COMP4121 Advanced Algorithms

Aleks Ignjatović

School of Computer Science and Engineering
University of New South Wales

More randomised algorithms: Gaussian Annulus, Random Projection and Johnson Lindenstrauss Lemmas

Details to be found on pages 12-27 of the Blum, Hopcroft and Kannan textbook

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Let us consider a Gaussian random variable X with zero mean $(E[X] = \mu = 0)$ and variance $V[X] = v = {}^1/_{2\pi}$; its density is given by

$$\frac{1}{\sqrt{2\pi\, v}} e^{-\frac{x^2}{2v}} = e^{-\pi\, x^2}$$

- Assume that we use such X to generate independently the coordinates $(x_1, \ldots, x_d)$ of a random vector x from $\mathbb{R}^d$.
- Since $E[X] = 0$ we have $E[X^2] = E[(X - E[X])^2] = V[X] = {}^1/_{2\pi}$.
- Thus, also $E\left[\frac{X_1^2+\ldots+X_d^2}{d}\right] = d\,V[X]/d = V[X]$.
- We denote by $|x|$ the norm (here just the length) of a vector x, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$.
- We now have $E\left[\frac{|x|^2}{d}\right] = V[X] = {}^1/_{2\pi}$.
- Thus the expected value of the square of the length of such a random vector is $E[|x|^2] = {}^d/_{2\pi}$.
- So, on average, $|x| \approx \frac{\sqrt{d}}{\sqrt{2\pi}} = \Theta(\sqrt{d})$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Let us consider a Gaussian random variable X with zero mean ($E[X] = \mu = 0$) and variance $V[X] = v = 1/2\pi$; its density is given by

$$\frac{1}{\sqrt{2\pi\, v}} e^{-\frac{x^2}{2v}} = e^{-\pi\, x^2}$$

- Assume that we use such X to generate independently the coordinates $(x_1, \ldots, x_d)$ of a random vector $\vec{x}$ from $\mathbb{R}^d$.

- Since $E[X] = 0$ we have $E[X^2] = E[(X - E[X])^2] = V[X] = 1/2\pi$.

- Thus, also $E\left[\frac{X_1^2 + \ldots + X_d^2}{d}\right] = d\,V[X]/d = V[X]$.

- We denote by $|x|$ the norm (here just the length) of a vector x, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$.

- We now have $E\left[\frac{|x|^2}{d}\right] = V[X] = 1/2\pi$.

- Thus the expected value of the square of the length of such a random vector is $E[|x|^2] = d/2\pi$.
- So, on average, $|x| \approx \frac{\sqrt{d}}{\sqrt{2\pi}} = \Theta(\sqrt{d})$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Let us consider a Gaussian random variable X with zero mean ($E[X] = \mu = 0$) and variance $V[X] = v = {}^1/_{2\pi}$; its density is given by

$$\frac{1}{\sqrt{2\pi\, v}} e^{-\frac{x^2}{2v}} = e^{-\pi\, x^2}$$

- Assume that we use such X to generate independently the coordinates $(x_1, \ldots, x_d)$ of a random vector $\vec{x}$ from $\mathbb{R}^d$.
- Since $E[X] = 0$ we have $E[X^2] = E[(X - E[X])^2] = V[X] = {}^1/_{2\pi}$.
- Thus, also $E\left[\frac{X_1^2 + \ldots + X_d^2}{d}\right] = d\,V[X]/d = V[X]$.
- We denote by $|x|$ the norm (here just the length) of a vector x, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$.
- We now have $E\left[\frac{|x|^2}{d}\right] = V[X] = {}^1/_{2\pi}$.
- Thus the expected value of the square of the length of such a random vector is $E[|x|^2] = {}^d/_{2\pi}$.
- So, on average, $|x| \approx \frac{\sqrt{d}}{\sqrt{2\pi}} = \Theta(\sqrt{d})$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Let us consider a Gaussian random variable X with zero mean $(E[X] = \mu = 0)$ and variance $V[X] = v = 1/2\pi$; its density is given by

$$\frac{1}{\sqrt{2\pi\, v}} e^{-\frac{x^2}{2v}} = e^{-\pi\, x^2}$$

- Assume that we use such X to generate independently the coordinates $(x_1, \ldots, x_d)$ of a random vector $\vec{x}$ from $\mathbb{R}^d$.
- Since $E[X] = 0$ we have $E[X^2] = E[(X - E[X])^2] = V[X] = 1/2\pi$.
- Thus, also $E\left[\frac{X_1^2 + \ldots + X_d^2}{d}\right] = d\, V[X]/d = V[X]$.
- We denote by $|x|$ the norm (here just the length) of a vector x, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$,
- We now have $E\left[\frac{|x|^2}{d}\right] = V[X] = 1/2\pi$.
- Thus the expected value of the square of the length of such a random vector is $E[|x|^2] = d/2\pi$.
- So, on average, $|x| \approx \frac{\sqrt{d}}{\sqrt{2\pi}} = \Theta(\sqrt{d})$.

- Let us consider a Gaussian random variable X with zero mean ($E[X] = \mu = 0$) and variance $V[X] = v = 1/2\pi$; its density is given by

$$\frac{1}{\sqrt{2\pi\, v}} e^{-\frac{x^2}{2v}} = e^{-\pi\, x^2}$$

- Assume that we use such X to generate independently the coordinates $(x_1, \ldots, x_d)$ of a random vector $\vec{x}$ from $\mathbb{R}^d$.
- Since $E[X] = 0$ we have $E[X^2] = E[(X - E[X])^2] = V[X] = 1/2\pi$.
- Thus, also $E\left[\frac{X_1^2 + \ldots + X_d^2}{d}\right] = d\, V[X]/d = V[X]$.
- We denote by $|x|$ the norm (here just the length) of a vector x, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$,
- We now have $E\left[\frac{|x|^2}{d}\right] = V[X] = 1/2\pi$.
- Thus the expected value of the square of the length of such a random vector is $E[|x|^2] = d/2\pi$.
- So, on average, $|x| \approx \frac{\sqrt{d}}{\sqrt{2\pi}} = \Theta(\sqrt{d})$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Let us consider a Gaussian random variable X with zero mean ($E[X] = \mu = 0$) and variance $V[X] = v = 1/2\pi$; its density is given by

$$\frac{1}{\sqrt{2\pi\, v}} e^{-\frac{x^2}{2v}} = e^{-\pi\, x^2}$$

- Assume that we use such X to generate independently the coordinates $(x_1, \ldots, x_d)$ of a random vector $\vec{x}$ from $\mathbb{R}^d$.
- Since $E[X] = 0$ we have $E[X^2] = E[(X - E[X])^2] = V[X] = 1/2\pi$.
- Thus, also $E\left[\frac{X_1^2 + \ldots + X_d^2}{d}\right] = d\, V[X]/d = V[X]$.
- We denote by $|x|$ the norm (here just the length) of a vector x, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$,
- We now have $E\left[\frac{|x|^2}{d}\right] = V[X] = 1/2\pi$.
- Thus the expected value of the square of the length of such a random vector is $E[|x|^2] = d/2\pi$.
- So, on average, $|x| \approx \frac{\sqrt{d}}{\sqrt{2\pi}} = \Theta(\sqrt{d})$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Let us consider a Gaussian random variable X with zero mean ($E[X] = \mu = 0$) and variance $V[X] = v = {}^{1}/_{2\pi}$; its density is given by

$$\frac{1}{\sqrt{2\pi\, v}} e^{-\frac{x^2}{2v}} = e^{-\pi\, x^2}$$

- Assume that we use such X to generate independently the coordinates $(x_1, \ldots, x_d)$ of a random vector $\vec{x}$ from $\mathbb{R}^d$.
- Since $E[X] = 0$ we have $E[X^2] = E[(X - E[X])^2] = V[X] = {}^{1}/_{2\pi}$.
- Thus, also $E\left[\frac{X_1^2 + \ldots + X_d^2}{d}\right] = d\, V[X]/d = V[X]$.
- We denote by $|x|$ the norm (here just the length) of a vector x, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$,
- We now have $E\left[\frac{|x|^2}{d}\right] = V[X] = {}^{1}/_{2\pi}$.
- Thus the expected value of the square of the length of such a random vector is $E[|x|^2] = {}^{d}/_{2\pi}$.
- So, on average, $|x| \approx \frac{\sqrt{d}}{\sqrt{2\pi}} = \Theta(\sqrt{d})$.

- Also, by the Law of Large Numbers,

$$\mathcal{P}\left(\left|\frac{x_1^2+\ldots+x_d^2}{d} - V[X]\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2}$$

- Thus,

$$\mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2} \quad \text{i.e.,} \quad \mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| \leq \varepsilon\right) \geq 1 - \frac{V[X^2]}{d\,\varepsilon^2}$$

- This implies that, if d is large, then with high probability

$$\frac{|x|^2}{d} \approx \frac{1}{2\pi} \quad \Rightarrow \quad |x| \approx \sqrt{\frac{d}{2\pi}} = \Theta(\sqrt{d})$$

i.e., thus generated point is at a distance $\Theta(\sqrt{d})$ from the origin.

- If we chose 2 points independently, then

$$E[\langle x, y\rangle] = E[x_1y_1 + \ldots + x_dy_d] = d\,E[XY] = d\,E[X]E[Y] = 0$$

- This means that the expected value of the scalar product of any two vectors with independently randomly chosen coordinates is zero.
- **So, intuitively, vectors with randomly chosen coordinates have approximately the same length $\Theta(\sqrt{d})$ and any two such random vectors are likely to be almost orthogonal!**

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Also, by the Law of Large Numbers,

$$\mathcal{P}\left(\left|\frac{x_1^2+\ldots+x_d^2}{d} - V[X]\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2}$$

- Thus,

$$\mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2} \quad \text{i.e.,} \quad \mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| \leq \varepsilon\right) \geq 1 - \frac{V[X^2]}{d\,\varepsilon^2}$$

- This implies that, if d is large, then with high probability

$$\frac{|x|^2}{d} \approx \frac{1}{2\pi} \quad \Rightarrow \quad |x| \approx \sqrt{\frac{d}{2\pi}} = \Theta(\sqrt{d})$$

i.e., thus generated point is at a distance $\Theta(\sqrt{d})$ from the origin.

- If we chose 2 points independently then

$$E[\langle x, y\rangle] = E[x_1y_1 + \ldots + x_dy_d] = d\,E[XY] = d\,E[X]E[Y] = 0$$

- This means that the expected value of the scalar product of any two vectors with independently randomly chosen coordinates is zero.
- **So, intuitively, vectors with randomly chosen coordinates have approximately the same length $\Theta(\sqrt{d})$ and any two such random vectors are likely to be almost orthogonal!**

- Also, by the Law of Large Numbers,

$$\mathcal{P}\left(\left|\frac{x_1^2+\ldots+x_d^2}{d}-V[X]\right|>\varepsilon\right)\leq\frac{V[X^2]}{d\,\varepsilon^2}$$

- Thus,

$$\mathcal{P}\left(\left|\frac{|x|^2}{d}-\frac{1}{2\pi}\right|>\varepsilon\right)\leq\frac{V[X^2]}{d\,\varepsilon^2}\quad\text{i.e.,}\quad\mathcal{P}\left(\left|\frac{|x|^2}{d}-\frac{1}{2\pi}\right|\leq\varepsilon\right)\geq 1-\frac{V[X^2]}{d\,\varepsilon^2}$$

- This implies that, if d is large, then with high probability

$$\frac{|x|^2}{d}\approx\frac{1}{2\pi}\quad\Rightarrow\quad|x|\approx\sqrt{\frac{d}{2\pi}}=\Theta(\sqrt{d})$$

i.e., thus generated point is at a distance $\Theta(\sqrt{d})$ from the origin.

- If we chose 2 points independently, then

$$E[\langle x,y\rangle]=E[x_1y_1+\ldots+x_dy_d]=d\,E[XY]=d\,E[X]E[Y]=0$$

- This means that the expected value of the scalar product of any two vectors with independently randomly chosen coordinates is zero.
- **So, intuitively, vectors with randomly chosen coordinates have approximately the same length $\Theta(\sqrt{d})$ and any two such random vectors are likely to be almost orthogonal!**

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Also, by the Law of Large Numbers,

$$\mathcal{P}\left(\left|\frac{x_1^2+\ldots+x_d^2}{d} - V[X]\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2}$$

- Thus,

$$\mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2} \quad \text{i.e.,} \quad \mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| \leq \varepsilon\right) \geq 1 - \frac{V[X^2]}{d\,\varepsilon^2}$$

- This implies that, if d is large, then with high probability

$$\frac{|x|^2}{d} \approx \frac{1}{2\pi} \quad \Rightarrow \quad |x| \approx \sqrt{\frac{d}{2\pi}} = \Theta(\sqrt{d})$$

i.e., thus generated point is at a distance $\Theta(\sqrt{d})$ from the origin.

- If we chose 2 points independently, then

$$E[\langle x, y\rangle] = E[x_1 y_1 + \ldots + x_d y_d] = d\,E[XY] = d\,E[X]E[Y] = 0$$

- This means that the expected value of the scalar product of any two vectors with independently randomly chosen coordinates is zero.
- **So, intuitively, vectors with randomly chosen coordinates have approximately the same length $\Theta(\sqrt{d})$ and any two such random vectors are likely to be almost orthogonal!**

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Also, by the Law of Large Numbers,

$$\mathcal{P}\left(\left|\frac{x_1^2+\ldots+x_d^2}{d} - V[X]\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2}$$

- Thus,

$$\mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2} \quad \text{i.e.,} \quad \mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| \leq \varepsilon\right) \geq 1 - \frac{V[X^2]}{d\,\varepsilon^2}$$

- This implies that, if d is large, then with high probability

$$\frac{|x|^2}{d} \approx \frac{1}{2\pi} \quad \Rightarrow \quad |x| \approx \sqrt{\frac{d}{2\pi}} = \Theta(\sqrt{d})$$

  i.e., thus generated point is at a distance $\Theta(\sqrt{d})$ from the origin.
- If we chose 2 points independently, then

$$E[\langle x, y\rangle] = E[x_1y_1 + \ldots + x_dy_d] = d\,E[XY] = d\,E[X]E[Y] = 0$$

- This means that the expected value of the scalar product of any two vectors with independently randomly chosen coordinates is zero.
- So, intuitively, vectors with randomly chosen coordinates have approximately the same length $\Theta(\sqrt{d})$ and any two such random vectors are likely to be almost orthogonal!

- Also, by the Law of Large Numbers,

$$\mathcal{P}\left(\left|\frac{x_1^2+\ldots+x_d^2}{d} - V[X]\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2}$$

- Thus,

$$\mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2} \quad \text{i.e.,} \quad \mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| \leq \varepsilon\right) \geq 1 - \frac{V[X^2]}{d\,\varepsilon^2}$$

- This implies that, if d is large, then with high probability

$$\frac{|x|^2}{d} \approx \frac{1}{2\pi} \quad \Rightarrow \quad |x| \approx \sqrt{\frac{d}{2\pi}} = \Theta(\sqrt{d})$$

i.e., thus generated point is at a distance $\Theta(\sqrt{d})$ from the origin.

- If we chose 2 points independently, then

$$E[\langle x, y\rangle] = E[x_1 y_1 + \ldots + x_d y_d] = d\,E[XY] = d\,E[X]E[Y] = 0$$

- This means that the expected value of the scalar product of any two vectors with independently randomly chosen coordinates is zero.
- **So, intuitively, vectors with randomly chosen coordinates have approximately the same length $\Theta(\sqrt{d})$ and any two such random vectors are likely to be almost orthogonal!**

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Also, by the Law of Large Numbers,

$$\mathcal{P}\left(\left|\frac{x_1^2+\ldots+x_d^2}{d} - V[X]\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2}$$

- Thus,

$$\mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| > \varepsilon\right) \leq \frac{V[X^2]}{d\,\varepsilon^2} \quad \text{i.e.,} \quad \mathcal{P}\left(\left|\frac{|x|^2}{d} - \frac{1}{2\pi}\right| \leq \varepsilon\right) \geq 1 - \frac{V[X^2]}{d\,\varepsilon^2}$$

- This implies that, if d is large, then with high probability

$$\frac{|x|^2}{d} \approx \frac{1}{2\pi} \quad \Rightarrow \quad |x| \approx \sqrt{\frac{d}{2\pi}} = \Theta(\sqrt{d})$$

  i.e., thus generated point is at a distance $\Theta(\sqrt{d})$ from the origin.
- If we chose 2 points independently, then

$$E[\langle x, y\rangle] = E[x_1y_1 + \ldots + x_dy_d] = d\,E[XY] = d\,E[X]E[Y] = 0$$

- This means that the expected value of the scalar product of any two vectors with independently randomly chosen coordinates is zero.
- **So, intuitively, vectors with randomly chosen coordinates have approximately the same length $\Theta(\sqrt{d})$ and any two such random vectors are likely to be almost orthogonal!**

High dimensional spaces are quite counter-intuitive!

- The above should hint that high dimensional spaces are quite counter-intuitive.
- We have to be care[illegible] transfer our 3D intuition to high dimensional spaces.
- We now show two strange facts about high dimensional balls:
  1. Most of the volume of a high dimensional ball is near any of its equators. By this we mean that if we cut a high dimensional ball with any two parallel hyper-planes symmetric with respect to the center and close to the center, then most of the volume of the ball is between these two hyper-planes.
  2. Most of the volume of a high dimensional ball is near its surface. By this we mean that if we consider a ball of radius r and a slightly smaller ball of radius $r(1-\epsilon)$ than most of the volume of the bigger ball is in the annulus outside the smaller ball.
- These facts are not just curiosities, but have importance for algorithms we will study.

High dimensional spaces are quite counter-intuitive!

- The above should hint that high dimensional spaces are quite counter-intuitive.
- We have to be careful not to transfer our 3D intuition to high dimensional spaces.

- We now show two strange facts about high dimensional balls:
  1. Most of the volume of a high dimensional ball is near any of its equators. By this we mean that if we cut a high dimensional ball with any two parallel hyper-planes symmetric with respect to the center and close to the center, then most of the volume of the ball is between these two hyper-planes.
  2. Most of the volume of a high dimensional ball is near its surface. By this we mean that if we consider a ball of radius r and a slightly smaller ball of radius $r(1-\epsilon)$ than most of the volume of the bigger ball is in the annulus outside the smaller ball.
- These facts are not just curiosities, but have importance for algorithms we will study.

High dimensional spaces are quite counter-intuitive!

- The above should hint that high dimensional spaces are quite counter-intuitive.
- We have to be care[illegible] transfer our 3D intuition to high dimensional spaces
- We now show two strange facts about high dimensional balls:
  1. Most of the volume of a high dimensional ball is near any of its equators. By this we mean that if we cut a high dimensional ball with any two parallel hyper-planes symmetric with respect to the center and close to the center, then most of the volume of the ball is between these two hyper-planes.
  2. Most of the volume of a high dimensional ball is near its surface. By this we mean that if we consider a ball of radius r and a slightly smaller ball of radius $r(1-\epsilon)$ than most of the volume of the bigger ball is in the annulus outside the smaller ball.
- These facts are not just curiosities, but have importance for algorithms we will study.

High dimensional spaces are quite counter-intuitive!

- The above should hint that high dimensional spaces are quite counter-intuitive.
- We have to be careful when we try to transfer our 3D intuition to high dimensional spaces.
- We now show two strange facts about high dimensional balls:
  1. Most of the volume of a high dimensional ball is near any of its equators. By this we mean that if we cut a high dimensional ball with any two parallel hyper-planes symmetric with respect to the center and close to the center, then most of the volume of the ball is between these two hyper-planes.
  2. Most of the volume of a high dimensional ball is near its surface. By this we mean that if we consider a ball of radius r and a slightly smaller ball of radius $r(1-\epsilon)$ than most of the volume of the bigger ball is in the annulus outside the smaller ball.
- These facts are not just curiosities, but have importance for algorithms we will study.

High dimensional spaces are quite counter-intuitive!

- The above should hint that high dimensional spaces are quite counter-intuitive.

- We have to be care[illegible] transfer our 3D intuition to high dimensional spaces

- We now show two strange facts about high dimensional balls:
  1. Most of the volume of a high dimensional ball is near any of its equators. By this we mean that if we cut a high dimensional ball with any two parallel hyper-planes symmetric with respect to the center and close to the center, then most of the volume of the ball is between these two hyper-planes.
  2. Most of the volume of a high dimensional ball is near its surface. By this we mean that if we consider a ball of radius r and a slightly smaller ball of radius $r(1-\epsilon)$ than most of the volume of the bigger ball is in the annulus outside the smaller ball.

- These facts are not just curiosities, but have importance for algorithms we will study.

High dimensional spaces are quite counter-intuitive!

- The above should hint that high dimensional spaces are quite counter-intuitive.
- We have to be careful not to transfer our 3D intuition to high dimensional spaces.
- We now show two strange facts about high dimensional balls:
  1. Most of the volume of a high dimensional ball is near any of its equators. By this we mean that if we cut a high dimensional ball with any two parallel hyper-planes symmetric with respect to the center and close to the center, then most of the volume of the ball is between these two hyper-planes.
  2. Most of the volume of a high dimensional ball is near its surface. By this we mean that if we consider a ball of radius r and a slightly smaller ball of radius $r(1-\epsilon)$ than most of the volume of the bigger ball is in the annulus outside the smaller ball.
- These facts are not just curiosities, but have importance for algorithms we will study.

High dimensional spaces are quite counter-intuitive!

- The above should hint that high dimensional spaces are quite counter-intuitive.
- We have to be care[illegible] transfer our 3D intuition to high dimensional spaces.
- We now show two strange facts about high dimensional balls:
  1. Most of the volume of a high dimensional ball is near any of its equators. By this we mean that if we cut a high dimensional ball with any two parallel hyper-planes symmetric with respect to the center and close to the center, then most of the volume of the ball is between these two hyper-planes.
  2. Most of the volume of a high dimensional ball is near its surface. By this we mean that if we consider a ball of radius r and a slightly smaller ball of radius $r(1-\epsilon)$ than most of the volume of the bigger ball is in the annulus outside the smaller ball.
- These facts are not just curiosities, but have importance for algorithms we will study.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- We first note that the volume of a sphere of radius r is proportional to r^d, in the sense that, if we denote by $V(d)$ the volume of the d-dimensional ball of radius 1, then the volume of a d-dimensional ball of radius r is equal to $r^d V(d)$. (This is clear for a d-dimensional cube and it follows for any other solid by the exhaustion method: we approximate the solid by a union of small disjoint hypercubes).
- This implies that the volume of a d-dimensional ball can be represented as the following integral:

$$V(d) = \int_{-1}^{1} \left(\sqrt{1 - x_1^2}\right)^{d-1} V(d-1)dx_1$$

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- We first note that the volume of a sphere of radius is proportional to r^d, in the sense that, if we denote by $V(d)$ the volume of the d-dimensional ball of radius 1, then the volu[illegible]imensional ball of radius r is equal to $r^d V(d)$. (This is clear [illegible]mensional cube and it follows for any other solid by the exhaustio[illegible] we approximate the solid by a union of small disjoint hypercubes).
- This implies that the [illegible] d-dimensional ball can be represented as the following integral:

$$V(d) = \int_{-1}^{1} \left(\sqrt{1 - x_1^2}\right)^{d-1} V(d-1)\,dx_1$$

- Let c be a constant. Consider the portion S of a unit ball consisting of all points such that their x_1 coordinate satisfies $x_1 \geq \frac{c}{\sqrt{d-1}}$.
- Let A be the volume of such a slice S.
- Then

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1-x_1^2}\right)^{d-1} V(d-1)\,dx_1$$

- (Recall that our goal is to show that two randomly and independently generated vectors are almost certainly almost orthogonal; to achieve this we need to perform some mumbo-jumbo whose purpose will be clear a bit later.)

- Let c be a constant. Consider the portion S of a unit ball consisting of all points such that their x_1 coordinate satisfies $x_1 \geq \frac{c}{\sqrt{d-1}}$.
- Let A be the volume of such a slice S.
- Then

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1-x_1^2}\right)^{d-1} V(d-1)dx_1$$

- (Recall that our goal is to show that two randomly and independently generated vectors are almost certainly almost orthogonal; to achieve this we need to perform some mumbo-jumbo whose purpose will be clear a bit later)

- Let c be a constant. Consider the portion S of a unit ball consisting of all points such that their x_1 coordinate satisfies $x_1 \geq \frac{c}{\sqrt{d-1}}$.
- Let A be the volume of such a slice S.
- Then

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1-x_1^2}\right)^{d-1} V(d-1)\,dx_1$$

- (Recall that our goal is to show that two randomly and independently generated vectors are almost certainly almost orthogonal; to achieve this we need to perform some mumbo-jumbo whose purpose will be clear a bit later)

- Let c be a constant. Consider the portion S of a unit ball consisting of all points such that their x_1 coordinate satisfies $x_1 \geq \frac{c}{\sqrt{d-1}}$.
- Let A be the volume of such a slice S.
- Then

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1-x_1^2}\right)^{d-1} V(d-1)\,dx_1$$

- (Recall that our goal is to show that two randomly and independently generated vectors are almost certainly almost orthogonal; to achieve this we need to perform some mumbo-jumbo whose purpose will be clear a bit later).

- We need the inequality $1 - x \leq e^{-x}$ for all $0 \leq x \leq 1$.
- For a proof, see the refresher notes on probability and statistics, available at the course web site.
- Using such an inequali[illegible]at we have

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1-x_1^2}\right)^{d-1} V(d-1)dx_1$$

we obtain

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(e^{-x_1^2}\right)^{\frac{d-1}{2}} V(d-1)dx_1$$

Over the slice S we have $x_1 \geq \frac{c}{\sqrt{d-1}}$ which implies $\frac{x_1\sqrt{d-1}}{c} \geq 1$ and this implies

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \frac{x_1\sqrt{d-1}}{c} e^{-x_1^2\frac{d-1}{2}} V(d-1)dx_1$$

- The purpose of adding this term is to allow substitution of variable x_1 with variable $u = x_1^2$; then $du = 2x_1 dx_1$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- We need the inequality $1 - x \le e^{-x}$ for all $0 \le x \le 1$.
- For a proof, see the refresher notes on probability and statistics, available at the course web site.
- Using such an inequality, that we have

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1 - x_1^2}\right)^{d-1} V(d-1)dx_1$$

we obtain

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(e^{-x_1^2}\right)^{\frac{d-1}{2}} V(d-1)dx_1$$

Over the slice S we have $x_1 \ge \frac{c}{\sqrt{d-1}}$ which implies $\frac{x_1\sqrt{d-1}}{c} \ge 1$ and this implies

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \frac{x_1\sqrt{d-1}}{c} e^{-x_1^2 \frac{d-1}{2}} V(d-1)dx_1$$

- The purpose of adding this term is to allow substitution of variable x_1 with variable $u = x_1^2$; then $du = 2x_1 dx_1$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- We need the inequality $1 - x \le e^{-x}$ for all $0 \le x \le 1$.
- For a proof, see the refresher notes on probability and statistics, available at the course web site.
- Using such an inequali... hat we have

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1 - x_1^2}\right)^{d-1} V(d-1) dx_1$$

we obtain

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(e^{-x_1^2}\right)^{\frac{d-1}{2}} V(d-1) dx_1$$

Over the slice S we have $x_1 \ge \frac{c}{\sqrt{d-1}}$ which implies $\frac{x_1\sqrt{d-1}}{c} \ge 1$ and this implies

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \frac{x_1\sqrt{d-1}}{c} e^{-x_1^2 \frac{d-1}{2}} V(d-1) dx_1$$

- The purpose of adding this term is to allow substitution of variable x_1 with variable $u = x_1^2$; then $du = 2x_1 dx_1$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- We need the inequality $1 - x \le e^{-x}$ for all $0 \le x \le 1$.
- For a proof, see the refresher notes on probability and statistics, available at the course web site.
- Using such an inequality on what we have

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1 - x_1^2}\right)^{d-1} V(d-1)dx_1$$

we obtain

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(e^{-x_1^2}\right)^{\frac{d-1}{2}} V(d-1)dx_1$$

Over the slice S we have $x_1 \ge \frac{c}{\sqrt{d-1}}$ which implies $\frac{x_1\sqrt{d-1}}{c} \ge 1$ and this implies

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \frac{x_1\sqrt{d-1}}{c} e^{-x_1^2 \frac{d-1}{2}} V(d-1)dx_1$$

- The purpose of adding this term is to allow substitution of variable x_1 with variable $u = x_1^2$; then $du = 2x_1 dx_1$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- We need the inequality $1 - x \leq e^{-x}$ for all $0 \leq x \leq 1$.
- For a proof, see the refresher notes on probability and statistics, available at the course web site.
- Using such an inequality and the fact that we have

$$A = \int_{\frac{c}{\sqrt{d-1}}}^{1} \left(\sqrt{1 - x_1^2}\right)^{d-1} V(d-1)dx_1$$

we obtain

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} (e^{-x_1^2})^{\frac{d-1}{2}} V(d-1)dx_1$$

Over the slice S we have $x_1 \geq \frac{c}{\sqrt{d-1}}$ which implies $\frac{x_1\sqrt{d-1}}{c} \geq 1$ and this implies

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \frac{x_1\sqrt{d-1}}{c} e^{-x_1^2\frac{d-1}{2}} V(d-1)dx_1$$

- The purpose of adding this term is to allow substitution of variable x_1 with variable $u = x_1^2$; then $du = 2x_1dx_1$.

Generating random points in d-dimensional spaces $\mathbb{R}^d$

- Thus, with $u = x_1^2$ we have $du = 2x_1 dx_1$ and so, from

$$A < \int_{\frac{c}{\sqrt{d-1}}}^{1} \frac{x_1\sqrt{d-1}}{c} e^{-x_1^2\frac{d-1}{2}} V(d-1)\,dx_1$$

we obtain

$$A < \int_{\frac{c^2}{d-1}}^{1} \frac{\sqrt{d-1}}{c} e^{-u\frac{d-1}{2}} V(d-1)\frac{1}{2}\,du$$

$$< \frac{\sqrt{d-1}}{2c} V(d-1) \int_{\frac{c^2}{d-1}}^{\infty} e^{-u\frac{d-1}{2}}\,du$$

$$= \frac{\sqrt{d-1}}{2c} V(d-1) \frac{1}{\frac{d-1}{2}} e^{-u\frac{d-1}{2}} \Big|_{u=\frac{c^2}{d-1}}^{u=\infty}$$

$$= \frac{\sqrt{d-1}}{2c} V(d-1) \frac{1}{\frac{d-1}{2}} e^{-\frac{d-1}{2}\frac{c^2}{d-1}}$$

$$= \frac{1}{c\sqrt{d-1}} V(d-1) e^{-\frac{c^2}{2}}$$

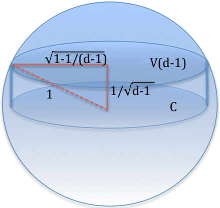
- So we have obtained that the volume A of the slice S of the unit ball which lies above the hyperplane $x_1 = \frac{c}{\sqrt{d-1}}$ satisfies

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

- We now want to show that this volume is small compared to half the volume of the entire ball. However, we do not want to use $V(d)$, we want an expression involving $V(d-1)$ which can cancel $V(d-1)$ in the above bound.
- Note that the volume of the whole hemisphere H is larger than the volume of a cylinder below the intersection of the ball and the plane $x_1 = \frac{1}{\sqrt{d-1}}$, which has the volume C which is the product of the d-1 dimensional "surface area" of its base which is equal to $V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}$ and its height which is $1/\sqrt{d-1}$.
- Thus, the volume C of such a cylinder

$$C = V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}\frac{1}{\sqrt{d-1}}$$

$$= V(d-1)\left(1-\frac{1}{d-1}\right)^{\frac{d-1}{2}}\frac{1}{\sqrt{d-1}}$$



- So we have obtained that the volume A of the slice S of the unit ball which lies above the hyperplane $x_1 = \frac{c}{\sqrt{d-1}}$ satisfies

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

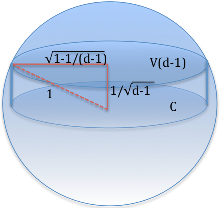
- We now want to show that this volume is small compared to half the volume of the entire ball. However, we do not want to use $V(d)$, we want an expression involving $V(d-1)$ which can cancel $V(d-1)$ in the above bound.

- Note that the volume of the whole hemisphere H is larger than the volume of a cylinder below the intersection of the ball and the plane $x_1 = \frac{1}{\sqrt{d-1}}$, which has the volume C which is the product of the d-1 dimensional "surface area" of its base which is equal to $V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}$ and its height which is $1/\sqrt{d-1}$.

- Thus, the volume C of such a cylinder is

$$C = V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}\frac{1}{\sqrt{d-1}}$$

$$= V(d-1)\left(1-\frac{1}{d-1}\right)^{\frac{d-1}{2}}\frac{1}{\sqrt{d-1}}$$



- So we have obtained that the volume A of the slice S of the unit ball which lies above the hyperplane $x_1 = \frac{c}{\sqrt{d-1}}$ satisfies

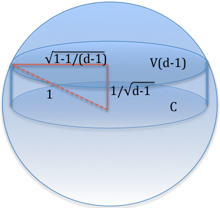
$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

- We now want to show [illegible] volume is small compared to half the volume of the entire ball. How[illegible] not want to use $V(d)$, we want an expression involving $V(d-1)$ which can cancel $V(d-1)$ in the above bound.
- Note that the volume of the whole hemisphere H is larger than the volume of a cylinder below the intersection of the ball and the plane $x_1 = \frac{1}{\sqrt{d-1}}$, which has the volume C which is the product of the d-1 dimensional "surface area" of its base which is equal to $V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}$ and its height which is ${}^{1}/\sqrt{d-1}$.

- Thus, the volume C of such a cylinder is

$$C = V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}\frac{1}{\sqrt{d-1}}$$

$$= V(d-1)\left(1-\frac{1}{d-1}\right)^{\frac{d-1}{2}}\frac{1}{\sqrt{d-1}}$$



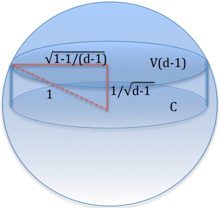
- So we have obtained that the volume A of the slice S of the unit ball which lies above the hyperplane $x_1 = \frac{c}{\sqrt{d-1}}$ satisfies

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

- We now want to show [illegible] volume is small compared to half the volume of the entire ball. How[illegible] not want to use $V(d)$, we want an expression involving $V(d-1)$ which can cancel $V(d-1)$ in the above bound.
- Note that the volume of the whole hemisphere H is larger than the volume of a cylinder below the intersection of the ball and the plane $x_1 = \frac{1}{\sqrt{d-1}}$, which has the volume C which is the product of the d-1 dimensional "surface area" of its base which is equal to $V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}$ and its height which is $1/\sqrt{d-1}$.
- Thus, the volume C of such a cylinder is

$$C = V(d-1)\left(\sqrt{1-\frac{1}{d-1}}\right)^{d-1}\frac{1}{\sqrt{d-1}}$$

$$= V(d-1)\left(1-\frac{1}{d-1}\right)^{\frac{d-1}{2}}\frac{1}{\sqrt{d-1}}$$



- We now need yet another inequality: for all $\alpha \geq 1$ and all $0 < x < 1$,

$$(1-x)^{\alpha} \geq 1 - \alpha x$$

(This inequality can be proved by considering $\beta(x) = (1-x)^{\alpha} - 1 + \alpha x$; again, $\beta(0) = 0$ and $\beta'(x) = -\alpha(1-x)^{\alpha-1} + \alpha = \alpha(1-(1-x)^{\alpha-1}) \geq 0$ which implies that $\beta(x)$ is non-decreasing and thus non-negative.)

- Applying the above inequality with $\alpha = (d-1)/2$ and $x = 1/(d-1)$ we obtain

$$\begin{aligned} C &= V(d-1)\left(1-\frac{1}{d-1}\right)^{\frac{d-1}{2}}\frac{1}{\sqrt{d-1}} \\ &\geq V(d-1)\left(1-\frac{1}{d-1}\,\frac{d-1}{2}\right)\frac{1}{\sqrt{d-1}} \\ &\geq V(d-1)\frac{1}{2\sqrt{d-1}} \end{aligned}$$

- We now need yet another inequality: for all $\alpha \geq 1$ and all $0 < x < 1$,

$$(1-x)^{\alpha} \geq 1 - \alpha x$$

(This inequality can b[illegible]ved by considering $\beta(x) = (1-x)^{\alpha} - 1 + \alpha x$; again, $\beta(0) = 0$ and $\beta'(x) = -\alpha(1-x)^{\alpha-1} + \alpha = \alpha(1-(1-x)^{\alpha-1}) \geq 0$ which implies that $\beta(x)$ is non-decreasing and thus non-negative.)

- Applying the above inequality with $\alpha = (d-1)/2$ and $x = 1/(d-1)$ we obtain

$$\begin{aligned} C &= V(d-1)\left(1 - \frac{1}{d-1}\right)^{\frac{d-1}{2}} \frac{1}{\sqrt{d-1}} \\ &\geq V(d-1)\left(1 - \frac{1}{d-1}\,\frac{d-1}{2}\right) \frac{1}{\sqrt{d-1}} \\ &\geq V(d-1)\,\frac{1}{2\sqrt{d-1}} \end{aligned}$$

- We now need yet another inequality: for all $\alpha \geq 1$ and all $0 < x < 1$,

$$(1-x)^{\alpha} \geq 1 - \alpha x$$

(This inequality can be proved by considering $\beta(x) = (1-x)^{\alpha} - 1 + \alpha x$; again, $\beta(0) = 0$ and $\beta'(x) = -\alpha(1-x)^{\alpha-1} + \alpha = \alpha(1-(1-x)^{\alpha-1}) \geq 0$ which implies that $\beta(x)$ is non-decreasing and thus non-negative.)
- Applying the above inequality with $\alpha = (d-1)/2$ and $x = 1/(d-1)$ we obtain

$$\begin{aligned} C &= V(d-1)\left(1 - \frac{1}{d-1}\right)^{\frac{d-1}{2}} \frac{1}{\sqrt{d-1}} \\ &\geq V(d-1)\left(1 - \frac{1}{d-1}\frac{d-1}{2}\right)\frac{1}{\sqrt{d-1}} \\ &\geq V(d-1)\frac{1}{2\sqrt{d-1}} \end{aligned}$$

- So we got that the volume A of the slice of the unit ball which is above the plane $x_1 = \frac{c}{\sqrt{d-1}}$ and the volume H of the whole hemisphere satisfy

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

$$H \geq V(d-1)\frac{1}{2\sqrt{d-1}}$$

- Thus, the ratio of the volume of the slice of the sphere and the whole hemisphere satisfies

$$\frac{A}{H} < \frac{V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}}{V(d-1)\frac{1}{2\sqrt{d-1}}} = \frac{2}{c}e^{-\frac{c^2}{2}}$$

- **Corollary:** If we pick uniformly at random a point from the unit ball, the probability that its coordinate x_1 satisfies $|x_1| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- **This also means that most of the volume of the unit ball is between the hyperplanes $x_1 = -\frac{c}{\sqrt{d-1}}$ and $x_1 = \frac{c}{\sqrt{d-1}}$, i.e., near the equator.**
- Also, the annulus consisting of all x such that $r \leq |x| \leq 1$ for $r < 1$ has volume $V(d) - r^d V(d) = V(d)(1 - r^d)$ so **if d is large almost all of the volume of the unit ball is near its surface.**

- So we got that the volume A of the slice of the unit ball which is above the plane $x_1 = \frac{c}{\sqrt{d-1}}$ and the volume H of the whole hemisphere satisfy

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

$$H \geq V(d-1)\frac{1}{2\sqrt{d-1}}$$

- Thus, the ratio of the volume of the slice of the sphere and the whole hemisphere satisfies

$$\frac{A}{H} < \frac{V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}}{V(d-1)\frac{1}{2\sqrt{d-1}}} = \frac{2}{c}e^{-\frac{c^2}{2}}$$

- **Corollary:** If we pick uniformly at random a point from the unit ball, the probability that its coordinate x_1 satisfies $|x_1| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.
- **This also means that most of the volume of the unit ball is between the hyperplanes $x_1 = -\frac{c}{\sqrt{d-1}}$ and $x_1 = \frac{c}{\sqrt{d-1}}$, i.e., near the equator.**
- Also, the annulus consisting of all x such that $r \leq |x| \leq 1$ for $r < 1$ has volume $V(d) - r^dV(d) = V(d)(1 - r^d)$ so **if d is large almost all of the volume of the unit ball is near its surface.**

- So we got that the volume A of the slice of the unit ball which is above the plane $x_1 = \frac{c}{\sqrt{d-1}}$ and the volume H of the whole hemisphere satisfy

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

$$H \geq V(d-1)\frac{1}{2\sqrt{d-1}}$$

- Thus, the ratio of the volume of the slice of the sphere and the whole hemisphere satisfies

$$\frac{A}{H} < \frac{V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}}{V(d-1)\frac{1}{2\sqrt{d-1}}} = \frac{2}{c}e^{-\frac{c^2}{2}}$$

- **Corollary:** If we pick uniformly at random a point from the unit ball, the probability that its coordinate x_1 satisfies $|x_1| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- **This also means that most of the volume of the unit ball is between the hyperplanes $x_1 = -\frac{c}{\sqrt{d-1}}$ and $x_1 = \frac{c}{\sqrt{d-1}}$, i.e., near the equator.**
- Also, the annulus consisting of all x such that $r \leq |x| \leq 1$ for $r < 1$ has volume $V(d) - r^d V(d) = V(d)(1 - r^d)$ so **if d is large almost all of the volume of the unit ball is near its surface.**

- So we got that the volume A of the slice of the unit ball which is above the plane $x_1 = \frac{c}{\sqrt{d-1}}$ and the volume H of the whole hemisphere satisfy

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

$$H \geq V(d-1)\frac{1}{2\sqrt{d-1}}$$

- Thus, the ratio of the volume of the slice of the sphere and the whole hemisphere satisfies

$$\frac{A}{H} < \frac{V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}}{V(d-1)\frac{1}{2\sqrt{d-1}}} = \frac{2}{c}e^{-\frac{c^2}{2}}$$

- **Corollary:** If we pick uniformly at random a point from the unit ball, the probability that its coordinate x_1 satisfies $|x_1| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- **This also means that most of the volume of the unit ball is between the hyperplanes $x_1 = -\frac{c}{\sqrt{d-1}}$ and $x_1 = \frac{c}{\sqrt{d-1}}$, i.e., near the equator.**

- Also, the annulus consisting of all x such that $r \leq |x| \leq 1$ for $r < 1$ has volume $V(d) - r^d V(d) = V(d)(1 - r^d)$ so **if d is large almost all of the volume of the unit ball is near its surface.**

- So we got that the volume A of the slice of the unit ball which is above the plane $x_1 = \frac{c}{\sqrt{d-1}}$ and the volume H of the whole hemisphere satisfy

$$A < V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}$$

$$H \geq V(d-1)\frac{1}{2\sqrt{d-1}}$$

- Thus, the ratio of the volume of the slice of the sphere and the whole hemisphere satisfies

$$\frac{A}{H} < \frac{V(d-1)\frac{e^{-\frac{c^2}{2}}}{c\sqrt{d-1}}}{V(d-1)\frac{1}{2\sqrt{d-1}}} = \frac{2}{c}e^{-\frac{c^2}{2}}$$

- **Corollary:** If we pick uniformly at random a point from the unit ball, the probability that its coordinate x_1 satisfies $|x_1| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- **This also means that most of the volume of the unit ball is between the hyperplanes $x_1 = -\frac{c}{\sqrt{d-1}}$ and $x_1 = \frac{c}{\sqrt{d-1}}$, i.e., near the equator.**

- Also, the annulus consisting of all x such that $r \leq |x| \leq 1$ for $r < 1$ has volume $V(d) - r^d V(d) = V(d)(1 - r^d)$ so **if d is large almost all of the volume of the unit ball is near its surface.**

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.
- Thus, the above two fact guarantee that with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e$ [illegible] ue smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogon[illegible]
- Next theorem generali[illegible] n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.
- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.
- The purpose is to reduce the dimensionality of the space.
- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball. Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $1 \geq |x_i| > 1 - \frac{2\ln n}{d}$;
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.
- Thus, the above two fact guarantee that **with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e$[illegible]ue smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogon**[illegible]
- Next theorem generali[illegible] n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.
- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.
- The purpose is to reduce the dimensionality of the space.
- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball. Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $1 \geq |x_i| > 1 - \frac{2 \ldots}{d}$;
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.
- Thus, the above two fact guarantee that **with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e$ [illegible] ue smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogonal.**
- Next theorem generali[illegible] n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.
- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.
- The purpose is to reduce the dimensionality of the space.
- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball.
  Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $1 \geq |x_i| > 1 - \frac{2 \ln n}{d}$;
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.

- Thus, the above two fact guarantee that **with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e$[illegible]ue smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogon**[illegible]
- Next theorem generali[illegible] n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.

- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.

- The purpose is to reduce the dimensionality of the space.

- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball.

  Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,

  1. $1 \geq |x_i| > 1 - \frac{2\ln n}{d}$;

  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6\ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.
- Thus, the above two fact guarantee that **with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e$ … ue smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogon…**
- Next theorem generali… n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.
- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.
- The purpose is to reduce the dimensionality of the space.
- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball. Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $1 \geq |x_i| > 1 - \frac{2 \ln n}{d}$,
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.
- Thus, the above two fact guarantee that **with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e^{-c^2/2}$ a value smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogonal.**
- Next theorem generalizes this to n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.
- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.
- The purpose is to reduce the dimensionality of the space.
- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball.
Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $1 \geq |x_i| > 1 - \frac{2 \ln n}{d}$;
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.
- Thus, the above two fact guarantee that **with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e$ ... ue smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogon...**
- Next theorem generali... n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.
- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.
- The purpose is to reduce the dimensionality of the space.
- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball.

  Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,

  1. $1 \geq |x_i| > 1 - \frac{2 \ln n}{d}$,
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.

- Thus, the above two fact guarantee that **with high probability** $|x| \approx 1$ **and** $|x'| \approx 1$ **as well as that the projection of** x' **onto** x **has, with high probability** $\geq 1 - \frac{2}{c}e$ [illegible] **ue smaller than** $\frac{c}{\sqrt{d-1}}$**, i.e., that** x **and** x' **are almost orthogon**[illegible]
- Next theorem generali[illegible] n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.

- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.

- The purpose is to reduce the dimensionality of the space.

- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball.

  Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,

  1. $1 \geq |x_i| > 1 - \frac{2 \ln n}{d}$,
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.

- If we randomly and independently choose 2 points x, x' from the unit ball, we can rotate the ball so that x is in the direction of x_1 coordinate.
- Thus, the above two fact guarantee that **with high probability $|x| \approx 1$ and $|x'| \approx 1$ as well as that the projection of x' onto x has, with high probability $\geq 1 - \frac{2}{c}e$[illegible]ue smaller than $\frac{c}{\sqrt{d-1}}$, i.e., that x and x' are almost orthogon**[illegible]
- Next theorem generali[illegible] n points.
- It shows that if we draw randomly a relatively small number n of vectors compared to the dimension d, we are almost guaranteed that they will be of approximately the same length and that form an almost orthogonal basis.
- They will span a space onto which we will project vectors from $\mathbb{R}^d$ into a space of dimension n.
- The purpose is to reduce the dimensionality of the space.
- **Theorem:** Assume we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball.
  Then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $1 \geq |x_i| > 1 - \frac{2\ln n}{d}$,
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6\ln n}{d-1}}$.

- **Proof:** Probability that $|x_i| < 1 - \varepsilon$ is

$$\frac{V(d)(1-\varepsilon)^d}{V(d)} = (1-\varepsilon)^d < e^{-\varepsilon d}$$

- Thus, taking $\varepsilon = \frac{2 \ln n}{d}$, we get that for each i

$$\mathcal{P}\left(|x_i| < 1 - \frac{2 \ln n}{d}\right) < e^{-\frac{2 \ln n}{d} d} = e^{-\ln n^2} = \frac{1}{n^2}$$

- Thus, the probability that $|x_i| < 1 - \frac{2 \ln n}{d}$ for at least one i is less than $n \cdot \frac{1}{n^2} = \frac{1}{n}$. This reasoning is called the *Union bound.*
- *Union bound:* if you have n sets A_i (not necessarily disjoint) such that $\mathcal{P}(x \in A_i) = p_i$ then $\mathcal{P}\left(x \in \cup_{i=1}^n A_i\right) \leq \sum_{i=1}^n p_i$.

- **Proof:** Probability that $|x_i| < 1 - \varepsilon$ is

$$\frac{V(d)(1-\varepsilon)^d}{V(d)} = (1-\varepsilon)^d < e^{-\varepsilon d}$$

- Thus, taking $\varepsilon = \frac{2 \ln n}{d}$ we have that for each i

$$\mathcal{P}\left(|x_i| < 1 - \frac{2 \ln n}{d}\right) < e^{-\frac{2 \ln n}{d} d} = e^{-\ln n^2} = \frac{1}{n^2}$$

- Thus, the probability that $|x_i| < 1 - \frac{2 \ln n}{d}$ for at least one i is less than $n \cdot \frac{1}{n^2} = \frac{1}{n}$. This reasoning is called the *Union bound.*
- *Union bound:* if you have sets A_i (not necessarily disjoint) such that $\mathcal{P}(x \in A_i) = p_i$ then $\mathcal{P}\left(x \in \cup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} p_i$.

- **Proof:** Probability that $|x_i| < 1 - \varepsilon$ is

$$\frac{V(d)(1-\varepsilon)^d}{V(d)} = (1-\varepsilon)^d < e^{-\varepsilon d}$$

- Thus, taking $\varepsilon = \frac{2 \ln n}{d}$ we have that for each i

$$\mathcal{P}\left(|x_i| < 1 - \frac{2 \ln n}{d}\right) < e^{-\frac{2\ln n}{d} d} = e^{-\ln n^2} = \frac{1}{n^2}$$

- Thus, the probability that $|x_i| < 1 - \frac{2 \ln n}{d}$ for at least one i is less than $n \cdot \frac{1}{n^2} = \frac{1}{n}$. This reasoning is called the *Union bound.*

- *Union bound:* if you have sets $A_1, \ldots, A_n$ (not necessarily disjoint) such that $\mathcal{P}(x \in A_i) = p_i$ then $\mathcal{P}\left(x \in \cup_{i=1}^n A_i\right) \le \sum_{i=1}^n p_i$.

- **Proof:** Probability that $|x_i| < 1 - \varepsilon$ is

$$\frac{V(d)(1-\varepsilon)^d}{V(d)} = (1-\varepsilon)^d < e^{-\varepsilon d}$$

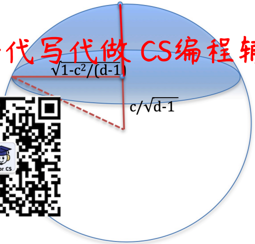
- Thus, taking $\varepsilon = \frac{2 \ln n}{d}$ we get that for each i

$$\mathcal{P}\left(|x_i| < 1 - \frac{2 \ln n}{d}\right) < e^{-\frac{2 \ln n}{d} d} = e^{-\ln n^2} = \frac{1}{n^2}$$

- Thus, the probability that $|x_i| < 1 - \frac{2 \ln n}{d}$ for at least one i is less than $n \cdot \frac{1}{n^2} = \frac{1}{n}$. This reasoning is called the *Union bound.*

- *Union bound:* if you have n sets $A_1, \ldots, A_n$ (not necessarily disjoint) such that $\mathcal{P}(x \in A_i) = p_i$ then $\mathcal{P}\left(x \in \cup_{i=1}^n A_i\right) \leq \sum_{i=1}^n p_i$.

- To prove the second claim we rotate the sphere so that x_j is the polar axis.



- We now note that, since $|x_j| \le 1$

$$|\langle x_i, x_j\rangle| = |x_i|\,|x_j|\,|\cos\angle(x_i,x_j)| \le |x_i|\,|\cos\angle(x_i,x_j)|$$

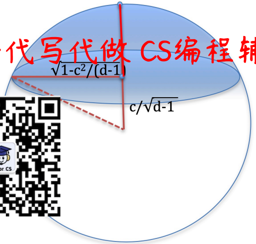
- Thus, $|\langle x_i, x_j\rangle| \ge \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ implies that the length of the orthogonal projection of x_i onto x_j, i.e., $|x_i||\cos\angle(x_i,x_j)|$ is larger than $\frac{\sqrt{6\ln n}}{\sqrt{d-1}}$.

- But we have seen in the previous slide that if we pick a point x_i uniformly at random a point from the unit ball, the probability that its coordinate x_j satisfies $|x_j| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- Letting $c = \sqrt{6\ln n}$, we obtain that for every pair i, j

$$\mathcal{P}\left(|\langle x_i, x_j\rangle| \ge \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \le \frac{2}{\sqrt{6\ln n}}e^{-\frac{6\ln n}{2}} = \frac{2}{\sqrt{6\ln n}}e^{-\ln n^3}$$

- To prove the second claim we rotate the sphere so that x_j is the polar axis.



- We now note that, since $|x_i|, |x_j| \leq 1$,

$$|\langle x_i, x_j\rangle| = |x_i|\,|x_j|\,|\cos\angle(x_i, x_j)| \leq |x_i||\cos\angle(x_i, x_j)|$$

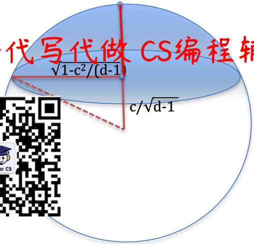
- Thus, $[|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ implies that the length of the orthogonal projection of x_i onto x_j, i.e., $|x_i||\cos\angle(x_i, x_j)|$ is larger than $\frac{\sqrt{6\ln n}}{\sqrt{d-1}}$.

- But we have seen in the previous corollary that if we pick a point x_i uniformly at random a point from the unit ball, the probability that its coordinate x_j satisfies $|x_j| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- Letting $c = \sqrt{6\ln n}$, we obtain that for every pair i, j

$$\mathcal{P}\left(|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6\ln n}}e^{-\frac{6\ln n}{2}} = \frac{2}{\sqrt{6\ln n}}e^{-\ln n^3}$$

- To prove the second claim we rotate the sphere so that x_j is the polar axis.



- We now note that, since $|x_i|, |x_j| \leq 1$,

$$|\langle x_i, x_j\rangle| = |x_i|\,|x_j|\,|\cos\angle(x_i,x_j)| \leq |x_i||\cos\angle(x_i,x_j)|$$

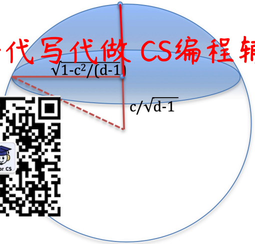
- Thus, $[|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ implies that the length of the orthogonal projection of x_i onto x_j, i.e., $|x_i||\cos\angle(x_i,x_j)|$ is larger than $\frac{\sqrt{6\ln n}}{\sqrt{d-1}}$.

- But we have seen in the previous corollary that if we pick a point x_i uniformly at random a point from the unit ball, the probability that its coordinate x_j satisfies $|x_j| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- Letting $c = \sqrt{6\ln n}$, we obtain that for every pair i, j

$$\mathcal{P}\left(|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6\ln n}}e^{-\frac{6\ln n}{2}} = \frac{2}{\sqrt{6\ln n}}e^{-\ln n^3}$$

- To prove the second claim we rotate the sphere so that x_j is the polar axis.



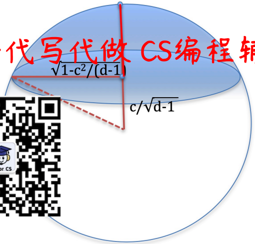
- We now note that, since $|x_i|, |x_j| \leq 1$,

$$|\langle x_i, x_j\rangle| = |x_i|\,|x_j|\,|\cos\angle(x_i, x_j)| \leq |x_i||\cos\angle(x_i, x_j)|$$

- Thus, $[|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ implies that the length of the orthogonal projection of x_i onto x_j, i.e., $|x_i||\cos\angle(x_i, x_j)|$ is larger than $\frac{\sqrt{6\ln n}}{\sqrt{d-1}}$.
- But we have seen in the previous corollary that if we pick a point x_i uniformly at random a point from the unit ball, the probability that its coordinate x_j satisfies $|x_j| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.
- Letting $c = \sqrt{6\ln n}$, we obtain that for every pair i, j

$$\mathcal{P}\left(|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6\ln n}}e^{-\frac{6\ln n}{2}} = \frac{2}{\sqrt{6\ln n}}e^{-\ln n^3}$$

- To prove the second claim we rotate the sphere so that x_j is the polar axis.



- We now note that, since $|x_i|, |x_j| \leq 1$,

$$|\langle x_i, x_j\rangle| = |x_i|\,|x_j|\,|\cos\angle(x_i, x_j)| \leq |x_i||\cos\angle(x_i, x_j)|$$

- Thus, $[|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ implies that the length of the orthogonal projection of x_i onto x_j, i.e., $|x_i||\cos\angle(x_i, x_j)|$ is larger than $\frac{\sqrt{6\ln n}}{\sqrt{d-1}}$.

- But we have seen in the previous corollary that if we pick a point x_i uniformly at random a point from the unit ball, the probability that its coordinate x_j satisfies $|x_j| > \frac{c}{\sqrt{d-1}}$ is less than $\frac{2}{c}e^{-\frac{c^2}{2}}$.

- Letting $c = \sqrt{6\ln n}$, we obtain that for every pair i, j

$$\mathcal{P}\left(|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6\ln n}}e^{-\frac{6\ln n}{2}} = \frac{2}{\sqrt{6\ln n}}e^{-\ln n^3}$$

- Simplifying the expression $\frac{2}{\sqrt{6\ln n}} e^{-\ln n^3}$ we obtain

$$\mathcal{P}\left(|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6\ln n}}\frac{1}{n^3} < \frac{1}{n^3}.$$

- There are $\binom{n}{2}$ pairs, so by the union bound, the probability that for at least one pair x_i, x_j the scalar product is large, i.e., $|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ is at most $\binom{n}{2}\frac{1}{n^3} = \frac{n(n-1)}{2}\frac{1}{n^3} < \frac{1}{2n}$.

- Thus, all inequalities of both kinds hold with a probability $(1-\frac{1}{n})(1-\frac{1}{2n}) = 1 - \frac{1}{n} - \frac{1}{2n} + \frac{1}{2n^2} > 1 - \frac{3}{2}\frac{1}{n}$.

- Thus, to summarise, we have proved that, if we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball, then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,

  1. $|x_i| > 1 - \frac{2\ln n}{d}$;

  2. $|\langle x_i, x_j\rangle| \leq \sqrt{\frac{6\ln n}{d-1}}$

- But how do we draw random points from a unit ball?
- It turns out that it is easier to drew points using a spherical Gaussian which is what we do next.

- Simplifying the expression $\frac{2}{\sqrt{6\ln n}}e^{-\ln n^3}$ we obtain

$$\mathcal{P}\left(|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6\ln n}}\frac{1}{n^3} < \frac{1}{n^3}.$$

- There are $\binom{n}{2}$ pairs, so by the union bound, the probability that for at least one pair x_i, x_j the scalar product is large, i.e., $|\langle x_i, x_j\rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ is at most $\binom{n}{2}\frac{1}{n^3} = \frac{n(n-1)}{2}\frac{1}{n^3} < \frac{1}{2n}$.
- Thus, all inequalities of both kinds hold with a probability $(1-\frac{1}{n})(1-\frac{1}{2n}) = 1 - \frac{1}{n} - \frac{1}{2n} + \frac{1}{2n^2} > 1 - \frac{3}{2}\frac{1}{n}$.
- Thus, to summarise, we have proved that, if we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball, then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $|x_i| > 1 - \frac{2\ln n}{d}$;
  2. $|\langle x_i, x_j\rangle| \leq \sqrt{\frac{6\ln n}{d-1}}$
- But how do we draw random points from a unit ball?
- It turns out that it is easier to drew points using a spherical Gaussian which is what we do next.

- Simplifying the expression $\frac{2}{\sqrt{6\ln n}} e^{-\ln n^3}$ we obtain

$$\mathcal{P}\left(|\langle x_i, x_j \rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}} \right) \leq \frac{2}{\sqrt{6\ln n}} \frac{1}{n^3} < \frac{1}{n^3}.$$

- There are $\binom{n}{2}$ pairs, so by the union bound, the probability that for at least one pair x_i, x_j the scalar product is large, i.e., $|\langle x_i, x_j \rangle| \geq \frac{\sqrt{6\ln n}}{\sqrt{d-1}}$ is at most $\binom{n}{2} \frac{1}{n^3} = \frac{n(n-1)}{2} \frac{1}{n^3} < \frac{1}{2n}$.

- Thus, all inequalities of both kinds hold with a probability $(1 - \frac{1}{n})(1 - \frac{1}{2n}) = 1 - \frac{1}{n} - \frac{1}{2n} + \frac{1}{2n^2} > 1 - \frac{3}{2}\frac{1}{n}$

- Thus, to summarise, we have proved that, if we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball, then with probability $1 - \frac{3}{2}\frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,

  1. $|x_i| > 1 - \frac{2\ln n}{d}$;

  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6\ln n}{d-1}}$.

- But how do we draw random points from a unit ball?
- It turns out that it is easier to drew points using a spherical Gaussian which is what we do next.

- Simplifying the expression $\frac{2}{\sqrt{6 \ln n}} e^{-\ln n^3}$ we obtain

$$\mathcal{P}\left(|\langle x_i, x_j \rangle| \geq \frac{\sqrt{6 \ln n}}{\sqrt{d-1}}\right) \leq \frac{2}{\sqrt{6 \ln n}} \frac{1}{n^3} < \frac{1}{n^3}.$$

- There are $\binom{n}{2}$ pairs, so by the union bound, the probability that for at least one pair x_i, x_j the scalar product is large, i.e., $|\langle x_i, x_j \rangle| \geq \frac{\sqrt{6 \ln n}}{\sqrt{d-1}}$ is at most $\binom{n}{2} \frac{1}{n^3} = \frac{n(n-1)}{2} \frac{1}{n^3} < \frac{1}{2n}$.
- Thus, all inequalities of both kinds hold with a probability $(1 - \frac{1}{n})(1 - \frac{1}{2n}) = 1 - \frac{1}{n} - \frac{1}{2n} + \frac{1}{2n^2} > 1 - \frac{3}{2} \frac{1}{n}$
- Thus, to summarise, we have proved that, if we draw independently and uniformly n points $x_1, \ldots, x_n$ from the unit ball, then with probability $1 - \frac{3}{2} \frac{1}{n}$ for all $1 \leq i, j \leq n$, $i \neq j$,
  1. $|x_i| > 1 - \frac{2 \ln n}{d}$;
  2. $|\langle x_i, x_j \rangle| \leq \sqrt{\frac{6 \ln n}{d-1}}$.
- But how do we draw random points from a unit ball?
- It turns out that it is easier to drew points using a spherical Gaussian which is what we do next.

Gaussian Annulus Theorem

- We need the following lemma which we do not prove; its proof is based on tail inequalities which can be found in the BHK book. Recal that for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ [illegible] $= \sqrt{x_1^2 + \ldots + x_d^2}$.
- **Lemma:** Let X be a [illegible]onal spherical Gaussian with zero mean and unit variance in each c[illegible]e., the with the probability density

$$f_X(x) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2}}$$

If we draw from such a distribution a random vector x, then the probability that x belongs to the annulus (hollow ball of outer radius $\sqrt{d} + \beta$ with the shell of thickness 2β)

$$\{x : \sqrt{d} - \beta \leq |x| \leq \sqrt{d} + \beta\} = \left\{x : \left||x| - \sqrt{d}\right| < \beta\right\}$$

is at least $1 - 3e^{-\frac{\beta^2}{96}}$.

- (For convenience of notation we denote $1/96$ by γ, so the probability that x belongs to the annulus is at least $1 - 3e^{-\gamma\beta^2}$).

Gaussian Annulus Theorem

- We need the following lemma which we do not prove; its proof is based on tail inequalities which can be found in the BHK book. Recal that for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ [illegible] $\sqrt{x_1^2 + \ldots + x_d^2}$.
- **Lemma:** Let X be a [illegible]onal spherical Gaussian with zero mean and unit variance in each c[illegible]., the with the probability density

$$f_X(x) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2}}$$

If we draw from such a distribution a random vector x, then the probability that x belongs to the annulus (hollow ball of outer radius $\sqrt{d} + \beta$ with the shell of thickness 2β)

$$\{x : \sqrt{d} - \beta \leq |x| \leq \sqrt{d} + \beta\} = \left\{x : \left||x| - \sqrt{d}\right| < \beta\right\}$$

is at least $1 - 3e^{-\frac{\beta^2}{96}}$.

- (For convenience of notation we denote $1/96$ by γ, so the probability that x belongs to the annulus is at least $1 - 3e^{-\gamma\beta^2}$).

Gaussian Annulus Theorem

- We need the following lemma which we do not prove; its proof is based on tail inequalities which can be found in the BHK book. Recal that for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ [illegible] $\sqrt{x_1^2 + \ldots + x_d^2}$.
- **Lemma:** Let X be a [illegible]onal spherical Gaussian with zero mean and unit variance in each c[illegible]e., the with the probability density

$$f_X(x) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2}}$$

If we draw from such a distribution a random vector x, then the probability that x belongs to the annulus (hollow ball of outer radius $\sqrt{d} + \beta$ with the shell of thickness 2β)

$$\{x : \sqrt{d} - \beta \leq |x| \leq \sqrt{d} + \beta\} = \left\{x : \left| |x| - \sqrt{d} \right| < \beta \right\}$$

is at least $1 - 3e^{-\frac{\beta^2}{96}}$.

- (For convenience of notation we denote $1/96$ by γ, so the probability that x belongs to the annulus is at least $1 - 3e^{-\gamma\beta^2}$).

- We need the following lemma which we do not prove; its proof is based on tail inequalities which can be found in the BHK book. Recal that for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$.
- **Lemma:** Let X be a d-dimensional spherical Gaussian with zero mean and unit variance in each direction, i.e., the with the probability density

$$f_X(x) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2}}$$

If we draw from such a distribution a random vector x, then the probability that x belongs to the annulus (hollow ball of outer radius $\sqrt{d} + \beta$ with the shell of thickness 2β)

$$\{x : \sqrt{d} - \beta \leq |x| \leq \sqrt{d} + \beta\} = \left\{x : \left||x| - \sqrt{d}\right| < \beta\right\}$$

is at least $1 - 3e^{-\frac{\beta^2}{96}}$.
- (For convenience of notation we denote $1/96$ by γ, so the probability that x belongs to the annulus is at least $1 - 3e^{-\gamma\beta^2}$).

- Let us choose k random vectors $u_1, \dots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).

- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1\rangle, \dots, \langle v, u_k\rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k}|v| \Big| \geq \varepsilon\sqrt{k}|v|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus, we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k} \Big| \geq \varepsilon\sqrt{k}\right) \leq 3e^{-\gamma k \varepsilon^2}$$

- Note that $\langle u_i, v\rangle = \sum_{j=1}^d u_{ij} v_j$; since u_{ij} are Gaussians, so is their linear combination; thus, $\langle u_i, v\rangle$ is also Gaussian.

- Since u_{ij} are of unit variance and independent, we obtain

$$V\left[\langle u_i, v\rangle\right] = V\left[\sum_{j=1}^d u_{i,j} v_j\right] = \sum_{j=1}^d V[u_{i,j}]v_j^2 = \sum_{j=1}^d v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v\rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \ldots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).

- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1 \rangle, \ldots, \langle v, u_k \rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k}|v| \Big| \geq \varepsilon \sqrt{k}|v| \right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k} \Big| \geq \varepsilon \sqrt{k} \right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- Note that $\langle u_i, v \rangle = \sum_{j=1}^{d} u_{ij} v_j$; since u_{ij} are Gaussians, so is their linear combination; thus, $\langle u_i, v \rangle$ is also Gaussian.
- Since u_{ij} are of unit variance and independent, we obtain

$$V\left[\langle u_i, v \rangle\right] = V\left[\sum_{j=1}^{d} u_{i,j} v_j\right] = \sum_{j=1}^{d} V[u_{i,j}] v_j^2 = \sum_{j=1}^{d} v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v \rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \ldots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).
- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1\rangle, \ldots, \langle v, u_k\rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big|\,|f(v)| - \sqrt{k}|v|\,\Big| \geq \varepsilon\sqrt{k}|v|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big|\,|f(v)| - \sqrt{k}\,\Big| \geq \varepsilon\sqrt{k}\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- Note that $\langle u_i, v\rangle = \sum_{j=1}^{d} u_{ij} v_j$; since u_{ij} are Gaussians, so is their linear combination; thus, $\langle u_i, v\rangle$ is also Gaussian.

- Since u_{ij} are of unit variance and independent, we obtain

$$V\left[\langle u_i, v\rangle\right] = V\left[\sum_{j=1}^{d} u_{i,j} v_j\right] = \sum_{j=1}^{d} V[u_{i,j}] v_j^2 = \sum_{j=1}^{d} v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v\rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \ldots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).

- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1\rangle, \ldots, \langle v, u_k\rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k}|v| \Big| \geq \varepsilon\sqrt{k}|v|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus, we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k} \Big| \geq \varepsilon\sqrt{k}\right) \leq 3e^{-\gamma k \varepsilon^2}$$

- Note that $\langle u_i, v\rangle = \sum_{j=1}^{d} u_{i,j} v_j$; since $u_{i,j}$ are Gaussians, so is their linear combination; thus, $\langle u_i, v\rangle$ is also Gaussian.
- Since u_{ij} are of unit variance and independent, we obtain

$$V[\langle u_i, v\rangle] = V\left[\sum_{j=1}^{d} u_{i,j} v_j\right] = \sum_{j=1}^{d} V[u_{i,j}] v_j^2 = \sum_{j=1}^{d} v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v\rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \dots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).
- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1\rangle, \dots, \langle v, u_k\rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big|\,|f(v)| - \sqrt{k}|v|\,\Big| \geq \varepsilon\sqrt{k}|v|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus, we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big|\,|f(v)| - \sqrt{k}\,\Big| \geq \varepsilon\sqrt{k}\right) \leq 3e^{-\gamma k \varepsilon^2}$$

- Note that $\langle u_i, v\rangle = \sum_{j=1}^d u_{i,j} v_j$; since $u_{i,j}$ are Gaussians, so is their linear combination; thus, $\langle u_i, v\rangle$ is also Gaussian.
- Since u_{ij} are of unit variance and independent, we obtain

$$V\left[\langle u_i, v\rangle\right] = V\left[\sum_{j=1}^d u_{i,j} v_j\right] = \sum_{j=1}^d V[u_{i,j}]v_j^2 = \sum_{j=1}^d v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v\rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \ldots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).
- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1 \rangle, \ldots, \langle v, u_k \rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k}|v| \Big| \geq \varepsilon \sqrt{k}|v| \right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus, we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k} \Big| \geq \varepsilon \sqrt{k} \right) \leq 3e^{-\gamma k \varepsilon^2}$$

- Note that $\langle u_i, v \rangle = \sum_{j=1}^{d} u_{i,j} v_j$; since $u_{i,j}$ are Gaussians, so is their linear combination; thus, $\langle u_i, v \rangle$ is also Gaussian.
- Since u_{ij} are of unit variance and independent, we obtain

$$V\left[\langle u_i, v \rangle\right] = V\left[\sum_{j=1}^{d} u_{i,j} v_j\right] = \sum_{j=1}^{d} V[u_{i,j}] v_j^2 = \sum_{j=1}^{d} v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v \rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \dots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).

- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1 \rangle, \dots, \langle v, u_k \rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k}|v| \Big| \geq \varepsilon \sqrt{k}|v| \right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus, we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k} \Big| \geq \varepsilon \sqrt{k} \right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- Note that $\langle u_i, v \rangle = \sum_{j=1}^{d} u_{i,j} v_j$; since u_{ij} are Gaussians, so is their linear combination; thus, $\langle u_i, v \rangle$ is also Gaussian.

- Since u_{ij} are of unit variance and independent, we obtain

$$V[\langle u_i, v \rangle] = V\left[\sum_{j=1}^{d} u_{i,j} v_j\right] = \sum_{j=1}^{d} V[u_{i,j}] v_j^2 = \sum_{j=1}^{d} v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v \rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \ldots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).
- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1\rangle, \ldots, \langle v, u_k\rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k}|v| \Big| \geq \varepsilon\sqrt{k}|v|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus, we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\Big| |f(v)| - \sqrt{k} \Big| \geq \varepsilon\sqrt{k}\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- Note that $\langle u_i, v\rangle = \sum_{j=1}^{d} u_{i,j} v_j$; since $u_{i,j}$ are Gaussians, so is their linear combination; thus, $\langle u_i, v\rangle$ is also Gaussian.
- Since u_{ij} are of unit variance and independent, we obtain

$$V\left[\langle u_i, v\rangle\right] = V\left[\sum_{j=1}^{d} u_{i,j} v_j\right] = \sum_{j=1}^{d} V[u_{i,j}]v_j^2 = \sum_{j=1}^{d} v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v\rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us choose k random vectors $u_1, \ldots, u_k \in \mathbb{R}^d$ by sampling randomly a spherical Gaussian with a unit variance (i.e., we generate k vectors with each of d many coordinatse obtained by sampling randomly and independently the one dimensional Gaussian).
- For any vector v define $f(v)$, called *the random projection*, by

$$f(v) = (\langle v, u_1 \rangle, \ldots, \langle v, u_k \rangle)$$

- **Theorem: (Random Projection Lemma)**

$$\mathcal{P}\left(\left| |f(v)| - \sqrt{k}|v| \right| \geq \varepsilon\sqrt{k}|v|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- **Proof:** Since $f(v)$ is a linear operator in v, we can divide both sides of the inner inequality with $|v|$. Thus, we can assume that $|v| = 1$ and we have to show that for $|v| = 1$

$$\mathcal{P}\left(\left| |f(v)| - \sqrt{k} \right| \geq \varepsilon\sqrt{k}\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- Note that $\langle u_i, v \rangle = \sum_{j=1}^{d} u_{i,j} v_j$; since $u_{i,j}$ are Gaussians, so is their linear combination; thus, $\langle u_i, v \rangle$ is also Gaussian.
- Since u_{ij} are of unit variance and independent, we obtain

$$V\left[\langle u_i, v \rangle\right] = V\left[\sum_{j=1}^{d} u_{i,j} v_j\right] = \sum_{j=1}^{d} V[u_{i,j}] v_j^2 = \sum_{j=1}^{d} v_j^2 = |v| = 1.$$

- Thus, $\langle u_i, v \rangle$ is also Gaussian of unit variance.
- It follows that the theorem is a direct consequence of the Gaussian Annulus Theorem with $\beta = \varepsilon\sqrt{k}$.

- Let us define $f^*(v) = f(v)/\sqrt{k}$;
- if we divide by $\sqrt{k}$ both sides of the inner inequality in

$$\mathcal{P}\left(\Big|\,|f(v)| - \sqrt{k}|v|\,\Big| \geq \varepsilon\sqrt{k}|v|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

  we obtain the following consequence:

- **Corollary:**

$$\mathcal{P}(|\,|f^*(v)| - |v|\,| \geq \varepsilon|v|) \leq 3e^{-\gamma\, k\, \varepsilon^2}$$

- Thus, the modified projection mapping $f^*(v)$ “almost preserves” length of vectors.
- This corollary is needed to prove the following important theorem with lots of practical applications.

- **Theorem: (Johnson-Lindenstrauss Lemma)** For any ε, $0 < \varepsilon < 1$, and any integer n, assume that k satisfies $k \geq \frac{3}{\gamma\,\varepsilon^2} \ln n$. Then for any set of n points given by vectors in $\mathbb{R}^d$, with the probability of at least $1 - \frac{3}{2n}$, the random pr $\mathbb{R}^d \to \mathbb{R}^k$ has the property that for ALL pairs of points v_i, v_j

$$||f^*(v_i - v_j)| - |v_i - v_j|| \leq \varepsilon |v_i - v_j|$$

Thus, $f^*(v)$ "almost" preserves distances between points given by vectors v_i, despite reduction of dimensionality from $d >> k$ to only k.

- **Proof:** By applying the Random Projection Lemma, we get that for any pair v_i, v_j,

$$\mathcal{P}\left(||f^*(v_i - v_j)| - |v_i - v_j|| > \varepsilon |v_i - v_j|\right) \leq 3e^{-\gamma\,k\,\varepsilon^2} \leq 3e^{-\gamma \frac{3}{\gamma\,\varepsilon^2} \ln n\,\varepsilon^2}$$
$$\leq 3e^{-\ln n^3} = \frac{3}{n^3}$$

- Since there are $\binom{n}{2} < \frac{n^2}{2}$ pairs, the probability that at least one of the above inequalities fails is less than $\frac{n^2}{2} \cdot \frac{3}{n^3} = 1 - \frac{3}{2n}$. (Union bound again).
- Note that k is a linear function of $\ln n$, so it grows slowly, allowing to handle efficiently lots of points v_i.

Johnson-Lindenstrauss Lemma

- **Theorem: (Johnson-Lindenstrauss Lemma)** For any ε, $0 < \varepsilon < 1$, and any integer n, assume that k satisfies $k \geq \frac{3}{\gamma\,\varepsilon^2} \ln n$. Then for any set of n points given by vectors [illegible] in $\mathbb{R}^d$, with the probability of at least $1 - \frac{3}{2n}$, the random pr[illegible] $\mathbb{R}^d \to \mathbb{R}^k$ has the property that for ALL pairs of points v_i, v_j

$$||f^*(v_i - v_j)| - |v_i - v_j|| \leq \varepsilon |v_i - v_j|$$

Thus, $f^*(v)$ "almost" preserves distances between points given by vectors v_i, despite reduction of dimensionality from $d >> k$ to only k.

- **Proof:** By applying the Random Projection Lemma, we get that for any pair v_i, v_j,

$$\mathcal{P}\left(||f^*(v_i - v_j)| - |v_i - v_j|| > \varepsilon |v_i - v_j|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2} \leq 3e^{-\gamma \frac{3}{\gamma\,\varepsilon^2} \ln n\, \varepsilon^2}$$

$$\leq 3e^{-\ln n^3} = \frac{3}{n^3}$$

- Since there are $\binom{n}{2} < \frac{n^2}{2}$ pairs, the probability that at least one of the above inequalities fails is less than $\frac{n^2}{2} \cdot \frac{3}{n^3} = 1 - \frac{3}{2n}$. (Union bound again).
- Note that k is a linear function of $\ln n$, so it grows slowly, allowing to handle efficiently lots of points v_i.

- **Theorem: (Johnson-Lindenstrauss Lemma)** For any ε, $0 < \varepsilon < 1$, and any integer n, assume that k satisfies $k \geq \frac{3}{\gamma\,\varepsilon^2} \ln n$. Then for any set of n points given by vectors $v_1, \ldots, v_n$ in $\mathbb{R}^d$, with the probability of at least $1 - \frac{3}{2n}$, the random projection $f^* : \mathbb{R}^d \to \mathbb{R}^k$ has the property that for ALL pairs of points v_i, v_j

$$||f^*(v_i - v_j)| - |v_i - v_j|| \leq \varepsilon |v_i - v_j|$$

Thus, $f^*(v)$ "almost" preserves distances between points given by vectors v_i, despite reduction of dimensionality from $d >> k$ to only k.

- **Proof:** By applying the Random Projection Lemma, we get that for any pair v_i, v_j,

$$\mathcal{P}\left(||f^*(v_i - v_j)| - |v_i - v_j|| > \varepsilon |v_i - v_j|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2} \leq 3e^{-\gamma \frac{3}{\gamma\,\varepsilon^2} \ln n\, \varepsilon^2}$$
$$\leq 3e^{-\ln n^3} = \frac{3}{n^3}$$

- Since there are $\binom{n}{2} < \frac{n^2}{2}$ pairs, the probability that at least one of the above inequalities fails is less than $\frac{n^2}{2} \cdot \frac{3}{n^3} = 1 - \frac{3}{2n}$. (Union bound again).
- Note that k is a linear function of $\ln n$, so it grows slowly, allowing to handle efficiently lots of points v_i.

- **Theorem: (Johnson-Lindenstrauss Lemma)** For any ε, $0 < \varepsilon < 1$, and any integer n, assume that k satisfies $k \geq \frac{3}{\gamma\,\varepsilon^2} \ln n$. Then for any set of n points given by vectors $v_1, \ldots, v_n$ in $\mathbb{R}^d$, with the probability of at least $1 - \frac{3}{2n}$, the random projection $f^* : \mathbb{R}^d \to \mathbb{R}^k$ has the property that for ALL pairs of points v_i, v_j

$$||f^*(v_i - v_j)| - |v_i - v_j|| \leq \varepsilon |v_i - v_j|$$

Thus, $f^*(v)$ "almost" preserves distances between points given by vectors v_i, despite reduction of dimensionality from $d >> k$ to only k.

- **Proof:** By applying the Random Projection Lemma, we get that for any pair v_i, v_j,

$$\mathcal{P}\left(||f^*(v_i - v_j)| - |v_i - v_j|| > \varepsilon|v_i - v_j|\right) \leq 3e^{-\gamma\, k\, \varepsilon^2} \leq 3e^{-\gamma \frac{3}{\gamma\,\varepsilon^2} \ln n\, \varepsilon^2}$$

$$\leq 3e^{-\ln n^3} = \frac{3}{n^3}$$

- Since there are $\binom{n}{2} < \frac{n^2}{2}$ pairs, the probability that at least one of the above inequalities fails is less than $\frac{n^2}{2} \cdot \frac{3}{n^3} = 1 - \frac{3}{2n}$. (Union bound again).
- Note that k is a linear function of $\ln n$, so it grows slowly, allowing to handle efficiently lots of points v_i.

- Main application of the Johnson-Lindenstrauss Lemma: nearest neighbour search in spaces of extremely high dimension, such as collections of documents, represented by the vector of the relative frequencies of each word from a dictionary (of key words, for example).

- Given a database of a large number of papers and a new paper p, find a paper p_i from the database which is the "most similar" to p, in the sense that the same keywords appear with approximately same frequencies.

- The dictionary of all keywords is very large, in thousands or even tens of thousands, so nearest neighbour search would be extremely slow.

- Even if your database had 100,000 entries, $\ln(100,000) < 12$

- Main application of the Johnson-Lindenstrauss Lemma: nearest neighbour search in spaces of extremely high dimension, such as collections of documents, represented by the vector of the relative frequencies of each word from a dictionary (of key words, for example).

- Given a database of a large number of papers and a new paper p, find a paper p_i from the database which is the "most similar" to p, in the sense that the same keywords appear with approximately same frequencies.

- The dictionary of all keywords is very large, in thousands or even tens of thousands, so nearest neighbour search would be extremely slow.

- Even if your database had 100,000 entries, $\ln(100,000) < 12$

Johnson-Lindenstrauss Lemma

- Main application of the Johnson-Lindenstrauss Lemma: nearest neighbour search in spaces of extremely high dimension, such as collections of documents, represented by the vector of the relative frequencies of each word from a dictionary (of key words, for example).
- Given a database of a large number of papers and a new paper p, find a paper p_i from the database which is the "most similar" to p, in the sense that the same keywords appear with approximately same frequencies.
- The dictionary of all keywords is very large, in thousands or even tens of thousands, so nearest neighbour search would be extremely slow.
- Even if your database had 100,000 entries, $\ln(100,000) < 12$

- Main application of th[illegible]Lindenstrauss Lemma: nearest neighbour search in spaces of ext[illegible] dimension, such as collections of documents, represented by the vector of the relative frequencies of each word from a dictionary (of key words, for example).
- Given a database of a large number of papers and a new paper p, find a paper p_i from the database which is the "most similar" to p, in the sense that the same keywords appear with approximately same frequencies.
- The dictionary of all keywords is very large, in thousands or even tens of thousands, so nearest neighbour search would be extremely slow.
- Even if your database had 100,000 entries, $\ln(100,000) < 12$

- Solution: choose a large enough k, i.e., $k \geq \frac{288}{\varepsilon^2}\ln(|DB| + |Q|)$ where $|DB|$ is the size of the database and $|Q|$ the expected number of queries to be made.
- Recall that $\gamma = 1/96$ … 288. However, with more careful analysis, the requirement $k \geq$ … can be reduced to $k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3}\ln n$.
- A good dictionary contains more than 100,000 words.
- So if you want to compare English documents according to the relative frequency of words, each document has to be represented as a vector in $100,000$ dimensional space!
- Assume that your database has 100,000 documents stored; you want to find the most similar documents in your database for a batch of $10,000$ new documents. How many random vectors should you choose to project onto?
- If we set $\varepsilon = 0.1$ we obtain

$$k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3}\ln(|DB| + |Q|) = \frac{24}{3 \times .01 - 2 \times .001} \times \ln(110,000) = 9950,$$

so more than 10-folds reduction of dimensionality.

- Solution: choose a large enough k, i.e., $k \geq \frac{288}{\varepsilon^2} \ln(|DB| + |Q|)$ where $|DB|$ is the size of the database and $|Q|$ the expected number of queries to be made.
- Recall that $\gamma = 1/96$ [illegible] 288. However, with more careful analysis, the requirement $k \geq$ [illegible] can be reduced to $k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln n$.
- A good dictionary contains more than 100,000 words.
- So if you want to compare English documents according to the relative frequency of words, each document has to be represented as a vector in $100,000$ dimensional space!
- Assume that your database has 100,000 documents stored; you want to find the most similar documents in your database for a batch of $10,000$ new documents. How many random vectors should you choose to project onto?
- If we set $\varepsilon = 0.1$ we obtain

$$k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln(|DB| + |Q|) = \frac{24}{3 \times .01 - 2 \times .001} \times \ln(110,000) = 9950,$$

so more than 10-folds reduction of dimensionality.

Applications of the Johnson-Lindenstrauss Lemma

- Solution: choose a large enough k, i.e., $k \geq \frac{288}{\varepsilon^2} \ln(|DB| + |Q|)$ where $|DB|$ is the size of the database and $|Q|$ the expected number of queries to be made.
- Recall that $\gamma = 1/96$ … 288. However, with more careful analysis, the requirement $k \geq$ … can be reduced to $k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln n$.
- A good dictionary contains more than 100,000 words.
- So if you want to compare English documents according to the relative frequency of words, each document has to be represented as a vector in $100,000$ dimensional space!
- Assume that your database has 100,000 documents stored; you want to find the most similar documents in your database for a batch of $10,000$ new documents. How many random vectors should you choose to project onto?
- If we set $\varepsilon = 0.1$ we obtain

$$k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln(|DB| + |Q|) = \frac{24}{3 \times .01 - 2 \times .001} \times \ln(110,000) = 9950,$$

so more than 10-folds reduction of dimensionality.

Applications of the Johnson-Lindenstrauss Lemma

- Solution: choose a large enough k, i.e., $k \geq \frac{288}{\varepsilon^2} \ln(|DB| + |Q|)$ where $|DB|$ is the size of the database and $|Q|$ the expected number of queries to be made.
- Recall that $\gamma = 1/96$ [illegible] 288. However, with more careful analysis, the requirement $k \geq$ [illegible] n be reduced to $k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln n$.
- A good dictionary contains more than 100,000 words.
- So if you want to compare English documents according to the relative frequency of words, each document has to be represented as a vector in $100,000$ dimensional space!
- Assume that your database has 100,000 documents stored; you want to find the most similar documents in your database for a batch of $10,000$ new documents. How many random vectors should you choose to project onto?
- If we set $\varepsilon = 0.1$ we obtain

$$k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln(|DB| + |Q|) = \frac{24}{3 \times .01 - 2 \times .001} \times \ln(110,000) = 9950,$$

so more than 10-folds reduction of dimensionality.

Applications of the Johnson-Lindenstrauss Lemma

- Solution: choose a large enough k, i.e., $k \geq \frac{288}{\varepsilon^2}\ln(|DB| + |Q|)$ where $|DB|$ is the size of the database and $|Q|$ the expected number of queries to be made.
- Recall that $\gamma = 1/96$ … 288. However, with more careful analysis, the requirement $k \geq$ … can be reduced to $k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3}\ln n$.
- A good dictionary contains more than 100,000 words.
- So if you want to compare English documents according to the relative frequency of words, each document has to be represented as a vector in $100,000$ dimensional space!
- Assume that your database has 100,000 documents stored; you want to find the most similar documents in your database for a batch of $10,000$ new documents. How many random vectors should you choose to project onto?
- If we set $\varepsilon = 0.1$ we obtain

$$k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3}\ln(|DB| + |Q|) = \frac{24}{3 \times .01 - 2 \times .001} \times \ln(110,000) = 9950,$$

so more than 10-folds reduction of dimensionality.

Applications of the Johnson-Lindenstrauss Lemma

- Solution: choose a large enough k, i.e., $k \geq \frac{288}{\varepsilon^2} \ln(|DB| + |Q|)$ where $|DB|$ is the size of the database and $|Q|$ the expected number of queries to be made.
- Recall that $\gamma = 1/96$ … 288. However, with more careful analysis, the requirement $k \geq$ … can be reduced to $k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln n$.
- A good dictionary contains more than 100,000 words.
- So if you want to compare English documents according to the relative frequency of words, each document has to be represented as a vector in $100,000$ dimensional space!
- Assume that your database has 100,000 documents stored; you want to find the most similar documents in your database for a batch of $10,000$ new documents. How many random vectors should you choose to project onto?
- If we set $\varepsilon = 0.1$ we obtain

$$k \geq \frac{24}{3\varepsilon^2 - 2\varepsilon^3} \ln(|DB| + |Q|) = \frac{24}{3 \times .01 - 2 \times .001} \times \ln(110,000) = 9950,$$

so more than 10-folds reduction of dimensionality.

- So we choose k rando[illegible] by choosing each coordinate of every vector using a unit variance [illegible]
- Pre-process your database by replacing each vector $x_j \in DB$, ($x_j \in \mathbb{R}^d$) with its scaled projection $f^*(x_j) = f(x_j)/\sqrt{k}$ to the $k << d$ dimensional subspace spanned by these k random vectors.
- As each query y arrives also replace $y \in \mathbb{R}^d$ with its projection $f^*(y) = f(y)/\sqrt{k}$.
- Do the search for the nearest neighbour of $f(y)/\sqrt{k}$ in the projected k dimensional space instead of the search for the nearest neighbour of y in the space of dimension $d >> k$.

Applications of the Johnson-Lindenstrauss Lemma

- So we choose k random vectors by choosing each coordinate of every vector using a unit variance G
- Pre-process your database by replacing each vector $x_j \in DB$, ($x_j \in \mathbb{R}^d$) with its scaled projection $f^*(x_j) = f(x_j)/\sqrt{k}$ to the $k << d$ dimensional subspace spanned by these k random vectors.

- As each query y arrives also replace $y \in \mathbb{R}^d$ with its projection $f^*(y) = f(y)/\sqrt{k}$.
- Do the search for the nearest neighbour of $f(y)/\sqrt{k}$ in the projected k dimensional space instead of the search for the nearest neighbour of y in the space of dimension $d >> k$.

Applications of the Johnson-Lindenstrauss Lemma

- So we choose k randor … y choosing each coordinate of every vector using a unit variance (…
- Pre-process your database by replacing each vector $x_j \in DB$, ($x_j \in \mathbb{R}^d$) with its scaled projection $f^*(x_j) = f(x_j)/\sqrt{k}$ to the $k << d$ dimensional subspace spanned by these k random vectors.
- As each query y arrives also replace $y \in \mathbb{R}^d$ with its projection $f^*(y) = f(y)/\sqrt{k}$.
- Do the search for the nearest neighbour of $f(y)/\sqrt{k}$ in the projected k dimensional space instead of the search for the nearest neighbour of y in the space of dimension $d >> k$.

- So we choose k randor ... y choosing each coordinate of every vector using a unit variance G ...
- Pre-process your database by replacing each vector $x_j \in DB$, $(x_j \in \mathbb{R}^d)$ with its scaled projection $f^*(x_j) = f(x_j)/\sqrt{k}$ to the $k << d$ dimensional subspace spanned by these k random vectors.
- As each query y arrives, also replace $y \in \mathbb{R}^d$ with its projection $f^*(y) = f(y)/\sqrt{k}$.
- Do the search for the nearest neighbour of $f(y)/\sqrt{k}$ in the projected k dimensional space instead of the search for the nearest neighbour of y in the space of dimension $d >> k$.