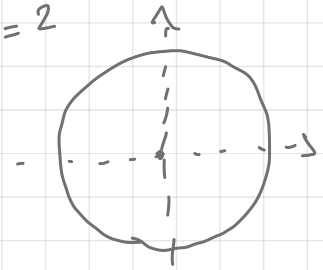
SURPRISING FACT THAT IN HIGH-DIMENSIONS MEASURES TEND TO CONCENTRATE AROUND CERTAIN PART OF THE SPACE

<u>EXAMPLE</u> LET $X \in \mathbb{R}^d$ BE UNIFORMLY DISTRIBUTED OVER THE UNIT EUCLIDEAN BALL $B(0,1) = \{ x \in \mathbb{R}^d : \|x\|^2 \leq 1 \}$

$x = (x_1, \dots, x_d)$, $\|x\|^2 = \sum_{i=1}^d x_i^2$

WHEN $d=1$, $B(0,1) = [-1, 1]$

$d = 2$



CONSIDER THE QUANTITY $\|X\| = \sqrt{\sum_{i=1}^d X_i^2}$

↳ DISTANCE OF X FROM ORIGIN

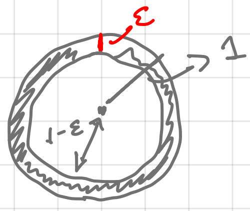
WHEN $d=1$, $\mathbb{E}[\|X\|] = \mathbb{E}[|X|] = 1/2$

BUT FOR GENERAL d, $\mathbb{E}[\|X\|] = \frac{d}{d+1} \to 1$ AS $d \to \infty$

= "MOST OF THE MASS" OF THIS DISTRIBUTION IS NEAR THE BOUNDARY OF THE BALL!!

TO SEE THIS, LET $\varepsilon \in (0,1)$ FIXED (SMALL)

$$\mathbb{P}(1-\varepsilon \le \|X\| \le 1) = ?$$



TO COMPUTE THIS, USE THE FACT THAT THE VOLUME (LEBESGUE MEASURE) OF $B(0,r)$ IN $\mathbb{R}^d$ ($r>0$) IS $\boxed{r^d v_d}$

$\hookrightarrow$ VOLUME OF $B(0,1) = \dfrac{\pi^{d/2}}{\Gamma(d/2+1)} \approx \left(\dfrac{2\pi e}{d}\right)^{d/2}$

$$\Gamma(x) = \int_0^\infty e^{-z} z^{x-1}\,dz \qquad x>0$$

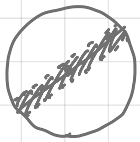
ANY EUCLIDEAN BALL OF VOLUME 1 HAS RADIUS OF ORDER $\sqrt{\dfrac{d}{2\pi e}}$ IN $\mathbb{R}^d$ (AS $d\to\infty$)

$$\mathbb{P}(1-\varepsilon \le \|X\| \le 1) = 1 - \frac{(1-\varepsilon)^d \cancel{v_d}}{\cancel{v_d}}$$

$$= 1 - (1-\varepsilon)^d$$

$$\ge 1 - e^{-\varepsilon d} \qquad \text{SINCE } 1-x \le e^{-x}$$

- ANOTHER SURPRISING RESULT: MOST OF THE MASS CONCENTRATES AROUND ANY $(d-1)$-DIMENSIONAL SUBSPACE



CONSIDER THE SLAB $= \left\{ x \in B(0,1) : |x_1| < \dfrac{c}{\sqrt{d}} \right\}$

SET OF POINTS IN $B(0,1)$ $\dfrac{c}{\sqrt{d}}$ - CLOSED TO SUBSPACE $\{x : x_1 = 0\}$

AS $d \to \infty$

$$\mathbb{P}(X \in \text{SLAB}) \to \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-x^2/2}\,dx$$

$$= P(Z \in (-c,c))$$

$$Z \sim N(0,1)$$

- CONSIDER THE UNIT CUBE IN $\mathbb{R}^d$: $[0,1]^d$. IN YOUR HW YOU SHOWED THAT MOST OF THE MASS CONCENTRATES IN SUCH A WAY THAT

$$\sqrt{\frac{d}{3}}(1-\varepsilon) \le \|X\| \le \sqrt{\frac{d}{3}}(1+\varepsilon) \qquad \text{FOR ANY } \varepsilon>0 \text{ AS } d\to\infty$$

$$X \sim \text{UNIFORM}([0,1]^d)$$

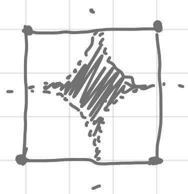
PUT A EUCLIDEAN BALL OF RADIUS $1/2$ AT EACH OF THE 2^d

CORNERS OF $[0,1]^d$. REMOVE THE CORRESPONDING PORTIONS.

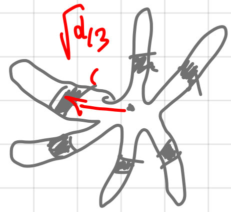
UNIT VOLUME BALL FOR NORM $\|\cdot\|_\infty$ $[\|x\|_\infty = \max |x_i|]$

HOW MUCH MASS IS LEFT? OVERALL WE REMOVE A MASS OF

$$\left(\frac{1}{2}\right)^d V_d \sim \left(\frac{1}{2}\right)^d \left(\frac{2\pi e}{d}\right)^{d/2} \to 0$$



SO IN HIGH d, THE SET $[0,1]^d$ LOOKS LIKE



$\to 2^d$ "TENTACLES"

CONCENTRATION UNIFORM DISTRIBUTION OVER SPHERE

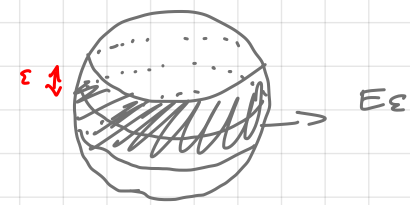
$$S^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$$

CONSIDER AN EQUATOR: $\{x \in S^{d-1} : x_1 = 0\}$

LET $E_\varepsilon = \{x \in S^{d-1} : |x_1| \leq \varepsilon\}$

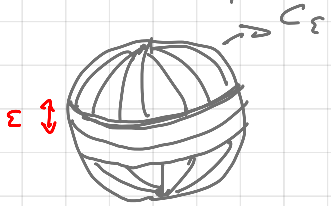
THEN $\mathbb{P}(X \in E_\varepsilon) \geq 1 - 2\exp\left\{-\frac{d\varepsilon^2}{2}\right\}$

UNIFORM DRAW FROM S^{d-1}



$\hookrightarrow$ MOST OF THE MASS CONCENTRATES NEAR ANY EQUATOR!

Pf/

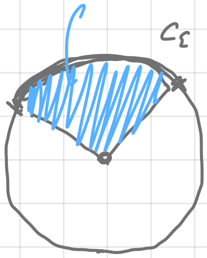


C_ε IS THE COMPLEMENT (IN S^{d-1}) OF E_ε IN, SAY, NORTHERN HEMISPHERE. WE WILL SHOW THAT $P(X \in C_\varepsilon) \leq \exp\left\{-\frac{d}{2}\varepsilon^2\right\}$

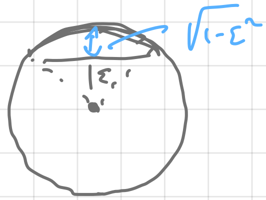
$$\mathbb{P}(X \in C_\varepsilon) = \frac{\mathrm{Vol}(\mathrm{Cone}(C_\varepsilon))}{\mathrm{Vol}(B(0,1))}$$

WHERE Vol IS THE LEBESGUE MEASURE IN $\mathbb{R}^d$ AND $\mathrm{Cone}(C_\varepsilon)$ IS

$B(0,1)$



NOTICE THAT $\mathrm{Vol}(\mathrm{Cone}(C_\varepsilon)) \leq \mathrm{Vol}(B(x', \sqrt{1-\varepsilon^2}))$



$\varepsilon < 1/2$

WHERE $x' = (\varepsilon, 0, \dots, 0)$

SO $\mathbb{P}(X \in C_\varepsilon) \leq \frac{(\sqrt{1-\varepsilon^2})^d \cancel{\sqrt{d}}}{\cancel{\sqrt{d}}} \leq \exp\{-\frac{d}{2}\varepsilon^2\}$

USING $1-x \leq e^{-x}$

■

■ CONCENTRATION OF GAUSSIAN MEASURE

LET $Z \sim N_d(0, I_d)$

$\hookrightarrow$ IDENTITY MATRIX

IF $X \sim N(\mu, \sigma^2)$ THEN ONE CAN SHOW THAT

$$\mathbb{P}(|X-\mu| \geq \varepsilon) \leq 2\bar{\Phi}(\varepsilon/\sigma) \quad \varepsilon > 0$$

$\hookrightarrow$ 1 - cdf OF $N(0,1)$ AT ε/σ

$\left[\bar{\Phi}(x) = \mathbb{P}(Y \geq x) \quad Y \sim N(0,1)\right]$

$$\leq 2\exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\}$$

PROPERTY OF GAUSSIANS WHEN $d = 1$!

LET $f: \mathbb{R}^d \to \mathbb{R}$ BE L - LIPSCHITZ $\left[|f(x) - f(y)| \leq L\|x-y\|\right]$

THEN

$$\mathbb{P}(|f(\underbrace{Z}_{N_d(0,\sigma^2 I_d)}) - \mathbb{E}[f(Z)]| \geq \varepsilon) \leq 2\bar{\Phi}\left(\frac{\varepsilon}{L\sigma}\right)$$

- LIPSCHITZ FUNCTIONS OF Z CONCENTRATE AROUND THEIR EXPECTED VALUE !
- THE RATE OF CONCENTRATION IS INDEPENDENT OF d !

  $\hookrightarrow$ THIS APPLIES TO GAUSSIAN STOCHASTIC PROCESSES OVER WELL BEHAVED (SEPARABLE) SPACES

- ORIGINALLY THE RESULT WAS PROVED NOT FOR CONCENTRATION AROUND $\mathbb{E}[f(Z)]$ BUT FOR MEDIAN OF $f(Z)$ [A MEDIAN m_f OF $f(Z)$ IS A NUMBER S.T. $\mathbb{P}(f(Z) \geq m_f) \geq 1/2$

  $\mathbb{P}(f(Z) \leq m_f) \geq 1/2$

THE "EASY" PROOF OF CONCENTRATION OF $f(Z)$ AROUND ITS MEDIAN

USES THIS DEEP RESULT ABOUT ISOPERIMETRIC OR EXTREMAL SETS:

LET H BE AN HALF-SPACE $\left[H = \{ x \in \mathbb{R}^d : x^T a \leq b \} \right]$ SOME $a \in \mathbb{R}^d$, $b \in \mathbb{R}$

LET A BE ANY BOREL SET IN $\mathbb{R}^d$ AND LET, FOR $\varepsilon > 0$,

$$A^\varepsilon = \{ x : \underset{\inf_{y \in A} \|x - y\|}{\underset{\parallel}{d(x, A)}} < \varepsilon \}$$

IF $P(A) = P(H)$ WHERE P HERE IS STANDARD GAUSSIAN DISTRIBUTION IN $\mathbb{R}^d$

THEN, FOR ALL $\varepsilon > 0$,

$$P(A^\varepsilon) \geq P(H^\varepsilon)$$