COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor

Lecture 12

Distributional JL Lemma: Let $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1-\epsilon)\|y\|_2 \leq \|\boldsymbol{M}y\|_2 \leq (1+\epsilon)\|y\|_2.$$

Distributional JL Lemma: Let $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1-\epsilon)\|y\|_2 \leq \|\boldsymbol{M}y\|_2 \leq (1+\epsilon)\|y\|_2.$$

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$

DISTRIBUTIONAL JL PROOF

Distributional JL Lemma: Let $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1-\epsilon)\|y\|_2 \leq \|\boldsymbol{M}y\|_2 \leq (1+\epsilon)\|y\|_2.$$

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$
- For any j, $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$.

Distributional JL Lemma: Let $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1-\epsilon)\|y\|_2 \leq \|\boldsymbol{M}y\|_2 \leq (1+\epsilon)\|y\|_2.$$

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$
- For any j, $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$. Last time, we deduced that

$$\tilde{y}_j^2 = (y_1^2 + \ldots + y_d^2)/m = \|y\|_2^2/m \text{ and so } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

Distributional JL Lemma: Let $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1-\epsilon)\|y\|_2 \leq \|\boldsymbol{M}y\|_2 \leq (1+\epsilon)\|y\|_2.$$

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$
- For any j, $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$. Last time, we deduced that

$$\tilde{y}_j^2 = (y_1^2 + \ldots + y_d^2)/m = \|y\|_2^2/m \text{ and so } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

- Main idea: $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$ is a weighted sum of independent random variables each with mean 0 and variance $1/m$. Haven't yet used the fact we were using Gaussians and that $\tilde{y}_j^2 \sim \mathcal{N}(0, \|y\|_2^2/m)$.

DISTRIBUTIONAL JL PROOF

Distributional JL Lemma: Let $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ have independent $\mathcal{N}(0, 1/m)$ entries. If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for any $y \in \mathbb{R}^d$, with probability at least $1 - \delta$

$$(1-\epsilon)\|y\|_2 \leq \|\boldsymbol{M}y\|_2 \leq (1+\epsilon)\|y\|_2.$$

- Let $\tilde{y} = \boldsymbol{M}y$ and $\boldsymbol{M}_j$ be the j^{th} row of $\boldsymbol{M}$
- For any j, $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$. Last time, we deduced that

$$\tilde{y}_j^2 = (y_1^2 + \ldots + y_d^2)/m = \|y\|_2^2/m \text{ and so } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

- Main idea: $\tilde{y}_j = \langle \boldsymbol{M}_j, y \rangle$ is a weighted sum of independent random variables each with mean 0 and variance $1/m$. Haven't yet used the fact we were using Gaussians and that $\tilde{y}_j^2 \sim \mathcal{N}(0, \|y\|_2^2/m)$.
- Remains to show that $\|\tilde{y}\|_2^2$ is tightly concentrated around $\|y\|_2^2$.

So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with :

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with :

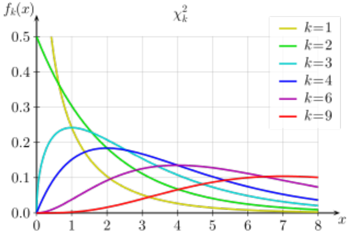
$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$\|\tilde{y}\|_2^2 = \sum_{i=1}^{m} \tilde{y}_j^2$ a Chi-Squared random variable with m degrees of freedom (a sum of m squared independent Gaussians)

So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with :

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$\|\tilde{y}\|_2^2 = \sum_{i=1}^{m} \tilde{y}_j^2$ a Chi-Squared random variable with m degrees of freedom (a sum of m squared independent Gaussians)



So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with :

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$\|\tilde{y}\|_2^2 = \sum_{i=1}^m \tilde{y}_j^2$ a Chi-Squared random variable with m degrees of freedom (a sum of m squared independent Gaussians)

Lemma: (Chi-Squared Concentration) Letting $\mathbf{Z}$ be a Chi-Squared random variable with m degrees of freedom,

$$\Pr\left[|\mathbf{Z} - \mathbb{E}\mathbf{Z}| \geq \epsilon \mathbb{E}\mathbf{Z}\right] \leq 2e^{-m\epsilon^2/8}.$$

So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with :

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$\|\tilde{y}\|_2^2 = \sum_{i=1}^m \tilde{y}_j^2$ a Chi-Squared random variable with m degrees of freedom (a sum of m squared independent Gaussians)

Lemma: (Chi-Squared Concentration) Letting $\mathbf{Z}$ be a Chi-Squared random variable with m degrees of freedom,

$$\Pr\left[|\mathbf{Z} - \mathbb{E}\mathbf{Z}| \geq \epsilon \mathbb{E}\mathbf{Z}\right] \leq 2e^{-m\epsilon^2/8}.$$

If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, with probability $1 - O(e^{-\log(1/\delta)}) \geq 1 - \delta$:

$$(1-\epsilon)\|y\|_2^2 \leq \|\tilde{y}\|_2^2 \leq (1+\epsilon)\|y\|_2^2.$$

So Far: Each entry of our compressed vector $\tilde{y}$ is Gaussian with :

$$\tilde{y}_j \sim \mathcal{N}(0, \|y\|_2^2/m) \text{ and } \mathbb{E}[\|\tilde{y}\|_2^2] = \|y\|_2^2$$

$\|\tilde{y}\|_2^2 = \sum_{i=1}^m \tilde{y}_j^2$ a Chi-Squared random variable with m degrees of freedom (a sum of m squared independent Gaussians)

Lemma: (Chi-Squared Concentration) Letting $\mathbf{Z}$ be a Chi-Squared random variable with m degrees of freedom,

$$\Pr\left[|\mathbf{Z} - \mathbb{E}\mathbf{Z}| \geq \epsilon \mathbb{E}\mathbf{Z}\right] \leq 2e^{-m\epsilon^2/8}.$$

If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, with probability $1 - O(e^{-\log(1/\delta)}) \geq 1 - \delta$:

$$(1-\epsilon)\|y\|_2^2 \leq \|\tilde{y}\|_2^2 \leq (1+\epsilon)\|y\|_2^2.$$

Gives the distributional JL Lemma and thus the classic JL Lemma!

JL LEMMA IS ESSENTIALLY OPTIMAL

ORTHOGONAL VECTORS

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.
- What is the largest set of mutually orthogonal unit vectors in d-dimensional space?

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.
- What is the largest set of mutually orthogonal unit vectors in d-dimensional space? Answer: d.

ORTHOGONAL VECTORS

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.
- What is the largest set of mutually orthogonal unit vectors in d-dimensional space? Answer: d.
- How large can a set of unit vectors in d-dimensional space be that have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$?

- Recall that we say two vectors x, y are orthogonal if $\langle x, y \rangle = 0$.
- What is the largest set of mutually orthogonal unit vectors in d-dimensional space? Answer: d.
- How large can a set of unit vectors in d-dimensional space be that have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$? Answer: $2^{\Omega(\epsilon^2 d)}$.

An exponentially large set of random vectors will be nearly pairwise orthogonal with high probability!

Claim: $2^{O(\epsilon^2 d)}$ random d-dimensional unit vectors will have all pairwise dot products $|\langle x, y\rangle| \leq \epsilon$ (be nearly orthogonal).

Claim: $2^{O(\epsilon^2 d)}$ random d-dimensional unit vectors will have all pairwise dot products $|\langle x, y\rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t$ have independent random entries $\pm\frac{1}{\sqrt{d}}$.

Claim: $2^{O(\epsilon^2 d)}$ random d-dimensional unit vectors will have all pairwise dot products $|\langle x, y\rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t$ have independent random entries $\pm\frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every x_i is always a unit vector.

Claim: $2^{O(\epsilon^2 d)}$ random d-dimensional unit vectors will have all pairwise dot products $|\langle x, y\rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t$ have independent random entries $\pm\frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every x_i is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j\rangle]$?

Claim: $2^{O(\epsilon^2 d)}$ random d-dimensional unit vectors will have all pairwise dot products $|\langle x, y \rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t$ have independent random entries $\pm\frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every x_i is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j \rangle]$? $\mathbb{E}[\langle x_i, x_j \rangle] = 0$

Claim: $2^{O(\epsilon^2 d)}$ random d-dimensional unit vectors will have all pairwise dot products $|\langle x, y\rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t$ have independent random entries $\pm\frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every x_i is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j\rangle]$? $\mathbb{E}[\langle x_i, x_j\rangle] = 0$
- By a Bernstein bound, $\Pr[|\langle x_i, x_j\rangle| \geq \epsilon] \leq 2e^{-\epsilon^2 d/6}$.

Claim: $2^{O(\epsilon^2 d)}$ random d-dimensional unit vectors will have all pairwise dot products $|\langle x, y\rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $x_1, \ldots, x_t$ have independent random entries $\pm\frac{1}{\sqrt{d}}$.

- What is $\|x_i\|_2$? Every x_i is always a unit vector.
- What is $\mathbb{E}[\langle x_i, x_j\rangle]$? $\mathbb{E}[\langle x_i, x_j\rangle] = 0$
- By a Bernstein bound, $\Pr[|\langle x_i, x_j\rangle| \geq \epsilon] \leq 2e^{-\epsilon^2 d/6}$.
- If $t = \frac{1}{2}e^{\epsilon^2 d/12}$, using a union bound over $\binom{t}{2} \leq \frac{1}{8}e^{\epsilon^2 d/6}$ possible pairs, with probability $\geq 3/4$ all will be nearly orthogonal.

We won't prove it but this is essentially optimal: In d dimensions, there can be at most $2^{O(\epsilon^2 d)}$ nearly orthogonal unit vectors.

Recall: The Johnson Lindenstrauss lemma states that if $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ is a random matrix (linear map) with $m = O\left(\frac{\log n}{\epsilon^2}\right)$, for $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all i, j:

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|\boldsymbol{M}x_i - \boldsymbol{M}x_j\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2.$$

Recall: The Johnson Lindenstrauss lemma states that if $\boldsymbol{M} \in \mathbb{R}^{m \times d}$ is a random matrix (linear map) with $m = O\left(\frac{\log n}{\epsilon^2}\right)$, for $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all i, j:

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|\boldsymbol{M}x_i - \boldsymbol{M}x_j\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2.$$

Implies: If $x_1, \ldots, x_n$ are nearly orthogonal unit vectors in d-dimensions (with pairwise dot products bounded by $\epsilon/8$), then

$$\frac{\boldsymbol{M}x_1}{\|\boldsymbol{M}x_1\|_2}, \ldots, \frac{\boldsymbol{M}x_n}{\|\boldsymbol{M}x_n\|_2}$$

are nearly orthogonal unit vectors in m-dimensions (with pairwise dot products bounded by ϵ).

Recall: The Johnson Lindenstrauss lemma states that if $\boldsymbol{M} \in \mathbb{R}^{m\times d}$ is a random matrix (linear map) with $m = O\left(\frac{\log n}{\epsilon^2}\right)$, for $x_1, \ldots, x_n \in \mathbb{R}^d$ with high probability, for all i, j:

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|\boldsymbol{M}x_i - \boldsymbol{M}x_j\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2.$$

Implies: If $x_1, \ldots, x_n$ are nearly orthogonal unit vectors in d-dimensions (with pairwise dot products bounded by $\epsilon/8$), then

$$\frac{\boldsymbol{M}x_1}{\|\boldsymbol{M}x_1\|_2}, \ldots, \frac{\boldsymbol{M}x_n}{\|\boldsymbol{M}x_n\|_2}$$

are nearly orthogonal unit vectors in m-dimensions (with pairwise dot products bounded by ϵ). Algebra is a bit messy but a good exercise to partially work through. Proof uses the fact that

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2\langle x_i, x_j\rangle \ .$$

Claim 1: n nearly orthogonal unit vectors can be projected to $m = O\left(\frac{\log n}{\epsilon^2}\right)$ dimensions and still be nearly orthogonal.

Claim 2: In m dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.

Claim 1: n nearly orthogonal unit vectors can be projected to $m = O\left(\frac{\log n}{\epsilon^2}\right)$ dimensions and still be nearly orthogonal.

Claim 2: In m dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.

- For both of these to hold it must be that $n \leq 2^{O(\epsilon^2 m)}$.

CONNECTION TO DIMENSIONALITY REDUCTION

Claim 1: n nearly orthogonal unit vectors can be projected to $m = O\left(\frac{\log n}{\epsilon^2}\right)$ dimensions and still be nearly orthogonal.

Claim 2: In m dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.

- For both of these to hold it must be that $n \leq 2^{O(\epsilon^2 m)}$.
- I.e., $n = 2^{\log n} \leq 2^{O(\epsilon^2 m)}$ and so $m = \Omega\left(\frac{\log n}{\epsilon^2}\right)$.

Claim 1: n nearly orthogonal unit vectors can be projected to $m = O\left(\frac{\log n}{\epsilon^2}\right)$ dimensions and still be nearly orthogonal.

Claim 2: In m dimensions, there can be at most $2^{O(\epsilon^2 m)}$ nearly orthogonal unit vectors.

- For both of these to hold it must be that $n \leq 2^{O(\epsilon^2 m)}$.
- I.e., $n = 2^{\log n} \leq 2^{O(\epsilon^2 m)}$ and so $m = \Omega\left(\frac{\log n}{\epsilon^2}\right)$.
- Tells us that the JL lemma is optimal up to constants.

Central Limit Theorem

INTERPRETATION AS A CENTRAL LIMIT THEOREM

Bernstein Inequality (Simplified): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ falling in [-1,1]. Let $\mu = \mathbb{E}[\sum \mathbf{X}_i]$, $\sigma^2 = \text{Var}[\sum \mathbf{X}_i]$, and $s \leq \sigma$. Then:

$$\Pr\left(\left|\sum_{i=1}^{n} \mathbf{X}_i - \mu\right| \geq s\sigma\right) \leq 2\exp\left(-\frac{s^2}{4}\right).$$

Bernstein Inequality (Simplified): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ falling in [-1,1]. Let $\mu = \mathbb{E}[\sum \mathbf{X}_i]$, $\sigma^2 = \text{Var}[\sum \mathbf{X}_i]$, and $s \leq \sigma$. Then:

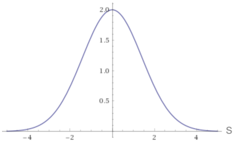
$$\Pr\left(\left|\sum_{i=1}^{n} \mathbf{X}_i - \mu\right| \geq s\sigma\right) \leq 2\exp\left(-\frac{s^2}{4}\right).$$

Can plot this bound for different s:

Bernstein Inequality (Simplified): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ falling in [-1,1]. Let $\mu = \mathbb{E}[\sum \mathbf{X}_i]$, $\sigma^2 = \text{Var}[\sum \mathbf{X}_i]$, and $s \leq \sigma$. Then:

$$\Pr\left(\left|\sum_{i=1}^{n} \mathbf{X}_i - \mu\right| \geq s\sigma\right) \leq 2\exp\left(-\frac{s^2}{4}\right).$$

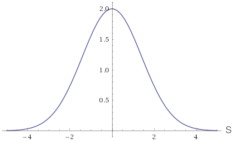
Can plot this bound for different s:



Bernstein Inequality (Simplified): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ falling in [-1,1]. Let $\mu = \mathbb{E}[\sum \mathbf{X}_i]$, $\sigma^2 = \text{Var}[\sum \mathbf{X}_i]$, and $s \leq \sigma$. Then:

$$\Pr\left(\left|\sum_{i=1}^{n} \mathbf{X}_i - \mu\right| \geq s\sigma\right) \leq 2\exp\left(-\frac{s^2}{4}\right).$$

Can plot this bound for different s:

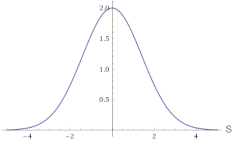


Looks a lot like a Gaussian (normal) distribution.

Bernstein Inequality (Simplified): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ falling in [-1,1]. Let $\mu = \mathbb{E}[\sum \mathbf{X}_i]$, $\sigma^2 = \text{Var}[\sum \mathbf{X}_i]$, and $s \leq \sigma$. Then:

$$\Pr\left(\left|\sum_{i=1}^{n} \mathbf{X}_i - \mu\right| \geq s\sigma\right) \leq 2\exp\left(-\frac{s^2}{4}\right).$$

Can plot this bound for different s:



Looks a lot like a Gaussian (normal) distribution.

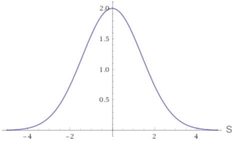
$$\mathcal{N}(0, \sigma^2) \text{ has density } p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}.$$

INTERPRETATION AS A CENTRAL LIMIT THEOREM

Bernstein Inequality (Simplified): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ falling in [-1,1]. Let $\mu = \mathbb{E}[\sum \mathbf{X}_i]$, $\sigma^2 = \text{Var}[\sum \mathbf{X}_i]$, and $s \leq \sigma$. Then:

$$\Pr\left(\left|\sum_{i=1}^{n} \mathbf{X}_i - \mu\right| \geq s\sigma\right) \leq 2\exp\left(-\frac{s^2}{4}\right).$$

Can plot this bound for different s:



Looks a lot like a Gaussian (normal) distribution.

$$\mathcal{N}(0, \sigma^2) \text{ has density } p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}.$$

$\mathcal{N}(0, \sigma^2)$ has density $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}$.

GAUSSIAN TAILS

$$\mathcal{N}(0, \sigma^2) \text{ has density } p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}.$$

Exercise: Using this can show that for $\mathbf{X} \sim \mathcal{N}(0, \sigma^2)$: for any $s \geq 0$,

$$\Pr\left(|\mathbf{X}| \geq s \cdot \sigma\right) \leq O(1) \cdot e^{-\frac{s^2}{2}}.$$

GAUSSIAN TAILS

$$\mathcal{N}(0, \sigma^2) \text{ has density } p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}.$$

Exercise: Using this can show that for $\mathbf{X} \sim \mathcal{N}(0, \sigma^2)$: for any $s \geq 0$,

$$\Pr\left(|\mathbf{X}| \geq s \cdot \sigma\right) \leq O(1) \cdot e^{-\frac{s^2}{2}}.$$

Essentially the same bound that Bernstein's inequality gives!

GAUSSIAN TAILS

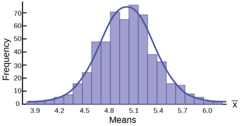
$$\mathcal{N}(0, \sigma^2) \text{ has density } p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}.$$

Exercise: Using this can show that for $\mathbf{X} \sim \mathcal{N}(0, \sigma^2)$: for any $s \geq 0$,

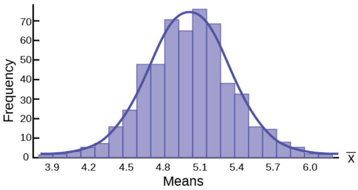
$$\Pr(|\mathbf{X}| \geq s \cdot \sigma) \leq O(1) \cdot e^{-\frac{s^2}{2}}.$$

Essentially the same bound that Bernstein's inequality gives!

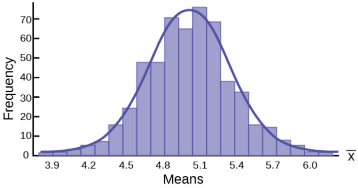
Central Limit Theorem Interpretation: Bernstein's inequality gives a quantitative version of the CLT. The distribution of the sum of *bounded* independent random variables can be upper bounded with a Gaussian (normal) distribution.



Stronger Central Limit Theorem: The distribution of the sum of *n bounded* independent random variables converges to a Gaussian (normal) distribution as *n* goes to infinity.

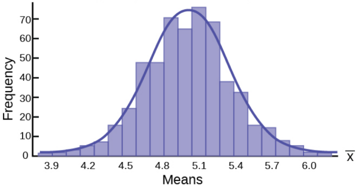


Stronger Central Limit Theorem: The distribution of the sum of *n bounded* independent random variables converges to a Gaussian (normal) distribution as *n* goes to infinity.



- Why is the Gaussian distribution is so important in statistics, science, ML, etc.?

Stronger Central Limit Theorem: The distribution of the sum of *n bounded* independent random variables converges to a Gaussian (normal) distribution as *n* goes to infinity.



- Why is the Gaussian distribution is so important in statistics, science, ML, etc.?
- Many random variables can be approximated as the sum of a large number of small and roughly independent random effects. Thus, their distribution looks Gaussian by CLT.