

Normal Distribution

Chris Piech

CS109, Stanford University



PSet 3 is out!

We put a lot of love into it,
and hope you enjoy

Uses discrete random variables

Uses continuous random variables

PS3 Simulate Distributions

Understanding the *process* that leads to different random variables is a great way to gain familiarity for what they mean.

For each random variable, write a function that simulates its generation process. Your function should return a number. The only probability function that you may use when coding your solution is `random.uniform(0, 1)`: a function that returns a uniform random in the range [0, 1].

We provide a solution to the Bernoulli random variable below:

$X \sim \text{Ber}(n)$ Returns 1 or 0 to indicate whether or not an underlying event was “successful.”

```
import random

def simulate_bernoulli(p = 0.4):
    if random.uniform(0, 1) < p:
        return 1
    return 0
```

$X \sim \text{Bin}(n = 20, p = 0.4)$ The number of successes after 20 independent experiments.

```
def simulate_binomial(n = 20, p = 0.4):
    # You will implement this function!
```

$X \sim \text{Geo}(p = 0.03)$ The number of trials until the first success.

```
def simulate_geometric(p = 0.03):
    # You will implement this function!
```

Answer Editor Solution

Agent:

```
3 def simulate_bernoulli(p = 0.4):
4     # We did this one for you!
5     if random.uniform(0, 1) < p:
6         return 1
7     return 0
8
9 def simulate_binomial(n = 20, p = 0.4):
10    # TODO: your code here
11    return 0
12
13 def simulate_geometric(p = 0.03):
14    # TODO: your code here
15    return 0
16
17 def simulate_negative_binomial(r = 5, p = 0.03):
18    # TODO: your code here
19    return 0
20
21 def simulate_poisson(lambda_parameter = 3.1):
22    # TODO: your code here
23    return 0
24
25 def simulate_exponential(lambda_parameter = 3.1):
26    # TODO: your code here
27    return 0
28
```

Run One Game Test Agent

Console

Simulate!

Great way to make sure you really understand the random variables

Don't use numpy or scipy!

Think about the poisson proof for poisson and exponential

PS3 Measles Test

To determine whether they have measles, 60 people have their blood tested. However, rather than testing each individual separately, it is decided to first place the people into groups of 6. The blood samples of the 6 people in each group will be pooled and analyzed together. If the test is negative, one test will suffice for the 6 people, whereas if the test is positive, each of the 6 people will also be individually tested and, in all, 7 tests will be made on this group. Note that we assume that the pooled test will be positive if at least one person in the pool has measles. Assume that the probability that a person has measles is 5% for all people, independently of each other, and compute the expected number of tests necessary for each group of 6 people.



Answer Editor Solution

Numeric Answer: 2.589 ✓ Check Answer

Explanation:

Block LaTeX Image B $\lt;/\gt$ I U

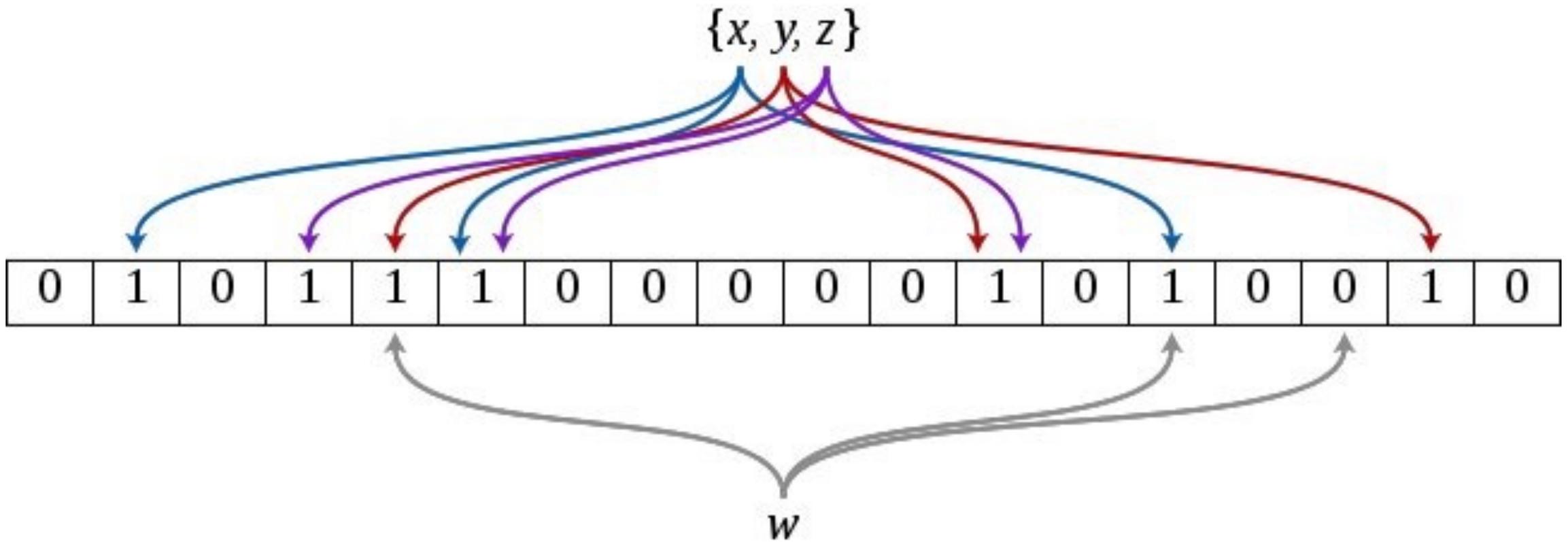
1. Define a random variable
2. Declare its distribution
3. Write the question as a probability statement using the random variable
4. Solve

Previous Question Next Question

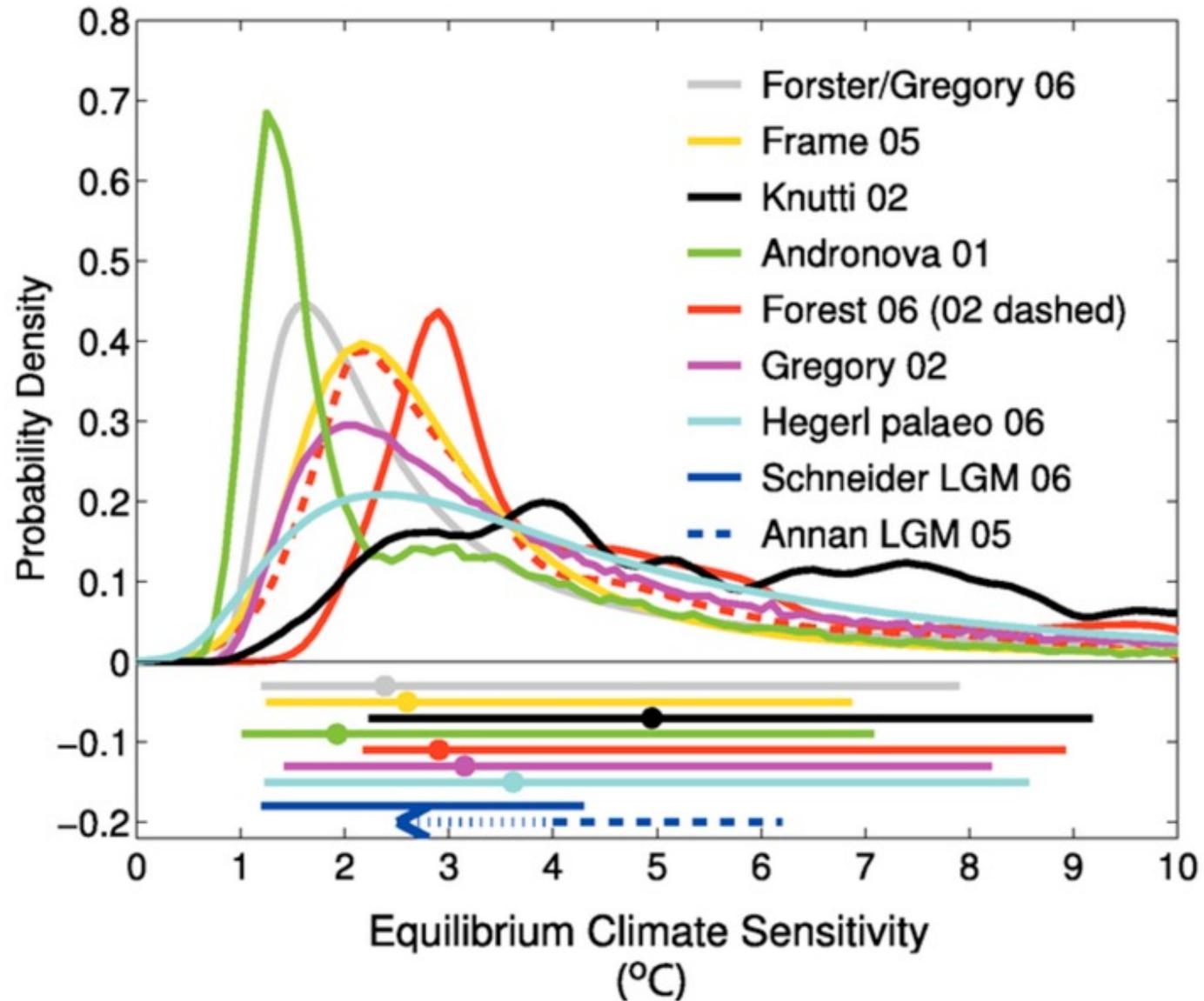
Many prototypical random variable questions.

Batch testing is a thing!

Bloom Filter



Climate Sensitivity



Which is Random?

Sequence 1:

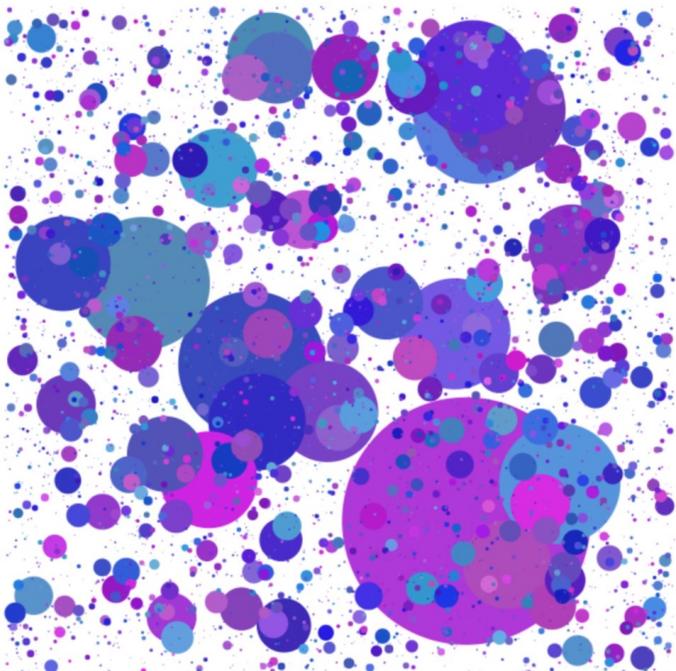
TTHHTHTTHTTTHTTTHTTTHTTHTHHTHHTHTHHTTTHTHTHTTTHTHHTTHTHHTHTTTHT
HTTHTTTHHHTHHTHTTHTHTTHTHHTHHHTTHTHTTTHTHTTHTHTHTHTTHTHTHHHTTHT
HTHHTHHHTHTHTTTHTTHTHTHTHTTHTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHT
TTHTHTTTHTHTHTHTHTHTTHTHHHTHTHTHTTHTHTHTTHTTTHTHTTHTTTHTHTTHTTHTTHT
TTHTHTHTHT

Sequence 2:

HTHHHTHTTHTTTTTTTTTTHHHTTTHTTTTTHTTTHHHTTHTHTTTTTTTHTHTTTTTTHHHHTH
THTTTHTTTHTTTHTTTTTHTHHTHHHTTTTTTHHHHTHHHTTTTTHTHTTTHHHHTHHHHHHHTT
HHTHHTHHHHHHHTTHTHTTTHTTTTTHTHTTHTTHTHTTHTTHTTHTTHTTHTTHTTHTTHTTHT
TTHTTTTTHTHTTHHHHTTTTTHTHHHHHTHTHTHTHHHTHTTTHHHTHHHHHHHTHHHTHTTT
HHHTTTHTHT

PS3 Algorithmic Art

We want to generate probabilistic artwork. We are going to use random variables to make a picture filled with circles. The goal of this problem is to give you a fun, and gentle introduction to continuous random variables. Here is an example of what the artwork could look like:



In our art, the circles are different sizes. Specifically, each circle's **radius** is drawn from a continuous random variable called a Pareto distribution (which is described below). The placement algorithm is by size: we sample 5000 circle sizes. Sort them by size. largest to

Answer Editor Solution

Numeric Answer: 0.22936657838275 ✓ Check Answer

```
1 import math
2 import numpy as np
3 import colorsys
4
5
6 def main():
7     draw_circle(50, 100, 200, random_color())
8
9
10 def draw_circle(radius, center_x, center_y, color):
11     """
12     do not touch! Does some fancy python -> js coding
13     context is a global which gives access the the
14     HTML5 canvas below
15     """
16     context.beginPath()
17     # context.strokeStyle = color
18     context.fillStyle=color
19     context.arc(center_x, center_y, radius, 0, 2 * math.pi)
20     context.fill()
21
22 def random_color():
23     """
24     Generates a random color in HSV space, then
```

Run

Canvas

Previous Question Next Question

Final exam question from last year

Meant to be play!

https://chrispiech.github.io/probabilityForComputerScientists/en/examples/algorithmic_art/

Algorithmic Art

chrispiech.github.io/probabilityForComputerScientists/en/examples/algorithmic_art/

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
 - Serendipity
 - Bacteria Evolution

Part 2: Random Variables

- Random Variables
- Probability Mass Functions
- Expectation
- Variance
- Bernoulli Distribution



Regenerate

In our art, the circles are different sizes. Specifically, each circle's **radius** is drawn from a Pareto distribution (which is described below). The placement algorithm is greedy: we sample 1000 circle sizes. Sort them by size, largest to smallest. Loop over the circle sizes and place circles one by one.

To place a circle on the canvas, we sample the location of the center of the circle. Both the x and y coordinates are uniformly distributed over the dimensions of the canvas. Once we have selected a prospective location we then check if there would be a collision with a circle that has already been placed. If there is a collision we keep trying new locations until you find one that has no collisions.

The Pareto Distribution

Pareto Random Variable

Notation: $X \sim \text{Pareto}(a)$

Description: A long tail distribution. Large values are rare and small values are common.

Parameters: $a \geq 1$, the shape parameter
Note there are other optional params. See [wikipedia](#)

Support: $x \in [0, 1]$

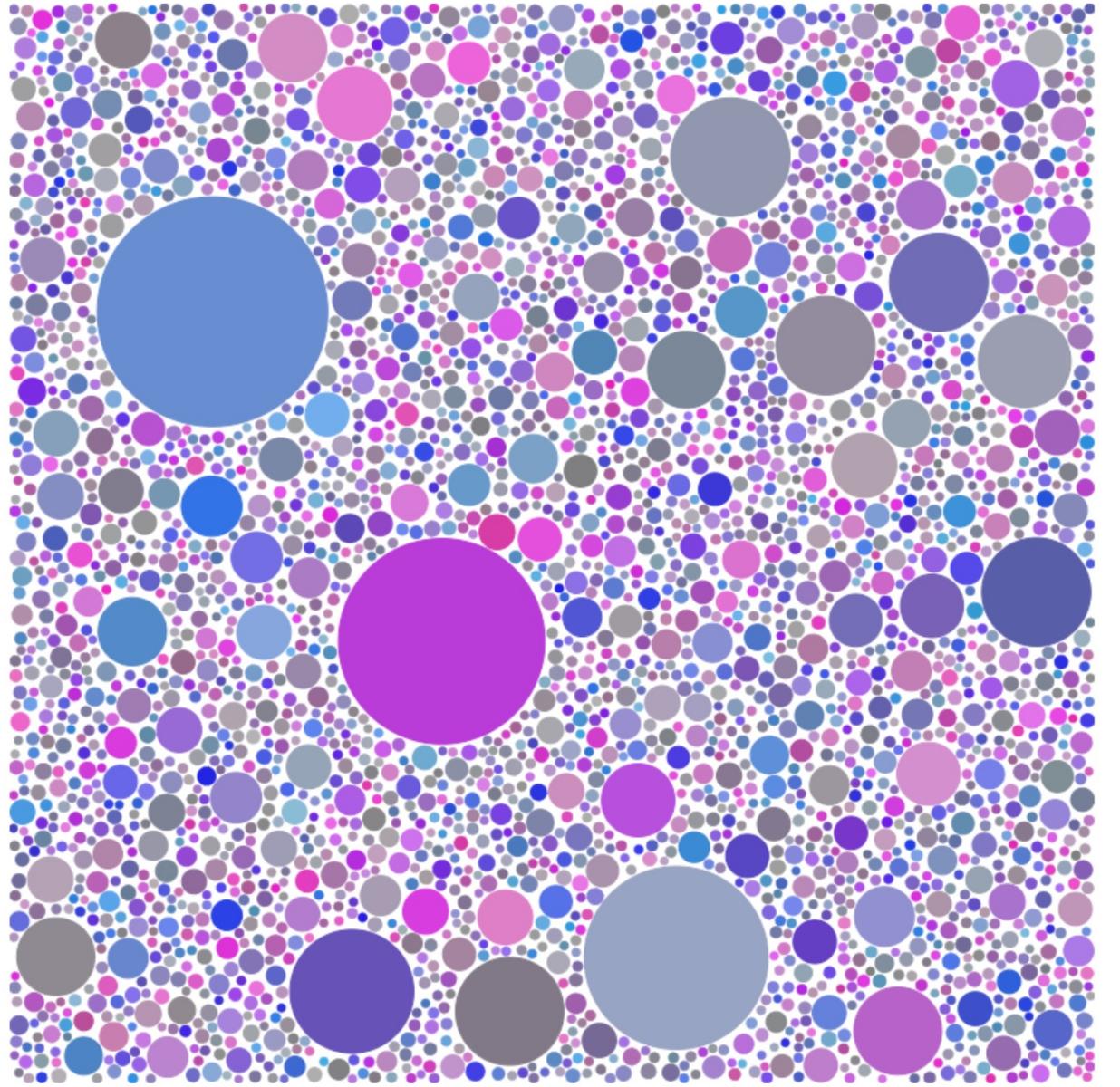
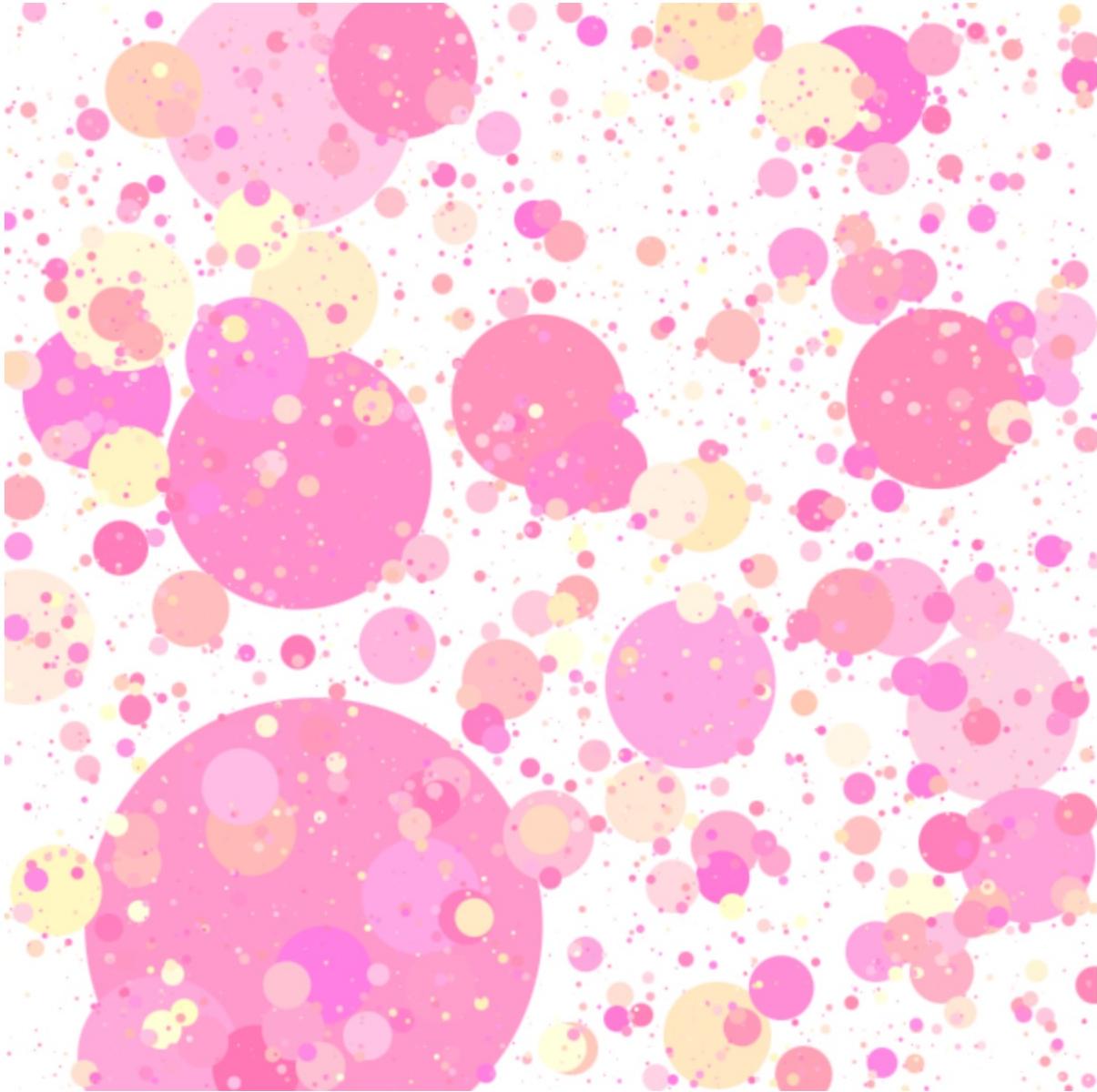
PDF equation: $f(x) = \frac{1}{x^{a+1}}$

CDF equation: $F(x) = 1 - \frac{1}{x^a}$

Sampling from a Pareto Distribution

How can we draw samples from a pareto? In python its simple: `stats.pareto.rvs(a)` however in JavaScript, or other languages, it might not be made transparent

$$y = 1 - \left(\frac{\beta}{x}\right)^\alpha$$
$$\left(\frac{\beta}{x}\right)^\alpha = 1 - y$$

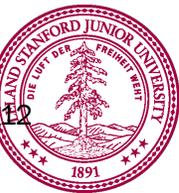
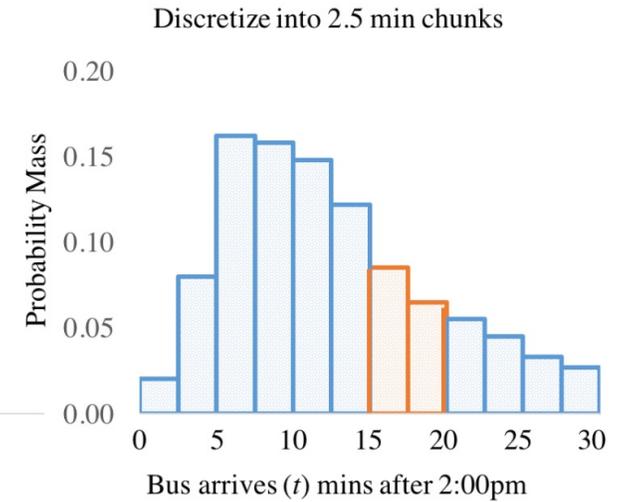
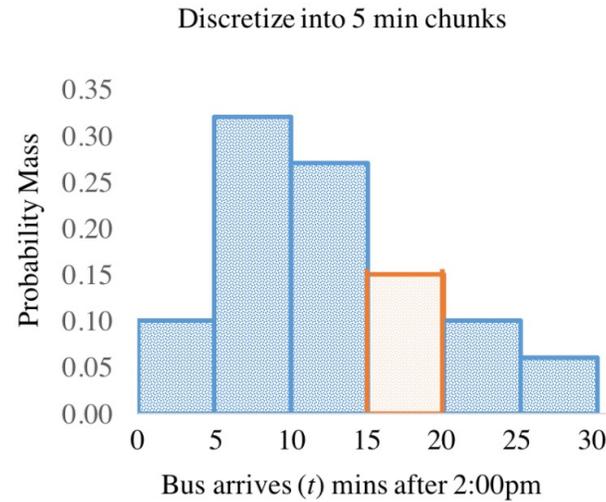
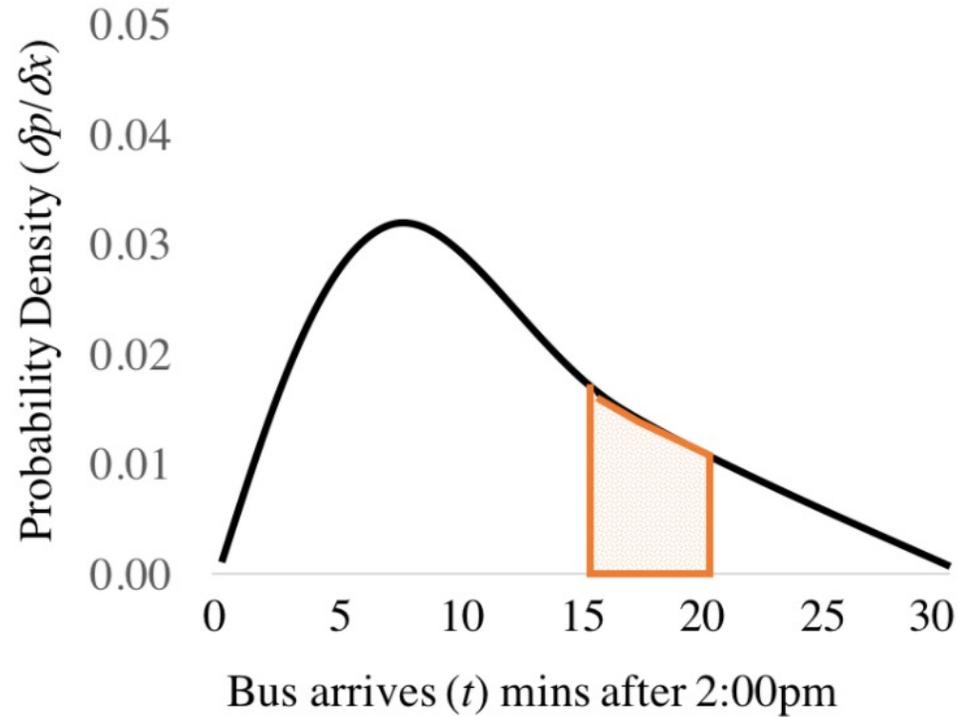


Please share your art on **Ed**

Review

Review: Probability Density Function

The limit at discretization size $\rightarrow 0$



Uniform distribution

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: $f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{for } x \in [\alpha, \beta] \\ 0 & \text{else} \end{cases}$

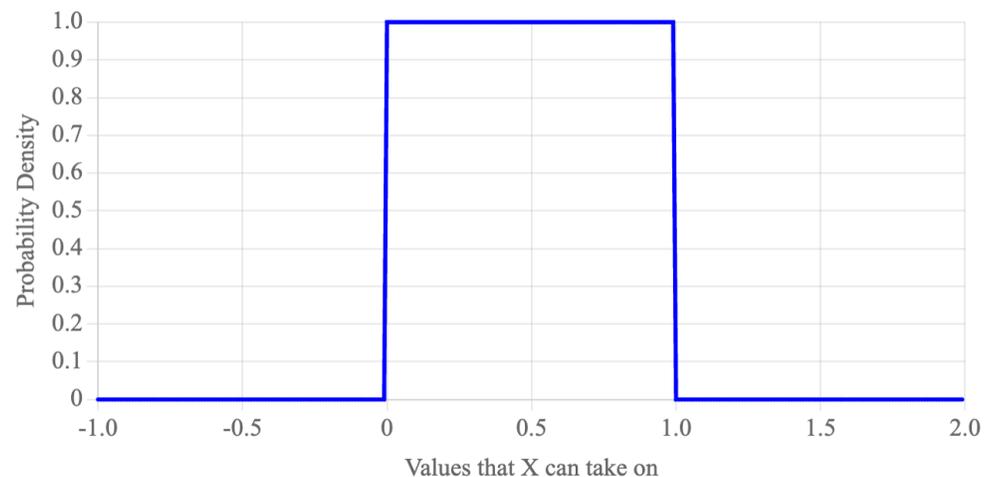
CDF equation: $F(x) = \begin{cases} \frac{x - \alpha}{\beta - \alpha} & \text{for } x \in [\alpha, \beta] \\ 0 & \text{for } x < \alpha \\ 1 & \text{for } x > \beta \end{cases}$

Expectation: $E[X] = \frac{1}{2}(\alpha + \beta)$

Variance: $\text{Var}(X) = \frac{1}{12}(\beta - \alpha)^2$

PDF graph:

Parameter α : Parameter β :



<https://chrispiech.github.io/probabilityForComputerScientists/en/part2/uniform/>



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!

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



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

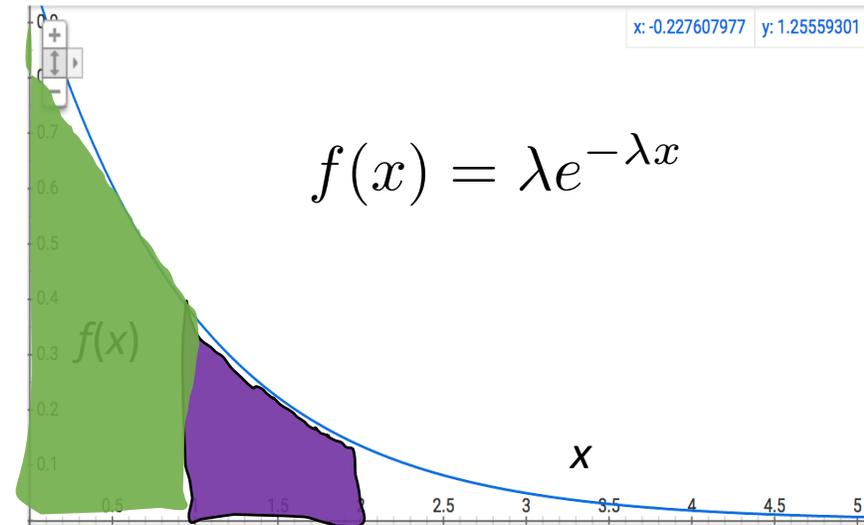
$$F(x) = P(X < x)$$



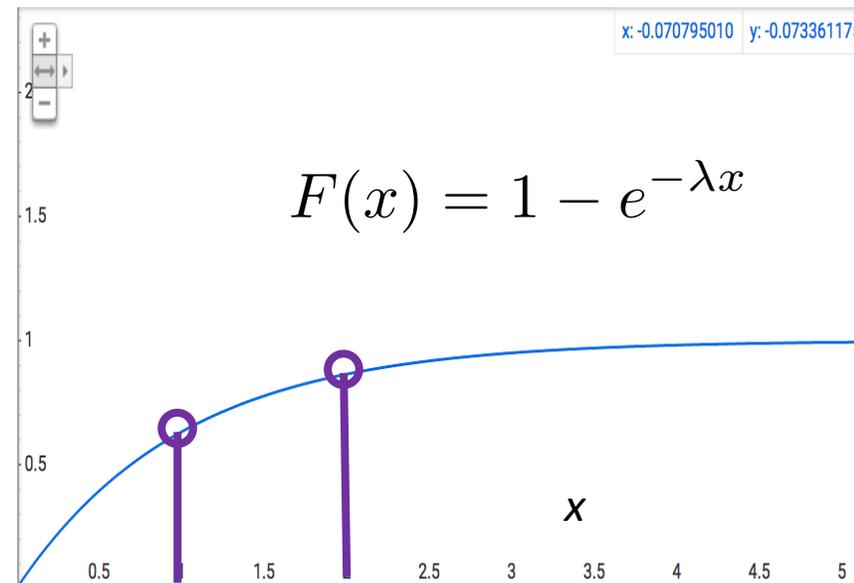
If you learn how to use a cumulative density function, you can avoid integrals!

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

Probability density function



Cumulative density function



$$F_X(x) = P(X < x)$$
$$= \int_{y=-\infty}^x f(y) dy$$

$$P(1 < X < 2)$$

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

or

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

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

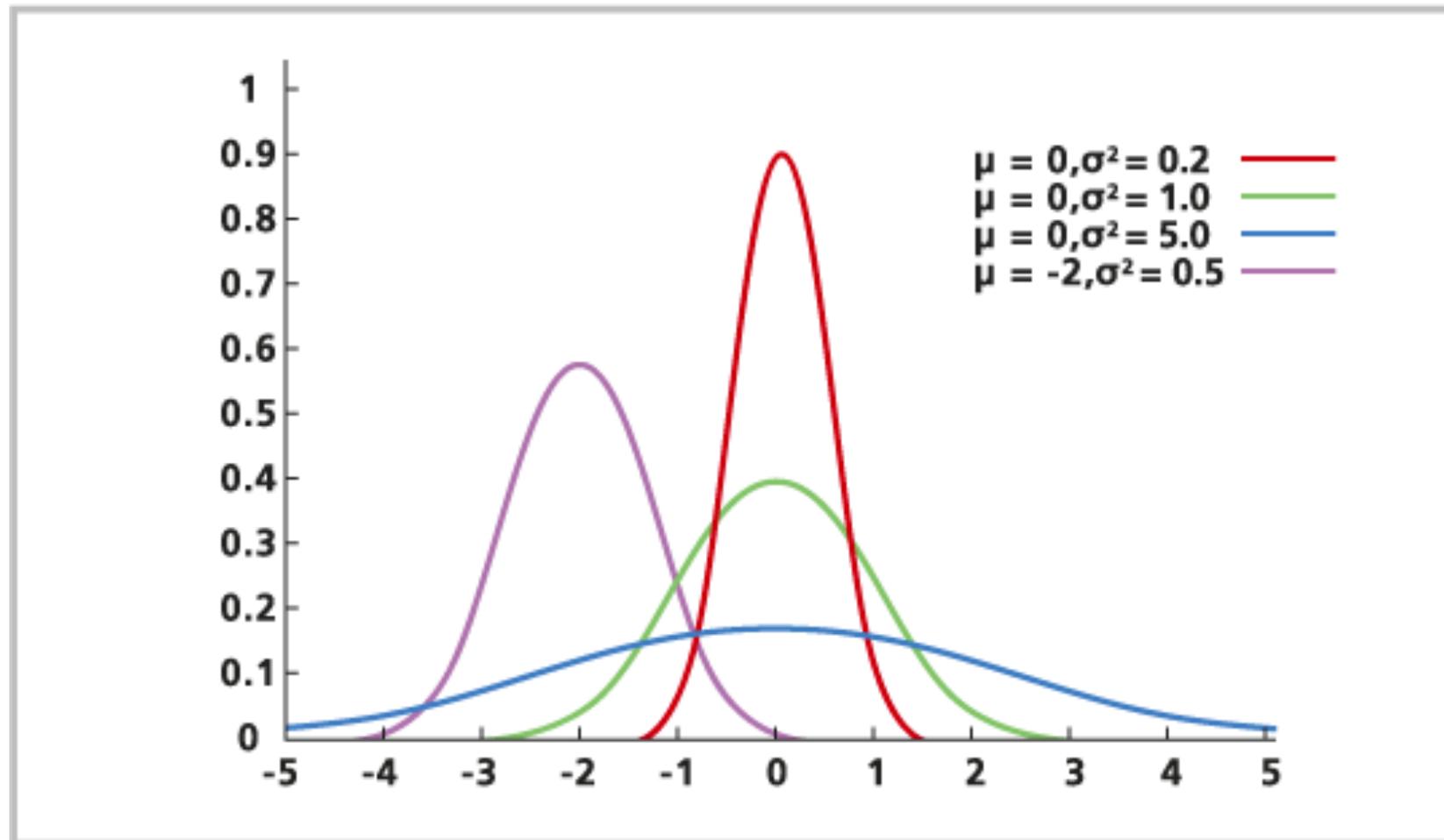
$$\approx 0.23$$



/Review

Big Day

NormCore: A Few Normal Examples



Normal Random Variable

def 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

$$E[X] = \mu$$

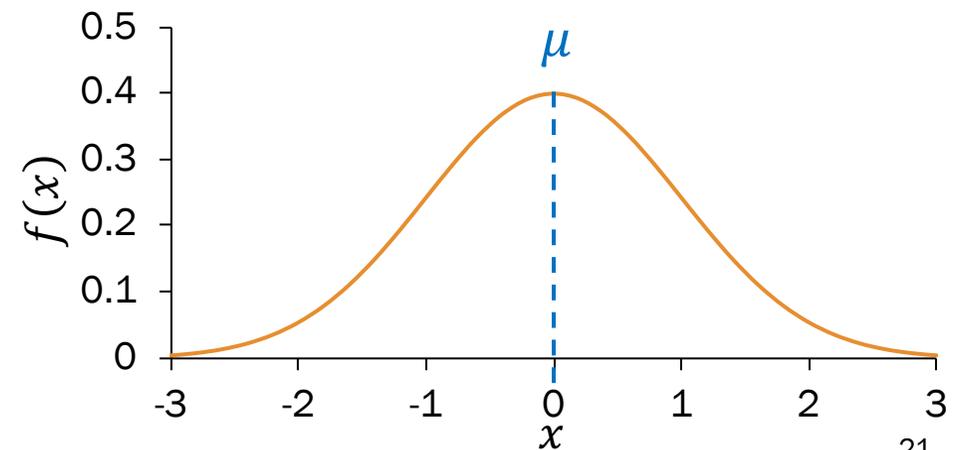
Variance

$$\text{Var}(X) = \sigma^2$$

Other names: **Gaussian** random variable

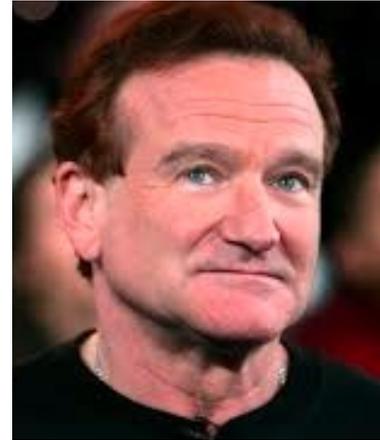
$$X \sim \mathcal{N}(\mu, \sigma^2)$$

mean variance



Carl Friedrich Gauss

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



Johann Carl Friedrich Gauss ([/ˈɡɑːs/](#); **German:** *Gauß* [\[ɡaʊ̯s\]](#) ([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's what they want you to believe...



“The simplest explanation is usually the best one”



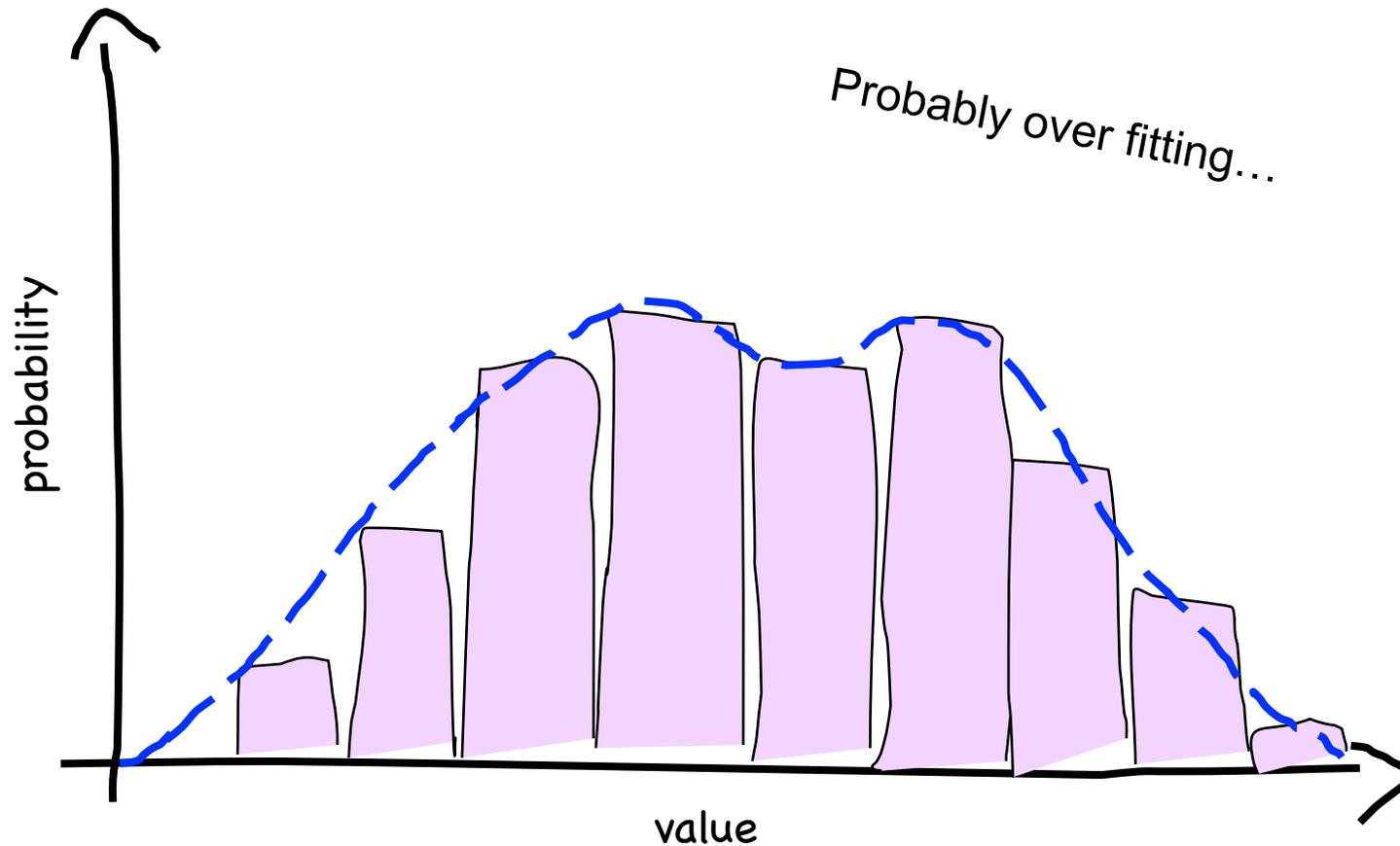
Ockham's razor

Shaving your hypothesis since 14th Century

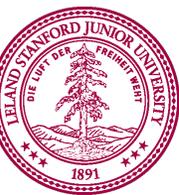


AMAZING!

Complexity is Tempting

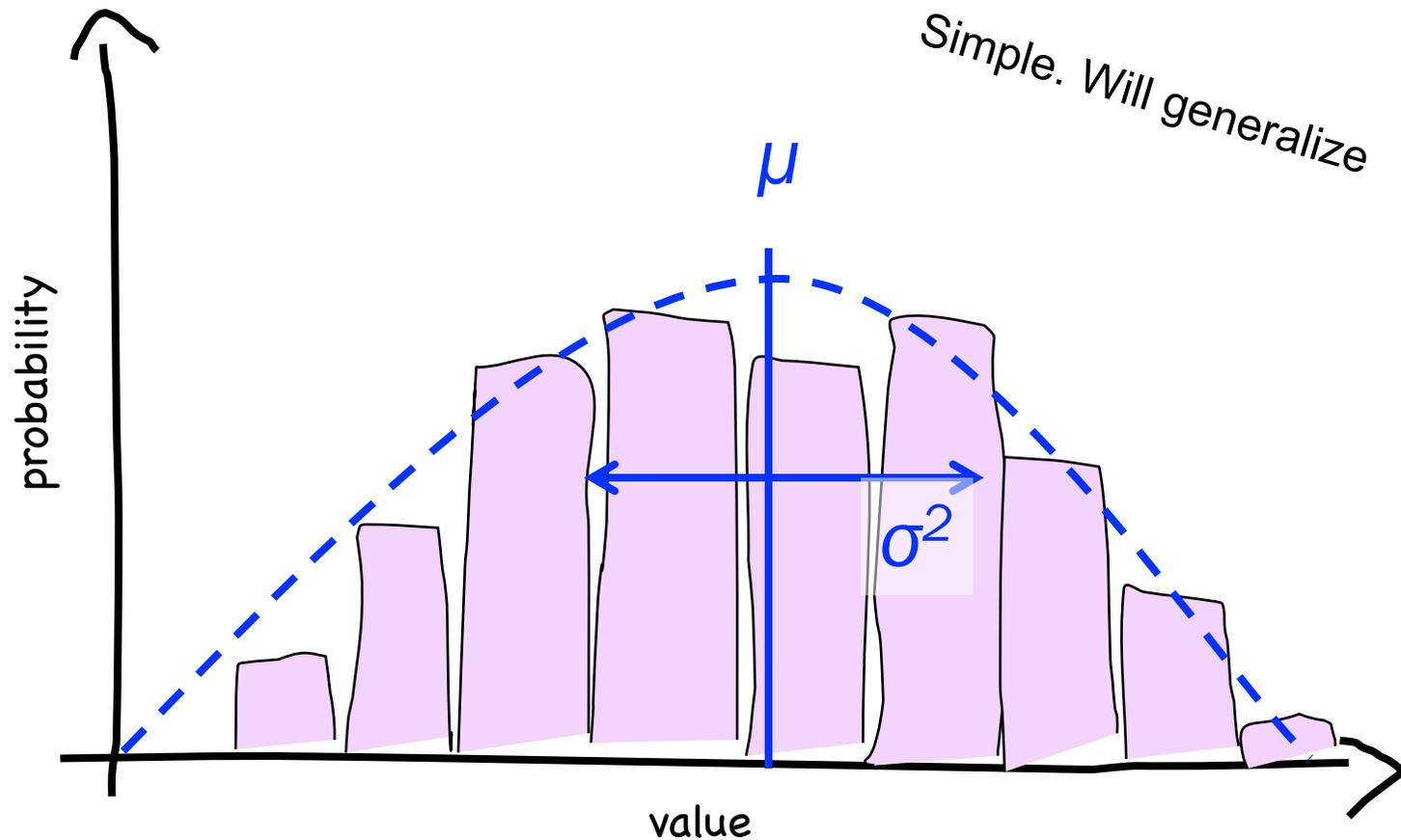


* That describes the training data, but will it **generalize**?



Fewest Assumptions

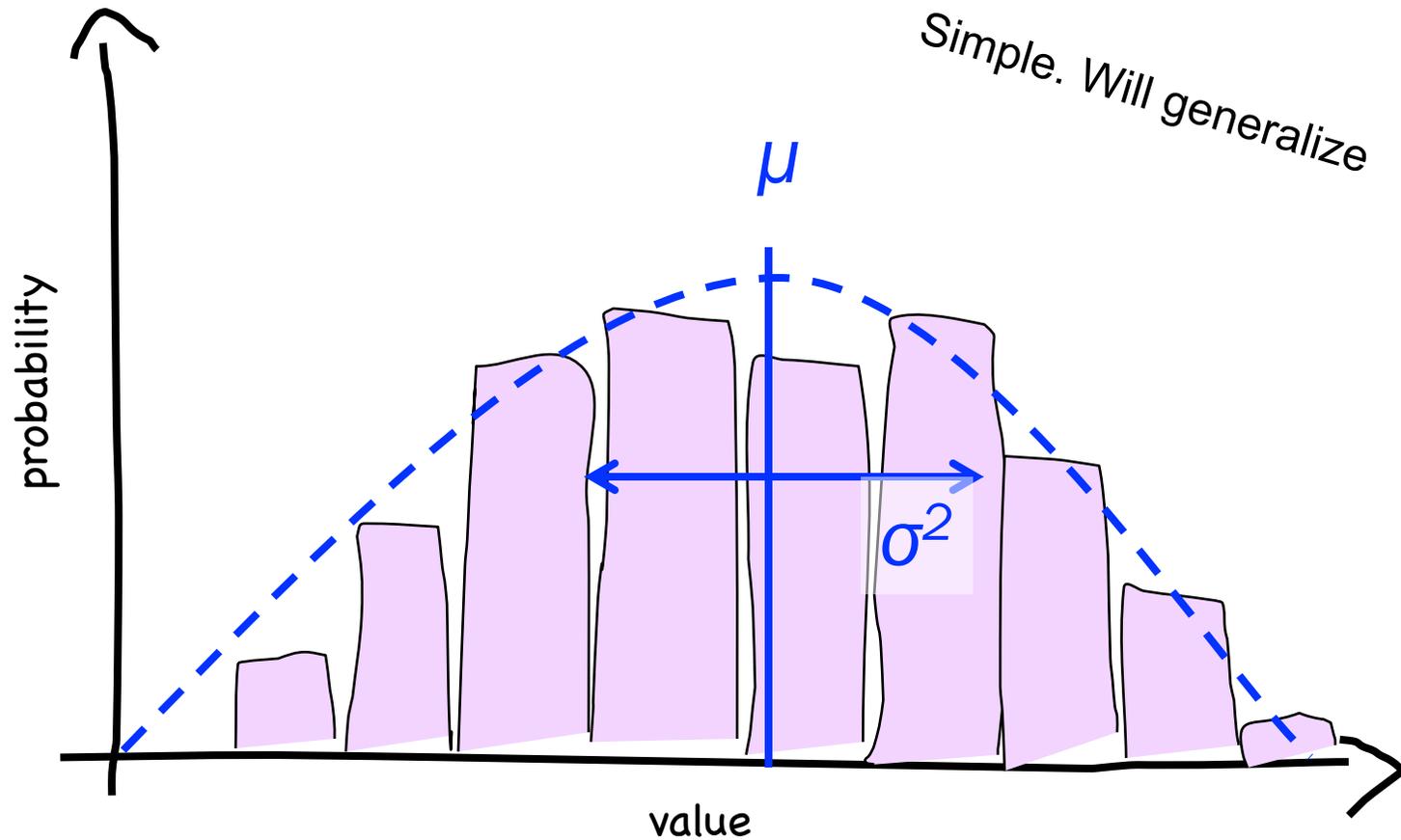
$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i)$$



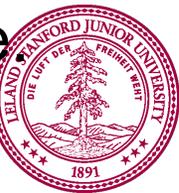
* A Gaussian **maximizes entropy** for a given mean and variance



Fewest Assumptions



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

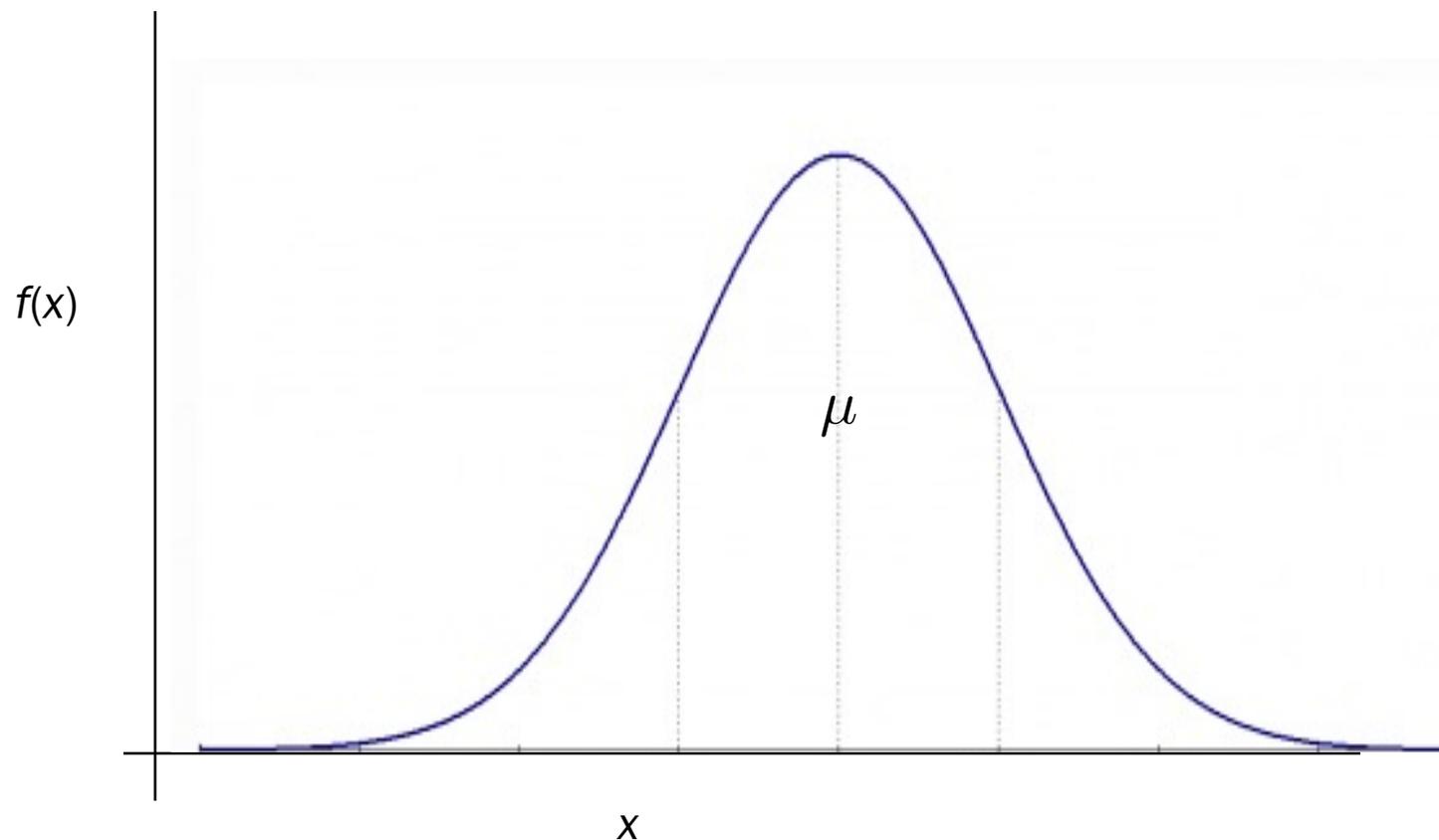


But also, it is easy to use

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

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

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

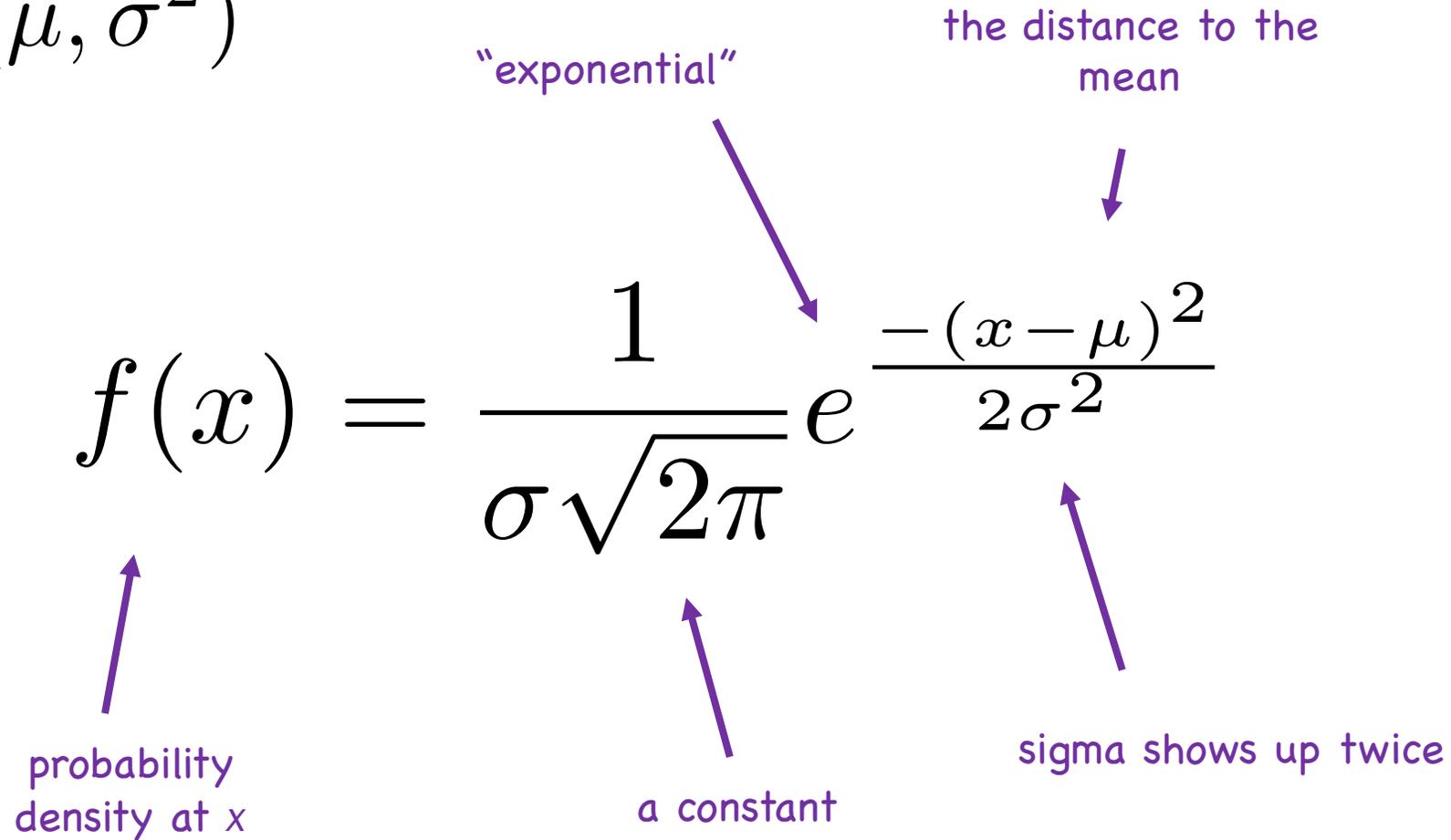
“exponential”

the distance to the mean

probability density at x

a constant

sigma shows up twice





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?



Campus bikes

You spend some minutes, X , traveling between classes.

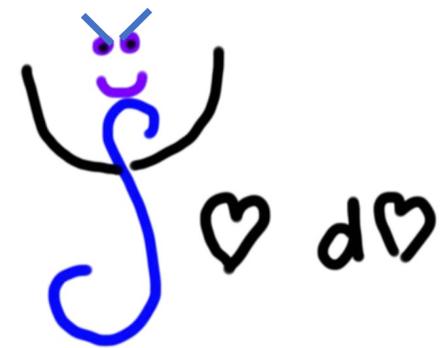
- Average time spent: $\mu = 4$ minutes
- Variance of time spent: $\sigma^2 = 2$ minutes²

Suppose X is normally distributed. What is the probability you spend ≥ 6 minutes traveling?

$$X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2)$$

$$P(X \geq 6) = \int_6^{\infty} f(x) dx = \int_6^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

(call me if you analytically solve this)



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 Normal is Normal

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

If $Y = aX + b$ then Y is also Normal

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

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

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



Special Linear Transform

If $Y = aX + b$ then Y 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 :

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

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

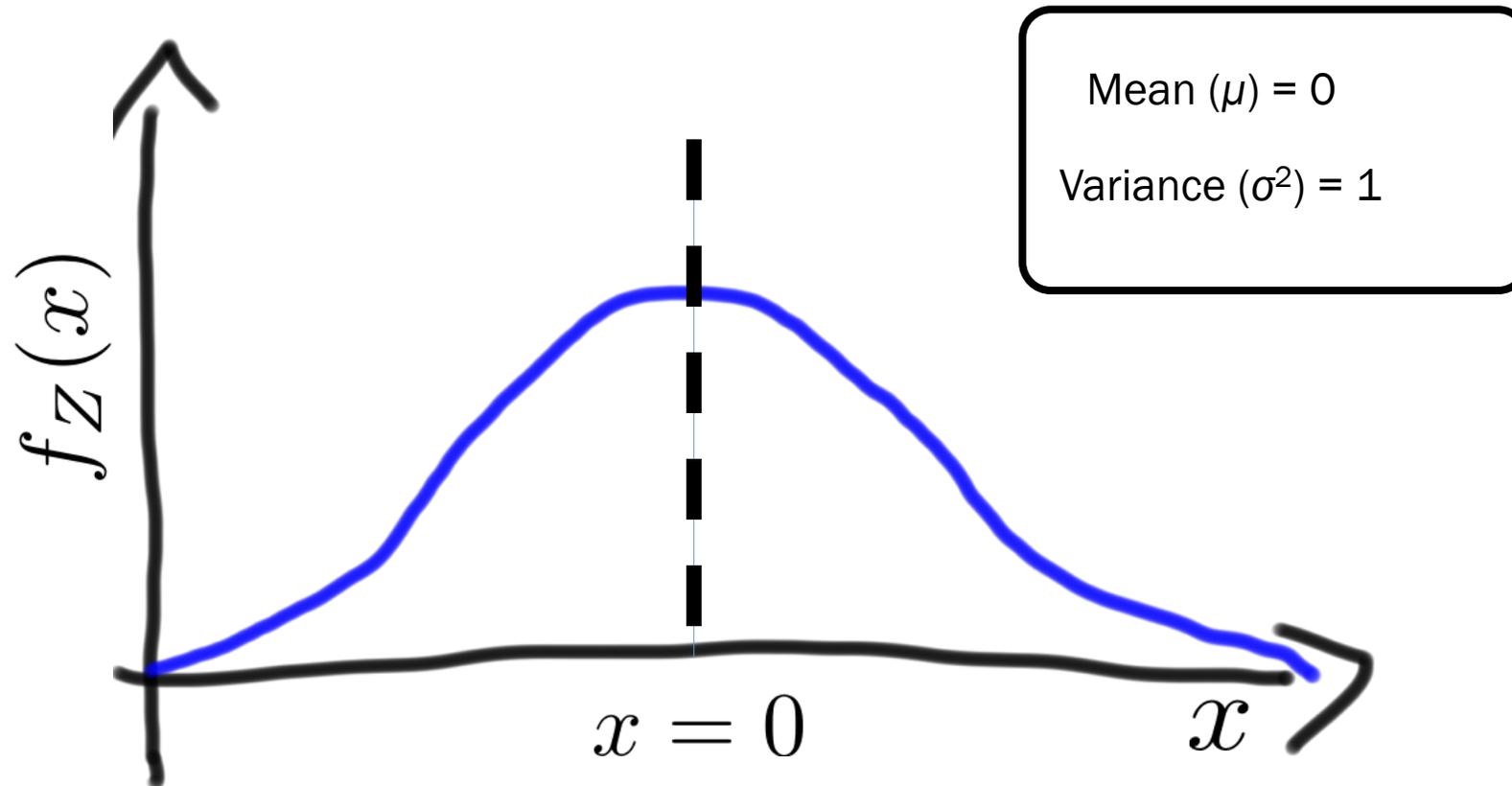
$$\sim \mathcal{N}\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}\right)$$

$$\sim \mathcal{N}(0, 1)$$



The Standard Normal

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



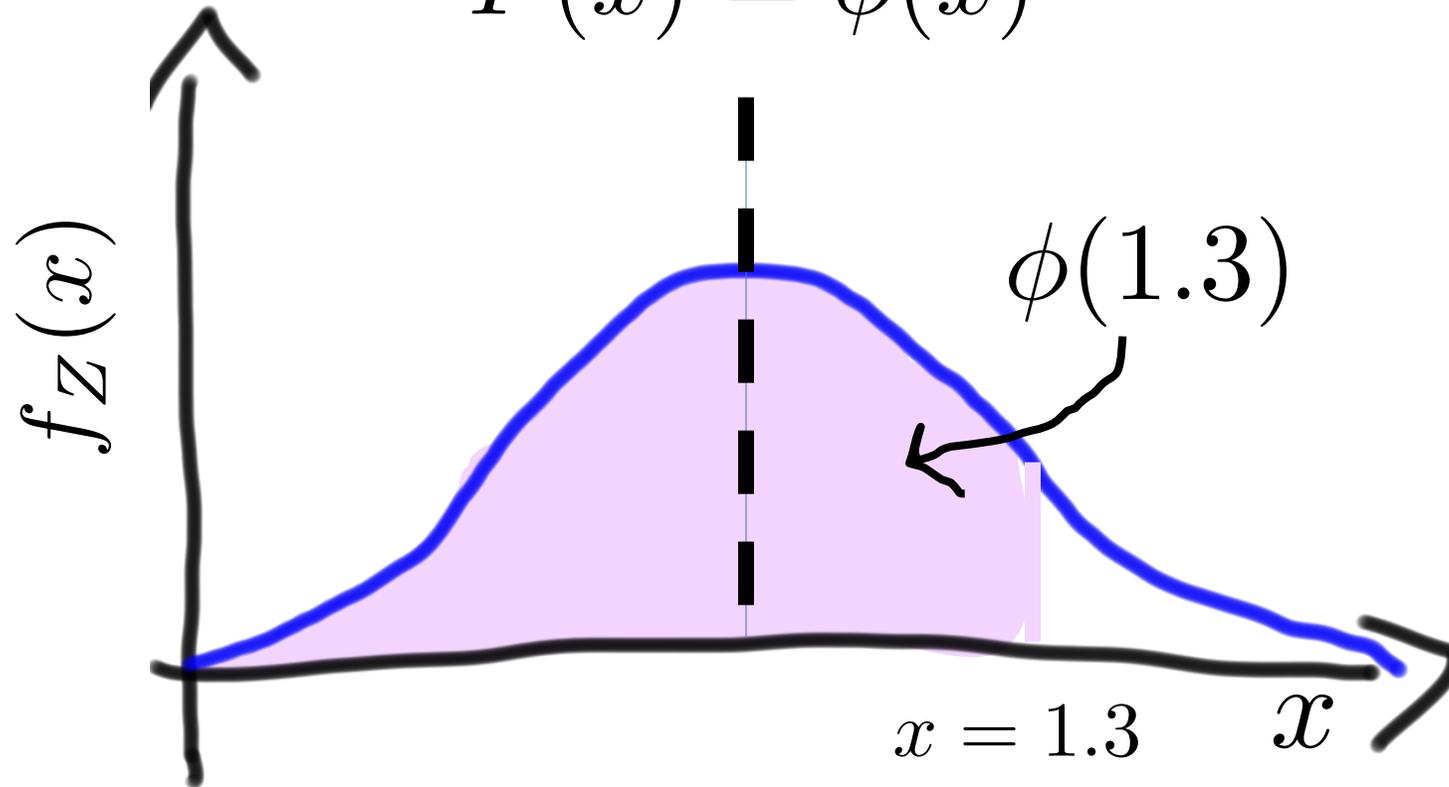
*This is the probability density function for the standard normal



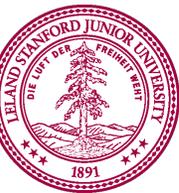
Phi

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

$$F(x) = \Phi(x)$$



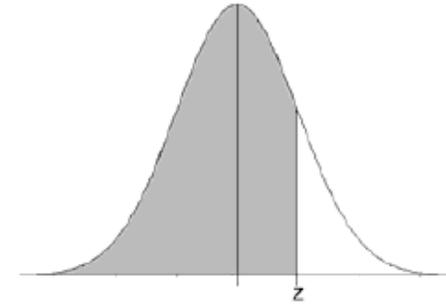
*This is the probability density function for the standard normal



Using Table of Φ

Standard Normal Cumulative Probability Table

$$\Phi(1.31) = 0.7054$$



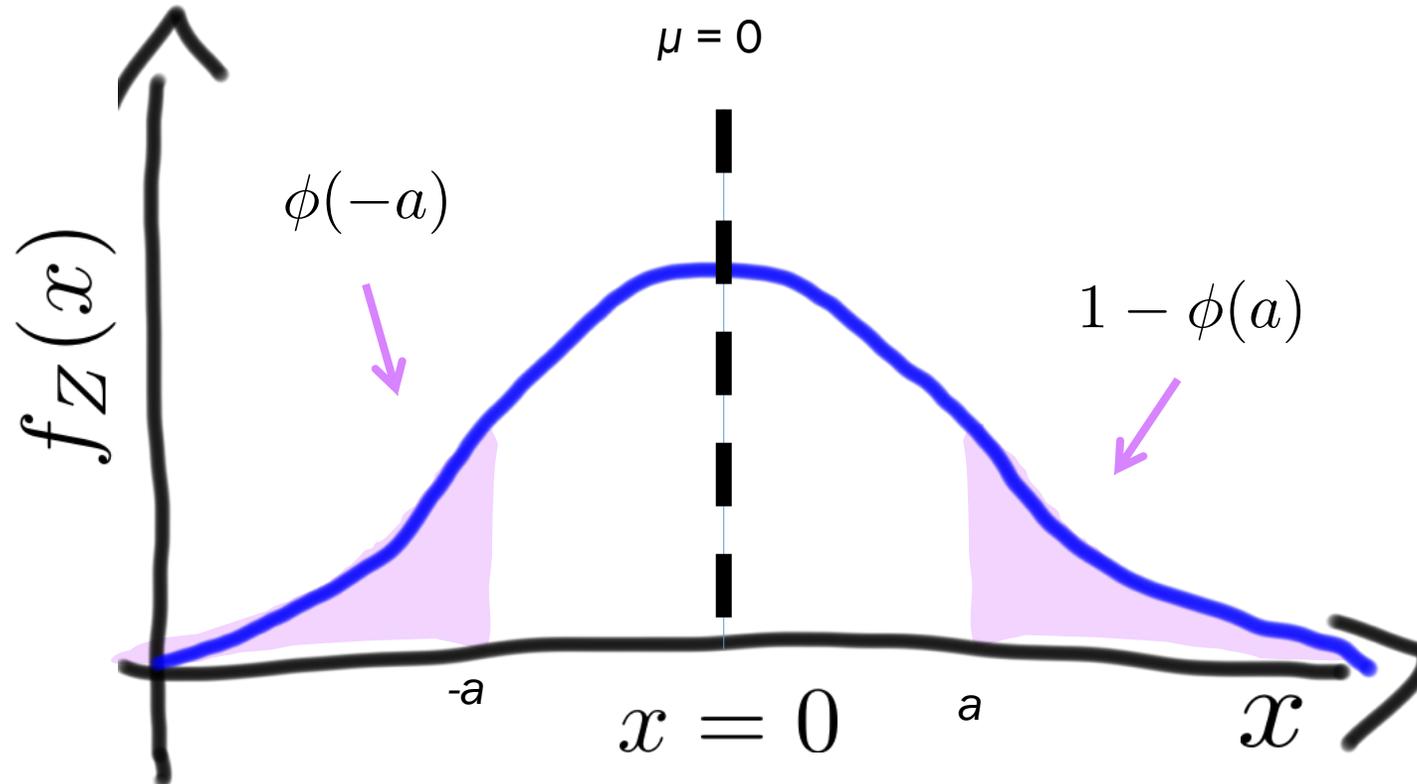
Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

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

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

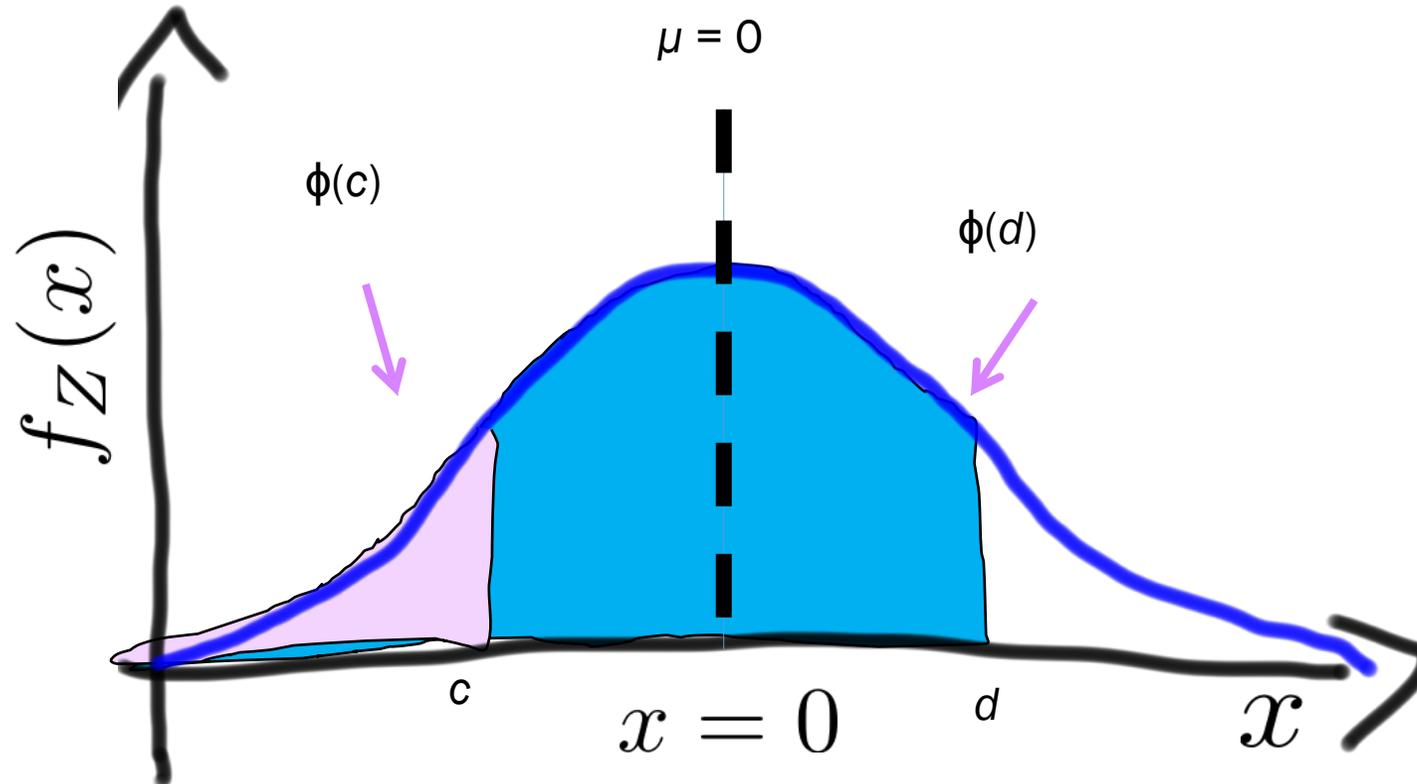


*This is the probability density function for the standard normal



Interval of Phi

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



Compute $F(x)$ via Transform

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

Use Z to compute $F(x)$

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





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

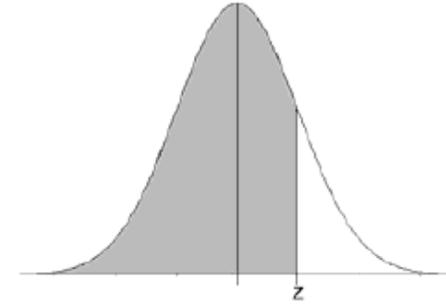
Table of $\Phi(Z)$ values in textbook, p. 201 and handout



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



Table is kinda old school



Using Programming Library

Every modern programming language has a normal library

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

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

* This is from Python's scipy library



I made one for you

The screenshot shows a web browser window with the address bar displaying `chrispiech.github.io/probabilityForComputerScientists/en/intro/calculators/`. The page features a dark sidebar on the left for the 'Course Reader for CS109'. The sidebar includes a search bar and a list of topics under 'Part 1: Core Probability' and 'Part 2: Random Variables'. The main content area contains four calculator tools, each with input fields and a blue button to execute the calculation.

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
 - Serendipity
 - Bacteria Evolution

Part 2: Random Variables

- Random Variables
- Probability Mass Functions
- Expectation
- Variance
- Bernoulli Distribution

Phi Calculator, $\Phi(x)$

x

`phi(x)`

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

y

`inverse_phi(y)`

Norm CDF Calculator

x

mu

std

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

Beta CDF Calculator

x

a

b



Campus bikes

You spend some minutes, X , traveling between classes.

- Average time spent: $\mu = 4$ minutes
- Variance of time spent: $\sigma^2 = 2$ minutes²

Suppose X is normally distributed. What is the probability you spend ≥ 6 minutes traveling?



$$X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2) \quad \times \quad P(X \geq 6) = \int_6^{\infty} f(x) dx \quad (\text{no analytic solution})$$

1. Compute $z = \frac{(x-\mu)}{\sigma}$

$$\begin{aligned} P(X \geq 6) &= 1 - F_x(6) \\ &= 1 - \Phi\left(\frac{6-4}{\sqrt{2}}\right) \\ &\approx 1 - \Phi(1.41) \end{aligned}$$

2. Look up $\Phi(z)$ in table

$$\begin{aligned} &1 - \Phi(1.41) \\ &\approx 1 - 0.9207 \\ &= 0.0793 \end{aligned}$$



Get your Gaussian On

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

1. $P(X > 0)$

- 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
 $\Phi(-z) = 1 - \Phi(z)$



Get your Gaussian On

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

Note standard deviation $\sigma = 4$.

How would you write each of the below probabilities as a function of the standard normal CDF, Φ ?

1. $P(X > 0)$ (we just did this)
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 $\Phi(-z) = 1 - \Phi(z)$



Get your Gaussian On

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

1. $P(X > 0)$
2. $P(2 < X < 5)$

- 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 $\Phi(-z) = 1 - \Phi(z)$



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. $P(|X - 3| > 6)$

Compute $z = \frac{(x-\mu)}{\sigma}$

$$P(X < -3) + P(X > 9)$$

$$= F(-3) + (1 - F(9))$$

$$= \Phi\left(\frac{-3-3}{4}\right) + \left(1 - \Phi\left(\frac{9-3}{4}\right)\right)$$

- 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 $\Phi(-x) = 1 - \Phi(x)$

Look up $\Phi(z)$ in table



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. $P(|X - 3| > 6)$

Compute $z = \frac{(x-\mu)}{\sigma}$

$$P(X < -3) + P(X > 9)$$

$$= F(-3) + (1 - F(9))$$

$$= \Phi\left(\frac{-3-3}{4}\right) + \left(1 - \Phi\left(\frac{9-3}{4}\right)\right)$$

Look up $\Phi(z)$ in table

$$= \Phi\left(-\frac{3}{2}\right) + \left(1 - \Phi\left(\frac{3}{2}\right)\right)$$

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

$$\approx 0.1337$$

- 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 $\Phi(-x) = 1 - \Phi(x)$



Imagine you are taking a quiz...
With no computer!!!

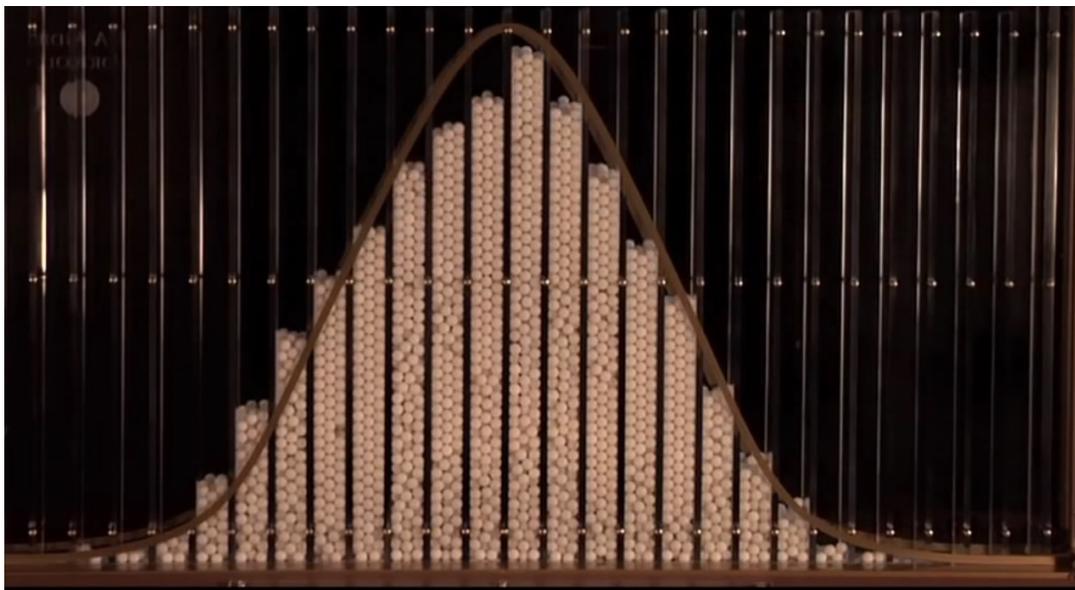
Website Testing

- 100 people are given a new website design
- $X = \#$ people whose time on site increases
 - CEO will endorse new design if $X \geq 65$ What is $P(\text{CEO endorses change} \mid \text{it has no effect})$?
 - $X \sim \text{Bin}(100, 0.5)$. Want to calculate $P(X \geq 65)$
 - Give a numerical answer...

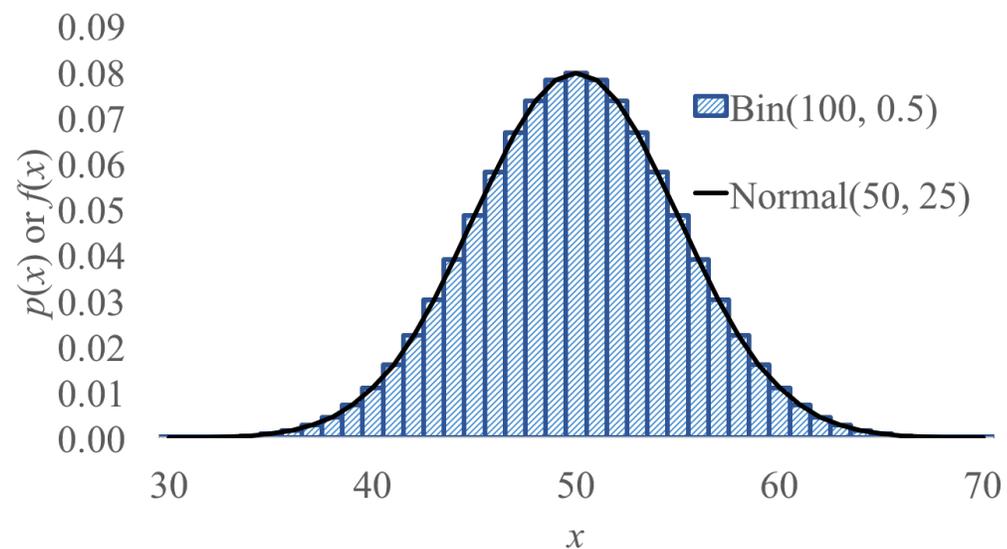
$$P(X \geq 65) = \sum_{i=65}^{100} \binom{100}{i} (0.5)^i (1 - 0.5)^{100-i}$$



Don't worry, Normal approximates Binomial



Galton Board



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



Website testing

- 100 people are given a new website design.
- $X = \#$ people whose time on site increases
- The design actually has no effect, so $P(\text{time on site increases}) = 0.5$ independently.
- CEO will endorse the new design if $X \geq 65$.

What is $P(\text{CEO endorses change})$? Give a numerical approximation.

Approach 1: Binomial

Define

$$X \sim \text{Bin}(n = 100, p = 0.5)$$

Want: $P(X \geq 65)$

Solve

$$P(X \geq 65) \approx 0.0018$$



(this approach is missing something important)

Approach 2: approximate with Normal

Define

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

Solve

$$\begin{aligned} P(X \geq 65) &\approx P(Y \geq 65) = 1 - F_Y(65) \\ &= 1 - \Phi\left(\frac{65-50}{5}\right) = 1 - \Phi(3) \approx 0.0013? \end{aligned}$$

$$\mu = np = 50$$

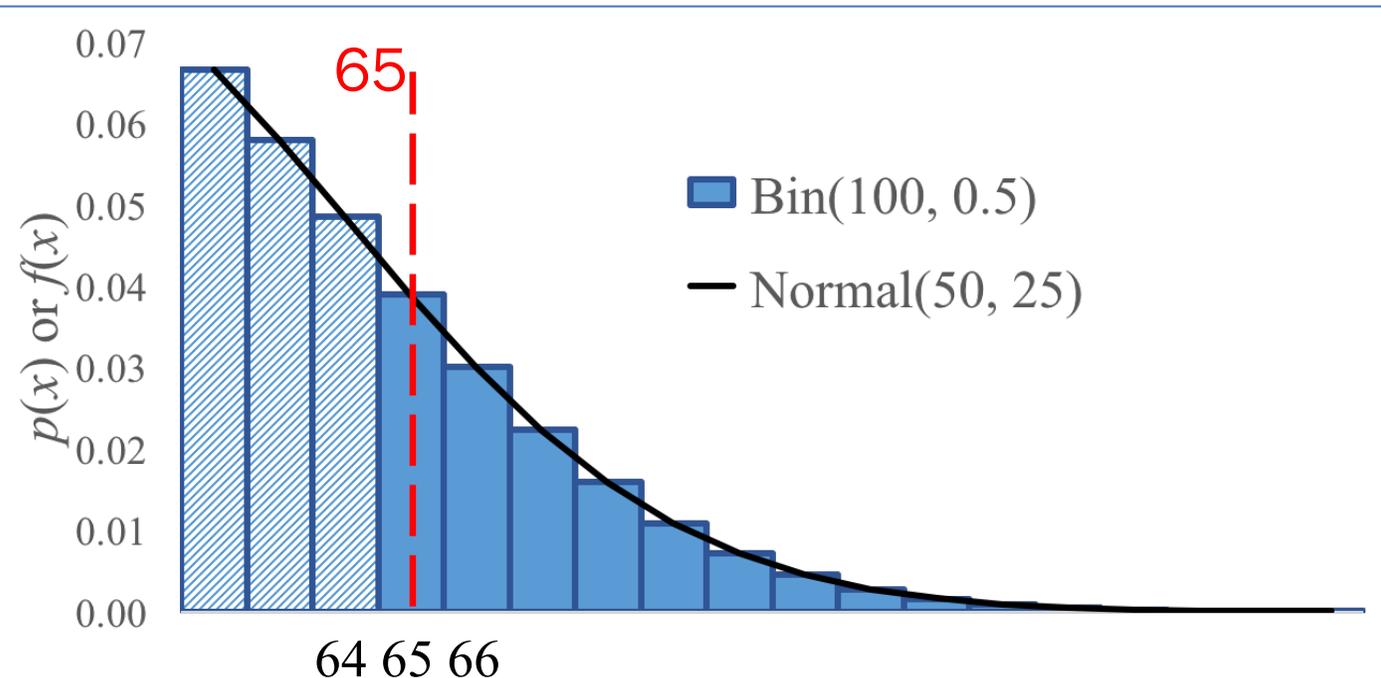
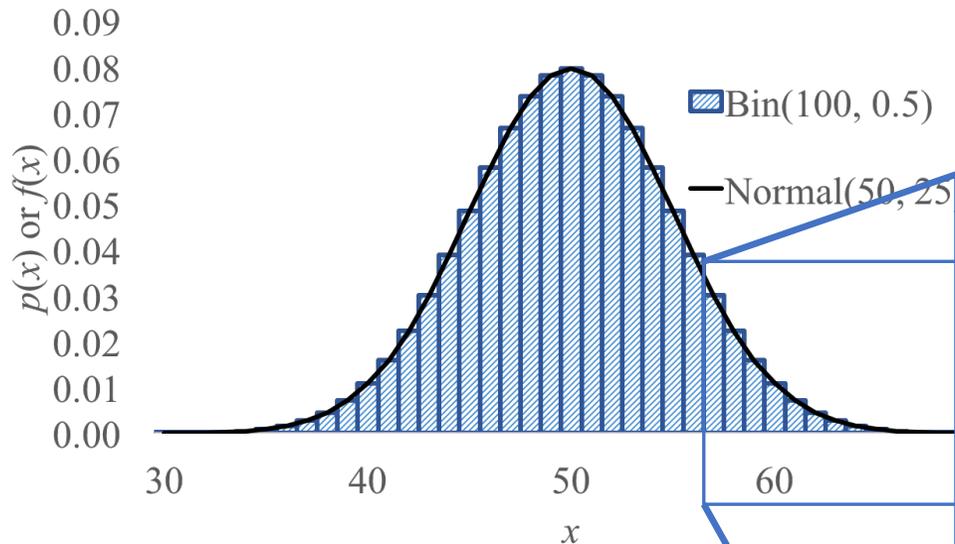
$$\sigma^2 = np(1-p) = 25$$

$$\sigma = \sqrt{25} = 5$$



Website testing (with continuity correction)

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



$$P(X \geq 65) \text{ Binomial}$$

$$\approx P(Y \geq 64.5) \text{ Normal}$$

$$\approx 0.0018 \quad \checkmark \text{ the better Approach 2}$$

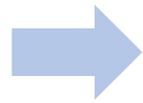
You must perform a continuity correction when approximating a Binomial RV with a Normal RV.



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

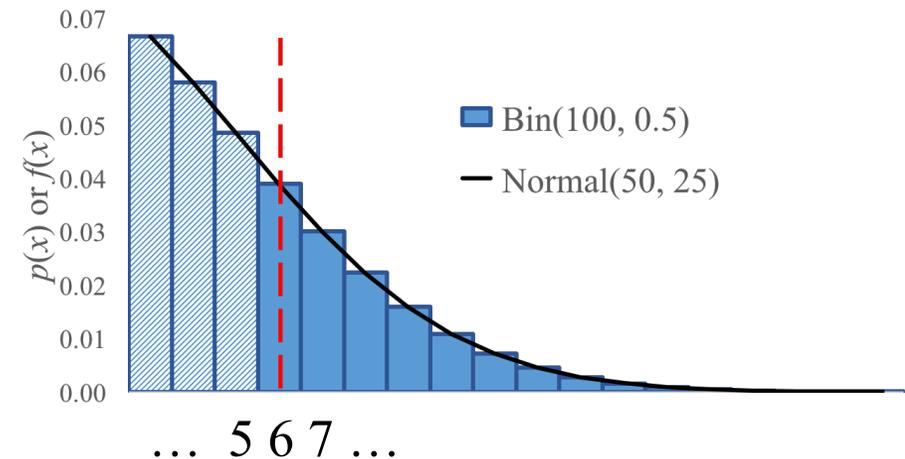
$$P(X = 6)$$

$$P(X \geq 6)$$

$$P(X > 6)$$

$$P(X < 6)$$

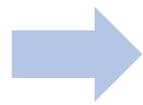
$$P(X \leq 6)$$



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

$$P(X = 6)$$

$$P(X \geq 6)$$

$$P(X > 6)$$

$$P(X < 6)$$

$$P(X \leq 6)$$

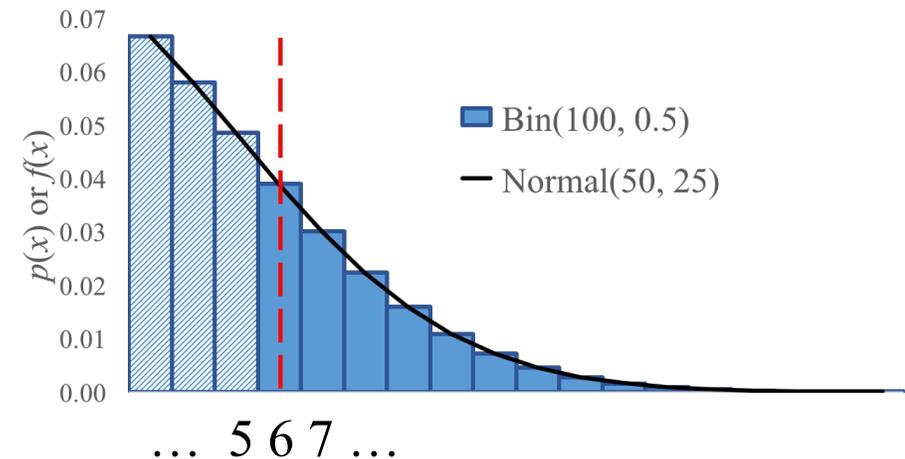
$$P(5.5 \leq Y \leq 6.5)$$

$$P(Y \geq 5.5)$$

$$P(Y \geq 6.5)$$

$$P(Y \leq 5.5)$$

$$P(Y \leq 6.5)$$



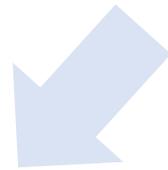
Normal Approximation

Who gets to approximate?

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

$$E[X] = np$$

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



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

$$\lambda = np$$

?



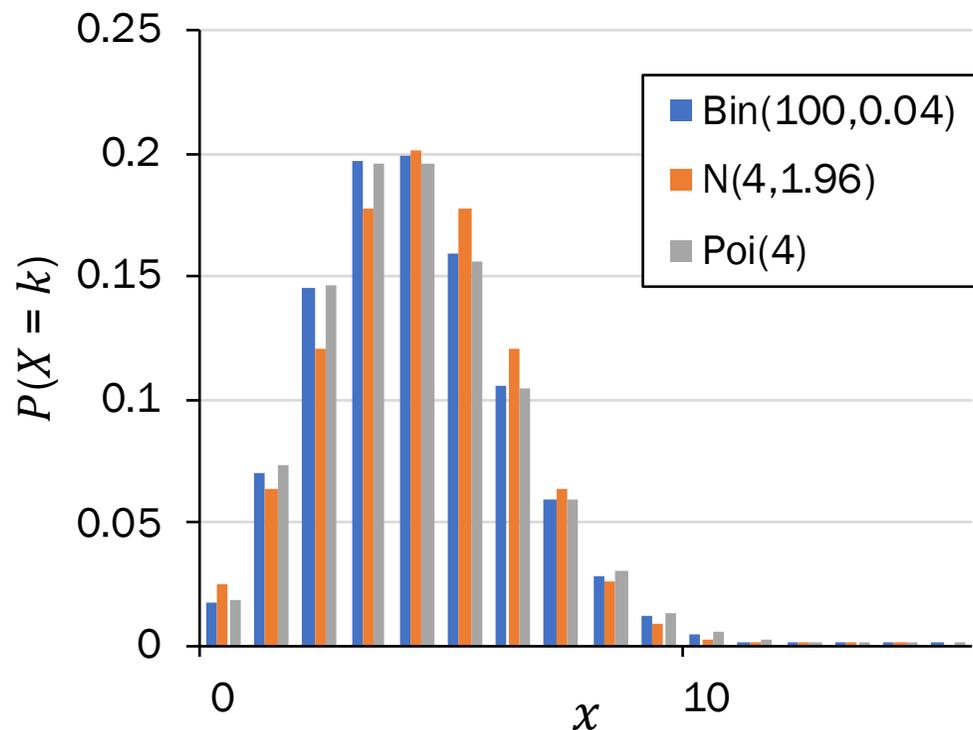
$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mu = np$$

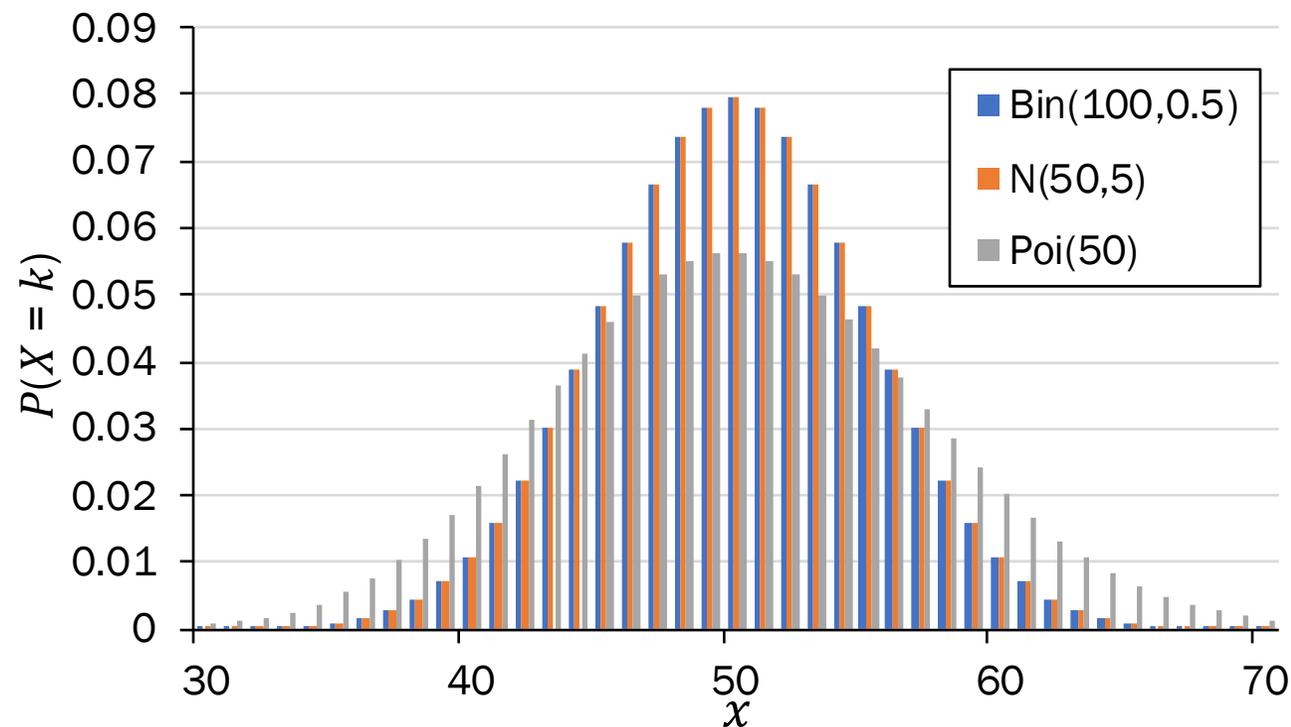
$$\sigma^2 = np(1 - p)$$



Who gets to approximate?



Poisson approximation
 n large (> 20), p small (< 0.05)
 slight dependence okay



Normal approximation
 n large (> 20), variance large ($np(1 - p) > 10$)
 independence

1. If there is a choice, either is fine.
2. When using Normal to approximate a discrete RV, use a continuity correction.



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

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

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

$$P(X > 1745) \approx P(Y \geq 1745.5) \quad \text{! Continuity correction}$$

Solve

$$\begin{aligned} P(Y \geq 1745.5) &= 1 - F(1745.5) \\ &= 1 - \Phi\left(\frac{1745.5 - 1686}{23.3}\right) \\ &= 1 - \Phi(2.54) \approx 0.0055 \end{aligned}$$

SciPy can do this



How many students should Stanford admit?

The Stanford Daily

NEWS · SPORTS · OPINIONS · ARTS & LIFE · THE GRIND · MULTIMEDIA · FEATURES · ARCHIVES

Class of 2018 admit rates lowest in University history

March 28, 2014 [16 Comments](#) [Tweet](#)

[Like 901](#)

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

Super Question:

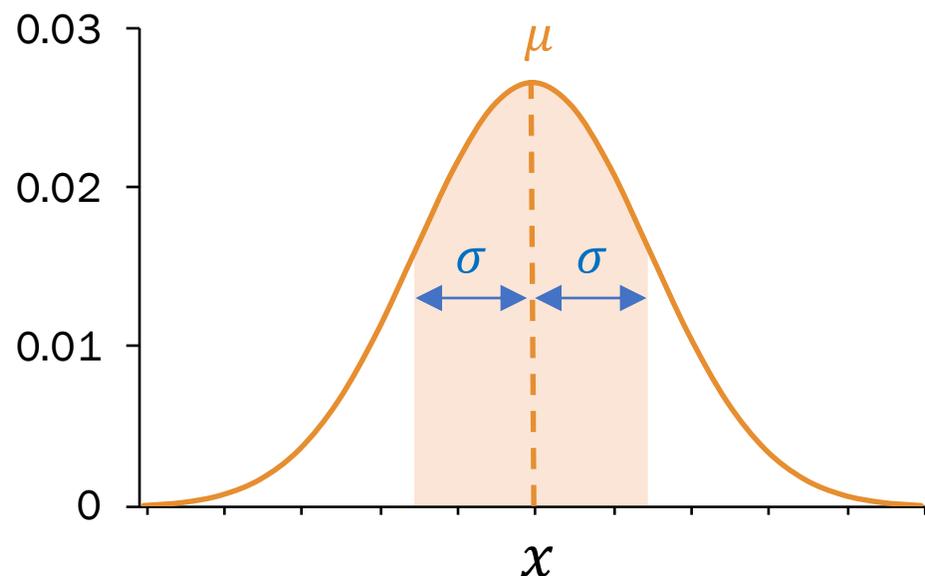
Why Be Normal? 68% rule

You may have heard the statement:

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

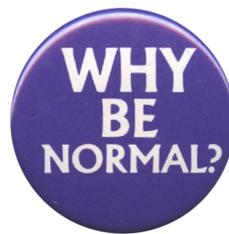
In general, this is only true of **normal distributions**:

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



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

Why Be Normal? 68% rule

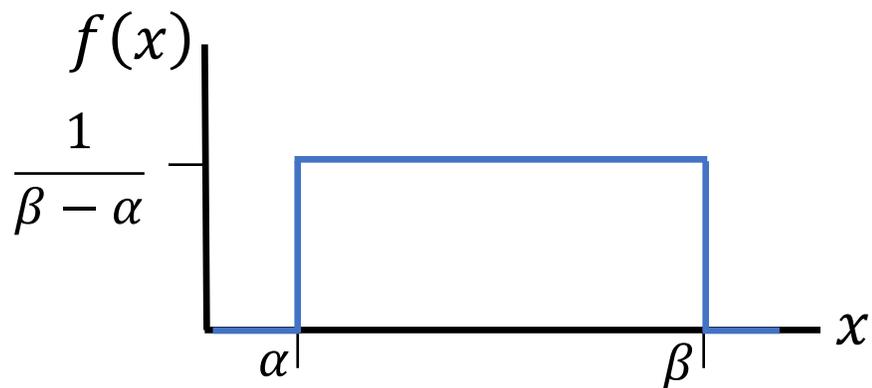


You may have heard the statement:

“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 $X \sim \text{Uni}(\alpha, \beta)$.



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

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

$$P(|X - \mu| < \sigma) = P(\mu - \sigma < X < \mu + \sigma)$$

$$= \frac{1}{\beta - \alpha} \cdot [(\mu + \sigma) - (\mu - \sigma)]$$

$$= \frac{1}{\beta - \alpha} [2\sigma] = \frac{1}{\beta - \alpha} \cdot \left[2 \cdot \frac{\beta - \alpha}{\sqrt{12}} \right]$$

$$= 2/\sqrt{12} \approx 0.58$$



Gaussian Sampling and ELO ratings

Basketball == Stats



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.

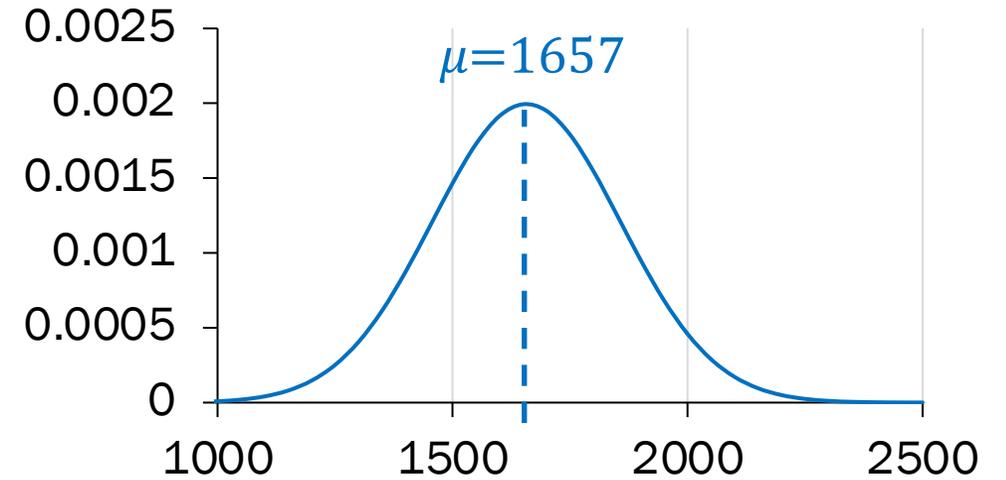


Arpad Elo

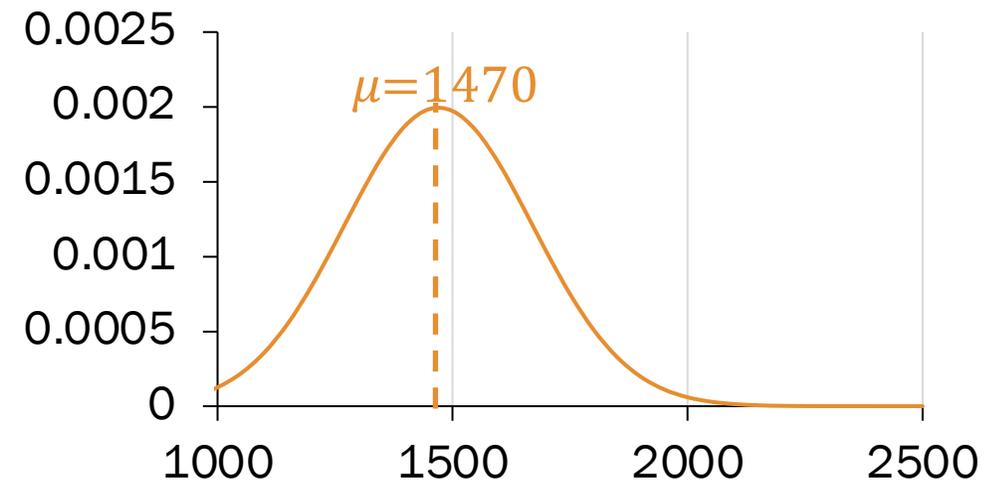
What is the probability that Warriors win this game?

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

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

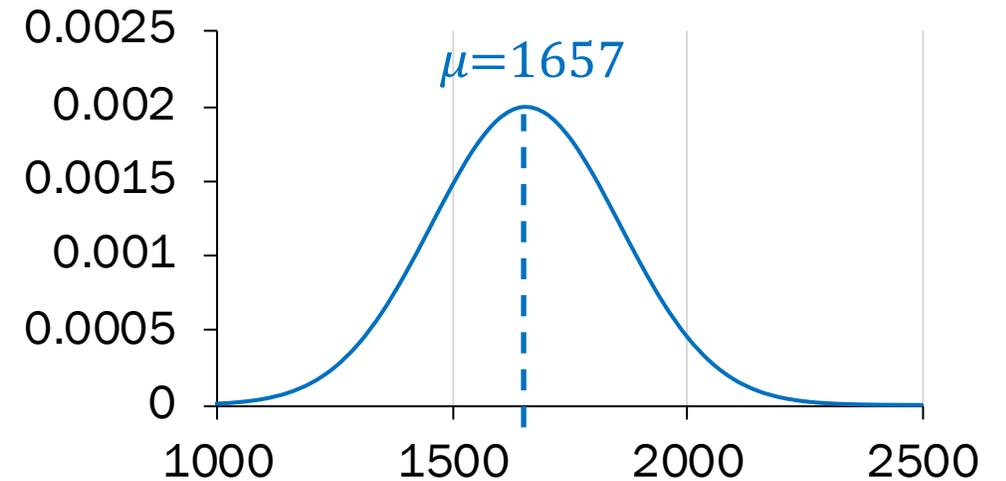
Want: $P(\text{Warriors win}) = P(A_W > A_O)$

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

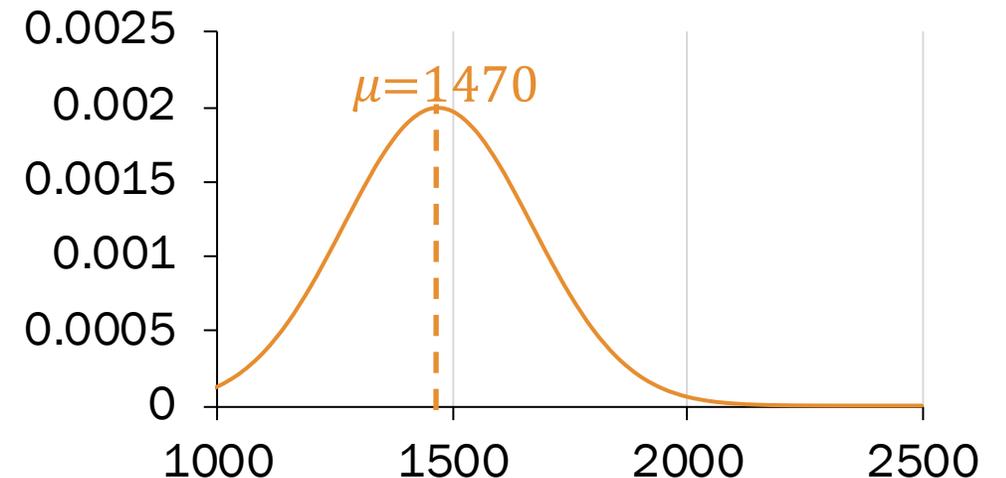
nSuccess = 0
for i in range(NTRIALS):
    w = stats.norm.rvs(WARRIORS_ELO, STDEV)
    o = stats.norm.rvs(OPPONENT_ELO, STDEV)
    if w > o:
        nSuccess += 1
print("Warriors sampled win fraction: ",
      float(nSuccess) / NTRIALS)
```

≈ 0.7488 , calculated by sampling

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



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



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!



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

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
```

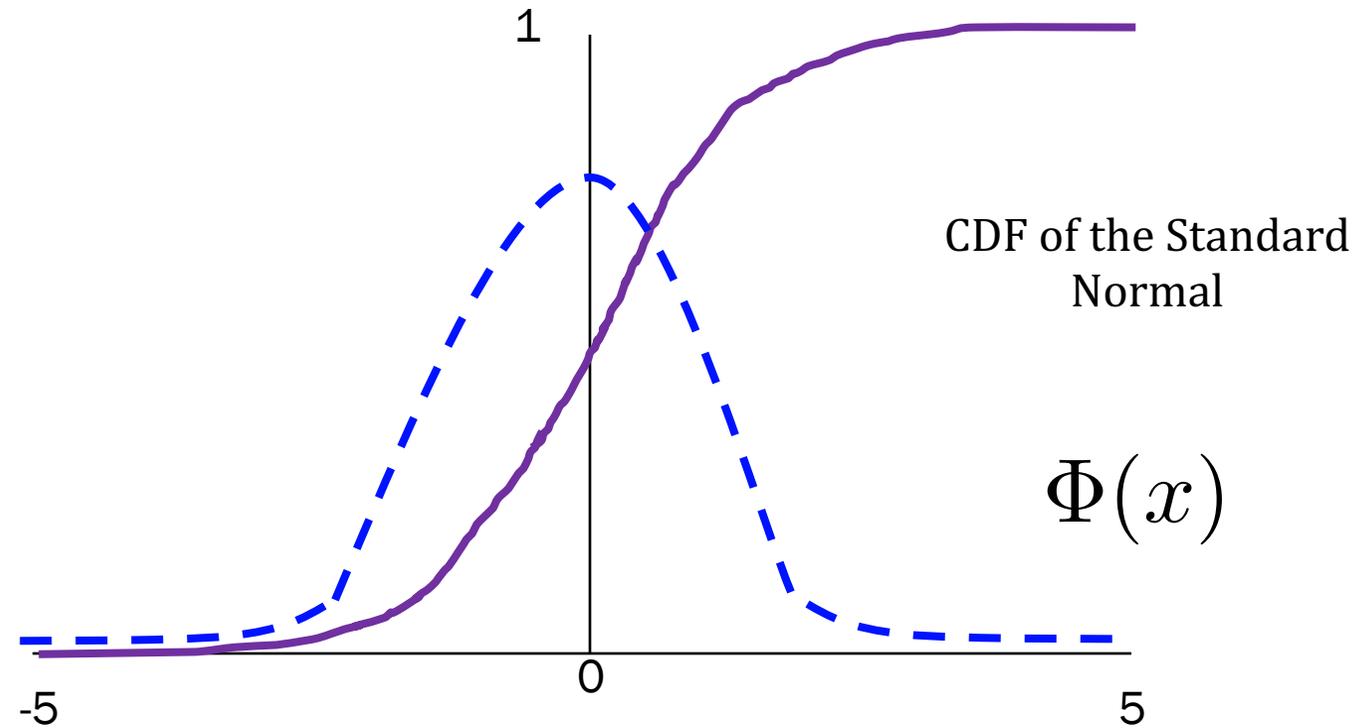
```
5.51485158841
```

```
4.94570606131
```

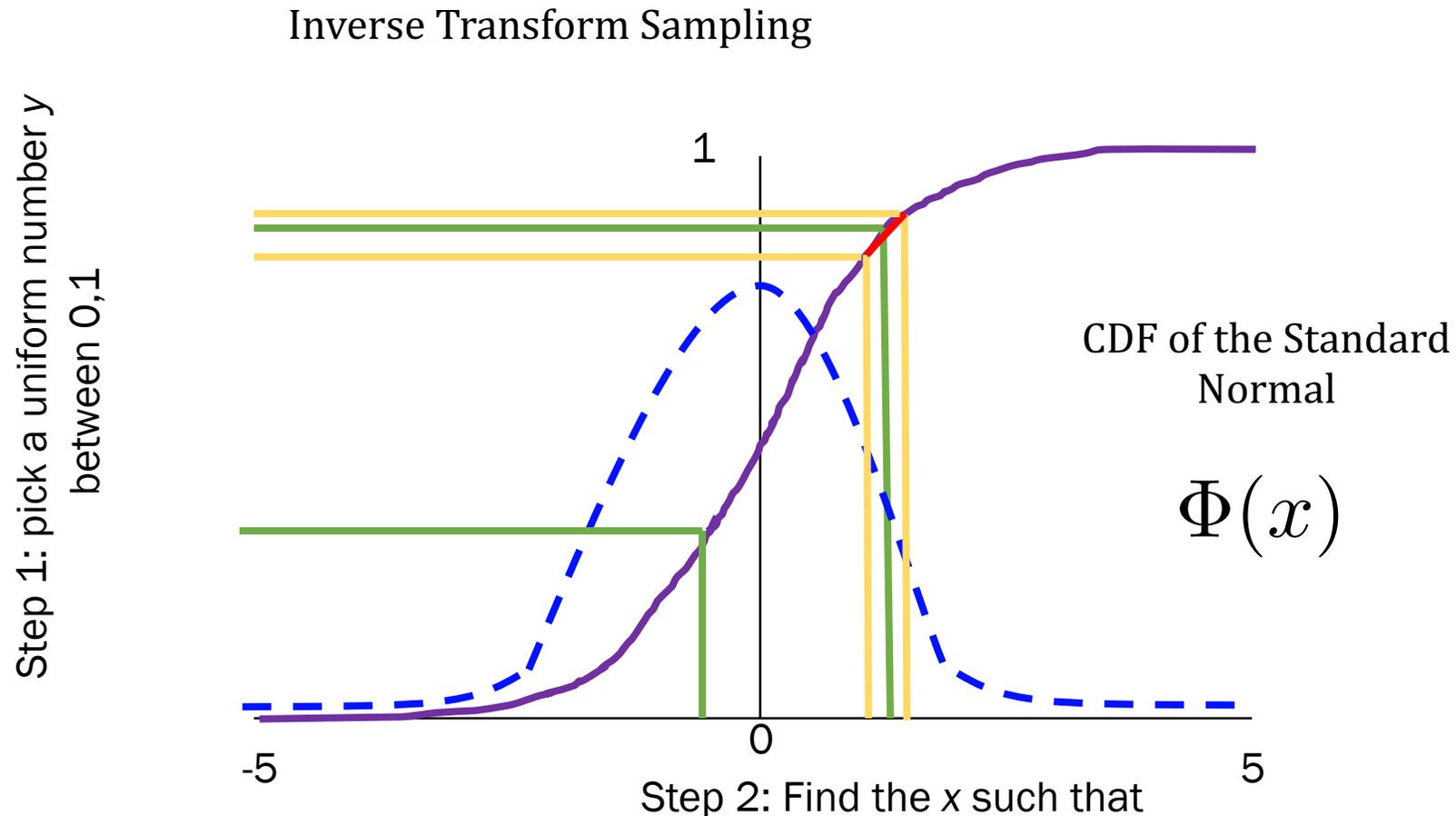
```
6.14724644482
```

```
4.73774184354
```

How Does a Computer Sample a Normal?



How Does a Computer Sample a Normal?



Sample 1:
1.201234

Sample 2:
-0.45123

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

Further reading: Box-Muller transform

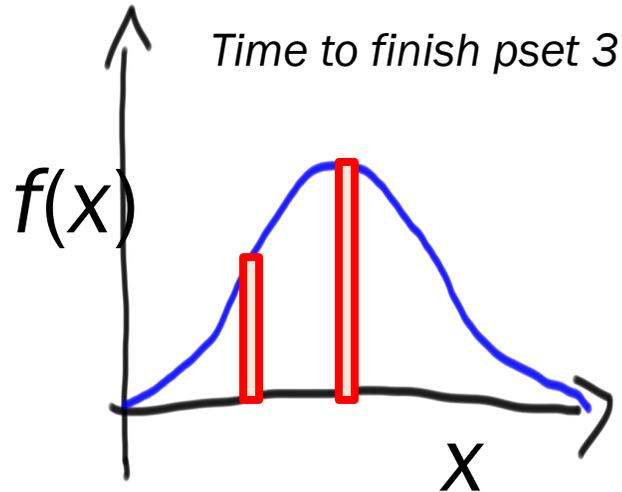


Relative values of a PDF

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?

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



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 $P(\text{drop}) < 0.0001$

Try to solve this by Wednesday...

