CS685/785 Foundation of Data Science

Lecture 3: High-Dimensional Space

Xi Li

Fall 2024

Table of Content

Table of Content

Review of Probability Theory and Inequalities

- Expectation and Variance of Random Variables
- Markov’s inequality
- Chebyshev’s inequality

Expectation

- $E[X]$: Mean, expected value, or expectation of a random variable X.
- If X is a continuous random variable with pdf $p(x)$:

$$E[x] = \int_{-\infty}^{+\infty} xp(x)\mathrm{d}x$$

- If X is a discrete random variable with probability $P(x)$:

$$E[x] = \sum_{x} xP(X = x)$$

Properties of Expectation

- For a random variable X and constants a, b

$$E[aX + b] = aE[X] + b$$

- Let X and Y be random variables

$$E[\mathrm{X} + Y] = E[X] + E[Y]$$

- More generally

$$E\left[\sum_i \mathrm{X}_i\right] = \sum_{\mathrm{i}} E[X_i]$$

- Let X and Y be **independent** random variables

$$E[\mathrm{X}Y] = E[X]E[Y]$$

Variance

- $Var[X]$: The variance of a random variable X. It measures how spread out it is.
- Definition:

$$\begin{aligned}\mathrm{Var}[x] &= E[(X - E[X])^2] \\ &= E[X^2] - (E[X])^2\end{aligned}$$

Properties of Variance

- For a random variable X and constants a, b

$$Var[aX + b] = a^2 Var[X]$$

- Let X and Y be **independent** random variables

$$Var[X + Y] = Var[X] + Var[Y]$$

- For more than 2 **independent** random variables

$$Var\left[\sum_i X_i\right] = \sum_i Var[X_i]$$

Markov's Inequality

Theorem 3.1 (Markov's inequality)

$Let\ x\ be\ a\ \boldsymbol{non-negative}\ random\ variable.$

$Then\ for\ a > 0,$

$$Prob(x \geq a) \leq \frac{E(x)}{a}.$$

Markov's Inequality

Theorem 3.1 (Markov's inequality)

$Let\ x\ be\ a\ nonnegative\ random\ variable.$

$Then\ for\ a > 0,$

$$Prob(x \geq a) \leq \frac{E(x)}{a}.$$

Proof:

Markov's Inequality

Theorem 3.1 (Markov's inequality)

$Let\ x\ be\ a\ nonnegative\ random\ variable.$

$Then\ for\ a > 0,$

$$Prob(x \geq a) \leq \frac{E(x)}{a}.$$

Proof: For a continuous non-negative random variable x with probability density function $p(x)$,

$$E(x) = \int_0^{\infty} xp(x)dx = \int_0^{a} xp(x)dx + \int_a^{\infty} xp(x)dx$$

Def. $E[x] = \int_{-\infty}^{+\infty} xp(x)\mathrm{d}x$

Markov's Inequality

Theorem 3.1 (Markov's inequality)

$Let\ x\ be\ a\ nonnegative\ random\ variable.$

$Then\ for\ a > 0,$

$$Prob(x \geq a) \leq \frac{E(x)}{a}.$$

Proof: For a continuous non-negative random variable x with probability density function $p(x)$,

$$E(x) = \int_0^\infty xp(x)dx = \int_0^a xp(x)dx + \int_a^\infty xp(x)dx$$

$$\geq \int_a^\infty xp(x)dx$$

$$x \geq a$$

Markov's Inequality

Theorem 3.1 (Markov's inequality)
Let x be a nonnegative random variable.
Then for $a > 0$,

$$Prob(x \geq a) \leq \frac{E(x)}{a}.$$

Proof: For a continuous non-negative random variable x with probability density function $p(x)$,

$$E(x) = \int_0^\infty xp(x)dx = \int_0^a xp(x)dx + \int_a^\infty xp(x)dx$$

$$\geq \int_a^\infty xp(x)dx \geq a\int_a^\infty p(x)dx = aProb(x \geq a)$$

Markov's Inequality

Theorem 3.1 (Markov's inequality)
$Let\ x\ be\ a\ \boldsymbol{non-negative}\ random\ variable.$
$Then\ for\ a > 0,$

$$Prob(x \geq a) \leq \frac{E(x)}{a}.$$

Corollary 3.2 $Prob(x \geq bE(x)) \leq \frac{1}{b}.$

Chebyshev's Inequality

Theorem 3.3 (Chebyshev's inequality)

Let x be a random variable.

Then for $c > 0$,

$$Prob(|x - E(x)| \geq c) \leq \frac{Var(x)}{c^2}.$$

Chebyshev's Inequality

Theorem 3.3 (Chebyshev's inequality)

$Let\ x\ be\ a\ random\ variable.$

$Then\ for\ c > 0,$

$$Prob(|x - E(x)| \geq c) \leq \frac{Var(x)}{c^2}.$$

Proof:

$$Prob(|x - E(x)| \geq c) = Prob(|x - E(x)|^2 \geq c^2)$$

Chebyshev's Inequality

Theorem 3.3 (Chebyshev's inequality)
$Let\ x\ be\ a\ random\ variable.$
$Then\ for\ c > 0,$

$$Prob(|x - E(x)| \geq c) \leq \frac{Var(x)}{c^2}.$$

Proof:

$$Prob(|x - E(x)| \geq c) = Prob(|x - E(x)|^2 \geq c^2)$$

Let $y = |x - E(x)|^2$. Note that y is a non-negative random variable and $E(y) = Var(x)$.

Def. Var$[x] = E[(X - E[X])^2]$

Chebyshev's Inequality

Theorem 3.3 (Chebyshev's inequality)

$Let\ x\ be\ a\ random\ variable.$

$Then\ for\ c > 0,$

$$Prob(|x - E(x)| \geq c) \leq \frac{Var(x)}{c^2}.$$

Proof:

$$Prob(|x - E(x)| \geq c) = Prob(|x - E(x)|^2 \geq c^2)$$

Let $y = |x - E(x)|^2$. Note that y is a non-negative random variable and $E(y) = Var(x)$.

$$Prob(|x - E(x)| \geq c) = Prob(y \geq c^2)$$

Chebyshev's Inequality

Theorem 3.3 (Chebyshev's inequality)
Let x be a random variable.
Then for $c > 0$,

$$Prob(|x - E(x)| \geq c) \leq \frac{Var(x)}{c^2}.$$

Proof:

$$Prob(|x - E(x)| \geq c) = Prob(|x - E(x)|^2 \geq c^2)$$

Let $y = |x - E(x)|^2$. Note that y is a non-negative random variable and $E(y) = Var(x)$.

$$Prob(|x - E(x)| \geq c) = \underbrace{Prob(y \geq c^2) \leq \frac{E(y)}{c^2}}_{\text{Markov's inequality}} = \frac{Var(x)}{c^2}$$

Law of Large Numbers (LLN)

Theorem 3.4 (Law of Large Numbers)

Let $x_1, x_2, \ldots, x_n$ *be* n ***independent samples*** *of a random variable* X*.*
Then

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(X)\right| \geq \epsilon\right) \leq \frac{Var(X)}{n\epsilon^2}$$

LLN -- An Intuitive Explanation

- $x_1, x_2, \ldots, x_n$: n independent samples of variable x
- $E(x)$: expected value of x (population mean)
- $\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n}$: sample mean

LLN states that $\bar{x} \rightarrow E(x), as\ n \rightarrow \infty.$

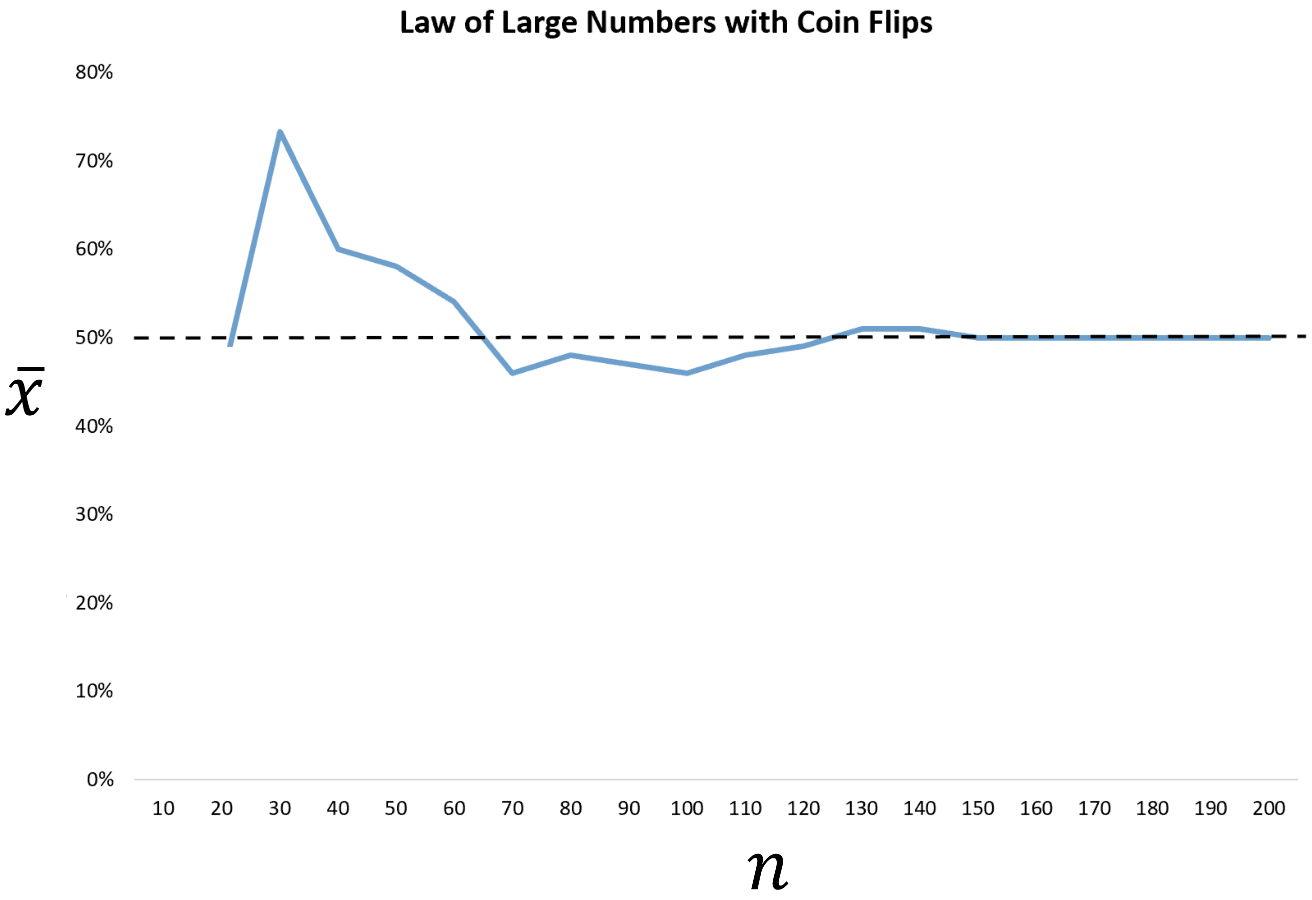
$$Prob\left(\left|\underbrace{\frac{x_1 + x_2 + \cdots + x_n}{n}}_{\text{Sample mean}} - E(X)\right| \geq \epsilon\right) \leq \frac{Var(X)}{n\epsilon^2}$$

LLN -- An Intuitive Explanation

- $X = \#$ heads after 100 tosses of a fair coin
- $E(\mathrm{X}) = 100 * 0.5 = 50$
- Trial 1: $x_1 = 55$
- Trial 2: $x_2 = 65$
- …
- $\bar{x} = \frac{55+65+45+\cdots+x_n}{n}$

LLN states that $\bar{x} \to 50$,

$as\ n \to \infty$.



LLN

We could make the following observations of LLN:

- The larger the variance $Var(x)$, the greater the probability that the error will exceed ϵ.
- The more the samples (the larger the n), the smaller the probability that the difference will exceed ϵ.
- The larger the ϵ (error tolerance), the smaller the difference will exceed ϵ.

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var(x)}{n\epsilon^2}$$

LLN -- Proof

Theorem 3.4 (Law of Large Numbers)

Let $x_1, x_2, \ldots, x_n$ *be* n ***independent samples*** *of a random variable* x.
Then

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var(x)}{n\epsilon^2}$$

Proof:

LLN -- Proof

Theorem 3.4 (Law of Large Numbers)

Let $x_1, x_2, \ldots, x_n$ *be* n ***independent samples*** *of a random variable* x*.*

Then

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var(x)}{n\epsilon^2}$$

Proof:

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var(\frac{x_1 + x_2 + \cdots + x_n}{n})}{\epsilon^2}$$

Chebyshev's inequality $Prob(|x - E(x)| \geq c) \leq \dfrac{Var(x)}{c^2}$

LLN -- Proof

Proof:

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)}{\epsilon^2}$$

$$= \frac{1}{n^2\epsilon^2} Var(x_1 + x_2 + \cdots + x_n)$$

$$Var[c\boldsymbol{X}] = c^2 Var[\boldsymbol{X}]$$

LLN -- Proof

Proof:

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)}{\epsilon^2}$$

$$= \frac{1}{n^2\epsilon^2} Var(x_1 + x_2 + \cdots + x_n)$$

$$Var\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} Var[X_i] \qquad = \frac{1}{n^2\epsilon^2}\sum_{i=1}^{n} Var(x_i)$$

LLN -- Proof

Proof:

$$Prob\left(\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - E(x)\right| \geq \epsilon\right) \leq \frac{Var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)}{\epsilon^2}$$

$$= \frac{1}{n^2\epsilon^2} Var(x_1 + x_2 + \cdots + x_n)$$

$$= \frac{1}{n^2\epsilon^2} \sum_{i=1}^{n} Var(x_i)$$

$$= \frac{1}{n^2\epsilon^2} nVar(x)$$

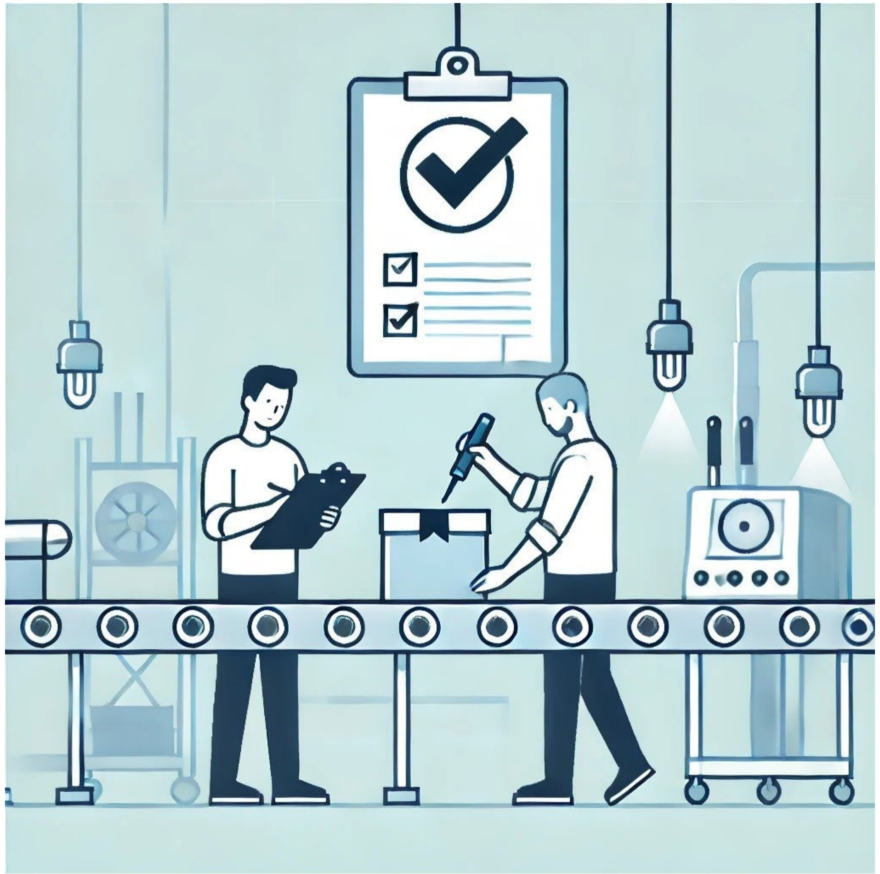
$$= \frac{Var(x)}{n\epsilon^2}$$

LLN Applications

- Manufacturing Quality Control
- Insurance Industry
- Gambling and Casinos

LLN in Manufacturing Quality Control

- It's time-consuming and costly to inspect every product.
- It's efficient and cheap to inspect a subset of products.
- Based on the LLN:
  - The sampling should be **random.**
  - **The more products** you test, **the more accurate** your estimation is.
- Consider a car manufacturer produces 10,000 vehicles per month. By LLN, It's reasonable to estimate the defect rate based on 100 randomly chosen vehicles.



LLN in Insurance Industry

- How does the insurance company know how much to charge people for coverage?
- Consider an insurance company having 100,000 auto policyholders
- Based on LLN:
    - Count the percentage of policyholder filing a claim: 5%
    - Estimate average cost of claim based on historical data: $10,000.
    - Predict total claim cost: $50,000,000.
    - Charge $1,000 per policyholder, profit = 100,000 * $1000 - $50,000,000 = $50,000,000

LLN in Gambling and Casinos

- Any individual game is unpredictable.
- However, in the long run, casinos are guaranteed to make money.
- For most games, the casino wins about 51-55% of the time.
- Based on the LLN:
    - A player might win big on occasion。
    - As more games are played, the average outcome converges to the expected value (profit).

Table of Content

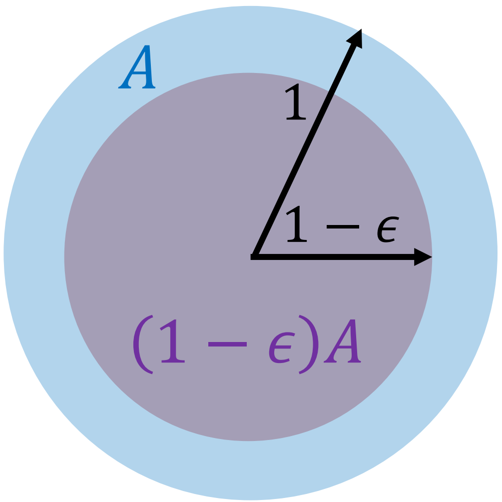
3.2 Geometry of High Dimensions

- Geometry behaves **counter-intuitively** in **high-dimensional** space!
- An important property -- **Almost all volume near the surface**:

  *Most of the **volume** of high-dimensional object is **near the surface**, rather than being uniformly distributed throughout the interior.*

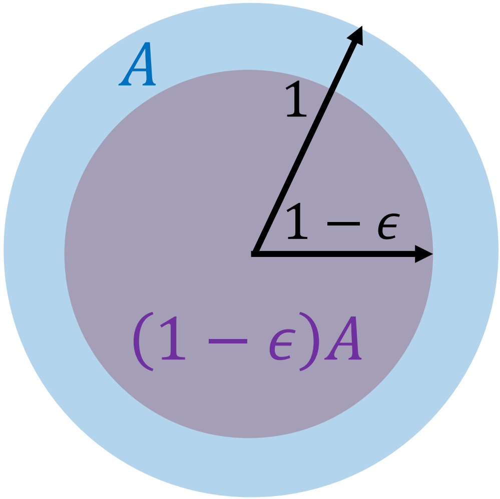
Almost all volume near the surface

- Consider an object $A \in R^d$
- Shrink A by ϵ: $(1-\epsilon)A = \{(1-\epsilon)x \mid x \in A\}$



Almost all volume near the surface

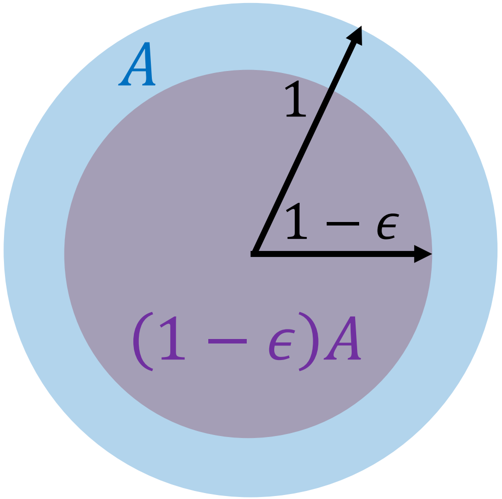
- Consider an object $A \in R^d$
- Shrink A by ϵ: $(1-\epsilon)A = \{(1-\epsilon)x \mid x \in A\}$
- Volume after shrinking: $Volume\big((1-\epsilon)A\big) = (1-\epsilon)^d Volume(A)$



Almost all volume near the surface

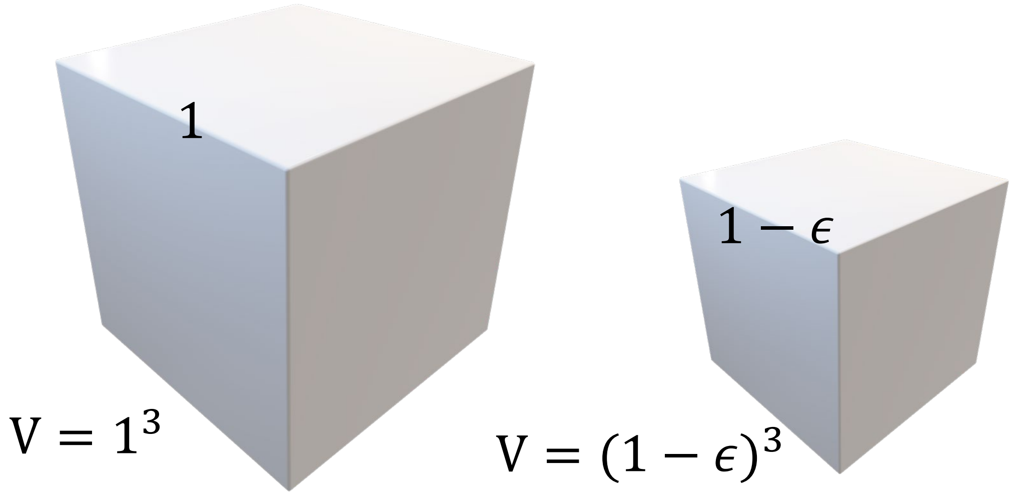
- Consider an object $A \in R^d$
- Shrink A by ϵ: $(1-\epsilon)A = \{(1-\epsilon)x \mid x \in A\}$
- Volume after shrinking: $Volume\big((1-\epsilon)A\big) = (1-\epsilon)^d Volume(A)$

Why this is true?



Almost all volume near the surface

- Consider an object $A \in R^d$
- Shrink A by ϵ: $(1-\epsilon)A = \{(1-\epsilon)x \mid x \in A\}$
- Volume after shrinking: $Volume\left((1-\epsilon)A\right) = (1-\epsilon)^d Volume(A)$
  - Consider A as a 3D cube, the volume of A shrinks by $(1-\epsilon)^3$
  - In d-dimensional space, partition A into infinitesimal cubes, and the volume of each cube shrink by $(1-\epsilon)^d$



Almost all volume near the surface

- Consider an object $A \in R^d$
- Shrink A by ϵ: $(1-\epsilon)A = \{(1-\epsilon)x \mid x \in A\}$
- Volume after shrinking:

$$\frac{Volume\big((1-\epsilon)A\big)}{Volume(A)} = (1-\epsilon)^d$$

Almost all volume near the surface

- Consider an object $A \in R^d$
- Shrink A by ϵ: $(1-\epsilon)A = \{(1-\epsilon)x \mid x \in A\}$
- Volume after shrinking:

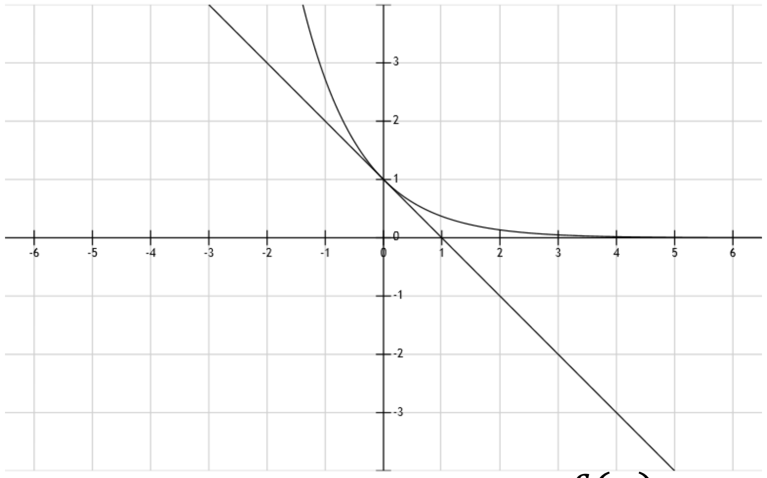
$$\frac{Volume\big((1-\epsilon)A\big)}{Volume(A)} = (1-\epsilon)^d$$

$$\leq e^{-\epsilon d}$$

$using\ the\ fact$

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

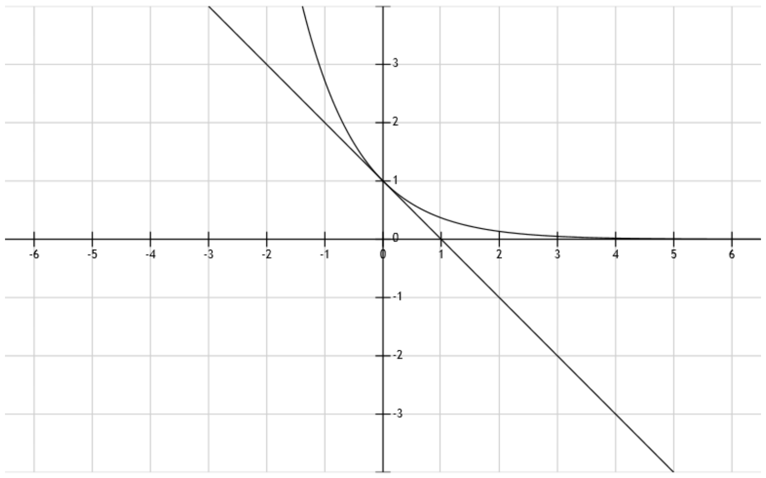
$f(x) = e^{-x}$



$f(x) = x$

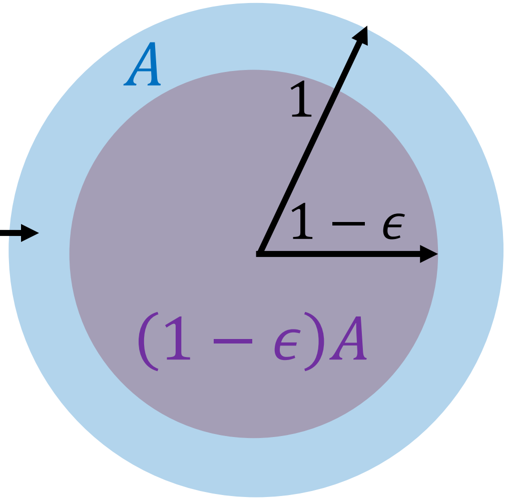
Almost all volume near the surface

$f(x) = e^{-x}$



$$\frac{Volume((1-\epsilon)A)}{Volume(A)} \le e^{-\epsilon d}$$

- $fix\ \epsilon,\ d \to \infty,\ the\ ratio \to 0$
- Most volume in the portion not belong to $(1-\epsilon)A$



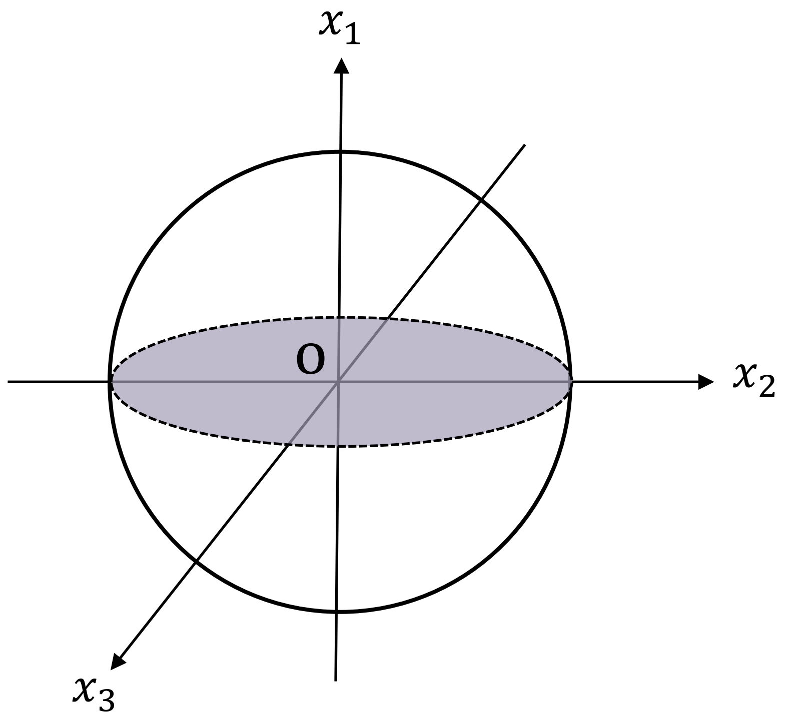
Almost all volume near the surface

- Ball at o $\in R^d$ with radius γ in d-dimensional space

$$B_\gamma(o) = \{x \in R^d \mid ||x - o|| < \gamma\}$$

- Unit ball: $\gamma = 1$ (consider o as origin for simplicity)

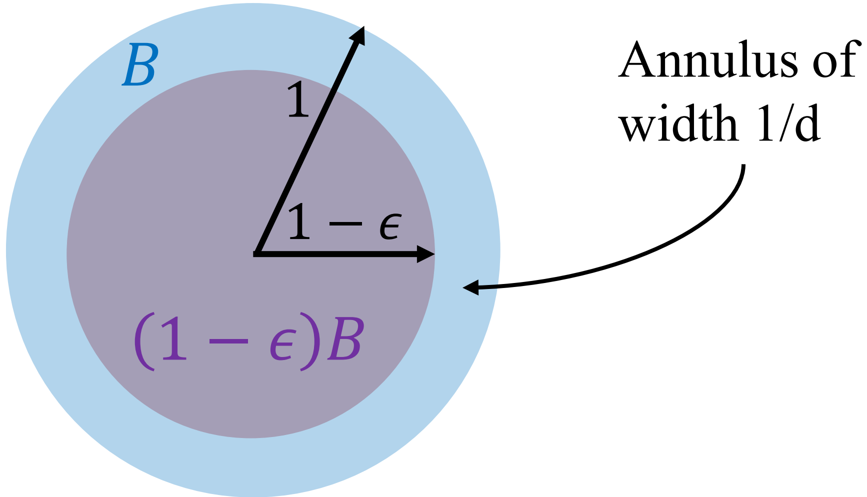
$$B = \{x \in R^d \mid ||x|| < 1\}$$



Almost all volume near the surface

- Let B denote the **unit ball** in d-dimensions: $B = \{x \in R^d \mid ||x|| < 1\}$
- $\frac{V((1-\epsilon)B)}{V(B)} \leq e^{-\epsilon d}$
- $1 - \frac{V((1-\epsilon)B)}{V(B)} \geq 1 - e^{-\epsilon d}$
- At least $1 - e^{-\epsilon d}$ points in $B \backslash (1-\epsilon)B$, (*i.e.*, an annulus of width ϵ)
- Particularly, $\epsilon = O(\frac{1}{d})$ for unit ball

  $\epsilon = O(\frac{\gamma}{d})$ for ball with radius γ



3.3 Properties of the Unit Ball

- Volume of the unit ball
- Volume near the equator
- Near Orthogonality

Volume of the unit ball

Volume of unit balls in d-dimensional space:

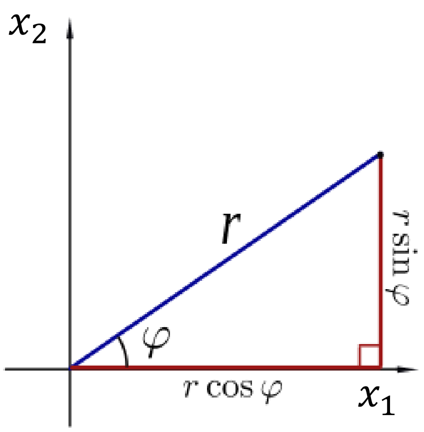
- $d = 1,\ V_1 = \int_{-1}^{1} 1 dx = 2$
- $d = 2,\ V_2 = \int_{x_1^2 + x_2^2 \leq 1} 1 dx_1 dx_2 = ?$

Volume of the unit ball

$V_2 = \int_{x_1^2+x_2^2\leq 1} 1 dx_1 dx_2 = ?$

Consider polar coordinates (γ, θ)

- $x_1 = \gamma \cos\theta, x_2 = \gamma \sin\theta$
- Jacobian Matrix $J = \begin{bmatrix} \frac{\partial x_1}{\partial \gamma} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial \gamma} & \frac{\partial x_2}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\gamma \sin\theta \\ \sin\theta & \gamma\cos\theta \end{bmatrix}$
- Scaling factor for coordinate system change: $\det J = \gamma \cos^2\theta + \gamma \sin^2\theta = \gamma$
- $dx_1 dx_2 \to \gamma d\gamma d\theta$
- $V_2 = \int_{\gamma=0}^{1} \int_{\theta=0}^{2\pi} \gamma d\gamma d\theta = \pi$

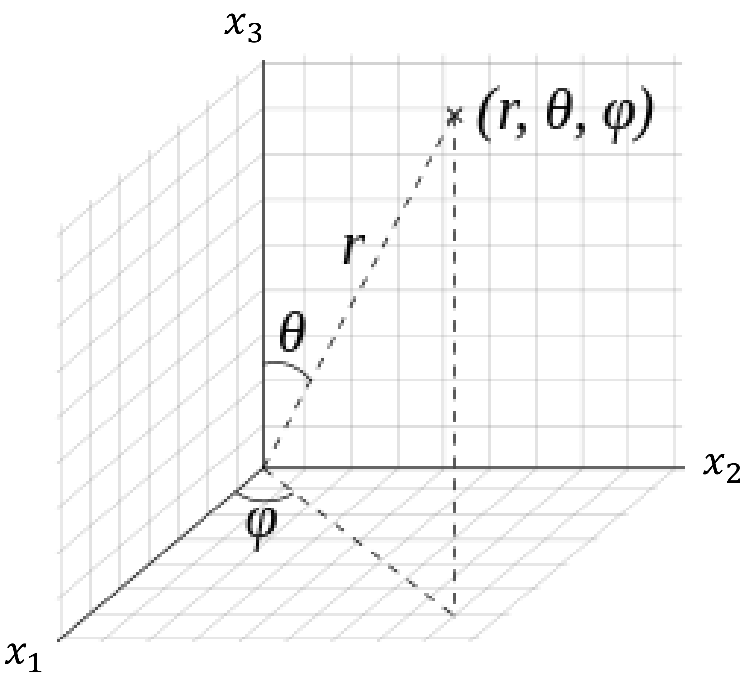


Volume of the unit ball

Volume of unit balls in d-dimensional space:

- $d = 1,\ V_1 = \int_{-1}^{1} 1 dx = 2$
- $d = 2,\ V_2 = \int_{x_1^2+x_2^2\le 1} 1 dx_1 dx_2 = \pi$
- $d = 3,\ V_3 = \int_{x_1^2+x_2^2+x_3^2\le 1} 1 dx_1 dx_2 dx_3 = \frac{4}{3}\pi$

$$x_1 = \gamma \sin\theta \cos\phi$$
$$x_2 = \gamma \sin\theta \sin\phi$$
$$x_3 = \gamma \cos\theta$$



Volume of the unit ball

Volume of unit balls in d-dimensional space:

- $d = 1,\ V_1 = \int_{-1}^{1} 1dx = 2$
- $d = 2,\ V_2 = \int_{x_1^2+x_2^2\le 1} 1dx_1dx_2 = \pi$
- $d = 3,\ V_3 = \int_{x_1^2+x_2^2+x_3^2\le 1} 1dx_1dx_2dx_3 = \frac{4}{3}\pi$
- …

Looks like the **volume V_d increases as d increases**, right?

Volume of the unit ball

Volume of unit balls in d-dimensional space:

- $d = 1,\ V_1 = \int_{-1}^{1} 1 dx = 2$
- $d = 2,\ V_2 = \int_{x_1^2 + x_2^2 \le 1} 1 dx_1 dx_2 = \pi$
- $d = 3,\ V_3 = \int_{x_1^2 + x_2^2 + x_3^2 \le 1} 1 dx_1 dx_2 dx_3 = \frac{4}{3}\pi$
- …

Looks like the volume V_d increases as d increases, right? **NOT TRUE!**

Volume of the unit ball

Volume of unit balls in d-dimensional space:

- $d = 1,\ V_1 = \int_{-1}^{1} 1dx = 2$
- $d = 2,\ V_2 = \int_{x_1^2+x_2^2\le 1} 1dx_1dx_2 = \pi$
- $d = 3,\ V_3 = \int_{x_1^2+x_2^2+x_3^2\le 1} 1dx_1dx_2dx_3 = \frac{4}{3}\pi$
- …

~~Looks like the volume V_d increases as d increases, right?~~

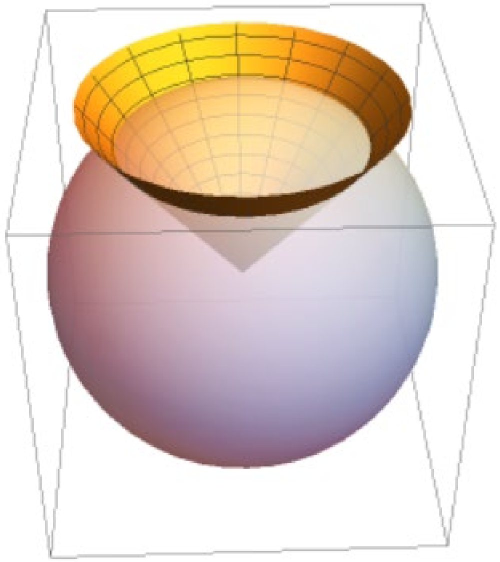
Actually, the volume $V_d \to 0$ as $d \to \infty$. **Counter-intuitive**

Volume of the unit ball

Closed form formulate of V_d:

- $V_d = \int_{x_1^2+x_2^2+\cdots+x_d^2\leq 1} 1 dx_1 dx_2 \ldots dx_d = \int_{S^d}\int_{\gamma=0}^{1} \gamma^{d-1} d\gamma d\Omega,$

  where S^d is the entire surface of a unit sphere, Ω is the solid angle (angular component of the volume integral)
- Visualization of solid angle in 3D space.



Volume of the unit ball

Closed form formulate of V_d:

- S^d: entire surface of a unit sphere; Ω: the solid angle
- $V_d = \int_{x_1^2+x_2^2+\cdots+x_d^2\leq 1} 1 dx_1 dx_2 \ldots dx_d = \int_{S^d}\int_{\gamma=0}^{1} \gamma^{d-1} d\gamma d\Omega$

$$= \frac{1}{d}\int_{S^d} d\Omega = \frac{1}{d}A(d)$$

where A(d) is the surface area of the unit ball (area of S^d)

Volume of the unit ball

Closed form formulate of V_d:

- S^d: entire surface of a unit sphere; Ω: the solid angle
- $V_d = \int_{x_1^2+x_2^2+\cdots+x_d^2 \le 1} 1 dx_1 dx_2 \ldots dx_d = \int_{S^d} \int_{\gamma=0}^{1} \gamma^{d-1} d\gamma d\Omega$

$$= \frac{1}{d} \int_{S^d} d\Omega = \frac{1}{d} A(d)$$

where A(d) is the surface area of the unit ball (area of S^d)

- A(d) = ?