Normal Distribution

Chris Piech

CS109, Stanford University

蛇

新年快乐

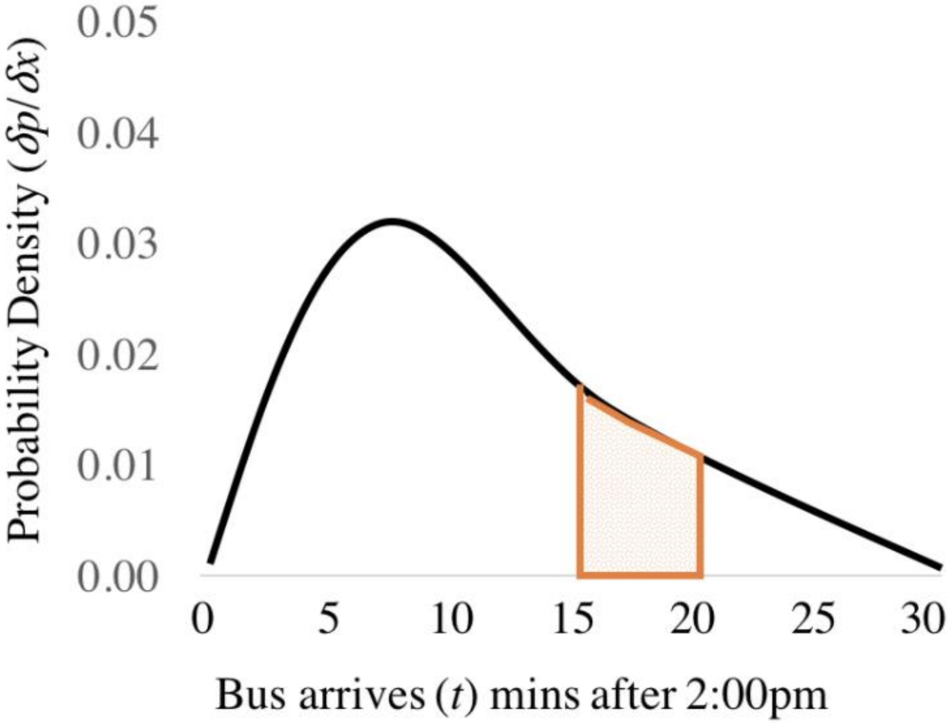
HAPPY CHINESE NEW YEAR

2025

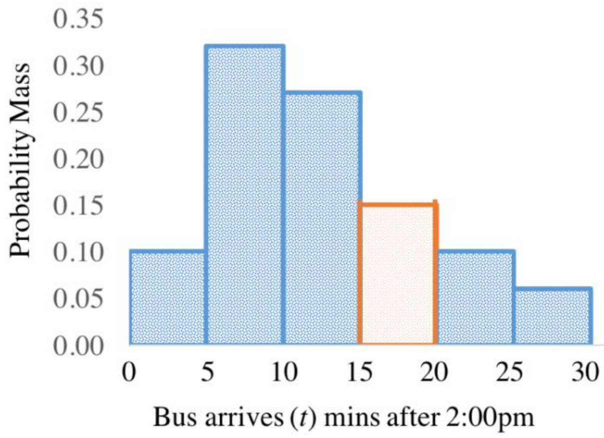
Review

Review: Probability Density Function

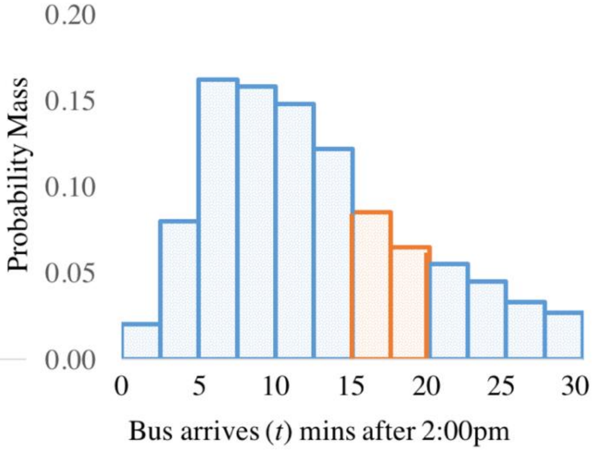
The limit at discretization size → 0



Discretize into 5 min chunks



Discretize into 2.5 min chunks



What do you get if you integrate over a probability *density* function?

A probability!

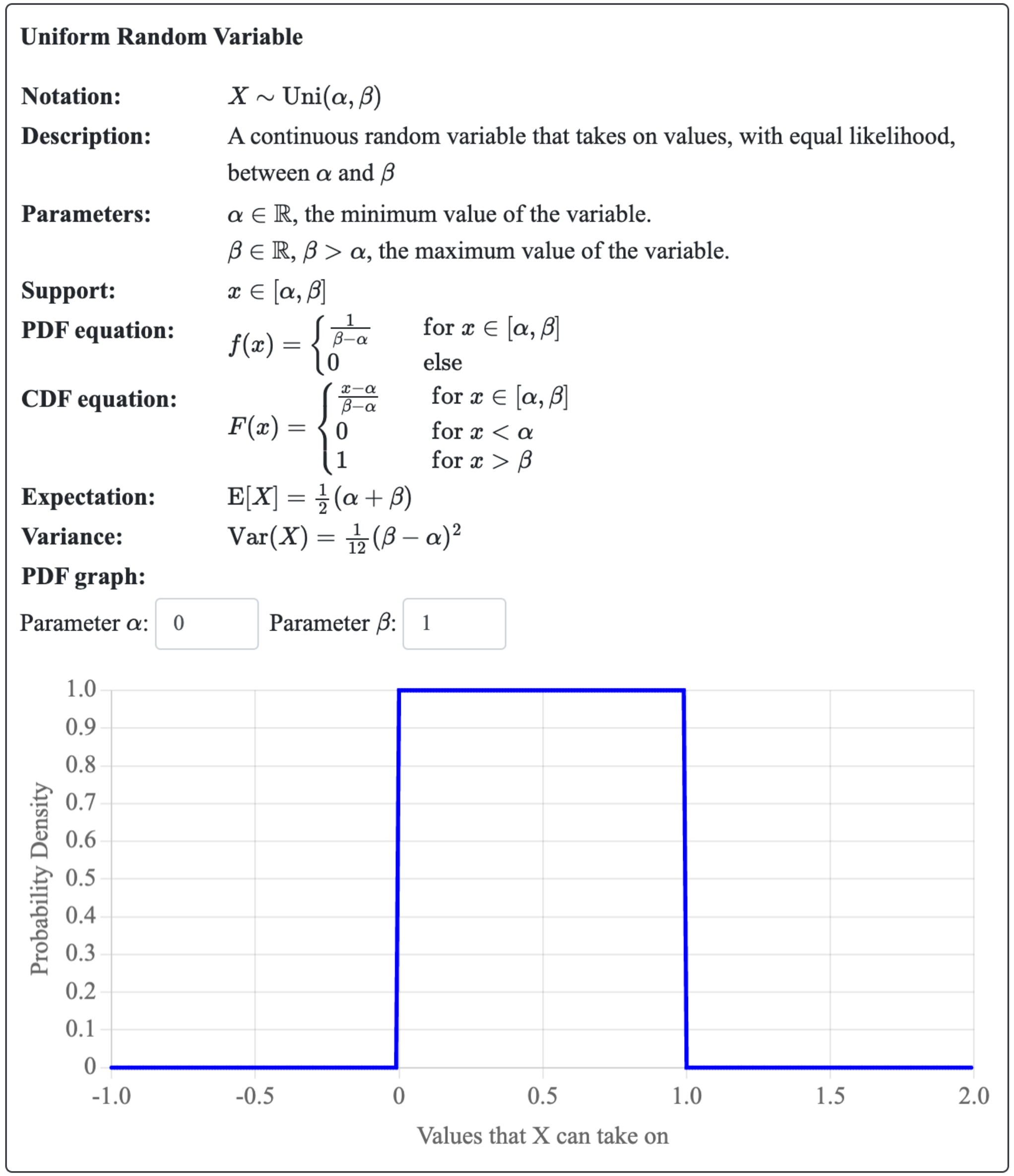
Review: Probability Density Function

The **probability density function** (PDF) of a continuous random variable represents the **derivative** of probability at a given point.

Units of probability *divided by units of X*. **Integrate it** to get probabilities!

$$P(a < X < b) = \int_{x=a}^{b} \boxed{f(X = x)}\, dx$$

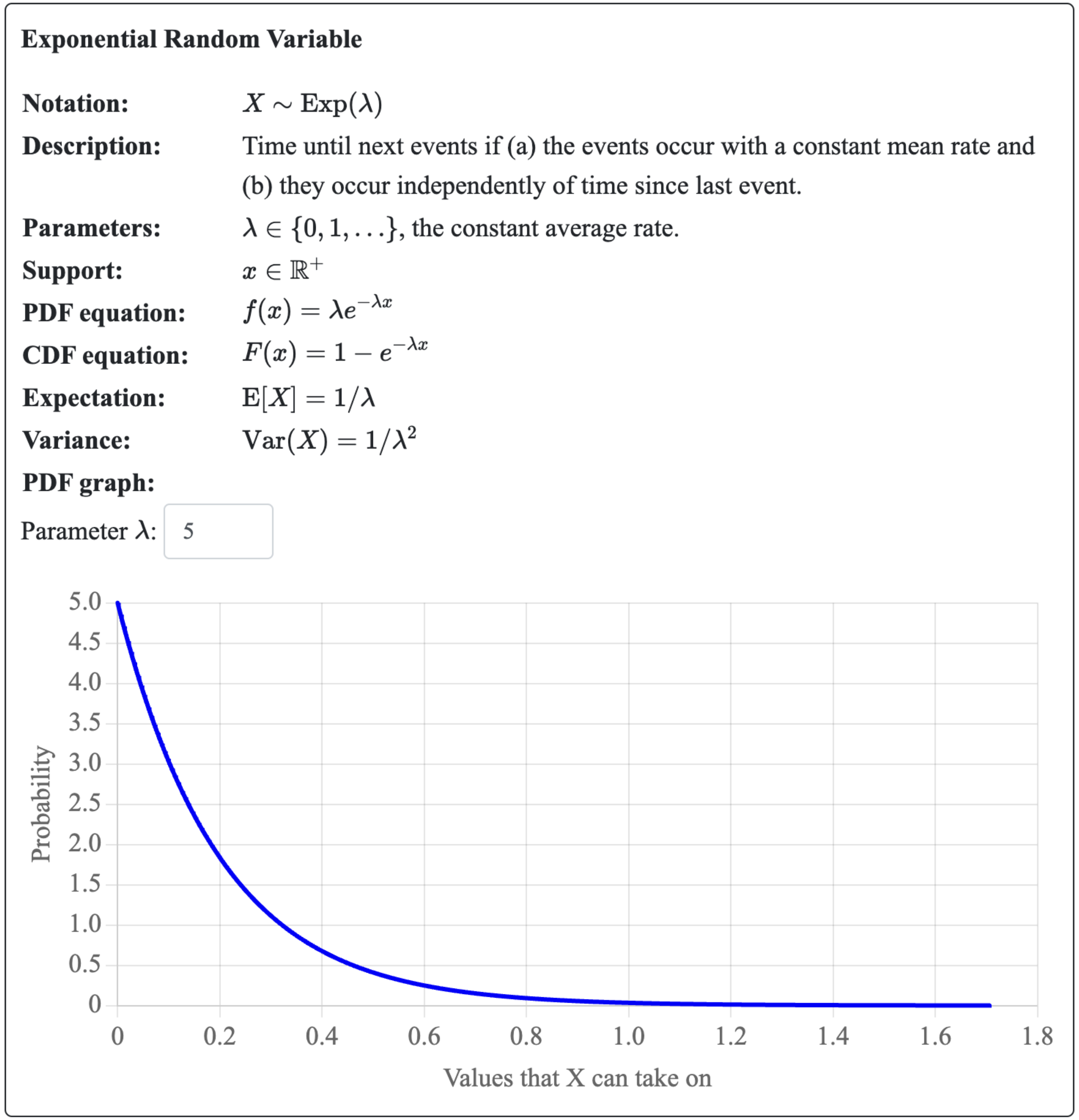
Uniform and Exponential Distributions



Uniform Random Variable

Notation:	$X \sim \text{Uni}(\alpha, \beta)$
Description:	A continuous random variable that takes on values, with equal likelihood, between α and β
Parameters:	$\alpha \in \mathbb{R}$, the minimum value of the variable. $\beta \in \mathbb{R}$, $\beta > \alpha$, the maximum value of the variable.
Support:	$x \in [\alpha, \beta]$
PDF equation:	$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{for } x \in [\alpha, \beta] \\ 0 & \text{else} \end{cases}$
CDF equation:	$F(x) = \begin{cases} \frac{x-\alpha}{\beta-\alpha} & \text{for } x \in [\alpha, \beta] \\ 0 & \text{for } x < \alpha \\ 1 & \text{for } x > \beta \end{cases}$
Expectation:	$\text{E}[X] = \frac{1}{2}(\alpha + \beta)$
Variance:	$\text{Var}(X) = \frac{1}{12}(\beta - \alpha)^2$
PDF graph:	

Parameter α: 0 Parameter β: 1



Exponential Random Variable

Notation:	$X \sim \text{Exp}(\lambda)$
Description:	Time until next events if (a) the events occur with a constant mean rate and (b) they occur independently of time since last event.
Parameters:	$\lambda \in \{0, 1, \dots\}$, the constant average rate.
Support:	$x \in \mathbb{R}^+$
PDF equation:	$f(x) = \lambda e^{-\lambda x}$
CDF equation:	$F(x) = 1 - e^{-\lambda x}$
Expectation:	$\text{E}[X] = 1/\lambda$
Variance:	$\text{Var}(X) = 1/\lambda^2$
PDF graph:	

Parameter λ: 5

Cumulative Density Function

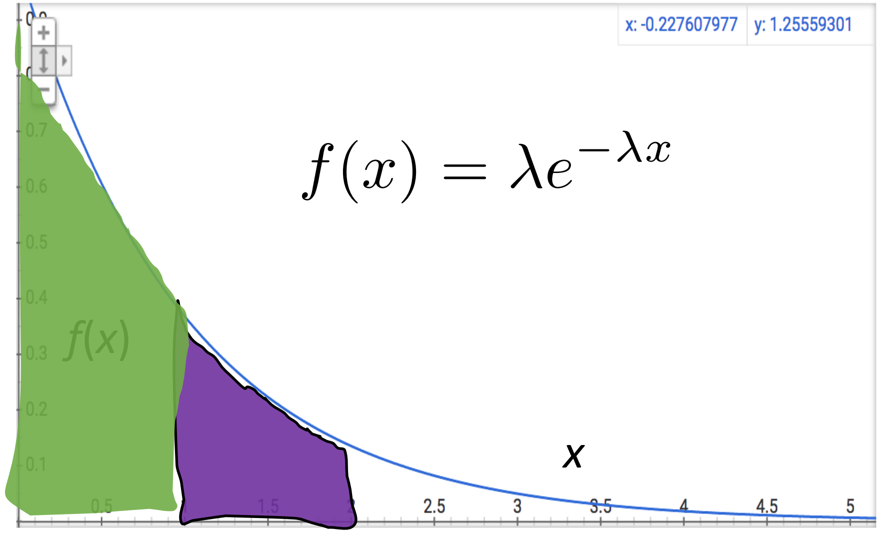
A cumulative density function (CDF) is a “closed form” equation for the probability that a random variable is less than a given value

$$F(x) = P(X < x)$$

If you learn how to use a cumulative density function, you can avoid integrals!

Using CDF Example. X is Exp($\lambda = 1$)

Probability
density
function

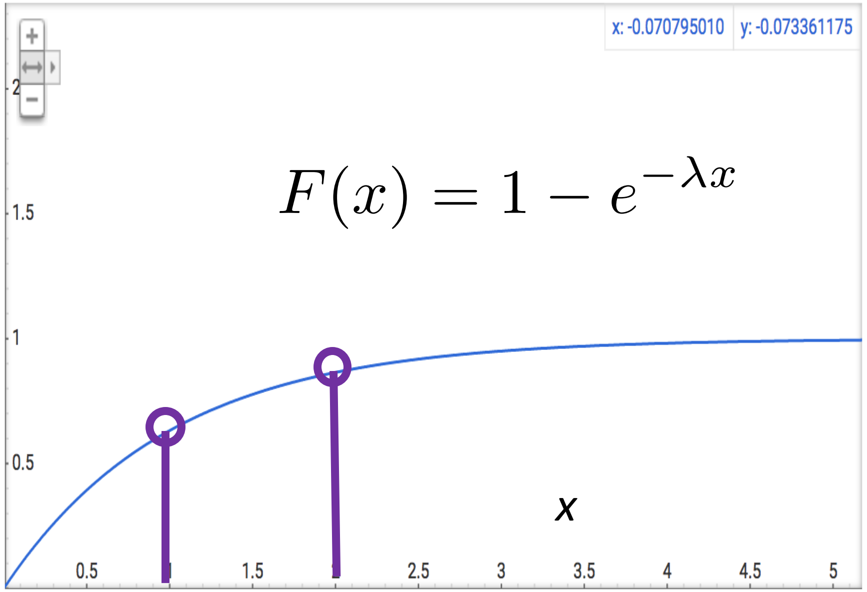


$$f(x) = \lambda e^{-\lambda x}$$

Cumulative
density function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\ dy$$



$$F(x) = 1 - e^{-\lambda x}$$

P(1 < X < 2)

$$= \int_{x=1}^{2} f(x)\ dx$$

or

$$= F(2) - F(1)$$

$$=(1 - e^{-2})$$

$$- (1 - e^{-1})$$

$$\approx 0.23$$

Did you know? Exponential is Memoryless!

$$F(x) = 1 - e^{-\lambda x}$$

$$X \sim \text{Exp}(\lambda)$$

X = time until the next event

$$\text{P}(X > s + t | X > s) = \text{P}(X > t)$$

What if s time has passed?

Did you know? Exponential is Memoryless!

$F(x) = 1 - e^{-\lambda x}$

$$X \sim \mathrm{Exp}(\lambda)$$

X = time until the next event

$$\mathrm{P}(X > s + t | X > s) = \mathrm{P}(X > t)$$

What if s time has passed?

Which is something we can prove:

$$\begin{aligned} \mathrm{P}(X > s + t | X > s) &= \frac{\mathrm{P}(X > s + t \text{ and } X > s)}{\mathrm{P}(X > s)} && \text{Def of conditional prob.} \\ &= \frac{\mathrm{P}(X > s + t)}{\mathrm{P}(X > s)} && \text{Because } X > s + t \text{ implies } X > s \end{aligned}$$

Did you know? Exponential is Memoryless!

$F(x) = 1 - e^{-\lambda x}$

$$X \sim \text{Exp}(\lambda)$$

X = time until the next event

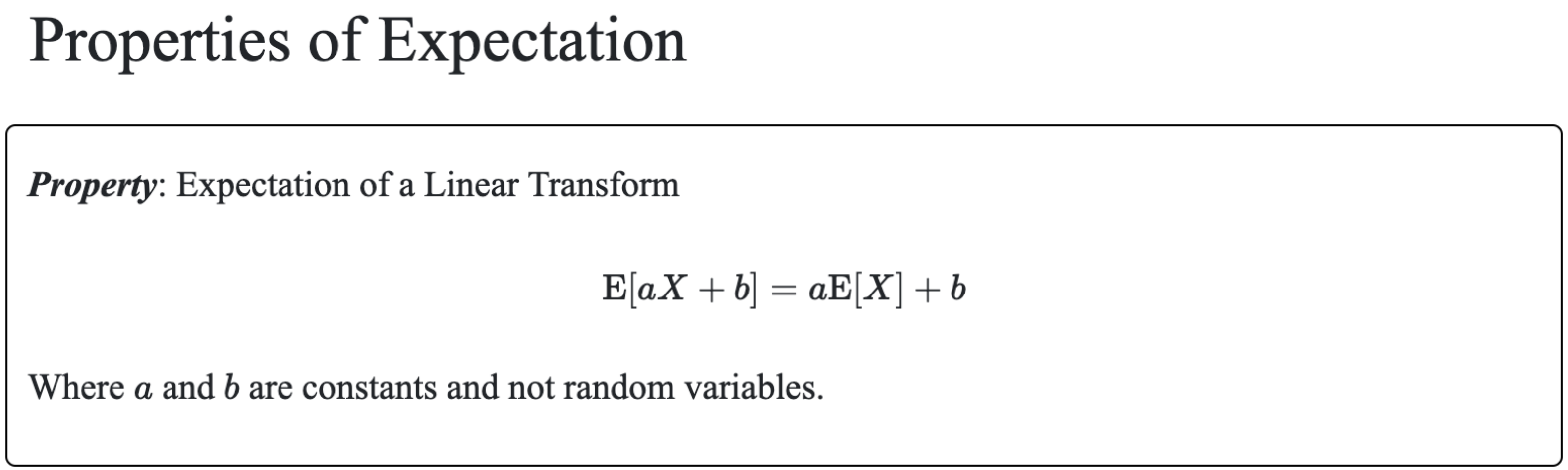
$$\text{P}(X > s + t | X > s) = \text{P}(X > t)$$

What if s time has passed?

Which is something we can prove:

$$\begin{aligned}
\text{P}(X > s + t | X > s) &= \frac{\text{P}(X > s + t \text{ and } X > s)}{\text{P}(X > s)} && \text{Def of conditional prob.} \\
&= \frac{\text{P}(X > s + t)}{\text{P}(X > s)} && \text{Because } X > s + t \text{ implies } X > s \\
&= \frac{1 - F_X(s + t)}{1 - F_X(s)} && \text{Def of CDF} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} && \text{By CDF of Exp} \\
&= e^{-\lambda t} && \text{Simplify} \\
&= 1 - F_X(t) && \text{By CDF of Exp} \\
&= \text{P}(X > t) && \text{Def of CDF}
\end{aligned}$$

I am going to use these two properties later in class today

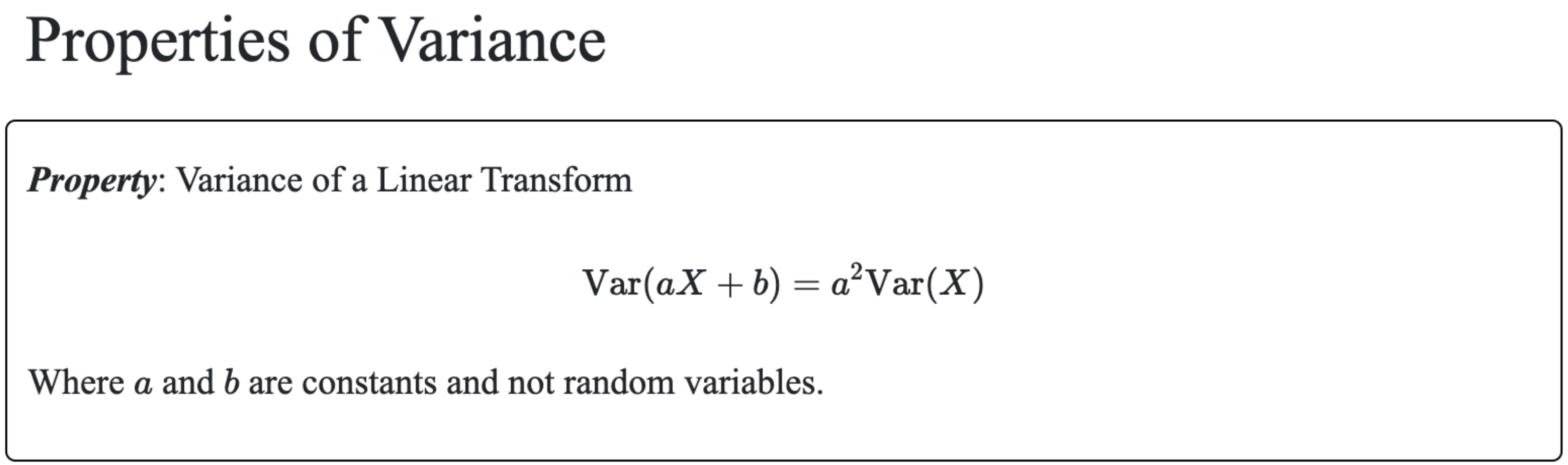


Properties of Expectation

Property: Expectation of a Linear Transform

$$\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$$

Where a and b are constants and not random variables.



Properties of Variance

Property: Variance of a Linear Transform

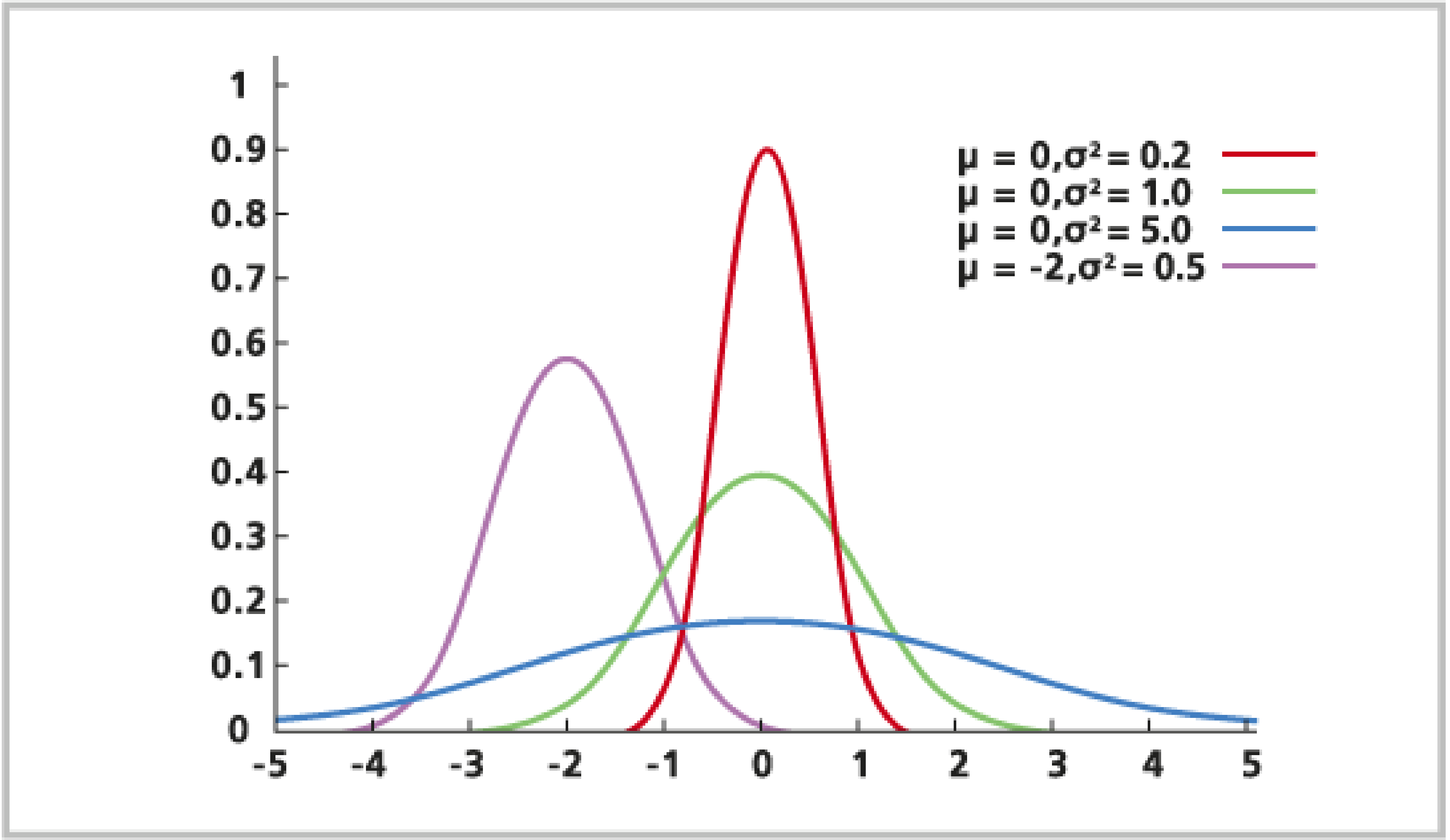
$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Where a and b are constants and not random variables.

/Review

Big Day

NormCore: A Few Normal Examples



Normal Random Variable

<u>def</u> An **Normal** random variable X is defined as follows:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Support: $(-\infty, \infty)$

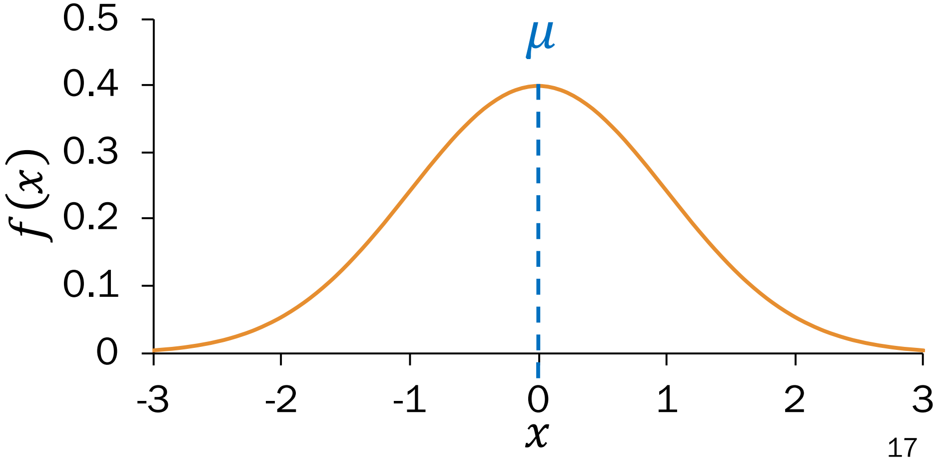
PDF	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
Expectation	$E[X] = \mu$
Variance	$\text{Var}(X) = \sigma^2$

Other names: **Gaussian** random variable

mean

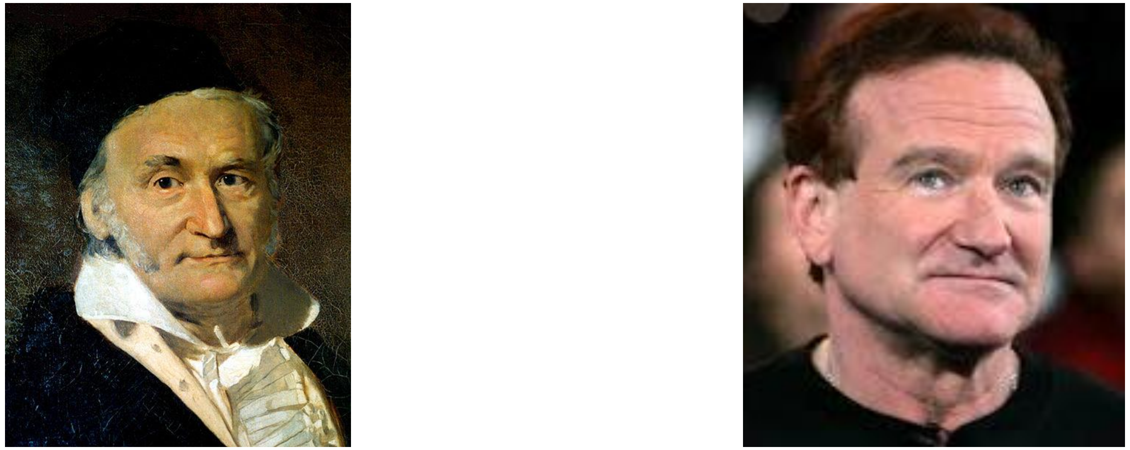
variance

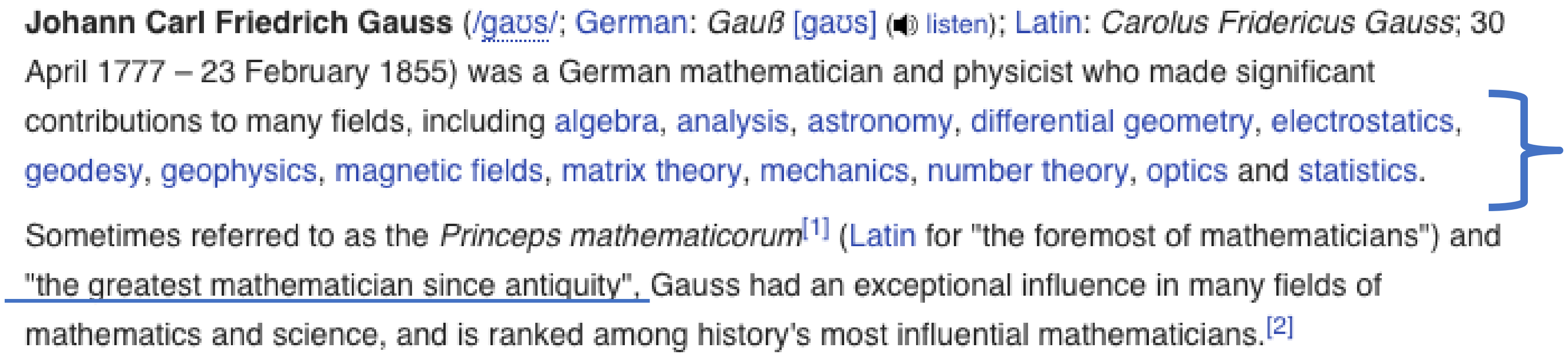
$$X \sim \mathcal{N}(\mu, \sigma^2)$$



Carl Friedrich Gauss

Carl Friedrich Gauss (1777-1855) was a remarkably influential German mathematician.





Johann Carl Friedrich Gauss (/gaʊs/; German: *Gauß* [gaʊs] (listen); Latin: *Carolus Fridericus Gauss*; 30 April 1777 – 23 February 1855) was a German mathematician and physicist who made significant contributions to many fields, including algebra, analysis, astronomy, differential geometry, electrostatics, geodesy, geophysics, magnetic fields, matrix theory, mechanics, number theory, optics and statistics.

Sometimes referred to as the *Princeps mathematicorum*[1] (Latin for "the foremost of mathematicians") and "the greatest mathematician since antiquity", Gauss had an exceptional influence in many fields of mathematics and science, and is ranked among history's most influential mathematicians.[2]

Did not invent Normal distribution but rather popularized it

Why the Normal?

These are log-normal

- Common for natural phenomena: height, weight, etc.

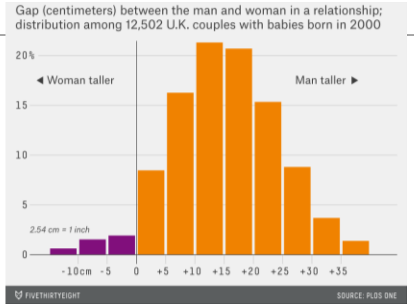
Most noise is assumed normal

- Most noise in the world is Normal
- Often results from the sum of many random variables

Only if they are equally weighted and independent

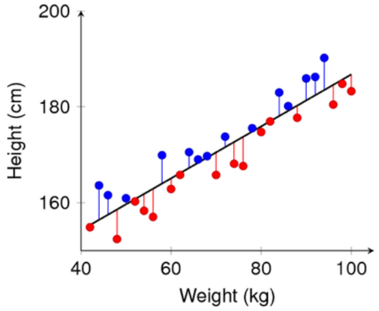
- Sample means are distributed normally

That is actually true…



Gap (centimeters) between the man and woman in a relationship; distribution among 12,502 U.K. couples with babies born in 2000

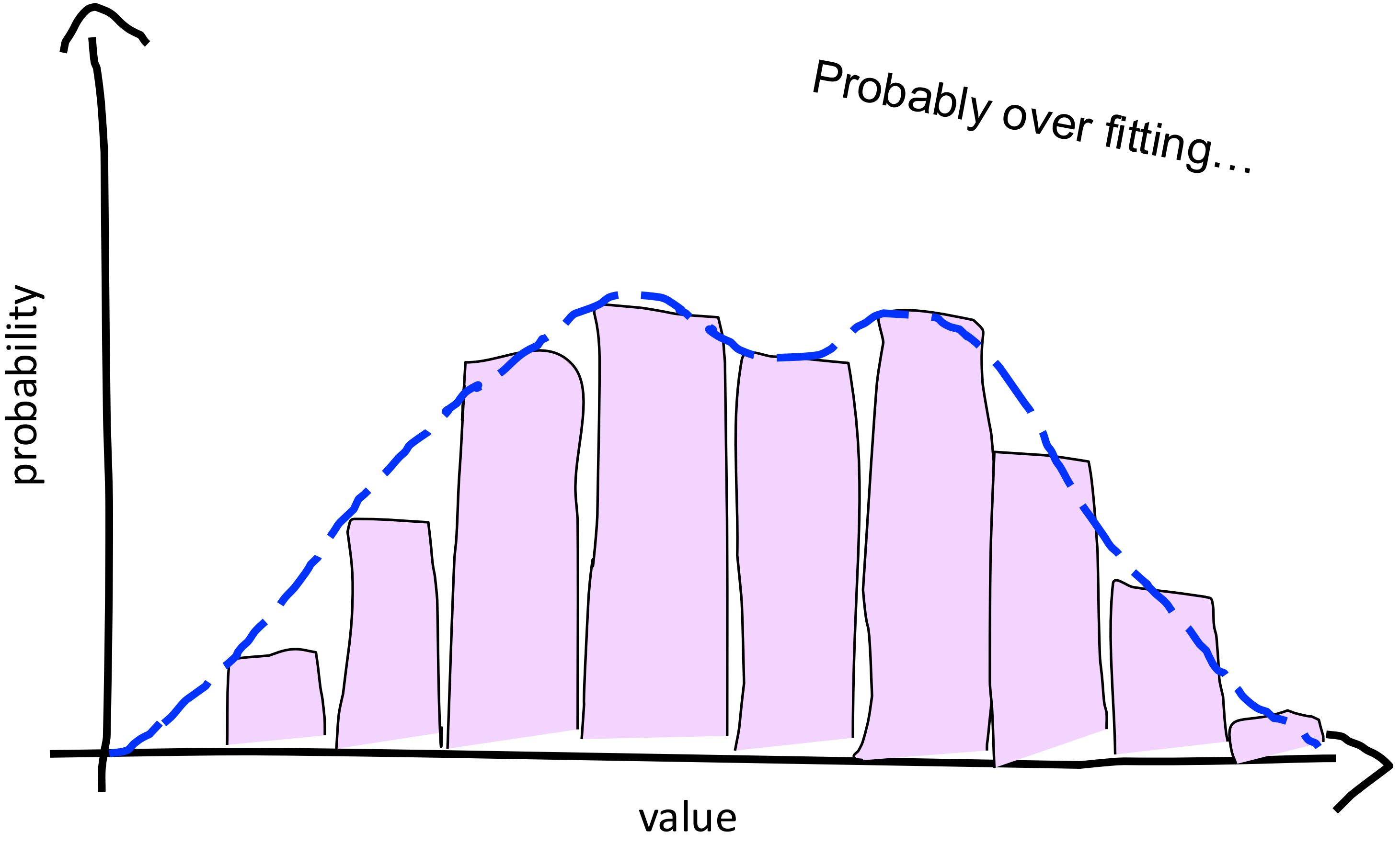
SOURCE: PLOS ONE



That's what they want you to believe…

“The simplest explanation is usually the best one”

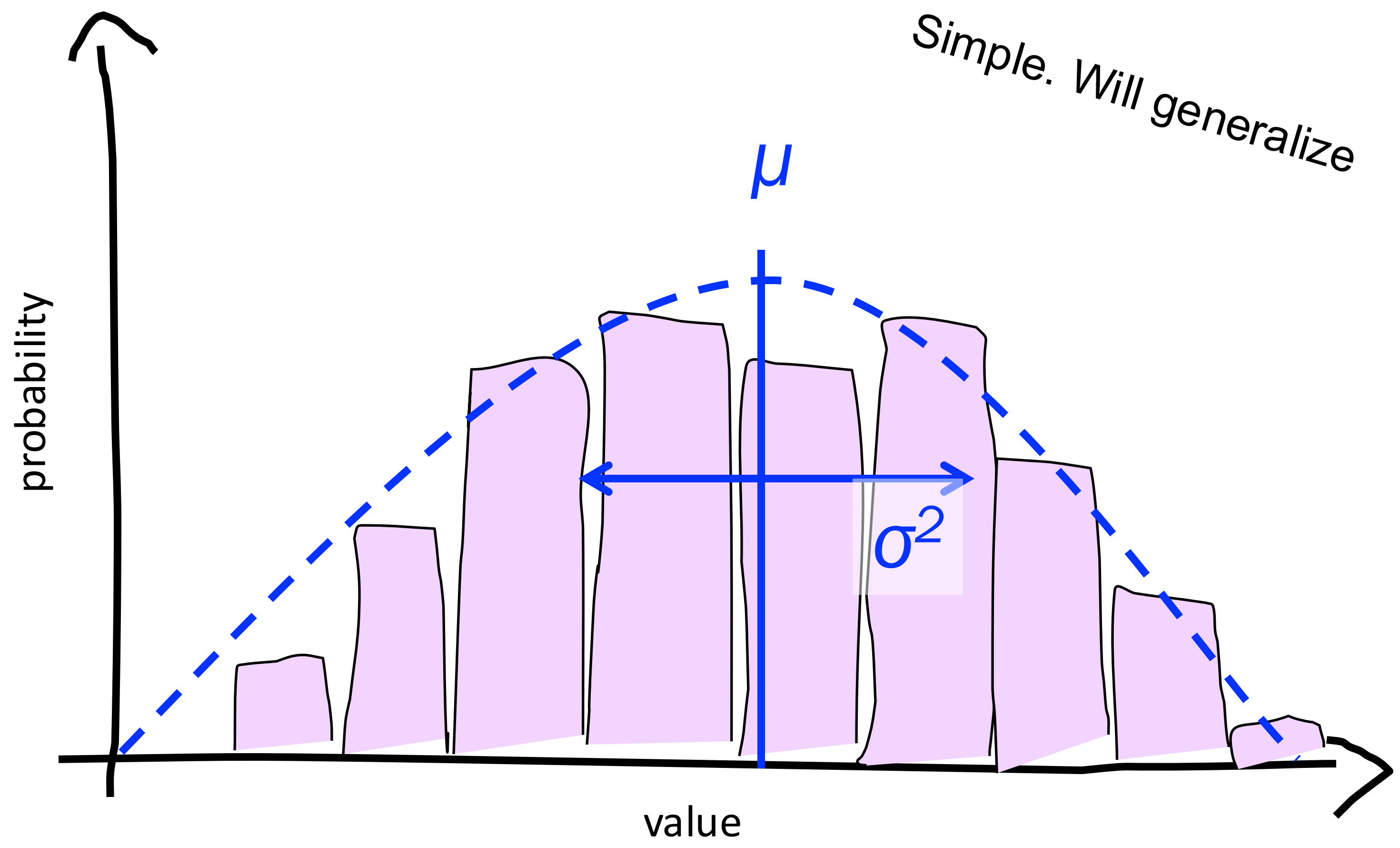
Complexity is Tempting



* That describes the training data, but will it **generalize**?

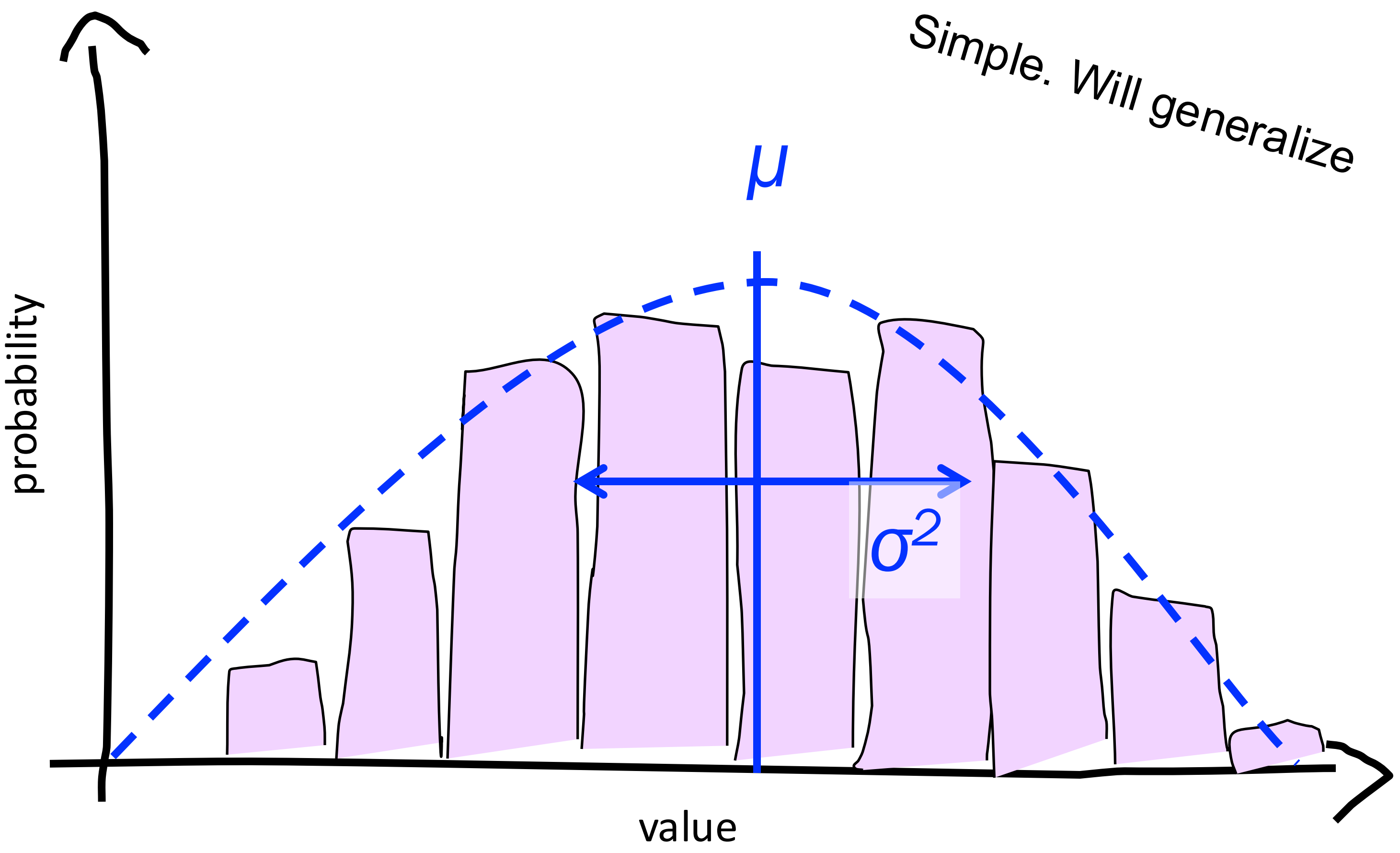
Fewest Assumptions

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$$



* A Gaussian **maximizes entropy** for a given mean and variance

Fewest Assumptions



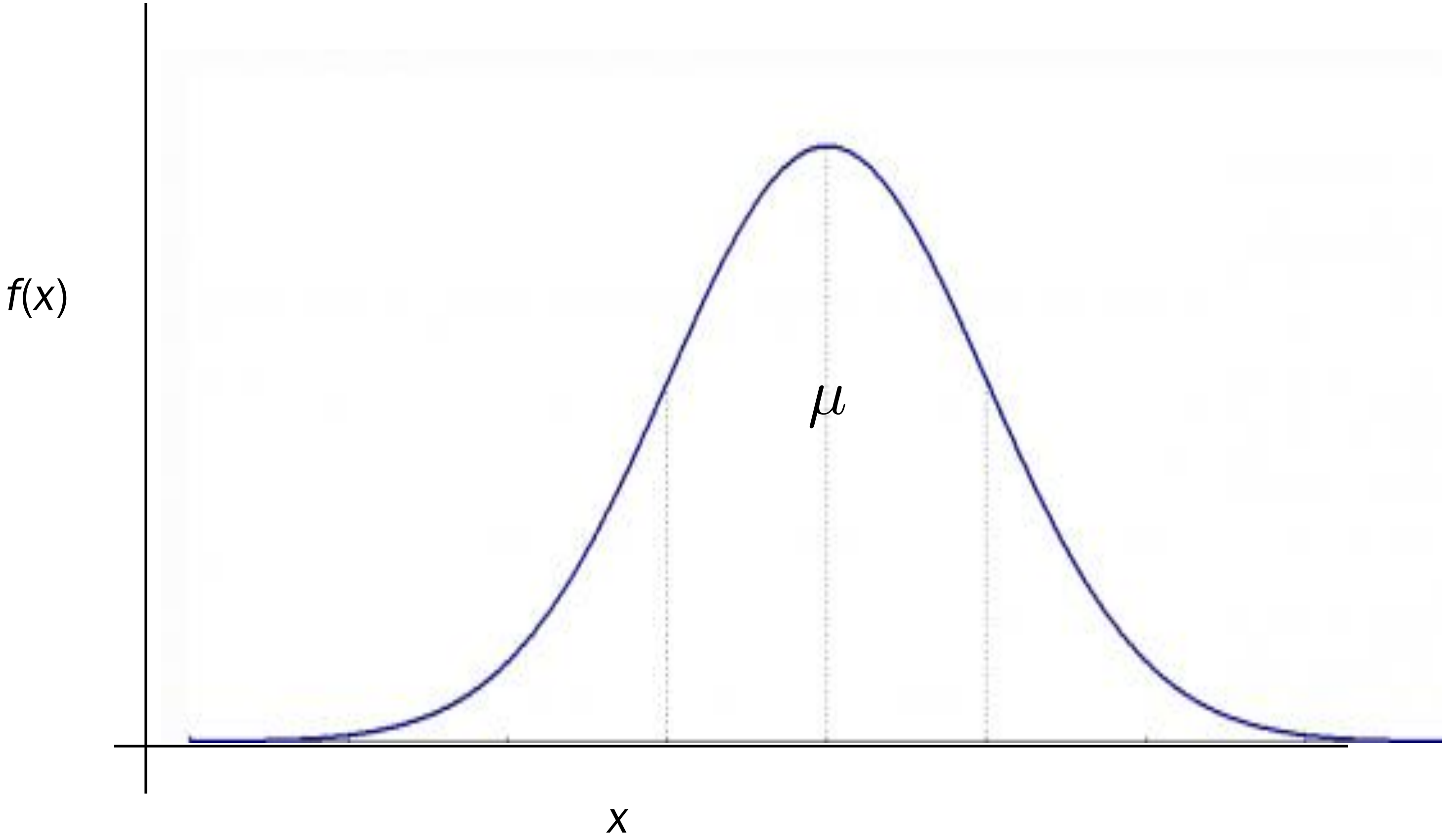
* A Gaussian makes the **fewest assumptions** after matching mean and variance.

Normal is Beautiful!

Normal Probability Density Function

$$\mathcal{N}(\mu, \sigma^2)$$

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



Anatomy of a Beautiful Equation

$$\mathcal{N}(\mu, \sigma^2)$$

the distance to the mean

“exponential”

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

sigma shows up twice

probability density
at *x*

a constant

Does it look less scary like this?

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

This means "e to the power of" and is common function in code math libraries

$$f(x) \propto \frac{1}{\sigma} \cdot \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

This means "proportional to". There is a constant but there are many cases where we don't care what it is!

What if you had to take the log of this function?

Lets go!

Let's Try It Out: Submarine Manufacturing

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet standards?

Let's Try It Out: Submarine Manufacturing

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

Let's Try It Out: Submarine Manufacturing

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = \int_{490}^{510} f(X = x)dx = \int_{490}^{510} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let’s Try It Out: Submarine Manufacturing

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = \int_{490}^{510} f(X = x)dx = \int_{490}^{510} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Campus bikes

$$P(490 \leq X \leq 510) = \int_{490}^{510} f(X = x)dx = \int_{490}^{510} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Loving, not scary

...except this time

No closed form for the integral

No closed form for F(x)

Spoiler: Numerically Solved CDF

$$\mathcal{N}(\mu, \sigma^2)$$

A function that has been solved for numerically

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

The cumulative density function of any normal

* We are going to spend the next few slides getting here

Linear Transform of a Normal is… Normal!

$$\text{Let } X \sim \mathcal{N}(\mu, \sigma^2)$$

$Y = aX + b$ is also Normal

$$\begin{aligned} E[Y] &= E[aX + b] \\ &= aE[X] + b \\ &= a\mu + b \end{aligned}$$

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(aX + b) \\ &= a^2\text{Var}(X) \\ &= a^2\sigma^2 \end{aligned}$$

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Aside: Celsius to Fahrenheit

$Y = aX + b \qquad Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Average temp in Palo Alto (on Jan 29th) in Celsius:

$$X \sim N(\mu = 13, \sigma^2 = 1)$$

What is the distribution in Fahrenheit?

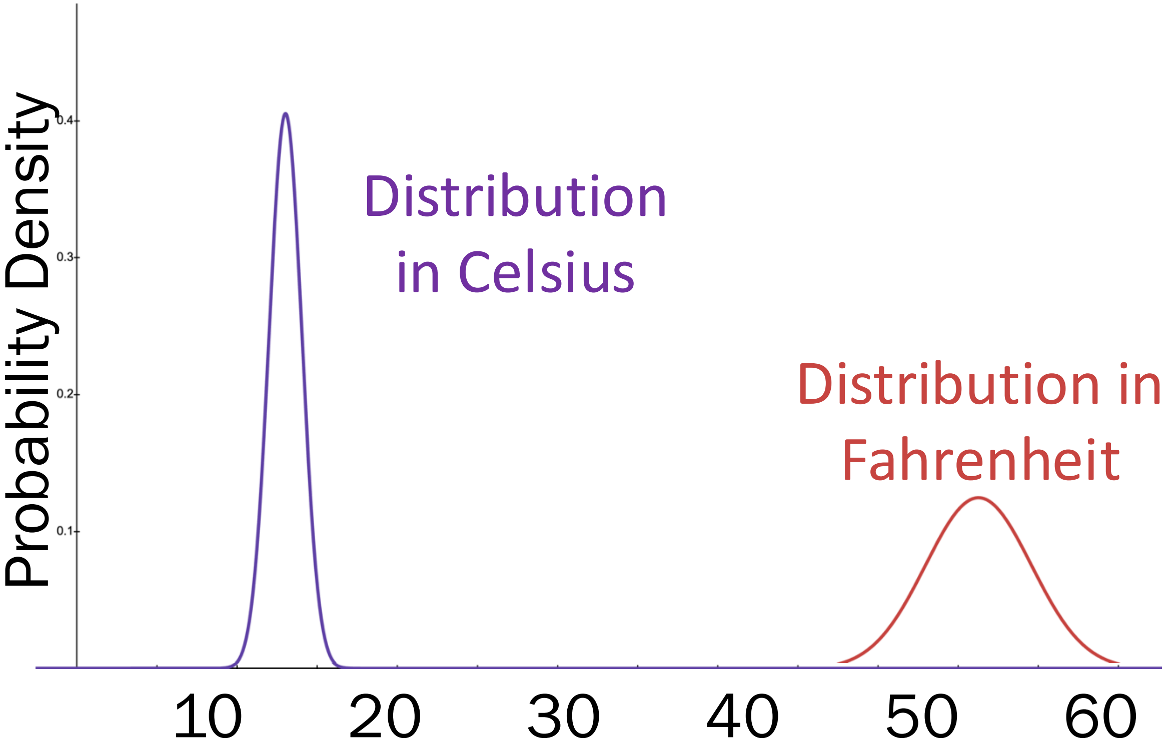
Let $Y = 1.8X + 32$

be the temperature in Fahrenheit.

Because this is a linear transform…

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

$$X \sim N(\mu = 55.4, \sigma^2 = 3.24)$$



Linear Transform of a Normal is… Normal!

Let $X \sim \mathcal{N}(\mu, \sigma^2)$

$Y = aX + b$ is also Normal

$$\begin{aligned} E[Y] &= E[aX + b] \\ &= aE[X] + b \\ &= a\mu + b \end{aligned}$$

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(aX + b) \\ &= a^2\text{Var}(X) \\ &= a^2\sigma^2 \end{aligned}$$

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

The cutest linear transform

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ $\quad Y = aX + b$ is also Normal

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

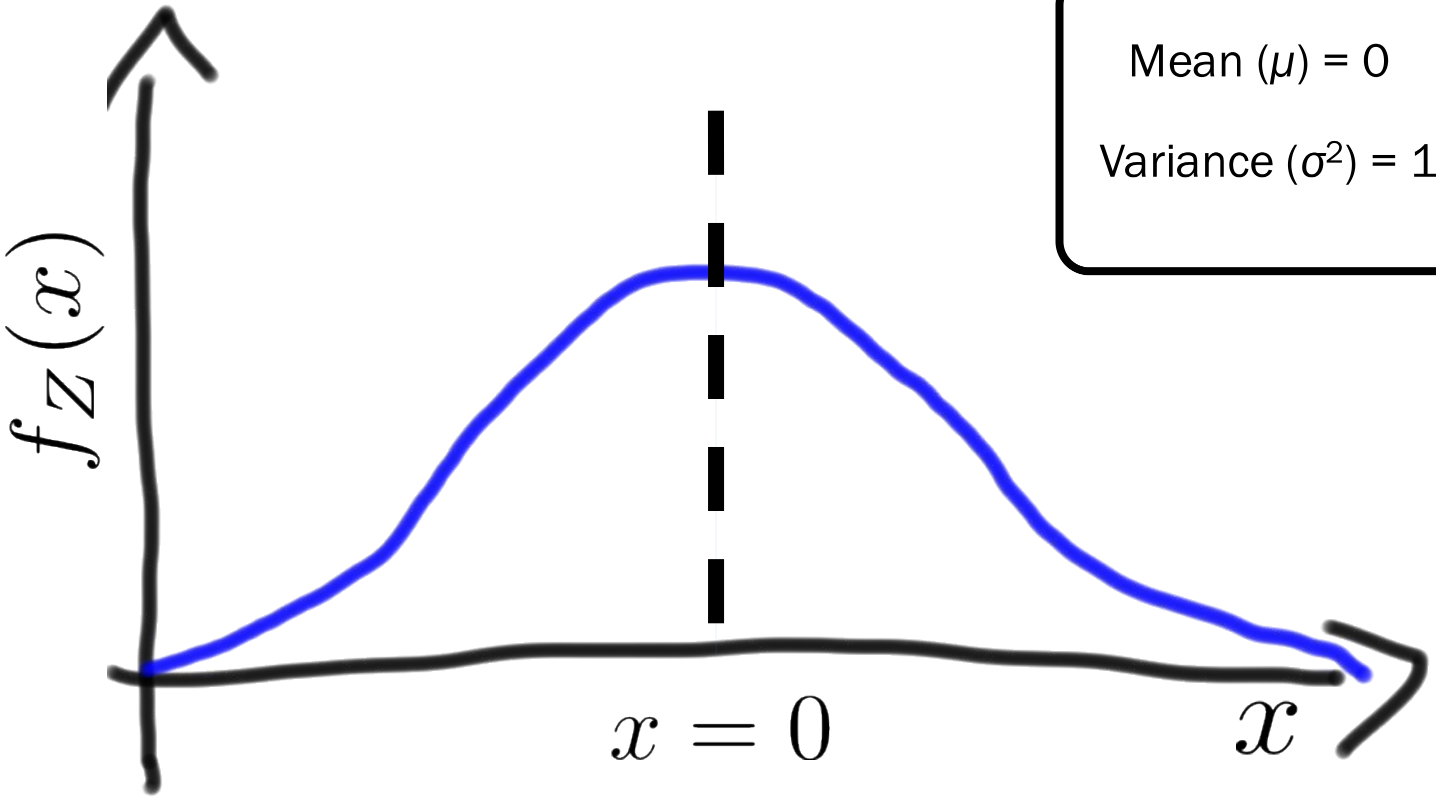
There is a special case of linear transform for any *X*:

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \qquad a = \frac{1}{\sigma} \quad b = -\frac{\mu}{\sigma}$$

$$\begin{aligned} Z &\sim \mathcal{N}(a\mu + b, a^2\sigma^2) \\ &\sim \mathcal{N}(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}) \\ &\sim \mathcal{N}(0, 1) \end{aligned}$$

The Standard Normal

$$Z \sim N(\mu = 0, \sigma^2 = 1)$$



Mean (μ) = 0

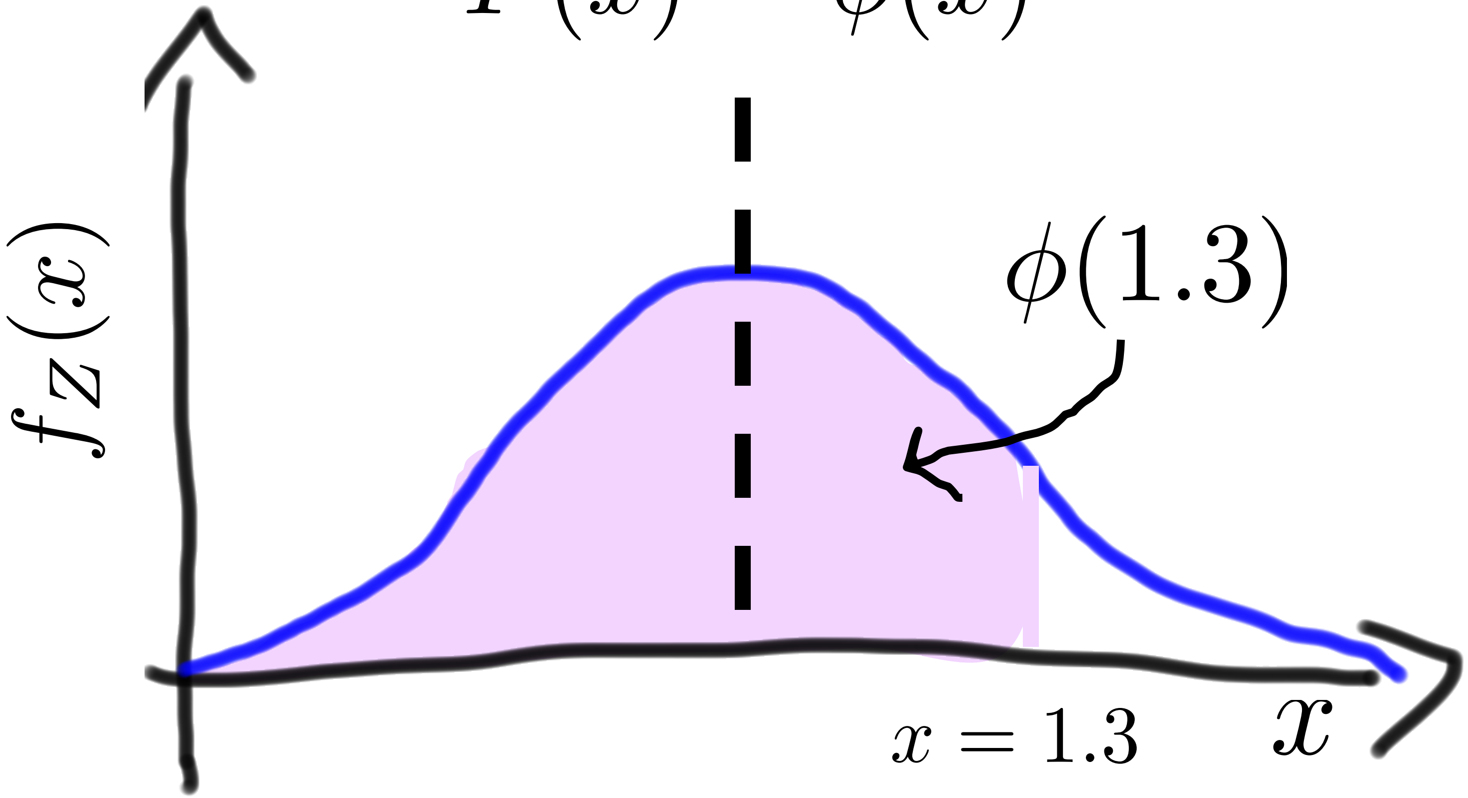
Variance (σ^2) = 1

*This is the probability density function for the standard normal

Phi

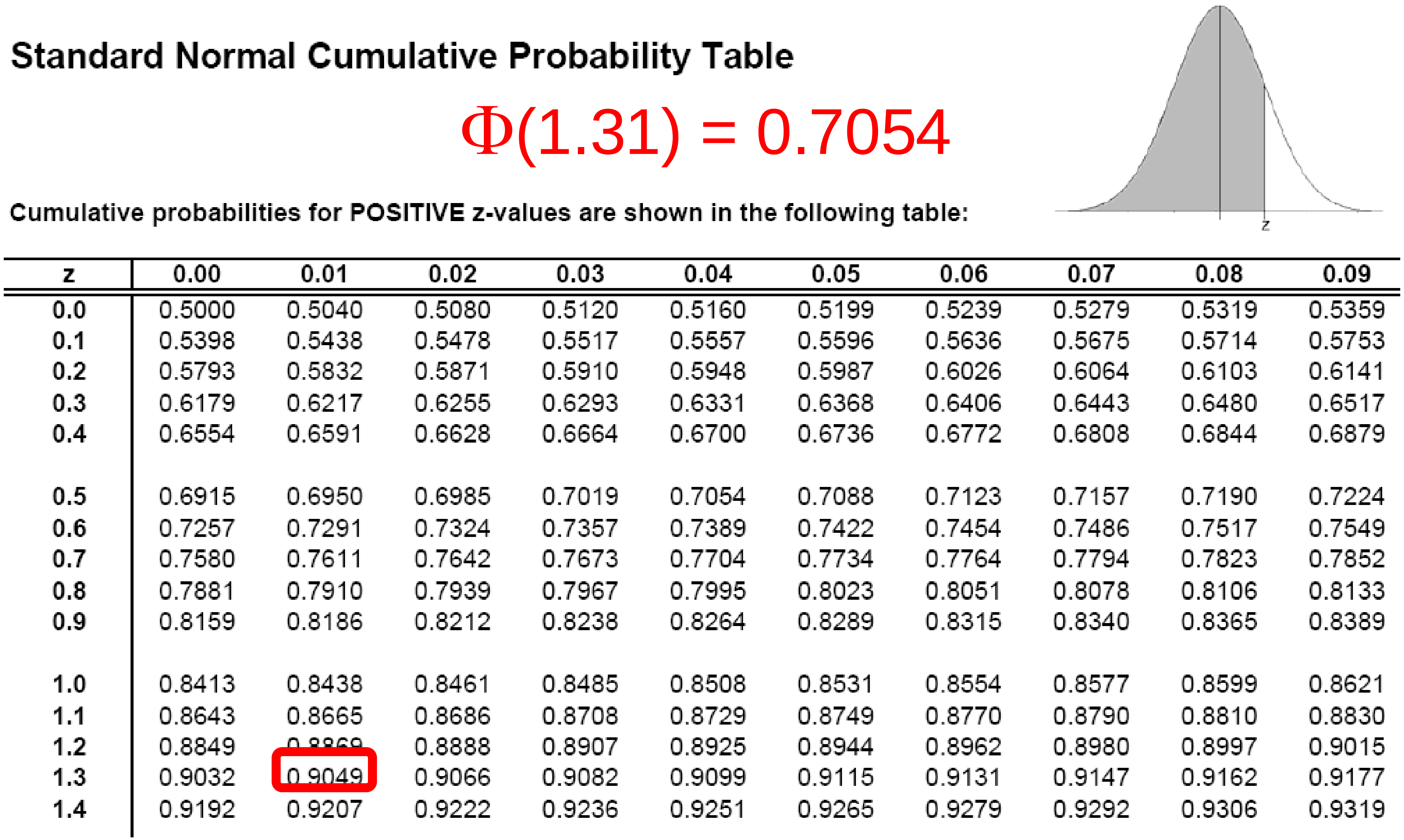
$$Z \sim N(\mu = 0, \sigma^2 = 1)$$

$$F(x) = \phi(x)$$



*This is the probability density function for the standard normal

Using Table of ϕ



Standard Normal Cumulative Probability Table

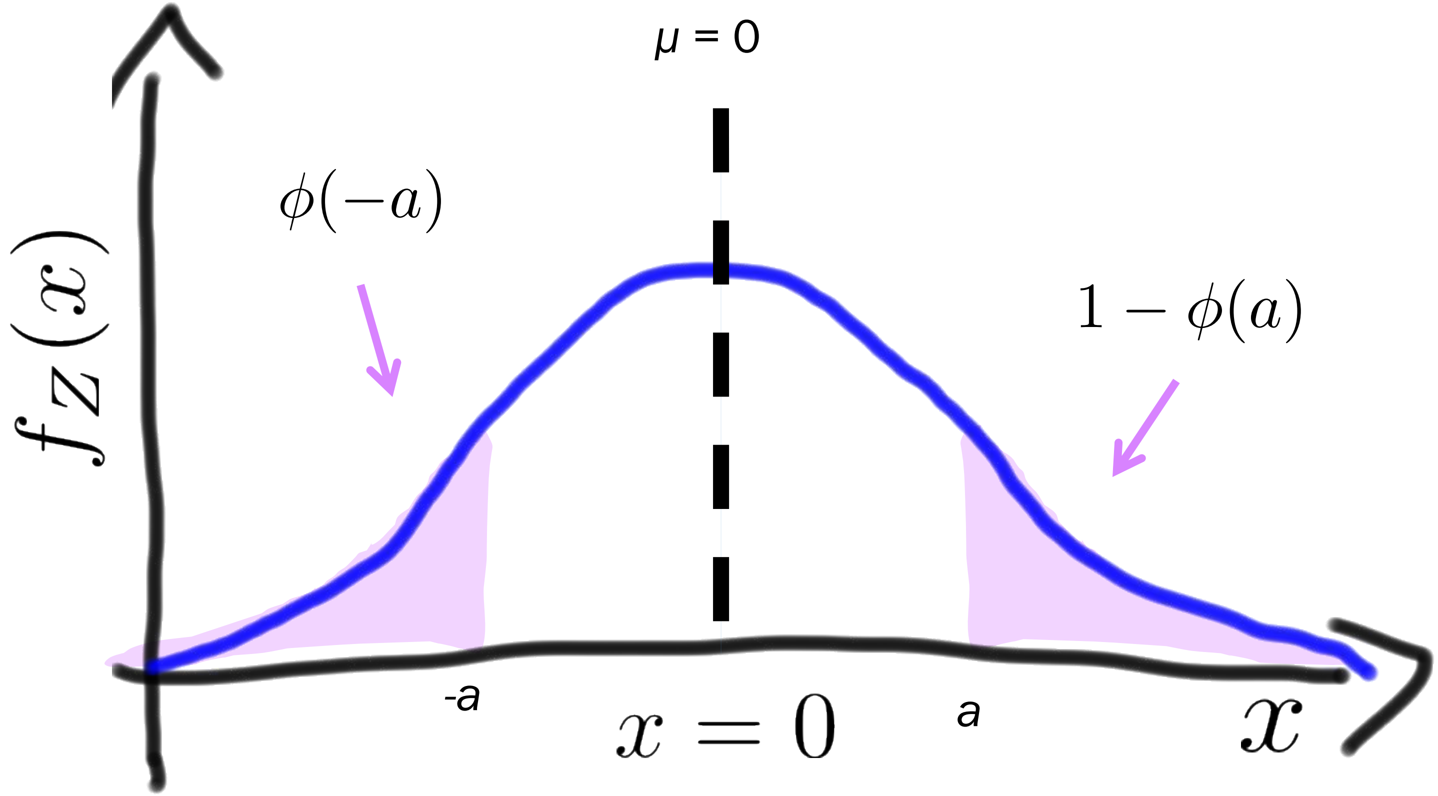
$\Phi(1.31) = 0.7054$

Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Symmetry of Phi

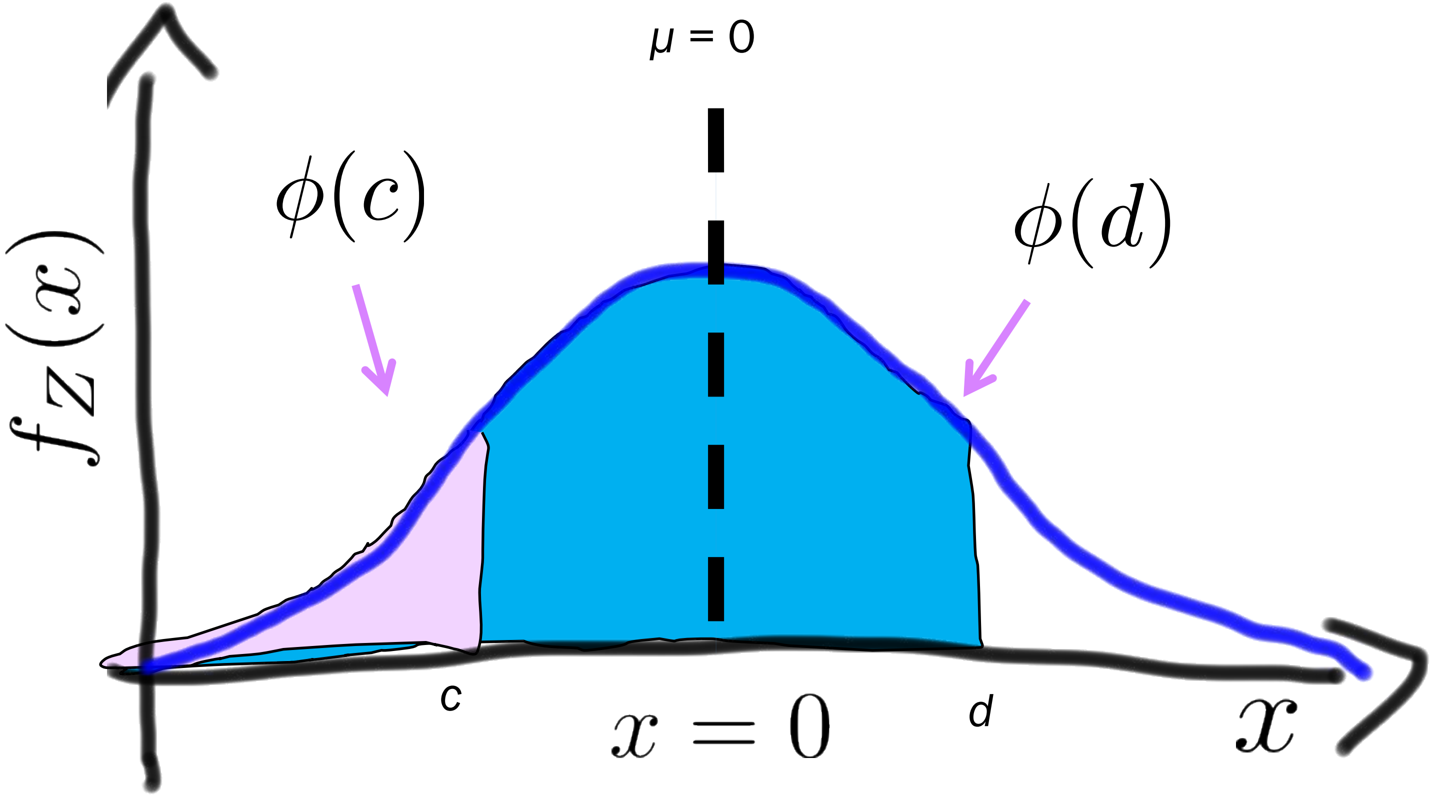
$$\phi(a) = 1 - \phi(a)$$



*This is the probability density function for the standard normal

Interval of Phi

$$P(c < Z < d) = \phi(d) - \phi(c)$$



Compute F(x) via Transform

$$\text{Let } X \sim \mathcal{N}(\mu, \sigma^2) \qquad Z = \frac{X - \mu}{\sigma}$$

Use *Z* to compute *F*(*x*)

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(X - \mu \leq x - \mu) \\
&= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{x - \mu}{\sigma}\right) \\
&\boxed{= \Phi\left(\frac{x - \mu}{\sigma}\right)}
\end{aligned}$$

For normal distribution, F(x) is computed using the phi transform.

And here we are

$$\mathcal{N}(\mu, \sigma^2)$$

CDF of Standard Normal: A function that has been solved for numerically

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

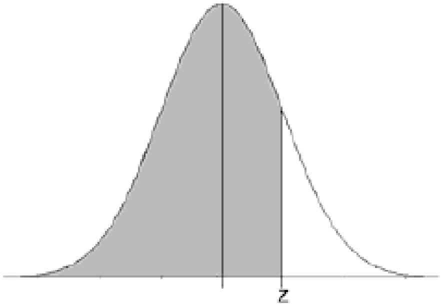
The cumulative density function (CDF) of any normal

Table of $\Phi(Z)$ values in textbook, p. 201 and handout

Using the Phi Table

Standard Normal Cumulative Probability Table

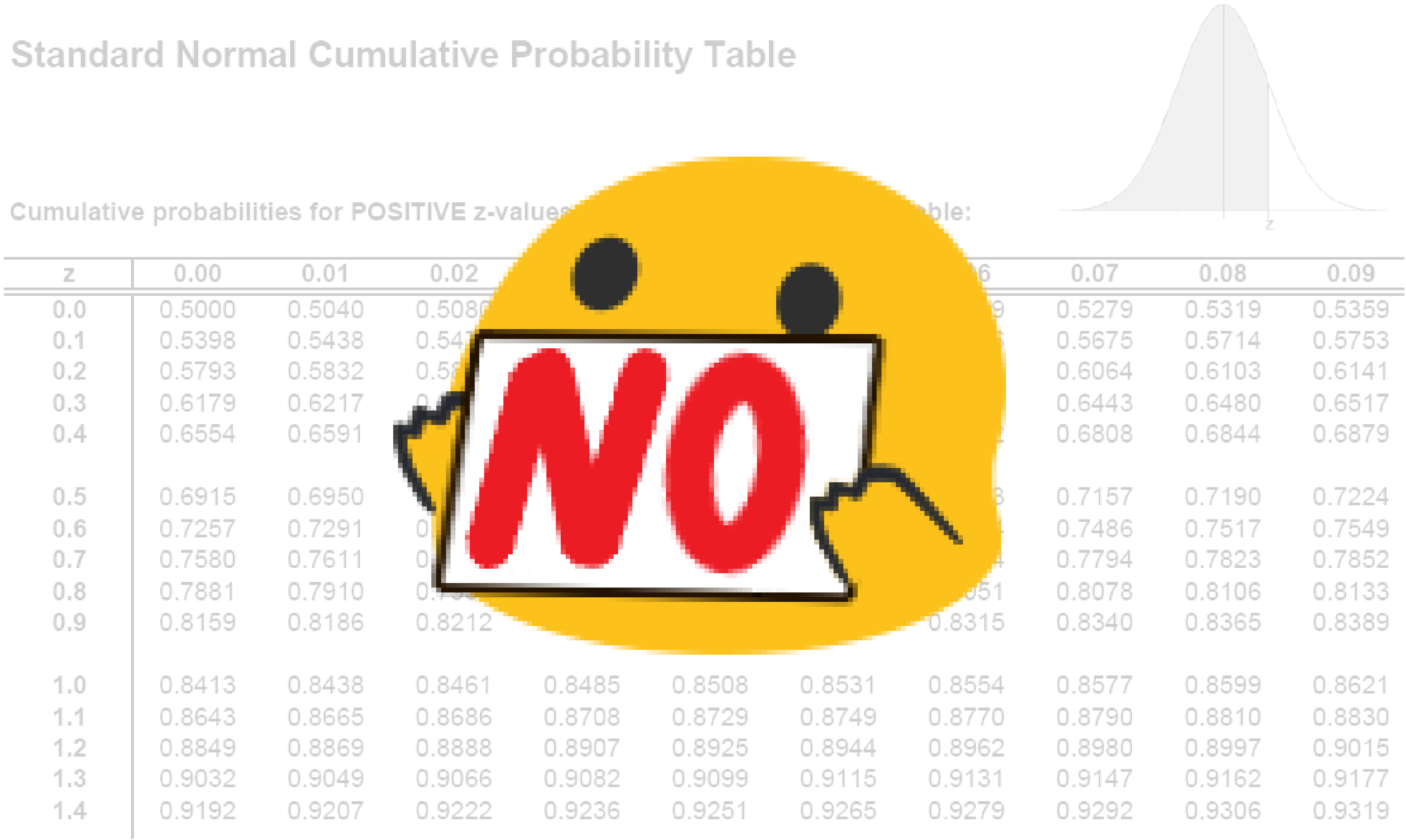
$\Phi(0.54) = 0.7054$



Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Do We Have To Use The Table??



Standard Normal Cumulative Probability Table

Cumulative probabilities for POSITIVE z-values [illegible] ble:

z	0.00	0.01	0.02	[illegible]	[illegible]	[illegible]	[illegible]	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	[illegible]	[illegible]	[illegible]	[illegible]	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	[illegible]	[illegible]	[illegible]	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Table is kinda old school

We Are Computer Scientists!

Every modern programming language has phi stored in a library:

```
from scipy import stats

stats.norm.cdf(x, mean, std)
```

$$= P(X < x) \text{ where } X \sim \mathcal{N}(\mu, \sigma^2)$$

We Are Computer Scientists!

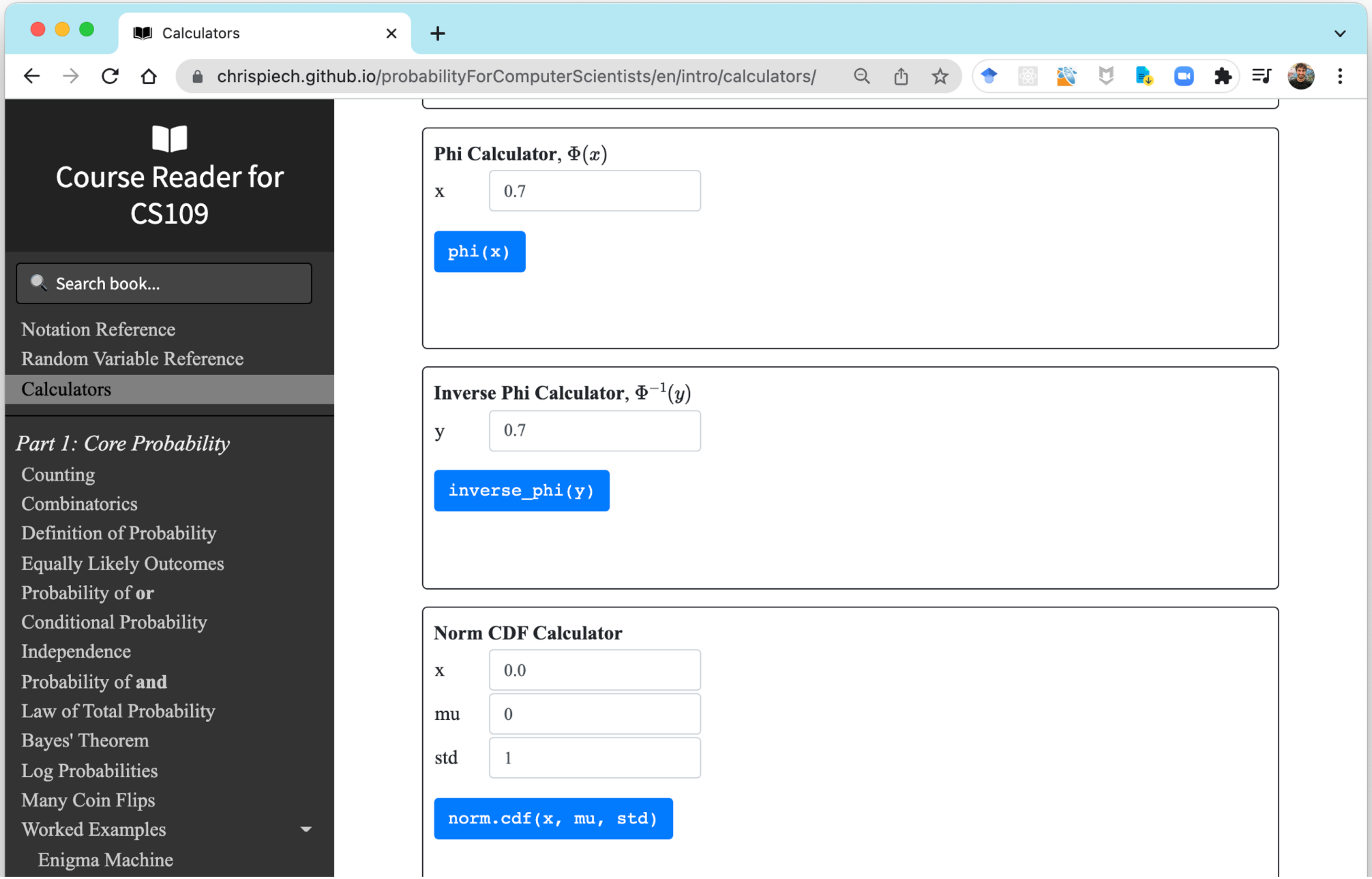
Every modern programming language has phi stored in a library:

```
from scipy import stats

stats.norm.cdf(x, mean, std)
```

$$= P(X < x) \text{ where } X \sim \mathcal{N}(\mu, \sigma^2)$$

not variance!!!

I Made One For You



Calculators

chrispiech.github.io/probabilityForComputerScientists/en/intro/calculators/

Course Reader for CS109

Search book...

- Notation Reference
- Random Variable Reference
- Calculators

Part 1: Core Probability

- Counting
- Combinatorics
- Definition of Probability
- Equally Likely Outcomes
- Probability of **or**
- Conditional Probability
- Independence
- Probability of **and**
- Law of Total Probability
- Bayes' Theorem
- Log Probabilities
- Many Coin Flips
- Worked Examples
  - Enigma Machine

Phi Calculator, $\Phi(x)$

x: 0.7

`phi(x)`

Inverse Phi Calculator, $\Phi^{-1}(y)$

y: 0.7

`inverse_phi(y)`

Norm CDF Calculator

x: 0.0

mu: 0

std: 1

`norm.cdf(x, mu, std)`

Practice: Submarine Manufacturing

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = \int_{490}^{510} f(X = x)dx$$

Practice: Submarine Manufacturing

If $X \sim \mathcal{N}(\mu, \sigma^2)$, $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

Now using the CDF!

What fraction of the panels you manufacture will meet standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$P(490 \leq X \leq 510) = ?$

Practice: Submarine Manufacturing

If $X \sim \mathcal{N}(\mu, \sigma^2)$, $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

Now using the CDF!

What fraction of the panels you manufacture will meet standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = P(X < 510) - P(X < 490) = \Phi\left(\frac{510 - 500}{6}\right) - \Phi\left(\frac{490 - 500}{6}\right)$$

subtract mean, divide by std. dev.

Practice: Submarine Manufacturing

If $X \sim \mathcal{N}(\mu, \sigma^2)$, $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Your team is tasked with producing the side panels for Deep Sea Submarines. Physics requires all panels to be built within 10 micros of 500. You check how precise your manufacturing is, and find these stats:

- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

Now using the CDF!

What fraction of the panels you manufacture will meet standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = P(X < 510) - P(X < 490) = \Phi\left(\frac{510 - 500}{6}\right) - \Phi\left(\frac{490 - 500}{6}\right)$$

$$= \Phi\left(\frac{5}{3}\right) - \left(1 - \Phi\left(\frac{5}{3}\right)\right) = 2 \ \Phi\left(\frac{5}{3}\right) - 1 \approx 0.904$$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 6)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-x) = 1 - \Phi(x)$

Are you ready for something different?

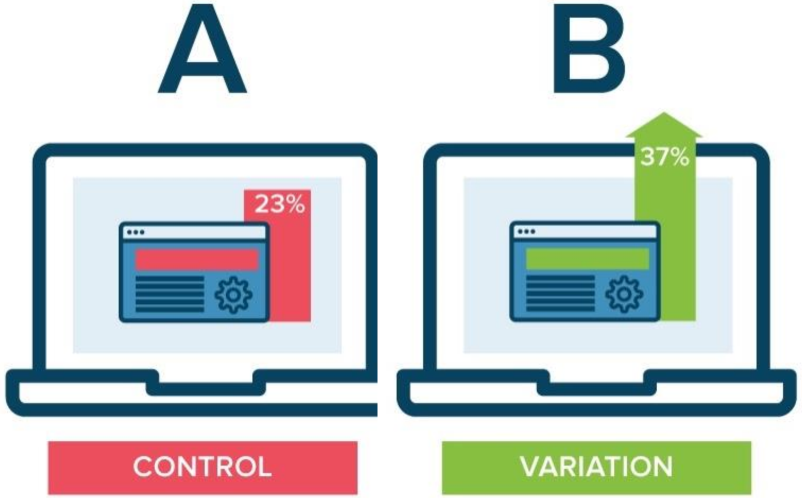
Pop quiz!
(jk)

Midterm: Website Testing

A new website design is tested out on 1M users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 501k.

What is P(CEO endorses change| it has no effect)?

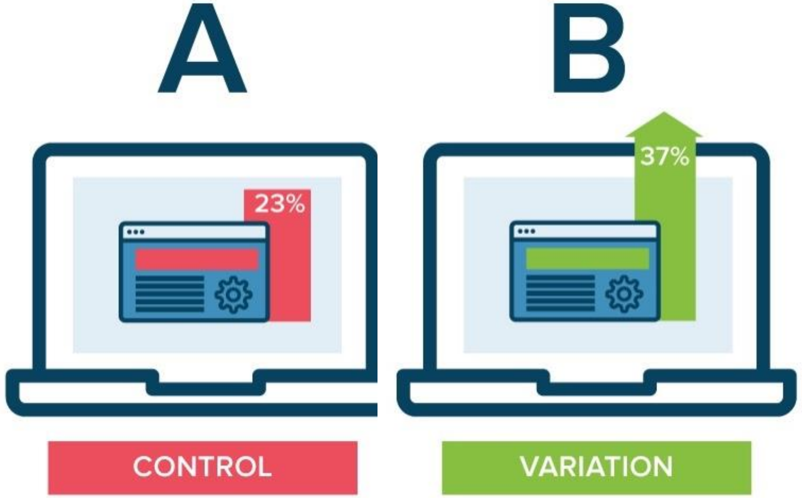


Midterm: Website Testing

A new website design is tested out on 1M users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 501k.

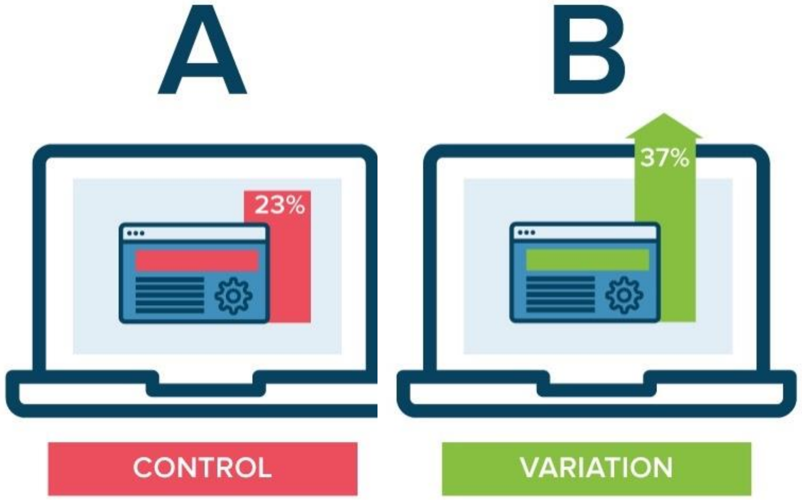
What is P(CEO endorses change| it has no effect)?



$$X \sim \text{Bin}(n = 10^6, p = 0.5)$$

$$P(X > 501000) = \sum_{i=501000}^{10^6} \binom{10^6}{i} (0.5)^i (0.5)^{10^6 - i}$$

Midterm: Website Testing



A new website design is tested out on 1M users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 501k.

What is P(CEO endorses change| it has no effect)?

$$X \sim \text{Bin}(n = 10^6, p = 0.5)$$

$$P(X > 501000) = \sum_{i=501000}^{10^6} \binom{10^6}{i} (0.5)^i (0.5)^{10^6 - i}$$

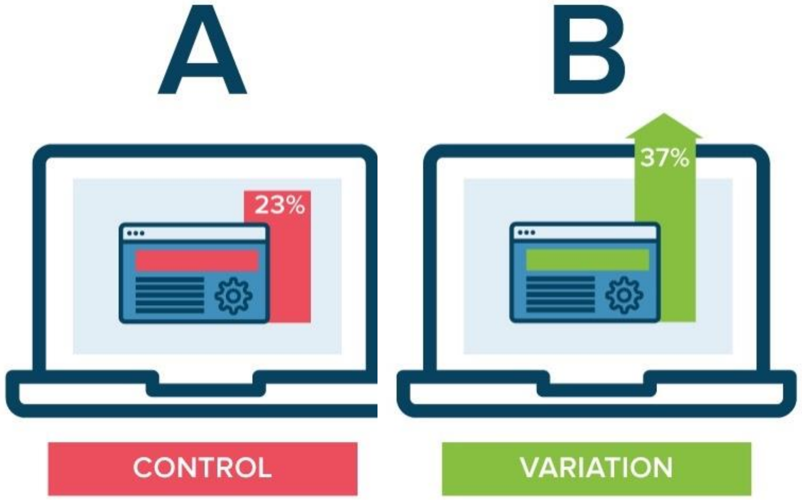
```
>>> math.comb(1000000,501000)
```

Midterm: Website Testing

A new website design is tested out on 1M users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 501k.

What is P(CEO endorses change| it has no effect)?



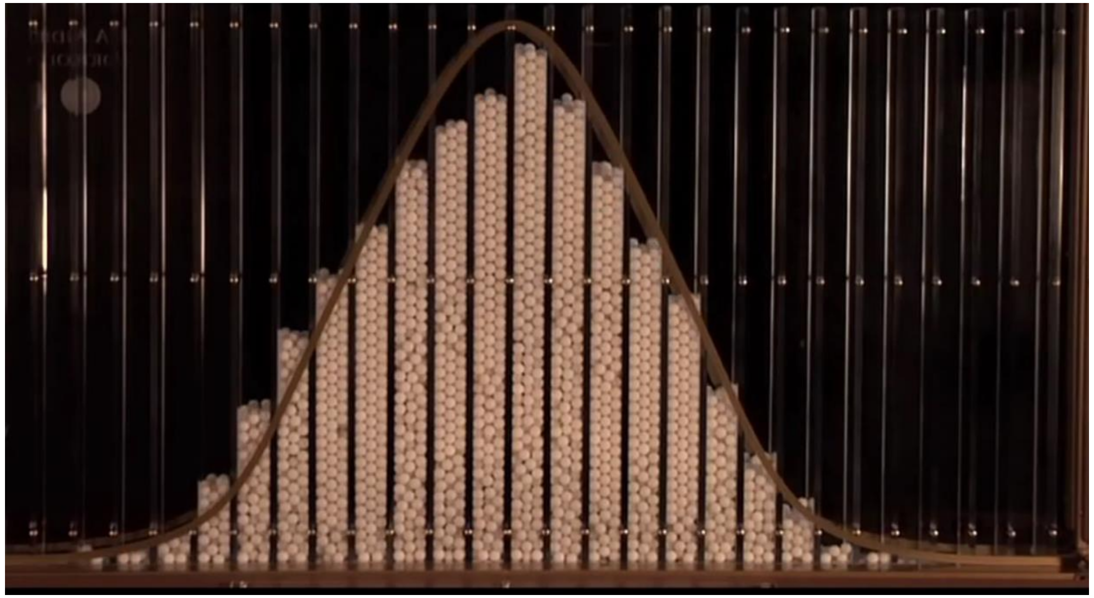
$$X \sim \text{Bin}(n = 10^6, p = 0.5)$$

$$P(X > 501000) = \sum_{i=501000}^{10^6} \binom{10^6}{i} (0.5)^i (0.5)^{10^6 - i}$$

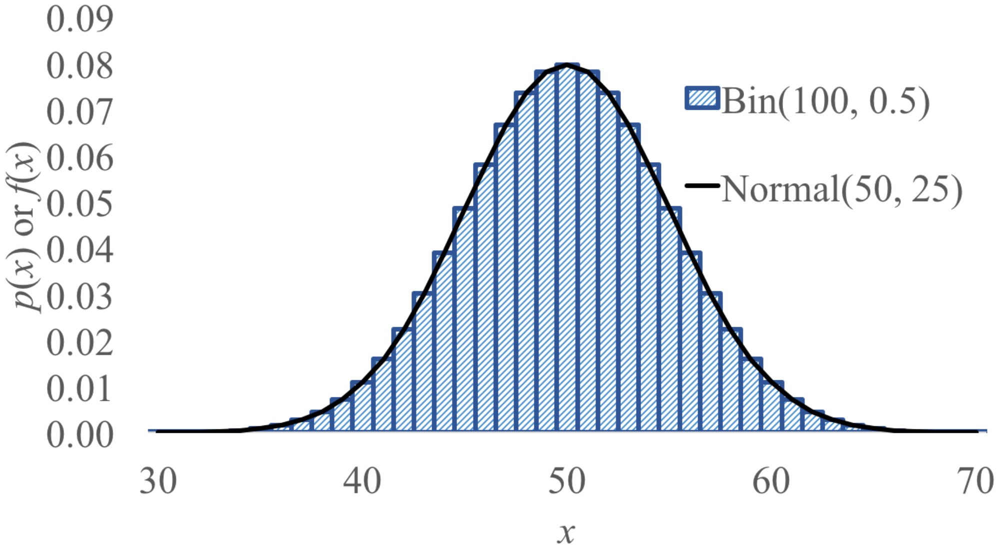
```
>>> math.comb(1000000,501000)
ValueError: Exceeds the limit (4300 digits) for
integer string conversion; use
sys.set_int_max_str_digits() to increase the limit
>>>
```

$$P(X > 505000) = \sum_{i=505000}^{10^6} \binom{10^6}{i} (0.5)^i (0.5)^{10^6-}$$

Don’t worry, Normal approximates Binomial

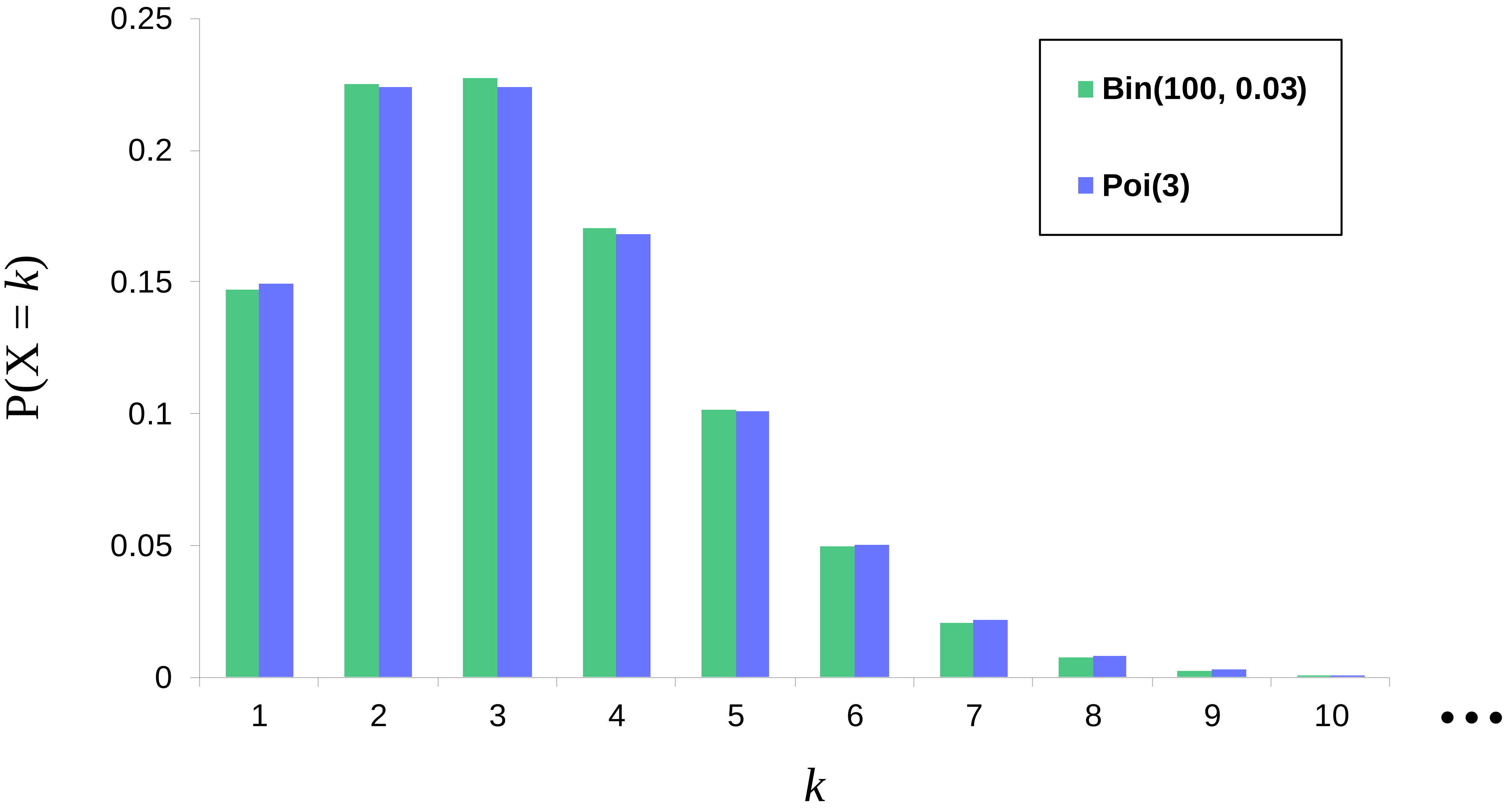


Galton Board

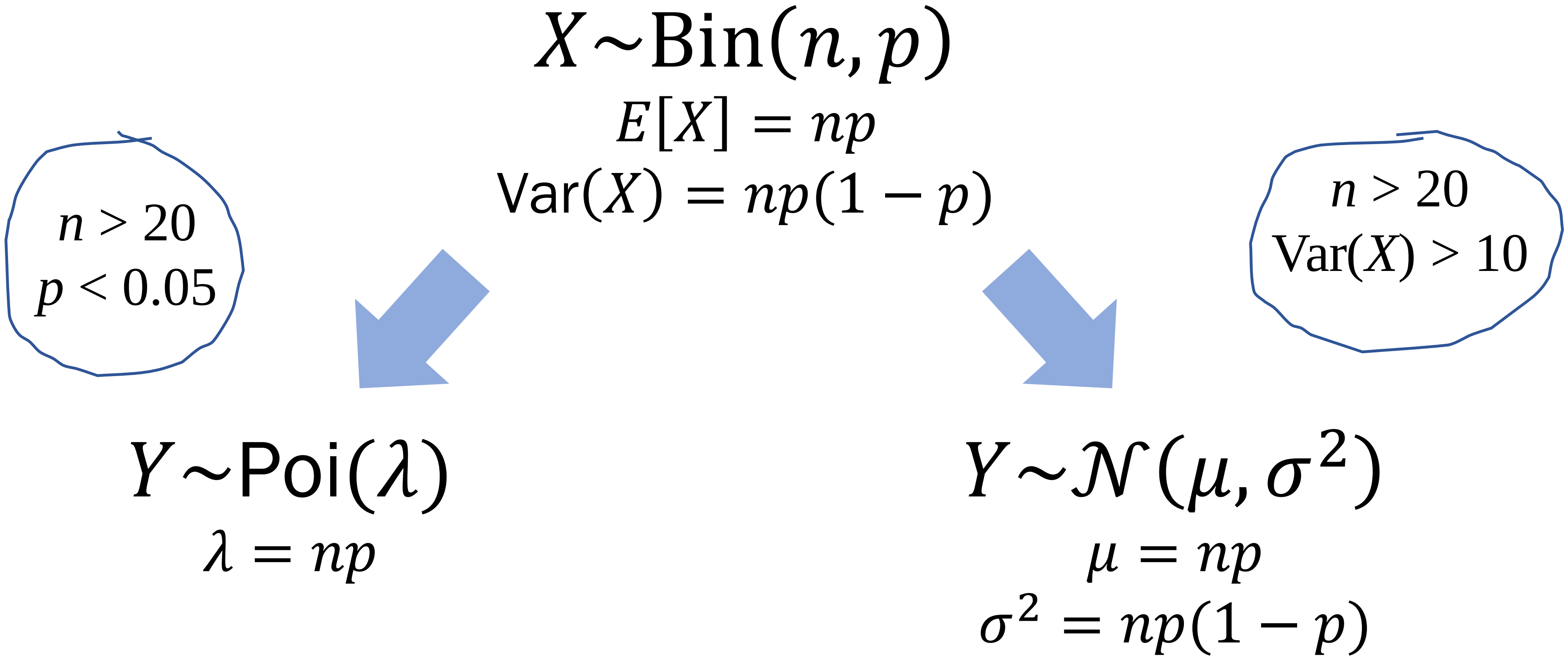


(We’ll explain ***why*** in 2 weeks’ time)

Poisson Approximates Binomial, With Extreme *n* and *p*



Two Ways To Approximate The Binomial



$$X \sim \text{Bin}(n, p)$$

$$E[X] = np$$

$$\text{Var}(X) = np(1-p)$$

$n > 20$
$p < 0.05$

$n > 20$
$\text{Var}(X) > 10$

$$Y \sim \text{Poi}(\lambda)$$

$$\lambda = np$$

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

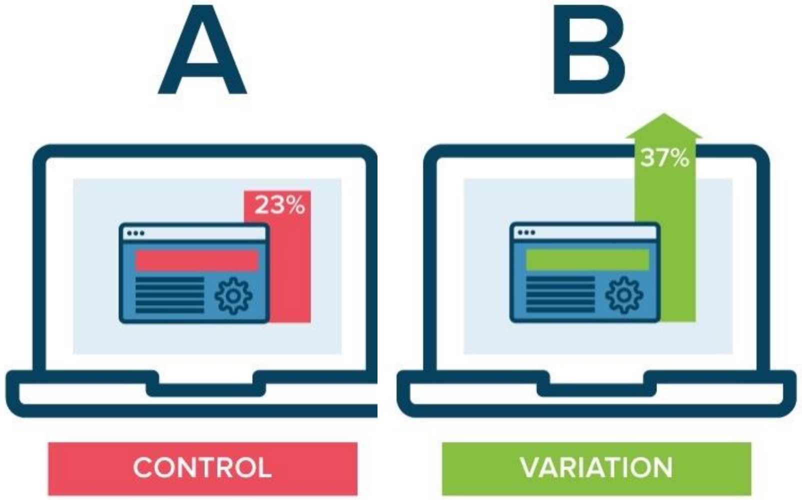
$$\mu = np$$

$$\sigma^2 = np(1-p)$$

Poisson approximation for big *n*, small *p*.

Normal approximation for big *n*, medium *p*.

Midterm: Website Testing



A new website design is tested out on 1M users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 501k.

What is P(CEO endorses change| it has no effect)?

$$X \sim \text{Bin}(n = 10^6, p = 0.5)$$

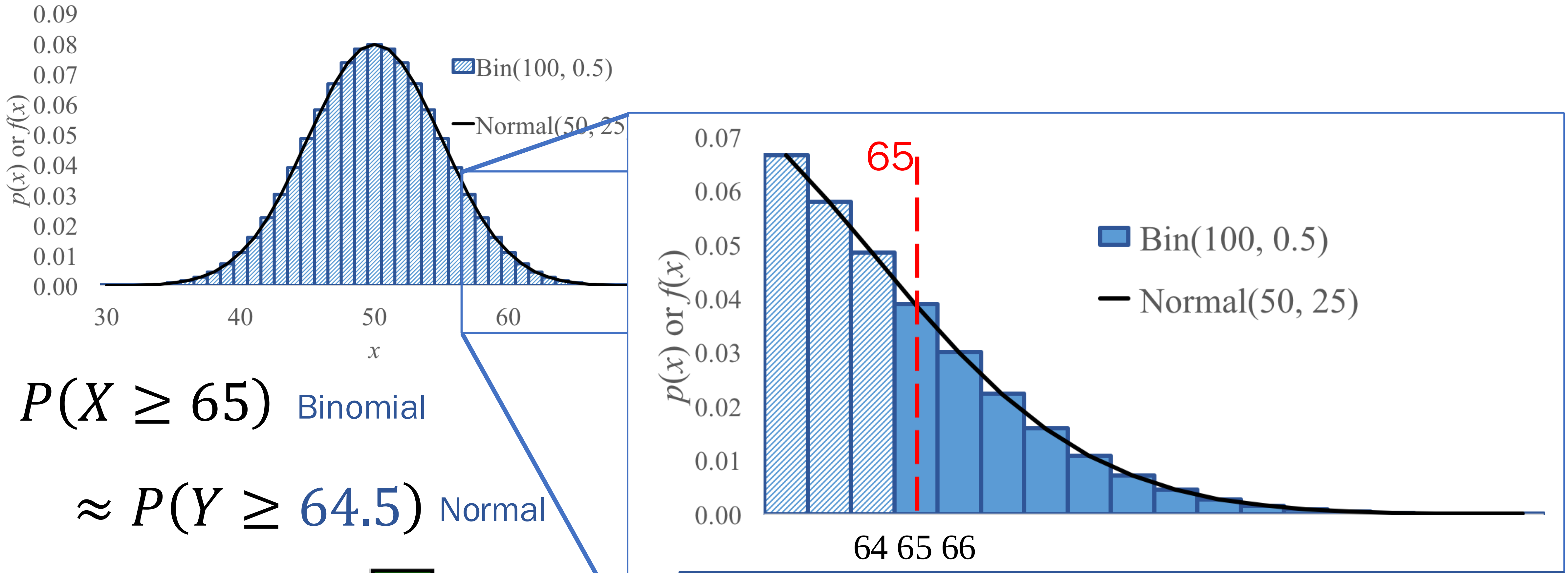
$$Y \sim N(\mu = \underbrace{500000}_{n \cdot p}, \sigma^2 = \underbrace{250000}_{n \cdot p \cdot (1-p)})$$

$$\begin{aligned} P(X > 501000) &\approx P(Y > 501000) \\ &\approx 1 - P(Y < 501000) \\ &\approx 1 - F_Y(501000) \approx 0.02275 \end{aligned}$$

Correct answer is 0.02270

Normal Approximation (with continuity correction)

In our website testing, $Y \sim \mathcal{N}(50, 25)$ approximates $X \sim \text{Bin}(100, 0.5)$.



$P(X \geq 65)$ Binomial

$\approx P(Y \geq 64.5)$ Normal

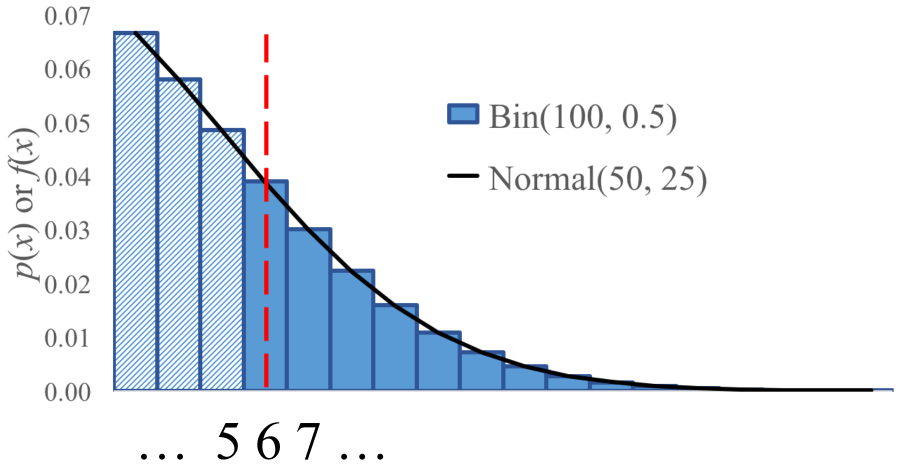
≈ 0.0018 ✓

You must perform a **continuity correction** when approximating a Binomial RV with a Normal RV.

Continuity correction

If $Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

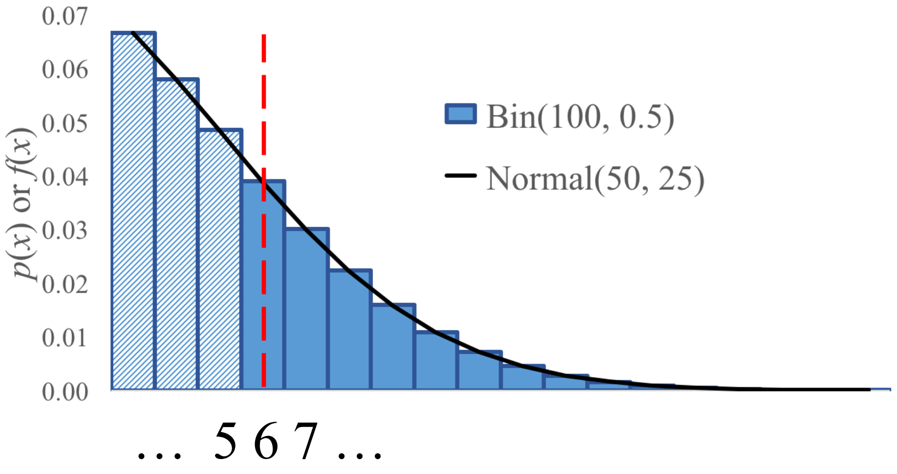
Discrete (e.g., Binomial) probability question	⇨	Continuous (Normal) probability question
$P(X = 6)$		
$P(X \geq 6)$		
$P(X > 6)$		
$P(X < 6)$		
$P(X \leq 6)$		



Continuity correction

If $Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

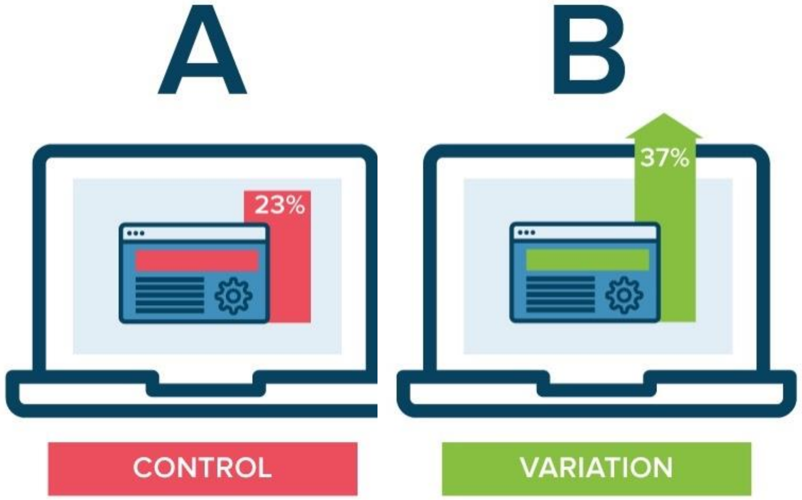
Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	$P(5.5 \le Y \le 6.5)$
$P(X \ge 6)$	$P(Y \ge 5.5)$
$P(X > 6)$	$P(Y \ge 6.5)$
$P(X < 6)$	$P(Y \le 5.5)$
$P(X \le 6)$	$P(Y \le 6.5)$



Midterm: Website Testing

A new website design is tested out on 1M users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 501k.



What is P(CEO endorses change| it has no effect)?

$$X \sim \text{Bin}(n = 10^6, p = 0.5)$$

$$Y \sim N(\mu = \underset{n \cdot p}{500000}, \sigma^2 = \underset{n \cdot p \cdot (1-p)}{250000})$$

$$\begin{aligned} P(X > 501000) &\approx P(Y > 501000.5) \\ &\approx 1 - P(Y < 501000.5) \\ &\approx 1 - F_Y(501000.5) \approx 0.02270 \end{aligned}$$

YOU ARE AMAZING!

Just Invented the Normal Approximation

Stanford Admissions (a while back)

Stanford accepts 2480 students.

- Each admitted student matriculates w.p. 0.68 (independent trials)
- Let $X = \#$ of students who will attend

What is $P(X > 1745)$? *Give a numerical approximation.*

Strategy:

A. Just Binomial
B. Poisson
C. Normal
D. None/other

(by yourself)

Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student matriculates w.p. 0.68 (independent trials)
- Let X = # of students who will attend

What is $P(X > 1745)$? *Give a numerical approximation.*

Strategy:

A. Just Binomial — not an approximation (also computationally expensive)

B. Poisson — $p = 0.68$, not small enough

[✓ C.] Normal — ✓ Variance $np(1-p) = 540 > 10$

D. None/other

Define an approximation

Let $Y \sim \mathcal{N}\left(E[X], \text{Var}(X)\right)$

$$E[X] = np = 1686$$

$$\text{Var}(X) = np(1-p) \approx 540 \rightarrow \sigma = 23.3$$

$$P(X > 1745) \approx P(Y \geq 1745.5)$$

⚠ Continuity correction

Solve

SciPy can do this

$$P(Y \geq 1745.5) = 1 - \boxed{F(1745.5)}$$

$$= 1 - \Phi\left(\frac{1745.5 - 1686}{23.3}\right)$$

$$= 1 - \Phi(2.54) \approx 0.0055$$

How many students should Stanford admit?

The Stanford Daily

Class of 2018 admit rates lowest in University history

March 28, 2014 16 Comments

Alex Zivkovic
Desk Editor

Stanford admitted 2,138 students to the Class of 2018 in this year's admissions cycle, producing – at 5.07 percent – the lowest admit rate in University history.

The University received a total of 42,167 applications this year, a record total and a 8.6 percent increase over last year's figure of 38,828. Stanford accepted 748 students

Admit rate: 4.3%
Yield rate: 81.9%

Pedagogical Pause

Great questions!
Great thinkers start with great questions. Ask away!!!

Super Question:

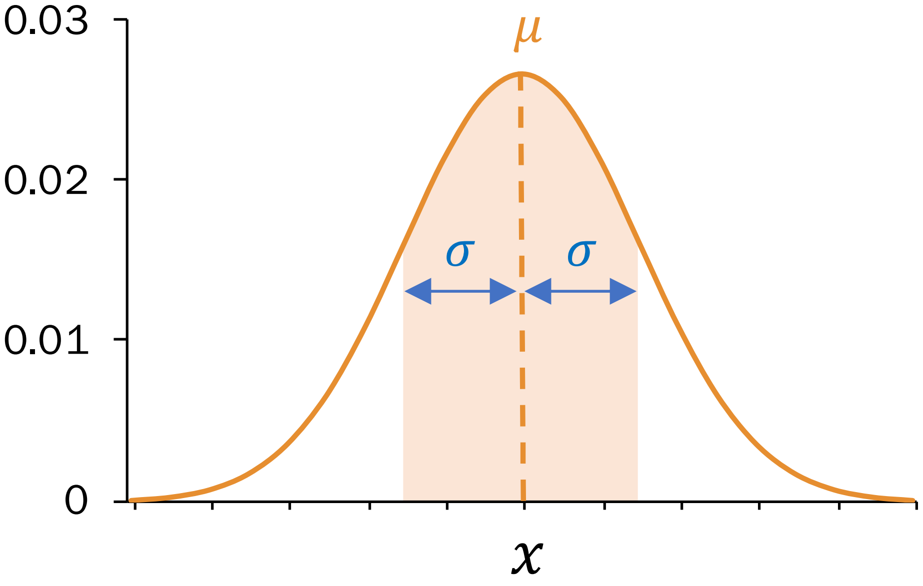
Why Be Normal? 68% rule

You may have heard the statement:

> "68% of the class will fall within 1 standard deviation of the exam average."

In general, this is only true of normal distributions:

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with CDF F.



$$P(|X-\mu| < \sigma) = P(\mu - \sigma < X < \mu + \sigma)$$
$$= F(\mu+\sigma) - F(\mu-\sigma)$$
$$= \Phi\left(\frac{(\mu+\sigma)-\mu}{\sigma}\right) - \Phi\left(\frac{(\mu-\sigma)-\mu}{\sigma}\right)$$
$$= \Phi(1) - \Phi(-1) = \Phi(1) - \left(1 - \Phi(1)\right)$$
$$= 2\Phi(1) - 1 \quad \approx 2(0.8413) - 1 = 0.6826$$

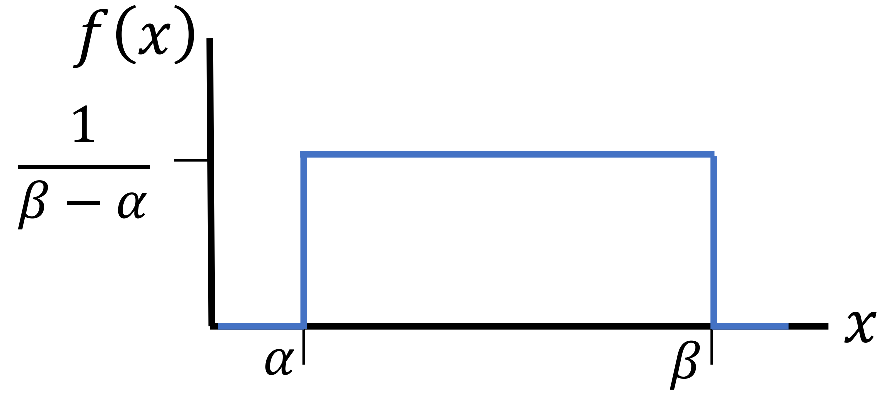
Why Be Normal? 68% rule

You may have heard the statement:

> “68% of the class will fall within 1 standard deviation of the exam average.”

In general, this is only true of normal distributions:

Counterexample: Let $X \sim \text{Uni}(\alpha, \beta)$.



$$\mu = E[X] = \frac{\alpha + \beta}{2}$$

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12} \qquad \sigma = \text{SD}(X) = \frac{\beta - \alpha}{\sqrt{12}}$$

$$\begin{aligned} P(|X - \mu| < \sigma) &= P(\mu - \sigma < X < \mu + \sigma) \\ &= \frac{1}{\beta - \alpha} \cdot [(\mu + \sigma) - (\mu - \sigma)] \\ &= \frac{1}{\beta - \alpha}[2\sigma] = \frac{1}{\beta - \alpha} \cdot \left[2 \cdot \frac{\beta - \alpha}{\sqrt{12}}\right] \\ &= 2/\sqrt{12} \approx 0.58 \end{aligned}$$

Challenge

11.11

光棍节

SINGLE'S DAY

SALE

Enough Servers?

You receive $R \sim N(\mu = 10^6, \sigma = 10^4)$ requests in the busiest min

You are going to buy N servers

Each server can handle 10,000 requests per min, otherwise you drop requests

What is the smallest value of N such that P(drop) < 0.0001

Extra Content

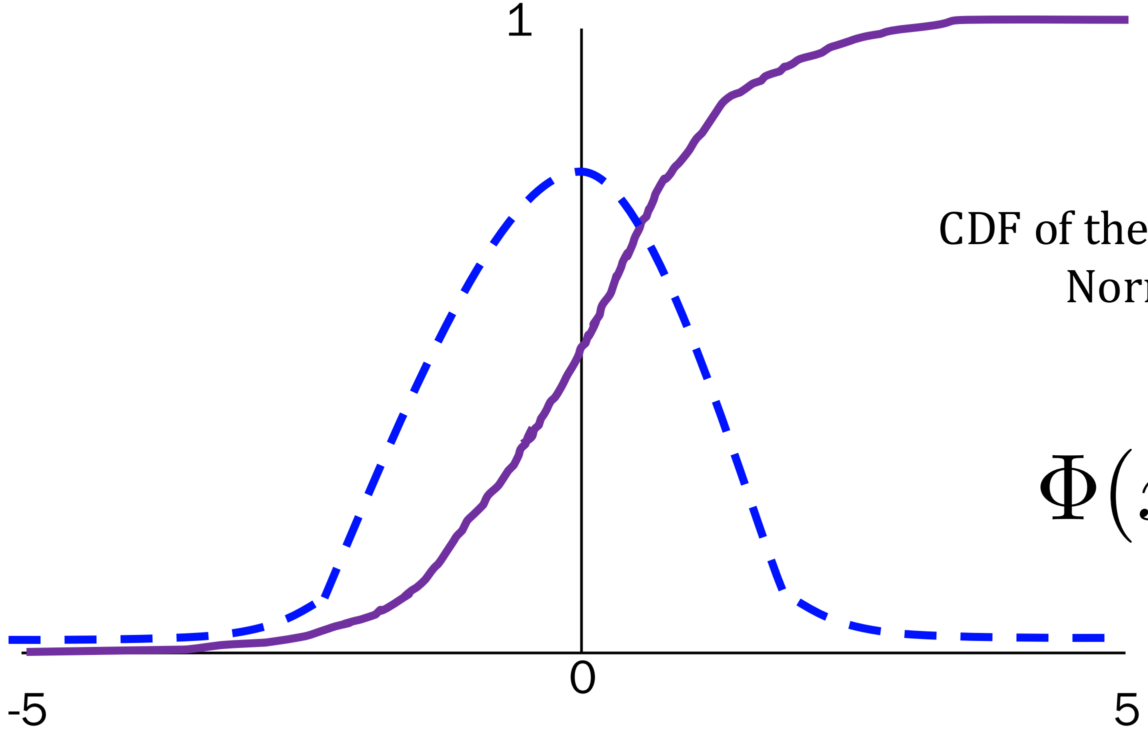
How does python sample from a Gaussian?

```
from random import *

for i in range(10):
  mean = 5
  std = 1
  sample = gauss(mean, std)
  print sample
```

How does this work?

```
3.79317794179
5.19104589315
4.209360629
5.39633891584
7.10044176511
6.72655475942
5.51485158841
4.94570606131
6.14724644482
4.73774184354
```

How Does a Computer Sample a Normal?



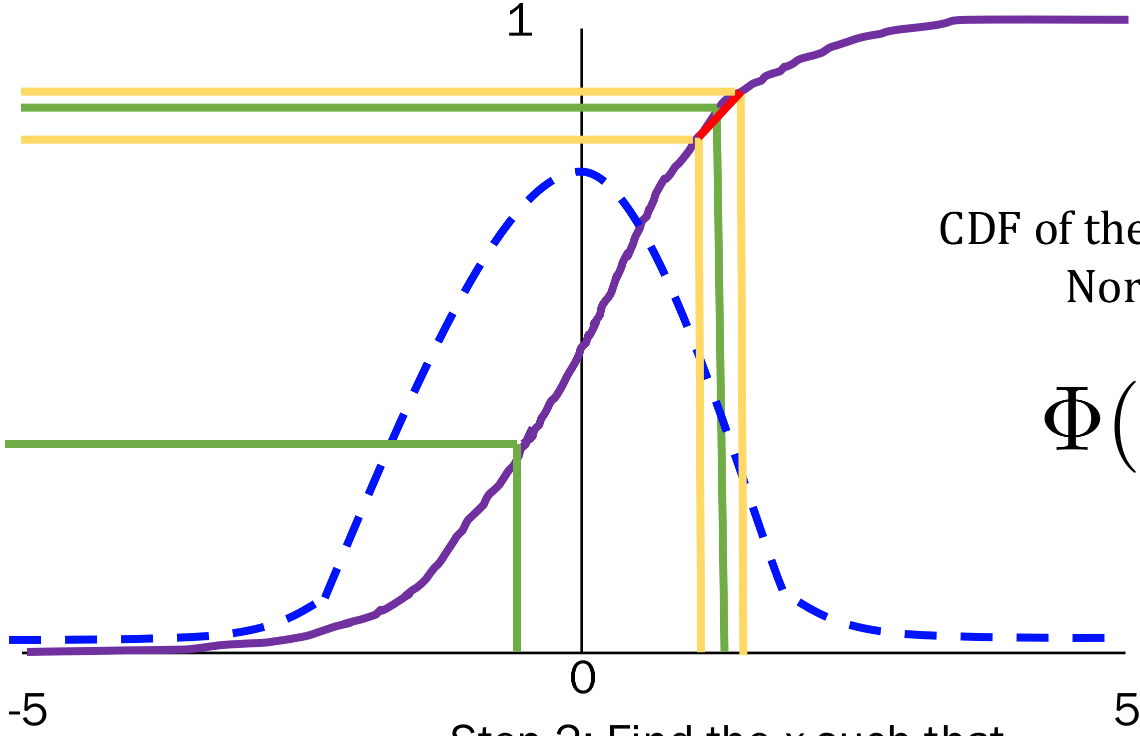
CDF of the Standard Normal

$$\Phi(x)$$

How Does a Computer Sample a Normal?

Inverse Transform Sampling

Step 1: pick a uniform number *y* between 0,1



Step 2: Find the *x* such that

$$\Phi(x) = y$$

$$x = \Phi^{-1}(y)$$

CDF of the Standard Normal

$$\Phi(x)$$

Sample 1:
1.201234

Sample 2:
-0.45123

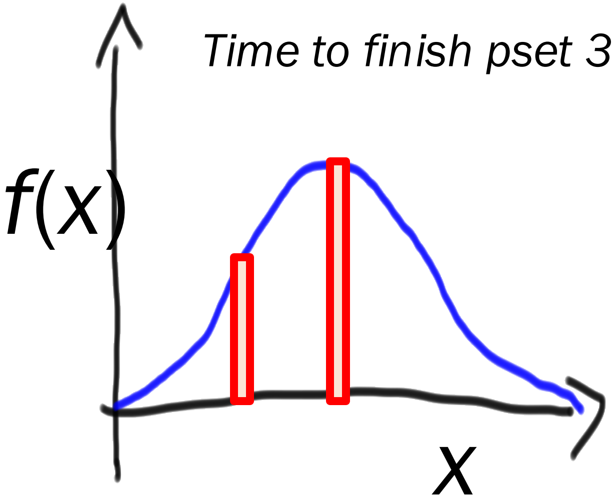
Further reading: Box–Muller transform

Relative values of a PDF

Relative Probability of Continuous Variables

X = time to finish pset 3

$X \sim N(\mu = 10, \sigma^2 = 2)$

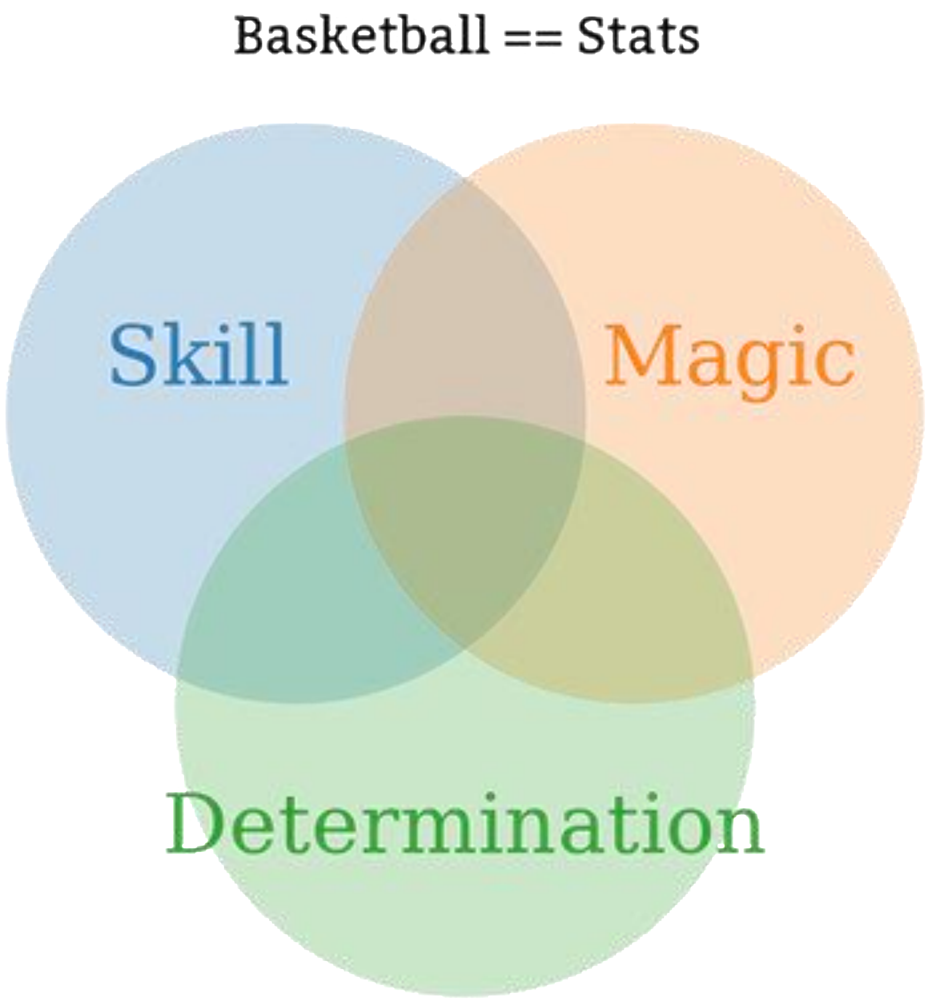


How much more likely are you to complete in 10 hours than in 5?

$$
\begin{aligned}
\frac{P(X=10)}{P(X=5)} &= \frac{\varepsilon f(X=10)}{\varepsilon f(X=5)} \\
&= \frac{f(X=10)}{f(X=5)} \\
&= \frac{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(5-\mu)^2}{2\sigma^2}}} \\
&= \frac{\frac{1}{\sqrt{4\pi}} e^{-\frac{(10-10)^2}{4}}}{\frac{1}{\sqrt{4\pi}} e^{-\frac{(5-10)^2}{4}}} \\
&= \frac{e^0}{e^{-\frac{25}{4}}} = 518
\end{aligned}
$$

Gaussian and ELO

Gaussian Sampling and ELO ratings



What is the probability that the Warriors win?
How do you model zero-sum games?

Gaussian Sampling and ELO ratings

Each team has an ELO score S, calculated based on its past performance.

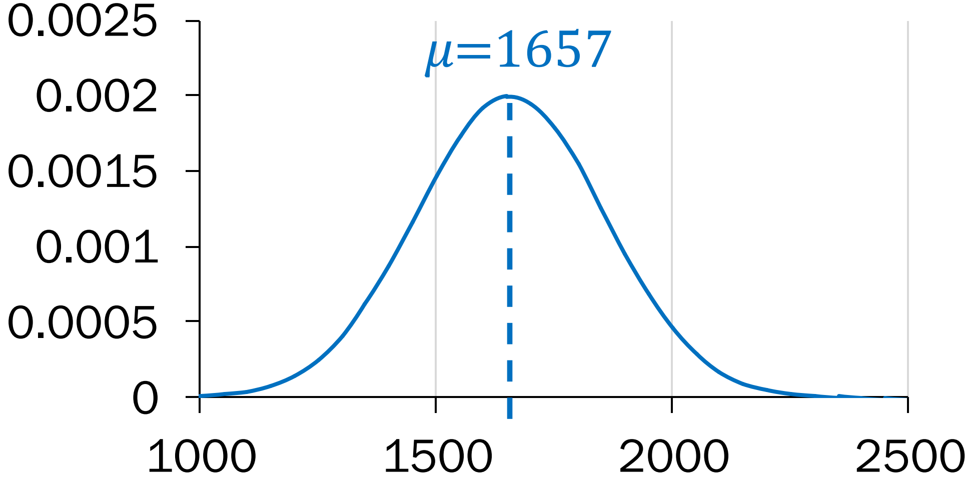
- Each game, a team has ability $A \sim \mathcal{N}(S, 200^2)$.
- The team with the higher sampled ability wins.

What is the probability that Warriors win this game?

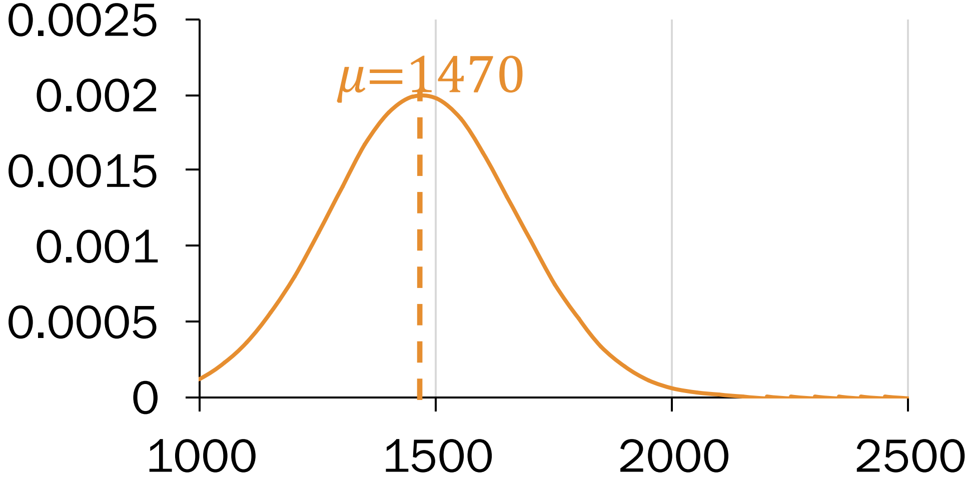
Want: $P(\text{Warriors win}) = P(A_W > A_O)$

Arpad Elo

Warriors' $A_W \sim \mathcal{N}(S = 1657, 200^2)$



Opponent's $A_O \sim \mathcal{N}(S = 1470, 200^2)$



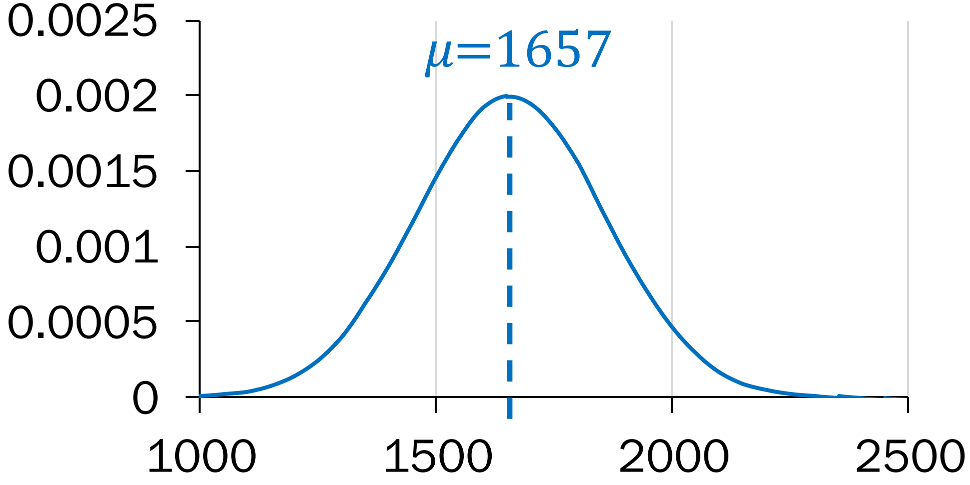
Gaussian Sampling and ELO ratings

Want: $P(\text{Warriors win}) = P(A_W > A_O)$

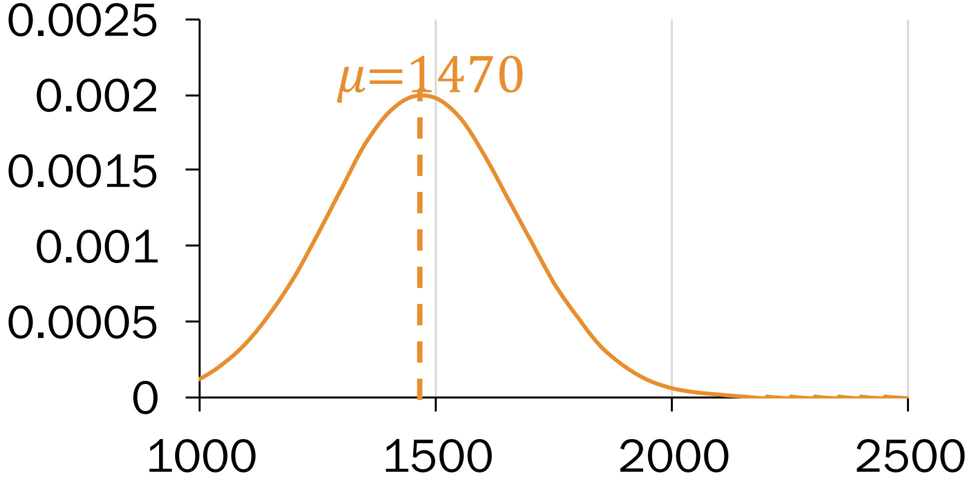
```
from scipy import stats
WARRIORS_ELO = 1657
OPPONENT_ELO = 1470
STDEV = 200
NTRIALS = 10000

nSuccess = 0
for i in range(NTRIALS):
  w = stats.norm.rvs(WARRIORS_ELO, STDEV)
  o = stats.norm.rvs(OPPONENT_ELO, STDEV)
  if w > o:
    nSuccess += 1
print("Warriors sampled win fraction: ",
      float(nSuccess) / NTRIALS)
```

≈ 0.7488, calculated by sampling

Warriors' $A_W \sim \mathcal{N}(S = 1657, 200^2)$



Opponent's $A_O \sim \mathcal{N}(S = 1470, 200^2)$



Is there a better way?

$$P(A_W > A_o)$$

- This is a probability of an event involving ***two continuous*** random variables!
- We’ll solve this problem analytically in two weeks’ time.

Big goal for next time: Events involving two ***discrete*** random variables.
Stay tuned!