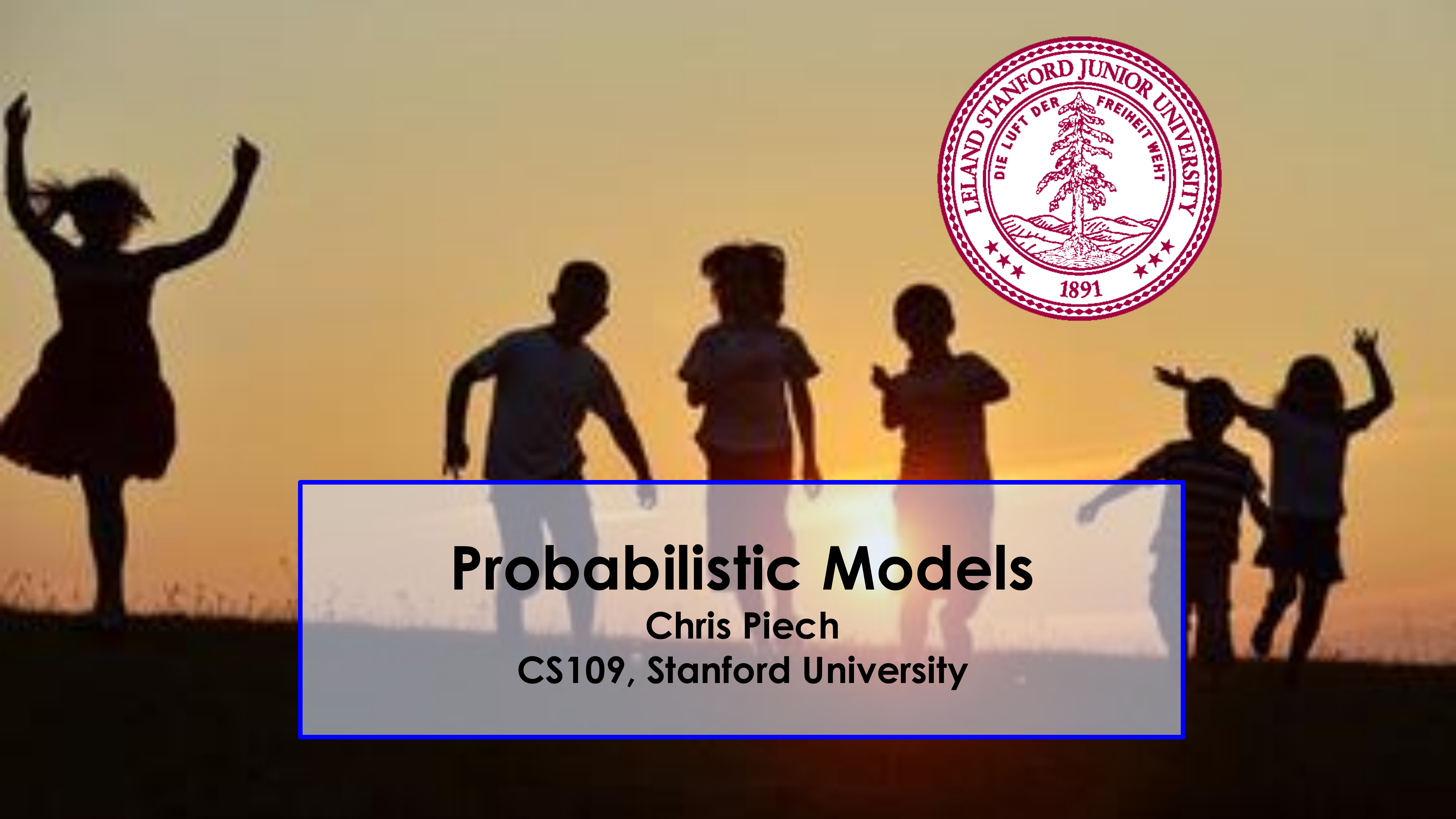
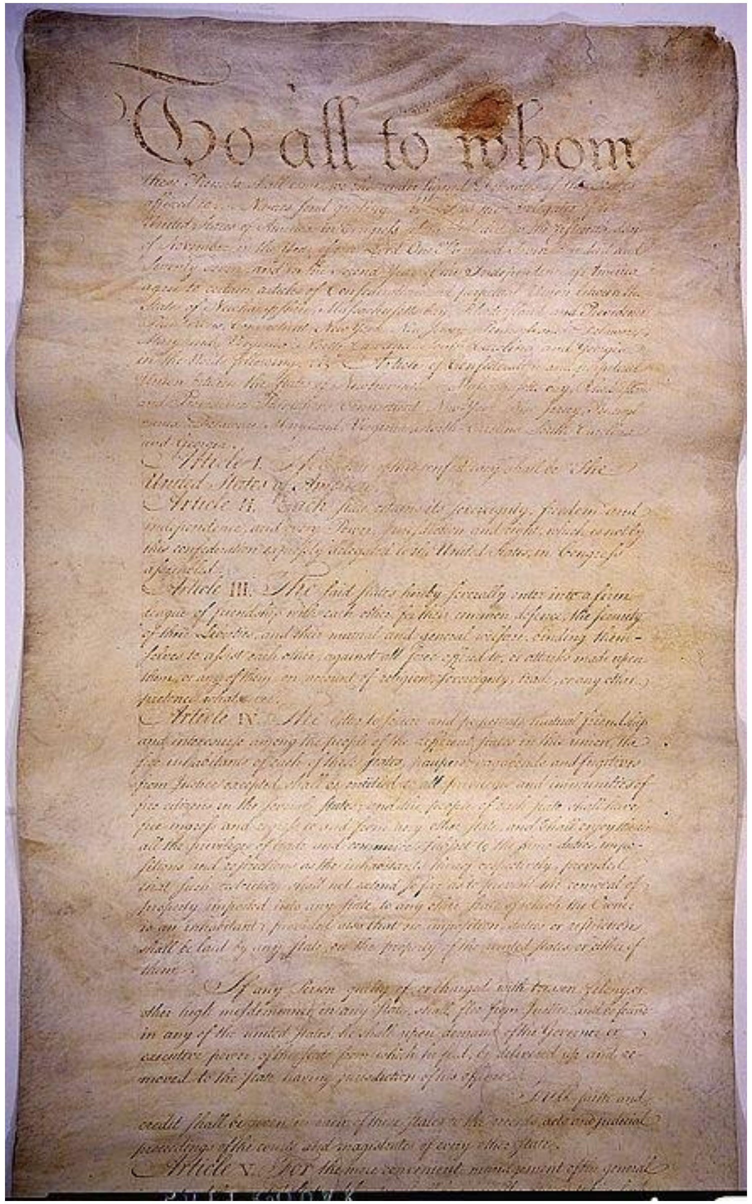
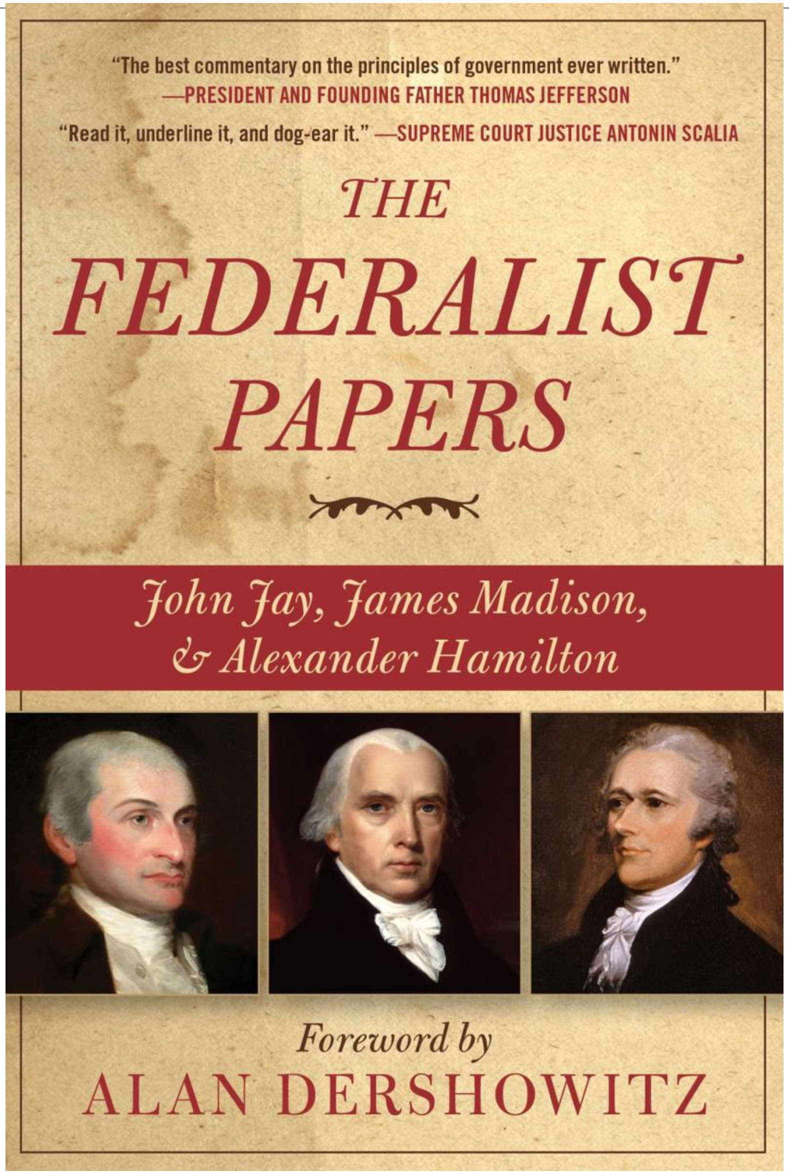


Probabilistic Models

Chris Piech
CS109, Stanford University

Exciting Day!





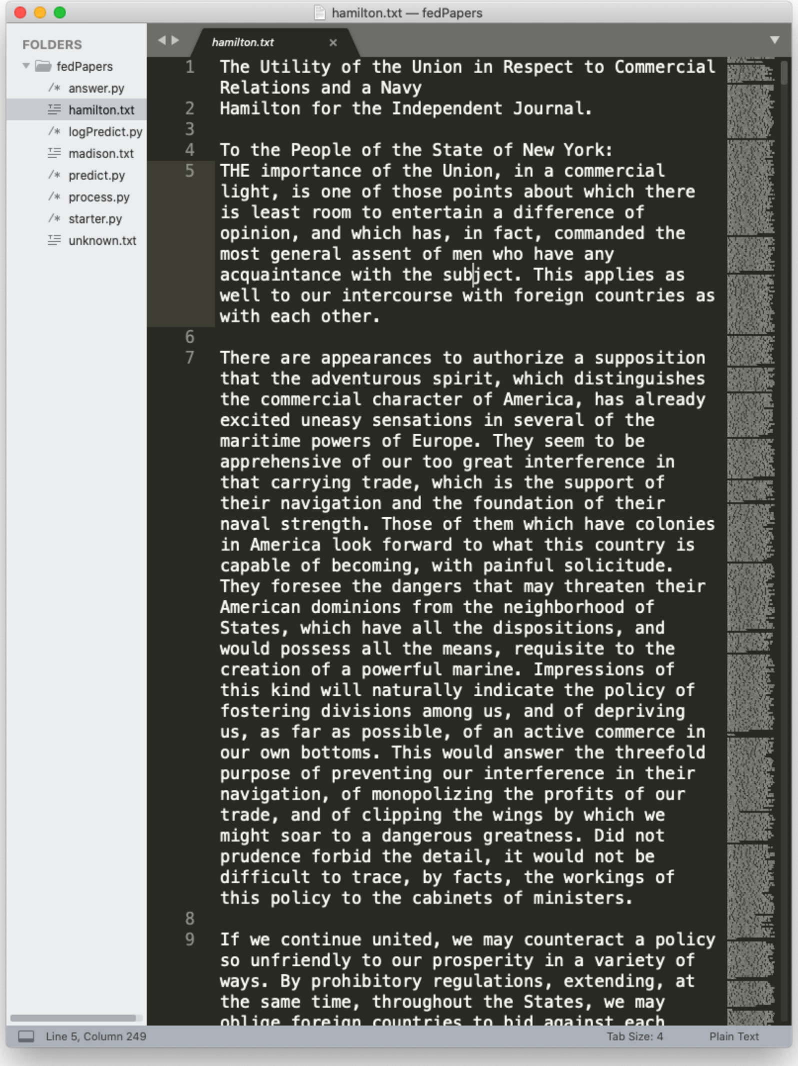
Who wrote Federalist Paper 53?

madison.txt

To the People of the State of New York:

AMONG the numerous advantages promised by a wellconstructed Union, none deserves to be more accurately developed than its tendency to break and control the violence of faction. The friend of popular governments never finds himself so much alarmed for their character and fate, as when he contemplates their propensity to this dangerous vice. He will not fail, therefore, to set a due value on any plan which, without violating the principles to which he is attached, provides a proper cure for it. The instability, injustice, and confusion introduced into the public councils, have, in truth, been the mortal diseases under which popular governments have everywhere perished; as they continue to be the favorite and fruitful topics from which the adversaries to liberty derive their most specious declamations. The valuable improvements made by the American constitutions on the popular models, both ancient and modern, cannot certainly be too much admired; but it would be an unwarrantable partiality, to contend that they have as effectually obviated the danger on this side, as was wished and expected. Complaints are everywhere heard from our most considerate and virtuous citizens, equally the friends of public and private faith, and of public and personal liberty, that our governments are too unstable, that the public good is disregarded in the conflicts of rival parties, and that measures are too often decided, not according to the rules of justice and the rights of the minor party, but by the superior force of an interested and overbearing majority. However anxiously we may wish that these complaints had no foundation, the evidence, of known facts will not permit us to deny that they are in some degree true. It will be found, indeed, on a candid review of our situation, that some of the distresses under which we labor have been erroneously charged on the operation of our governments; but it will be found, at the same time, that other causes will not alone account for many of our heaviest misfortunes; and, particularly, for that prevailing and increasing distrust of public

hamilton.txt



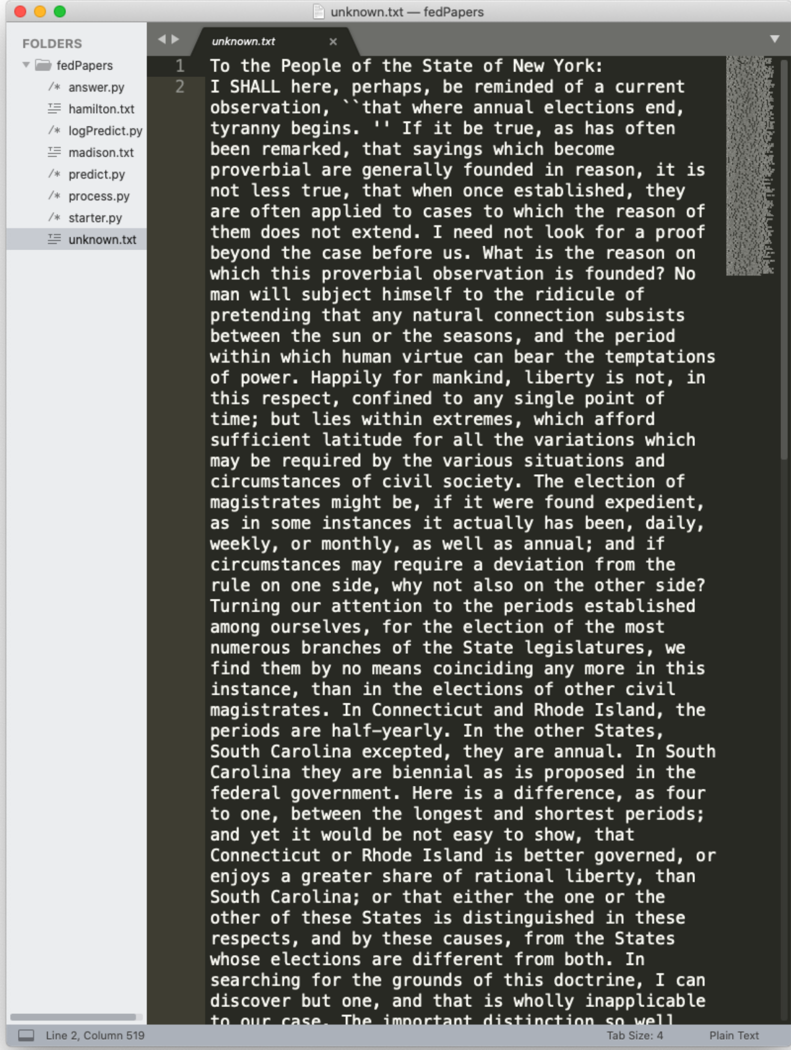
The Utility of the Union in Respect to Commercial Relations and a Navy
Hamilton for the Independent Journal.

To the People of the State of New York:
THE importance of the Union, in a commercial light, is one of those points about which there is least room to entertain a difference of opinion, and which has, in fact, commanded the most general assent of men who have any acquaintance with the subject. This applies as well to our intercourse with foreign countries as with each other.

There are appearances to authorize a supposition that the adventurous spirit, which distinguishes the commercial character of America, has already excited uneasy sensations in several of the maritime powers of Europe. They seem to be apprehensive of our too great interference in that carrying trade, which is the support of their navigation and the foundation of their naval strength. Those of them which have colonies in America look forward to what this country is capable of becoming, with painful solicitude. They foresee the dangers that may threaten their American dominions from the neighborhood of States, which have all the dispositions, and would possess all the means, requisite to the creation of a powerful marine. Impressions of this kind will naturally indicate the policy of fostering divisions among us, and of depriving us, as far as possible, of an active commerce in our own bottoms. This would answer the threefold purpose of preventing our interference in their navigation, of monopolizing the profits of our trade, and of clipping the wings by which we might soar to a dangerous greatness. Did not prudence forbid the detail, it would not be difficult to trace, by facts, the workings of this policy to the cabinets of ministers.

If we continue united, we may counteract a policy so unfriendly to our prosperity in a variety of ways. By prohibitory regulations, extending, at the same time, throughout the States, we may oblige foreign countries to bid against each

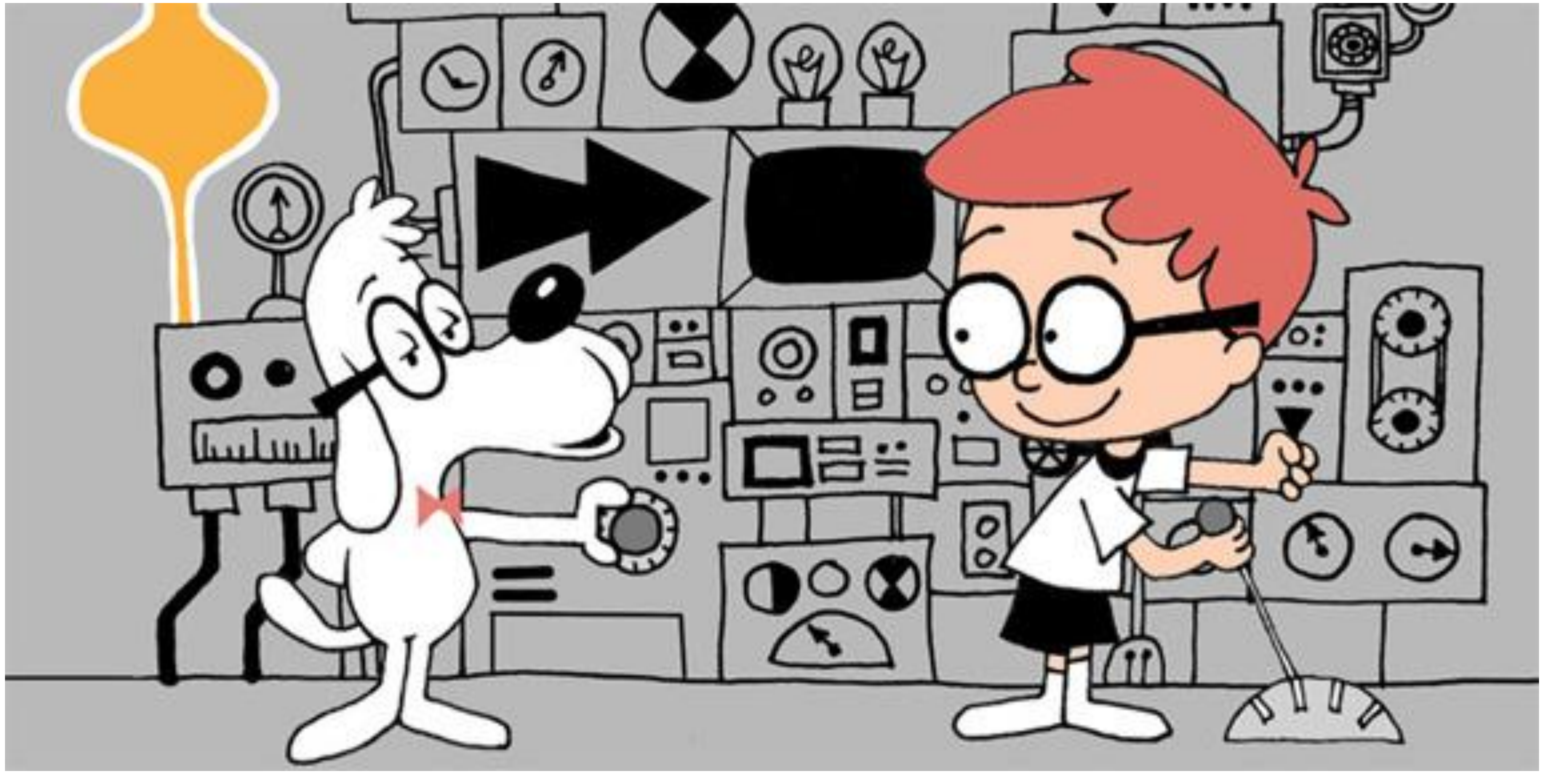
unknown.txt



To the People of the State of New York:
I SHALL here, perhaps, be reminded of a current observation, ``that where annual elections end, tyranny begins. '' If it be true, as has often been remarked, that sayings which become proverbial are generally founded in reason, it is not less true, that when once established, they are often applied to cases to which the reason of them does not extend. I need not look for a proof beyond the case before us. What is the reason on which this proverbial observation is founded? No man will subject himself to the ridicule of pretending that any natural connection subsists between the sun or the seasons, and the period within which human virtue can bear the temptations of power. Happily for mankind, liberty is not, in this respect, confined to any single point of time; but lies within extremes, which afford sufficient latitude for all the variations which may be required by the various situations and circumstances of civil society. The election of magistrates might be, if it were found expedient, as in some instances it actually has been, daily, weekly, or monthly, as well as annual; and if circumstances may require a deviation from the rule on one side, why not also on the other side? Turning our attention to the periods established among ourselves, for the election of the most numerous branches of the State legislatures, we find them by no means coinciding any more in this instance, than in the elections of other civil magistrates. In Connecticut and Rhode Island, the periods are half-yearly. In the other States, South Carolina excepted, they are annual. In South Carolina they are biennial as is proposed in the federal government. Here is a difference, as four to one, between the longest and shortest periods; and yet it would be not easy to show, that Connecticut or Rhode Island is better governed, or enjoys a greater share of rational liberty, than South Carolina; or that either the one or the other of these States is distinguished in these respects, and by these causes, from the States whose elections are different from both. In searching for the grounds of this doctrine, I can discover but one, and that is wholly inapplicable to our case. The important distinction so well

First, some review

Recall the good times



Permutations

$n!$

How many ways are there to order n objects?

Fixed Sized Buckets

How many ways are there to put ***n* distinct objects** into ***r* buckets** such that:

n_1 go into bucket 1

n_2 go into bucket 2

...

n_r go into bucket r?

Fixed Sized Buckets

How many ways are there to put ***n* distinct objects** into ***r* buckets** such that:

n_1 go into bucket 1

n_2 go into bucket 2

...

n_r go into bucket r?

Ways to put elements into fixed size containers

How many ways are there to put ***n* distinct objects** into ***r* buckets** such that:

n_1 go into bucket 1
n_2 go into bucket 2
...
n_r go into bucket r?

This is called the multinomial notation

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

Note: Multinomial > Binomial

Counting unordered objects

Binomial coefficient

How many ways are there
to order ***n*** heads
and (***n-k***) tails

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

Called the binomial coefficient
because of something from Algebra

Multinomial coefficient

How many ways are there
to order n_1 outcomes of type 1
n_2 outcomes of type 2
n_3 outcomes of type 3…
n_r outcomes of type r

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

Multinomials generalize
Binomials for counting.

Where are we in CS109?

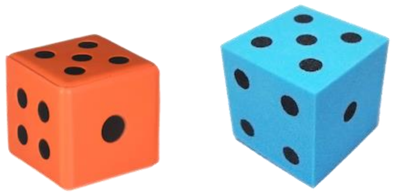
Overview of Topics

- Counting Theory
- Core Probability
- Random Variables
- Probabilistic Models
- Uncertainty Theory
- Machine Learning

How do you represent a joint distribution?

Roll two 6-sided dice, yielding values X and Y.

What is the joint PMF of X and Y?



$$P(X = a, Y = b) = 1/36 \qquad (a, b) \in \{(1,1), \dots, (6,6)\}$$

		X					
		1	2	3	4	5	6
Y	1	1/36	…	…	…	…	1/36
	2	…	…	…	…		
	3	…	…	…	…	…	…
	4	…	…	…	…	…	…
	5	…	…	…	…	…	…
	6	1/36	…	…	…	…	1/36

$P(X = 4, Y = 3)$

Probability table

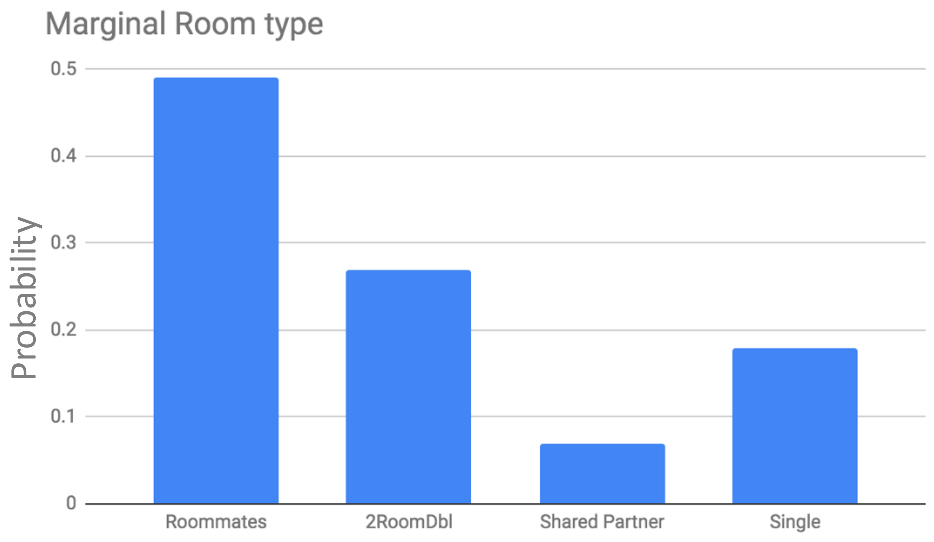
- All possible outcomes for several discrete RVs
- Not parametric (e.g., parameter p in Ber(p))

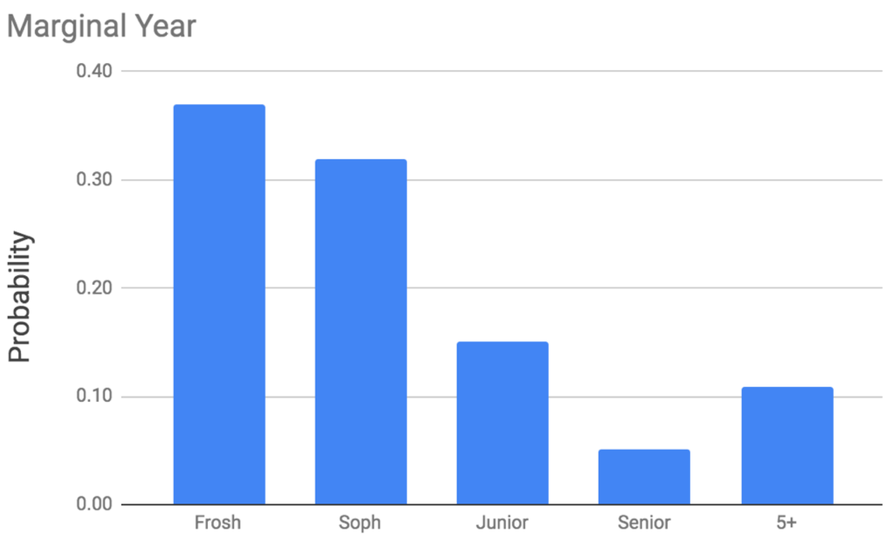
How do you represent a joint distribution?

Joint

	Roommates	2RoomDbl	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00

Marginals





Key limitation of the joint table:
it is too big

What about 3 Random Variables?

D is disease, ***S*** is can smell, ***F*** is fever status

$D = 0$

	$S = 0$	$S = 1$
$F = \text{none}$	0.024	0.783
$F = \text{low}$	0.003	0.092
$F = \text{high}$	0.001	0.046

$D = 1$

	$S = 0$	$S = 1$
$F = \text{none}$	0.006	0.014
$F = \text{low}$	0.005	0.011
$F = \text{high}$	0.004	0.011

$$P(D = 1) = \sum_f \sum_s P(D = 1, F = f, S = s)$$

What about 10 Random Variables?

Imagine you have **10 discrete** RVs which can each take on **5 values**

$$\text{\# unique assignments} = 5^{10}$$

10 million entries in your joint table.

So, we are going to need models …

… **probabilistic models** …

Roll 100 dice.

X_1 = How many 1s?

X_2 = How many 2s?

X_3 = How many 3s?

X_4 = How many 4s?

X_5 = How many 5s?

X_6 = How many 6s?

How big is the joint table?

Sometimes the structure of the variables suggests a more efficient representation

Multinomial RV

Probability

Binomial RV

What is the probability of getting k successes and $n - k$ failures in n trials?

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Binomial # of ways of ordering the successes

Probability of each ordering of k successes is equal + mutually exclusive

Multinomial RV

What is the probability of getting c_1 of outcome 1, c_2 of outcome 2, …, and c_m of outcome m in n trials?

Multinomial RVs generalize Binomial RVs

Recall the Binomial Derivation: What if more than two outcomes?

```
(H, H, H, H, T, T, T, T, T, T)
(H, H, H, T, H, T, T, T, T, T)
(H, H, H, T, T, H, T, T, T, T)
(H, H, H, T, T, T, H, T, T, T)
(H, H, H, T, T, T, T, H, T, T)
(H, H, H, T, T, T, T, T, H, T)
(H, H, H, T, T, T, T, T, T, H)
(H, H, T, H, H, T, T, T, T, T)
(H, H, T, H, T, H, T, T, T, T)
(H, H, T, H, T, T, H, T, T, T)
(H, H, T, H, T, T, T, H, T, T)
(H, H, T, H, T, T, T, T, H, T)
(H, H, T, H, T, T, T, T, T, H)
(H, H, T, T, H, H, T, T, T, T)
(H, H, T, T, H, T, H, T, T, T)
(H, H, T, T, H, T, T, H, T, T)
(H, H, T, T, H, T, T, T, H, T)
(H, H, T, T, H, T, T, T, T, H)
(H, H, T, T, T, H, H, T, T, T)
(H, H, T, T, T, H, T, H, T, T)
(H, H, T, T, T, H, T, T, H, T)
```

$$\begin{aligned}
P(E_{128}) &= p \cdot p \cdot p \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot (1-p) \\
&= p^4 \cdot (1-p)^6
\end{aligned}$$

$$\begin{aligned}
P(\text{exactly } k \text{ heads}) &= \sum_{i=1}^{N} P(E_i) && \text{Mutual Exclusion} \\
&= \sum_{i=1}^{N} p^k \cdot (1-p)^{n-k} && \text{Sub in } P(E_i) \\
&= N \cdot p^k \cdot (1-p)^{n-k} && \text{Sum } N \text{ times} \\
&= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} && \text{Perm of indistinct objects}
\end{aligned}$$

Three Outcomes A, B, C

```
(A, A, A, C, C, C, A, C, B, C)
(A, A, A, C, C, C, A, C, C, B)
(A, A, A, C, C, C, B, A, C, C)
(A, A, A, C, C, C, B, C, A, C)
(A, A, A, C, C, C, B, C, C, A)
(A, A, A, C, C, C, C, A, B, C)
(A, A, A, C, C, C, C, A, C, B)
(A, A, A, C, C, C, C, B, A, C)
(A, A, A, C, C, C, C, B, C, A)
(A, A, A, C, C, C, C, C, A, B)
(A, A, A, C, C, C, C, C, B, A)
(A, A, B, A, A, C, C, C, C, C)
(A, A, B, A, C, A, C, C, C, C)
(A, A, B, A, C, C, A, C, C, C)
(A, A, B, A, C, C, C, A, C, C)
(A, A, B, A, C, C, C, C, A, C)
(A, A, B, A, C, C, C, C, C, A)
(A, A, B, C, A, A, C, C, C, C)
(A, A, B, C, A, C, A, C, C, C)
(A, A, B, C, A, C, C, A, C, C)
(A, A, B, C, A, C, C, C, A, C)
```

$$p_A = 0.6$$
$$p_B = 0.1$$
$$p_C = 0.3$$

First here is a simulator where you can try rolling this dice 10 times:

Dice Roll Simulator

Number of rolls n: 10

Simulate

Simulator results:

C, C, B, A, A, A, C, C, C, A

Totals:
A: 4
B: 1
C: 5

What is the probability of exactly 4 As, 1 B and 5 Cs?

Multinomial Random Variable?

Consider an experiment of n independent trials:

- Each trial results in one of m outcomes. $P(\text{outcome } i) = p_i$, $\sum_{i=1}^{m} p_i = 1$
- Let X_i = # trials with outcome i

Joint PMF

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

where $\sum_{i=1}^{m} c_i = n$ and $\sum_{i=1}^{m} p_i = 1$

Probability of each ordering is equal + mutually exclusive

Multinomial Random Variable?

Consider an experiment of n independent trials:

- Each trial results in one of m outcomes. $P(\text{outcome } i) = p_i$, $\sum_{i=1}^{m} p_i = 1$
- Let X_i = # trials with outcome i

Joint PMF

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

where $\sum_{i=1}^{m} c_i = n$ and $\sum_{i=1}^{m} p_i = 1$

Multinomial # of ways of ordering the outcomes

Probability of each ordering is equal + mutually exclusive

Sometimes the structure of the variables suggests a more efficient representation

I roll 6 dice. What is more probable:

A) I roll 6 “sixes”

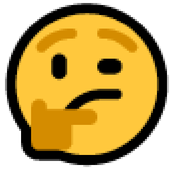
B) I roll exactly one of each number

Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes



Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes

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

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

Hello dice rolls, my old friends

A 6-sided die is rolled 7 times.

What is the probability of getting:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, \overbrace{X_6 = 3}^{\text{\# of times a six appears}})$$

$$= \binom{7}{1,1,0,2,0,\underbrace{3}_{\text{choose where the sixes appear}}} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^3 = 420 \left(\frac{1}{6}\right)^7$$

probability of rolling a six

this many times

Parameters of a Multinomial RV?

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

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

p: probability of success outcome on a single trial

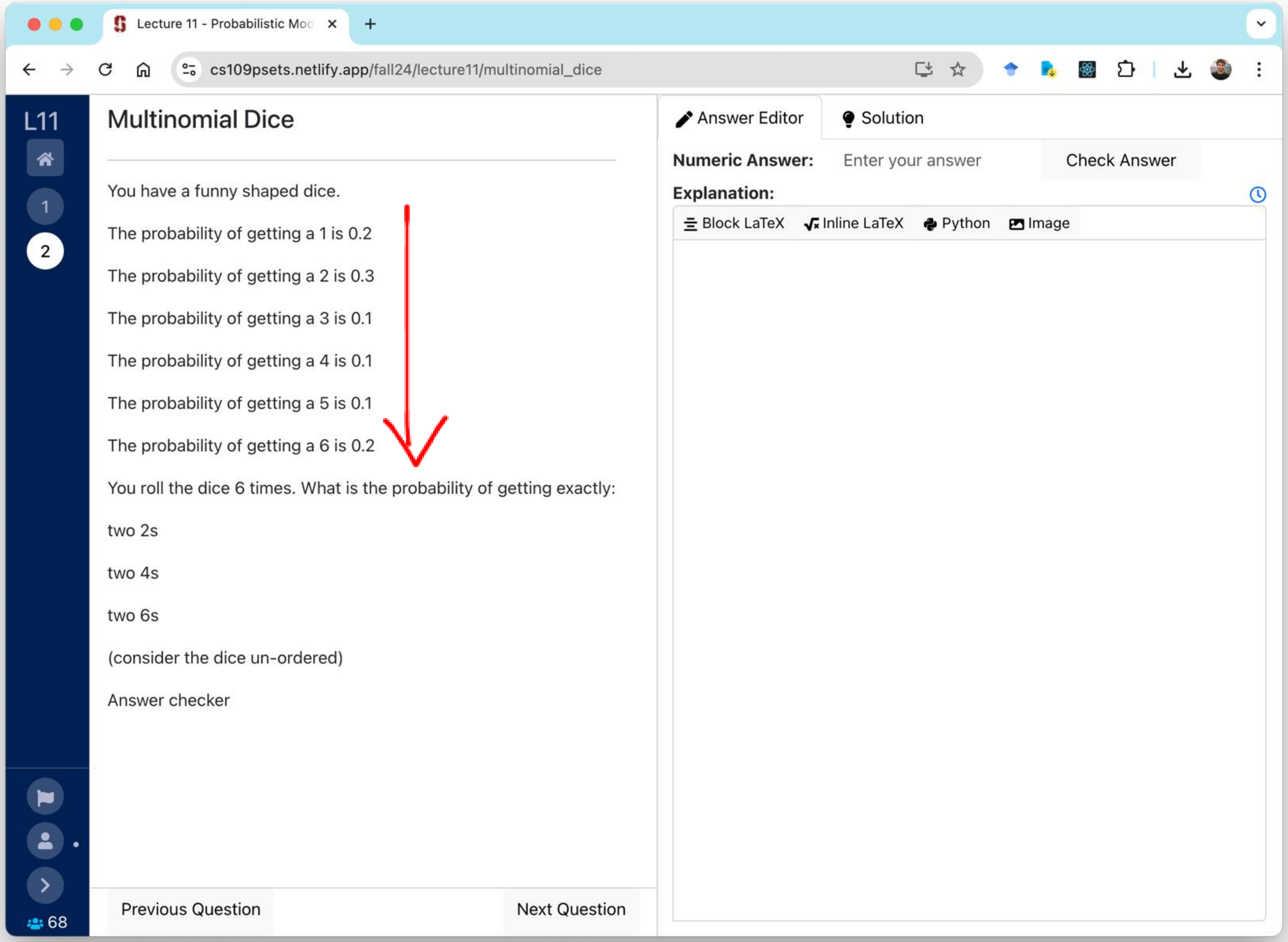
A Multinomial RV has parameters $n, p_1, p_2, \ldots, p_m$ (Note $p_m = 1 - \sum_{i=1}^{m-1} p_i$)

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

p_i: probability of outcome i on a single trial

Where do we get p_i from?

Most useful when probabilities are not equal



The Federalist Papers

Intro to Natural Language Processing

Probabilistic text analysis

Ignoring the order of words…

What is the probability of any given word that you write in English?

- $P(\text{word} = \text{"the"}) > P(\text{word} = \text{"pokemon"})$
- $P(\text{word} = \text{"Stanford"}) > P(\text{word} = \text{"Cal"})$

Probabilities of *counts* of words = Multinomial distribution

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.)

Model text as 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

$$\underbrace{P\left(\begin{array}{l}\text{Viagra} = 3\\ \text{Free} = 2\\ \text{Risk} = 1\\ \text{Credit-card: } 2\\ \ldots\\ \text{For} = 2\end{array} \middle| \text{spam}\right)}_{\text{Probability of seeing this document} \mid \text{spam}} \qquad \overbrace{\frac{n!}{3!\,2!\,\ldots\,2!}\, p_{\text{viagra}}^{3}\, p_{\text{free}}^{2} \cdots p_{\text{for}}^{2}}^{\text{It's a Multinomial!}}$$

The probability of a word in spam email being viagra

Who wrote the federalist papers?

Old and New Analysis

Authorship of the Federalist Papers

- 85 essays advocating ratification of the US constitution
- Written under the pseudonym "Publius" (really, Alexander **Hamilton**, James **Madison**, John **Jay**)

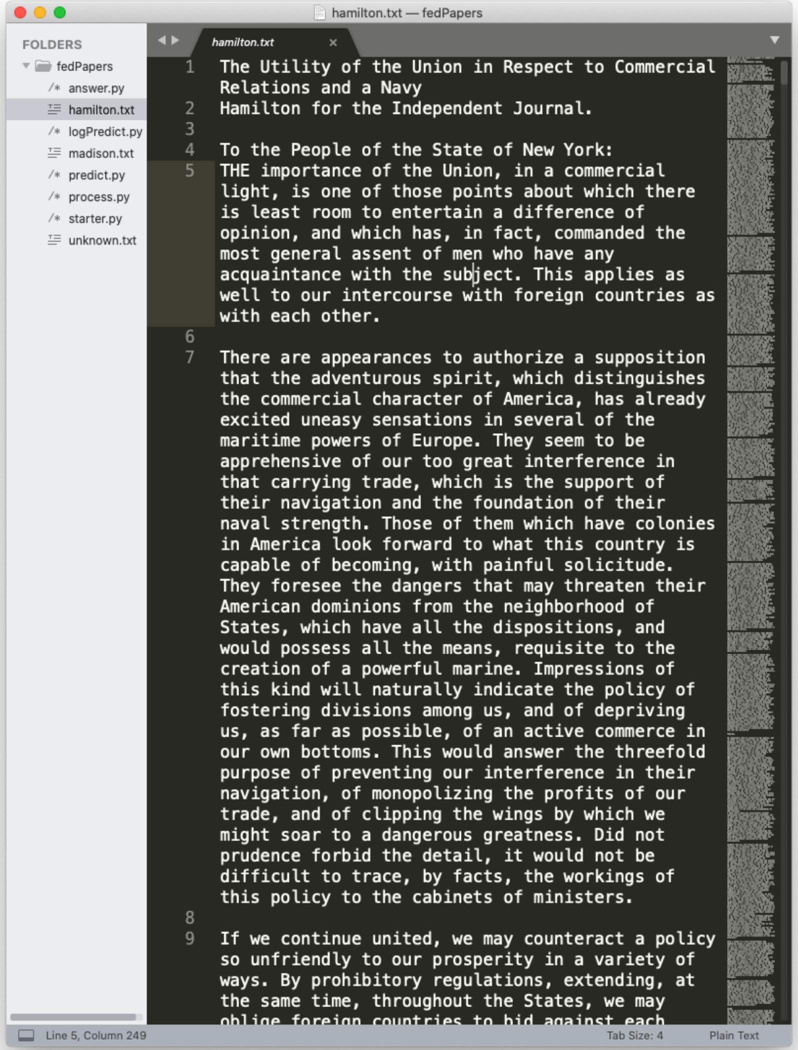
Who wrote which essays?

- Analyze probability of words in each essay and compare against word distributions from known writings of three authors

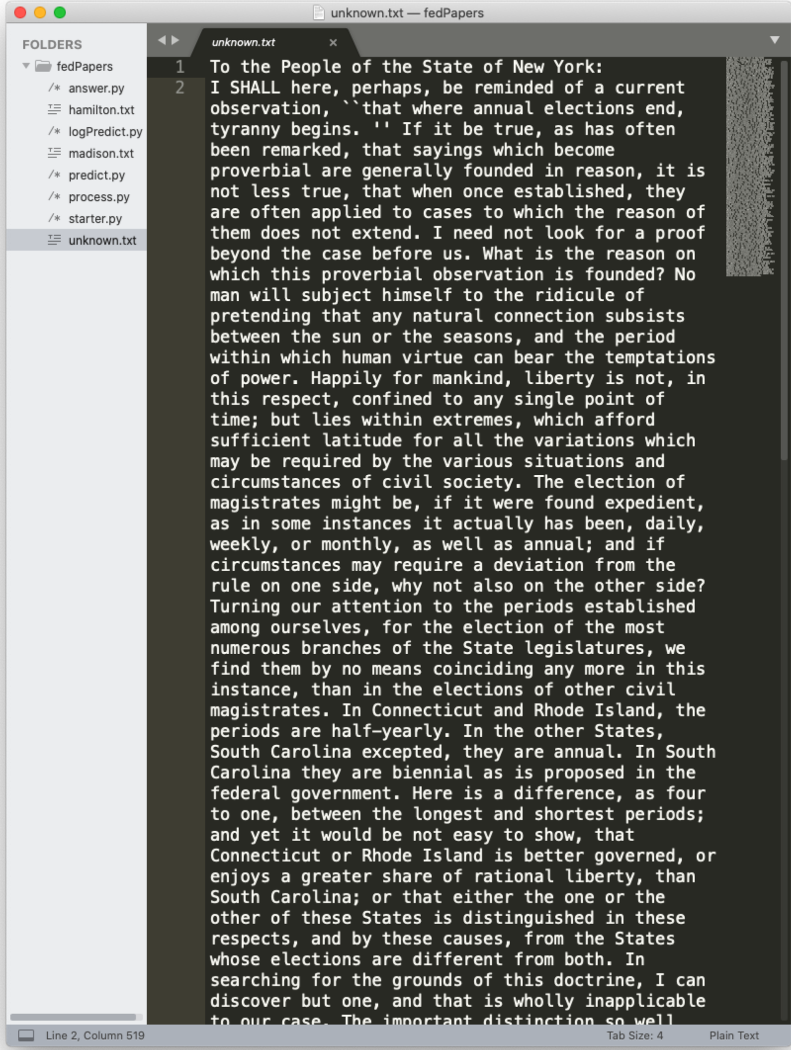
Who wrote Federalist Paper 53?

madison.txt

hamilton.txt



unknown.txt



Where to start?

We have words, we want to know probability of authorship. We also know probability of words given author…

Well hello again…

Who wrote Federalist Paper 53?

Prob Document given Hamilton

Prior belief it was Hamilton

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Prob Hamilton given Document

Prob of the document???

Who wrote Federalist Paper 53?

Model document as a
multinomial where we care
about count of words

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Who wrote Federalist Paper 53?

Loop over unique words

Prob hamilton would write word i

Prior belief it was Hamilton

Number of times word i is in the doc

$$P(H|D) = \frac{\binom{n}{c_1 \ldots c_k} \cdot \prod_i h_i^{c_i} \cdot P(H)}{P(D)}$$

Prob Hamilton given Document

Prob of the document???

Who wrote Federalist Paper 53?

Prob that Hamilton wrote it

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

$$= \frac{P(H) \cdot \binom{n}{c_1 \ldots c_m} \cdot \prod_i h_i^{c_i}}{P(D)}$$

Prob that Madison wrote it

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

$$= \frac{P(M) \cdot \binom{n}{c_1 \ldots c_m} \cdot \prod_i m_i^{c_i}}{P(D)}$$

$$\frac{P(H|D)}{P(M|D)} = \frac{P(H) \cdot \binom{n}{c_1 \ldots c_k} \cdot \prod_i h_i^{c_i}}{P(D)} \Big/ \frac{P(M) \cdot \binom{n}{c_1 \ldots c_k} \cdot \prod_i m_i^{c_i}}{P(D)}$$

$$= \frac{\prod_i m_i^{c_i}}{\prod_i h_i^{c_i}}$$

To the code

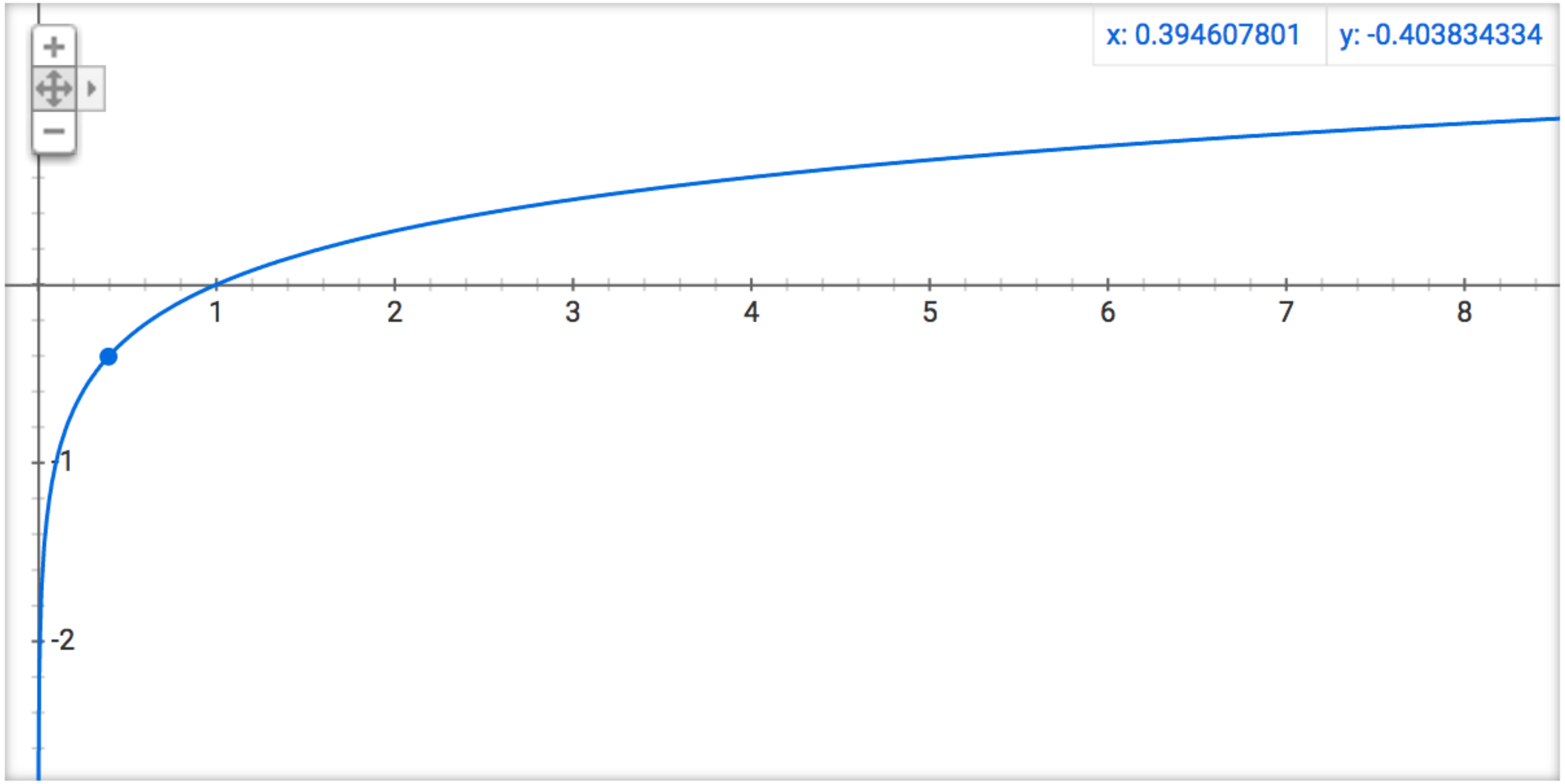
What happened?

All our probabilities are zero…

Log Review

$$e^y = x \qquad\qquad \log(x) = y$$

Graph for log(x)



More info

Log Identities

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$\log(a/b) = \log(a) - \log(b)$$

$$\log(a^n) = n \cdot \log(a)$$

Products become sums!

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$\log(\prod_i a_i) = \sum_i \log(a_i)$$

* Spoiler alert: This is important because the product of many small numbers gets hard for computers to represent.

Use logs when probabilities become too small!

The doc

Number of times word *i* shows up in the doc

Hamilton wrote it

Maddison wrote it

$$\frac{P(H|D)}{P(M|D)} = \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}}$$

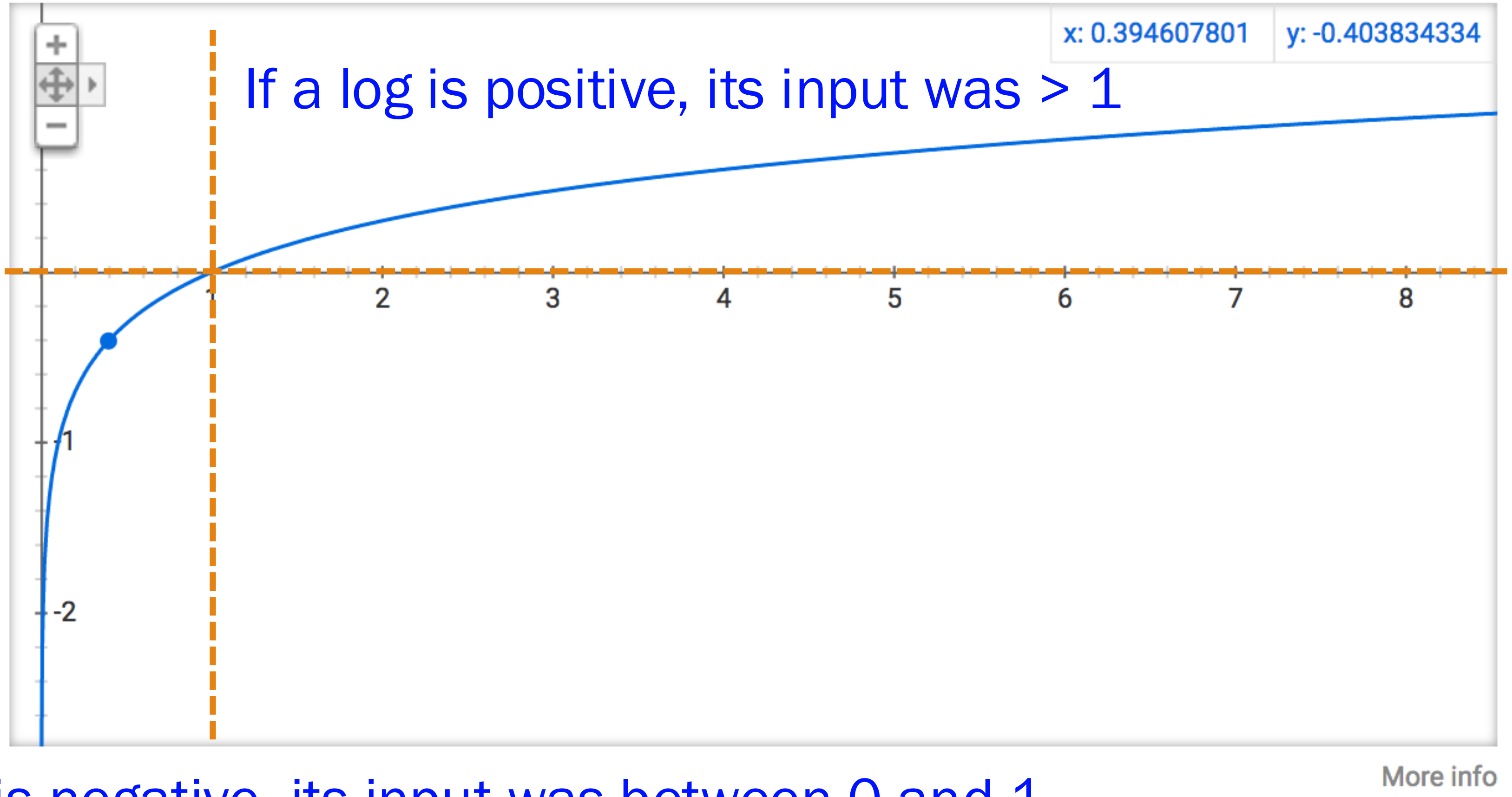
h_i = Prob Hamilton writes word *i*

m_i = Prob Maddison writes word *i*

$$\begin{aligned}\log \frac{P(H|D)}{P(M|D)} &= \log \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}} \\ &= \sum_i \log \underline{h_i^{c_i}} - \sum_i \log m_i^{c_i} \\ &= \boxed{\sum_i c_i \cdot \log h_i} - \boxed{\sum_i c_i \log m_i}\end{aligned}$$

What does it mean if a log value is positive / negative

Graph for log(x)



If a log is negative, its input was between 0 and 1

Use logs when probabilities become too small!

$$
\begin{aligned}
\log \frac{P(H|D)}{P(M|D)} &= \log \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}} \\
&= \sum_i \log h_i^{c_i} - \sum_i \log m_i^{c_i} \\
&= \sum_i c_i \cdot \log h_i - \sum_i c_i \log m_i \\
&= -1344
\end{aligned}
$$

Hamilton Term	-12925
Madison Term	-11581

$$\frac{P(H|D)}{P(M|D)} < 1$$

Madison wrote it!

Great Question
(Bonus if we have time)

How does python sample from a Gaussian?

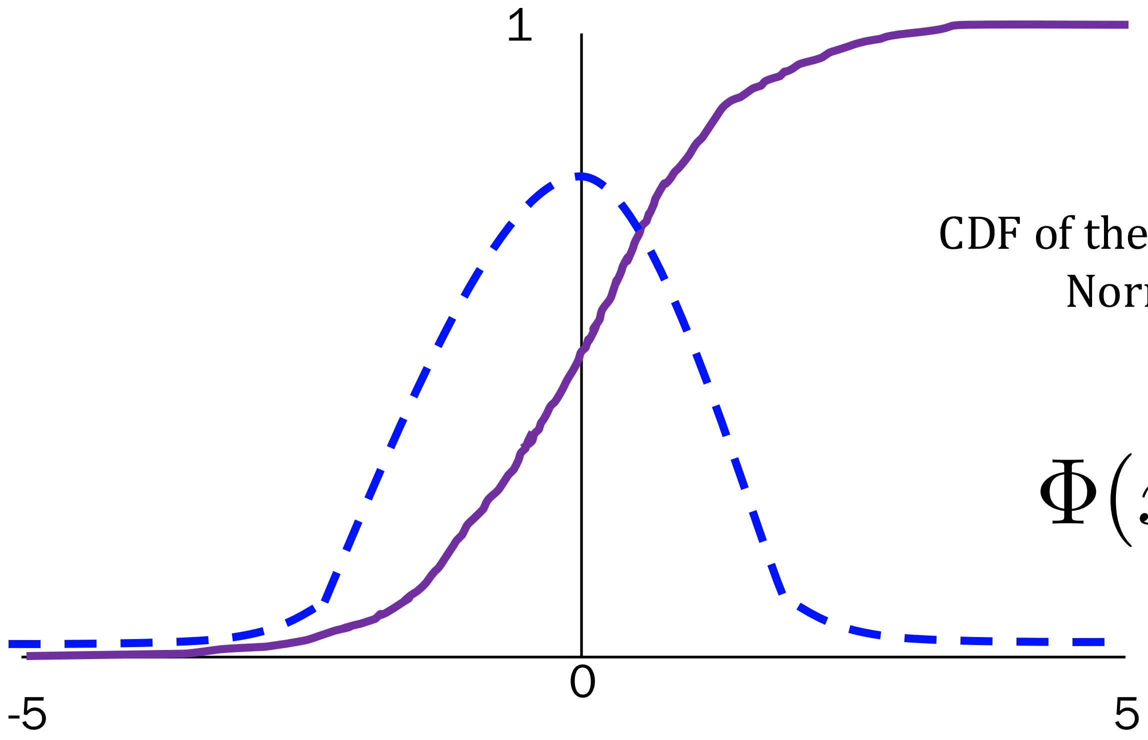
```
from random import *

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

How does this work?

3.79317794179
5.19104589315
4.209360629
5.39633891584
7.10044176511
6.72655475942
5.51485158841
4.94570606131
6.14724644482
4.73774184354

How Does a Computer Sample a Normal?

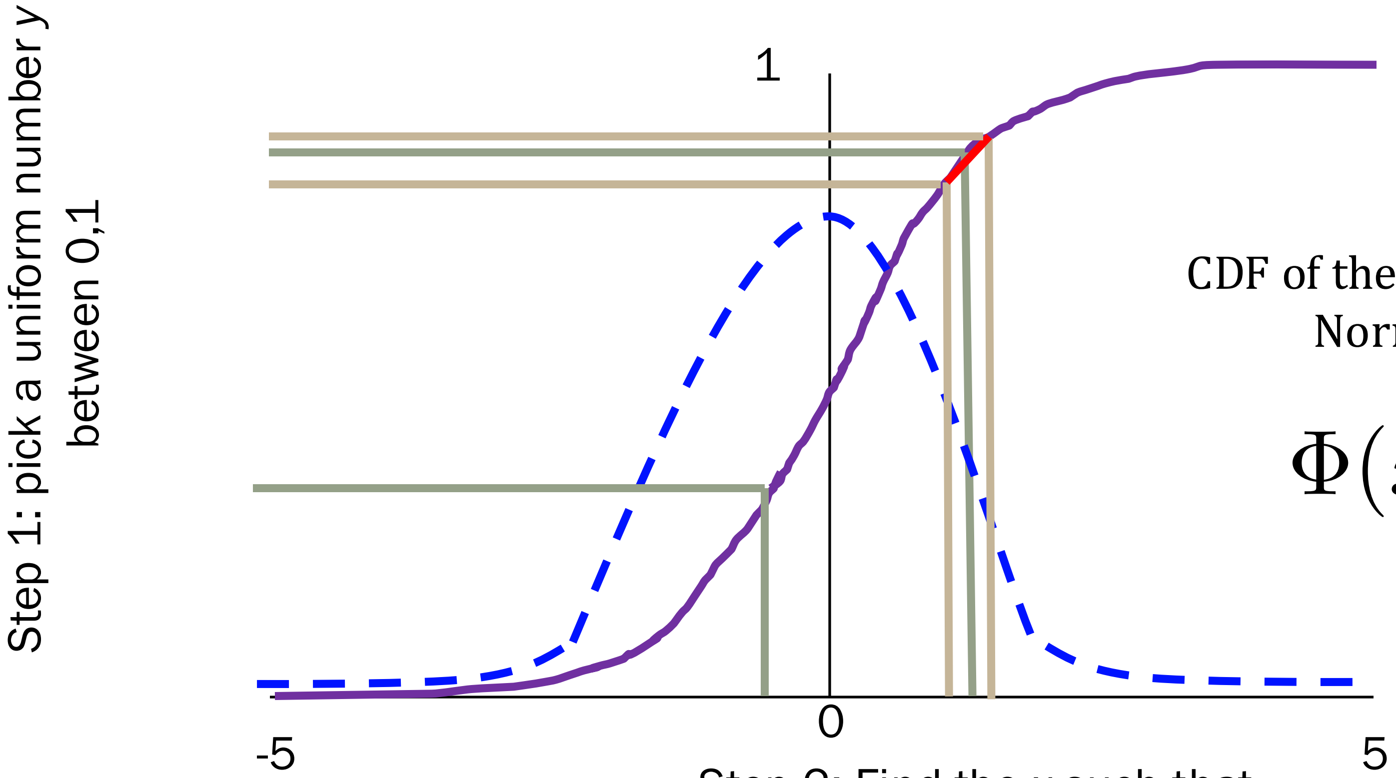


CDF of the Standard Normal

$$\Phi(x)$$

How Does a Computer Sample a Normal?

Inverse Transform Sampling



Step 2: Find the *x* such that

$$\Phi(x) = y$$

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

Sample 1:
1.201234

Sample 2:
-0.45123

Further reading: Box–Muller transform

Have a Great Weekend