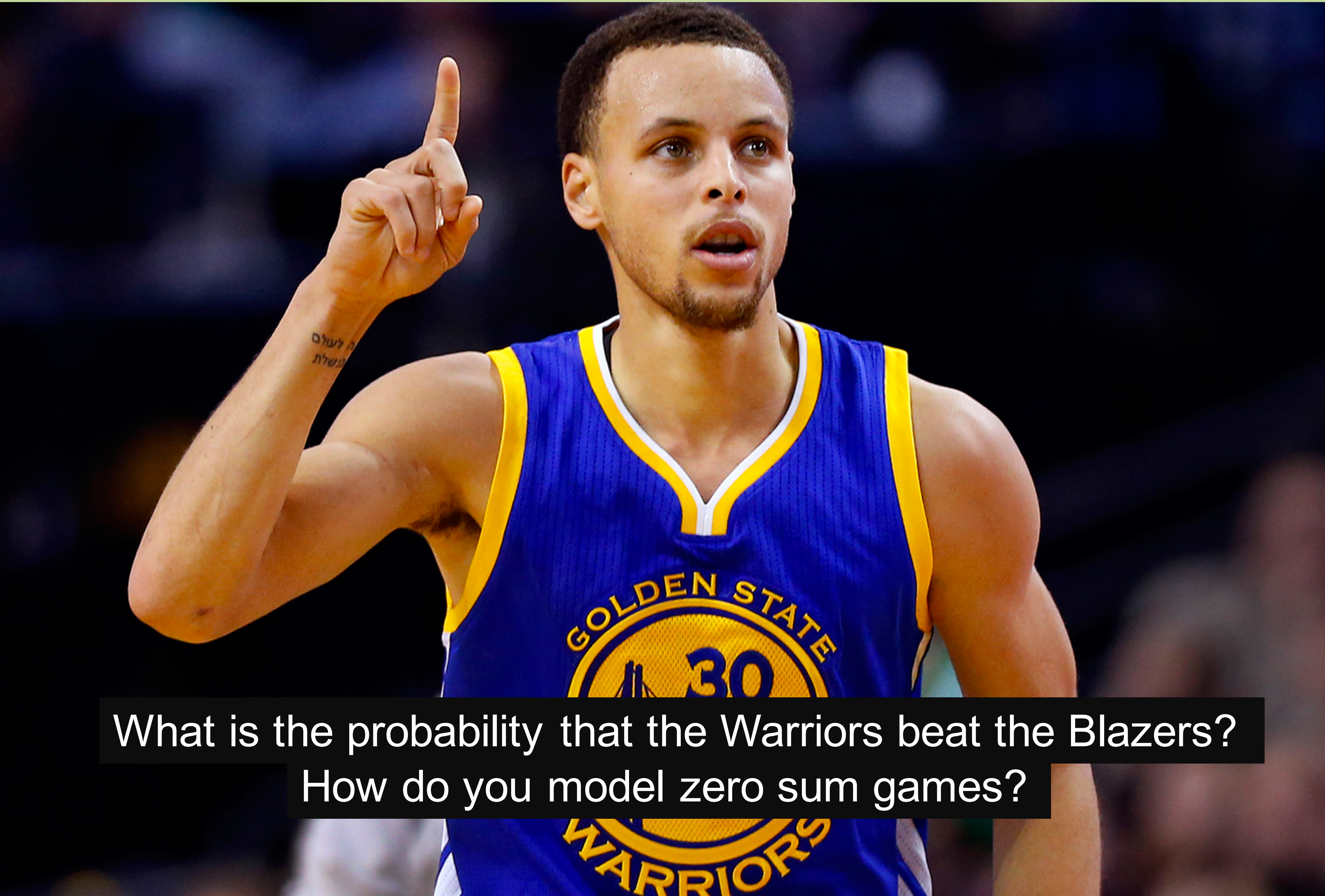
Gaussian

Chris Piech
CS109, Stanford University

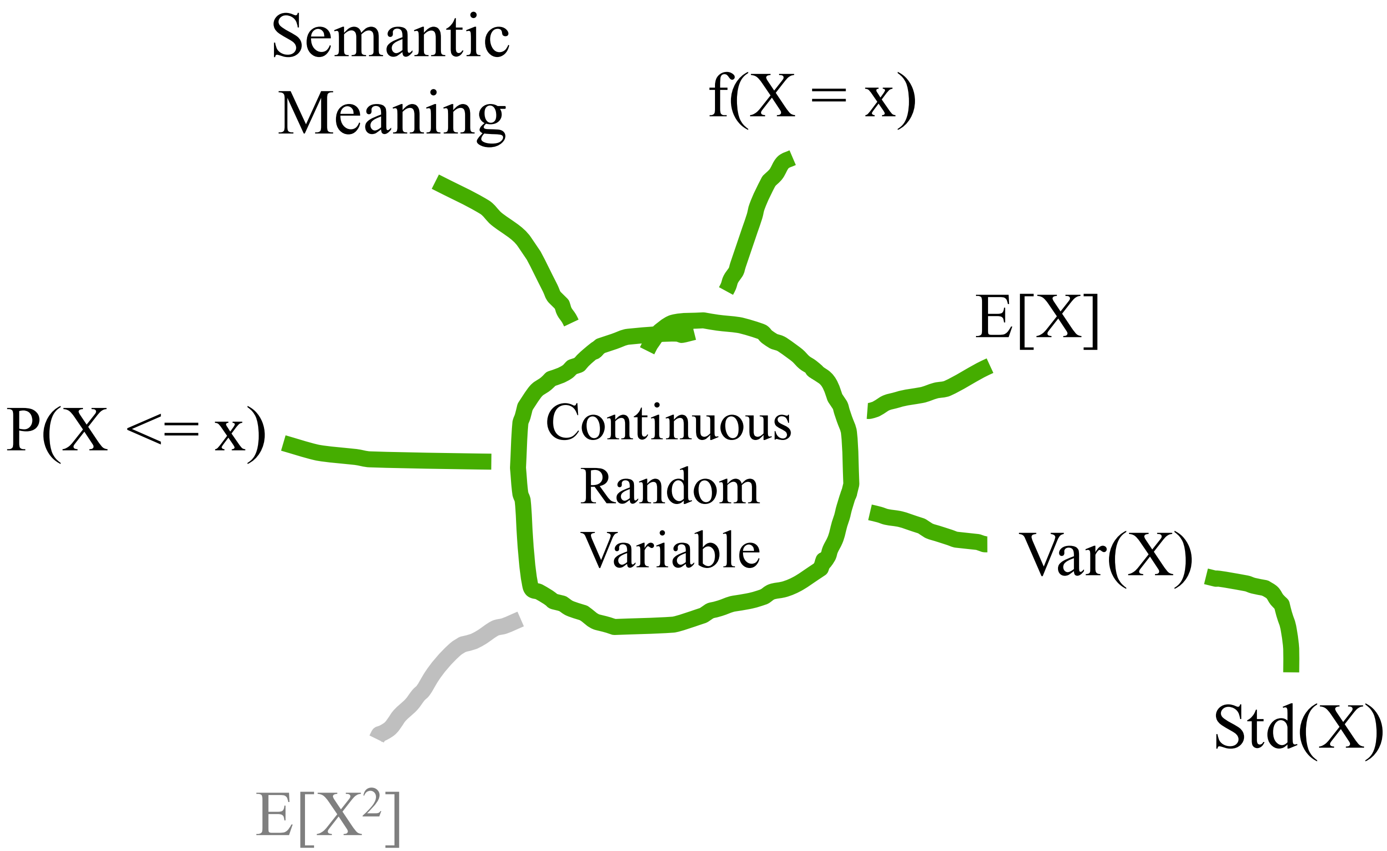
Will the Warriors Win?



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

Continuous Random Variables

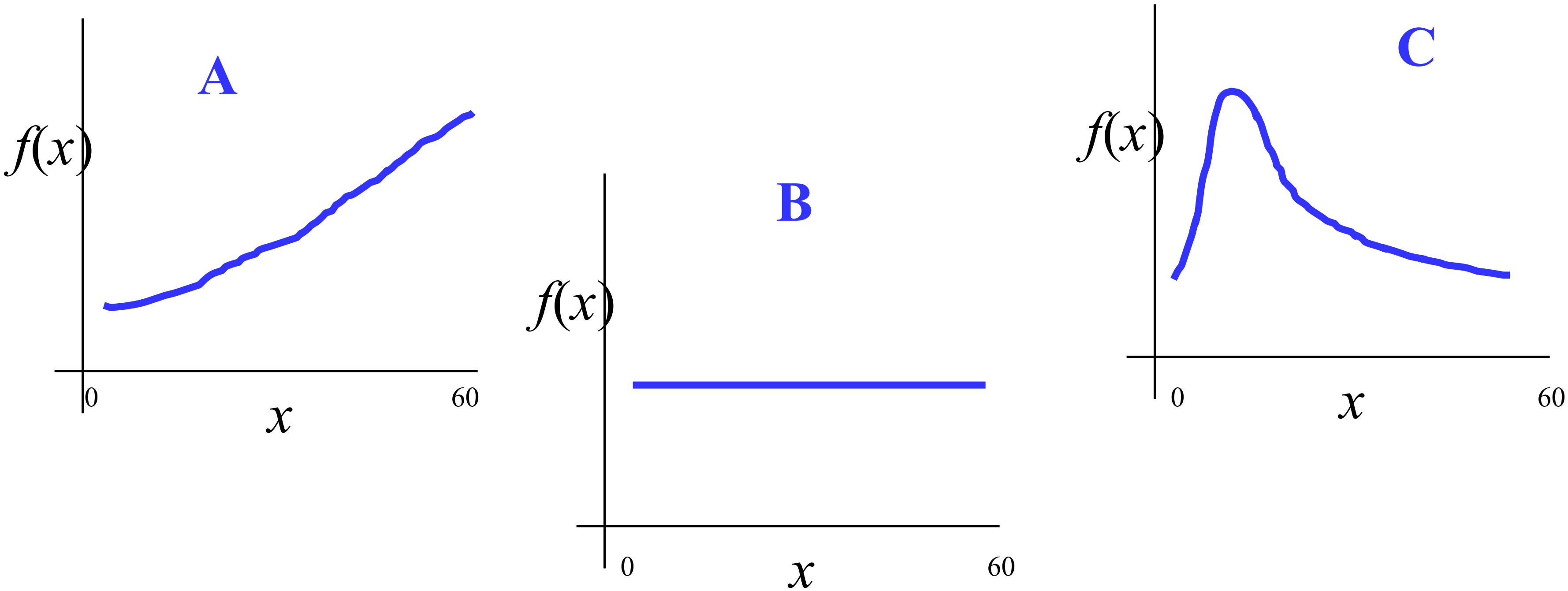
Fundamental Properties



Probability Density Function

Probability density functions articulate relative belief.

Let X be a random variable which is the # of minutes after 2pm that you arrive at a bus station:



Which of these represent that you think your arrival is more likely to be close to 3

Probability Density Function

- Say f is a **Probability Density Function** (PDF)
  - $f(x)$ is **not** a probability, it is probability/units of X
  - Answer probability questions via a subinterval over X

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

  - Because X must take on some value:

$$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

For a continuous variable, the relative likelihood of values is expressed using a “density function” (*f*)

Notation

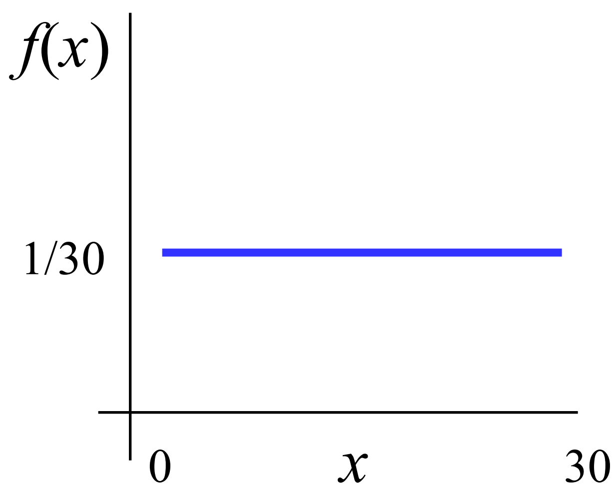
$p(a)$ or $p_X(a)$ Probability Mass Function (discrete) $P(X = a)$

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

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

Riding the Marguerite

- Say the Marguerite bus stops at the Gates bldg. at 15 minute intervals (2:00, 2:15, 2:30, etc.)
  - Passenger arrives at stop uniformly between 2-2:30pm
  - X ~ Uni(0, 30)



Riding the Marguerite

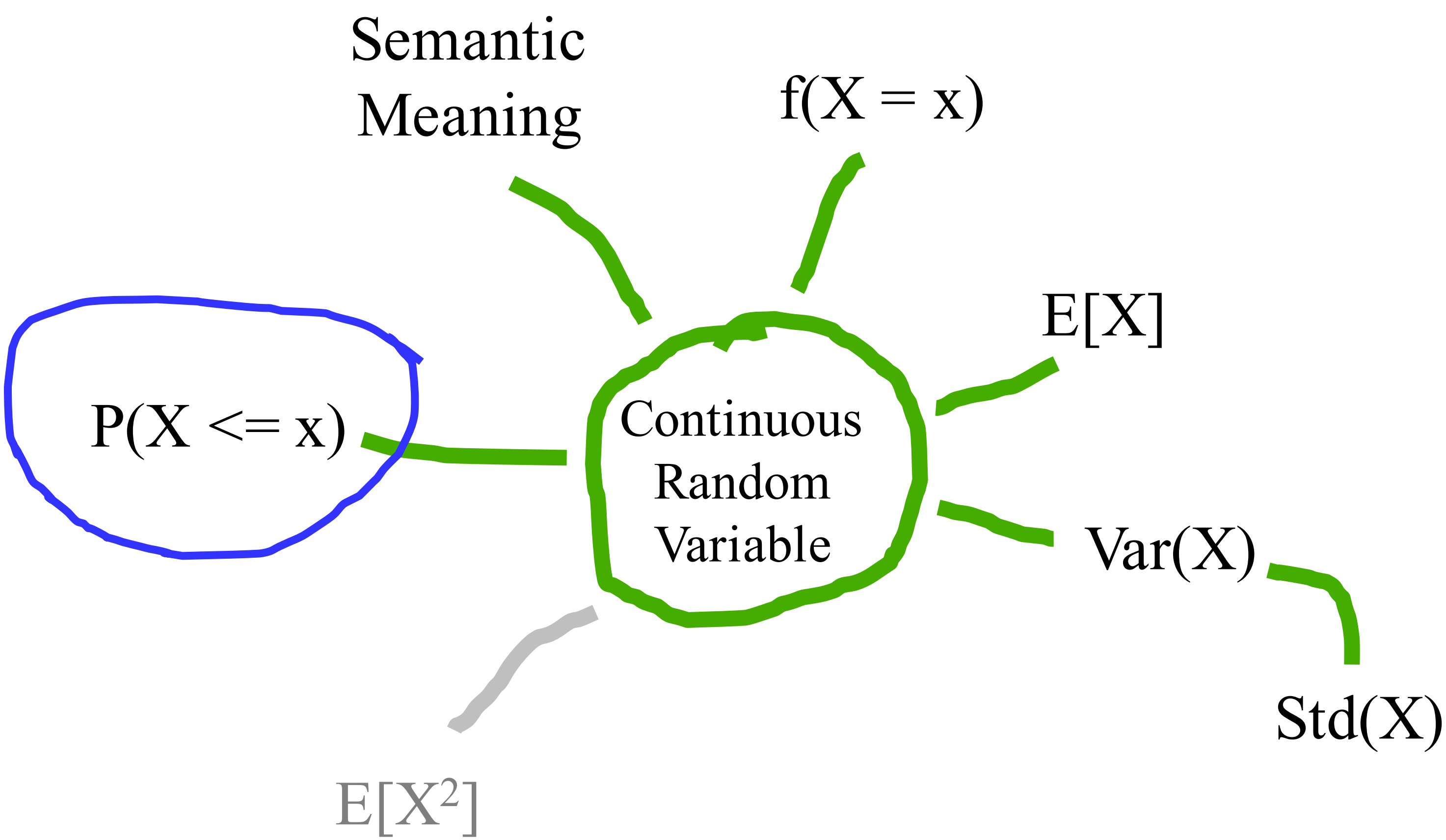
- Say the Marguerite bus stops at the Gates bldg. at 15 minute intervals (2:00, 2:15, 2:30, etc.)
  - Passenger arrives at stop uniformly between 2-2:30pm
  - X ~ Uni(0, 30)
- P(Passenger waits < 5 minutes for bus)?
  - Must arrive between 2:10-2:15pm or 2:25-2:30pm

$$P(10 < X < 15) + P(25 < x < 30) = \int_{10}^{15} \frac{1}{30}\,dx + \int_{25}^{30} \frac{1}{30}\,dx = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$$

- P(Passenger waits > 14 minutes for bus)?
  - Must arrive between 2:00-2:01pm or 2:15-2:16pm

$$P(0 < X < 1) + P(15 < x < 16) = \int_{0}^{1} \frac{1}{30}\,dx + \int_{15}^{16} \frac{1}{30}\,dx = \frac{1}{30} + \frac{1}{30} = \frac{1}{15}$$

Fundamental Properties



Cumulative Distribution Function

- For a continuous random variable X, the **Cumulative Distribution Function** (CDF) is:

$$F(a) = P(X < a) = P(X \le a) = \int_{-\infty}^{a} f(x)dx$$

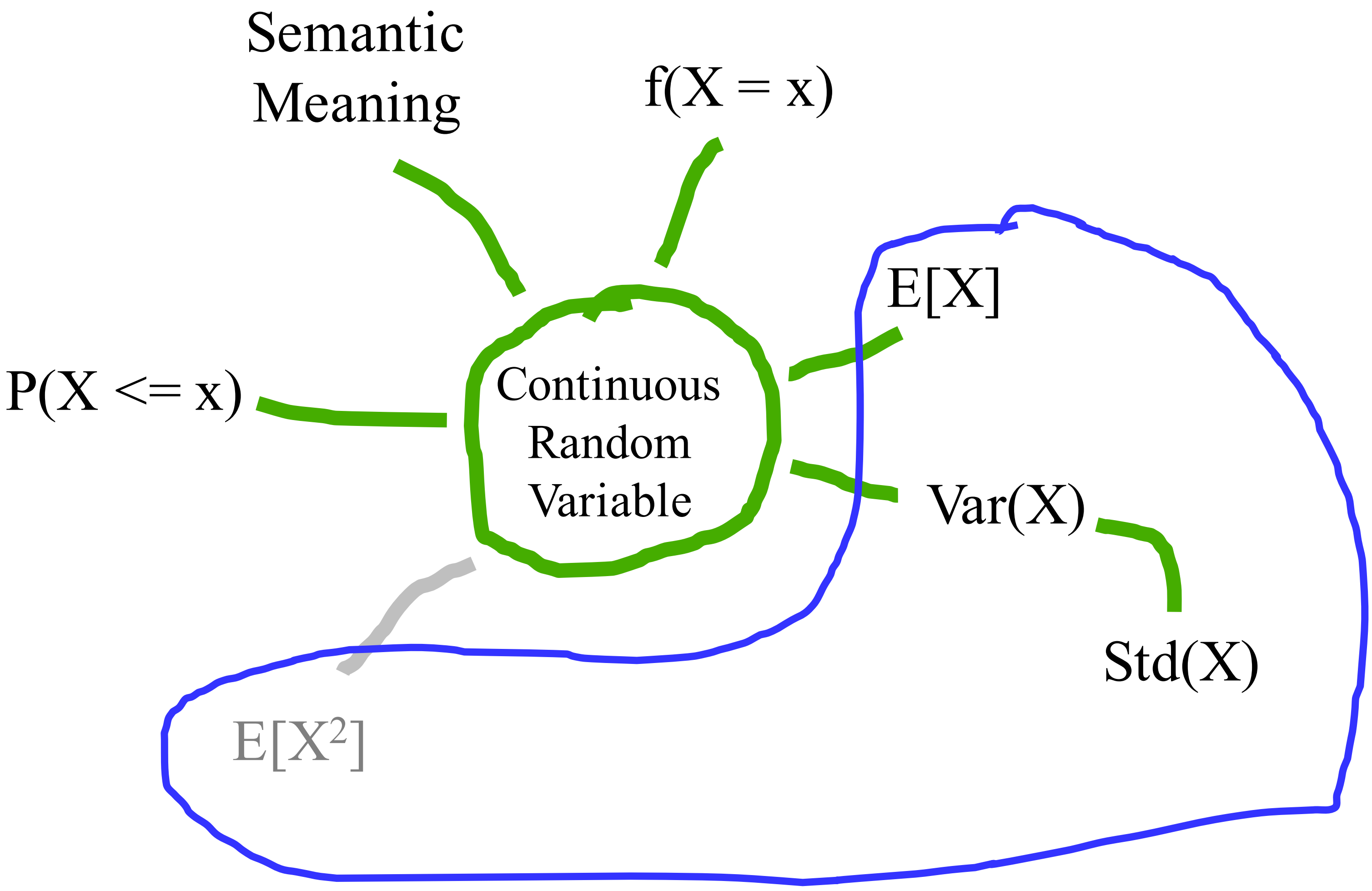
- Density f is derivative of CDF F: $f(a) = \frac{d}{da}F(a)$
- For continuous f and small ε :

$$P(a - \frac{\varepsilon}{2} \le X \le a + \frac{\varepsilon}{2}) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)dx \approx \varepsilon f(a)$$

  - So, ratio of probabilities can still be meaningful:
    - $P(X = 1)/P(X = 2) \approx (\varepsilon f(1))/(\varepsilon f(2)) = f(1)/f(2)$

For a continuous variable, most probability questions can be calculated using the Cumulative Distribution Function (*F*)

Fundamental Properties



Expectation and Variance

For discrete RV X:

$$E[X] = \sum_x x\, p(x)$$

$$\boxed{E[g(X)] = \sum_x g(x)\, p(x)}$$

$$E[X^n] = \sum_x x^n\, p(x)$$

For continuous RV X:

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx}$$

$$E[X^n] = \int_{-\infty}^{\infty} x^n\, f(x)\, dx$$

For both discrete and continuous RVs:

$$E[aX + b] = aE[X] + b$$

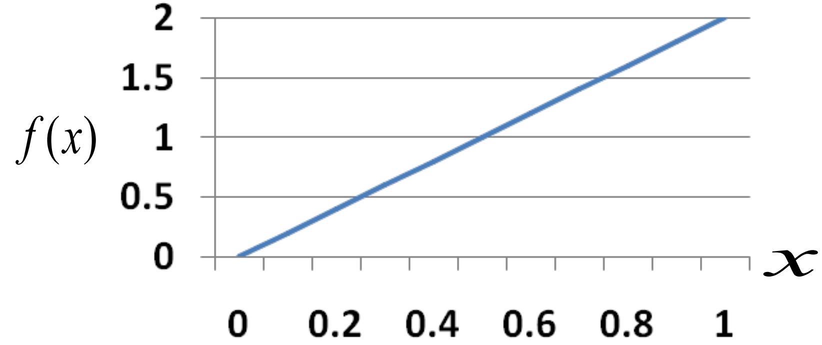
$$\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

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

Linearly Increasing Density

- X is a continuous random variable with PDF:

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



  - What is E[X]?

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)dx = \int_0^1 2x^2\, dx = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3}$$

  - What is Var(X)?

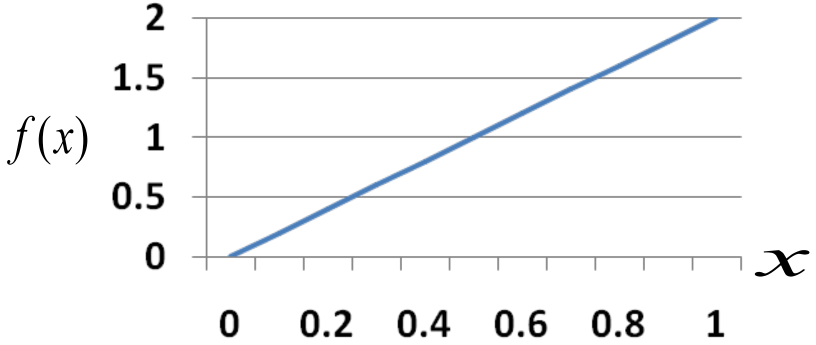
$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 2x^3\, dx = \frac{1}{2}x^4\Big|_0^1 = \frac{1}{2}$$

$$Var(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

Why 2?

- X is a continuous random variable with PDF:

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



What about $f(x) = 3x$?

Not a valid PDF

$$\int_0^1 2x\ dx = x^2|_0^1 = 1$$

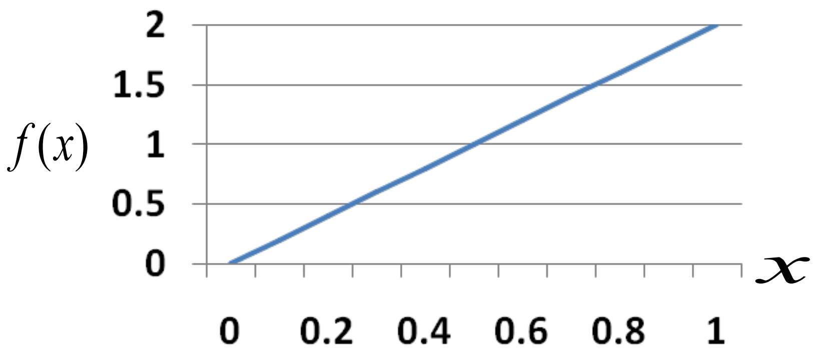
valid PDF

$$\int_0^1 3x\ dx = \frac{3}{2}x^2|_0^1 = \frac{3}{2}$$

Linearly Increasing Density

- X is a continuous random variable with PDF:

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



  - What is E[X]?

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 2x^2 dx = \frac{2}{3} x^3 \Big|_0^1 = \frac{2}{3}$$

  - What is Var(X)?

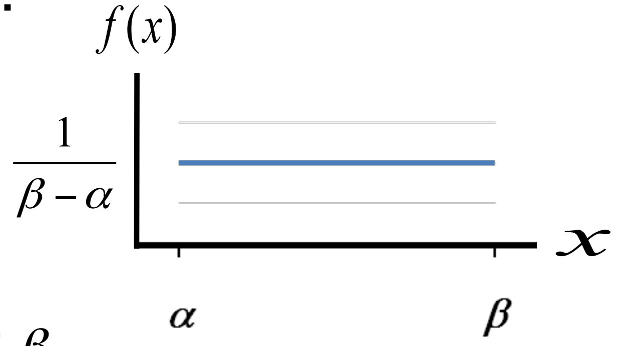
$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2} x^4 \Big|_0^1 = \frac{1}{2}$$

$$Var(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

Uniform Random Variable Revisited

- X is a **Uniform Random Variable**: X ~ Uni(α, β)
  - Probability Density Function (PDF):

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$



    - Sometimes defined over range $\alpha < x < \beta$

  - $P(a \le x \le b) = \int_a^b f(x)dx = \frac{b-a}{\beta-\alpha}$ (for $\alpha \le a \le b \le \beta$)
  - $E[X] = \int_{-\infty}^{\infty} x f(x)dx = \int_{\alpha}^{\beta} \frac{x}{\beta-\alpha} dx = \frac{x^2}{2(\beta-\alpha)}\Bigg|_{\alpha}^{\beta} = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \frac{\alpha+\beta}{2}$
  - $Var(X) = \frac{(\beta-\alpha)^2}{12}$

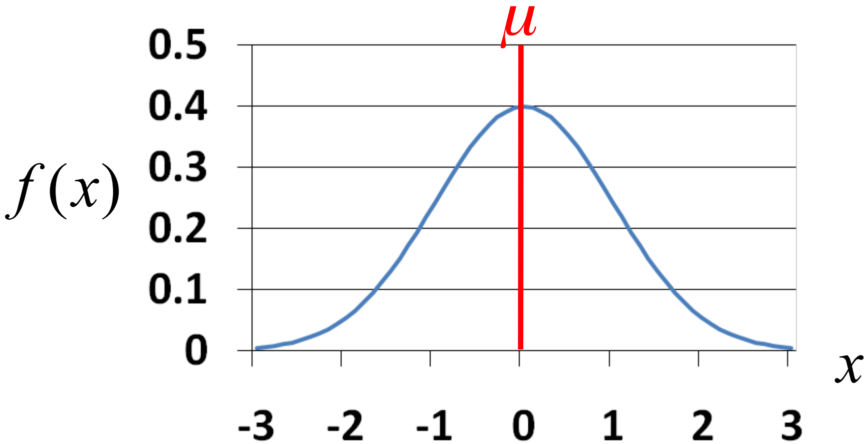
Big Day

The Normal Distribution

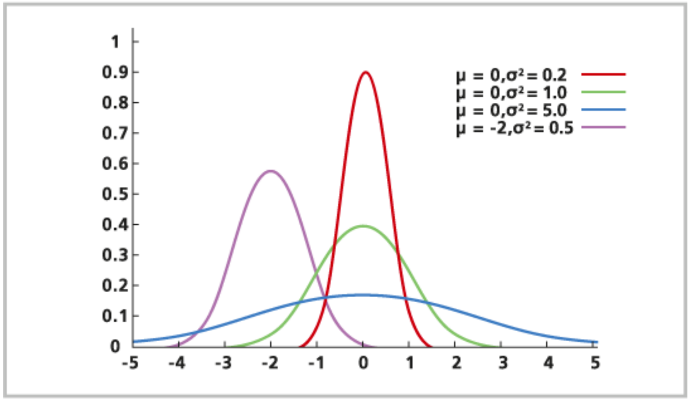
- X is a **Normal Random Variable**: $X \sim N(\mu, \sigma^2)$
  - Probability Density Function (PDF):

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

  - $E[X] = \mu$
  - $Var(X) = \sigma^2$



  - Also called "Gaussian"
  - Note: $f(x)$ is symmetric about μ



Why use the normal?

- Common for natural phenomena: heights, weights, etc.
- Often results from the sum of multiple variables
- Most noise is Normal.
- Sample means are distributed normally.

Or that is what they want you to believe

But I Encourage you to be Critical

These are log-normal

- Common for natural phenomena: heights, weights, etc.

Only if they are equally weighted

- Often results from the sum of multiple variables

Most noise is assumed normal

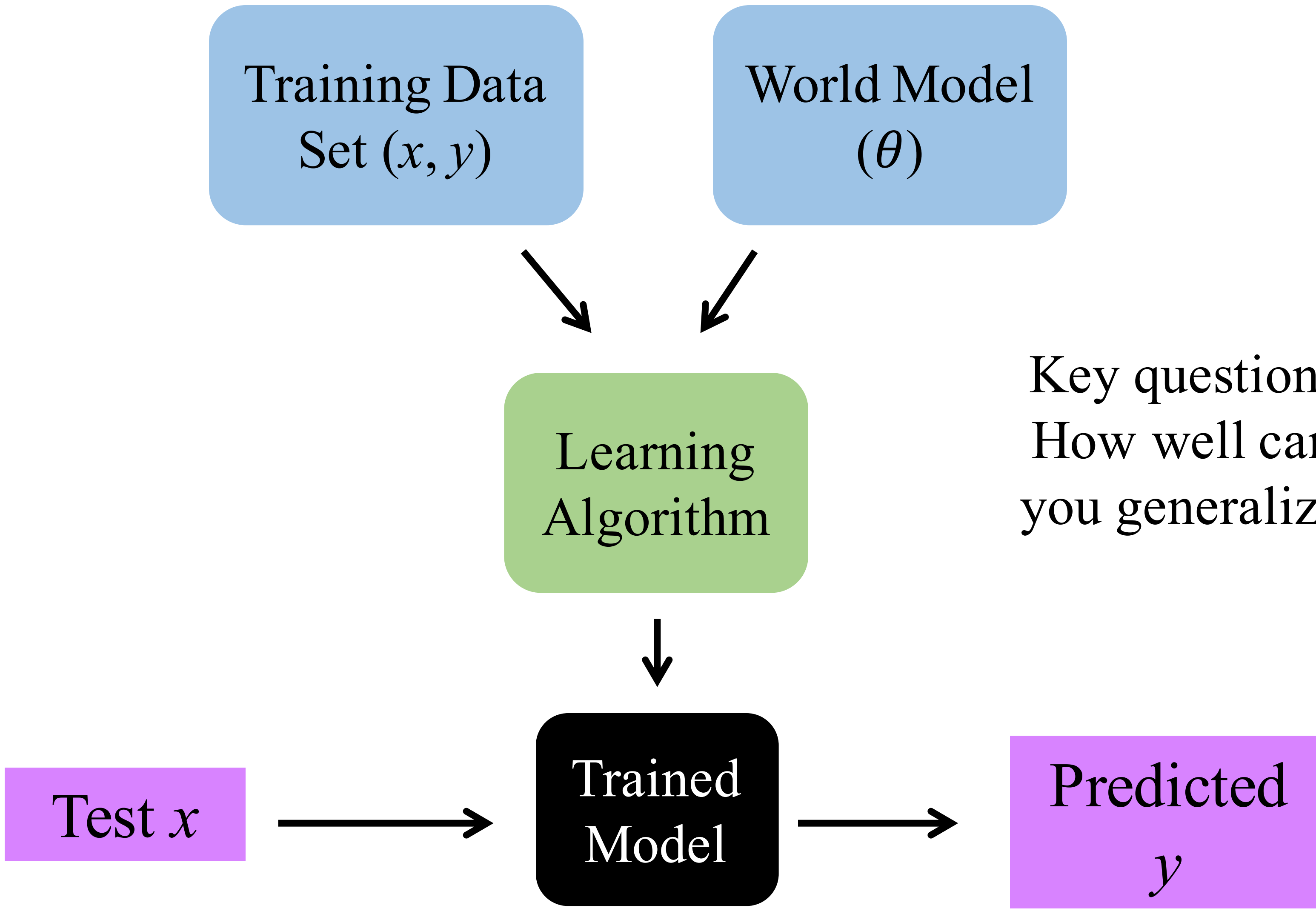
- Most noise is Normal.

- Sample means are distributed normally.

It is the most important distribution

Because of a deeper truth…

Supervised Learning

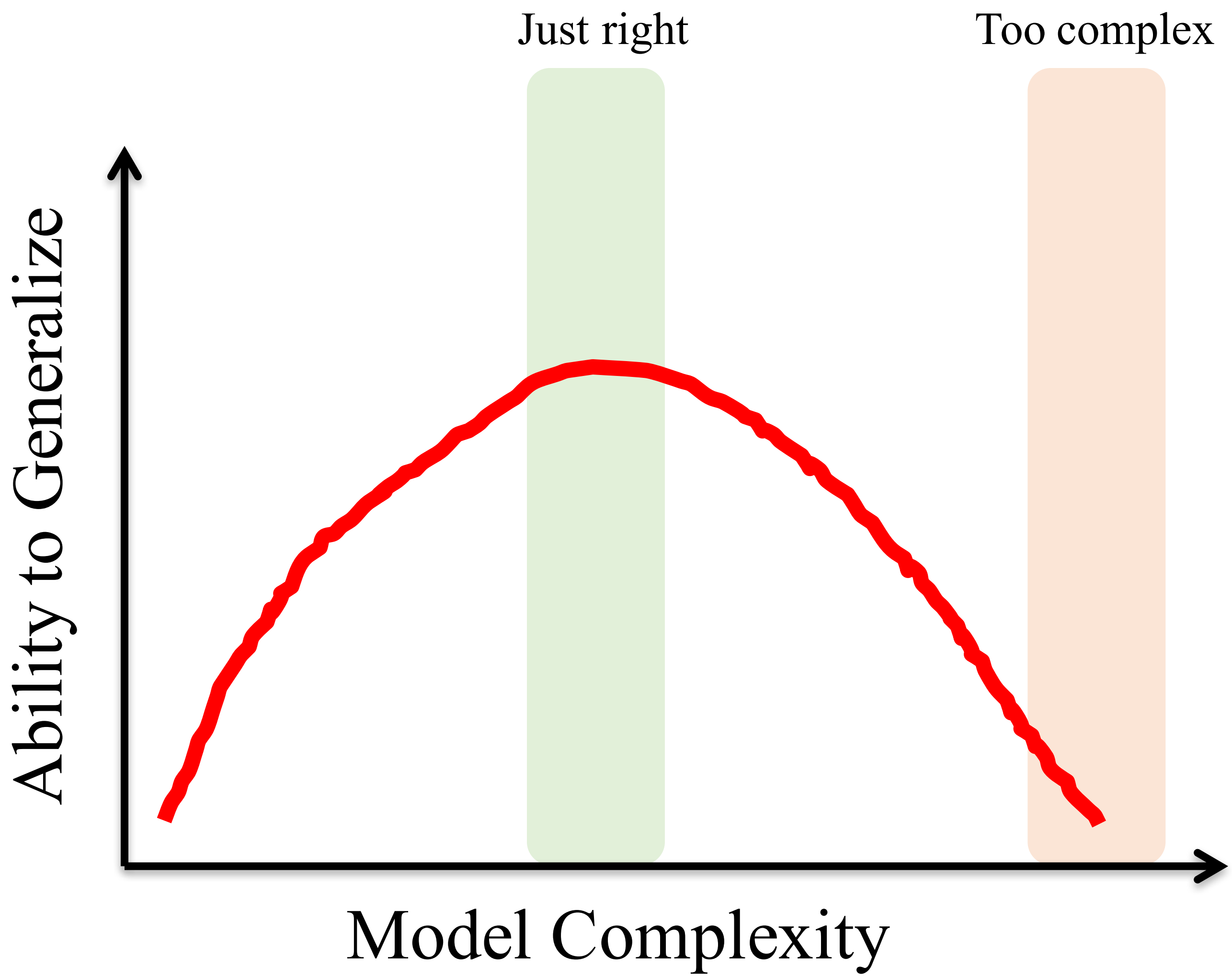


Key question: How well can you generalize

Occam’s Razor

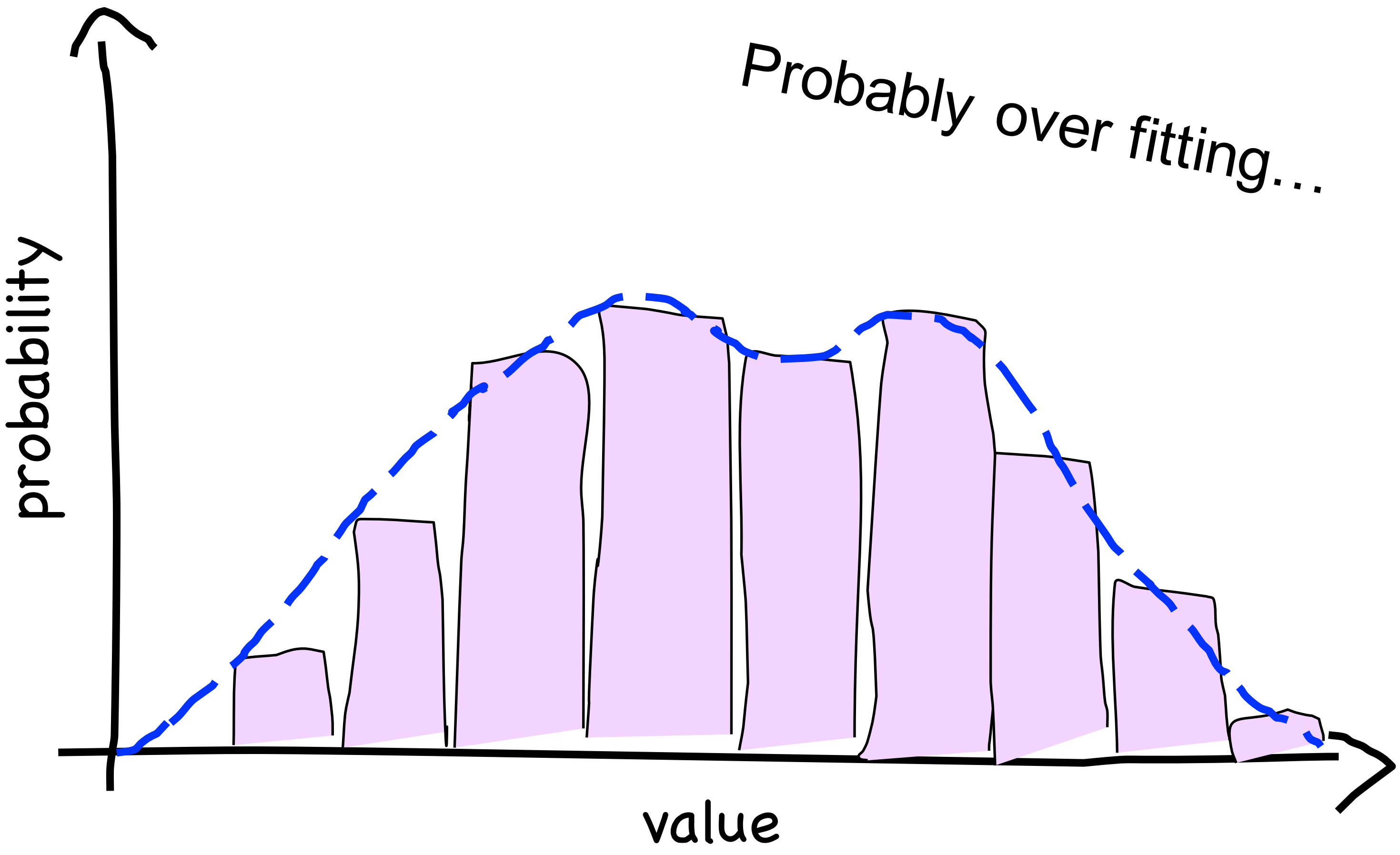
“The simplest explanation is usually the best one”

Overfiting



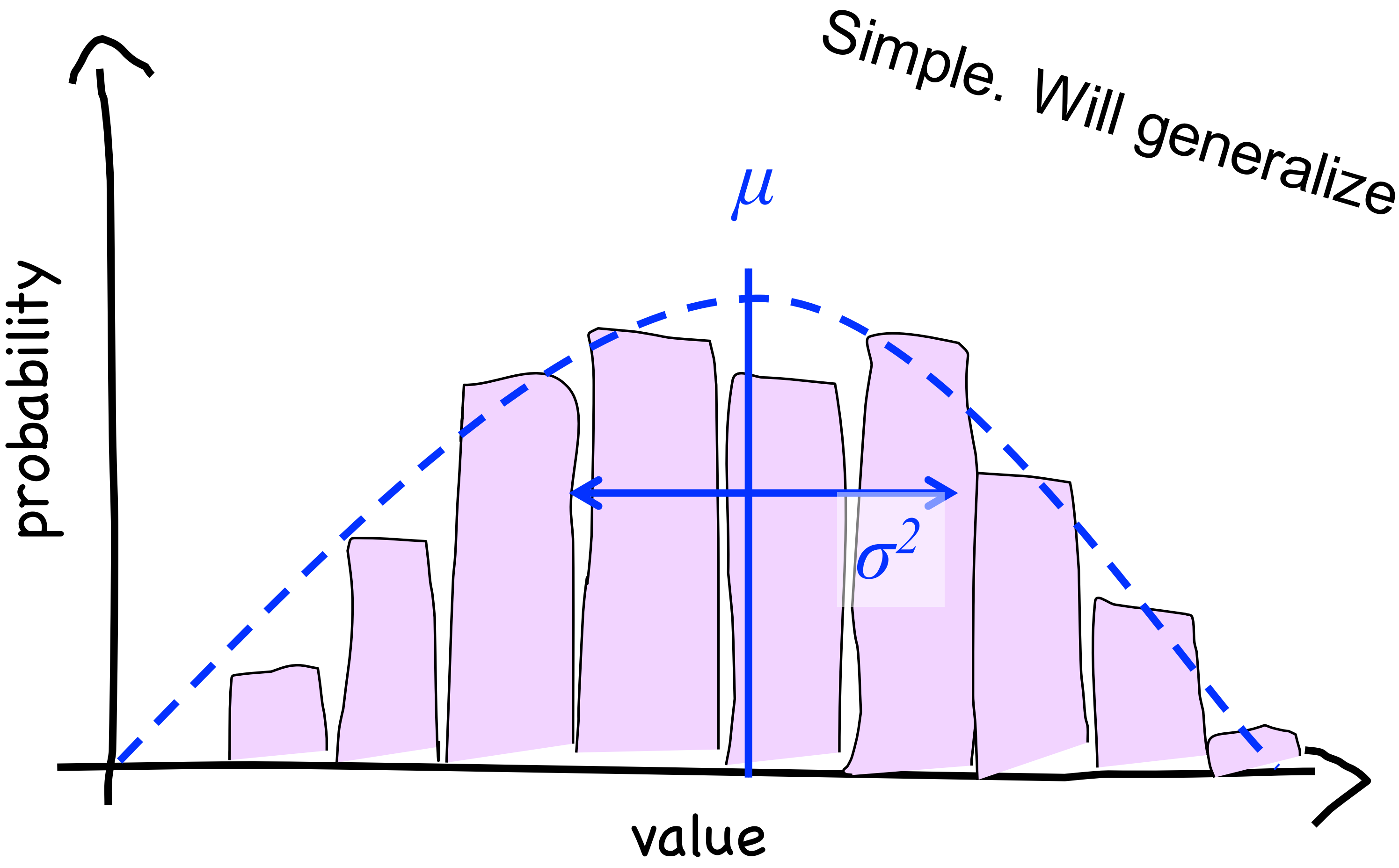
Just right
Too complex
Ability to Generalize
Model Complexity

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

Carl Friedrich Gauss

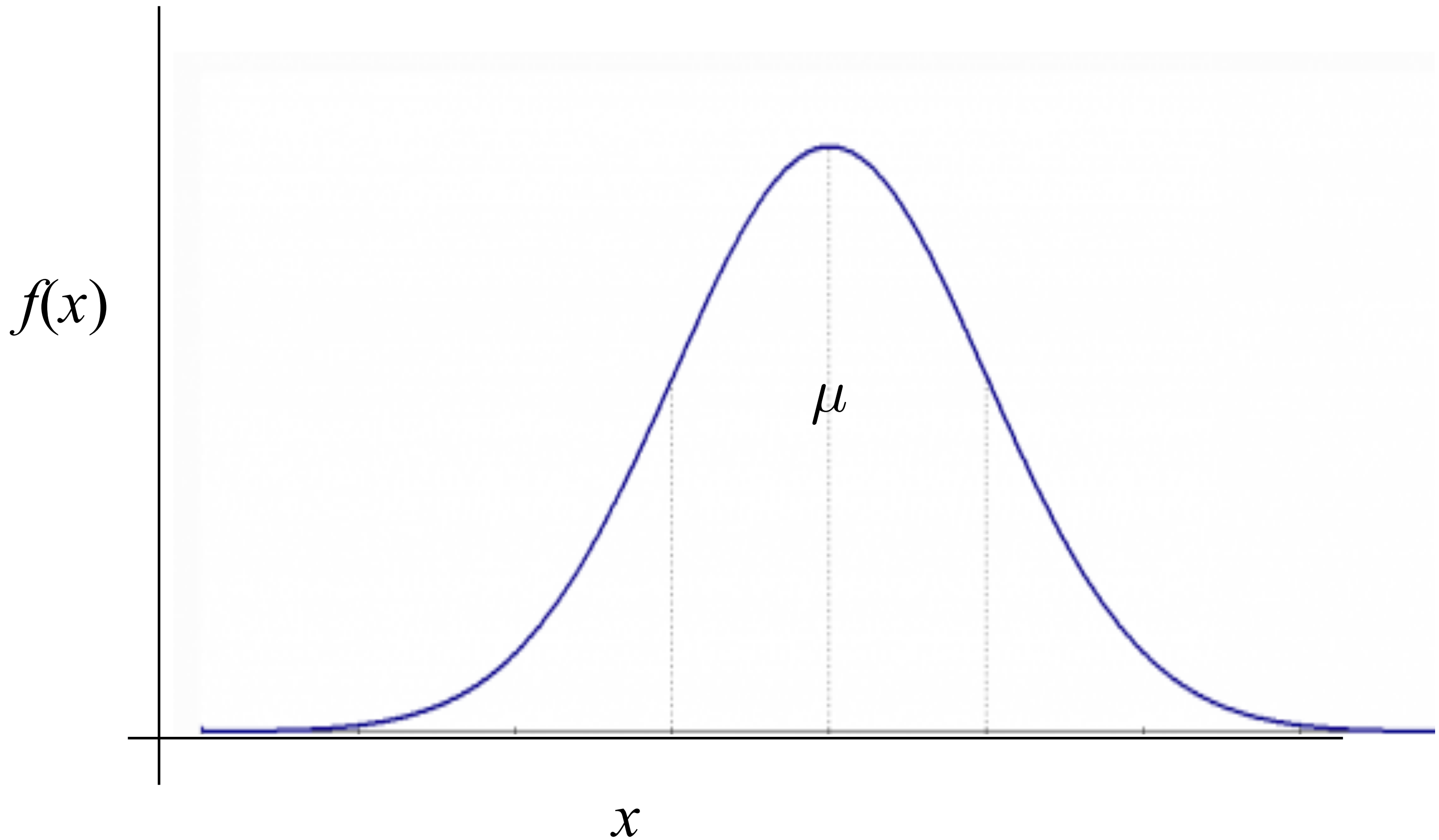
- Carl Friedrich Gauss (1777-1855) was a remarkably influential German mathematician

- Started doing groundbreaking math as teenager
  - Did not invent Normal distribution, but popularized it
- He looked like Martin Sheen
  - Who is, of course, Charlie Sheen’s father

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

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

"exponential"

the distance to the mean

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

probability density at x

a constant

sigma shows up twice

Let's try and integrate it!

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

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

* Call me if you find an equation for this

No closed form for the integral

No closed form for $F(x)$

Spoiler Alert

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

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

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

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

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

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

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

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:

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

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

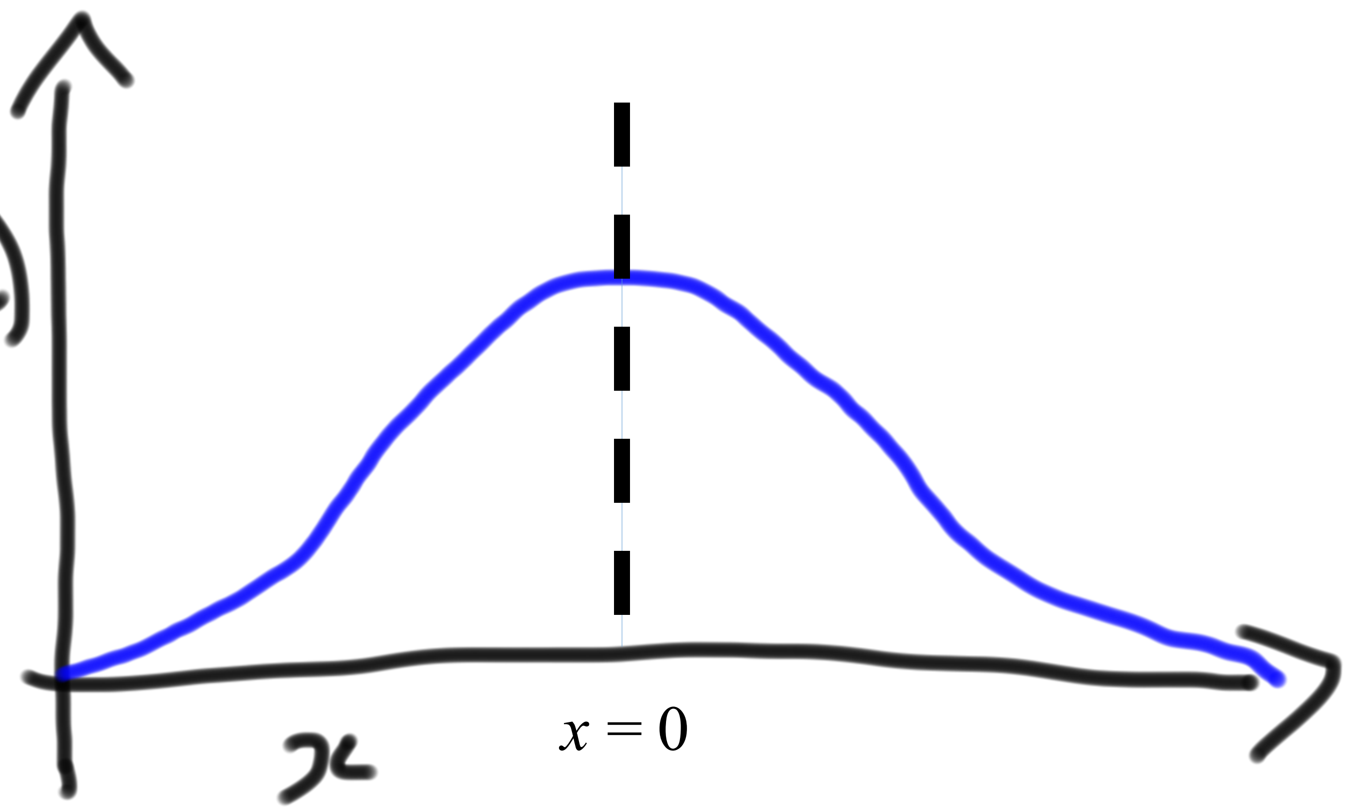
Standard (Unit) Normal Variable

- Z is a **<u>Standard (or Unit) Normal RV</u>**: $Z \sim \mathrm{N}(0, 1)$
  - $\mathrm{E}[\mathrm{Z}] = m = 0$ $\quad$ $\mathrm{Var}(\mathrm{Z}) = s^2 = 1$ $\quad$ $\mathrm{SD}(\mathrm{Z}) = s = 1$
  - CDF of Z, $F_Z(z)$ does not have closed form
  - We denote $F_Z(z)$ as $\Phi(z)$: "phi of z"

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

  - By symmetry: $\Phi(-z) = \mathrm{P}(Z \leq -z) = \mathrm{P}(Z \geq z) = 1 - \Phi(z)$

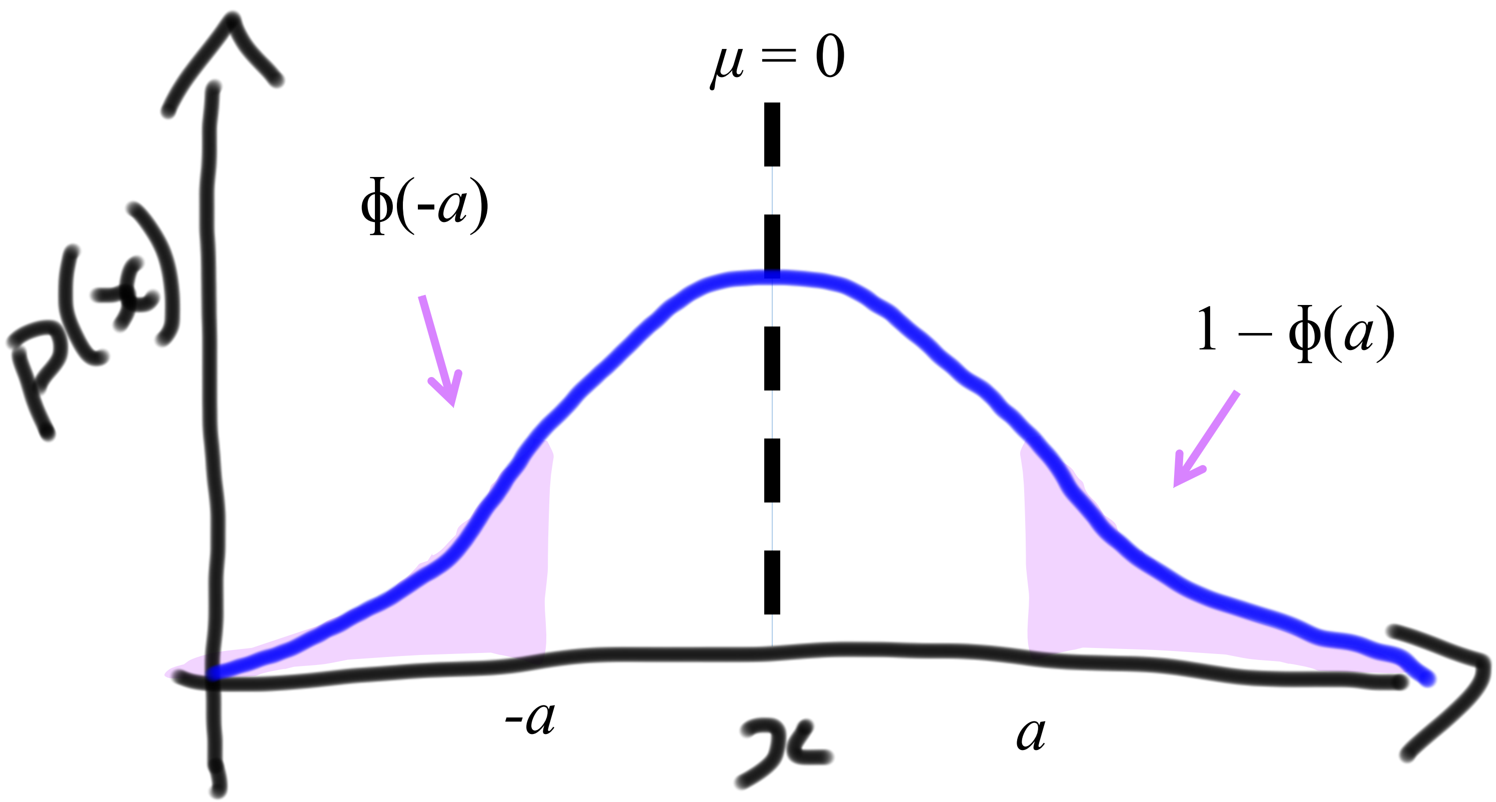
The Standard Normal



*This is the probability density function for the standard normal

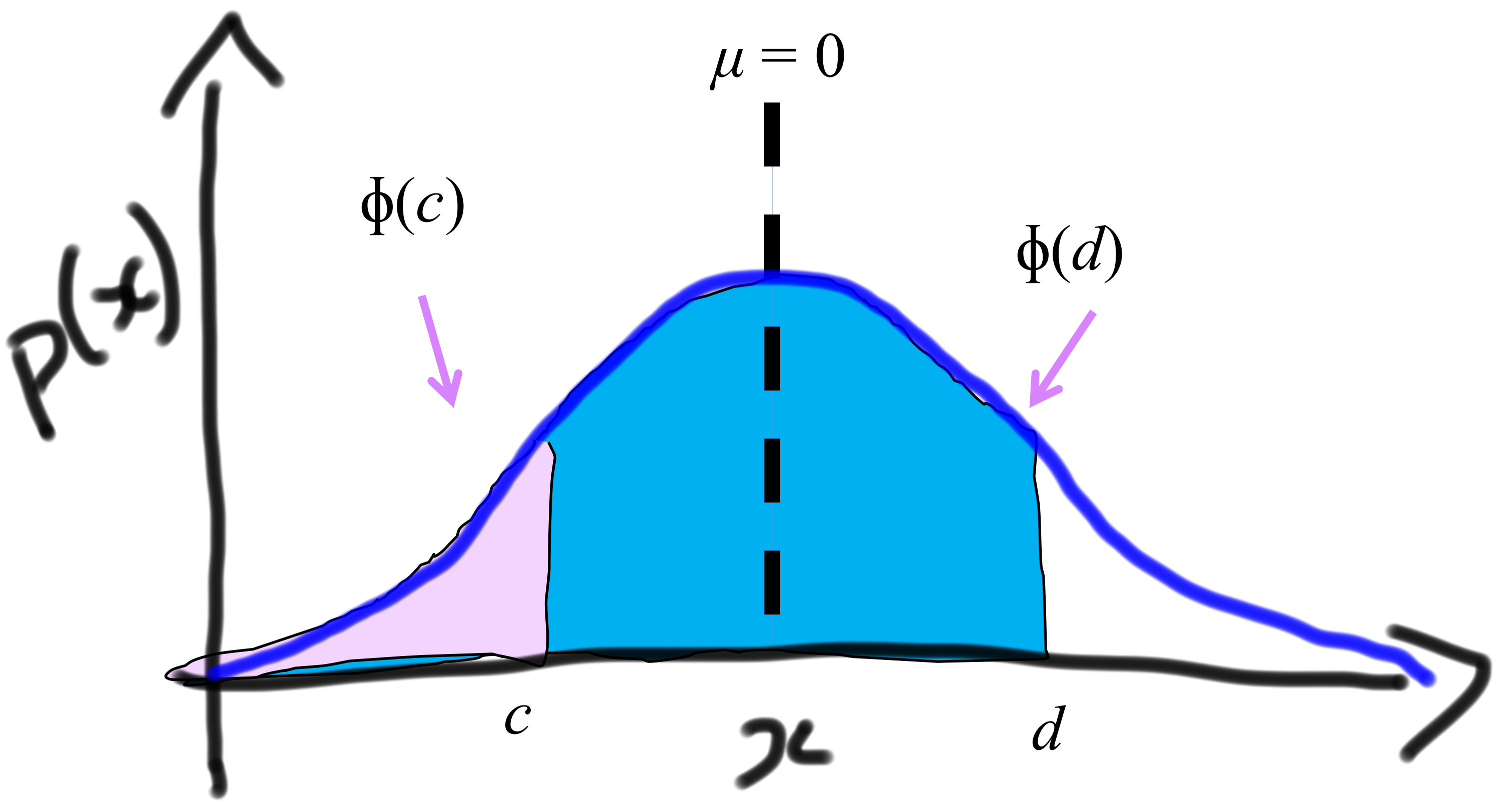
Symmetry of Phi

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



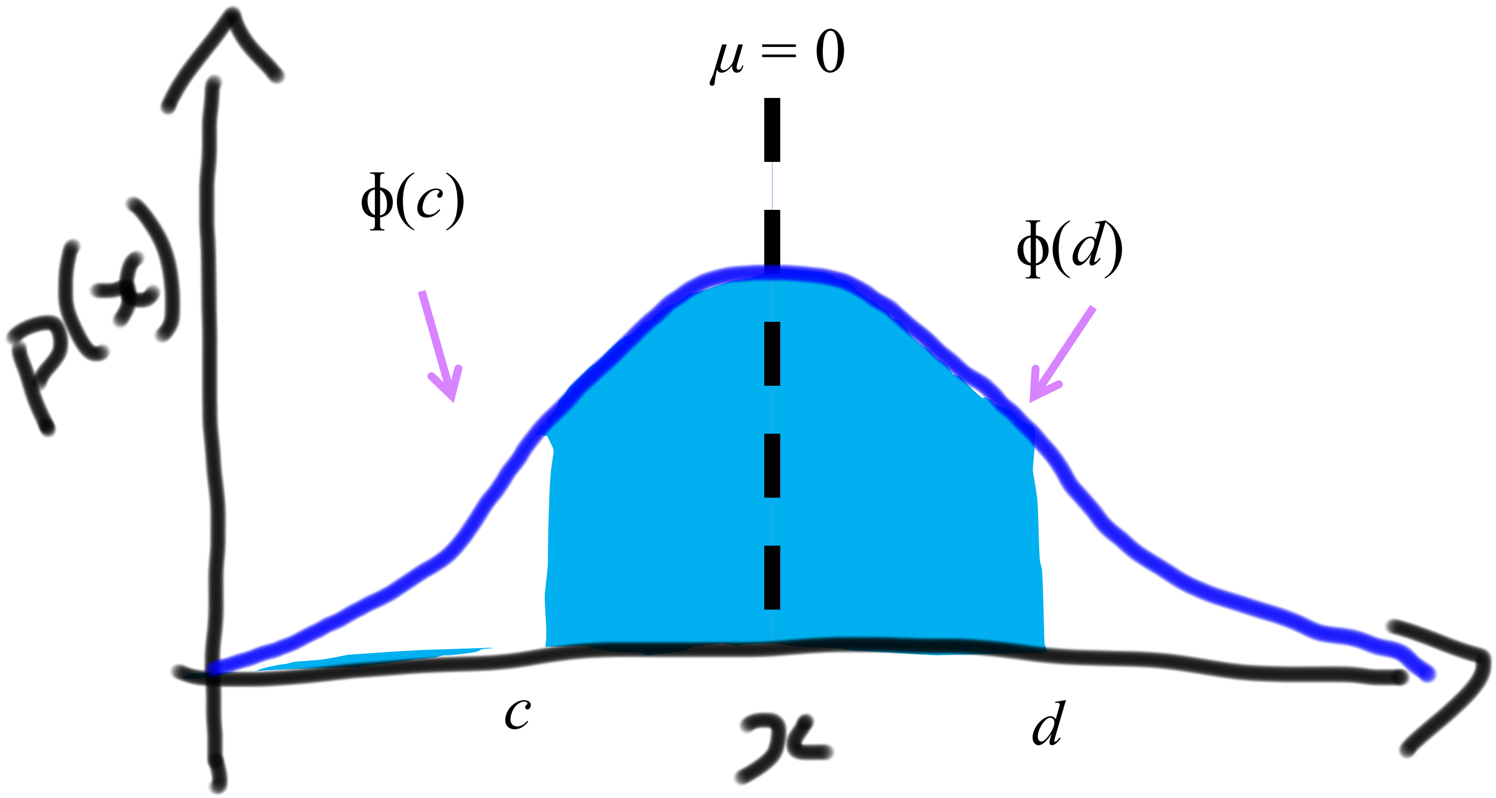
*This is the probability density function for the standard normal

Interval of Phi



Interval of Phi

$$\phi(d) - \phi(c)$$



Compute F(x) via Transform

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

Use Z to compute $F(x)$

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

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

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

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

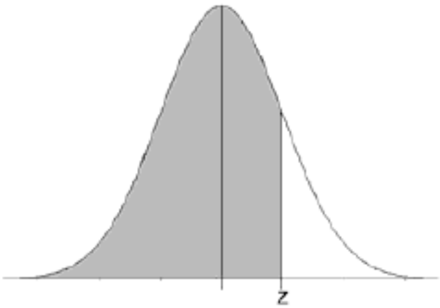
The cumulative density function (CDF) of any normal

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	$\boxed{0.7054}$	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

Kinda old school

Using Programming Library

Every modern programming language has a normal library

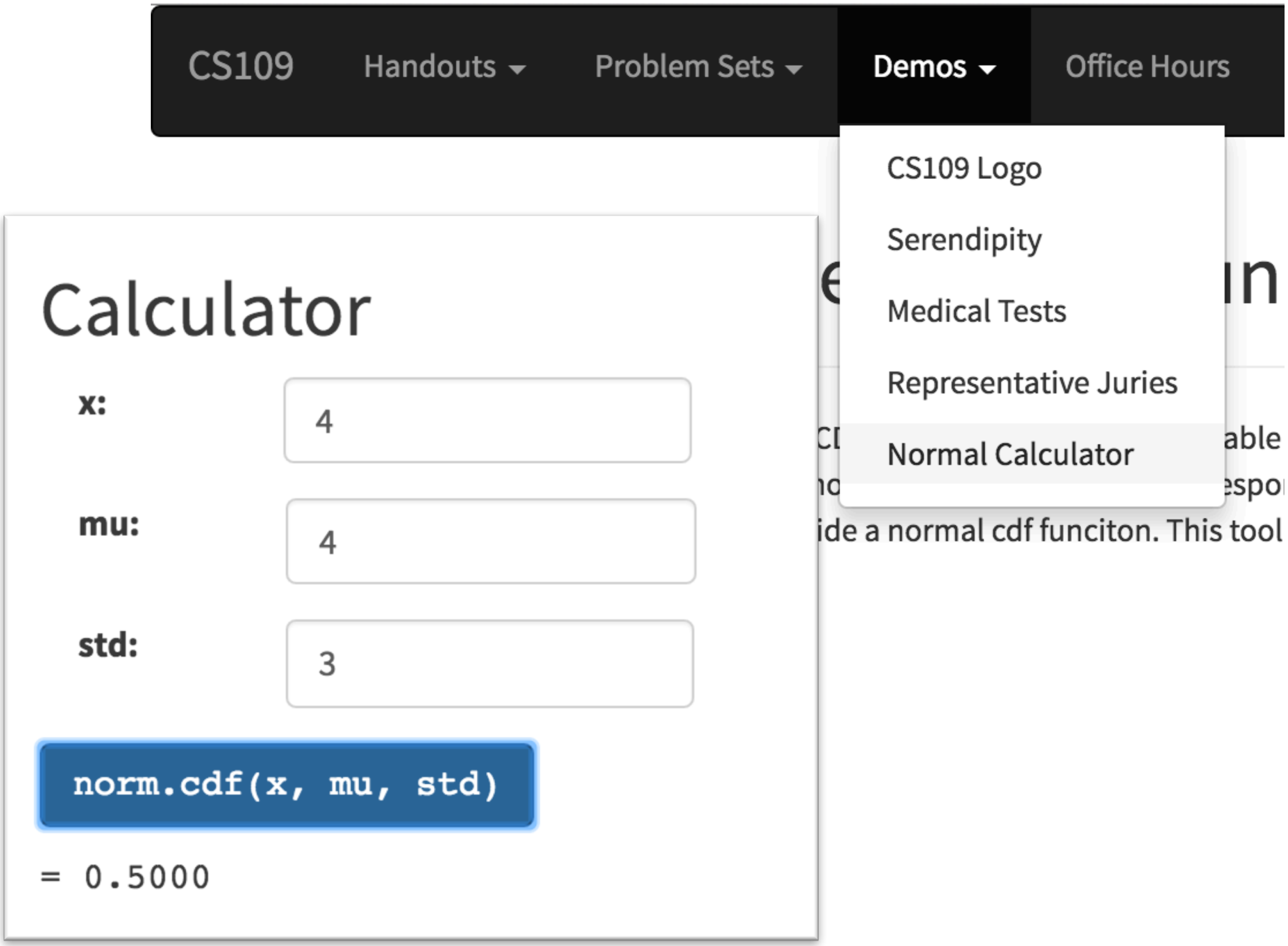
```
norm.cdf(x, mean, std)
```

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

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

* This is from Python's scipy library

I made one for you



Get Your Gaussian On

- X ~ N(3, 16) μ = 3 σ^2 = 16 σ = 4
  - What is P(X > 0)?

$$P(X > 0) = P(\frac{X-3}{4} > \frac{0-3}{4}) = P(Z > -\frac{3}{4}) = 1 - P(Z \leq -\frac{3}{4})$$

$$1 - \Phi(-\tfrac{3}{4}) = 1 - (1 - \Phi(\tfrac{3}{4})) = \Phi(\tfrac{3}{4}) = 0.7734$$

  - What is P(2 < X < 5)?

$$P(2 < X < 5) = P(\frac{2-3}{4} < \frac{X-3}{4} < \frac{5-3}{4}) = P(-\frac{1}{4} < Z < \frac{2}{4})$$

$$\Phi(\tfrac{2}{4}) - \Phi(-\tfrac{1}{4}) = \Phi(\tfrac{1}{2}) - (1 - \Phi(\tfrac{1}{4})) = 0.6915 - (1 - 0.5987) = 0.2902$$

  - What is P(|X – 3| > 6)?

$$P(X < -3) + P(X > 9) = P(Z < \frac{-3-3}{4}) + P(Z > \frac{9-3}{4})$$

$$\Phi(-\tfrac{3}{2}) + (1 - \Phi(\tfrac{3}{2})) = 2(1 - \Phi(\tfrac{3}{2})) = 2(1 - 0.9332) = 0.1337$$

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)?

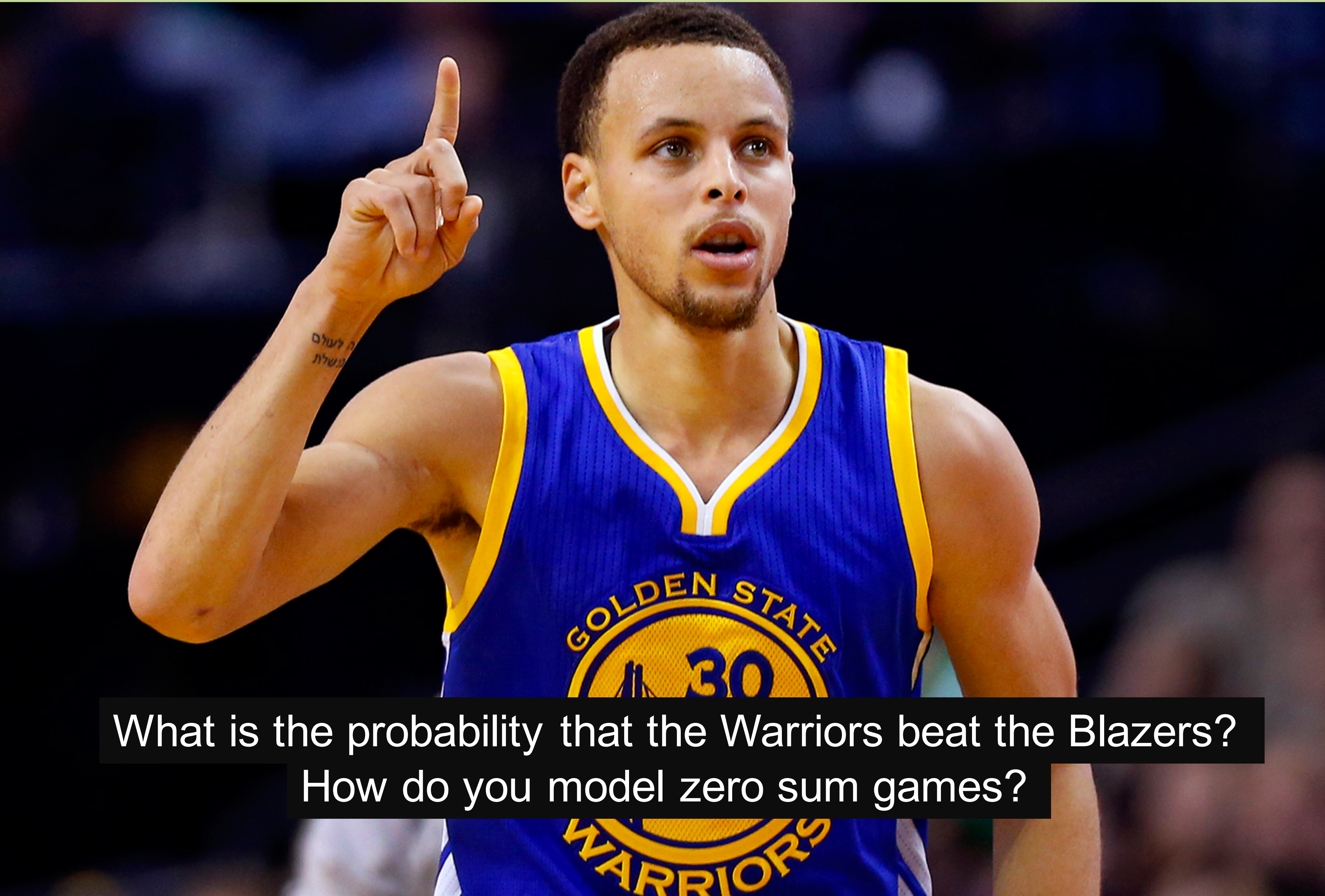
$$P(2+Y<0.5)=P(Y<-1.5)=\Phi(-1.5)=1-\Phi(1.5)\approx 0.0668$$

  - 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 beat the Blazers?

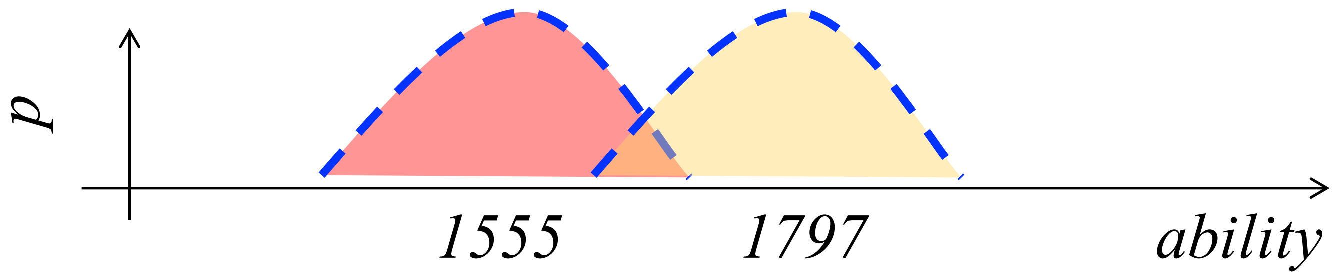
How do you model zero sum games?

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

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



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

ELO Ratings

```
from random import *

WARRIORS_ELO = 1797
BLAZERS_ELO = 1555
VAR = 200 * 200

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

print float(nSuccess) / NTRIALS
```

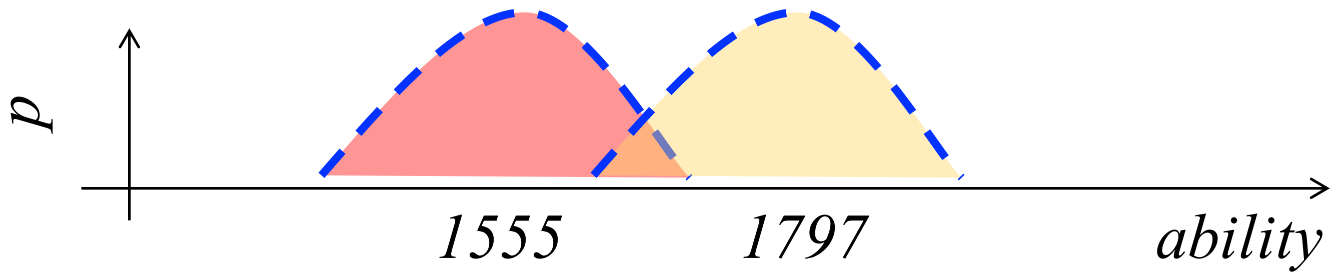
ELO Ratings

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

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



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

$$\approx 0.87$$

Calculated via sampling

Poll of polls?

French Elections:

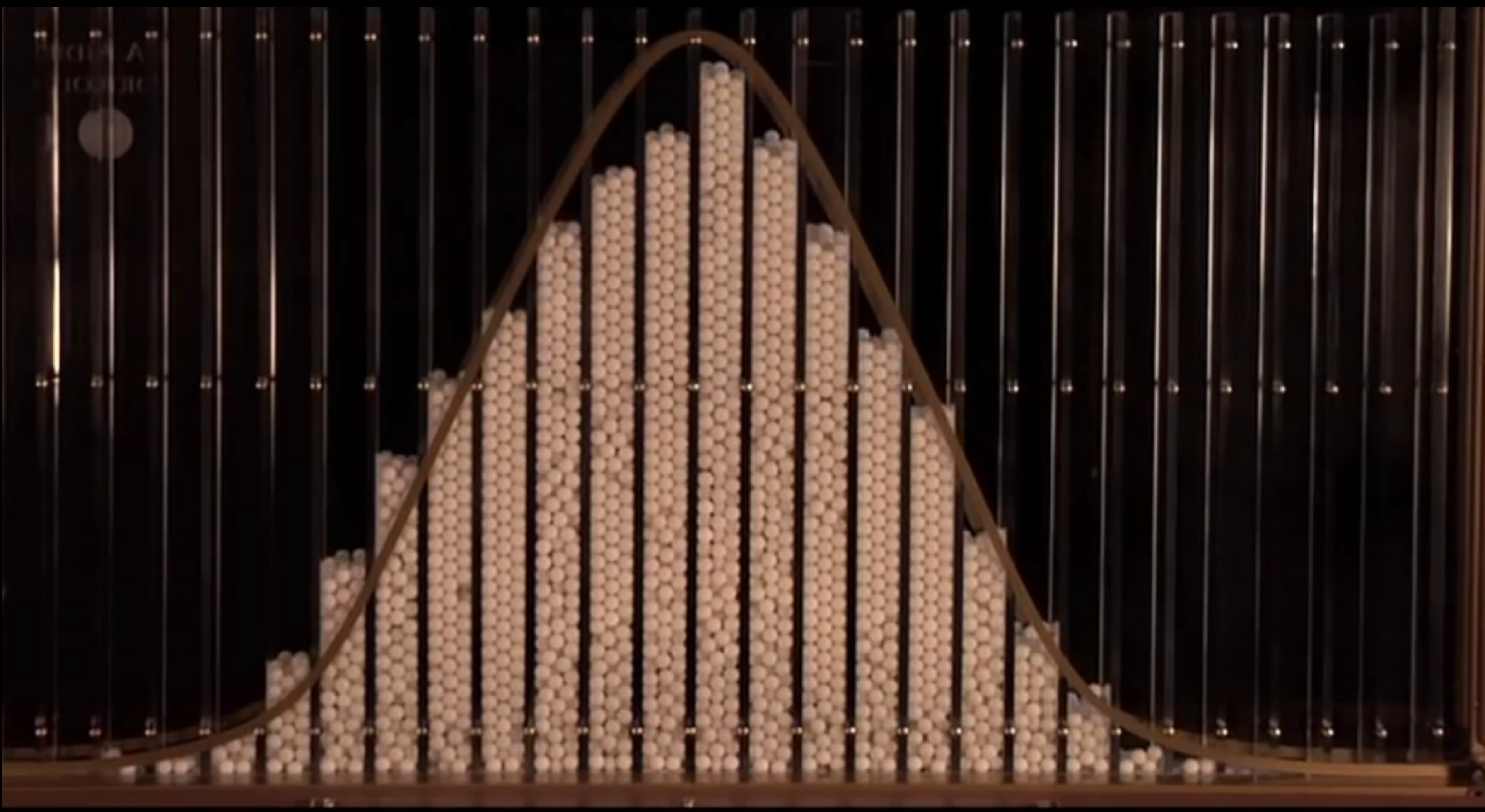
Poll	Le Pen	Macron	Fillon	Hamon	Mélenchon	Other
Ipsos/CEVIPOF/LeMonde Apr 19 – Apr 20 2,048 Registered Voters	22	**24**	19	8	19	10
Elabe/BFMTV/L'Express Apr 19 – Apr 20 1,445 Registered Voters	22	**24**	20	7	20	9
OpinionWay/ORPI/Radio Classíque/Les Echos Apr 18 – Apr 20 2,269 Registered Voters	22	**23**	21	8	18	7
Harris/France Télévisions/L'emission Politique Apr 18 – Apr 20 962 Registered Voters	21	**25**	20	8	19	9
BVA/Orange/Presse Regionale Apr 18 – Apr 19 1,427 Registered Voters	23	**24**	19	9	19	8

Credit: fivethirtyeight.com

What is the probability that Le Pen / Macron wins?

Gaussian for Binomial

Remember this?



There is a deep reason for the Binomial/Normal approximation…

Normal Approximation of Binomial

- X ~ Bin(n, p)
  - E[X] = np Var(X) = $np(1 - p)$
  - Poisson approx. good: n large (> 20), p small (< 0.05)
  - For large n: X ≈ Y ~ N(E[X], Var(X)) = N(np, $np(1 - p)$)
  - Normal approx. good : Var(X) = $np(1 - p) \geq 10$

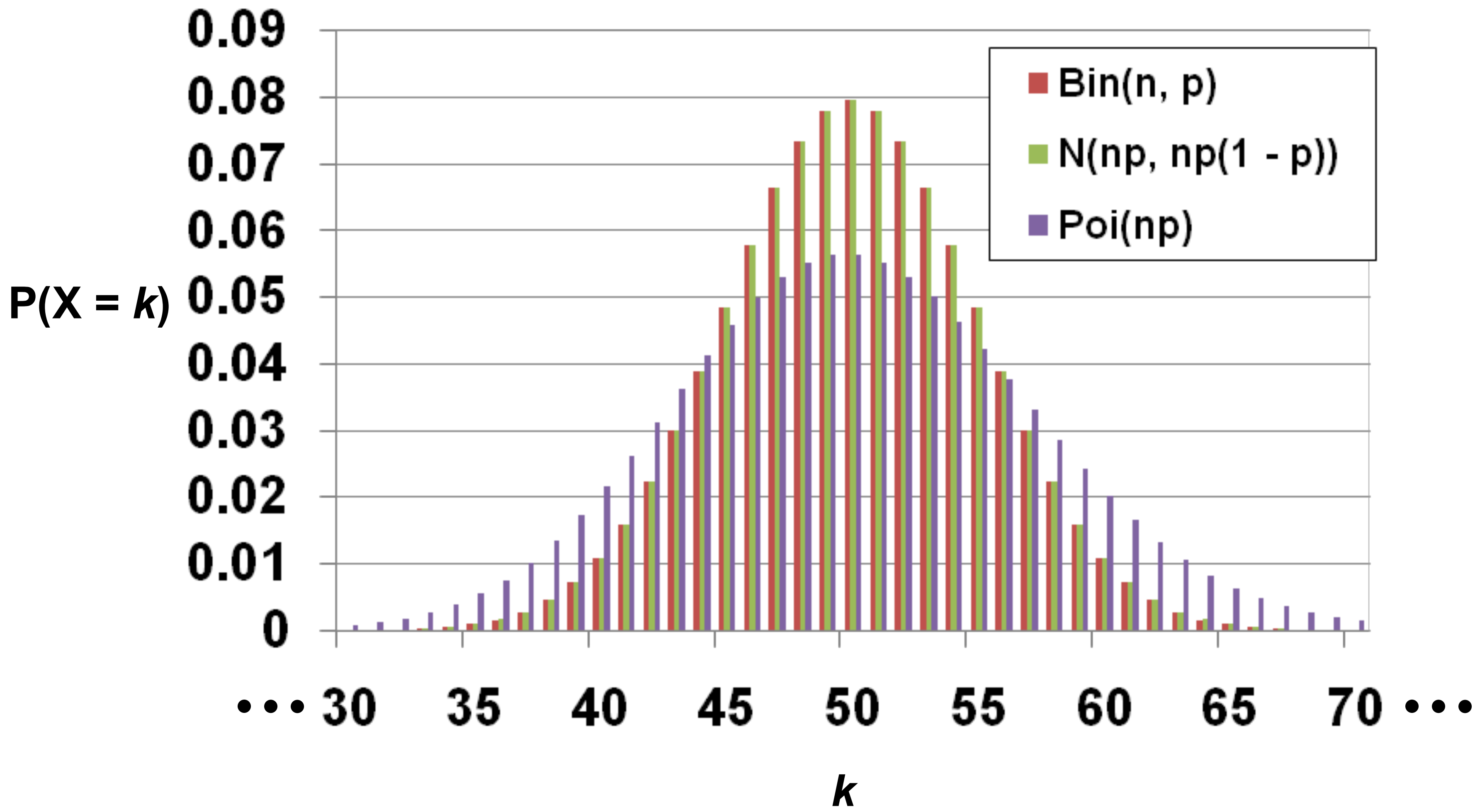
$$P(X = k) \approx P\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right) = \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p)}}\right)$$

"Continuity correction"

  - DeMoivre-Laplace Limit Theorem:
    - S_n: number of successes (with prob. p) in n independent trials

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \xrightarrow{n \to \infty} \Phi(b) - \Phi(a)$$

Comparison when n = 100, p = 0.5



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)

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

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

$$P(X \geq 65) \approx P(Y > 64.5)$$

$$P(Y > 64.5) = P\left(\frac{Y-50}{5} > \frac{64.5-50}{5}\right) = P(Z > 2.9) = 1 - \Phi(2.9) \approx 0.0019$$

  - Using Binomial:

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

Stanford Admissions

- Stanford accepts 2480 students
  - Each accepted student has 68% chance of attending
  - X = # students who will attend. X ~ Bin(2480, 0.68)
  - What is P(X > 1745)?

$$np = 1686.4 \quad np(1-p) \approx 539.65 \quad \sqrt{np(1-p)} \approx 23.23$$

  - Use Normal approximation: Y ~ N(1686.4, 539.65)

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

$$P(Y > 1745.5) = P\left(\frac{Y - 1686.4}{23.23} > \frac{1745.5 - 1686.4}{23.23}\right) = 1 - \Phi(2.54) \approx 0.0055$$

  - Using Binomial:

$$P(X > 1745) \approx 0.0053$$

Changes in Stanford Admissions

- Stanford Daily, March 28, 2014

  "Class of 2018 Admit Rates Lowest in University History" by Alex Zivkovic

  "Fewer students were admitted to the Class of 2018 than the Class of 2017, due to the increase in Stanford's yield rate which has increased over 5 percent in the past four years, according to Colleen Lim M.A. '80, Director of Undergraduate Admission."