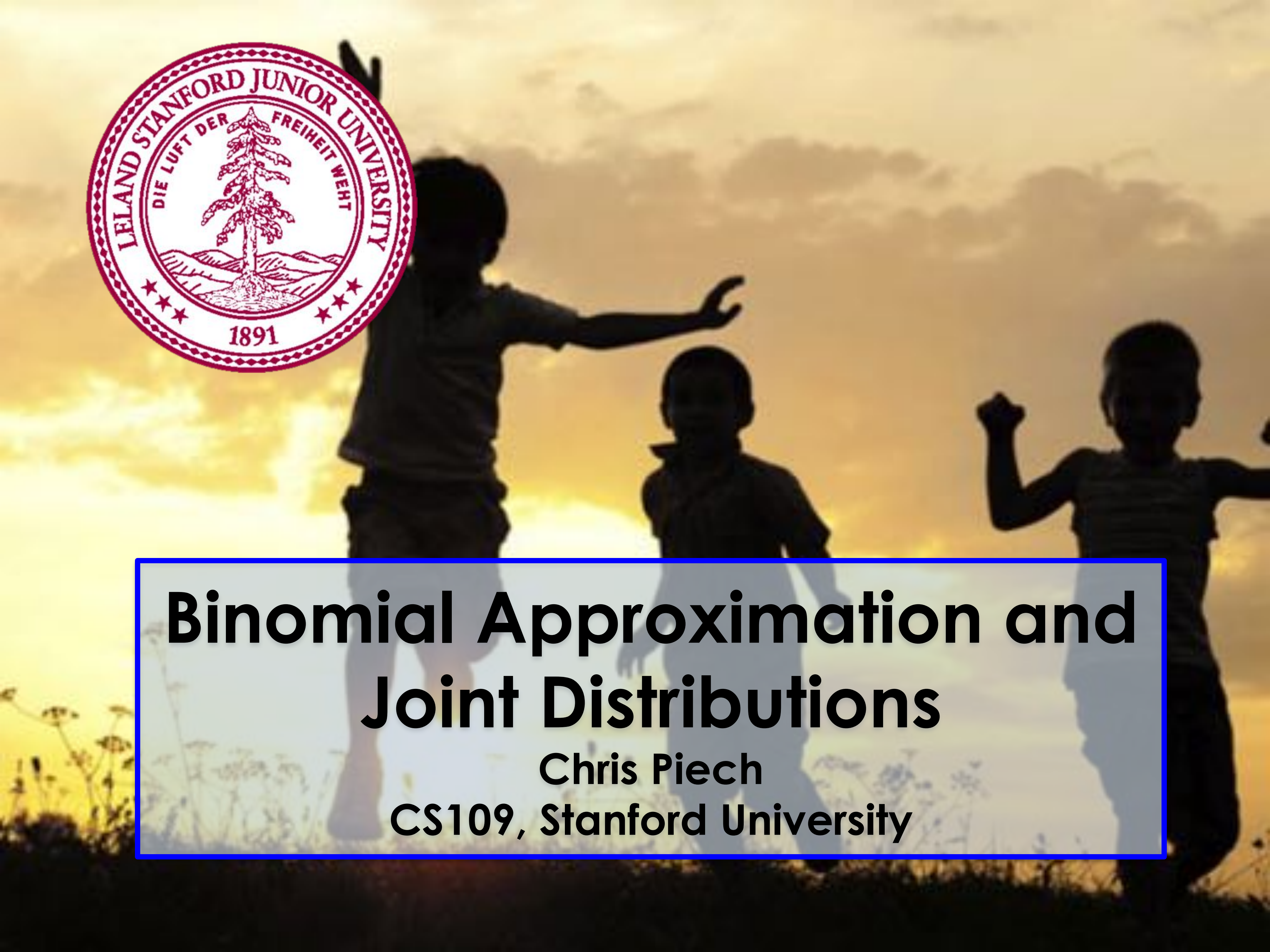


Binomial Approximation and Joint Distributions

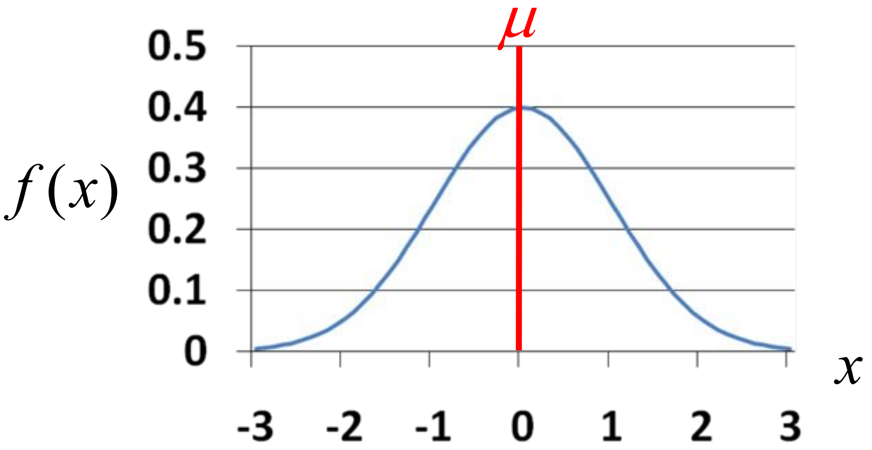
Chris Piech
CS109, Stanford University

Review

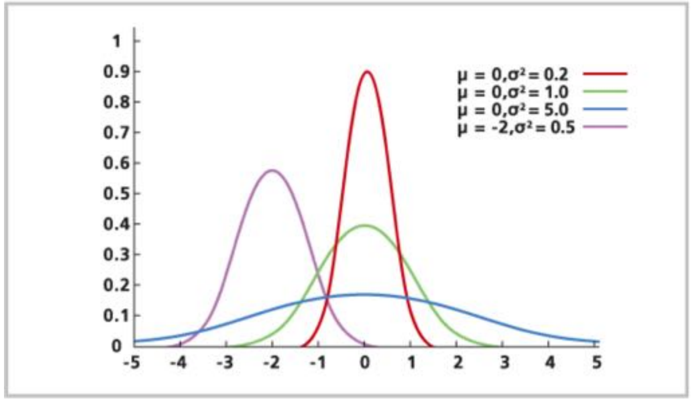
The Normal Distribution

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

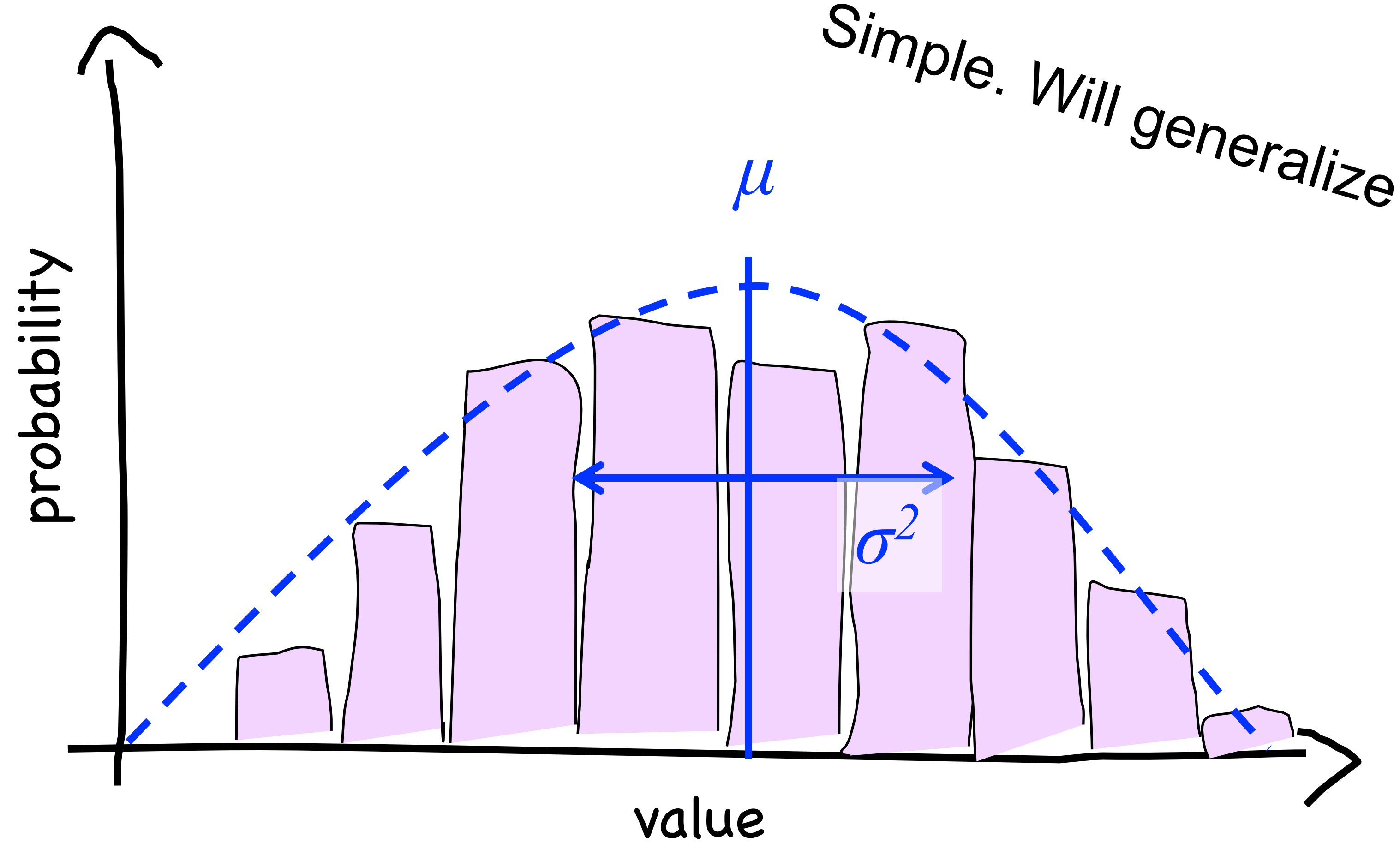
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{where} \quad -\infty < x < \infty$$



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

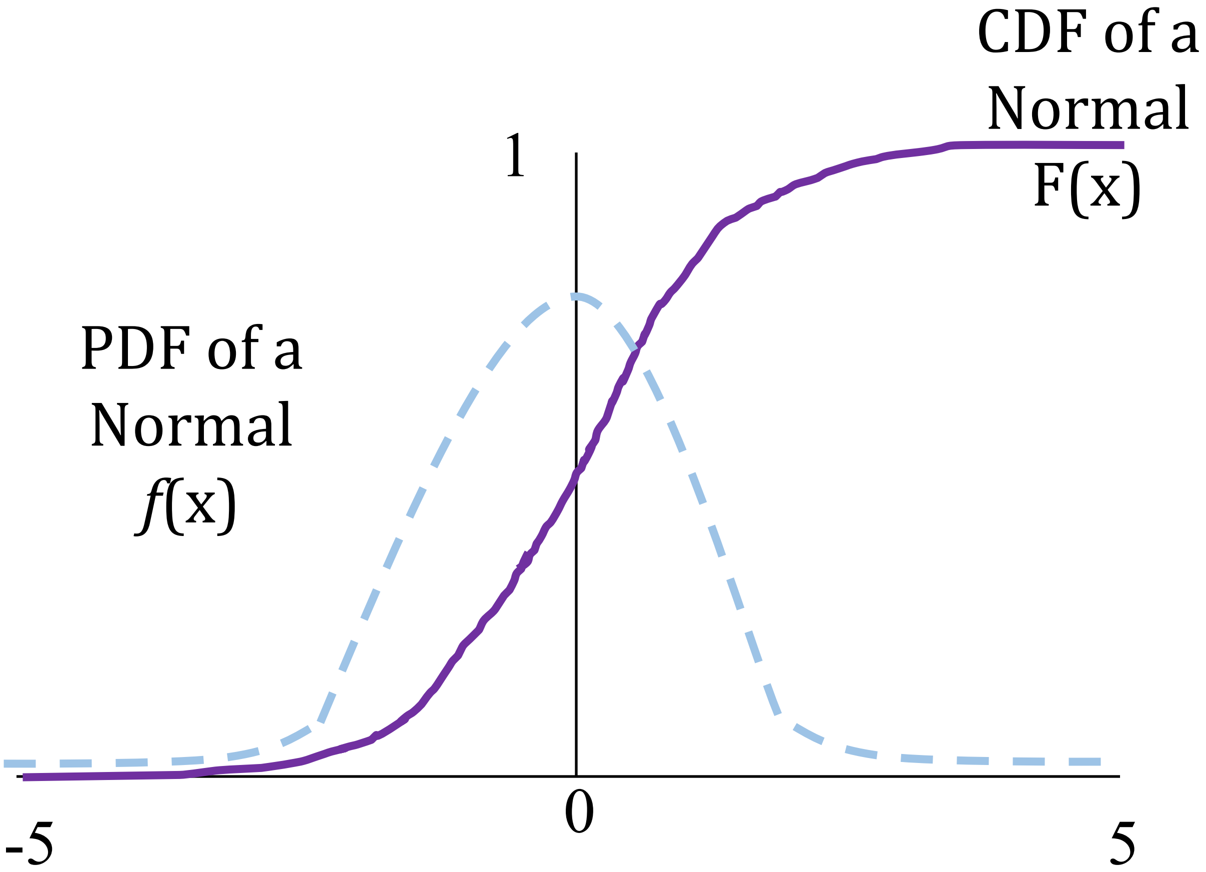


Simplicity is Humble



* A Gaussian maximizes entropy for a given mean and variance

Density vs Cumulative



$f(x)$ = derivative of probability

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

Probability Density Function

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

"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

Cumulative Density Function

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

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

Great questions!

68% rule only for Gaussians?

68% Rule?

What is the probability that a normal variable $X \sim N(\mu, \sigma^2)$ has a value within one standard deviation of its mean?

$$P(\mu - \sigma < X < \mu + \sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + \sigma - \mu}{\sigma}\right)$$

$$= P(-1 < Z < 1)$$

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

$$= \Phi(1) - [1 - \Phi(1)]$$

$$= 2\Phi(1) - 1$$

$$= 2[0.8413] - 1 = 0.683$$

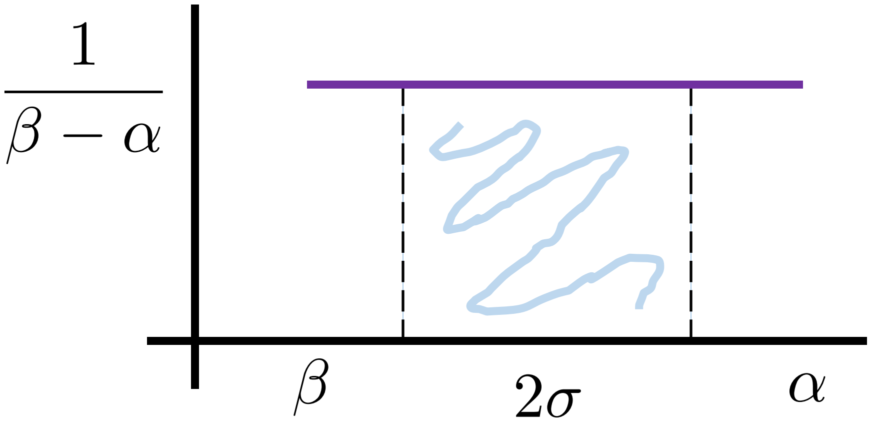
Only applies to normal

68% Rule?

Counter example: Uniform $X \sim Uni(\alpha, \beta)$

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

$$\sigma = \sqrt{Var(X)}$$
$$= \frac{\beta - \alpha}{\sqrt{12}}$$



$$P(\mu - \sigma < X < \mu + \sigma)$$
$$= \frac{1}{\beta - \alpha}\left[\frac{2(\beta - \alpha)}{\sqrt{12}}\right]$$
$$= \frac{2}{\sqrt{12}}$$
$$= 0.58$$

How does python sample from a Gaussian?

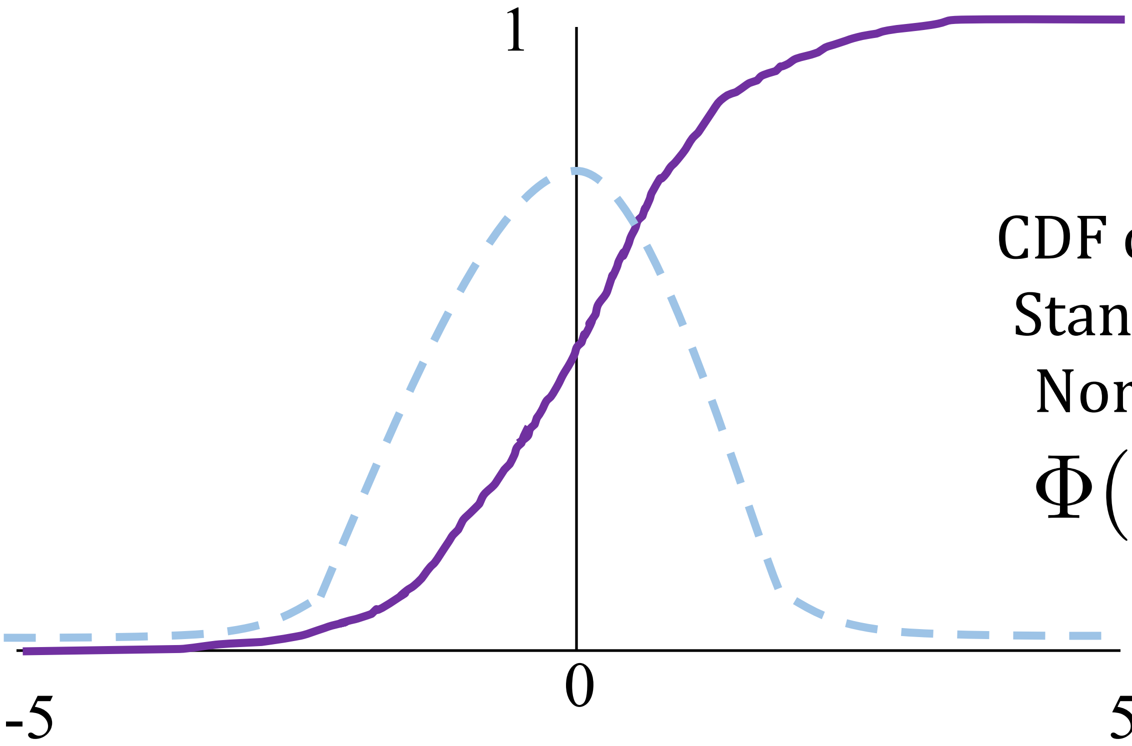
```
from random import *

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 Normal?

Inverse Transform Sampling

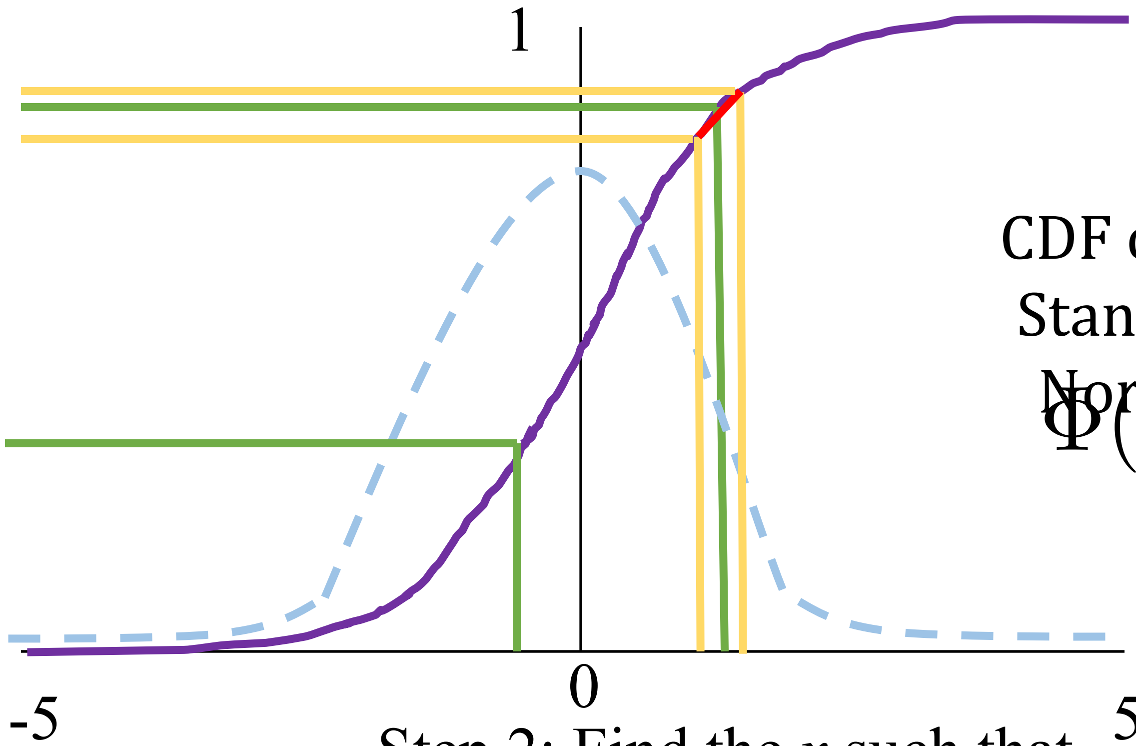


CDF of the Standard Normal

$\Phi(x)$

How Does a Computer Sample Normal?

Inverse Transform Sampling



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

CDF of the Standard Normal $\Phi(x)$

Step 2: Find the x such that

$$\Phi(x) = y$$

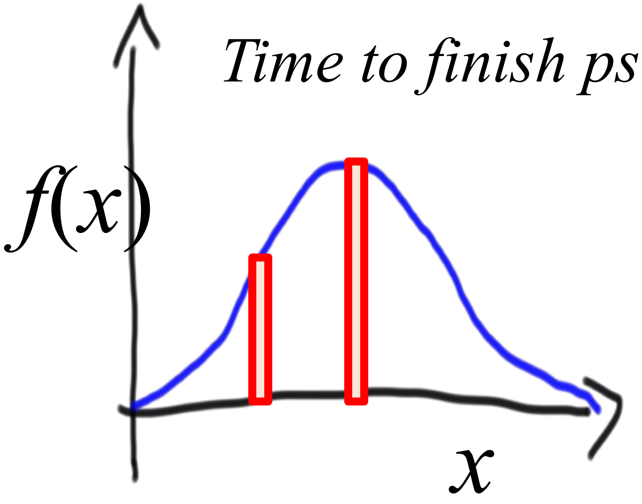
$$x = \Phi^{-1}(y)$$

Further reading: Box–Muller transform

Continuous RV Relative Probability

X = time to finish pset 3

X ~ N(10, 2)



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

$$\frac{P(X=10)}{P(X=5)} = \frac{\varepsilon f(X=10)}{\varepsilon f(X=5)}$$

$$= \frac{f(X=10)}{f(X=5)}$$

$$= \frac{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(5-\mu)^2}{2\sigma^2}}}$$

$$= \frac{\frac{1}{\sqrt{4\pi}} e^{-\frac{(10-10)^2}{4}}}{\frac{1}{\sqrt{4\pi}} e^{-\frac{(5-10)^2}{4}}}$$

$$= \frac{e^0}{e^{-\frac{25}{4}}} = 518$$

Imagine you are sitting a test…

Website Testing

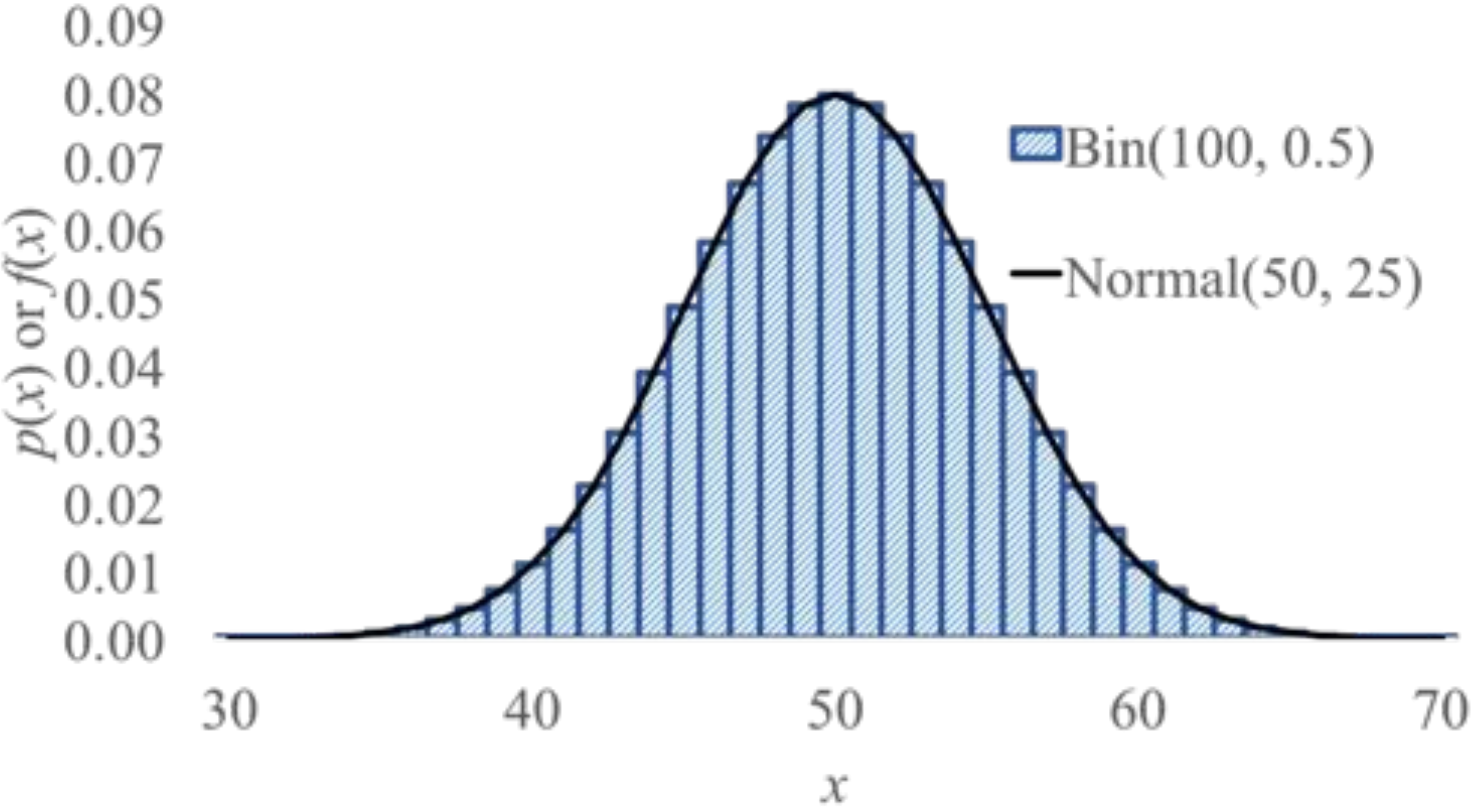
- 100 people are given a new website design
  - X = # people whose time on site increases
  - CEO will endorse new design if X ≥ 65 What is P(CEO endorses change| it has no effect)?
  - X ~ Bin(100, 0.5). Want to calculate P(X ≥ 65)
  - Give a numerical answer…

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

Normal Approximates Binomial

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

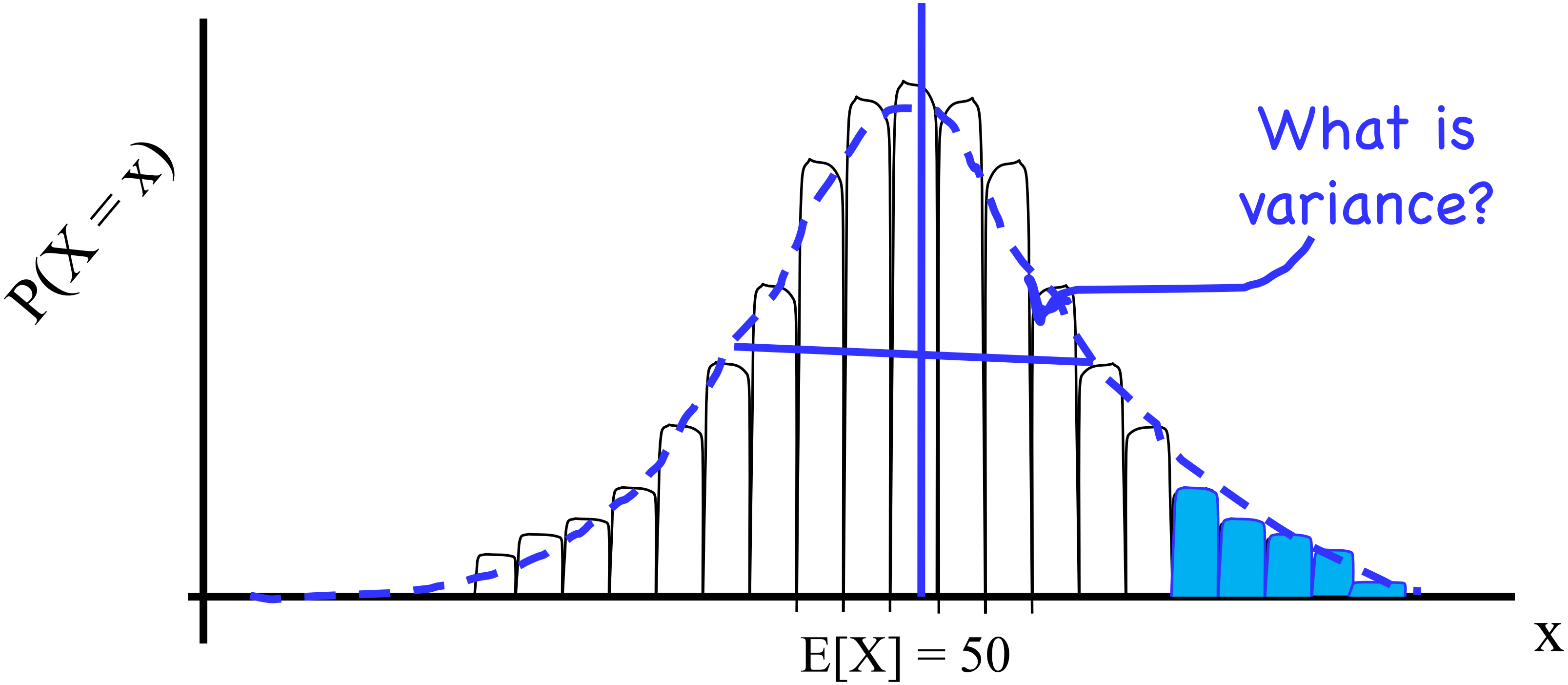
Normal Approximates Binomial



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)

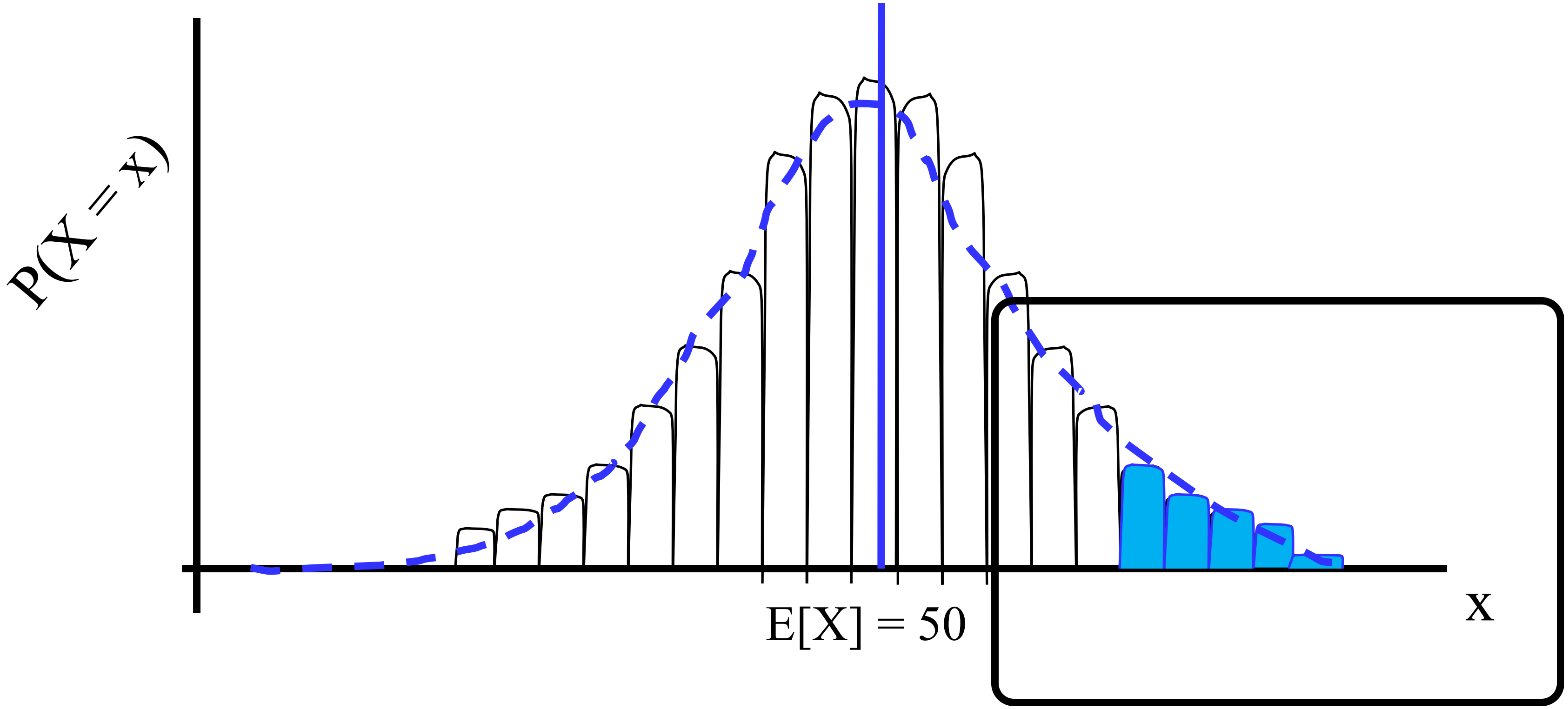
$$np = 50 \quad np(1-p) = 25 \quad \sqrt{np(1-p)} = 5$$

  - 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



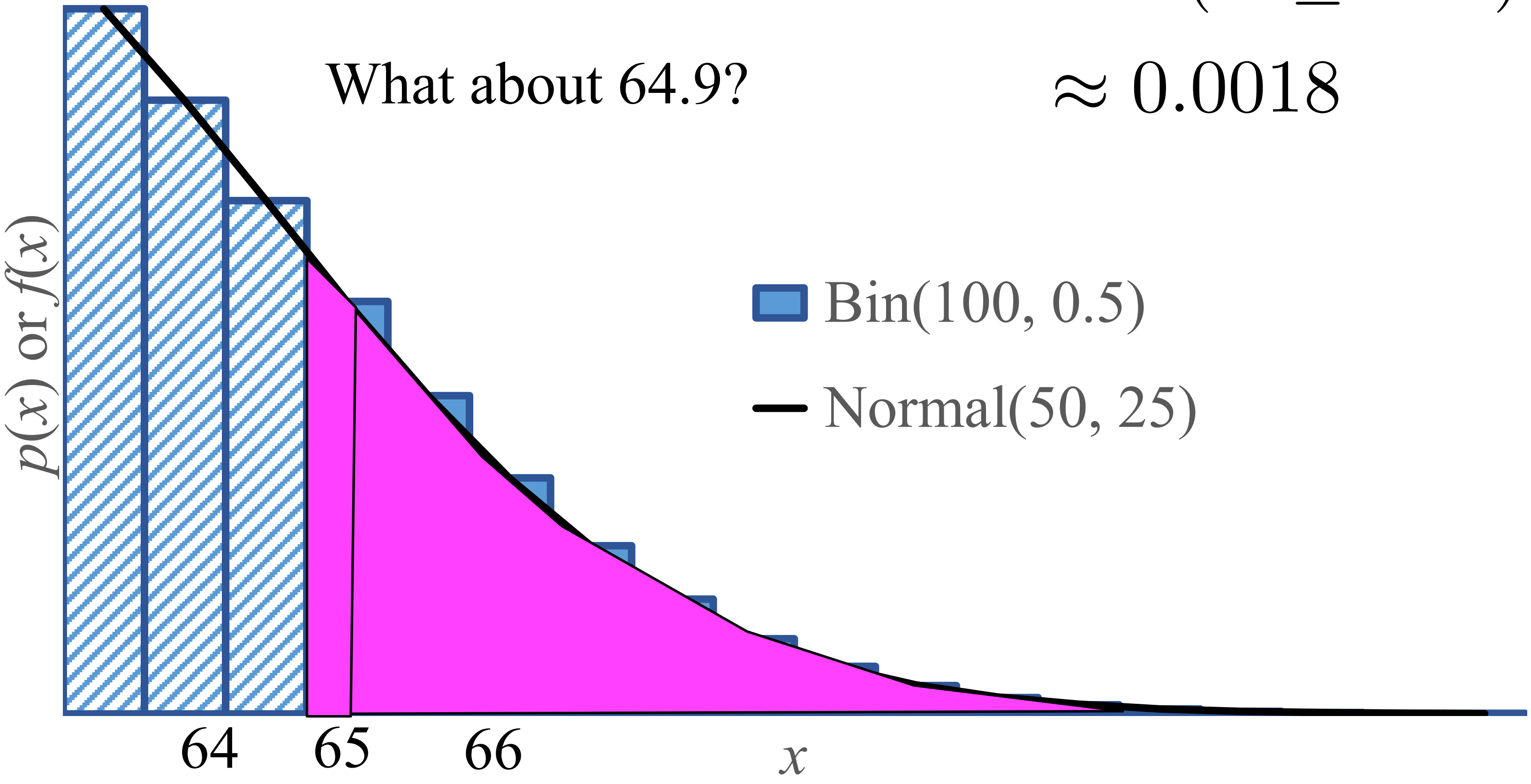
Continuity Correction

If Y (normal) approximates X (binomial)

$$P(X \geq 65)$$

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

$$\approx 0.0018$$



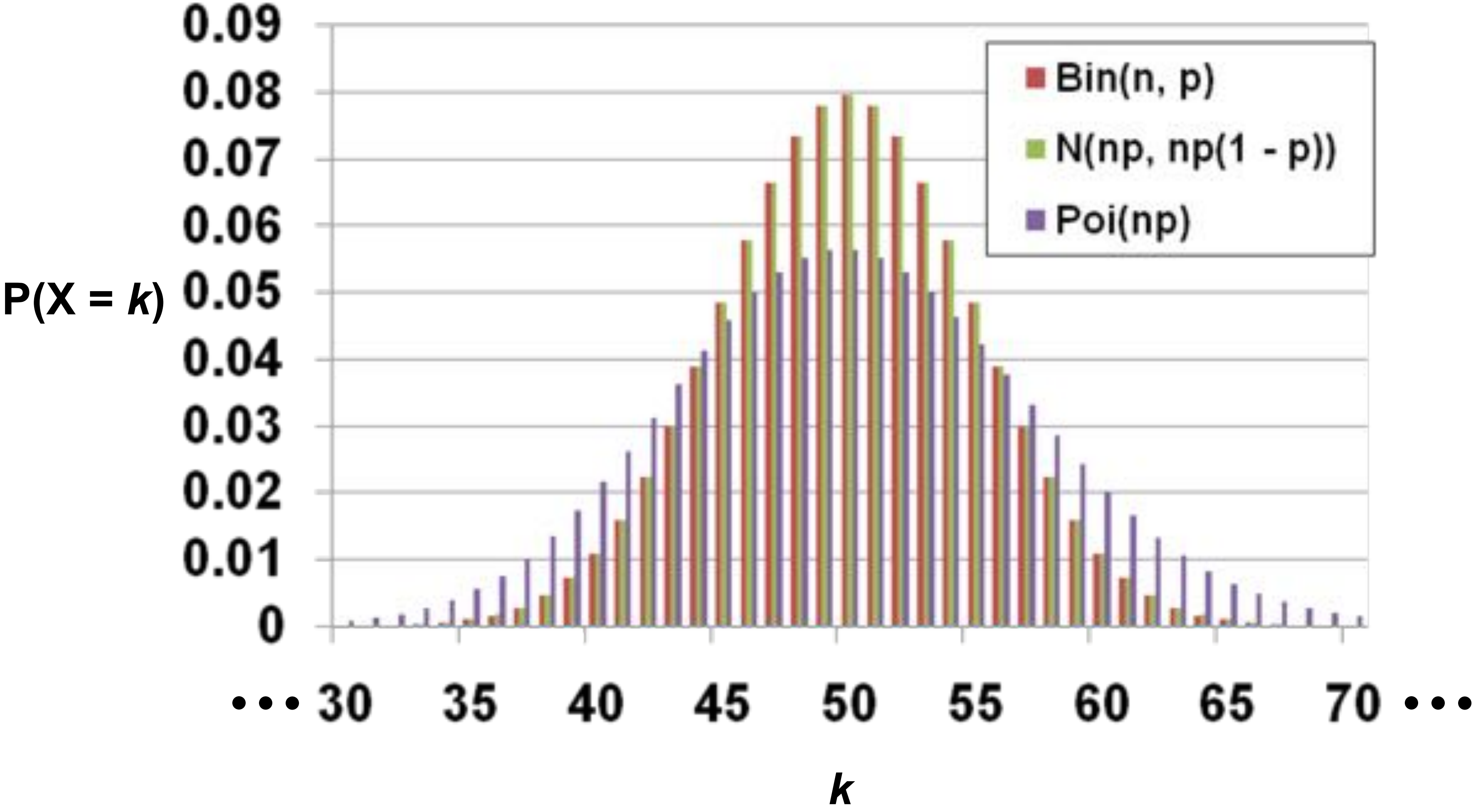
Continuity Correction

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

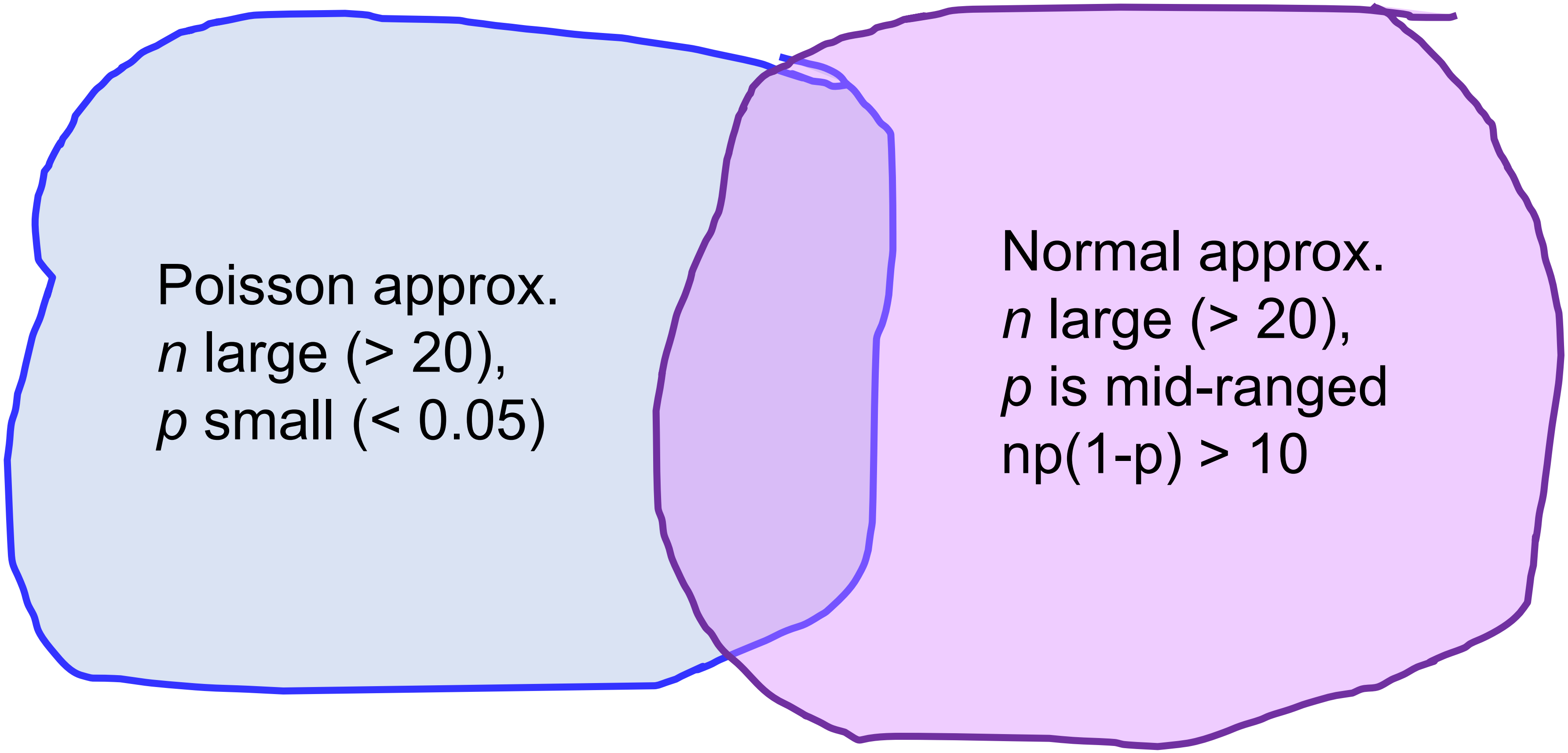
* Note: Binomial is always defined in units of “1”

Comparison when $n = 100$, $p = 0.5$



Who Gets to Approximate?

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



Poisson approx.
n large (> 20),
p small (< 0.05)

Normal approx.
n large (> 20),
p is mid-ranged
np(1-p) > 10

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)

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

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

$$\approx 0.08$$

Changes in Stanford Admissions

Class of 2021 Admit Rates Lowest in University History

“Fewer students were admitted to the Class of 2021 than the Class of 2019, 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.”

68% 10 years ago

84% last year

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

How mysterious and curious. You must wait a few classes ☺.

Joint Distributions

CS109 Joint

Go to this URL: https://goo.gl/Jh3Eu4

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

	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

X Y	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

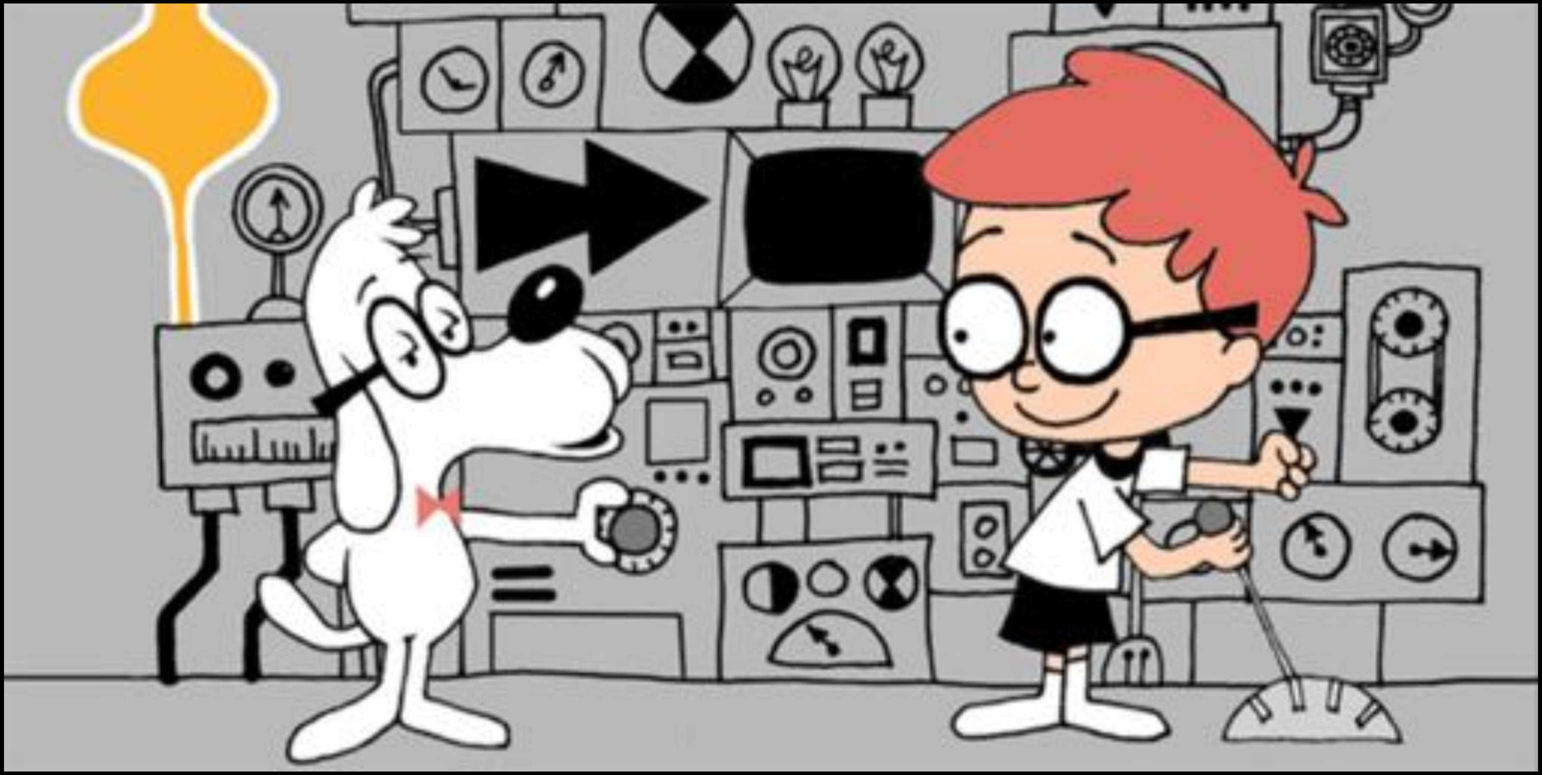
X Y	0	1	2	3	$p_Y(y)$
0	0.16	0.12	0.07	0.04	0.39
1	0.12	0.14	0.12	0	0.38
2	0.07	0.12	0	0	0.19
3	0.04	0	0	0	0.04
$p_X(x)$	0.39	0.38	0.19	0.04	1.00

Marginal distributions

CS109 Joint Results

Go to this URL: https://goo.gl/Jh3Eu4

Way Back



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)
…
n_r are the same (indistinguishable)?

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

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

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

Probability of exactly *i* successes

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

End Way Back

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 $\binom{n}{c_1, c_2, \ldots, c_m} = \frac{n!}{c_1! c_2! \cdots c_m!}$

Hello Die Rolls, My Old Friends

- 6-sided die is rolled 7 times
  - Roll results: 1 one, 1 two, 0 three, 2 four, 0 five, 3 six

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3)$$

$$= \frac{7!}{1!1!0!2!0!3!}\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^3 = 420\left(\frac{1}{6}\right)^7$$

- This is generalization of Binomial distribution
  - Binomial: each trial had 2 possible outcomes
  - Multinomial: each trial has *m* possible outcomes

Probabilistic Text Analysis

- Ignoring order of words, what is probability of any given word you write in English?
  - P(word = “the”) > P(word = “transatlantic”)
  - P(word = “Stanford”) > P(word = “Cal”)
  - Probability of each word is just multinomial distribution
- What about probability of those same words in someone else’s writing?
  - P(word = “probability” | writer = you) >

    P(word = “probability” | writer = non-CS109 student)
  - After estimating P(word | writer) from known writings, use Bayes’ Theorem to determine P(writer | word) for new writings!

A Document is a Large Multinomial

According to the Global Language Monitor there are 988,968 words in the english language used on the internet.

Text is a Multinomial

Example document:
“Pay for Viagra with a credit-card. Viagra is great. So are credit-cards. Risk free Viagra. Click for free.”
n = 18

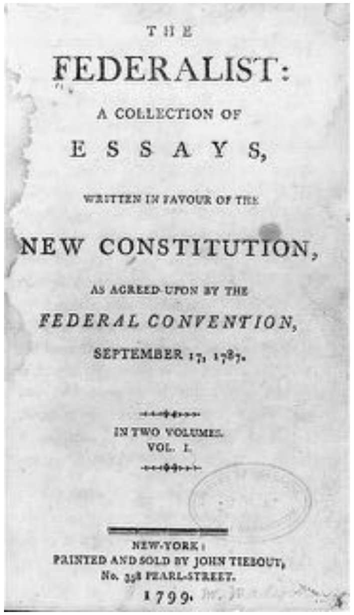
$$\underbrace{P\left(\begin{array}{l} \text{Viagra} = 2 \\ \text{Free} = 2 \\ \text{Risk} = 1 \\ \text{Credit-card: } 2 \\ \ldots \\ \text{For} = 2 \end{array} \middle| \text{spam}\right)}_{\text{Probability of seeing this document } | \text{ spam}} = \overbrace{\frac{n!}{2!2!\ldots 2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \cdots p_{\text{for}}^2}^{\text{It's a Multinomial!}}$$

The probability of a word in spam email being viagra

Who wrote the federalist papers?

Old and New Analysis

- Authorship of “Federalist Papers”
    - 85 essays advocating ratification of US constitution
    - Written under pseudonym “Publius”
        - Really, Alexander Hamilton, James Madison and John Jay
    - Who wrote which essays?
        - Analyzed probability of words in each essay versus word distributions from known writings of three authors



THE
FEDERALIST:
A COLLECTION OF
ESSAYS,
WRITTEN IN FAVOUR OF THE
NEW CONSTITUTION,
AS AGREED UPON BY THE
FEDERAL CONVENTION,
SEPTEMBER 17, 1787.

IN TWO VOLUMES.
VOL. I.

NEW-YORK:
PRINTED AND SOLD BY JOHN TIEBOUT,
No. 358 PEARL-STREET.
1799.

Let’s write a program!