Warmup Exercise / Detour: Prove this Lemma

Before we get started with today's lecture, let's warm up by proving a lemma that will be useful later.

Suppose X is an RV with PDF $f(x)$. If $a > 0$, and $b \in \mathbb{R}$, if we define $Y = aX + b$, then what is the PDF of Y?

We have that: $F_Y(y) = P(Y \leq y)$

CC BY NC SA

Warmup Exercise / Detour: Prove this Lemma

Before we get started with today's lecture, let's warm up by proving a lemma that will be useful later.

Suppose X is an RV with PDF $f(x)$. If $a > 0$, and $b \in \mathbb{R}$, if we define $Y = aX + b$, then what is the PDF of Y?

We have that: $F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$

How do we get the PDF $f_Y(y)$ from the CDF $F_Y(y)$? Differentiate!

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X\left(\frac{y-b}{a}\right)}{dy} = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}$$

If $\mathrm{Y} = \mathrm{aX} + \mathrm{b}$, then

$$f_Y(y) = f_X\left(\frac{\mathrm{y}-\mathrm{b}}{\mathrm{a}}\right) \cdot \frac{1}{\mathrm{a}}$$

CC BY NC SA

Data Colada

Data Colada “is a blog dedicated to investigative analysis and replication of academic research, focusing in particular on the validity of findings in the social sciences.”

- The blog is run by Uri Simonsohn at ESADE Business School in Barcelona, Leif Nelson at Haas here at Berkeley, and Joseph Simmons at the Wharton School at Penn.

Their work has found evidence of serious academic misconduct, even among world class researchers at universities like Harvard and Duke. See this video.

Today, we’ll get a glimpse into one of the mathematical techniques that make it possible to sniff out this type of fraud.

CC BY NC SA

Thought Experiment #1

Suppose we run our classic CS70 experiment: We take all students' homework, shuffle it, and hand it back randomly. Let X_i be the number of students who get their homework back on the ith experiment.

Suppose we have 1000 high school classes across the country do this experiment. Each reports back their results.

Suppose that the sum we get back is $\mathrm{S} = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. How likely is it that we get this value or less? This value or more? Is there any reason to be suspicious of these results?

CC BY NC SA

Empirical Observation

Lecture 25, CS70 Fall 2026

CC BY NC SA

Sums of i.i.d. Random Variables: $S = X_1 + X_2 + \cdots + X_n$

$X_i \sim$ Bernoulli(p)

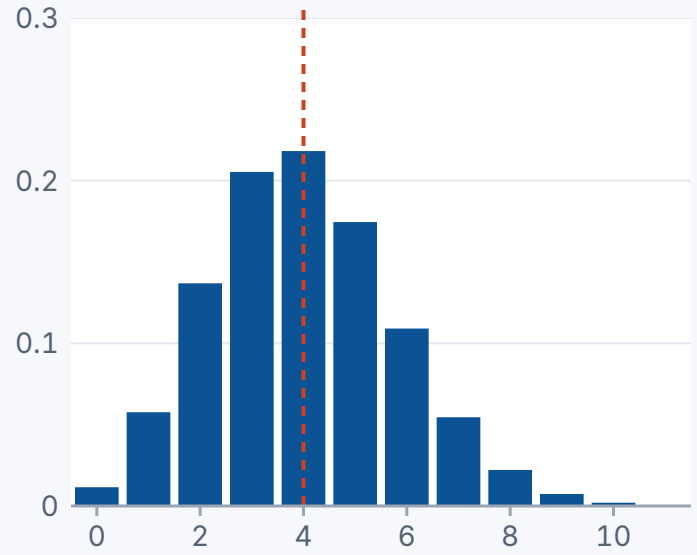
1 with probability p, else 0

n (number added) **20**

p **0.20**

$\mathbb{E}[S] = np$ **4**

$\mathrm{Var}(S) = np(1-p)$ **3.2**



$X_i \sim$ Uniform(k)

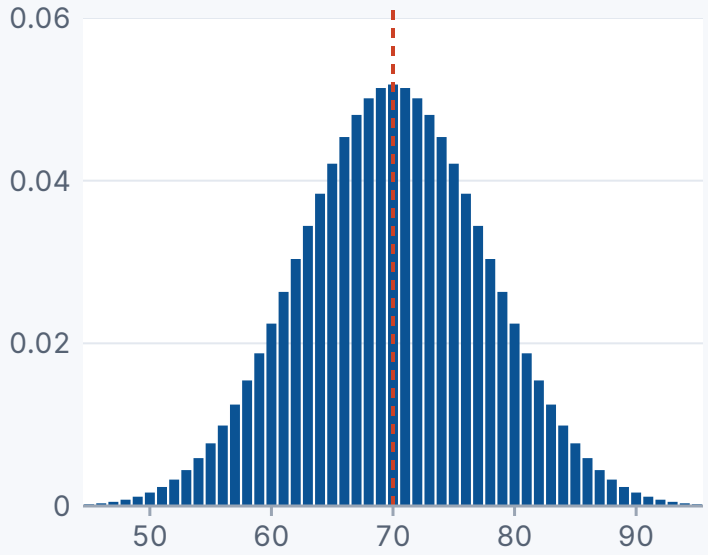
uniform on {1, 2, ..., k}

n (number added) **20**

k **6**

$\mathbb{E}[S] = n(k+1)/2$ **70**

$\mathrm{Var}(S) = n(k^2-1)/12$ **58.33**



$X_i \sim$ Geometric(p)

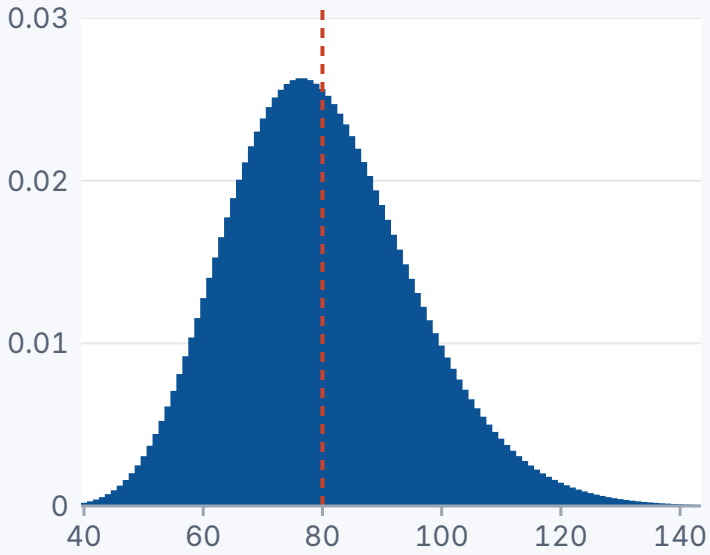
flips until the first heads

n (number added) **20**

p **0.25**

$\mathbb{E}[S] = n/p$ **80**

$\mathrm{Var}(S) = n(1-p)/p^2$ **240**



$X_i \sim$ FixedPoint(m)

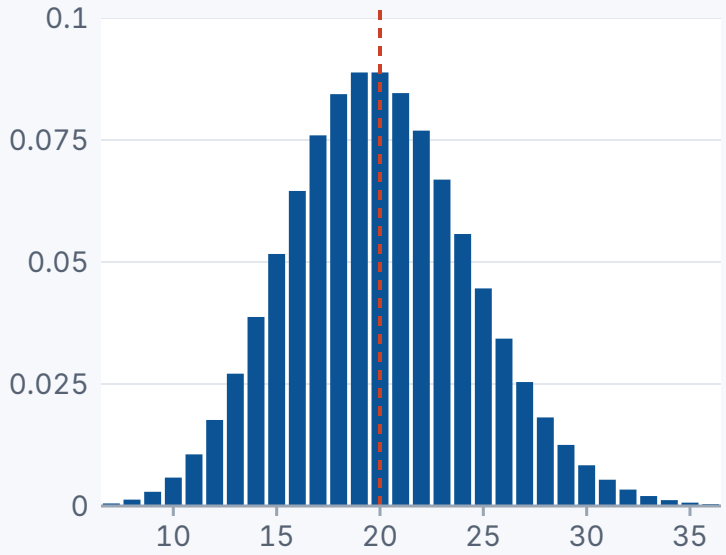
students who get their own homework back

n (number added) **20**

m **5**

$\mathbb{E}[S] = n$ **20**

$\mathrm{Var}(S) = n$ **20**



CC BY NC SA

Standardized Sums: $\tilde{S} = (S - \mathbb{E}[S])/\sigma(S)$

$X_i \sim$ Bernoulli(p)

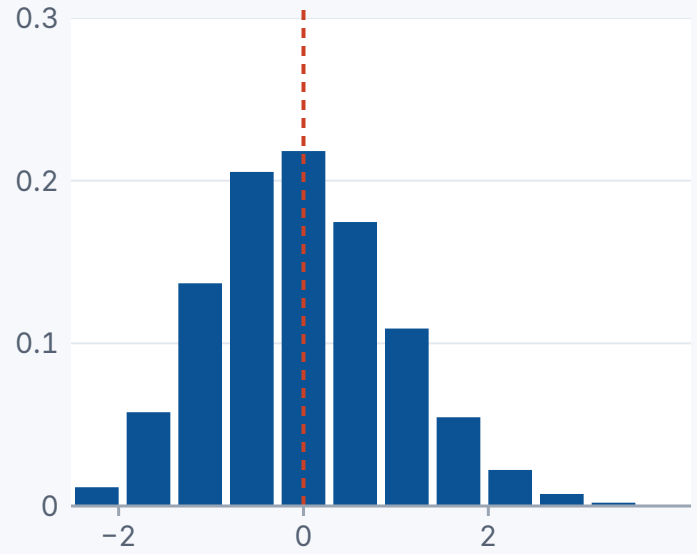
1 with probability p, else 0

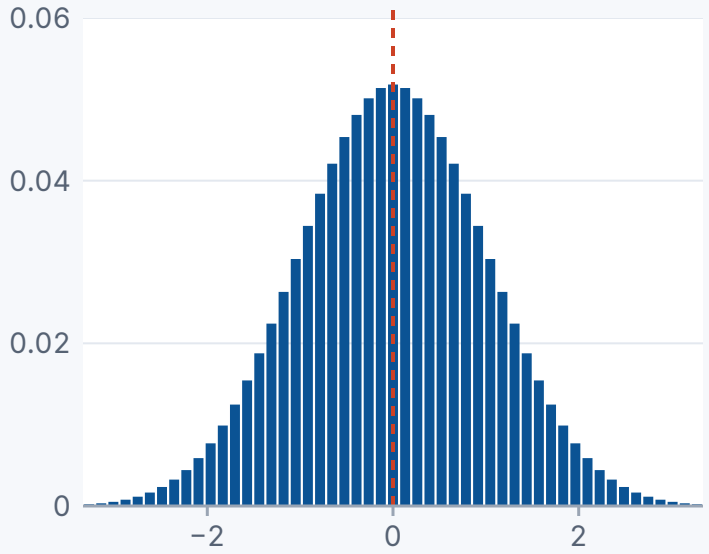
n (number added) **20**

p **0.20**

$\mathbb{E}[S] = np$ **4**

$\mathrm{Var}(S) = np(1-p)$ **3.2**





$X_i \sim$ Uniform(k)

uniform on {1, 2, ..., k}

n (number added) **20**

k **6**

$\mathbb{E}[S] = n(k+1)/2$ **70**

$\mathrm{Var}(S) = n(k^2-1)/12$ **58.33**

$X_i \sim$ Geometric(p)

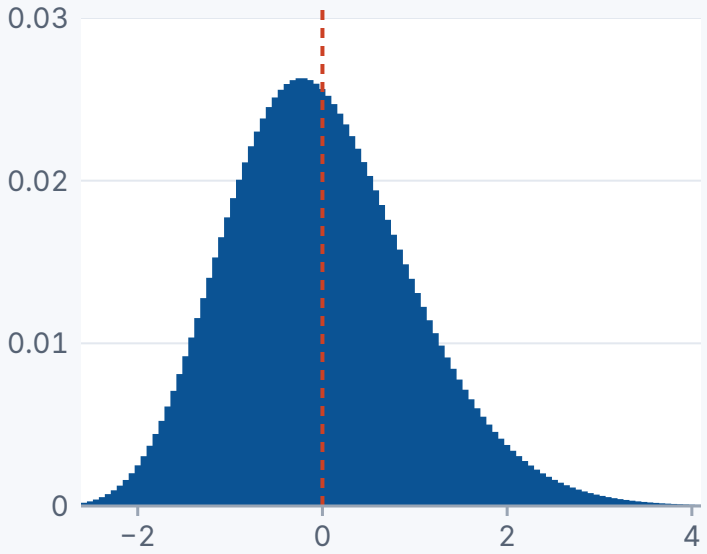
flips until the first heads

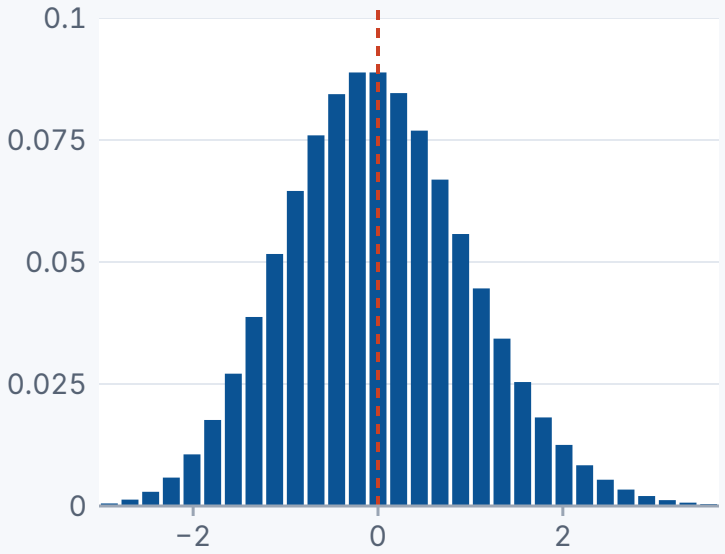
n (number added) **20**

p **0.25**

$\mathbb{E}[S] = n/p$ **80**

$\mathrm{Var}(S) = n(1-p)/p^2$ **240**





$X_i \sim$ FixedPoint(m)

students who get their own homework back

n (number added) **20**

m **5**

$\mathbb{E}[S] = n$ **20**

$\mathrm{Var}(S) = n$ **20**

CC BY NC SA

Standardized Sums in the Large n Limit

Standardized sum of n i.i.d. Geometric(p)

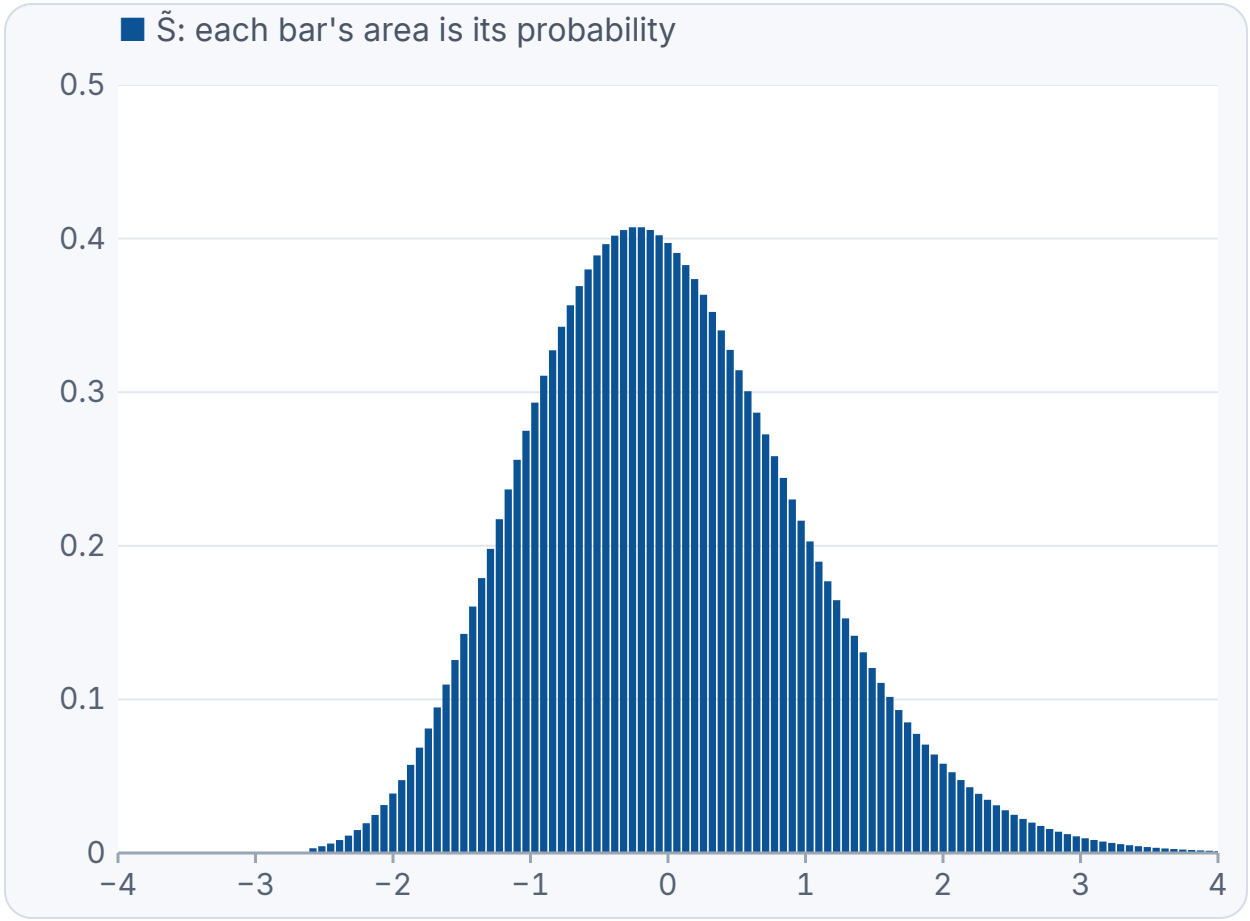
$\mathbb{E}[S] = n/p = 80$ $\sigma(S) = \sqrt{n(1-p)}/p = 15.49$

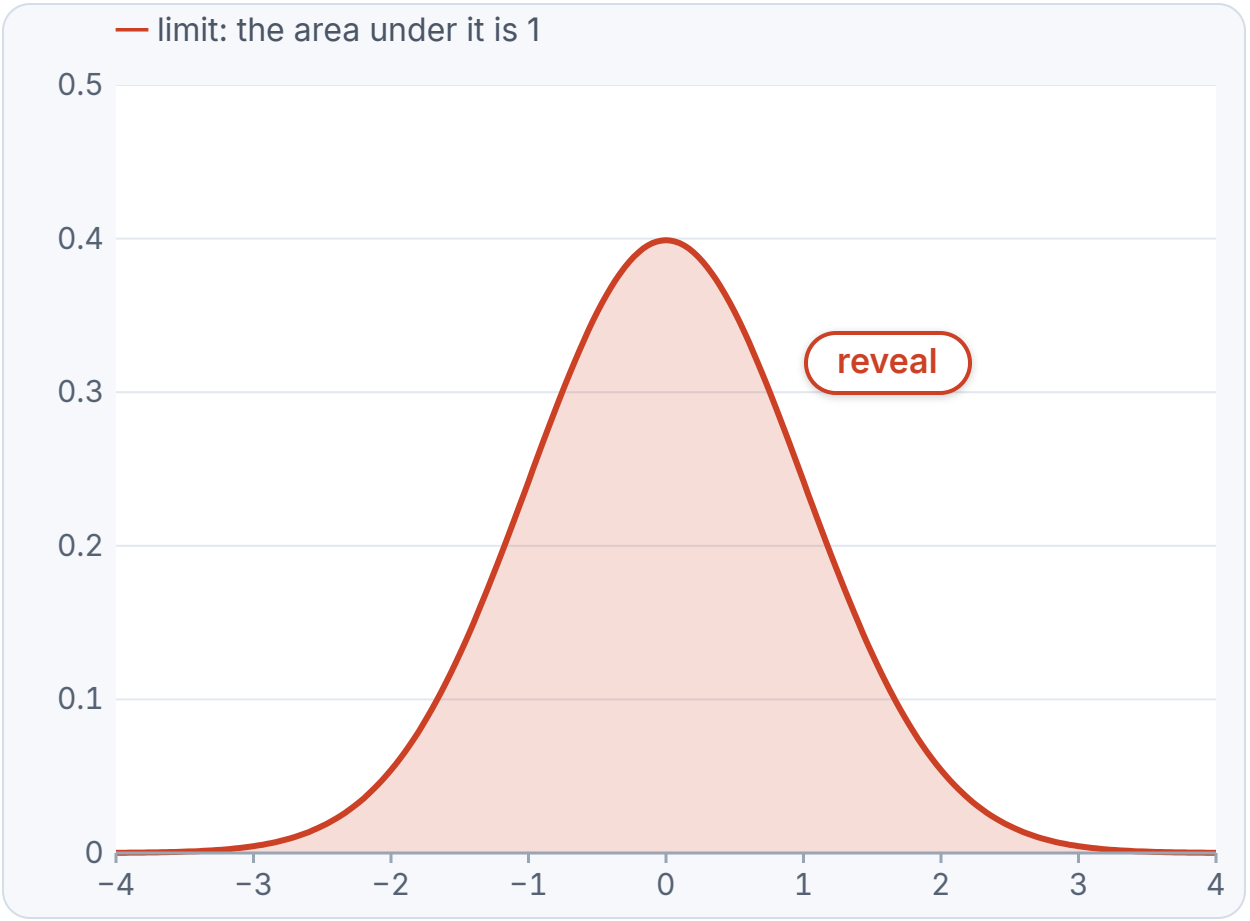
n (number added) 20 p 0.25

Limit as $n \to \infty$

$\mathbb{E} = 0$, Var = 1

Overlay

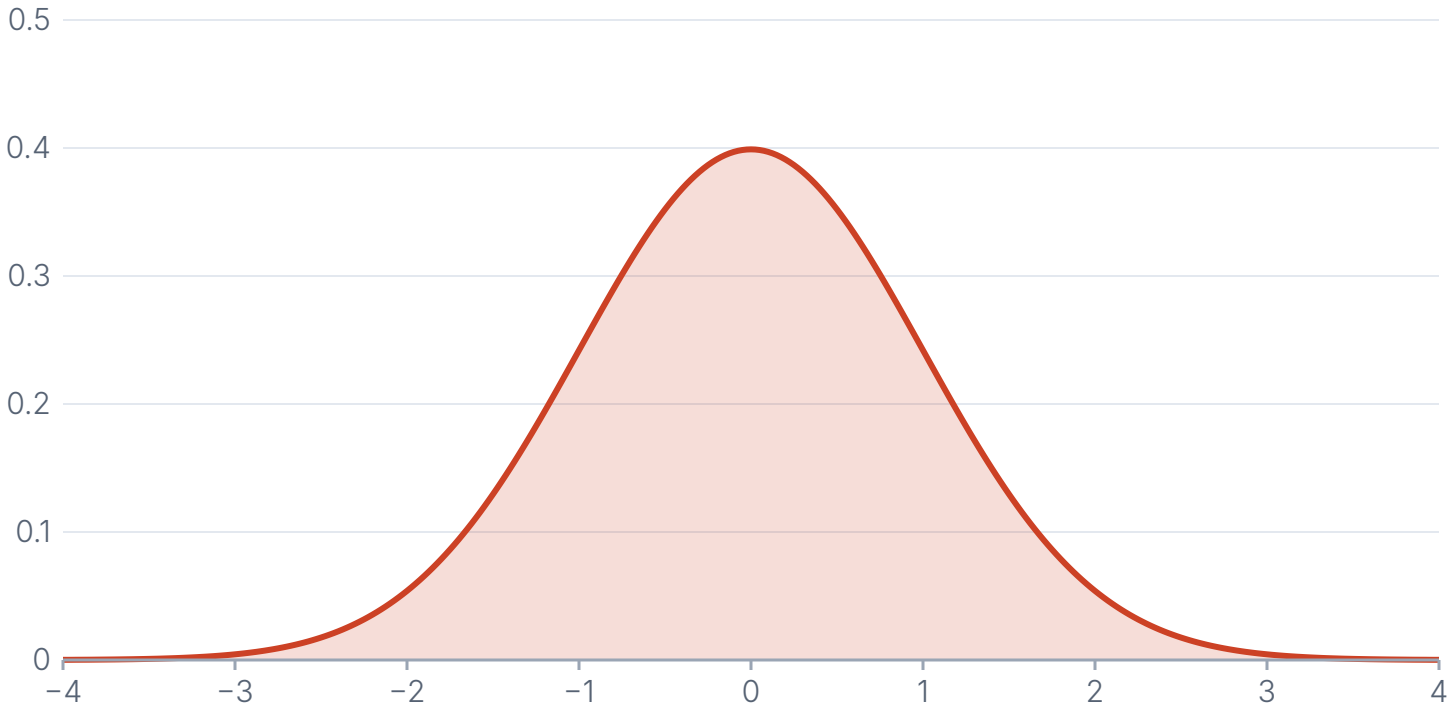




CC BY NC SA

Large Sums of i.i.d. Random Variables

If we add n i.i.d. discrete random variables, then for large n, the sum's distribution is approximately that of a continuous random variable with a (scaled and shifted) version of the PDF shown below.



This distribution is called the **Gaussian**. Let's explore its properties.

cc BY NC SA

Gaussian Distribution

Lecture 25, CS70 Fall 2026

CC BY NC SA

The Gaussian Distribution

The Gaussian distribution is probably the most famous continuous distribution.

- Arises naturally in many theoretical contexts.
- Useful for modeling real world phenomena.
- Has very convenient mathematical properties.

CC BY NC SA

Normal (Gaussian) Distribution

CC BY NC SA

Normal (Gaussian) Distribution

We say that $\mathbf{X} \sim \mathbf{Gaussian}(\mu, \sigma^2)$ or $\mathbf{X} \sim \mathbf{Normal}(\mu, \sigma^2)$, or more simply $\mathbf{X} \sim \mathbf{N}(\mu, \sigma^2)$ if its PDF is:

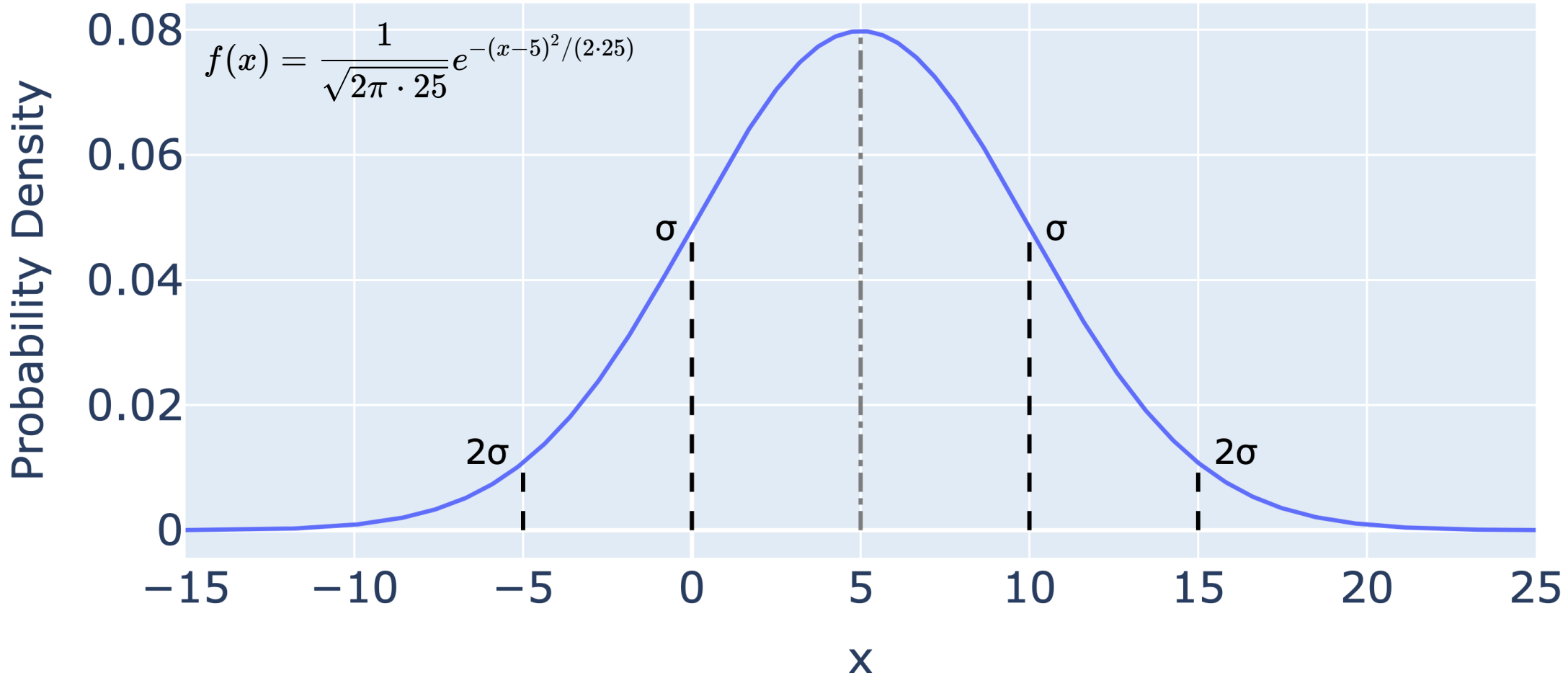
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

I will use Gaussian, Normal, and N interchangeably.

CC BY NC SA

The Gaussian PDF Visually

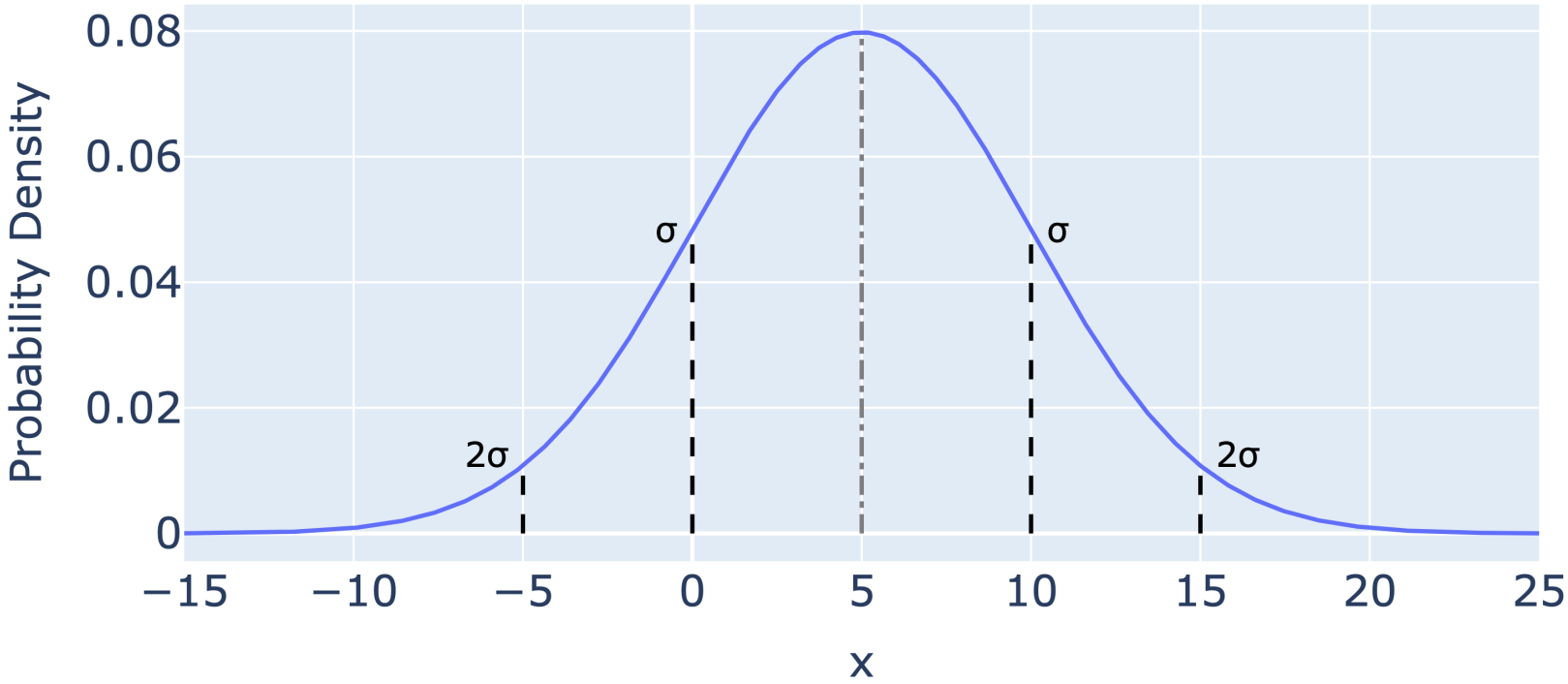
The PDF of $X \sim \text{Gaussian}(5,\ 25)$ is below:



CC BY NC SA

Computing the Probability of an Event

If we have a random variable $X \sim \mathrm{N}(5,\ 25)$, what is $P(X \leq 0)$?

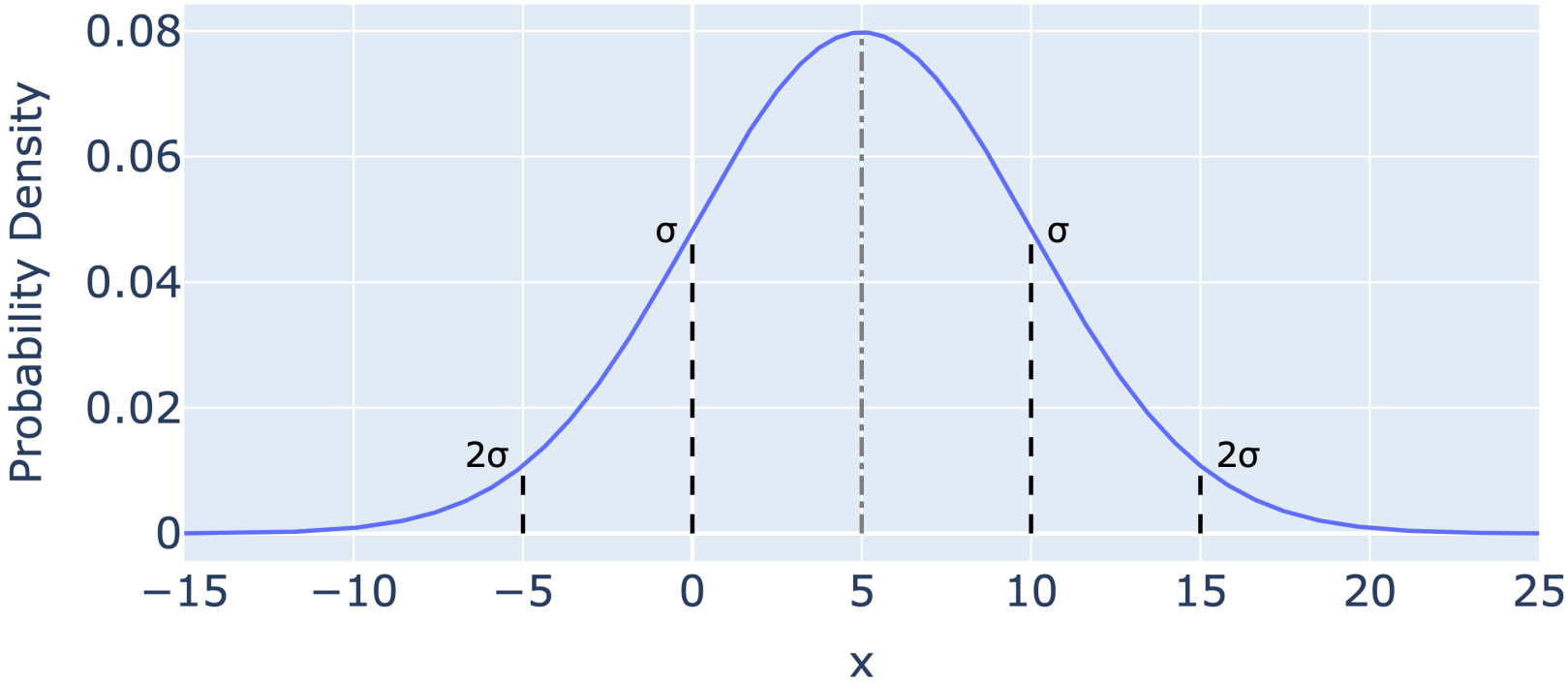


CC BY NC SA

Computing the Probability of an Event

If we have a random variable $X \sim \mathrm{N}(5,\ 25)$, what is $P(X \leq 0)$?

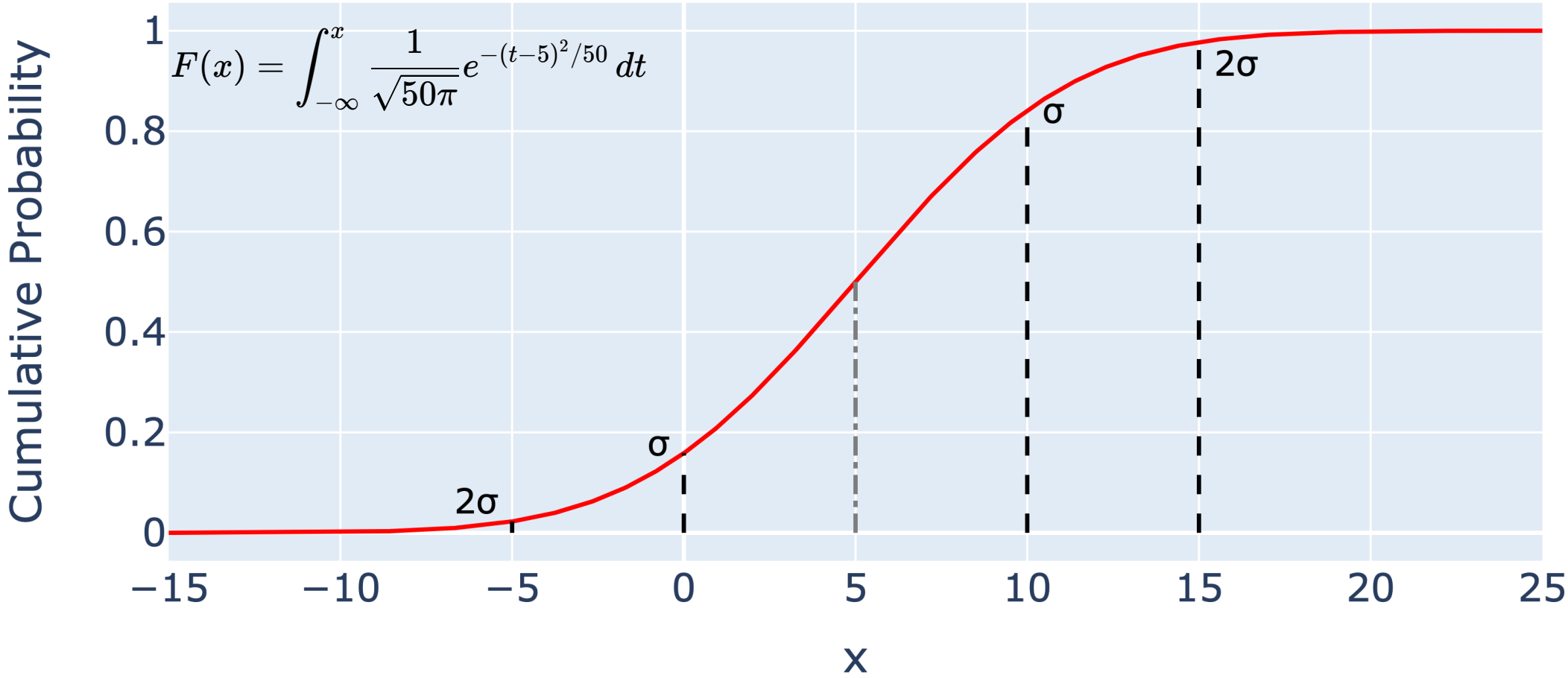
$$\int_{-\infty}^{0} \frac{1}{\sqrt{50\pi}} e^{-(x-5)^2/50}\, dx \quad \approx \quad 0.1587$$



The CDF of a Gaussian

The CDF of $X \sim \text{Normal}(5,\ 25)$ is given below.

- $P(X \leq 0) \approx 0.1587$
- Note: There is no closed-form expression for the CDF.

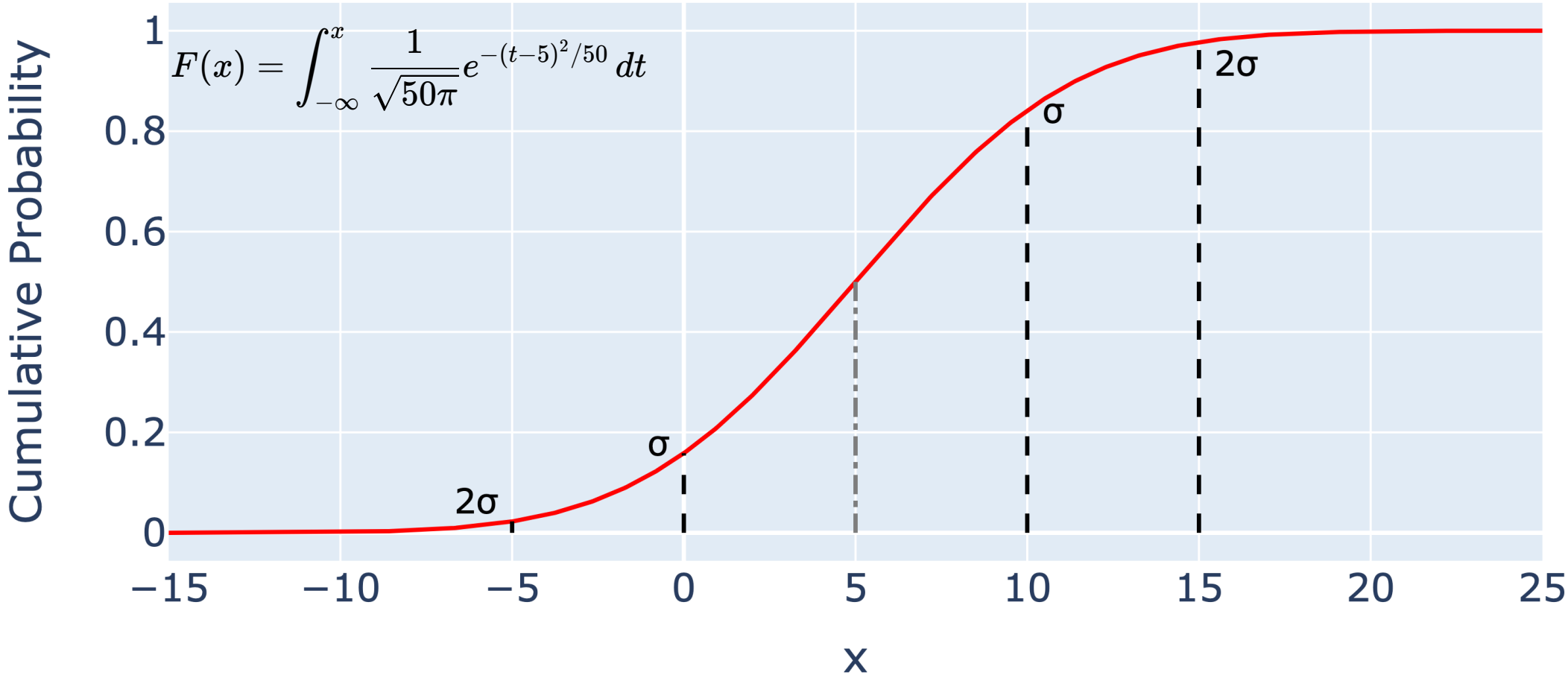


CC BY NC SA

The CDF of a Gaussian

Or symbolically we can write

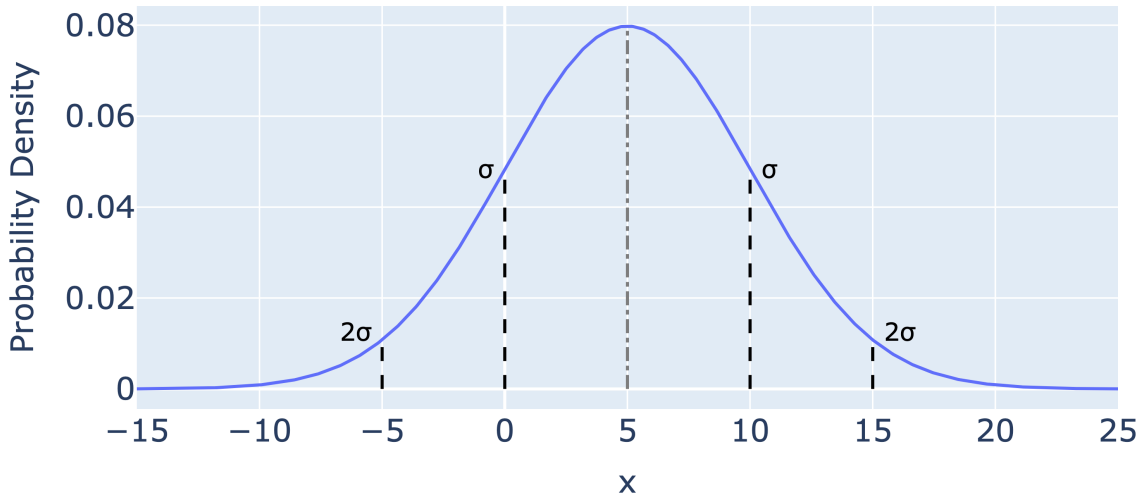
- $P(X \leq 0) = F_X(\mu - \sigma) \approx 0.1587$
- Other quantities: $F_X(\mu) = 0.5$, $F_X(\mu + 2\sigma) \approx 0.977$



CC BY NC SA

The Mean and Variance of a Gaussian

Suppose we have $X \sim \mathrm{N}(5,\ 25)$, as shown below.



Then, as we'll prove later:

- The mean is 5.
- The variance is 25, so the standard deviation is 5.

More generally, the mean of $\mathrm{N}(\mu,\ \sigma^2)$ is μ, and the variance is σ^2.

And standardizing gives a standard Gaussian: $\dfrac{X - \mu}{\sigma} \sim \mathrm{N}(0,\ 1)$.

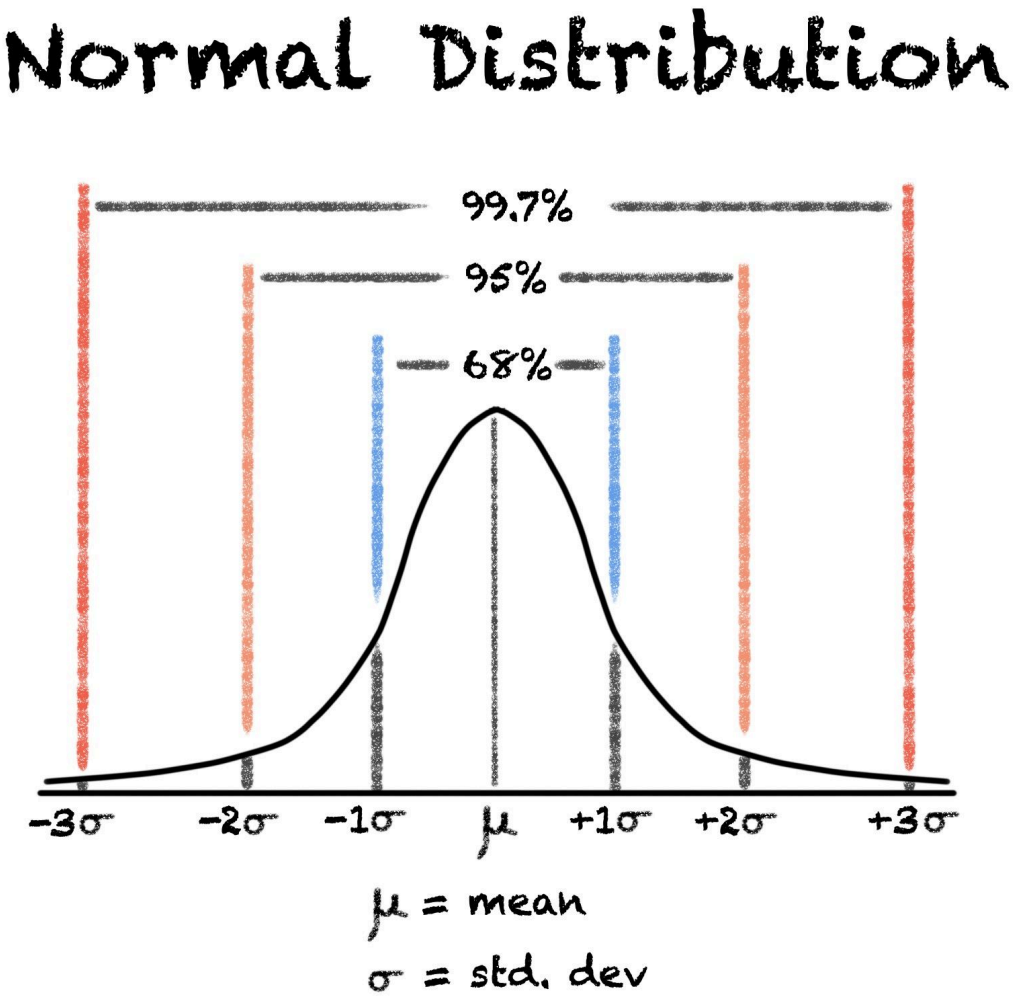
CC BY NC SA

The 68–95–99.7 rule

The 68-95-99.7 rule tells us that for a normally distributed random variable:

- ~68% of the mass is within one standard deviation.
- 95% of the mass is within two standard deviations.
- 99.7% of the mass is within three standard deviations.

The "Z-score" is the number of standard deviations you are from the mean.



Link to blog

CC BY NC SA

Normal (Gaussian) Distribution

$\mathrm{X} \sim \mathrm{Gaussian}(\mu, \sigma^2)$ or $\mathrm{X} \sim \mathrm{Normal}(\mu, \sigma^2)$, or more simply $\mathrm{X} \sim \mathrm{N}(\mu, \sigma^2)$ if its PDF is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

If we have a Gaussian RV $\tilde{X} \sim \mathrm{N}(0, 1)$, we say that $\tilde{X}$ is a **standard** normal. It has PDF:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

CC BY NC SA

Normal (Gaussian) Distribution

If we have a Gaussian RV $\tilde{X} \sim \mathrm{N}(0, 1)$, we say that $\tilde{X}$ is a **standard** normal. It has PDF:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

The standard Gaussian CDF is so important that it's often written (in probability contexts) as simply Φ. This is a capital Greek "Phi".

- Example: $\Phi(3) \approx 0.9987$ is $P(X \leq 3)$, where $X \sim N(0, 1)$.
- Interpretation: Chance of being at most 3 standard deviations above the mean is 99.87%.

CC BY NC SA

The Gaussian CDF Φ

Example: Assume heights are normally distributed. Average height of a male in this dataset (https://ourworldindata.org/human-height) is ~69 inches, and standard deviation is ~3 inches.

- Chance of being as tall as Josh Hug or shorter (64 inches) is $\Phi\left(\frac{-5}{3}\right) = 4.78\%$
- Chance of being as tall as John DeNero (76 inches?) or taller: $1 - \Phi\left(\frac{7}{3}\right) = 1\%$

joshh.ug/E



CC BY NC SA

Affine Gaussian Transformations

Lecture 25, CS70 Fall 2026

CC BY NC SA

Question

Suppose we have $X \sim \text{Exponential}(\lambda)$.

- Is $X + 1$ an exponential?
- Is $X/2$ an exponential?

Reminder: the CDF of X is

$$F_X(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

joshh.ug/A

CC BY NC SA

Question

Suppose we have $X \sim \text{Exponential}(\lambda)$.

- Is $X + 1$ an exponential?
- Is $X/2$ an exponential?

Reminder: the CDF of X is

$$F_X(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Only $X/2$ is: $X/2 \sim \text{Exponential}(2\lambda)$. $X + 1$ is not, since it is never less than 1.

But for a Gaussian random variable X, both $X + k$ and cX are Gaussian.

CC BY NC SA

Gaussian

Gaussian properties to prove:

1. If $X \sim N(0,\ 1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$
2. If $\mathrm{Y} \sim N(\mu, \sigma^2)$, and $\mathrm{X} = (Y - \mu)/\sigma$, then $\mathrm{X} \sim N(0,\ 1)$

CC BY NC SA

Property 1

If $X \sim N(0,1)$, and we define $Y = \sigma X + \mu$, then $Y \sim N(\mu, \sigma^2)$. In other words:

- If we add a constant μ to a standard Gaussian RV, then we get a new Gaussian RV whose first parameter is μ.
- If we multiply a standard Gaussian RV by a constant σ, then we get a new Gaussian RV whose second parameter is σ^2.

To prove this, we'll use our warmup lemma!

Warmup Lemma

If $\mathrm{Y} = \mathrm{aX} + \mathrm{b}$, then

$$f_Y(y) = f_X\left(\frac{\mathrm{y} - \mathrm{b}}{\mathrm{a}}\right) \cdot \frac{1}{\mathrm{a}}$$

CC BY NC SA

Property 1

If $X \sim N(0,1)$, and we define $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$

Warmup Lemma

If $\mathrm{Y} = \mathrm{aX} + \mathrm{b}$, then

$$f_Y(y) = f_X\left(\frac{\mathrm{y} - \mathrm{b}}{\mathrm{a}}\right) \cdot \frac{1}{\mathrm{a}}$$

CC BY NC SA

Property 1

If $X \sim N(0,1)$, and we define $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$

Proof: We have that $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$f_Y(y) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{y-\mu}{\sigma}\right)^2/2}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}$$

This is the PDF for $\mathrm{N}(\mu,\ \sigma^2)$

Warmup Lemma

If $\mathrm{Y} = \mathrm{aX} + \mathrm{b}$, then

$$f_Y(y) = f_X\left(\frac{\mathrm{y} - \mathrm{b}}{\mathrm{a}}\right) \cdot \frac{1}{\mathrm{a}}$$

CC BY NC SA

Property 2

We've shown that if $X \sim N(0,1)$, and we define $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$.

CC BY NC SA

Property 2

We've shown that if $X \sim N(0,1)$, and we define $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$.

Naturally, if $Y \sim N(\mu, \sigma^2)$, we can do the reverse transformation to recover the original X. That is, if $Y \sim N(\mu,\ \sigma^2)$, then:

$$X = \frac{Y - \mu}{\sigma} \sim N(0,\ 1)$$

CC BY NC SA

Reflection on Gaussian Properties

Gaussian properties we've proven:

1. If $X \sim N(0,\ 1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$ ✔
2. If $\mathrm{Y} \sim N(\mu, \sigma^2)$, and $\mathrm{X} = (Y - \mu)/\sigma$, then $\mathrm{X} \sim N(0,\ 1)$ ✔

In other words, any affine transformation of a Gaussian is also Gaussian.

- Not true for many RVs, e.g. $\mathrm{X} \sim \mathrm{Exponential}(\lambda)$, but $X + 1$ is not exponential.

CC BY NC SA

Gaussian Parameter Interpretation

Lecture 25, CS70 Fall 2026

CC BY NC SA

Gaussian

Let's next show the following properties:

3. The mean of $\mathbf{X} \sim N(0,1)$ is 0
4. The variance of $X \sim N(0,1)$ is 1
5. The mean of $\mathbf{X} \sim N(\mu, \sigma^2)$ is μ
6. The variance of $X \sim N(\mu, \sigma^2)$ is σ^2

CC BY NC SA

Expectation of a Standard Gaussian Random Variable

If $X \sim \text{Gaussian}(0,\ 1)$, then $\mathbb{E}[X] = 0$. Proof:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

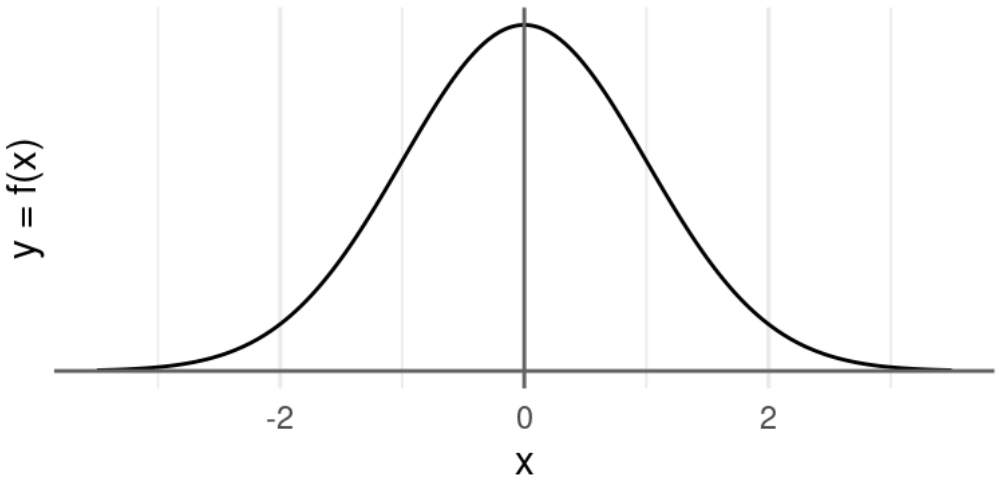


Figure from Stat 20

CC BY NC SA

Expectation of a Standard Gaussian Random Variable

If $X \sim \text{Gaussian}(0, 1)$, then $\mathbb{E}[X] = 0$. Proof:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-x^2/2} dx$$

Cancel out exactly!

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} x \cdot e^{-x^2/2} dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} x \cdot e^{-x^2/2} dx = 0$$

Another view: The expected value of any non-pathological distribution symmetric around zero is zero. The PDF of a standard Gaussian looks like this:

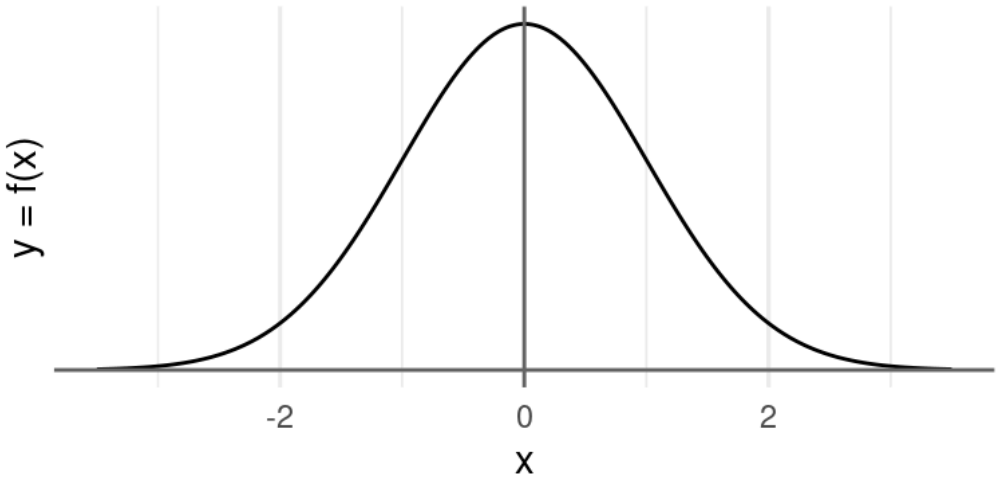


Figure from Stat 20

Variance of a Standard Gaussian Random Variable

If $X \sim \text{Gaussian}(0,\ 1)$, then $\text{Var}(X) = \mathbb{E}[X^2] = 1$. Proof:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, and we already know $\mathbb{E}[X]^2 = 0$

$$\mathbb{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \cdot e^{-x^2/2} dx$$

Integrate by parts where $u(x) = -x$ and $v'(x) = -xe^{-x^2/2}$

$$= \frac{1}{\sqrt{2\pi}} (-x) \cdot (e^{-x^2/2}) \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -1 \cdot -e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{-x}{e^{x^2/2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f(x)\, dx$$

What are A and B?

https://joshh.ug/B

CC BY NC SA

Variance of a Standard Gaussian Random Variable

If $X \sim \text{Gaussian}(0,\ 1)$, then $\text{Var}(X) = \mathbb{E}[X^2] = 1$. Proof:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, and we already know $\mathbb{E}[X]^2 = 0$

$$\mathbb{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \cdot e^{-x^2/2} dx$$

Integrate by parts where $u(x) = -x$ and $v'(x) = -xe^{-x^2/2}$

$$= \frac{1}{\sqrt{2\pi}} (-x) \cdot (e^{-x^2/2}) \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -1 \cdot -e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{-x}{e^{x^2/2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f(x)\, dx$$

$$= 0 + 1$$

Intuition or L'Hôpital

This is just the PDF of a Gaussian, which must integrate to 1.

CC BY NC SA

Gaussian

Gaussian properties to prove:

3. The mean of $\mathbf{X} \sim N(0,1)$ is 0 ✔
4. The variance of $X \sim N(0,1)$ is 1 ✔
5. The mean of $\mathbf{X} \sim N(\mu, \sigma^2)$ is μ
6. The variance of $X \sim N(\mu, \sigma^2)$ is σ^2

Now let's show properties 5 and 6.

CC BY NC SA

Proof of Properties 5 and 6

Suppose $X \sim N(0,\ 1)$, and $Y = \sigma X + \mu$.

Goals:

Show the mean of $N(\mu,\ \sigma^2)$ is μ.

Show the variance of $N(\mu,\ \sigma^2)$ is σ^2.

CC BY NC SA

Proof of Properties 5 and 6

Suppose $X \sim N(0,\ 1)$, and $Y = \sigma X + \mu$.

Goals:
Show the mean of $N(\mu,\ \sigma^2)$ is μ.
Show the variance of $N(\mu,\ \sigma^2)$ is σ^2.

That means that $\mathbb{E}[Y] = \mathbb{E}[\sigma X + \mu] = \sigma\mathbb{E}[X] + \mu$

$= 0 + \mu = \mu$ Property 3: $\mathbb{E}[X] = 0$

And $\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + \mu) = \sigma^2 \mathrm{Var}(X)$

$= \sigma^2$ Property 4: $\mathrm{Var}(X) = 1$

CC BY NC SA

Gaussian

Gaussian properties we've proven:

1. If $X \sim N(0,\ 1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$ ✓
2. If $\mathrm{Y} \sim N(\mu, \sigma^2)$, and $\mathrm{X} = (Y - \mu)/\sigma$, then $\mathrm{X} \sim N(0,\ 1)$ ✓
3. The mean of $\mathrm{X} \sim N(0, 1)$ is 0 ✓
4. The variance of $X \sim N(0, 1)$ is 1 ✓
5. The mean of $\mathrm{X} \sim N(\mu, \sigma^2)$ is μ ✓
6. The variance of $X \sim N(\mu, \sigma^2)$ is σ^2 ✓

CC BY NC SA

Reflection on Gaussian Properties

Gaussian properties we've proven:

1. If $X \sim N(0,\ 1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu,\ \sigma^2)$ ✓
2. If $\mathbf{Y} \sim N(\mu, \sigma^2)$, and $\mathbf{X} = (Y - \mu)/\sigma$, then $\mathbf{X} \sim N(0,\ 1)$ ✓
5. The mean of $\mathbf{X} \sim N(\mu, \sigma^2)$ is μ ✓
6. The variance of $X \sim N(\mu, \sigma^2)$ is σ^2 ✓

Thus we have that:

- Any affine transformation of a Gaussian is also Gaussian.
- A Gaussian RV has exactly two free parameters which are its mean and variance.
  - Not true for most other RVs, e.g. if $X \sim \mathbf{Geometric}(p)$, then we have that $\mathbf{Var}(X) = (1-p)/p^2$, $\mathbb{E}[X] = 1/p$. Can't set independently.

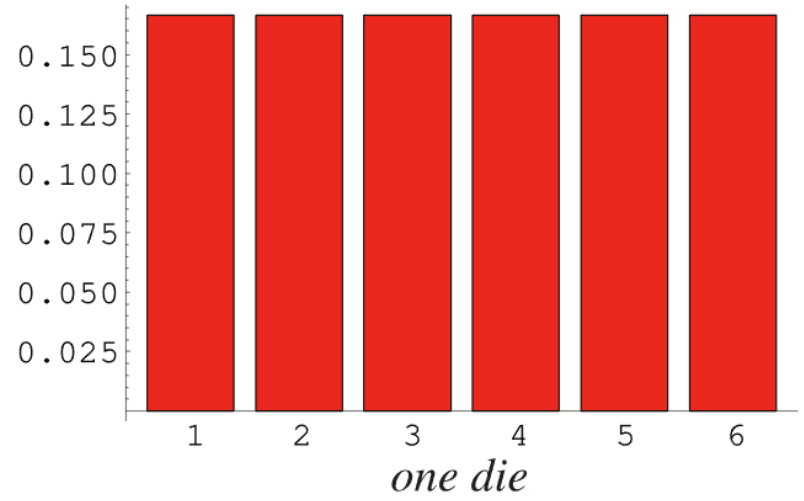
CC BY NC SA

The Central Limit Theorem

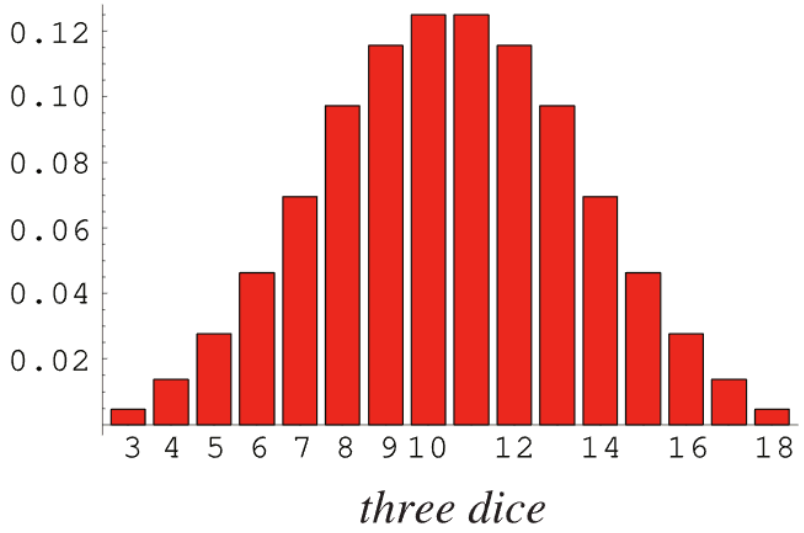
Lecture 25, CS70 Fall 2026

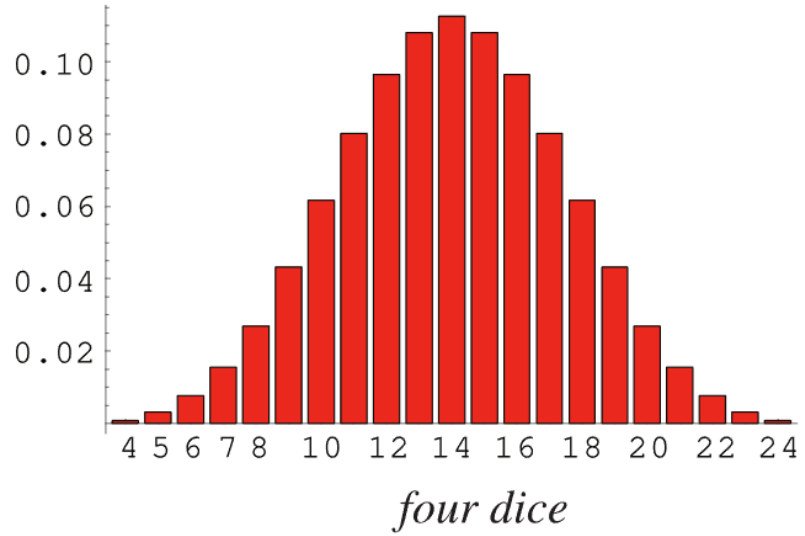
CC BY NC SA

Summing Dice







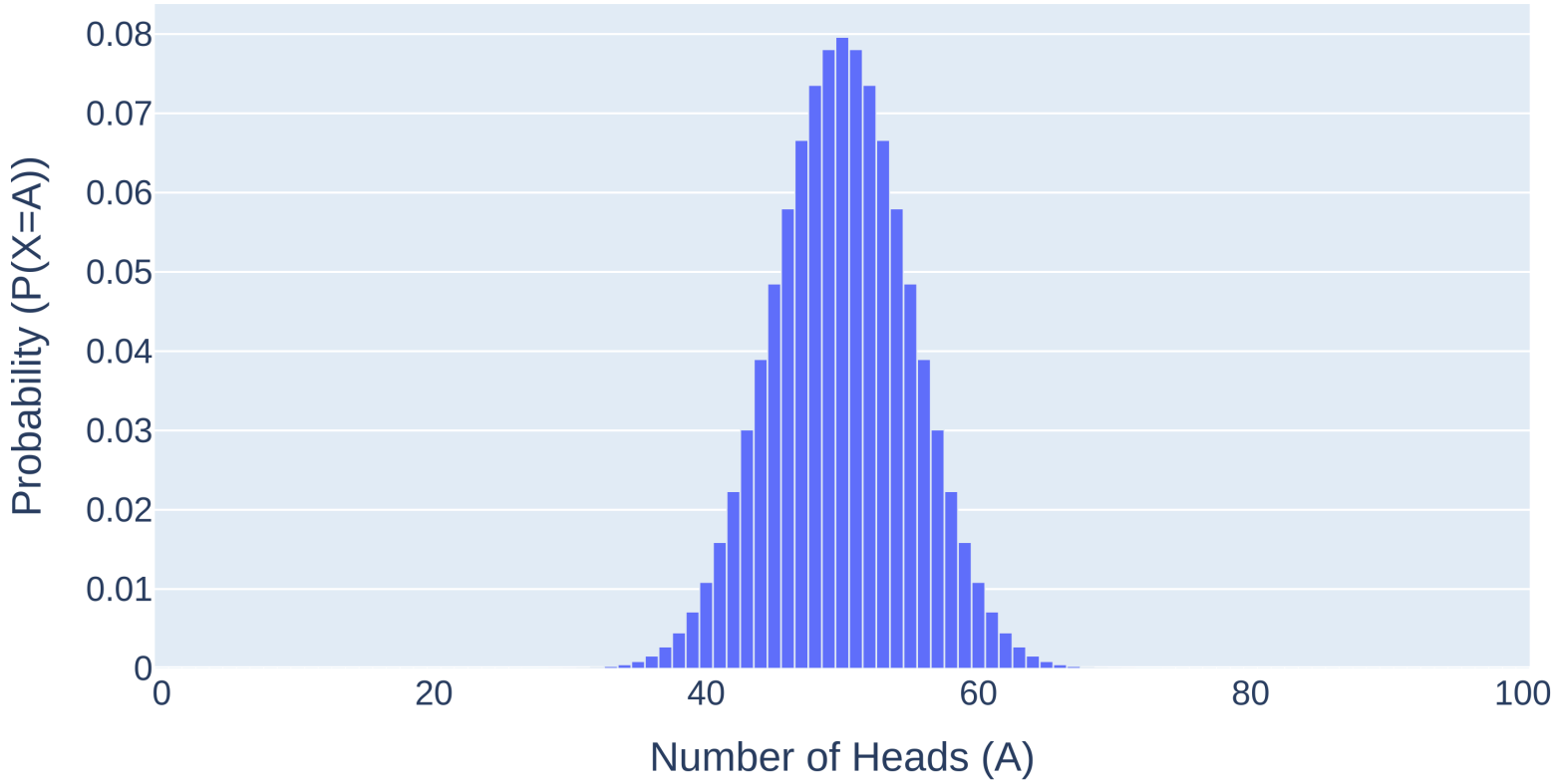


It's not a coincidence that summing four dice looks kind of Gaussian!

The Binomial Distribution

We've also seen the Binomial distribution, which is the sum of Bernoulli random variables. For example the Binomial distribution for $n = 100$, $p = 0.5$ is given below.

- Again, it looks kinda Gaussian.



CC BY NC SA

This Trend Holds

Amazingly, it doesn't matter what you start with!

- The sum of a bunch of i.i.d. exponentials, geometrics, that weird distribution where you hand back homeworks, etc. is always going to look Gaussian!

We've already seen a similar phenomenon once called the weak law of large numbers. Let's review.

CC BY NC SA

Review: Weak Law of Large Numbers

Recall the **Weak Law of Large Numbers**.

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu < \infty$, and $\mathbf{Var}(X_i) = \sigma^2 < \infty$, then for every $\epsilon > 0$, we have that:

$$\lim_{n \to \infty} P\left(\left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq \epsilon \right) = 0$$

Informal interpretation: As we run the same experiment over and over, the difference between the empirical average and the expectation converges to zero.

CC BY NC SA

The Central Limit Theorem

Our new idea is the **Central Limit Theorem**.

If X_1, X_2, $X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu < \infty$, and $\mathbf{Var}(X_i) = \sigma^2 < \infty$, then as $n \to \infty$, we have that:

$$\frac{X_1 + X_2 + \cdots + X_n - \mu n}{\sigma \cdot \sqrt{n}} \to N(0, 1)$$

Informal interpretation: If we add a bunch of i.i.d. random variables, the resulting sum looks like a normal distribution.

- By subtracting out $\mu \cdot n$, the mean becomes zero.
- By dividing by $\sigma \cdot \sqrt{n}$, the standard deviation becomes one.

CC BY NC SA

The Central Limit Theorem

Our new idea is the **Central Limit Theorem**.

If $X_1,\ X_2,\ X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu < \infty$, and $\mathbf{Var}(X_i) = \sigma^2 < \infty$, then as $n \to \infty$, we have that:

$$\frac{X_1 + X_2 + \cdots + X_n - \mu n}{\sigma \cdot \sqrt{n}} \to N(0, 1)$$

What if we don't subtract out the mean and divide out the $\sigma\sqrt{n}$? Then:

$$S = X_1 + X_2 + \cdots + X_n \to N(\mu n,\ \sigma^2 n)$$

CC BY NC SA

Example Usage of the Central Limit Theorem

Suppose we run our classic CS70 experiment: We take all students' homework, shuffle it, and hand it back randomly. Let X_i be the number of students who get their homework back on the ith experiment.

Suppose that the sum we get back from 1000 high schools is $\mathbf{S} = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. What is the Z-score for this result?

- Recall: $\mathbb{E}[X_i] = 1$ and $\mathbf{Var}(X_i) = 1$

CC BY NC SA

Example Usage of the Central Limit Theorem

Suppose we run our classic CS70 experiment: We take all students' homework, shuffle it, and hand it back randomly. Let X_i be the number of students who get their homework back on the ith experiment.

Suppose that the sum we get back from 1000 high schools is $\mathbf{S} = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. What is the Z-score for this result?

- Recall: $\mathbb{E}[X_i] = 1$ and $\mathbf{Var}(X_i) = 1$
- First we compute $\mathbb{E}[S] = 1000$ and $\mathbf{Var}(S) = 1000$. This works since the X_i are independent (so uncorrelated).
- To compute the Z-score, we have $\frac{1231-\mu}{\sigma} = \frac{1231-1000}{31.6} = 7.31$

This is an outrageously high Z-score!

CC BY NC SA

Example Usage of the Central Limit Theorem

Suppose we run our classic CS70 experiment: We take all students' homework, shuffle it, and hand it back randomly. Let X_i be the number of students who get their homework back on the ith experiment.

Suppose that the sum we get back from 1000 high schools is $\mathbf{S} = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. What is the Z-score for this result? 7.31.

- $P(S \geq 1{,}231 \mid \textbf{nothing weird happened}) = 1 - \Phi(7.31) \approx 1.33 \times 10^{-13}$

In other words, if this were just chance, a sum this large would come up about once in 7.5 trillion tries.

- This is known in experimental science as a "p-value".
- If the "null hypothesis" that nothing weird happened is true, a result this extreme has probability $p \approx 1.33 \times 10^{-13}$.

CC BY NC SA

Question

Suppose we run an experiment to look for a predicted physics mechanism. If the mechanism does not exist, we'd expect about 120 neutrinos from background alone.

We observe $X = 174$. If the mechanism does not exist, $P(X \geq 174) = 0.0000023$.

Which of these are true? Select all that apply.

A. The probability that the mechanism does not exist is 0.0000023.

B. The probability that our result was just a fluke is 0.0000023.

C. The probability that the mechanism exists is 0.9999977.

D. If the mechanism does not exist, a result at least this extreme happens with probability 0.0000023.

joshh.ug/F

CC BY NC SA

Extra: Summing Gaussians

Lecture 25, CS70 Fall 2026

CC BY NC SA

The Sum of Independent Gaussians is Gaussian

Another remarkable property of the Gaussian distribution is that the sum of two independent Gaussians is always also Gaussian.

Which of the following distributions have this property?

- Bernoulli
- Binomial
- Uniform

https://joshh.ug/C

- This is a tricky question!



CC BY NC SA

The Sum of Independent Gaussians is Gaussian

Another remarkable property of the Gaussian distribution is that the sum of two independent Gaussians is always also Gaussian.

Which of the following distributions have this property?

- Bernoulli, no except for the degenerate case where $p = 0$.
- Binomial, yes but only if they have the same p.
- Uniform, no except for the degenerate case where there is only one possible outcome.

(so all of them can be yes or no)

CC BY NC SA

Detour: Joint Distribution of Two Independent Standard Normals

Earlier, we saw that if X and Y are independent, then $f(x,\ y) = f_X(x) \cdot f_Y(y)$.

Thus, the joint distribution of two standard normal (a.k.a. Gaussian) RVs is:

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

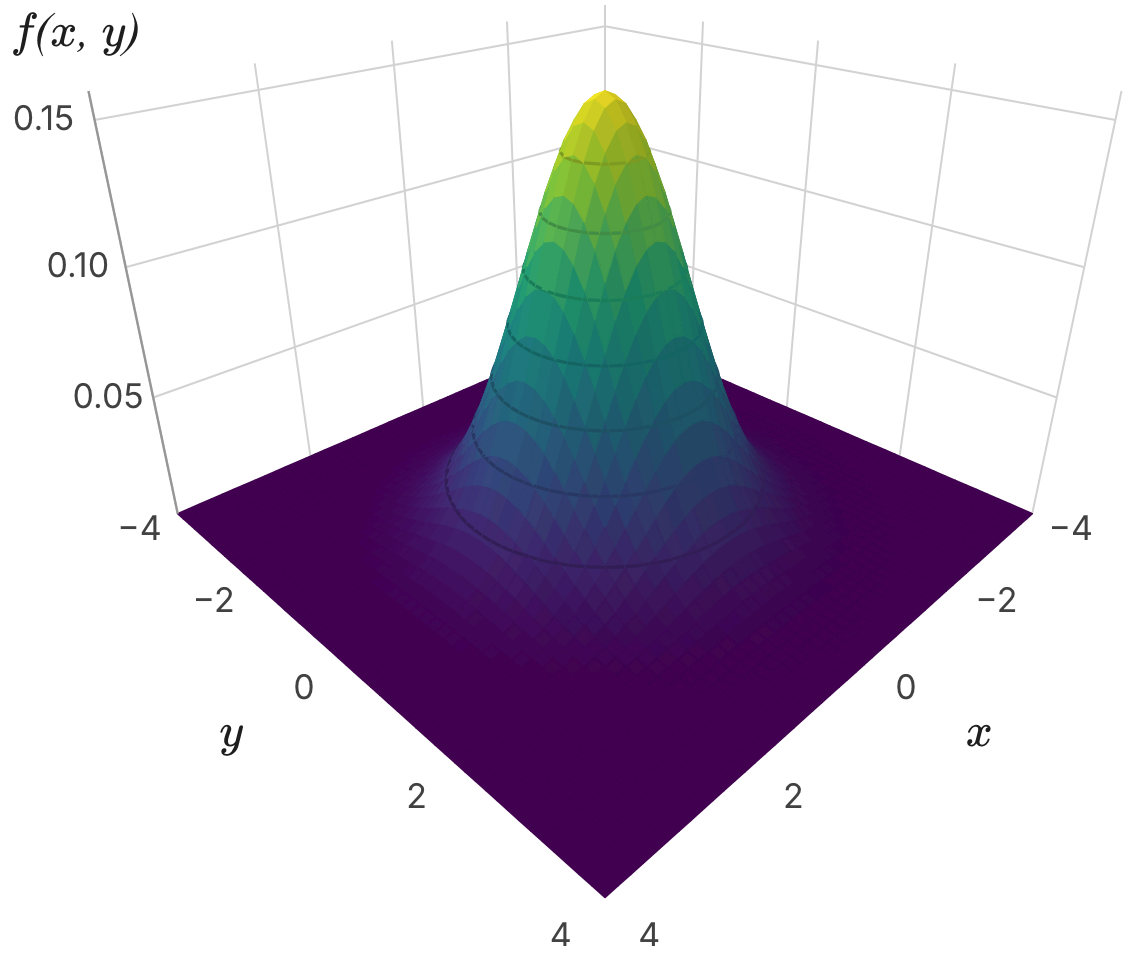
$$= \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

CC BY NC SA

Detour: Visualizing the Joint Distribution of Two Independent Standard Normals

Below is a plot of their joint PDF. Drag it to turn it.

- Interactive version: https://joshh.ug/cs70/joint_gaussian_pdf.html

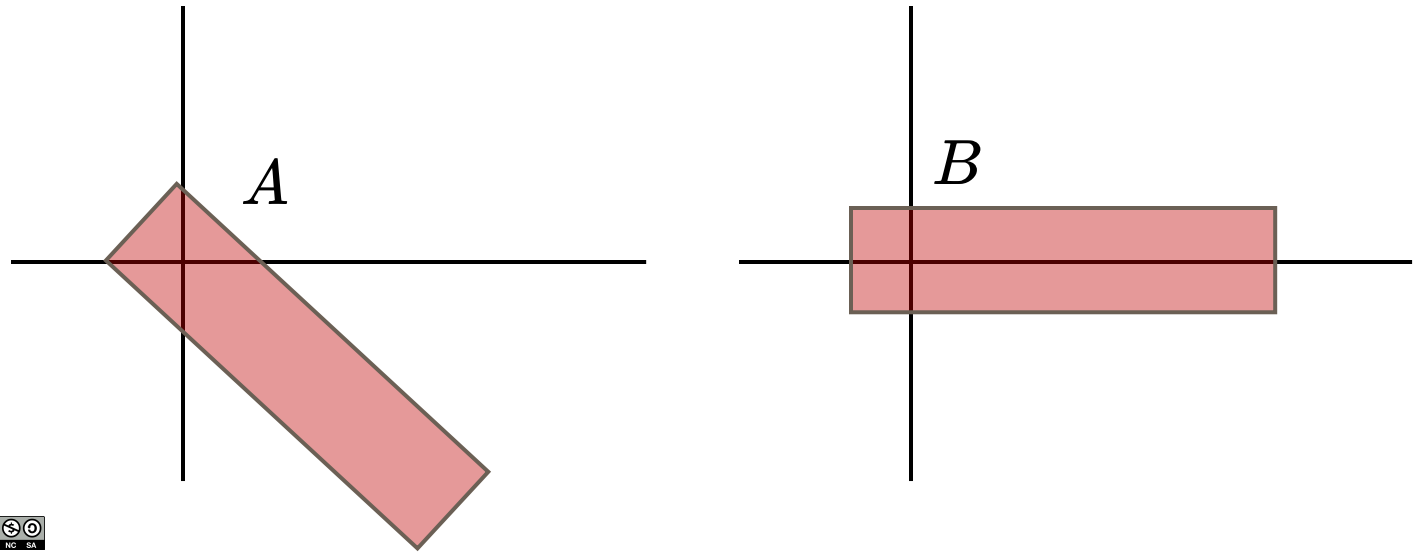


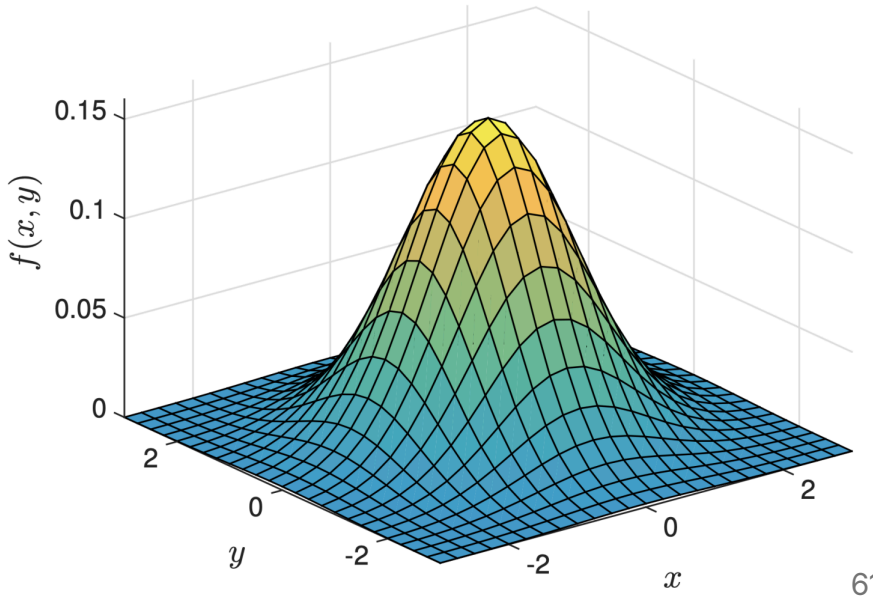
CC BY NC SA

Rotational Symmetry of Joint Normals

To prove the sum of two independent standard Gaussians is also Gaussian, we'll rely on the rotational symmetry of the joint distribution of two normals.

By rotational symmetry, we mean the probability depends only on the radius, not on the angle. Example: The two events A and B below have the same probability.

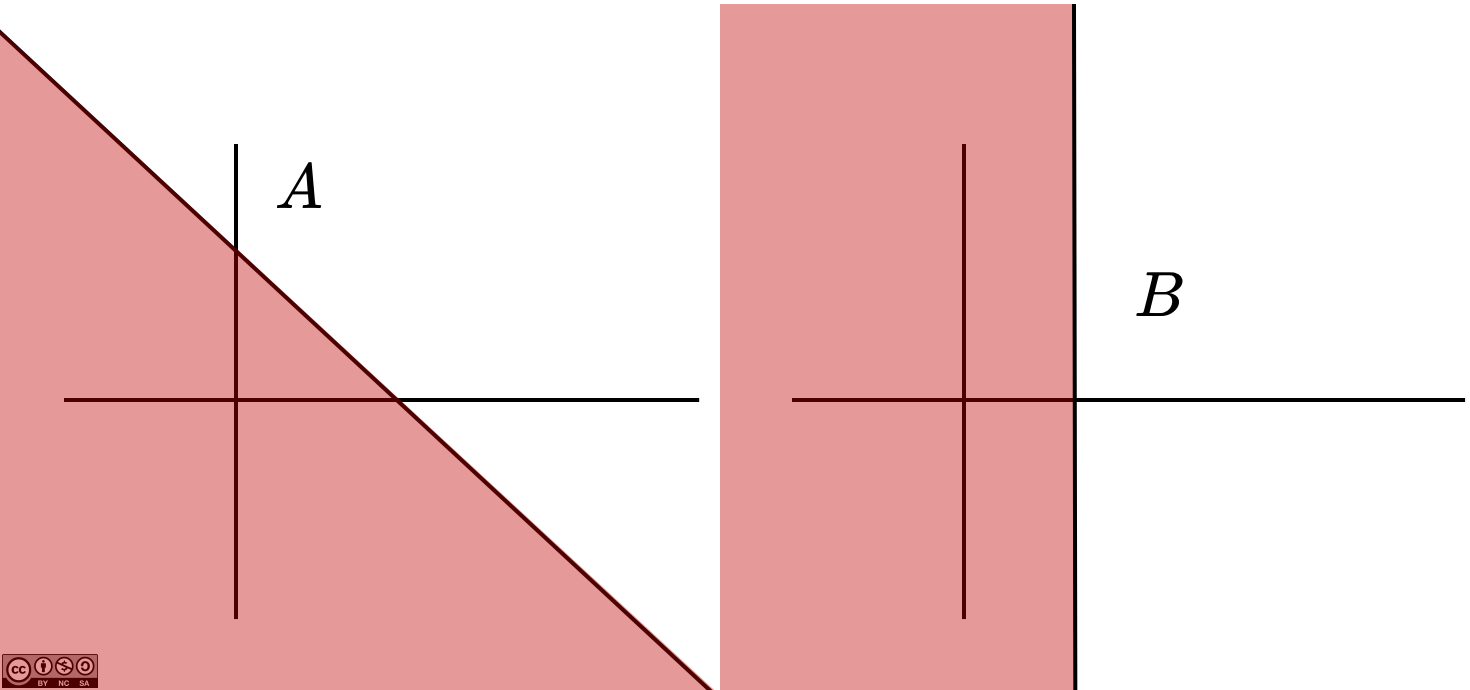


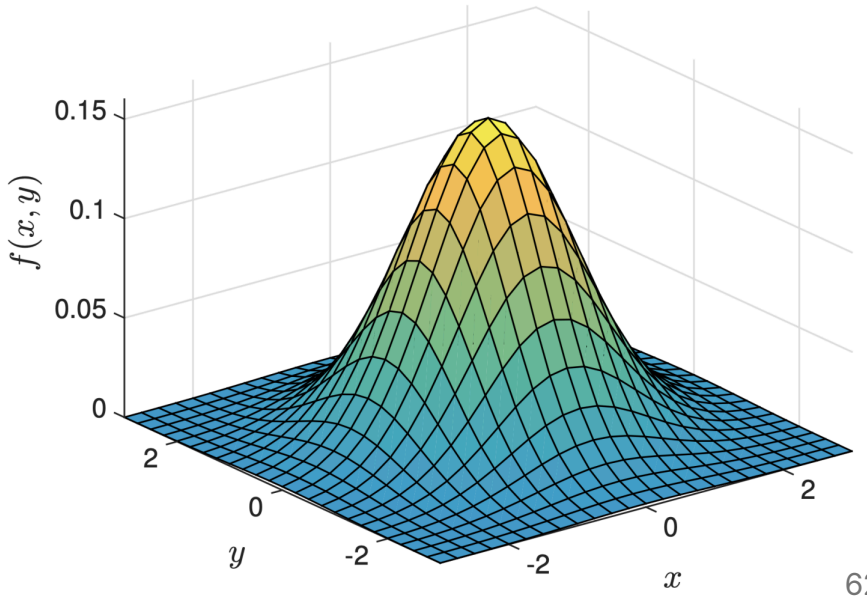


Rotational Symmetry of Joint Normals

To prove the sum of two independent standard Gaussians is also Gaussian, we'll rely on the rotational symmetry of the joint distribution of two normals.

By rotational symmetry, we mean the probability depends only on the radius, not on the angle. Example: The two events A and B below have the same probability.



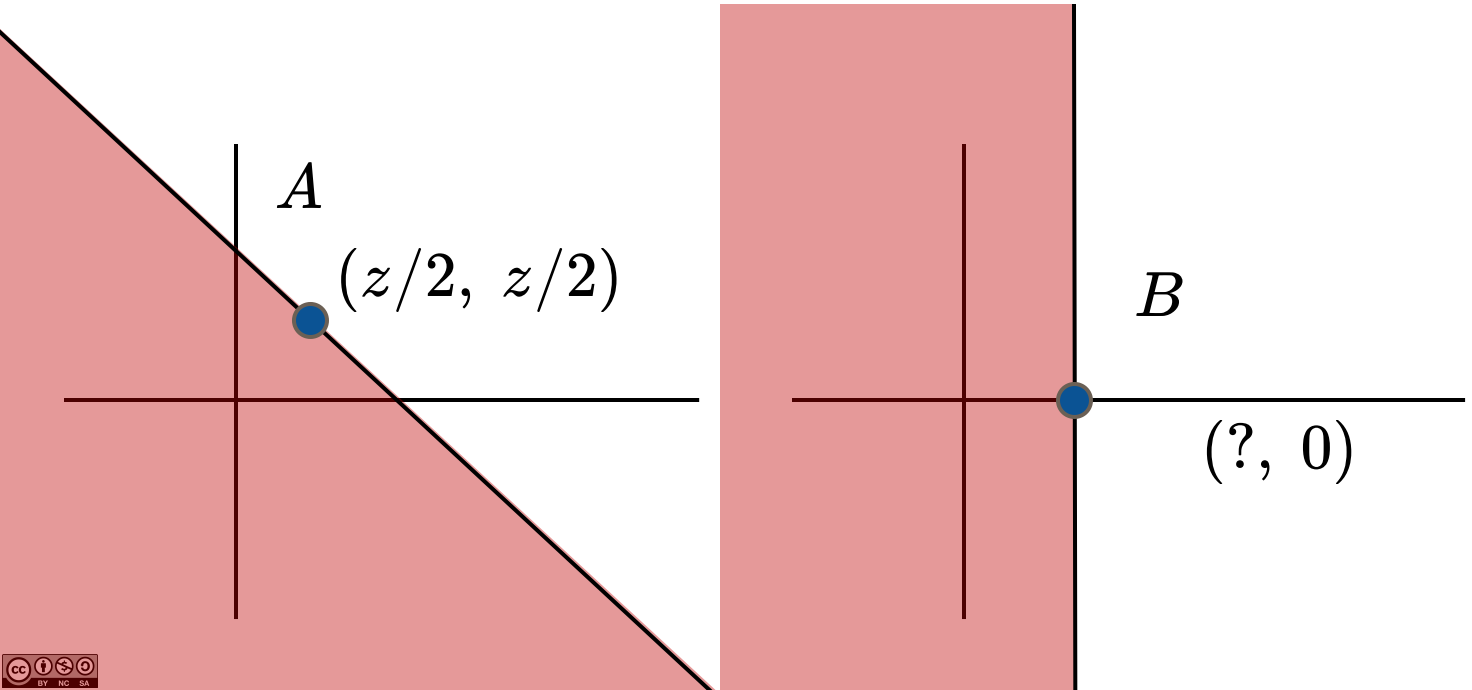


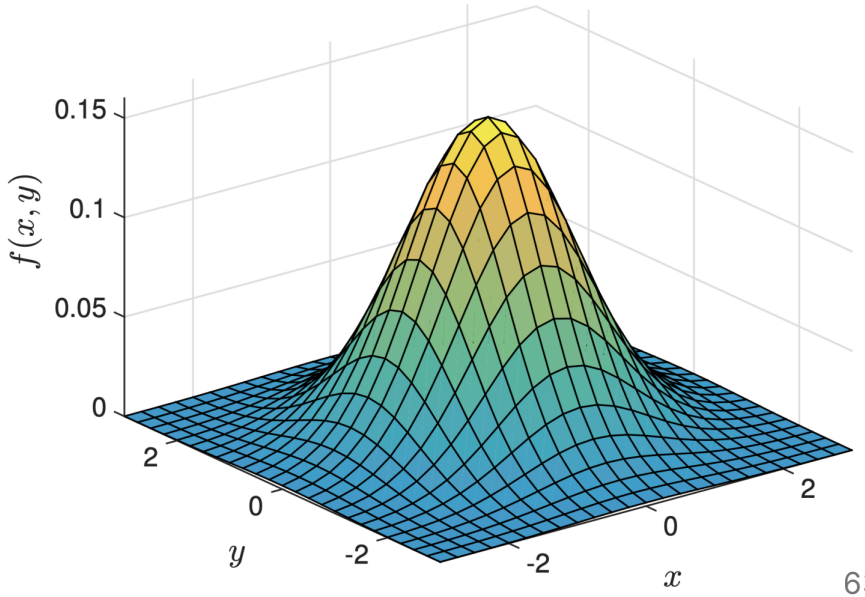
Rotational Symmetry of Joint Normals

To prove the sum of two independent standard Gaussians is also Gaussian, we'll rely on the rotational symmetry of the joint distribution of two normals.

By rotational symmetry, we mean the probability depends only on the radius, not on the angle. Example: The two events A and B below have the same probability.

https://joshh.ug/D

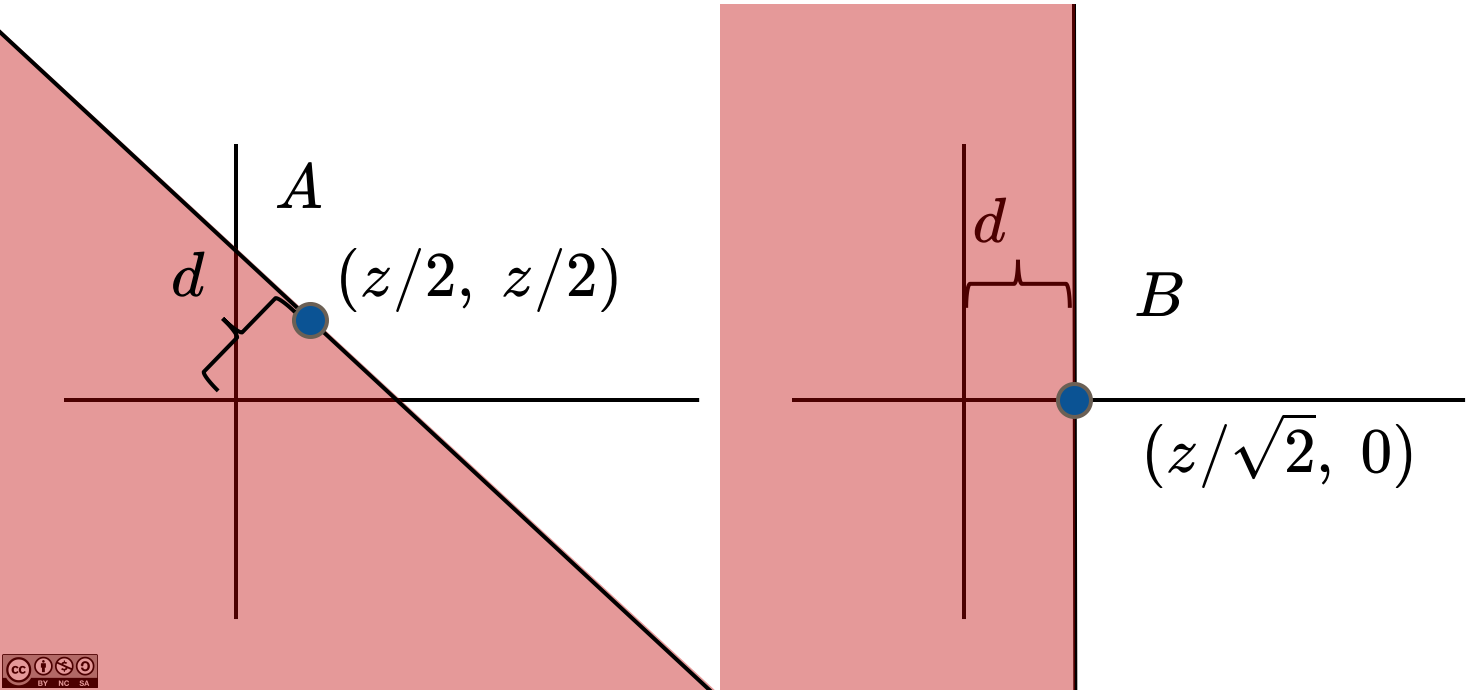




Rotational Symmetry of Joint Normals

To prove the sum of two independent standard Gaussians is also Gaussian, we'll rely on the rotational symmetry of the joint distribution of two normals.

By rotational symmetry, we mean the probability depends only on the radius, not on the angle. Example: The two events A and B below have the same probability.

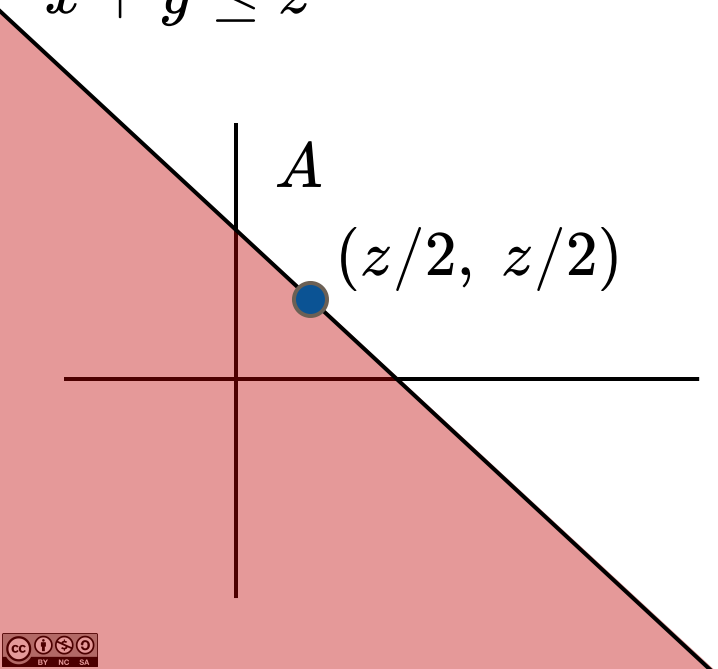


$$d = \sqrt{\frac{z^2}{4} + \frac{z^2}{4}} = \sqrt{\frac{z^2}{2}} = \frac{z}{\sqrt{2}}$$

Proof: Sum of Two Independent Standard Gaussians is Gaussian

Theorem: If $X \sim N(0,1)$ and $Y \sim N(0,\ 1)$ are independent, then $Z = X + Y \sim N(0,\ 2)$

Let $Z = X + Y$, then $P(Z \leq z) = P(X + Y \leq z) = P((X, Y) \in A)$



CC BY NC SA

Proof: Sum of Two Independent Standard Gaussians is Gaussian

Theorem: If $X \sim N(0,1)$ and $Y \sim N(0,\ 1)$ are independent, then $\mathbf{Z = X + Y \sim N(0,\ 2)}$

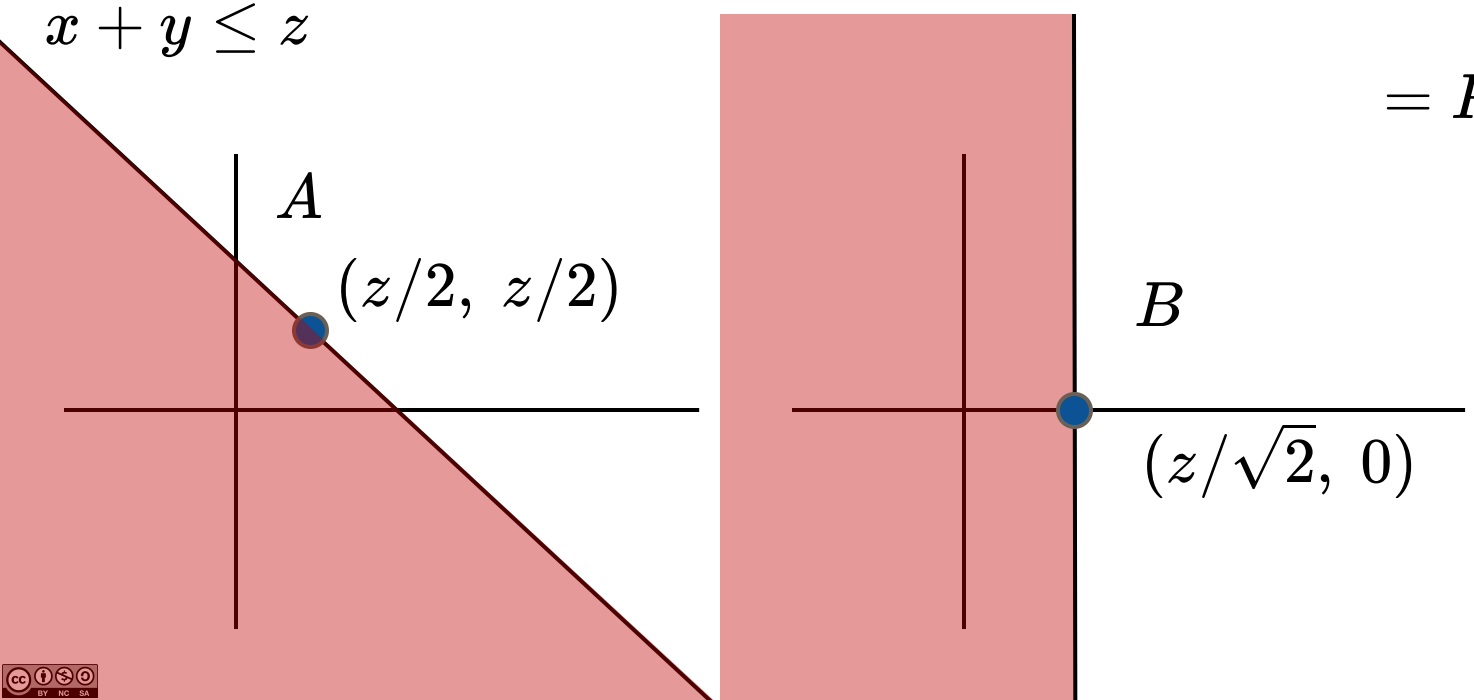
Let $Z = X + Y$, then $P(Z \le z) = P(X + Y \le z) = P((X,Y) \in A)$

$$= P((X,Y) \in B)$$

$$= P\left(X \le \frac{z}{\sqrt{2}}\right)$$

$$= P(\sqrt{2}X \le z) \quad \Rightarrow \quad Z = \sqrt{2}X$$

$$Z \sim \text{Gaussian}(0,2)$$



Note: If this seems a bit too abstract, maybe consider a specific z, e.g. $P(Z \le 3) = P(X \le 3/\sqrt{2})$.

Proof: Sum of Two Scaled Independent Standard Gaussians is Gaussian

Theorem: If $X \sim N(0,1)$ and $Y \sim N(0,\ 1)$ are independent, $\mathbf{Z} = aX + bY \sim N(0,\ a^2 + b^2)$

Let $Z = aX + bY$, then $P(Z \le z) = P(aX + bY \le z) = P((X,Y) \in A)$

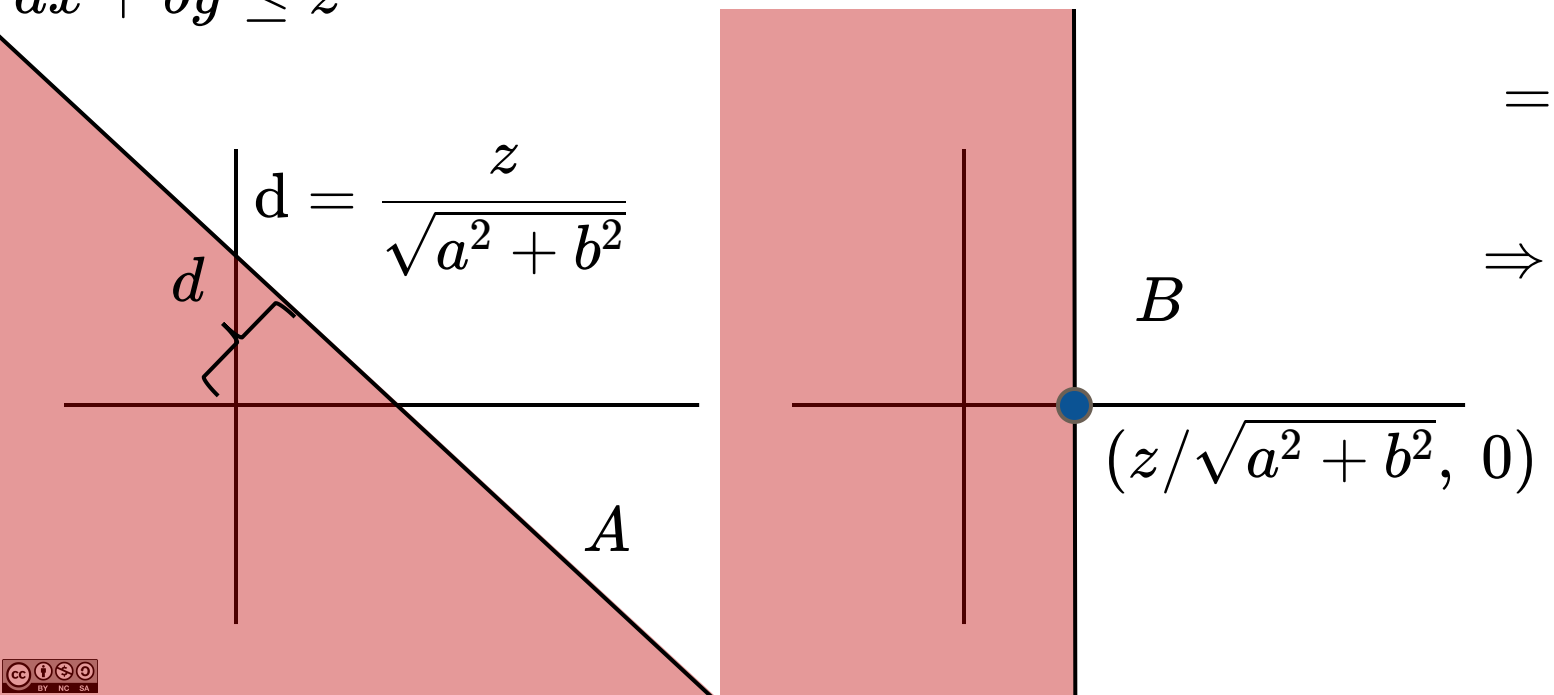
$$= P((X,Y) \in B)$$

$$= P\left(X \le \frac{z}{\sqrt{a^2+b^2}}\right)$$

$$= P(\sqrt{a^2+b^2}X \le z)$$

$$\Rightarrow Z = \sqrt{a^2+b^2}X$$

$$Z \sim \text{Gaussian}(0, a^2+b^2)$$



Proof: Sum of Two Scaled Independent Standard Gaussians is Gaussian

Theorem: If $X \sim N(0,1)$ and $Y \sim N(0,\ 1)$ are independent, $\mathbf{Z} = aX + bY \sim N(0,\ a^2 + b^2)$

Great! The sum of two zero mean Gaussians with variance a^2 and b^2 is easy to understand.

- It is also Gaussian.
- Its variance is just the sum of the variances, i.e. $a^2 + b^2$.

What if it's not zero mean?

- This is straightforward (next slide).

CC BY NC SA

Proof: Sum of Independent Gaussians is Gaussian (general case)

Corollary: If we have independent Gaussian random variables $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, and define $Z = X_1 + X_2$, then $Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Proof: Let $\tilde{X}_1 = \dfrac{X_1 - \mu_1}{\sigma_1}$, $\tilde{X}_2 = \dfrac{X_2 - \mu_2}{\sigma_2}$

We have that

$$Z = X_1 + X_2 = \sigma_1 \tilde{X}_1 + \mu_1 + \sigma_2 \tilde{X}_2 + \mu_2$$

$$= (\mu_1 + \mu_2) + \underbrace{\sigma_1 \tilde{X}_1 + \sigma_2 \tilde{X}_2}_{N(0, \sigma_1^2 + \sigma_2^2)}$$

$$\underbrace{\qquad\qquad}_{N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)}$$

CC BY NC SA

Whew... We're Done!

Today, we've shown that:

- The sum of n i.i.d. RVs is approximately Gaussian for large n (the CLT).
- A Gaussian is characterized by its mean and standard deviation alone.
- Any affine transformation of a Gaussian is also a Gaussian.
- Adding two independent Gaussians gives us another Gaussian with a nice relationship. The resulting Gaussian just has mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

These are handy properties that you'll see again in future classes.

- Admittedly, not the most important thing today, but hopefully you enjoyed the novel proof technique we used to get the third property.

CC BY NC SA

Conclusion

With that, we're all done with our brief semi-formal intro to continuous probability.

- The proof of the central limit theorem is beyond the scope of our class.
- Next week, we'll talk about Markov Chains and how the probability foundation we've built in this course can be used to interpret large language models.

CC BY NC SA