Cumulative Distribution 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!

$$F_X(x)$$

This is also shorthand notation for the CDF

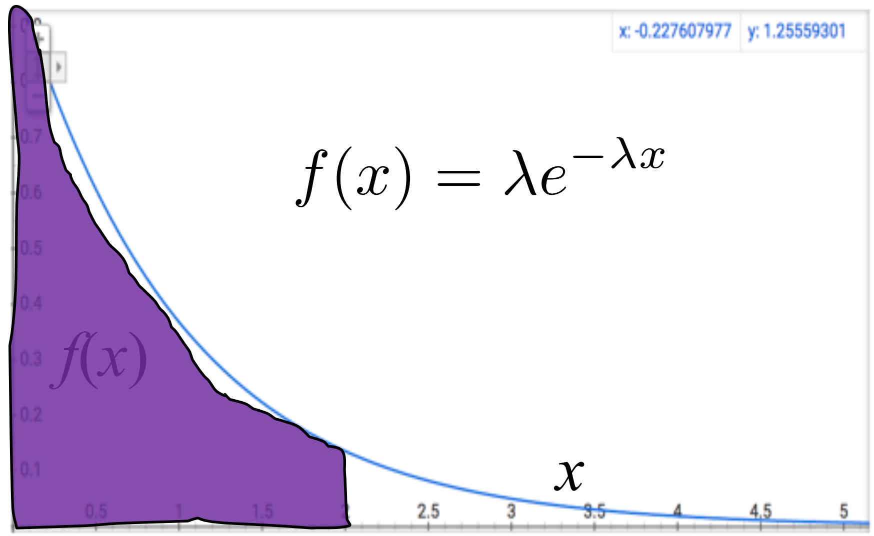
CDF of an Exponential

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

$$
\begin{aligned}
P(X < x) &= \int_{y=-\infty}^{x} f(y)\, dy \\
&= \int_{y=0}^{x} \lambda e^{-\lambda y}\, dy \\
&= \frac{\lambda}{\lambda}\Big[- e^{-\lambda y} \Big]_0^x \\
&= [-e^{-\lambda x}] - [-e^{\lambda 0}] \\
&= 1 - e^{-\lambda x}
\end{aligned}
$$

CDF: X ~ Exp(λ = 1)

Probability density function



P(X < 2)

$$= \int_{x=-\infty}^{2} f(x)\, dx$$

or

$$= F(2)$$

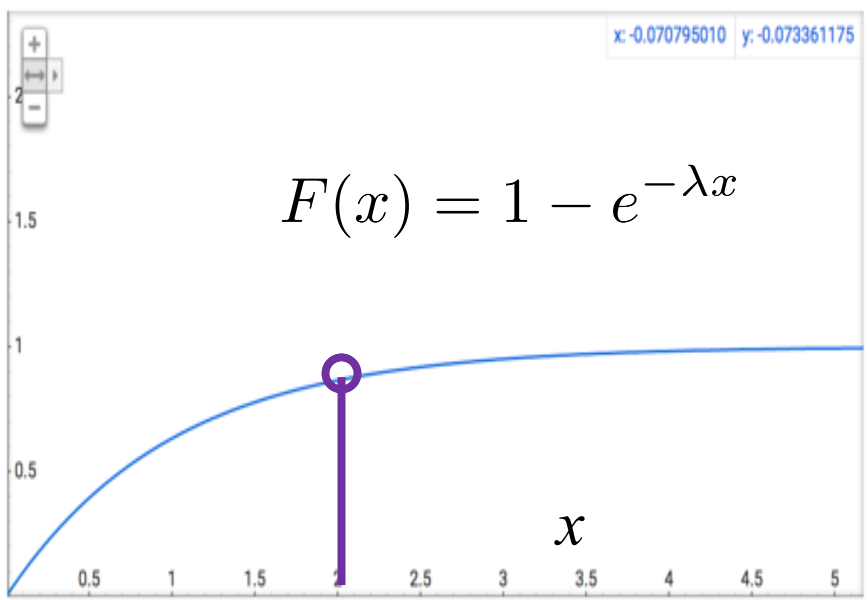
$$= 1 - e^{-2}$$

$$\approx 0.84$$

Cumulative distribution function

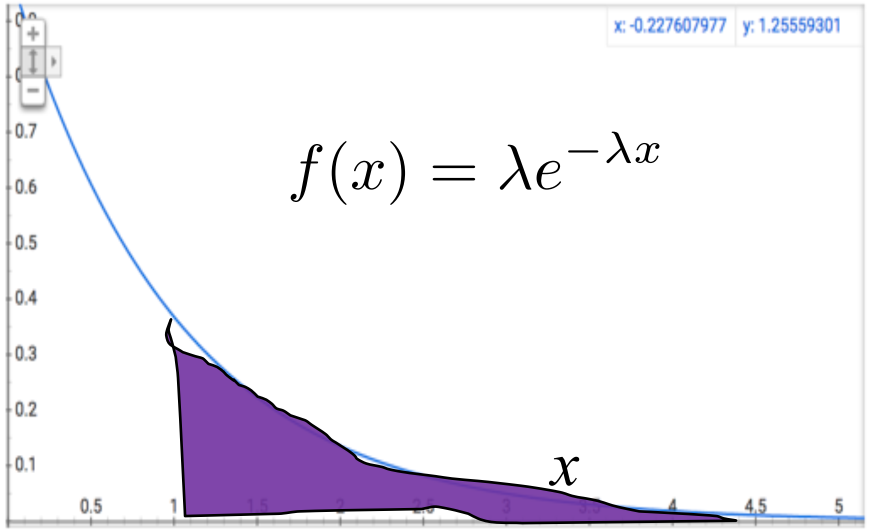
$$F_X(x) = P(X < x)$$

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



CDF: $X \sim \text{Exp}(\lambda = 1)$

Probability density function



$$P(X > 1)$$

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

or

$$= 1 - F(1)$$

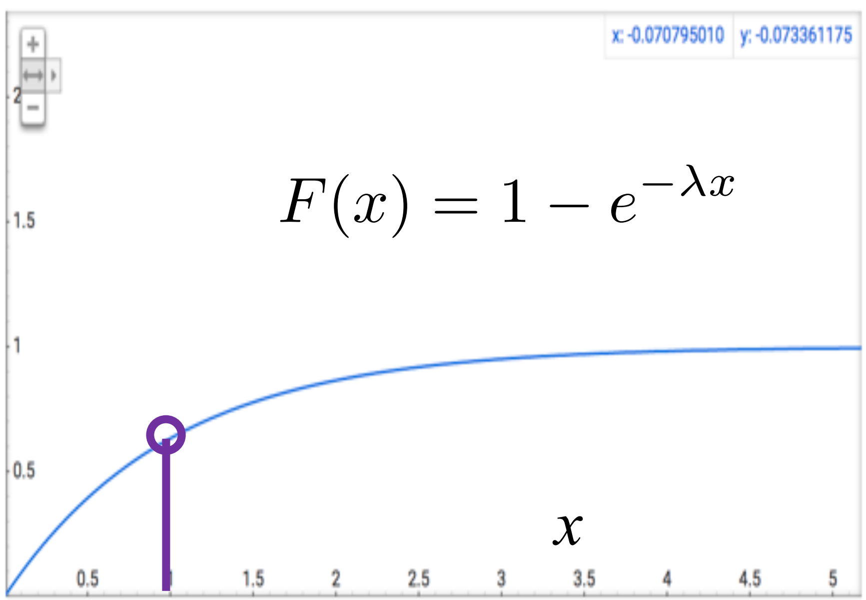
$$= e^{-1}$$

$$\approx 0.37$$

Cumulative distribution function

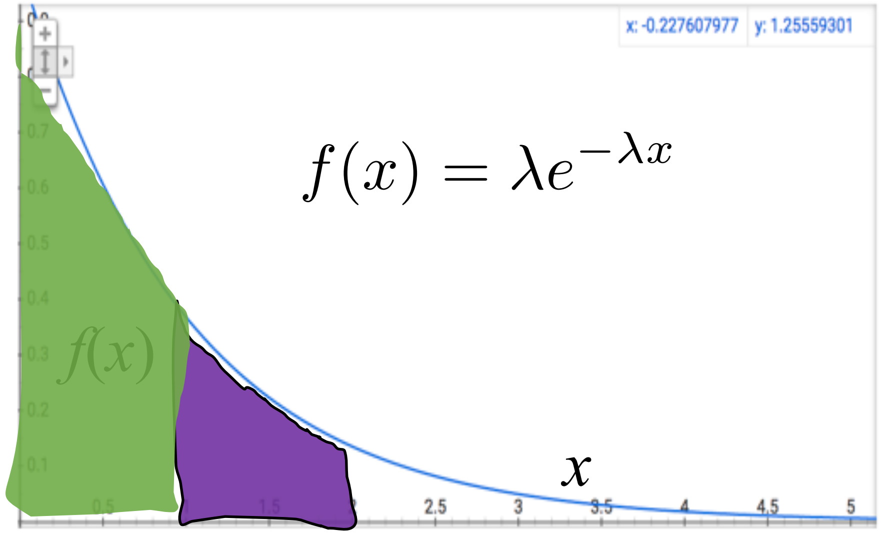
$$F_X(x) = P(X < x)$$

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



CDF: X ~ Exp(λ = 1)

Probability density function



$P(1 < X < 2)$

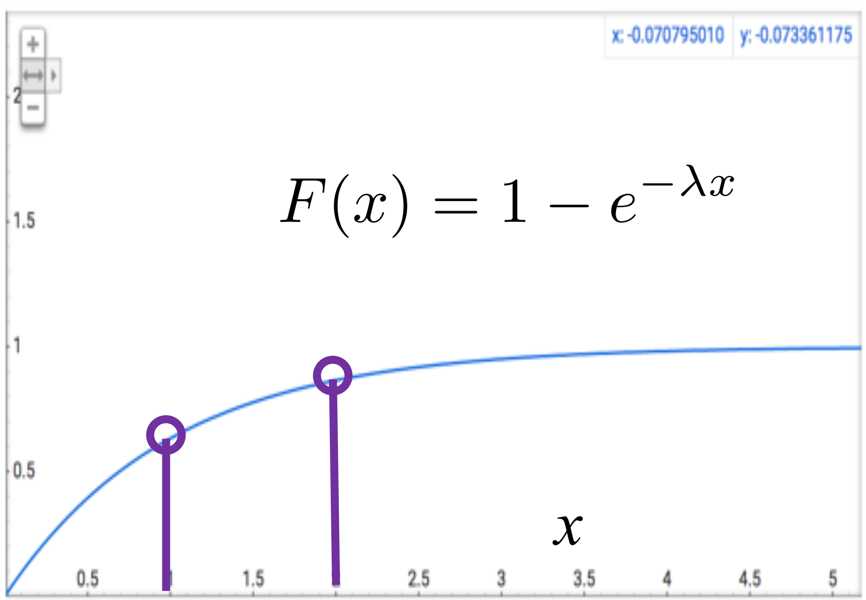
$$= \int_{x=1}^{2} f(x)\, dx$$

or

Cumulative distribution function

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

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



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

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

$$\approx 0.23$$

Probability of Earthquake in Next 4 Years?

Based on historical data, earthquakes of magnitude 8.0+ happen at a **rate of 0.002** per year*. What is the probability of **a major earthquake in the next 4 years?**

Y = Years until the next earthquake of magnitude 8.0+

$$Y \sim \text{Exp}(\lambda = 0.002) \qquad F(y) = 1 - e^{-0.002y}$$

$$\begin{aligned} P(Y < 4) &= F(4) \\ &= 1 - e^{-0.002 \cdot 4} \\ &\approx 0.008 \end{aligned}$$

*According to USGS, 2015

Notation

$p(a)$ or $p_X(a)$ Probability Mass Function (discrete) $P(X = a)$

$f(a)$ or $f_X(a)$ Probability Density Function (continuous) $f(X = a)$

$F(a)$ or $F_X(a)$ Cumulative Distribution Function $P(X \leq a)$

Exponential is Memoryless

- X = time until some event occurs
  - X ~ Exp(λ)
  - What is P(X > s + t | X > s)?

$$P(X > s + t \mid X > s) = \frac{P(X > s + t \text{ and } X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}$$

$$\frac{P(X > s + t)}{P(X > s)} = \frac{1 - F(s + t)}{1 - F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P(X > t)$$

$$\text{So, } P(X > s + t \mid X > s) = P(X > t)$$

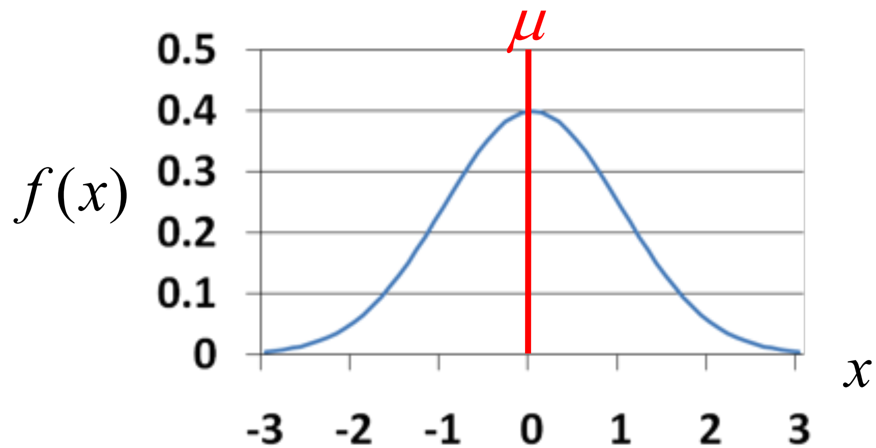
  - After initial period of time *s*, P(X > t | •) for waiting another *t* units of time until event is same as at start
  - "Memoryless" = no impact from preceding period *s*

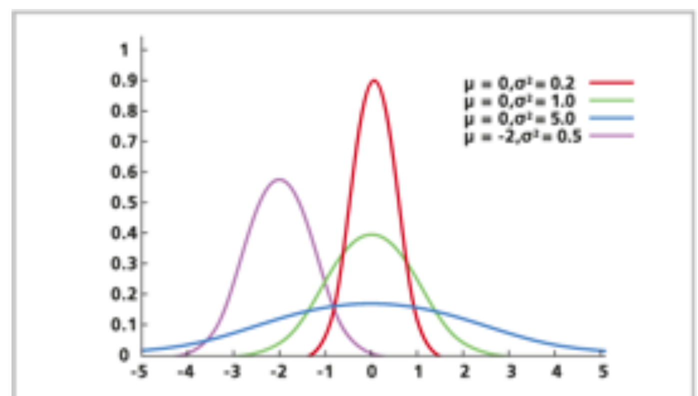
The Normal Distribution

- X is a **<u>Normal Random Variable</u>**: $X \sim \mathrm{N}(\mu, \sigma^2)$
  - Probability Density Function (PDF):

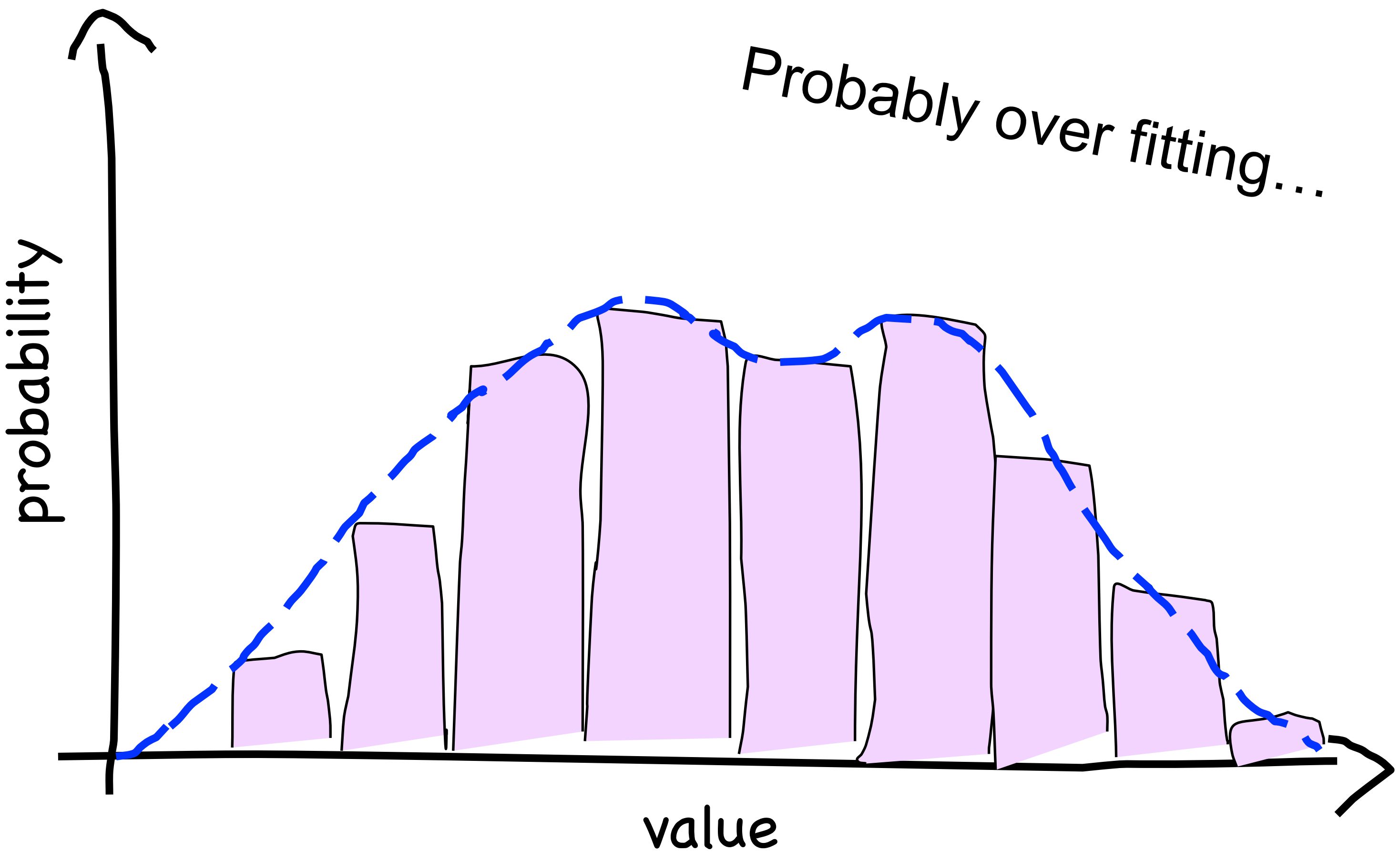
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{where} \quad -\infty < x < \infty$$

  - $E[X] = \mu$
  - $Var(X) = \sigma^2$
  - Also called “Gaussian”
  - Note: $f(x)$ is symmetric about μ



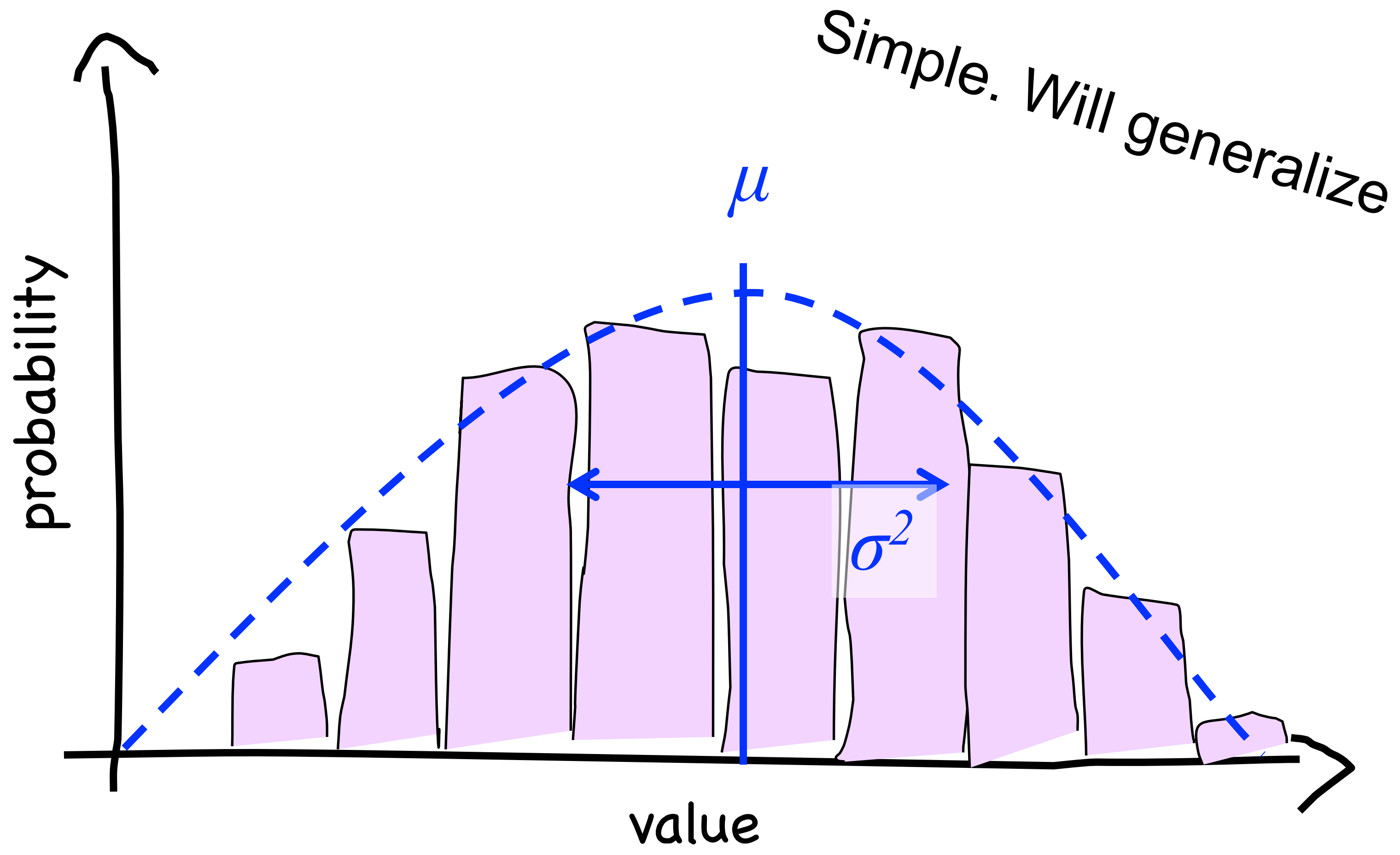


Complexity is Tempting



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

Simplicity is Humble

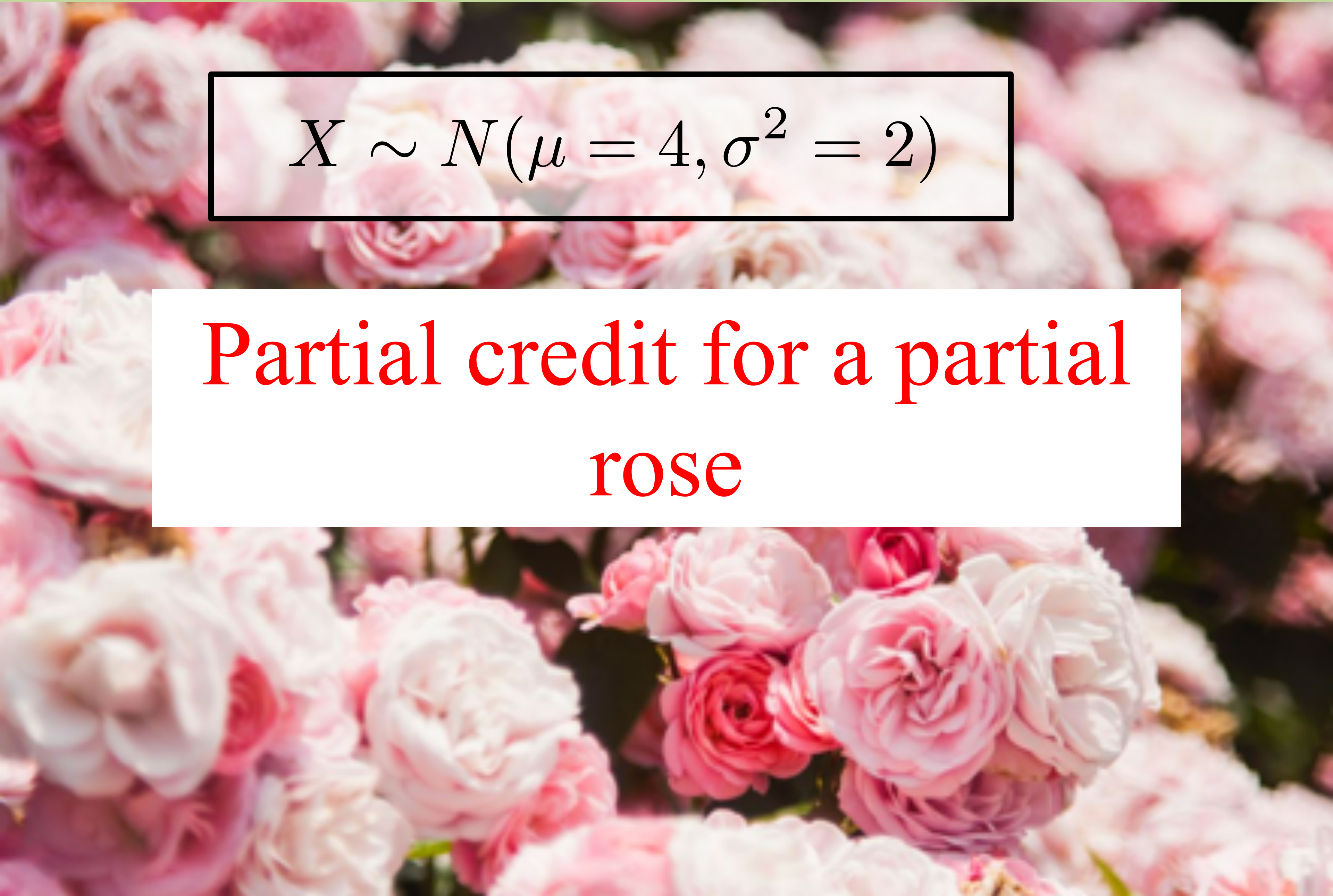


* A Gaussian maximizes entropy for a given mean and variance

Flowers on a Rose Bush

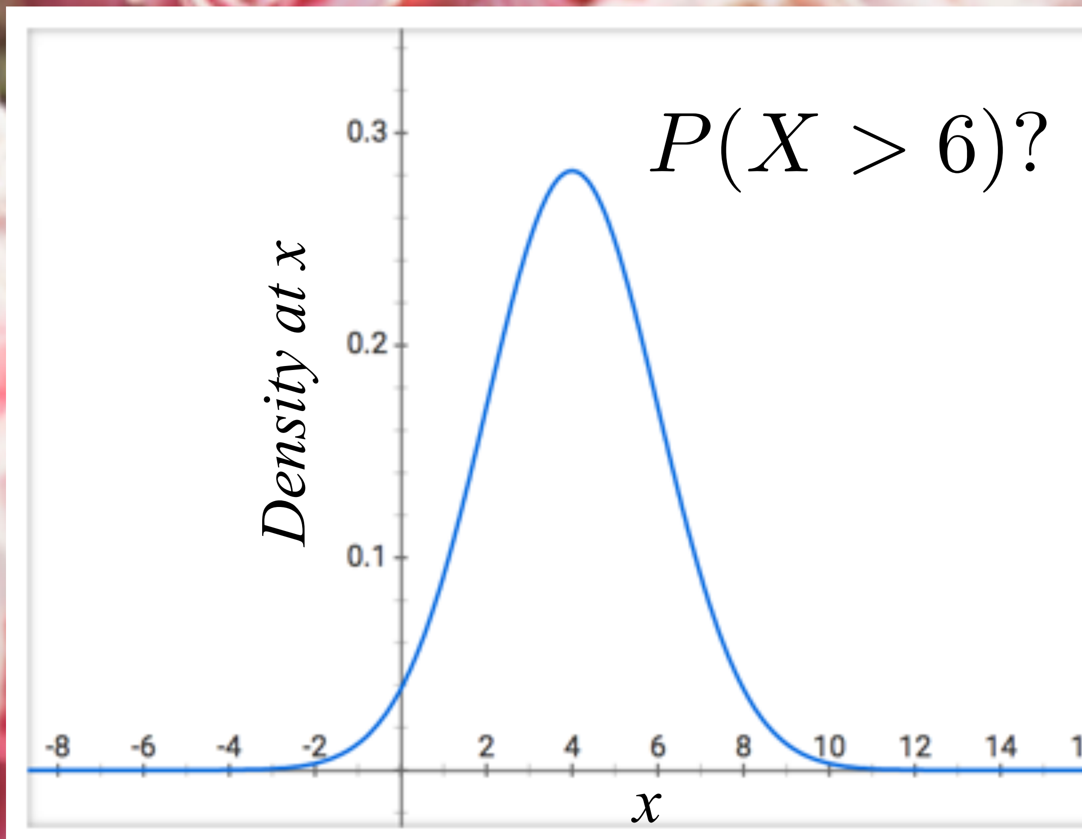
$$X \sim N(\mu = 4, \sigma^2 = 2)$$

Partial credit for a partial rose



Flowers on a Rose Bush

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



Let's try to integrate it!

$$P(a \leq X \leq b) =$$

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

No closed form for the integral

Linear Transform of Normal is Normal

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

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

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

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

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

Special Linear Transform

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

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

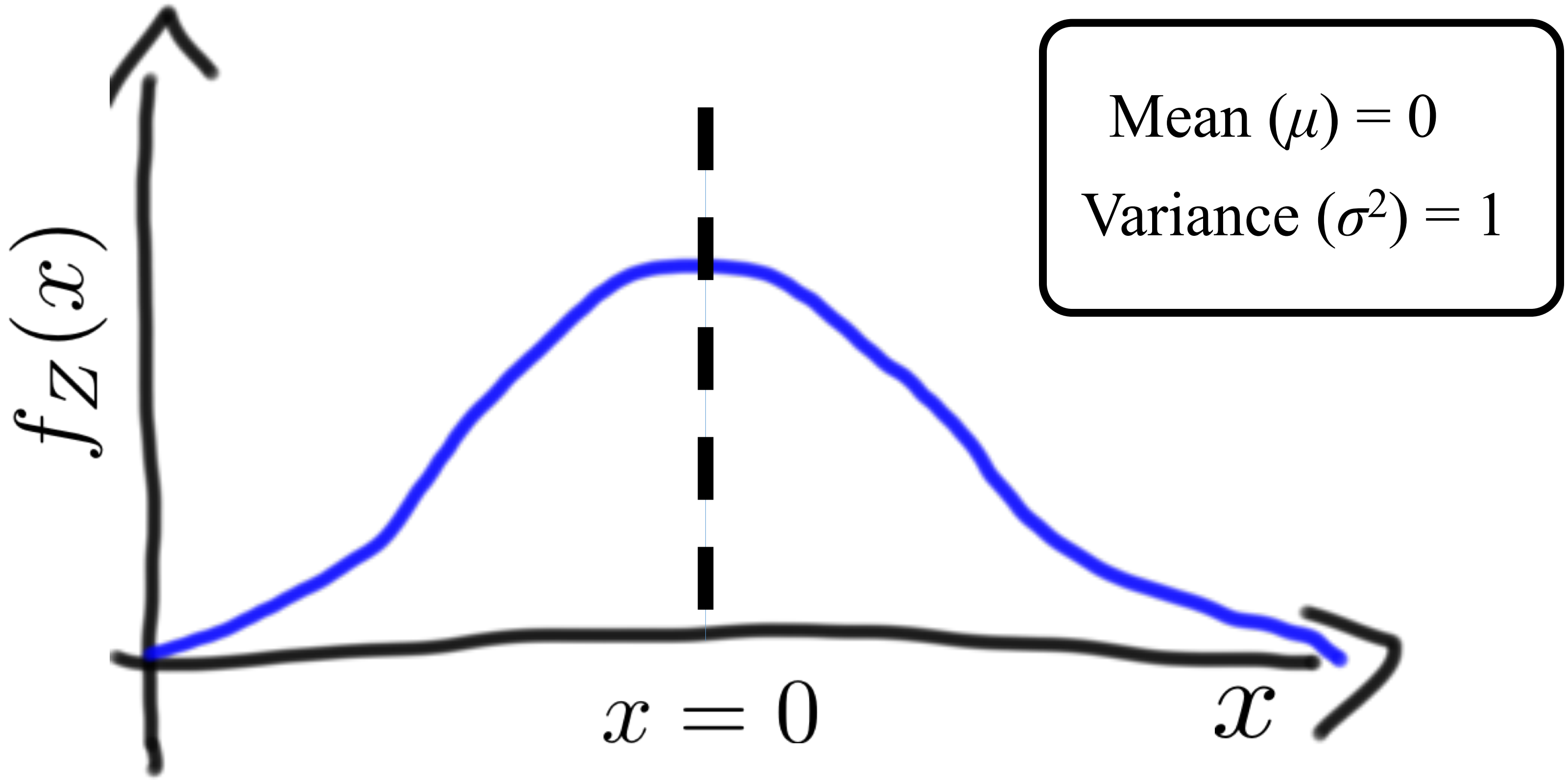
There is a special case of linear transform for any X:

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

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

The Standard Normal

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

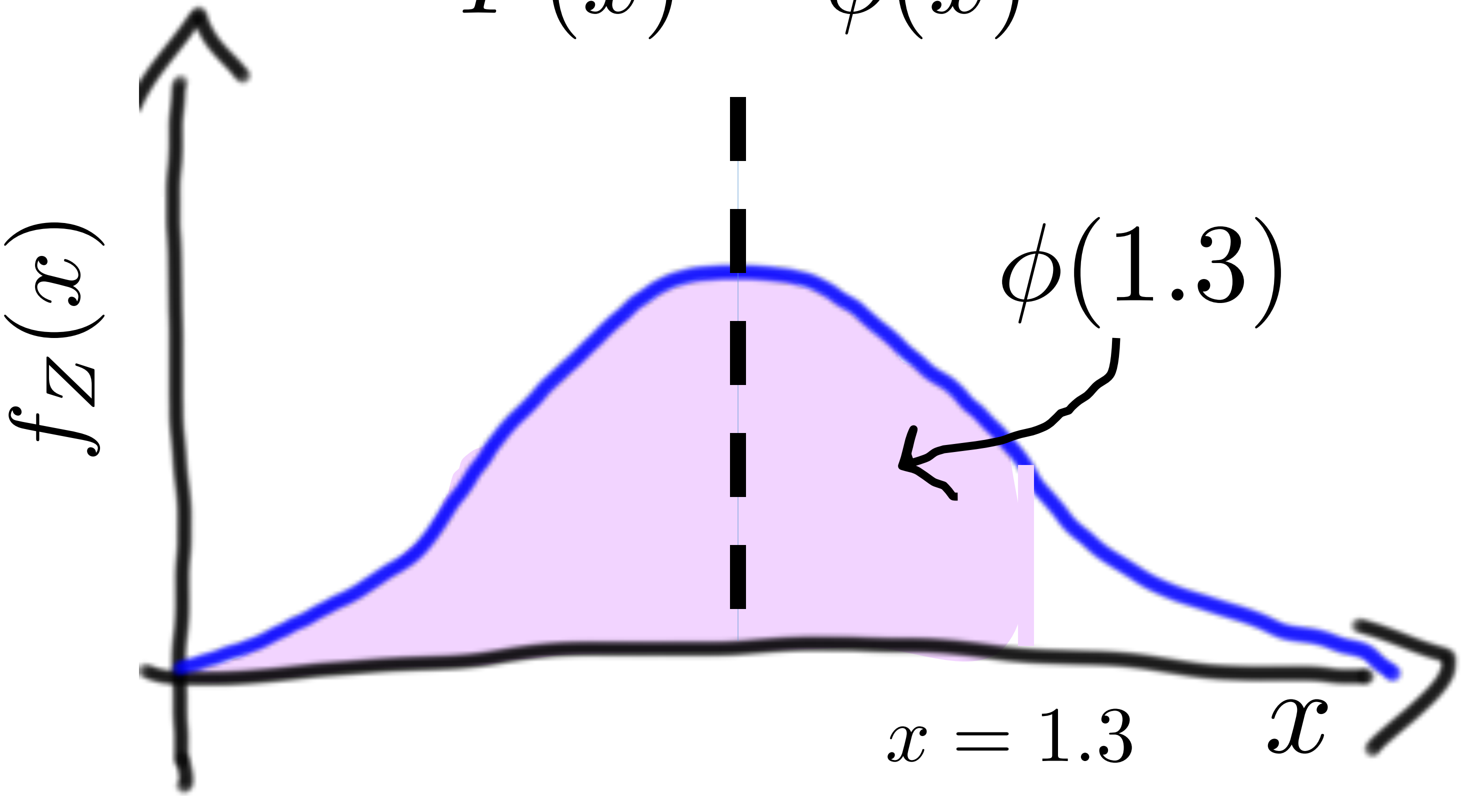


*This is the probability density function for the standard normal

Phi

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

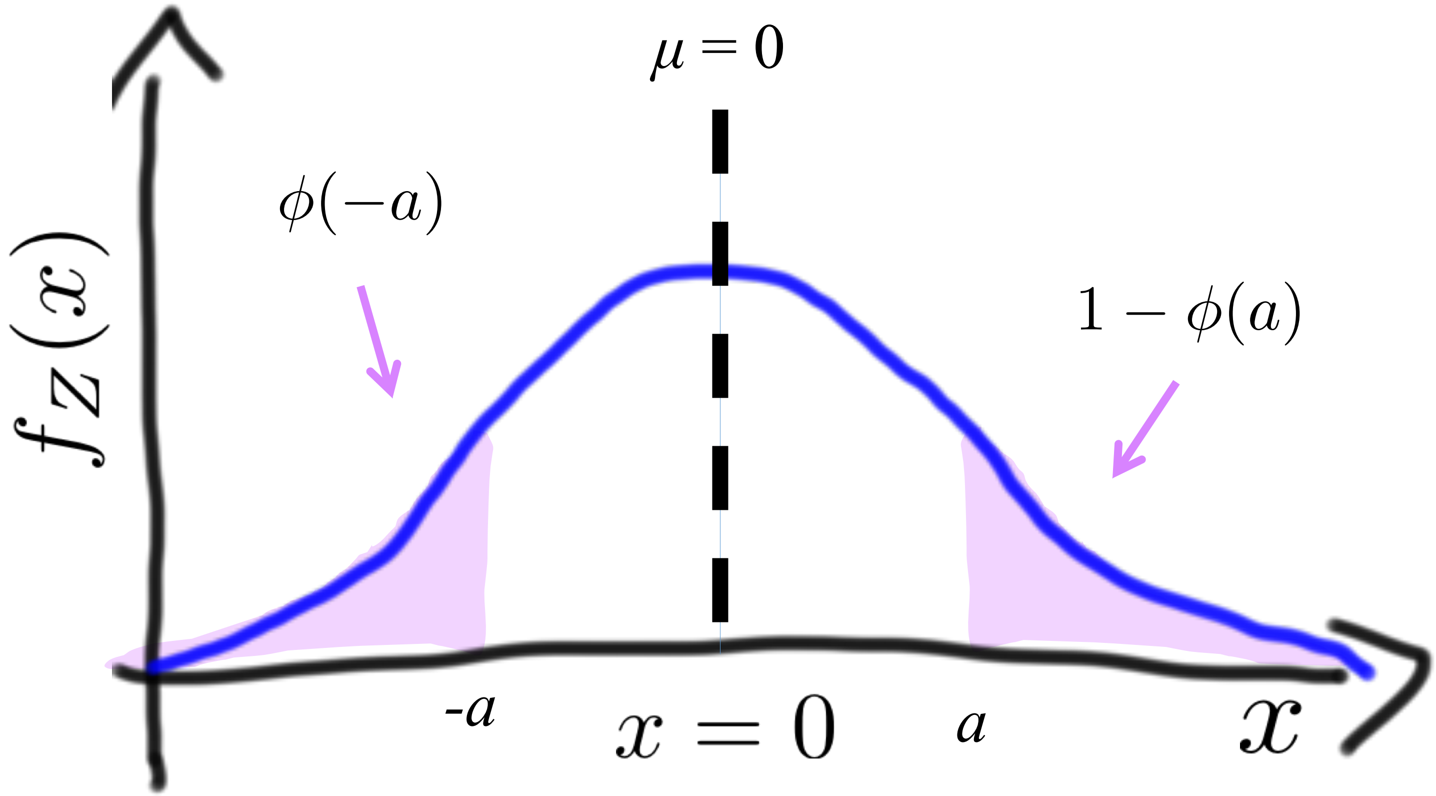
$$F(x) = \phi(x)$$



*This is the probability density function for the standard normal

Symmetry of Phi

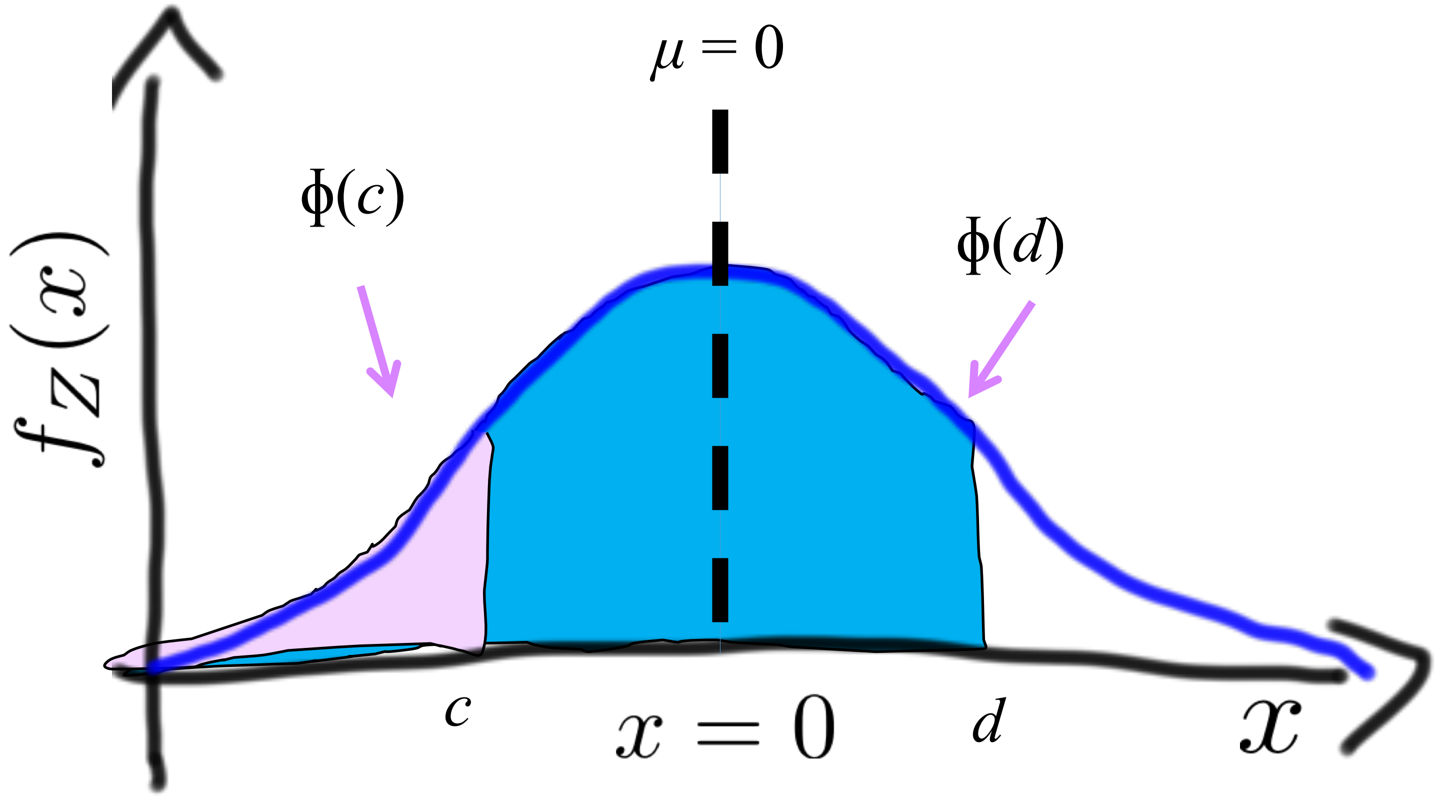
$$\varphi(-a) = 1 - \varphi(a)$$



*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

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

Use Z to compute $F(x)$

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

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

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

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

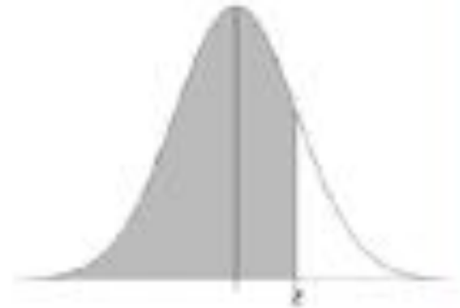
The cumulative density function (CDF) of any normal

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

Using Table of ϕ

Standard Normal Cumulative Probability Table

$\Phi(0.54) = 0.7054$

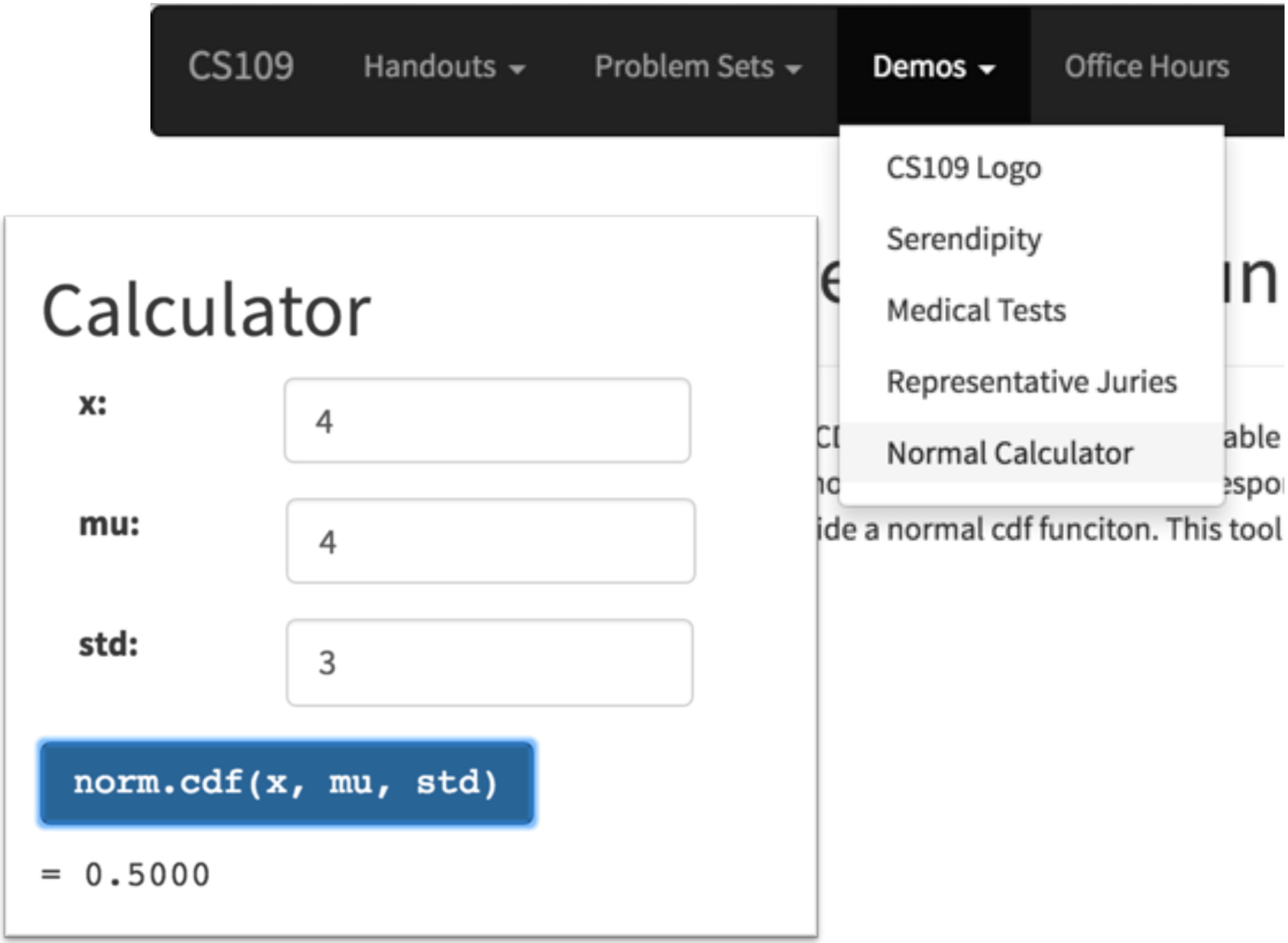


Cumulative probabilities for POSITIVE z-values are shown in the following table:

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

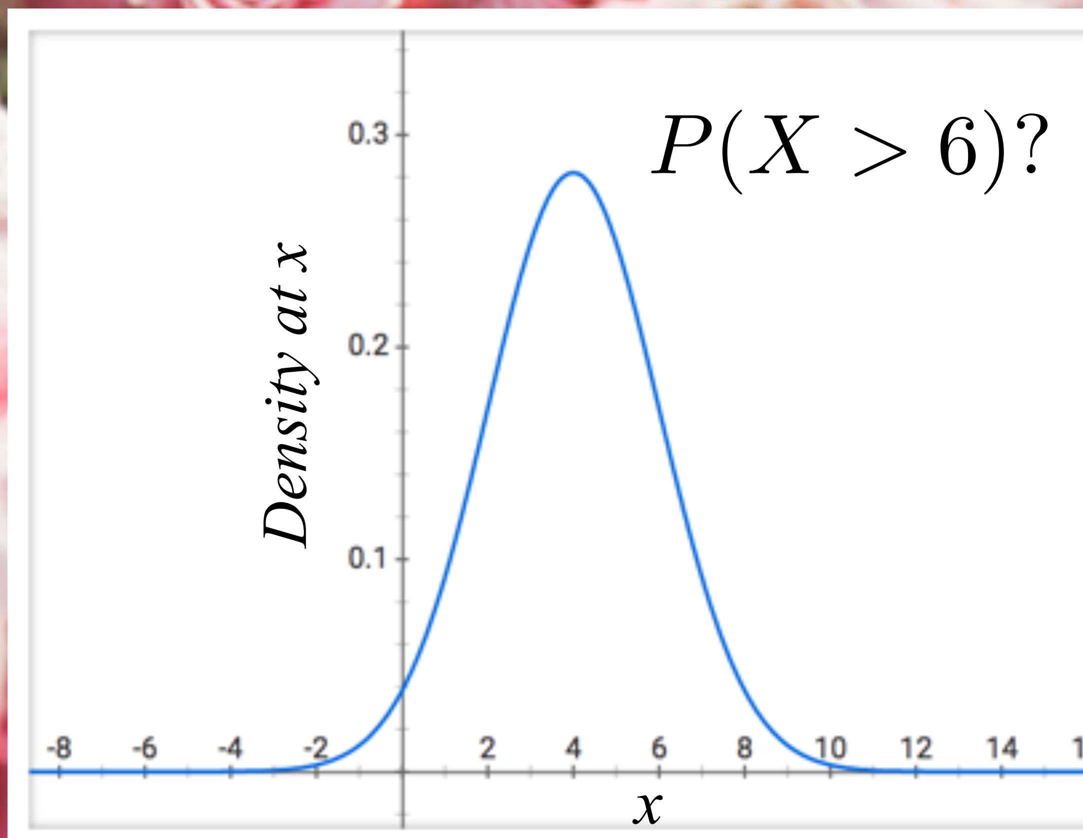
We have one for you



End Review

Flowers on a Rose Bush

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



Flowers on a Rose Bush

flowers on a rose bush

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

$$P(X > 6)?$$

$$P(X > 6) = 1 - F_X(6)$$

$$= 1 - \Phi\left(\frac{6-\mu}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{6-4}{\sqrt{2}}\right)$$

$$\approx 1 - \Phi\left(1.414\right)$$

$$\approx 0.079$$

For any normal:

$$F_X(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$$

Get Your Gaussian On

$X \sim N(\mu = 3, \sigma^2 = 16) \qquad \mu = 3 \qquad \sigma = 4$

$$\begin{aligned} P(X > 0) &= 1 - F_X(0) = 1 - \Phi\left(\frac{0 - \mu}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{-3}{4}\right) = 1 - (1 - \Phi\left(\frac{3}{4}\right)) \\ &= \Phi\left(\frac{3}{4}\right) = 0.7734 \end{aligned}$$

$$\begin{aligned} P(2 < X < 5) &= F_X(5) - F_X(2) = \Phi\left(\frac{5 - \mu}{\sigma}\right) - \Phi\left(\frac{2 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{5 - 3}{4}\right) - \Phi\left(\frac{2 - 3}{4}\right) = \Phi\left(\frac{2}{4}\right) - \Phi\left(\frac{-1}{4}\right) = 0.6915 \end{aligned}$$

$$\begin{aligned} P(|X - 3| > 6) &= P(X < -3) + P(X > 9) = F_X(-3) + (1 - F_X(9)) \\ &= \Phi\left(\frac{-3 - 3}{4}\right) + (1 - \Phi\left(\frac{9 - 3}{4}\right)) = 0.1337 \end{aligned}$$

Noisy Wires

- Send voltage of 2 or -2 on wire (to denote 1 or 0)
  - X = voltage sent
  - R = voltage received = X + Y, where noise Y ~ N(0, 1)
  - Decode R: if (R ≥ 0.5) then 1, else 0
  - What is P(error after decoding | original bit = 1)?

$$P(2+Y<0.5)=P(Y<-1.5)=\Phi(-1.5)=1-\Phi(1.5)\approx 0.0668$$

  - What is P(error after decoding | original bit = 0)?

$$P(-2+Y\geq 0.5)=P(Y\geq 2.5)=1-\Phi(2.5)\approx 0.0062$$

Gaussian for uncertainty

ELO Ratings

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

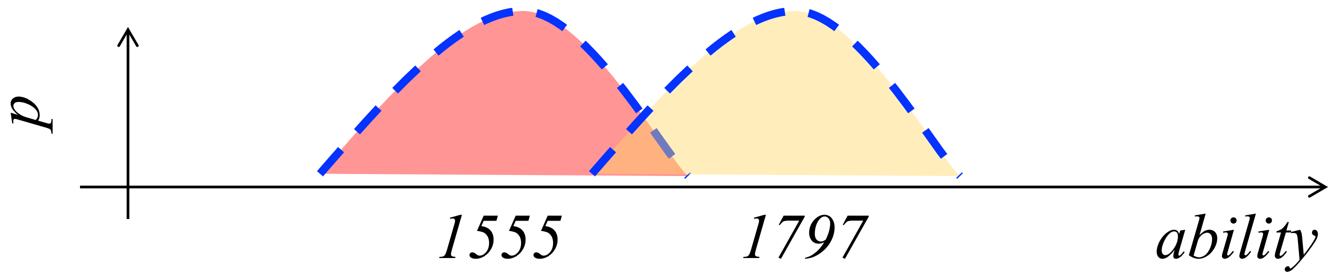
ELO Ratings

How it works:

- Each team has an "ELO" score S, calculated based on their past performance.
- Each game, the team has ability $A \sim N(S, 200^2)$
- The team with the higher sampled ability wins.

Arpad Elo

$A_B \sim \mathcal{N}(1555, 200^2)$ $A_W \sim \mathcal{N}(1797, 200^2)$



$$P(\text{Warriors win}) = P(A_W > A_B)$$

ELO Ratings

```
from random import *

WARRIORS_ELO = 1797
OPPONENT_ELO = 1555
VAR = 200 * 200

nSuccess = 0
for i in range(NTRIALS):
   w = gauss(WARRIORS_ELO, VAR)
   b = gauss(OPPONENT_ELO, VAR)
   if w > b:
      nSuccess += 1

print float(nSuccess) / NTRIALS
```

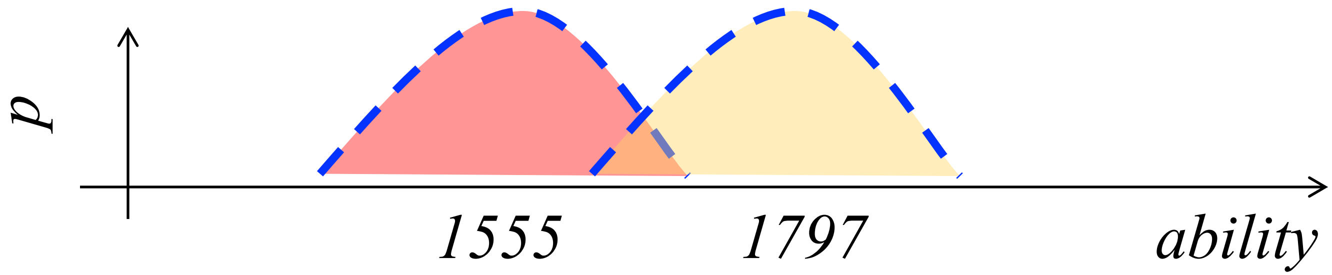
ELO Ratings

How it works:

- Each team has an "ELO" score *S*, calculated based on their past performance.
- Each game, the team has ability $A \sim N(S, 200^2)$
- The team with the higher sampled ability wins.

Arpad Elo

$A_B \sim \mathcal{N}(1555, 200^2)$ $\quad$ $A_W \sim \mathcal{N}(1797, 200^2)$



$$P(\text{Warriors win}) = P(A_W > A_B)$$
$$\approx 0.87$$

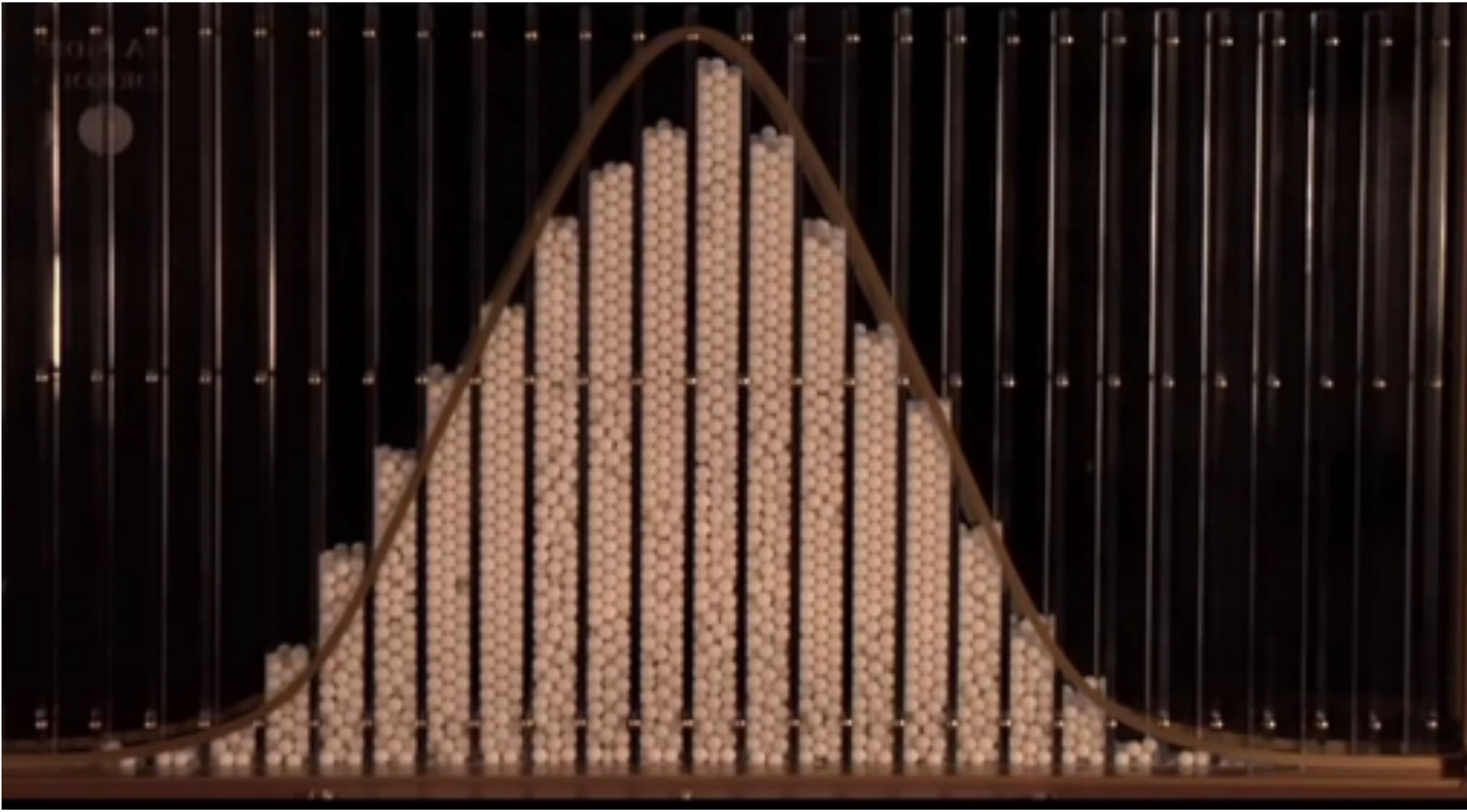
← Calculated via sampling

Website Testing

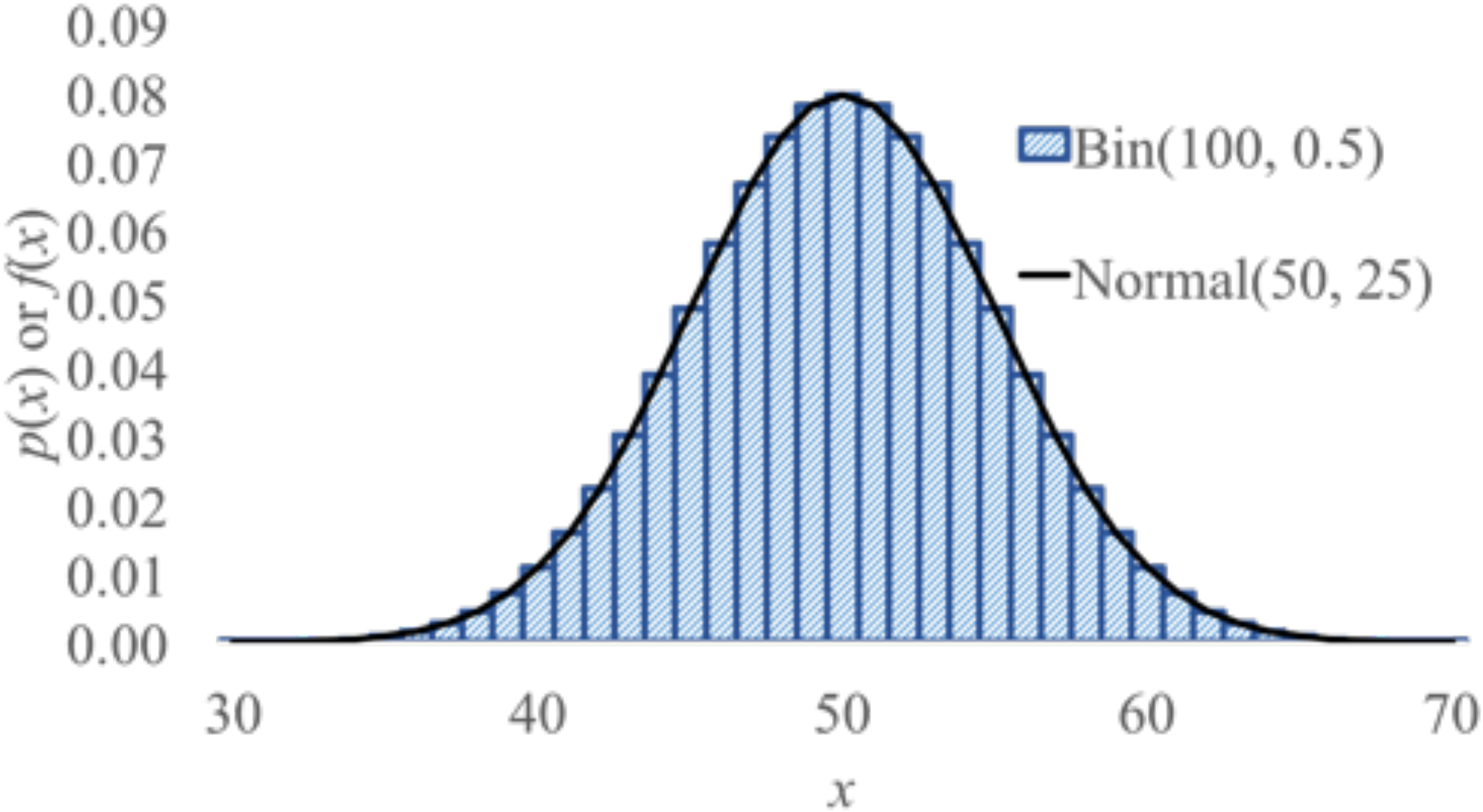
- 100 people are given a new website design
  - X = # people whose time on site increases
  - CEO will endorse new design if X ≥ 65 What is P(CEO endorses change| it has no effect)?
  - X ~ Bin(100, 0.5). Want to calculate P(X ≥ 65)
  - Give a numerical answer…

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

Normal Approximates Binomial



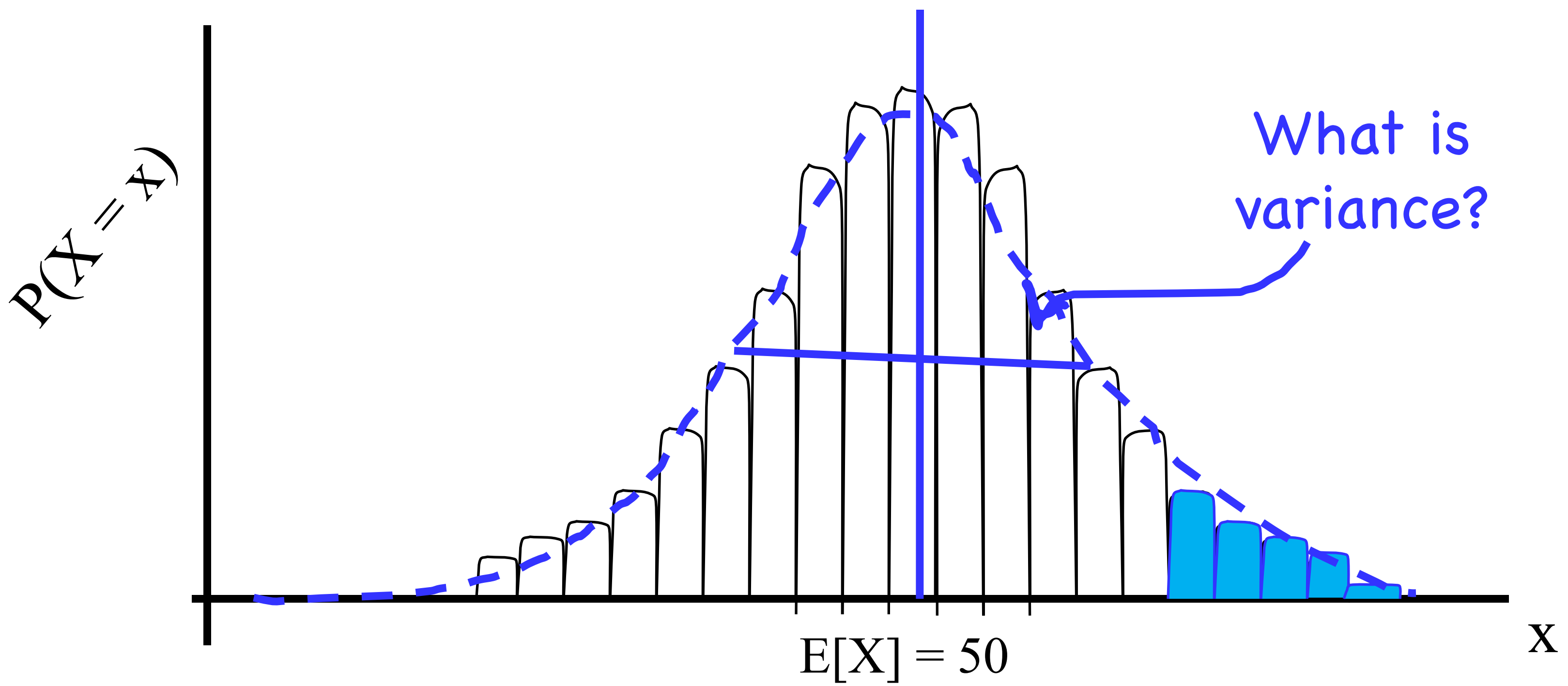
Normal Approximates Binomial



Let’s invent an approximation!

Website Testing

- 100 people are given a new website design
  - X = # people whose time on site increases
  - CEO will endorse new design if X ≥ 65 What is P(CEO endorses change| it has no effect)?
  - X ~ Bin(100, 0.5). Want to calculate P(X ≥ 65)



Website Testing

- 100 people are given a new website design
  - X = # people whose time on site increases
  - CEO will endorse new design if X ≥ 65 What is P(CEO endorses change| it has no effect)?
  - X ~ Bin(100, 0.5). Want to calculate P(X ≥ 65)

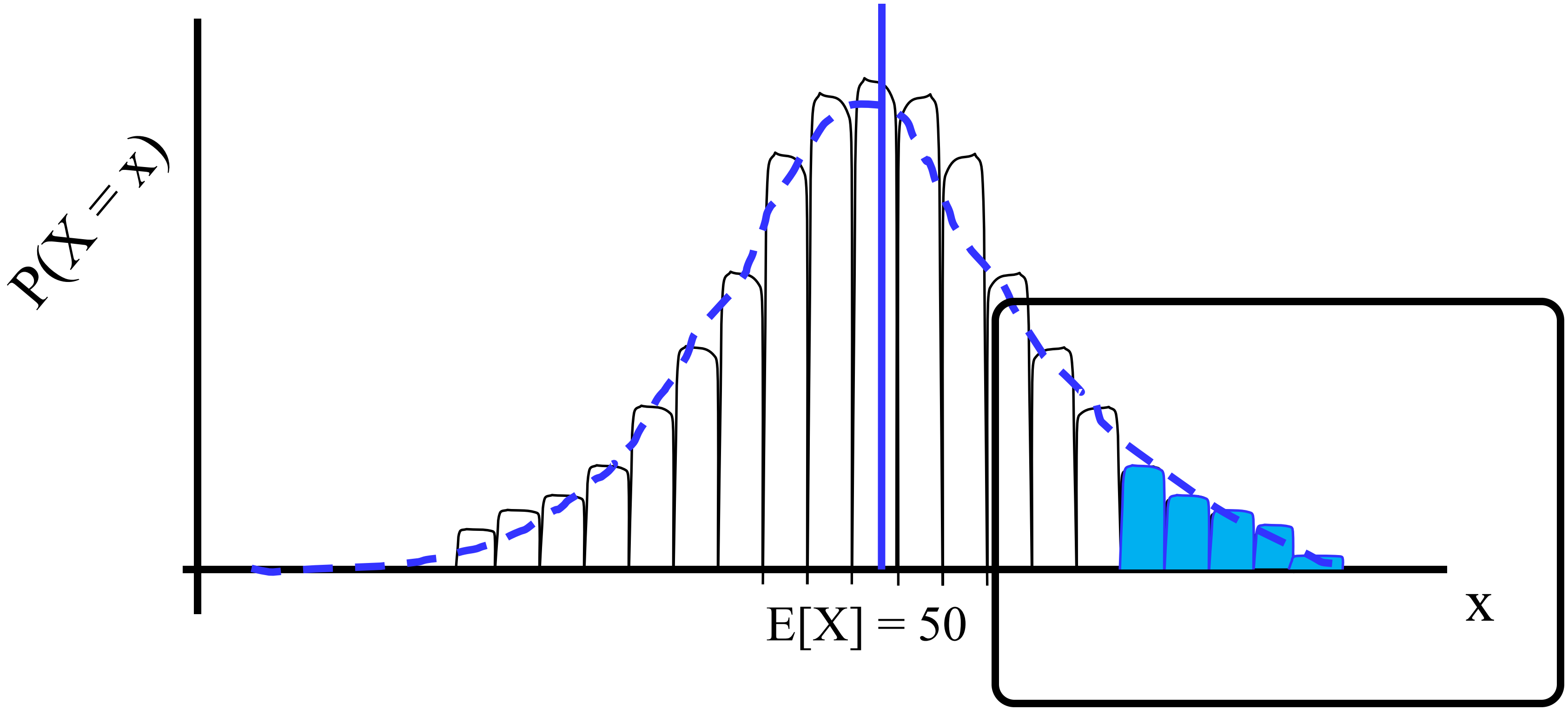
$$np = 50 \quad np(1-p) = 25 \quad \sqrt{np(1-p)} = 5$$

  - Use Normal approximation: Y ~ N(50, 25)

$$P(Y \geq 65) = P(\frac{Y-50}{5} > \frac{65-50}{5}) = P(Z > 3) = 1 - \phi(3) \approx 0.0013$$

  - Using Binomial: $P(X \geq 65) \approx 0.0018$

Website Testing



Continuity Correction

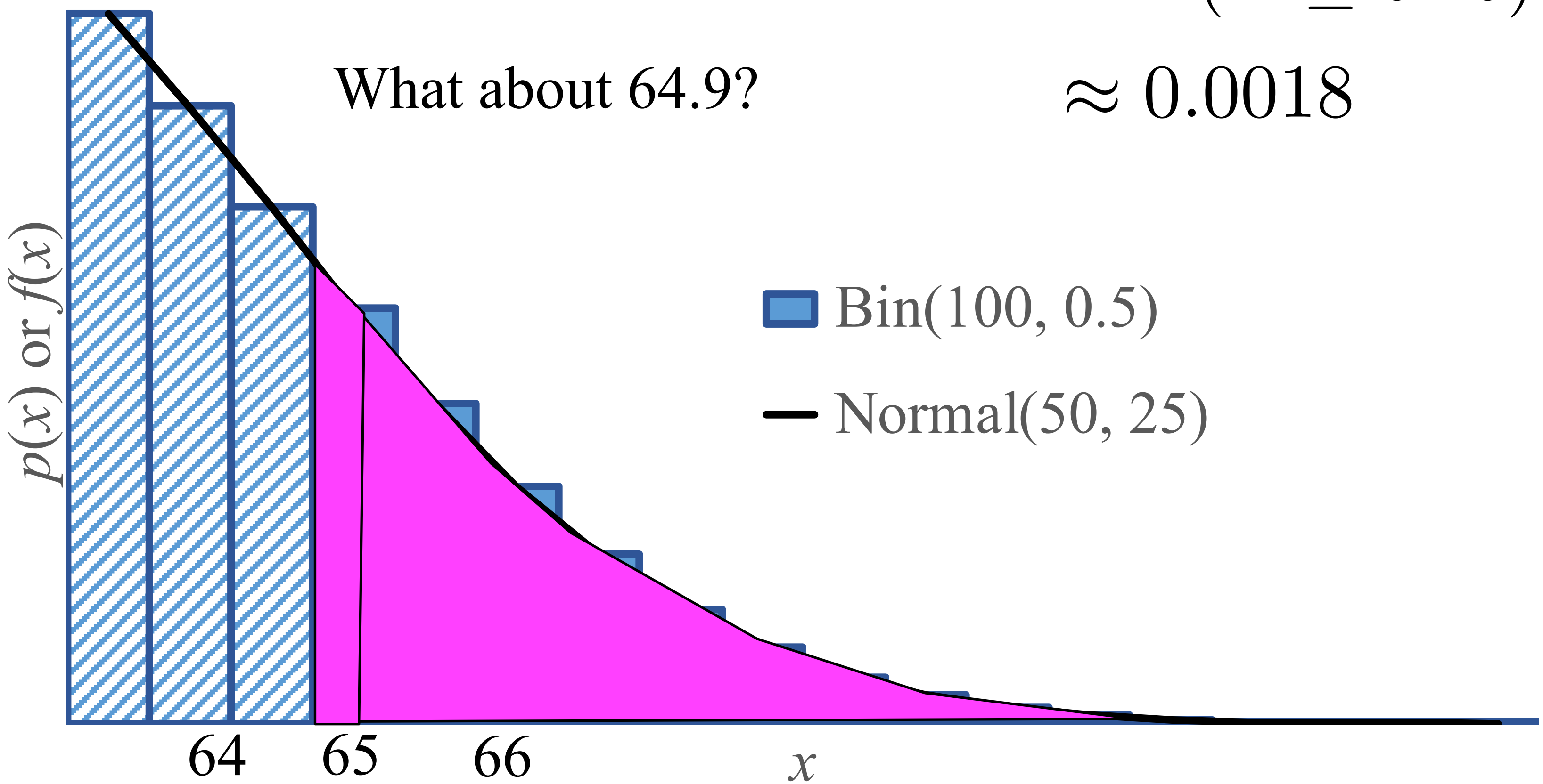
If Y (normal) approximates X (binomial)

$$P(X \geq 65)$$

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

$$\approx 0.0018$$

What about 64.9?



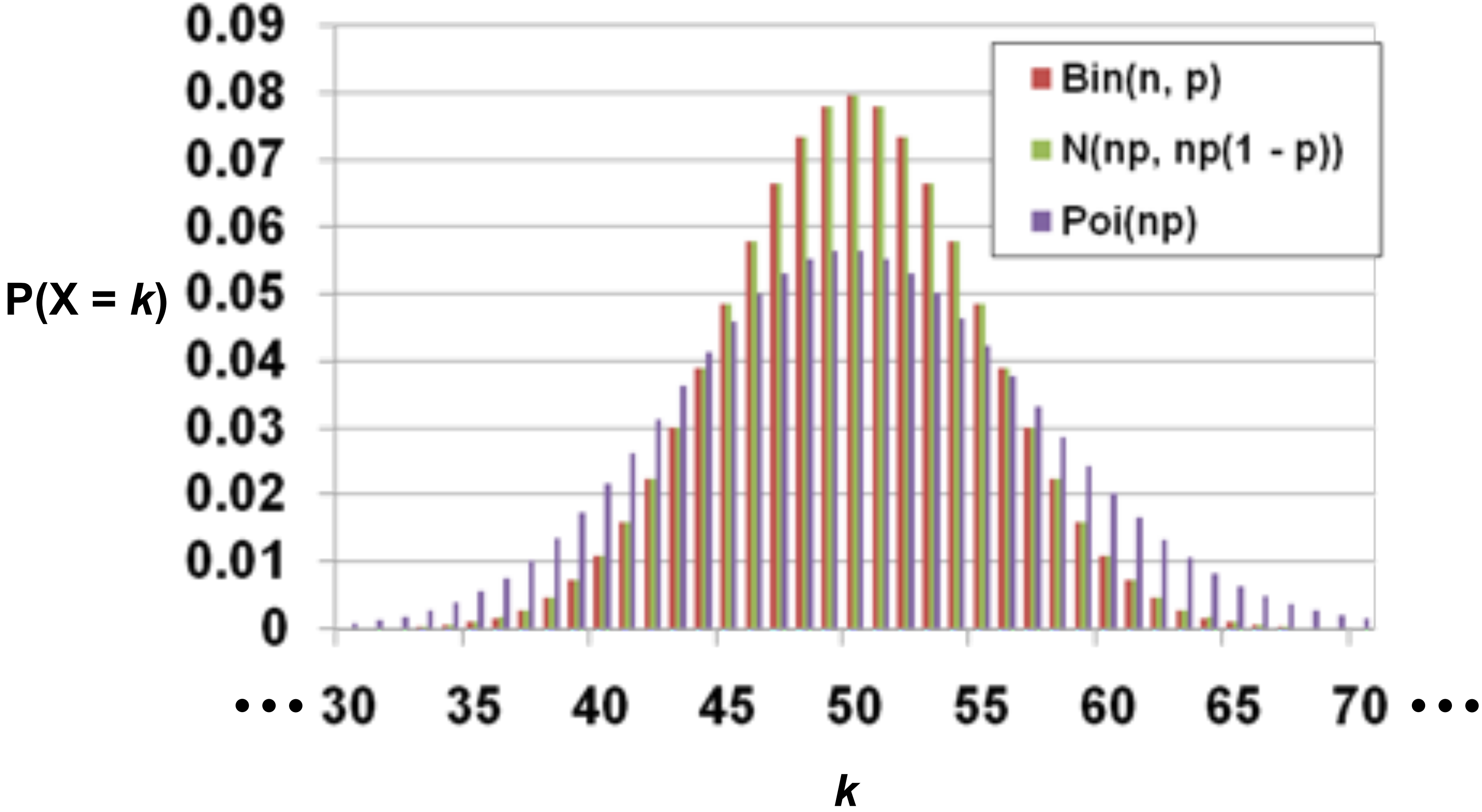
Continuity Correction

If Y (normal) approximates X (binomial)

Discrete (eg Binomial) probability question	Continuous (Normal) probability question
X = 6	5.5 < Y < 6.5
X >= 6	Y > 5.5
X > 6	Y > 6.5
X < 6	Y < 5.5
X <= 6	Y < 6.5

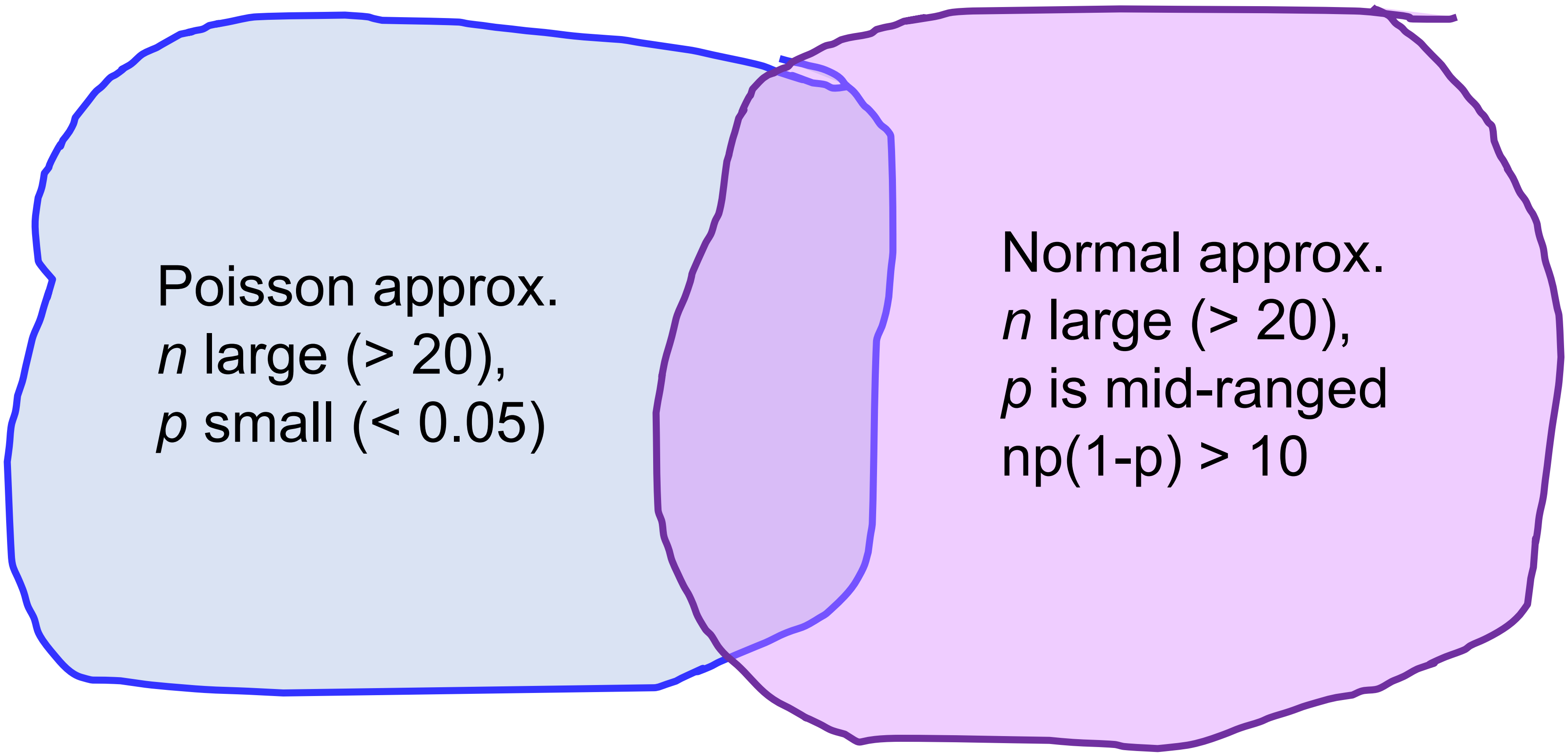
* Note: Binomial is always defined in units of “1”

Comparison when n = 100, p = 0.5



Who Gets to Approximate?

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



If there is a choice, go with the normal approximation

Stanford Admissions

- Stanford accepts 2050 students this year
  - Each accepted student has 84% chance of attending
  - X = # students who will attend. X ~ Bin(2050, 0.84)
  - What is P(X > 1745)?

$$np = 1722 \qquad np(1-p) = 276 \qquad \sqrt{np(1-p)} = 16.6$$

  - Use Normal approximation: Y ~ N(1722, 276)

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

$$P(Y \geq 1745.5) = P\left(\frac{Y - 1722}{16.6} > \frac{1745.5 - 1722}{16.6}\right) = P(Z > 1.4)$$

$$\approx 0.08$$


Stretch!
BOI

Joint Distributions

Noah Arthurs
CS109, Stanford University

Continuous Random Variables

Uniform Random Variable $X \sim Uni(\alpha, \beta)$

All values of *x* between alpha and beta are equally likely.

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

Aka Gaussian. Defined by mean and variance. Goldilocks distribution.

Exponential Random Variable $X \sim Exp(\lambda)$

Time until an event happens. Parameterized by lambda (same as Poisson).

Beta Random Variable

Coming soon!

Joint Distributions

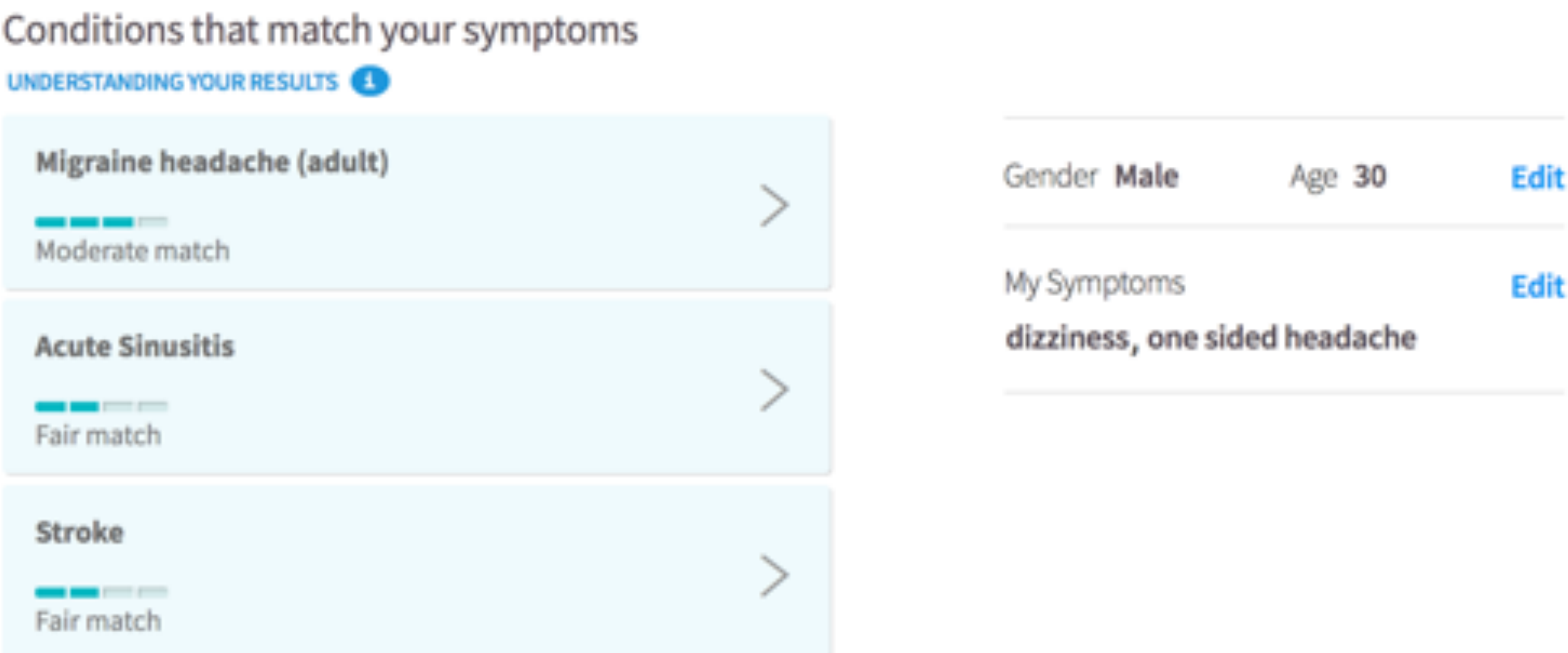
Many interesting problems…

Have multiple random variables interacting with one another

Multiple Random Variables

WebMD
HEALTH
NOW

Multiple Random Variables



Events occur with other events

Probability Table for Discrete

- States all possible outcomes with several discrete variables
- A probability table is not "parametric"
- If #variables is > 2, you can have a probability table, but you can't draw it on a slide

All values of A

All values of B

B \ A	0	1	2
0			
1		$P(A = 1, B = 1)$	
2			

Every outcome falls into a bucket

Here "," means "and"

Discrete Joint Mass Function

- For two discrete random variables *X* and *Y*, the **Joint Probability Mass Function** is:

$$p_{X,Y}(a,b) = P(X = a, Y = b)$$

- Marginal distributions:

$$p_X(a) = P(X = a) = \sum_y p_{X,Y}(a, y)$$

$$p_Y(b) = P(Y = b) = \sum_x p_{X,Y}(x, b)$$

- Example: X = value of die D_1, Y = value of die D_2

$$P(X = 1) = \sum_{y=1}^{6} p_{X,Y}(1, y) = \sum_{y=1}^{6} \frac{1}{36} = \frac{1}{6}$$

A Computer (or Three) In Every House

- Consider households in Silicon Valley
  - A household has X Macs and Y PCs
  - Can't have more than 3 Macs or 3 PCs

Y \ X	0	1	2	3	$p_Y(y)$
0	0.16	0.12	?	0.04	
1	0.12	0.14	0.12	0	
2	0.07	0.12	0	0	
3	0.04	0	0	0	
$p_X(x)$					

A Computer (or Three) In Every House

- Consider households in Silicon Valley
  - A household has X Macs and Y PCs
  - Can't have more than 3 Macs or 3 PCs

Y \ X	0	1	2	3	$p_Y(y)$
0	0.16	0.12	0.07	0.04	
1	0.12	0.14	0.12	0	
2	0.07	0.12	0	0	
3	0.04	0	0	0	
$p_X(x)$					

A Computer (or Three) In Every House

- Consider households in Silicon Valley
  - A household has X Macs and Y PCs
  - Can't have more than 3 Macs or 3 PCs

Y \ X	0	1	2	3	$p_Y(y)$
0	0.16	0.12	0.07	0.04	0.39
1	0.12	0.14	0.12	0	0.38
2	0.07	0.12	0	0	0.19
3	0.04	0	0	0	0.04
$p_X(x)$	0.39	0.38	0.19	0.04	1.00

Marginal distributions

Let's make our own!

An actual WebMD article:

Can a Pet Predict Your Personality?

1 / 25

Permutations

How many ways are there to order n distinct objects?

$$n!$$

Binomial

How many ways are there to make an unordered selection of r objects from n objects?

How many ways are there to order n objects such that:
r are the same (indistinguishable)
$(n-r)$ are the same (indistinguishable)?

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Called the "binomial" because of something from Algebra

Multinomial

How many ways are there to order n objects such that:

n_1 are the same (indistinguishable)

n_2 are the same (indistinguishable)

$\dots$

n_r are the same (indistinguishable)?

$$\frac{n!}{n_1!n_2!\dots n_r!} = \binom{n}{n_1, n_2, \dots, n_r}$$

Note: Multinomial > Binomial

Binomial Distribution

- Consider *n* independent trials of Ber(p) rand. var.
  - X is number of successes in *n* trials
  - X is a **Binomial** Random Variable: X ~ Bin(n, p)

Binomial # ways of ordering the successes

$$P(X=i)=p(i)=\binom{n}{i}p^i(1-p)^{n-i} \quad i=0,1,\ldots,n$$

Probability of exactly *i* successes

Probability of each ordering of *i* successes is equal + mutually exclusive

The Multinomial

- Multinomial distribution
  - n independent trials of experiment performed
  - Each trial results in one of m outcomes, with respective probabilities: $p_1, p_2, \ldots, p_m$ where $\sum_{i=1}^{m} p_i = 1$
  - X_i = number of trials with outcome i

$$P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} p_1^{c_1} p_2^{c_2} \ldots p_m^{c_m}$$

Joint distribution

Multinomial # ways of ordering the successes

Probabilities of each ordering are equal and mutually exclusive

where $\sum_{i=1}^{m} c_i = n$ and $\binom{n}{c_1, c_2, \ldots, c_m} = \frac{n!}{c_1! c_2! \cdots c_m!}$