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 $S = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. 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?

Review: Joint Distributions

Lecture 25, CS70 Summer 2025

Review: Joint Distributions (Discrete)

To refresh our memory on joint distributions, let's first consider the joint PMF of two discrete uniform random variables X and Y representing the roll of two independent four-sided dice.

What is the joint PMF of X and Y?

- Recall the PMF just gives the probability of each possible value for X and Y.
- How many rows are in the table specifying the PMF?
- What is the value in each row?

Review: Joint Distributions (Discrete)

To refresh our memory on joint distributions, let's first consider the joint PMF of two discrete uniform random variables X and Y representing the roll of two independent four-sided dice.

What is the joint PMF of X and Y?

- See table to the right.

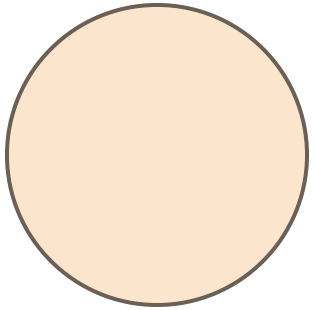
x	y	$P(X = x, Y = y)$
1	1	1/16
1	2	1/16
1	3	1/16
1	4	1/16
2	1	1/16
2	2	1/16
2	3	1/16
2	4	1/16
3	1	1/16
3	2	1/16
3	3	1/16
3	4	1/16
4	1	1/16
4	2	1/16
4	3	1/16
4	4	1/16

Review: Joint PDF of a Uniform Random Variables

Let's next consider the joint PDF of two uniform random variables X and Y representing throwing a dart at a target. The random variables represent the x and y coordinates.

As we saw last time, the joint PDF of X and Y is:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4\pi}, & \text{if } x^2 + y^2 \leq 4 \\ 0, & \text{otherwise} \end{cases}$$



2 foot radius

Independent Random Variables

We saw X and Y are independent continuous random variables if for all $a < b$ and $c < d$.

$$P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b) \cdot P(c \leq Y \leq d)$$

Or equivalently: If $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ for all x, y

The Gaussian Random Variable

Lecture 25, CS70 Summer 2025

Warmup Exercise / Detour: Prove this Lemma

Before we get talking about Gaussians, 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 $Y = aX + b$, then

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

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.

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:

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

X is a "standard" Normal (or "standard" Gaussian) if it has parameters $\mu = 0$ and $\sigma^2 = 1$.

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

Note: It's no coincidence that the parameters of a Gaussian use the same symbols we usually use for mean and variance. We'll come back to this.

Gaussian

Before we dig into what the Gaussian distribution means and why it matters, let's establish six properties:

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

We'll discuss these properties as we prove them.

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 $Y = aX + b$, then

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

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 $N(\mu, \sigma^2)$

Warmup Lemma

If $Y = aX + b$, then

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

Property 2

We've shown that if $X \sim N(0,1)$, and 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:

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

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 $\mathrm{X} \sim N(0, 1)$ ✓
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

Next up, let's consider properties 3 and 4.

Expectation of a Standard Gaussian Random Variable

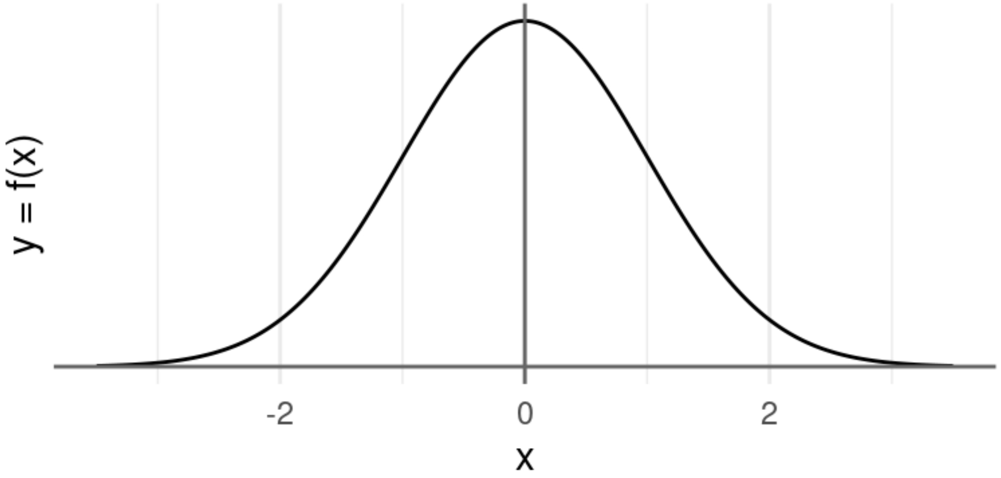
If $X \sim N(0,1)$, then $E[X] = 0$. Proof:

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

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

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

What is the result of this sum? Hint, the PDF looks like this:



Expectation of a Standard Gaussian Random Variable

If $X \sim N(0,1)$, then $E[X] = 0$. Proof:

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

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

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 distribution symmetric around zero is zero. The PDF of a standard Gaussian looks like this:

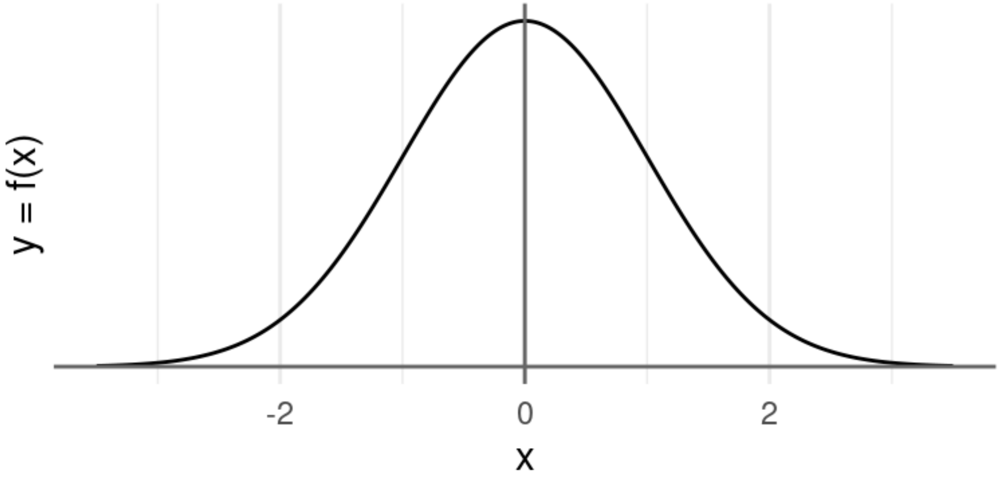


Figure from Stat 20

Variance of a Standard Gaussian Random Variable

If $X \sim N(0,1)$, then $\mathrm{var}(X) = E[X^2] = 1$. Proof:

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

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

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

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

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

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

Variance of a Standard Gaussian Random Variable

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

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

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

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

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

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

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

$$= 0 + 1$$

Intuition or L'Hôpital

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

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 $\mathrm{X} \sim N(0, 1)$ ✓
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

Lastly, let's show properties 5 and 6.

Proof of Properties 5 and 6

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

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

$$= 0 + \mu = \mu$$

Property 3: $E[X] = 0$

And $\text{var}(Y) = \text{var}(\sigma X + \mu) = \sigma^2 \text{var}(X)$

$$= \sigma^2$$

Property 4: $\text{var}(X) = 1$

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 $\mathrm{X} \sim N(0,1)$ ✓
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 ✓

Reflection on Gaussian Properties

Gaussian properties to prove:

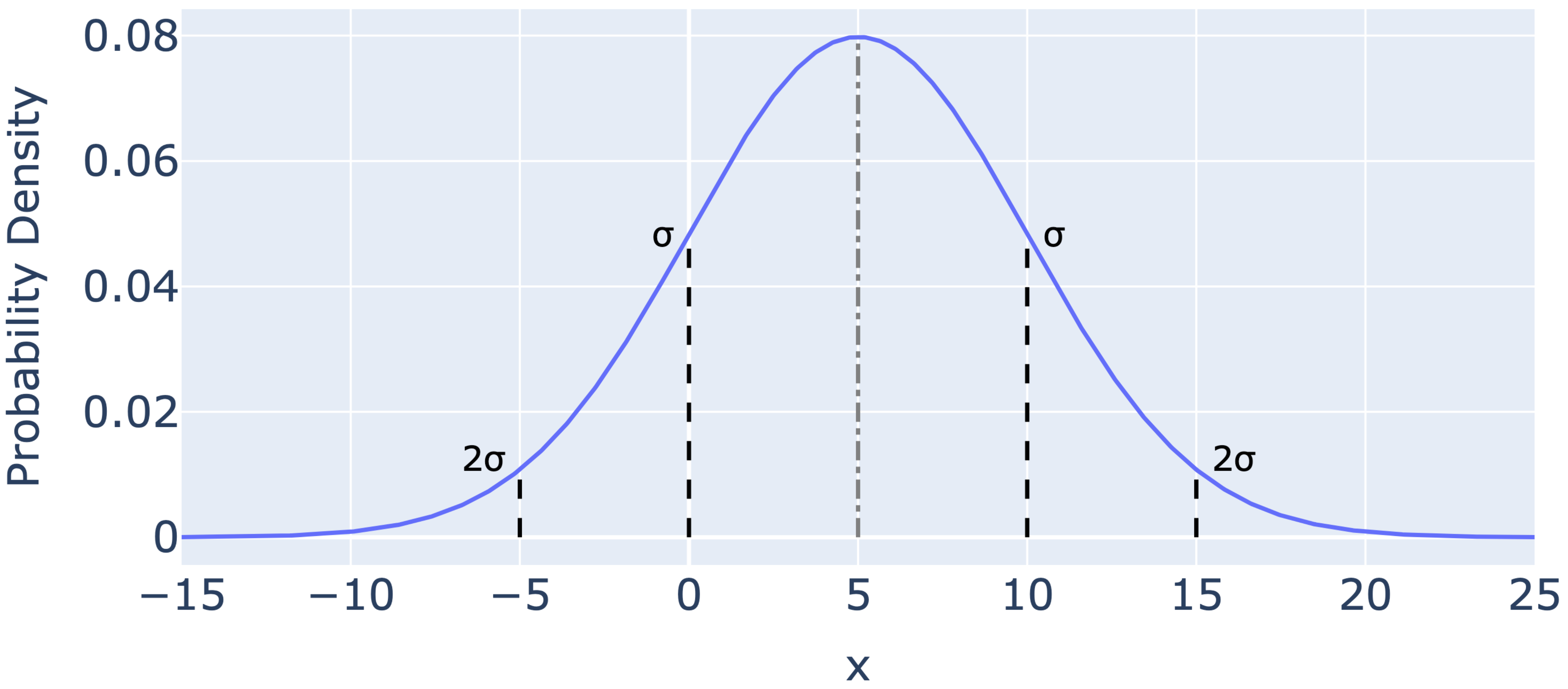
1. If $X \sim N(0,1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu, \sigma^2)$ ✓
2. If $\mathrm{Y} \sim N(\mu, \sigma^2)$, and $\mathrm{X} = (Y - \mu)/\sigma$, then $\mathrm{X} \sim N(0,1)$ ✓
5. The mean of $\mathrm{X} \sim N(\mu, \sigma^2)$ is μ ✓
6. The variance of $X \sim N(\mu, \sigma^2)$ is σ^2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

The properties above reveal two remarkable facts about the Gaussian:

- It has exactly two free parameters which are its mean and variance.
  - Not true for our other RVs, e.g. $\mathrm{var}\big(X \sim \mathrm{Geometric}(p)\big) = (1-p)/p^2$
- Any affine transformation of a Gaussian is also Gaussian.
  - Not true for many RVs, e.g., $\mathrm{X} \sim \mathrm{Exp}(\lambda)$, but $X + 1$ is not exponential.

The Gaussian PDF Visually

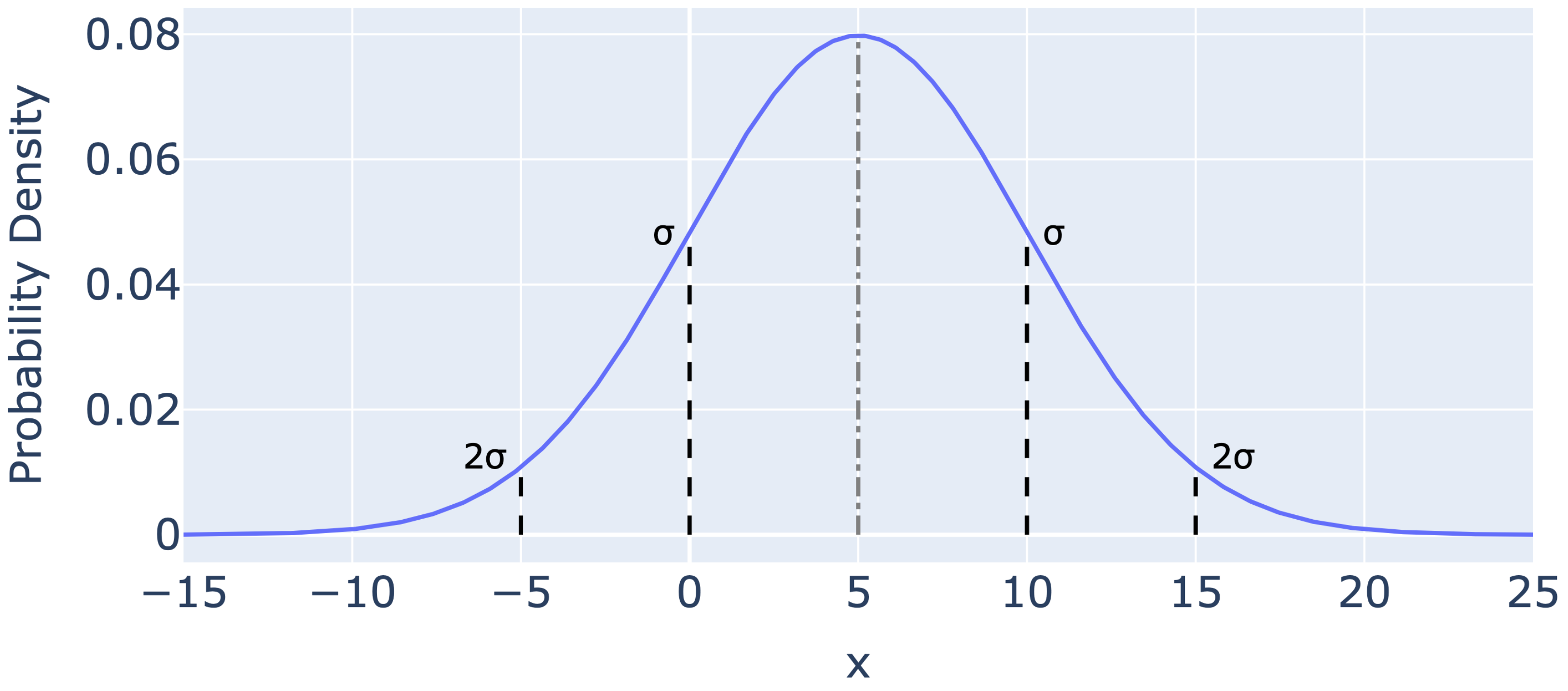
If we have a random variable $X \sim N(5, 25)$, we have the PDF below:



Computing the Probability of an Event

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

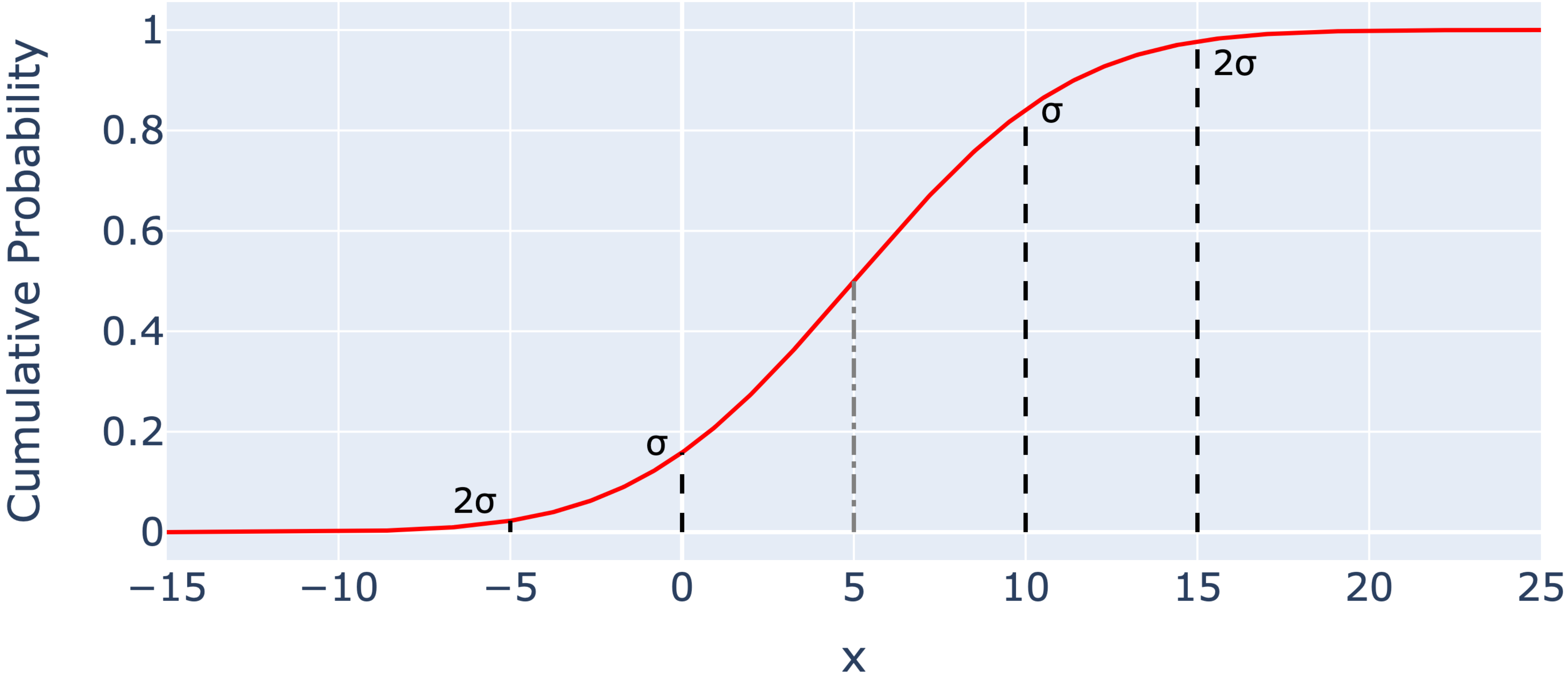
$$\int_{-\infty}^{0} \frac{1}{\sqrt{50\pi}} e^{(x-5)^2/50} \approx 0.1587$$



The CDF of a Gaussian

The CDF of $X \sim N(5, 25)$ is given below.

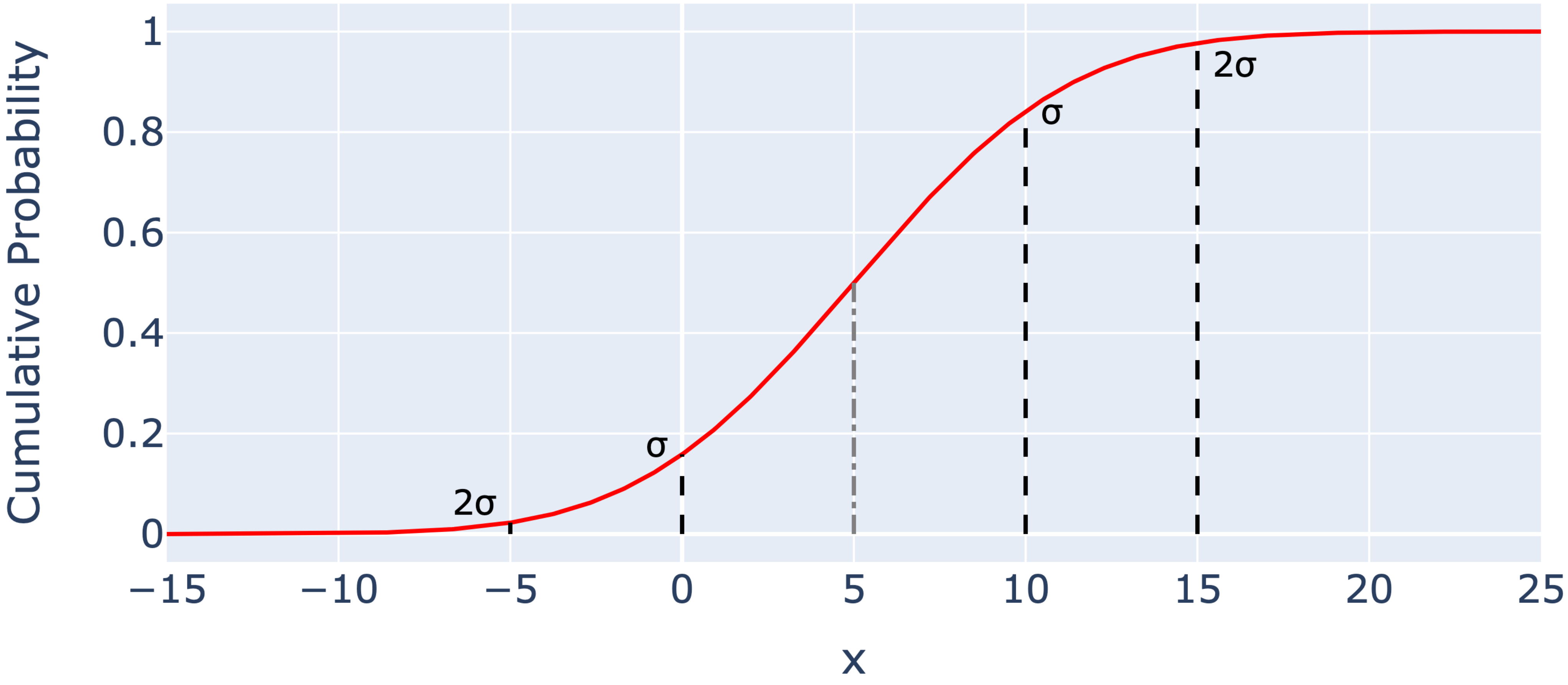
- $P(X \leq 0) \approx 0.1587$
- Note: There is no closed-form expression for the CDF.



The CDF of a Gaussian

Or symbolically we can write

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

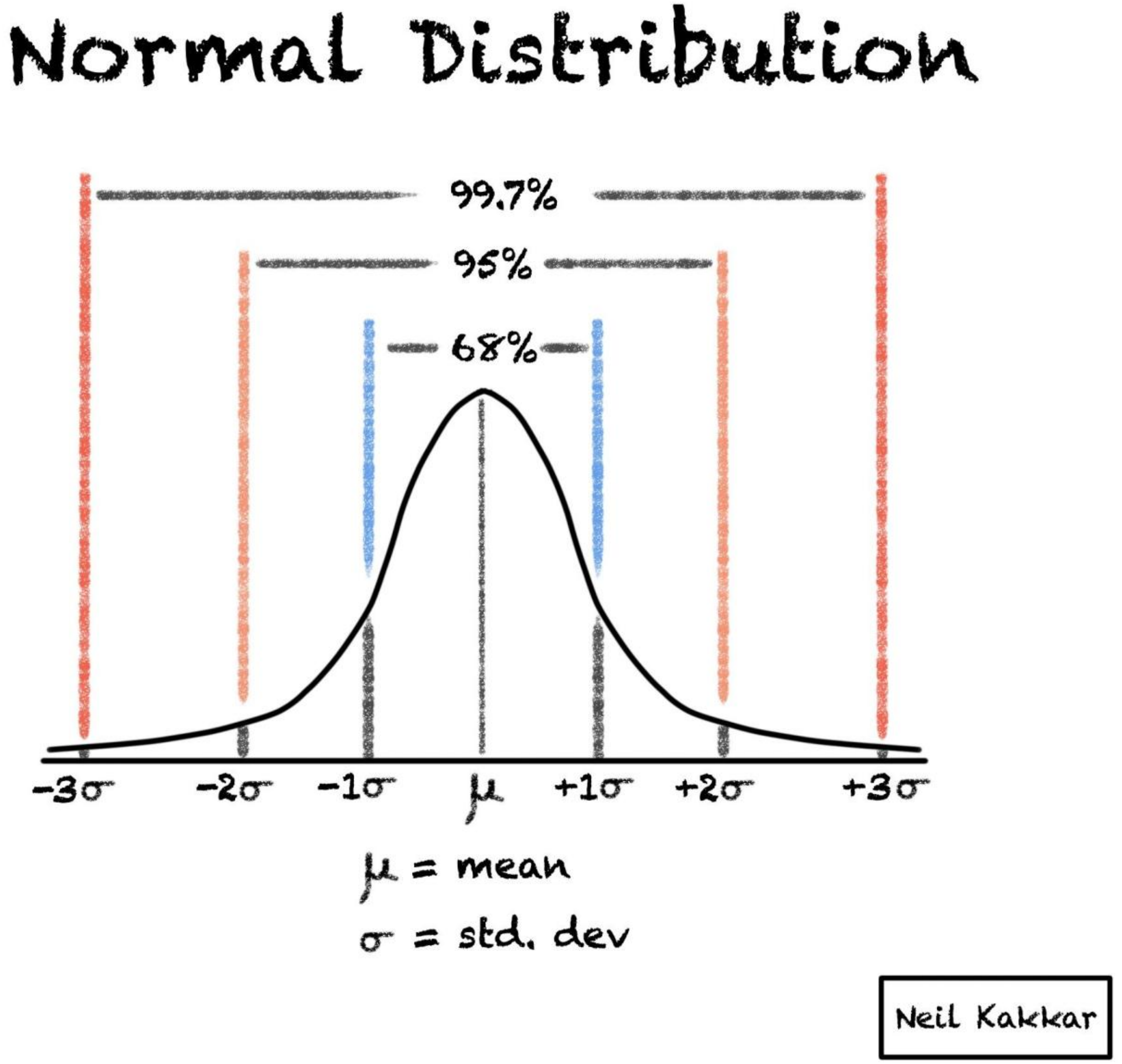


The 68–95–99.7 rule

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

The Gaussian CDF Φ

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

- Note, since we're thinking about a standard normal here, the Gaussian has $\mu = 0$ and $\sigma^2 = 1$, and the arguments to Φ are a z-score.
- $\Phi(3) = 0.997$ is $P(X \leq 3)$, where $X \sim N(0, 1)$.

Example: Assuming 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: $1 - \Phi\left(\frac{7}{3}\right) = 1\%$

Summing Gaussians

Lecture 25, CS70 Summer 2025

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.

Before we get there, which of the following distributions have this property?

- Bernoulli
- Binomial
- Uniform

- This is a tricky question!

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.

Before we get there, 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)

Detour: Joint Distribution of Two Independent Standard Normals

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

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

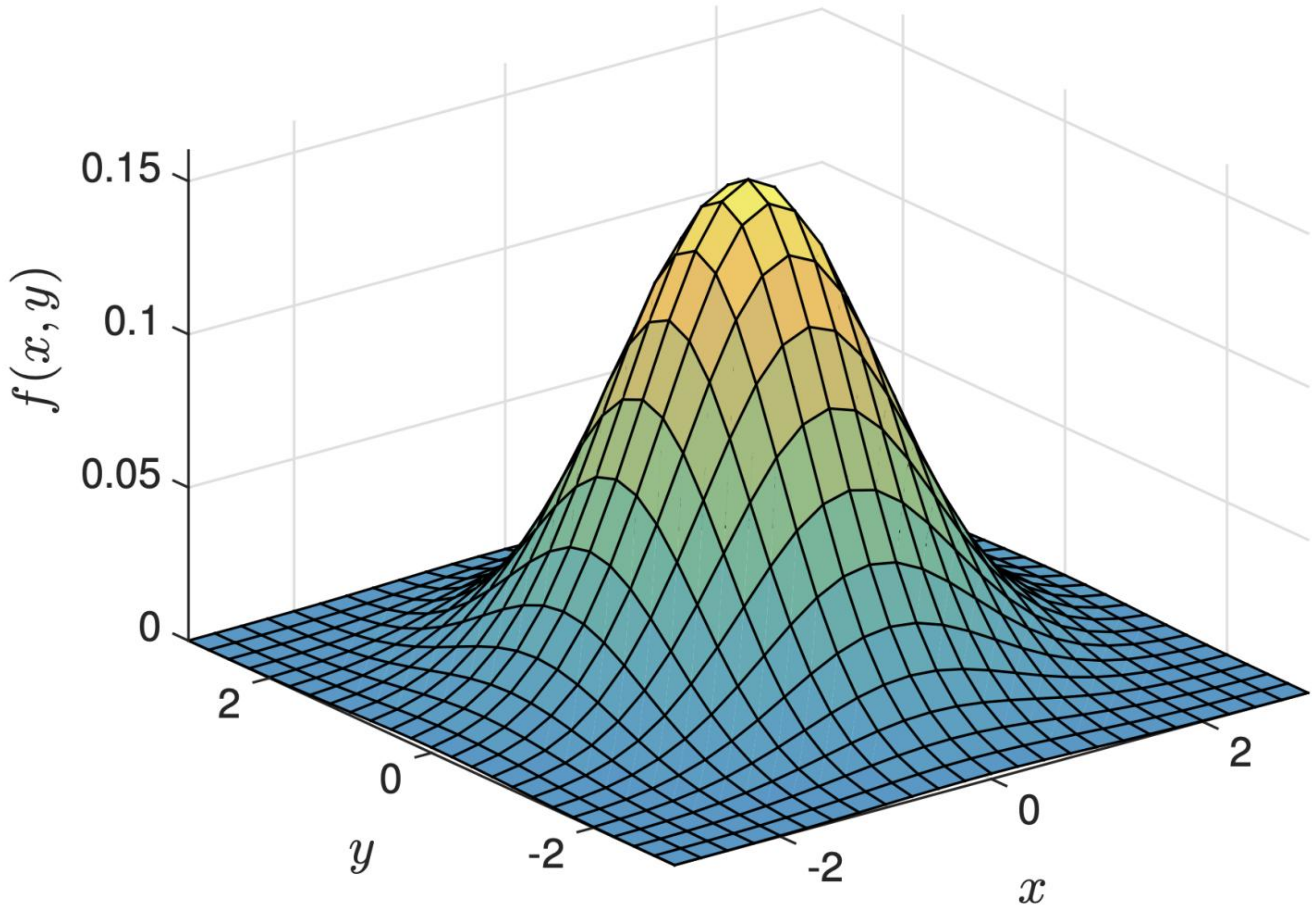
$$f(X, Y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

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

Detour: Visualizing the Joint Distribution of Two Independent Standard Normals

Below we see a plot of the joint PDF (from the notes*).

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

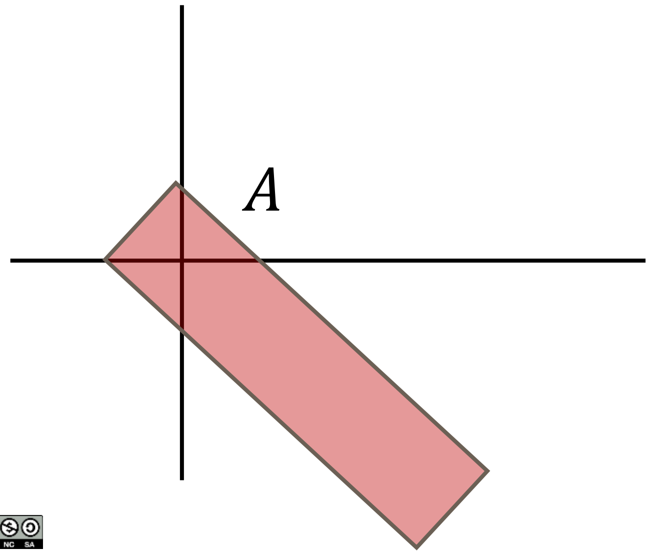


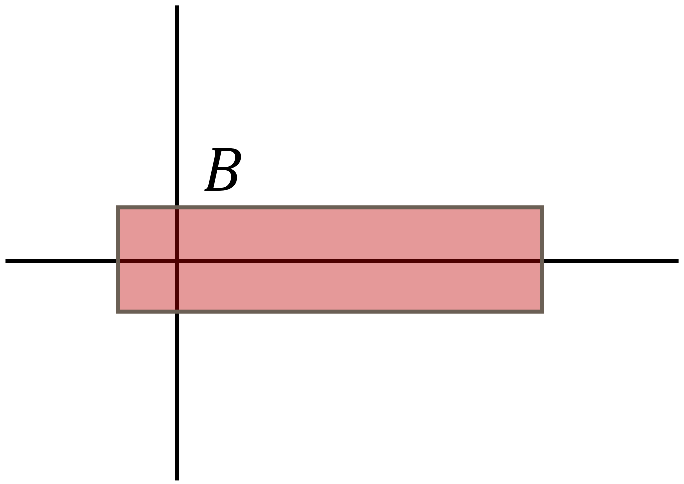
*: Figure ultimately comes from "Why Is the Sum of Independent Normal Random Variables Normal" by by B. Eisenberg and R. Sullivan, Mathematics Magazine, Vol. 81, No. 5.

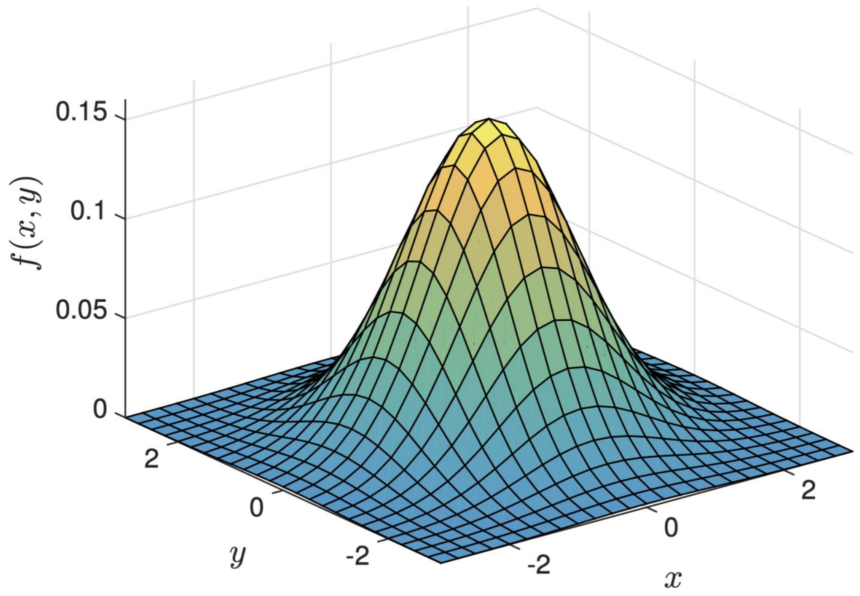
Rotational Symmetry of Joint Normals

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.



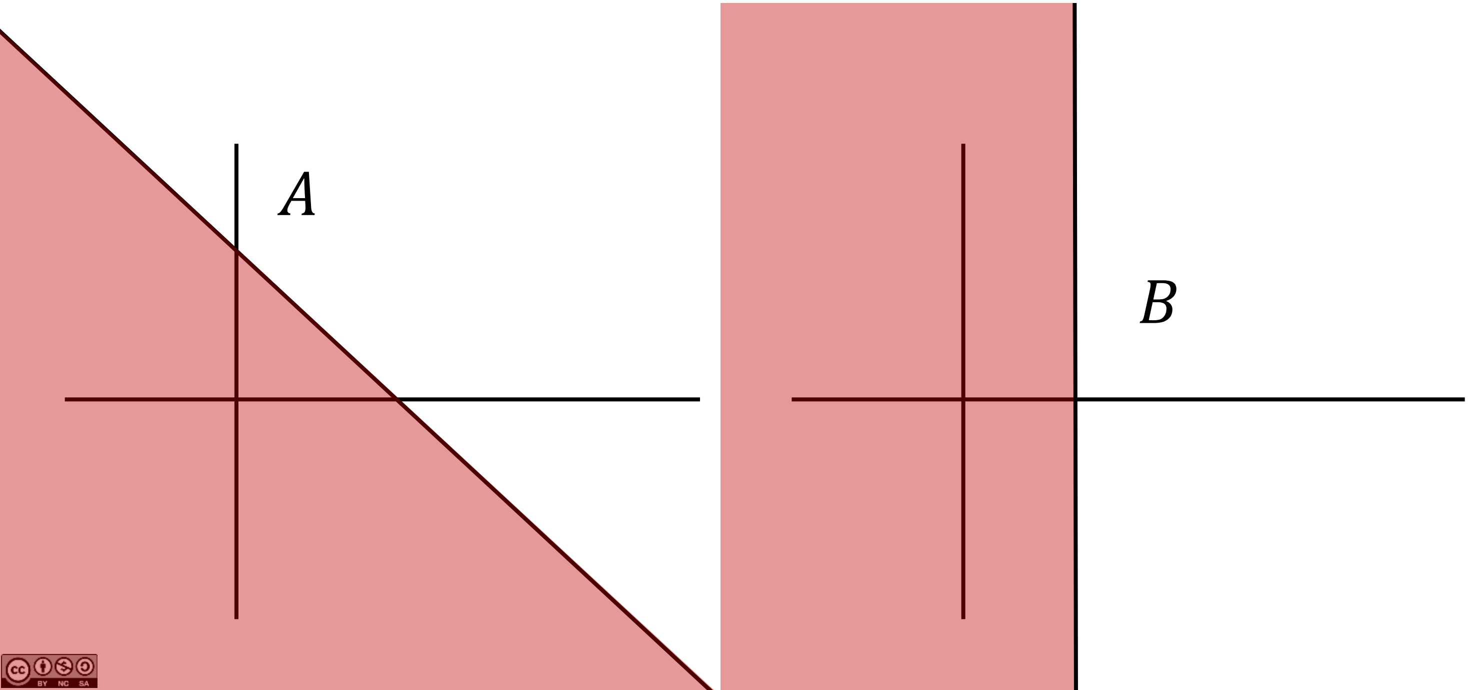


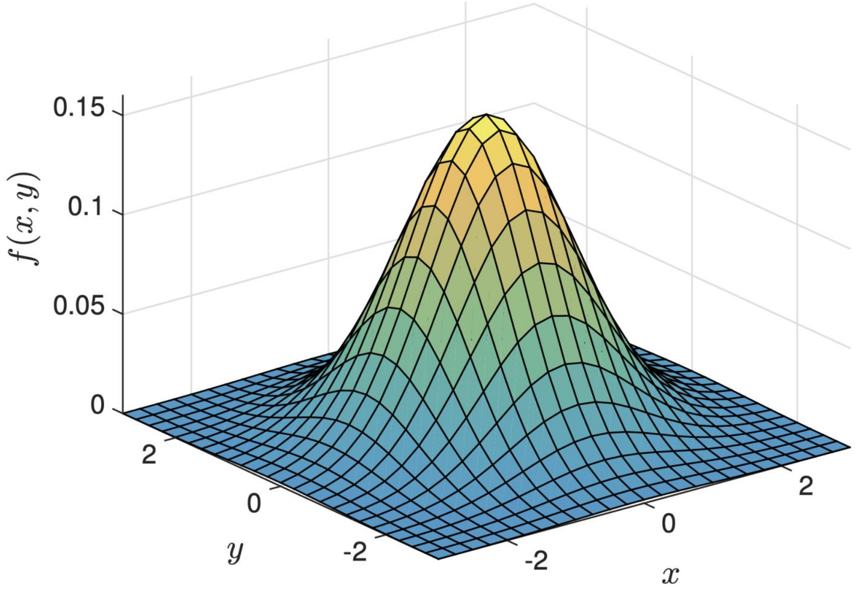


Rotational Symmetry of Joint Normals

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.

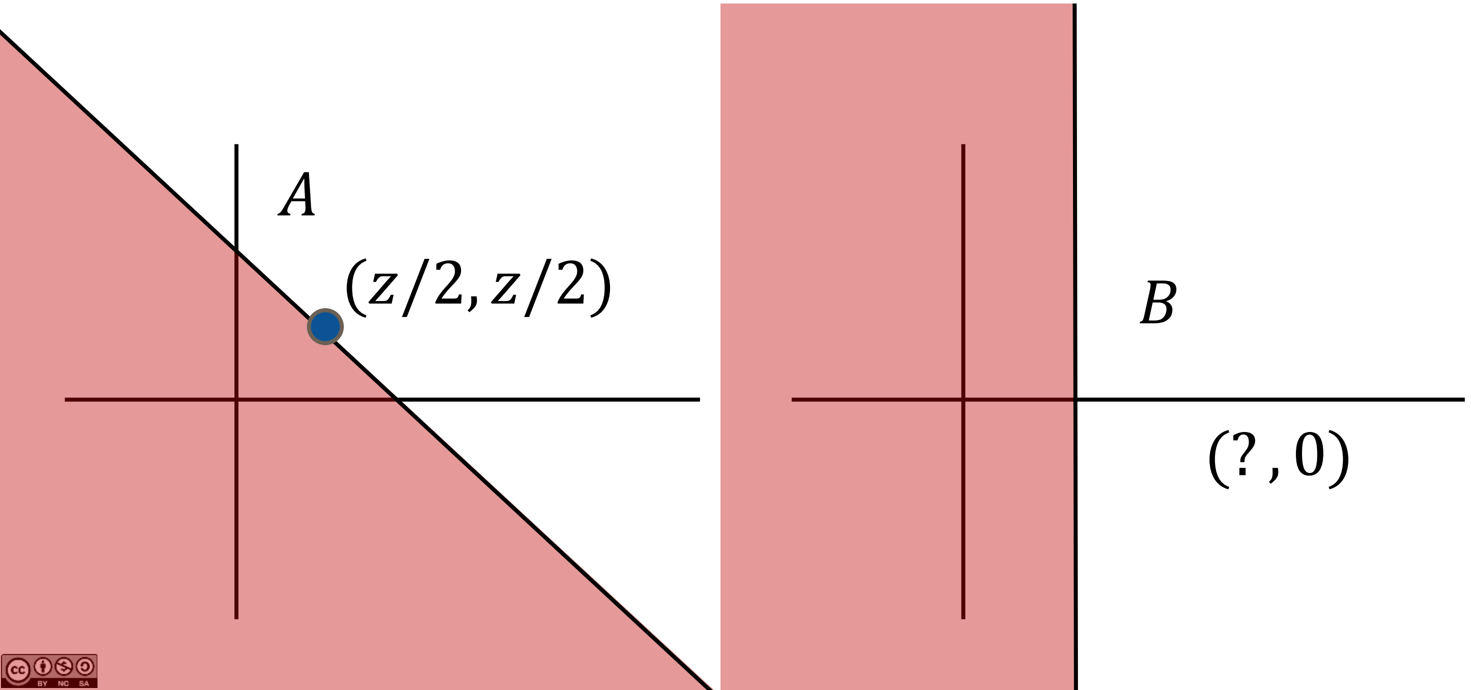


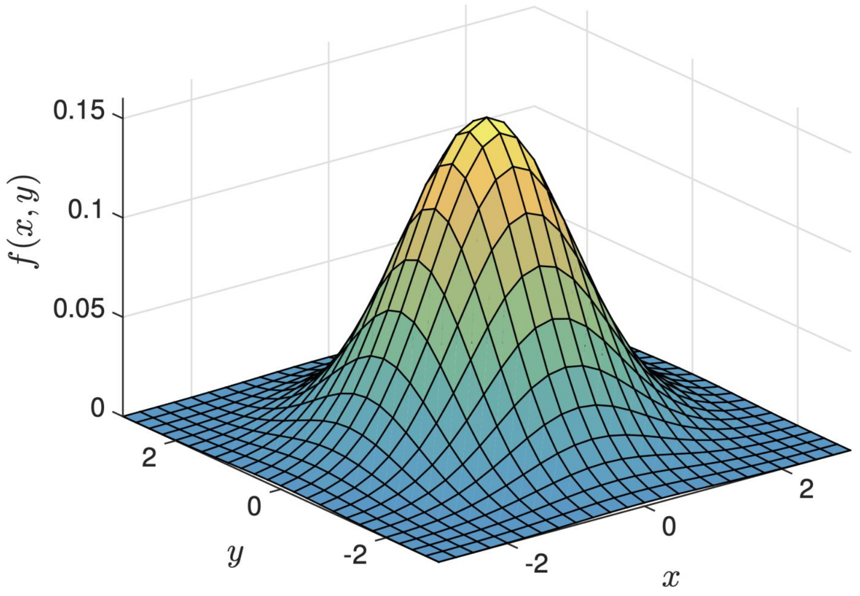


Rotational Symmetry of Joint Normals

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.

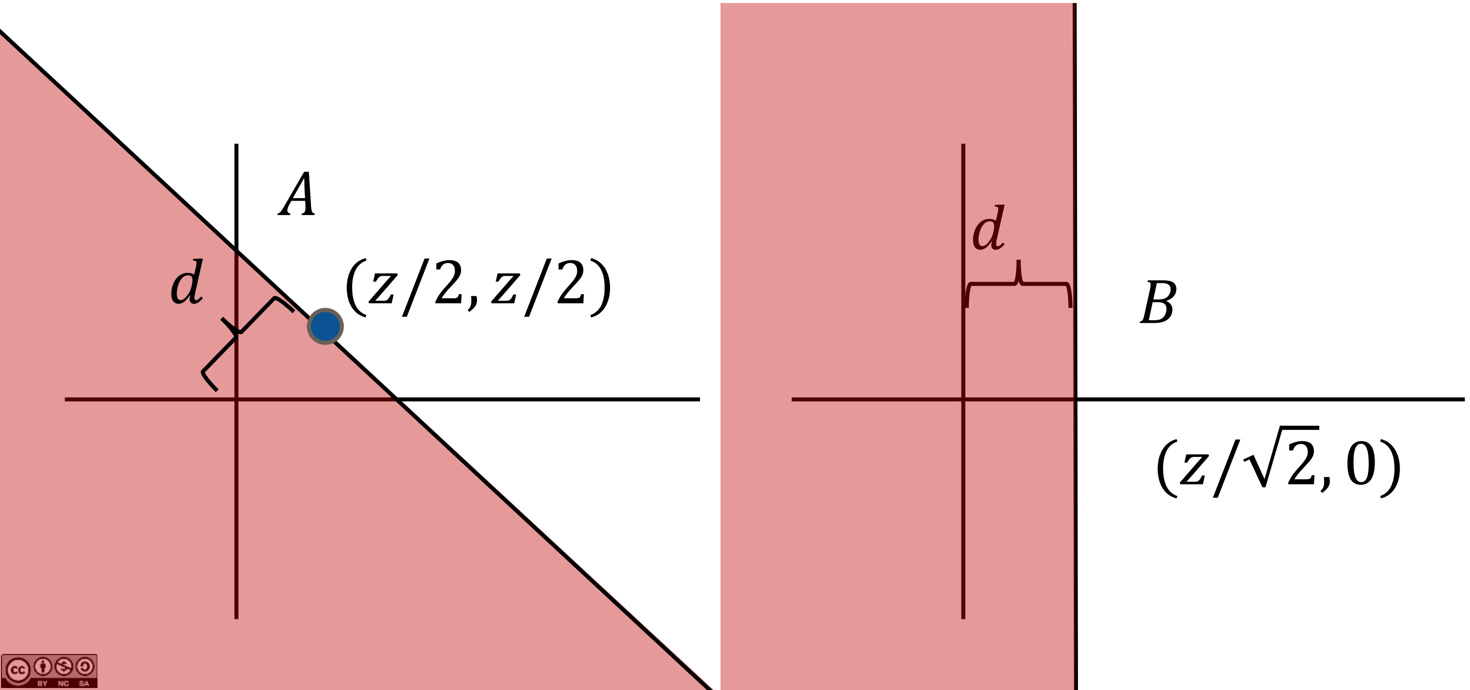




Rotational Symmetry of Joint Normals

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.

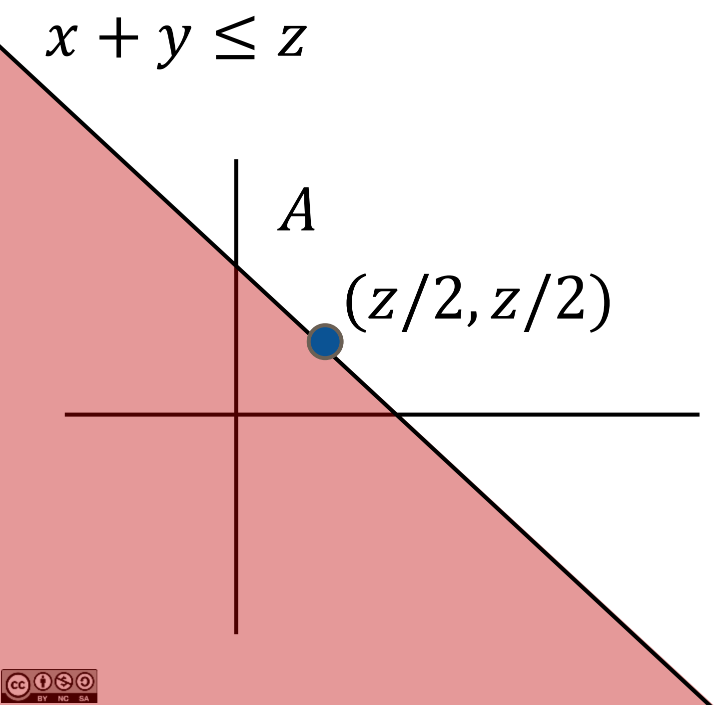


$$d = \sqrt{\frac{z^2}{4} + \frac{z^2}{4}} = \sqrt{\frac{z^2}{2}} = \frac{z}{\sqrt{2}}$$

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



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 $Z = X + Y \sim N(0,2)$

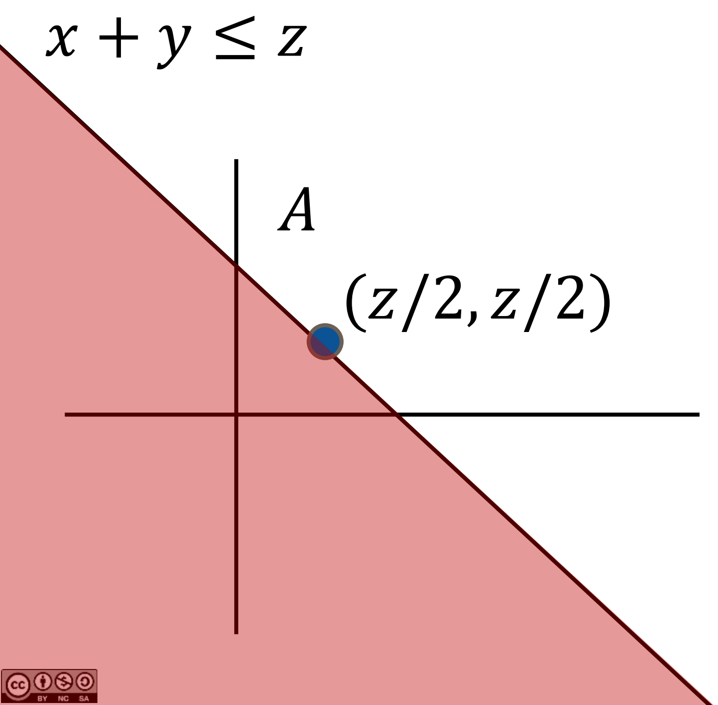
Let $Z = X + Y$, then $P(Z \leq z) = P(X + Y \leq z) = P((X,Y) \in A)$

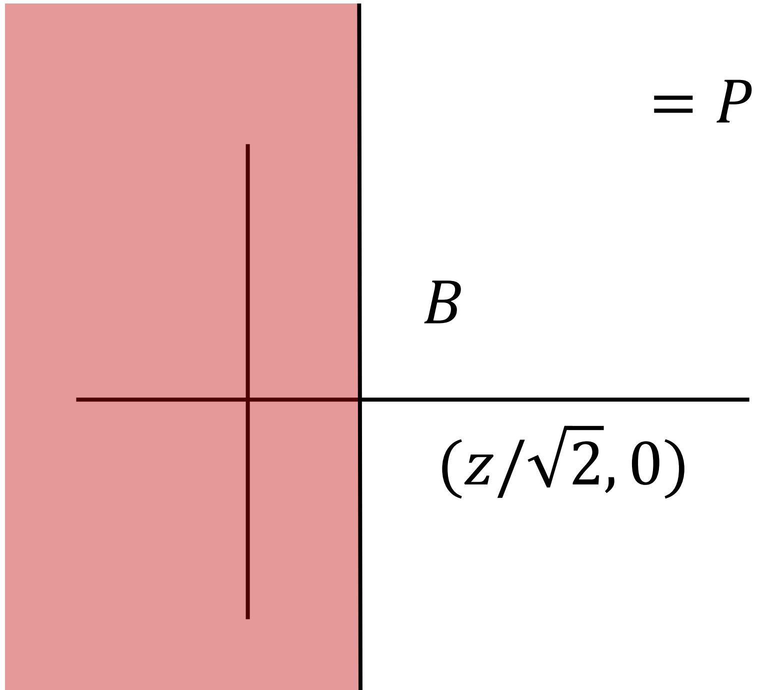
$$= P((X,Y) \in B)$$

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

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

$$Z \sim N(0,2)$$





Note: If this seems confusing, maybe consider a specific z, e.g., $P(Z \leq 3) = P(X \leq 3/\sqrt{2})$. Then consider how this also works for any choice of z.

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, $Z = aX + bY \sim N(0, a^2 + b^2)$

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

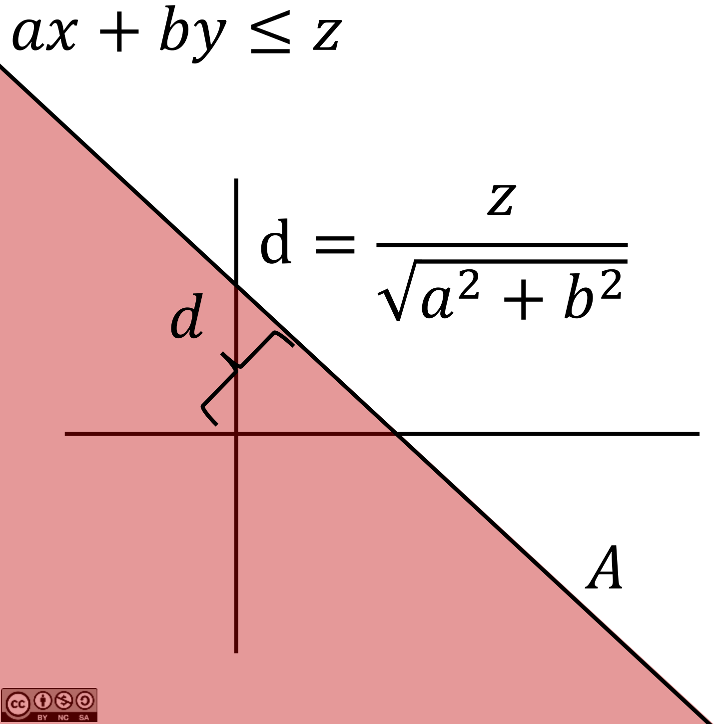
$$= P((X,Y) \in B)$$

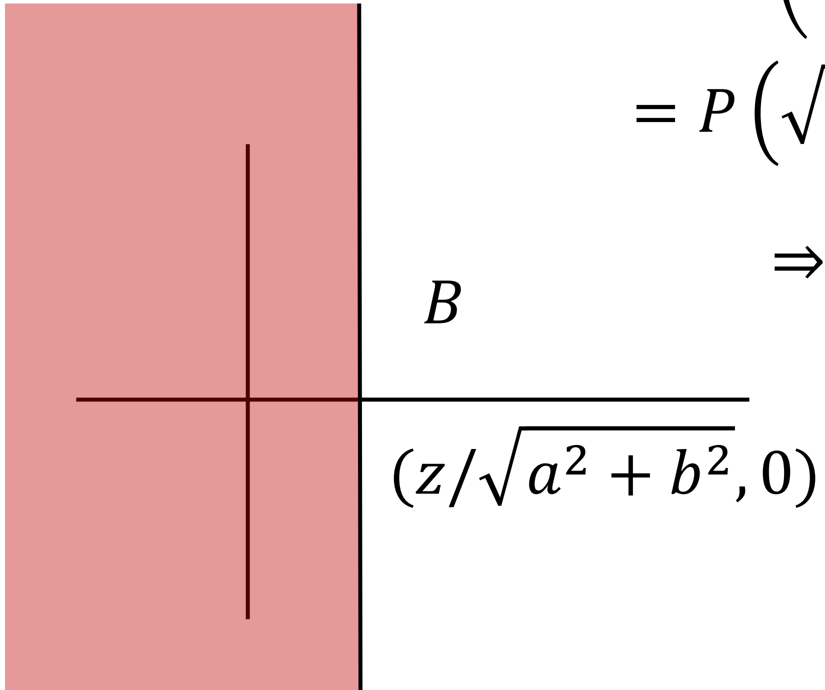
$$= P\left(X \leq \frac{z}{\sqrt{a^2+b^2}}\right)$$

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

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

$$Z \sim N(0, a^2 + b^2)$$





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, $\mathrm{Z} = aX + bY \sim N(0, a^2 + b^2)$

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).

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 $\tilde{X} = \dfrac{X_1 - \mu_1}{\sigma_1} = \dfrac{X_2 - \mu_2}{\sigma_2}$

We have that

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

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

Whew... We're Done!

So far we've shown that:

- A Gaussian is characterized by its mean and standard deviation alone.
- Adding two 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.

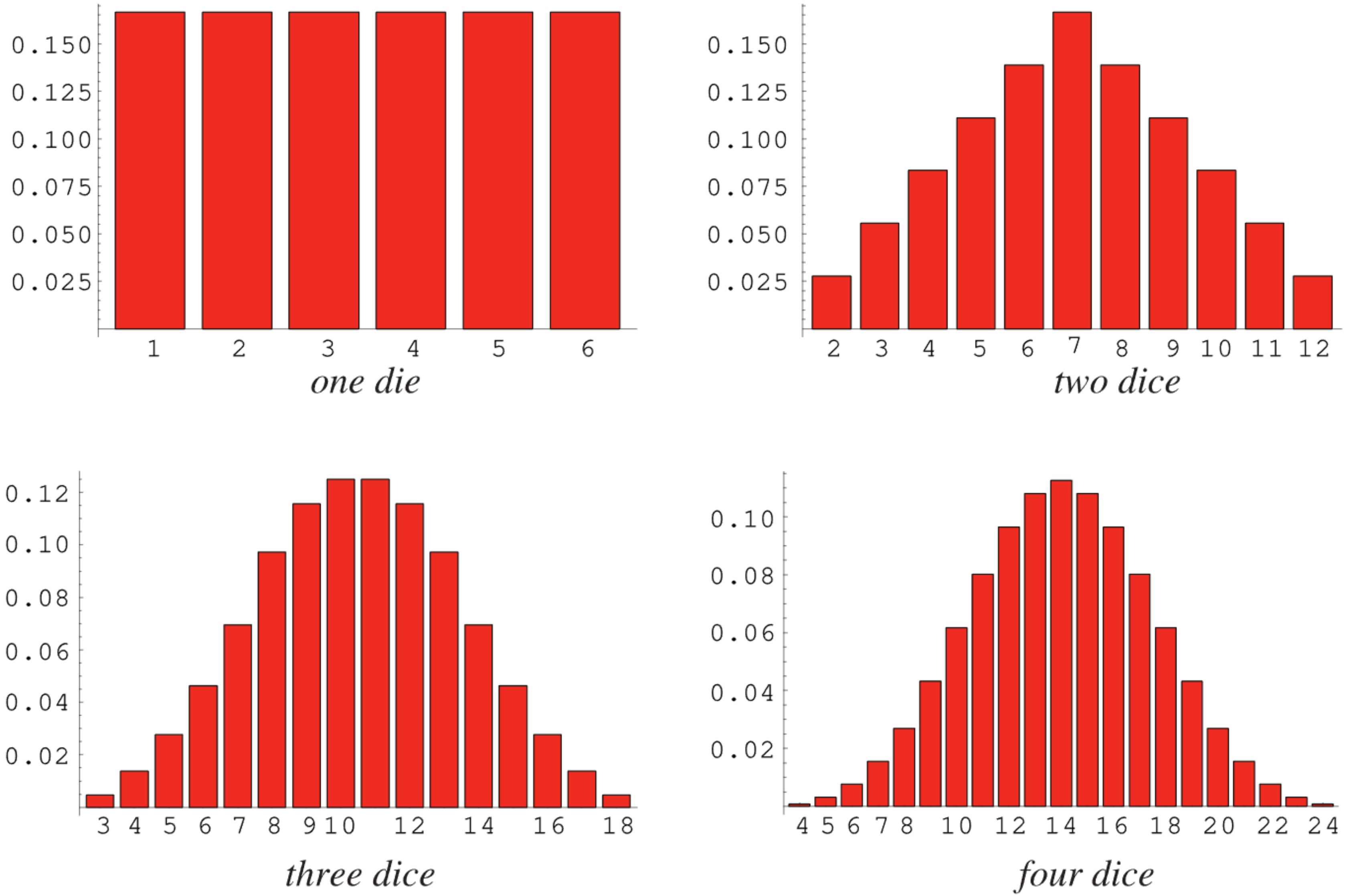
- Admittedly, not the most important thing today, but hopefully you enjoyed the novel proof technique we used to get there (proofs is the main point of 70!)

Next up, let's finally talk about some applications of Gaussian RVs, including experimental error (or even fraud) detection.

The Central Limit Theorem

Lecture 25, CS70 Summer 2025

Summing Dice

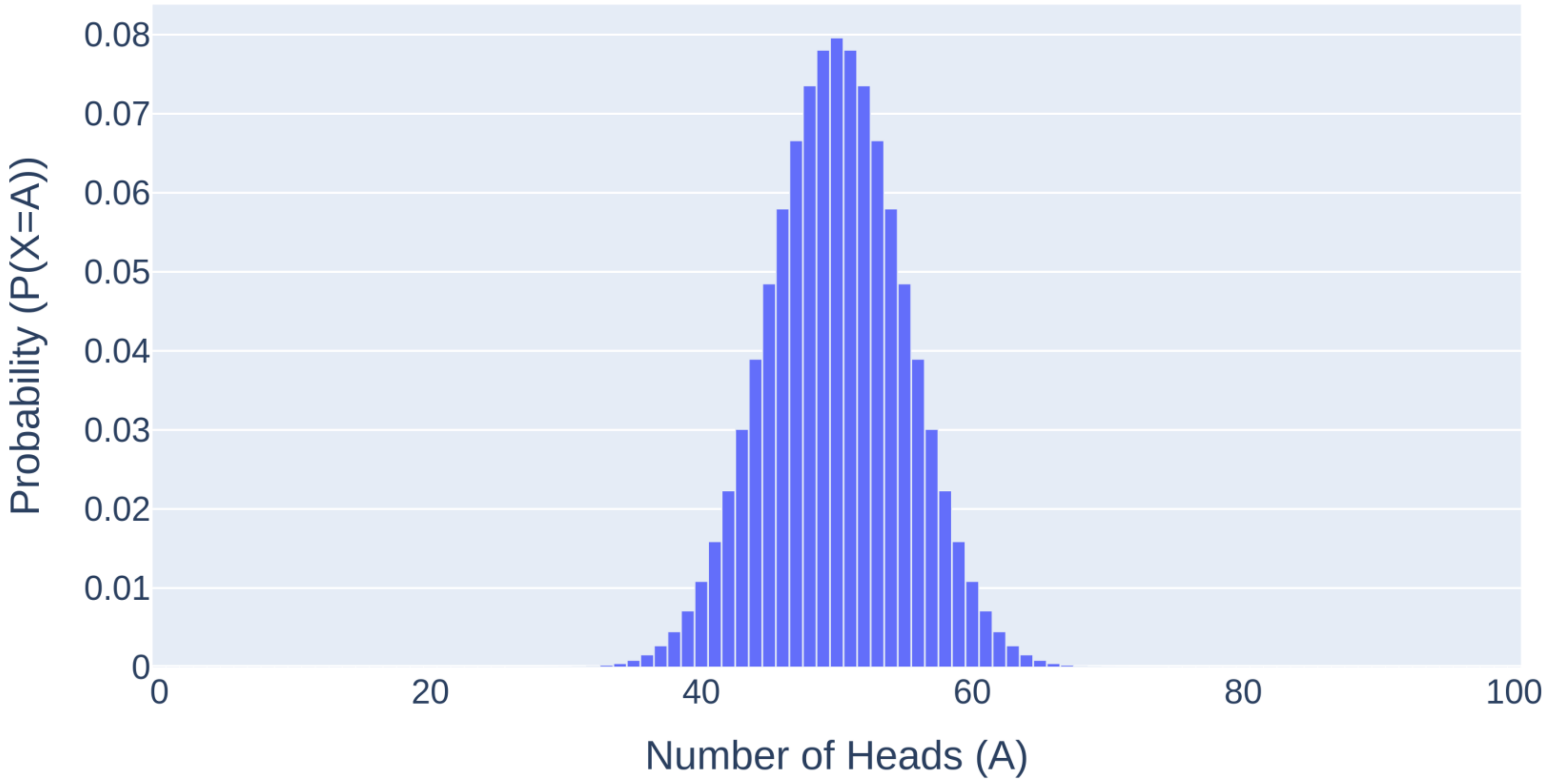


It’s not a coincidence that summing four dice looks kind of Gaussian!

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.



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.

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 $E[X_i] = \mu$, and $\mathrm{var}(X_i) = \sigma^2 < \infty$, 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.

The Central Limit Theorem

Our new idea is the The Central Limit Theorem.

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $E[X_i] = \mu < \infty$, and $\mathrm{var}(X_i) = \sigma^2 < \infty$, 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.

The Central Limit Theorem

Our new idea is the The Central Limit Theorem.

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $E[X_i] = \mu < \infty$, and $\mathrm{var}(X_i) = \sigma^2 < \infty$, 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:

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

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 $S = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. What is the Z-score for this result?

- Recall: $E[X_i] = 1$ and $\mathrm{var}(X_i) = 1$

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 $S = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. What is the Z-score for this result?

- Recall: $E[X_i] = 1$ and $\mathrm{var}(X_i) = 1$
- First, we compute $E[S] = 1000$ and $\mathrm{var}(S) = 1000$. This works since the RVs are uncorrelated and independent.
- 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!

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 $S = X_1 + X_2 + \cdots + X_{1000} = 1{,}231$. What is the Z-score for this result? 7.31.

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

In other words, there is about a 1 in 10 trillion chance that this was just a coincidence.

- This is known in experimental science as a "p-value".
- The "null hypothesis" that this nothing weird happened is correct with probability $p = 1.33 \times 10^{-13}$.