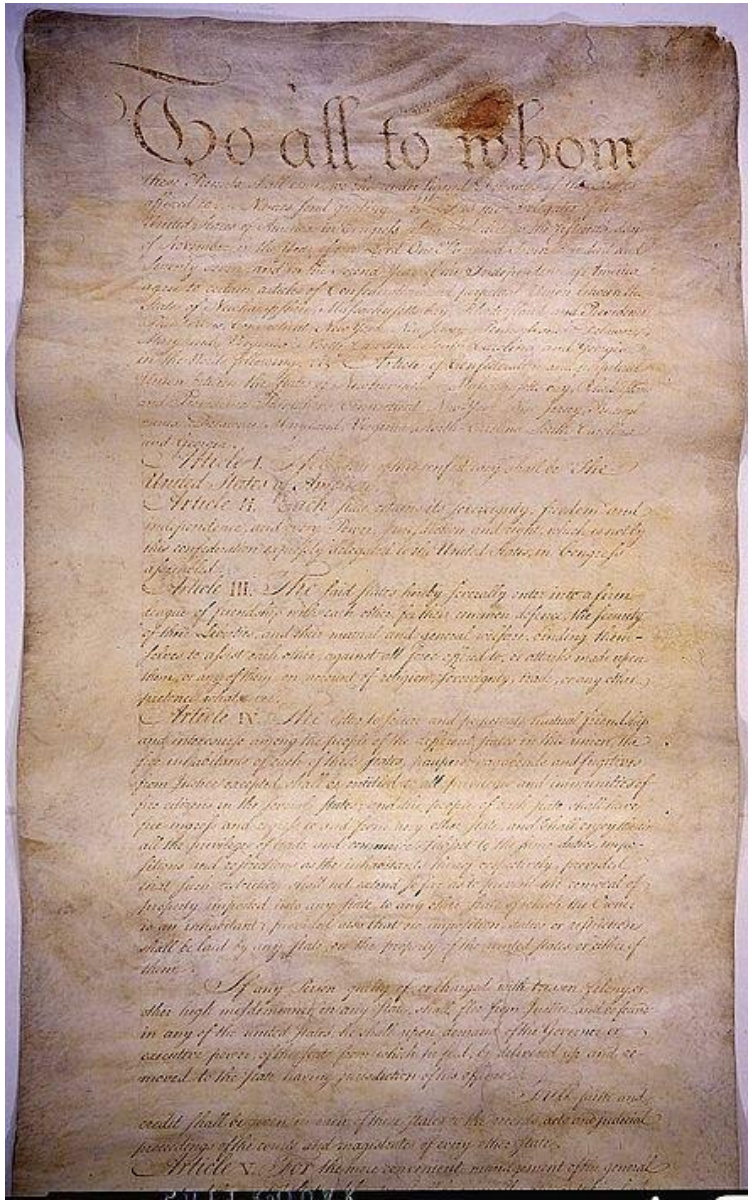
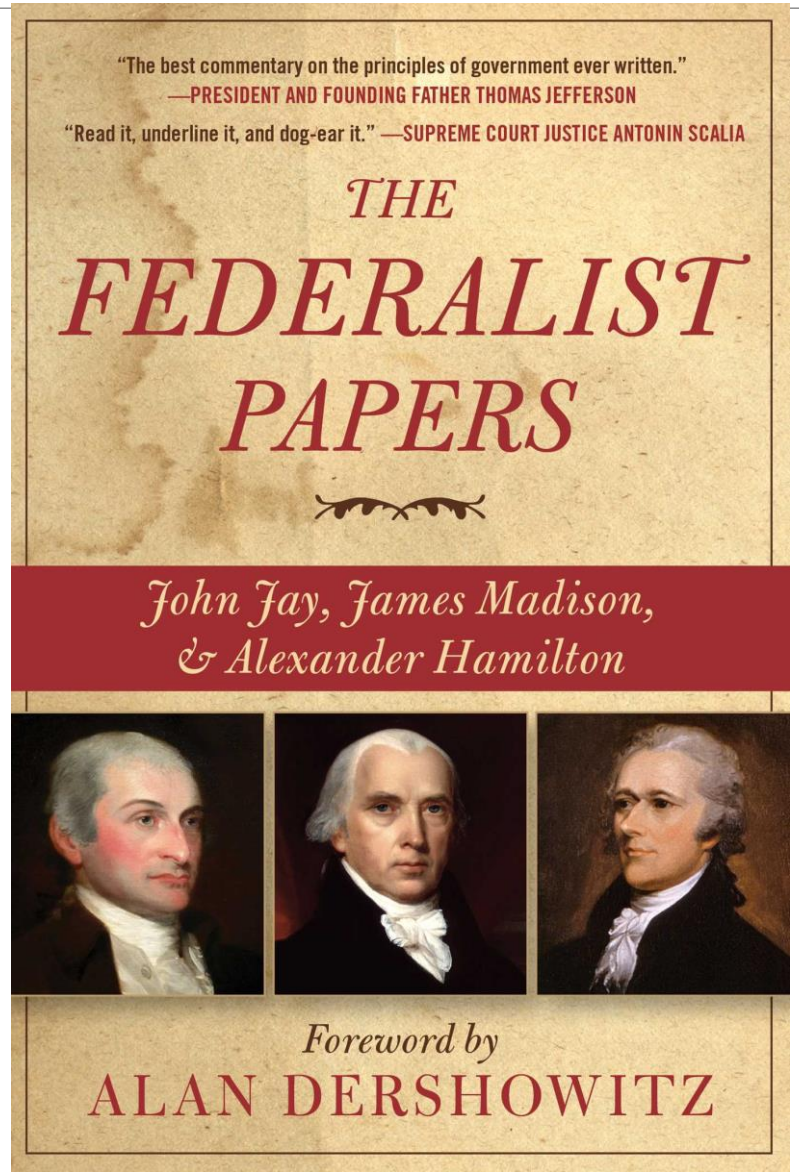
Multinomial

CS109, Stanford University

Exciting Day!

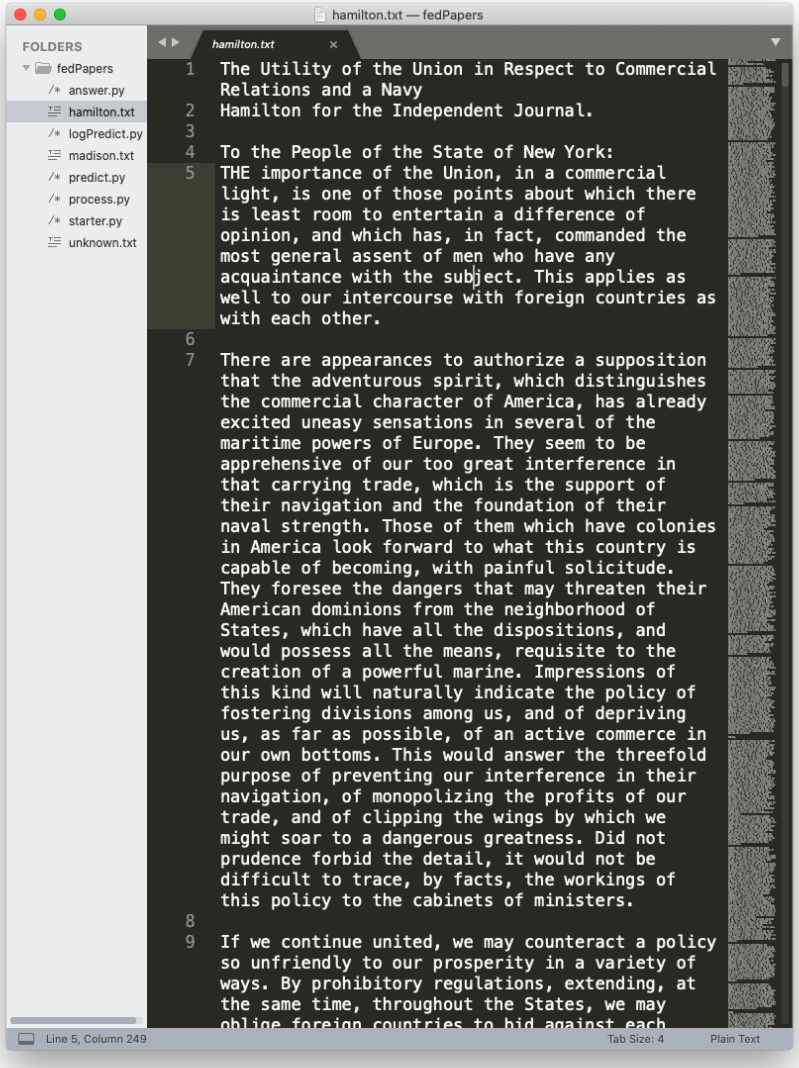




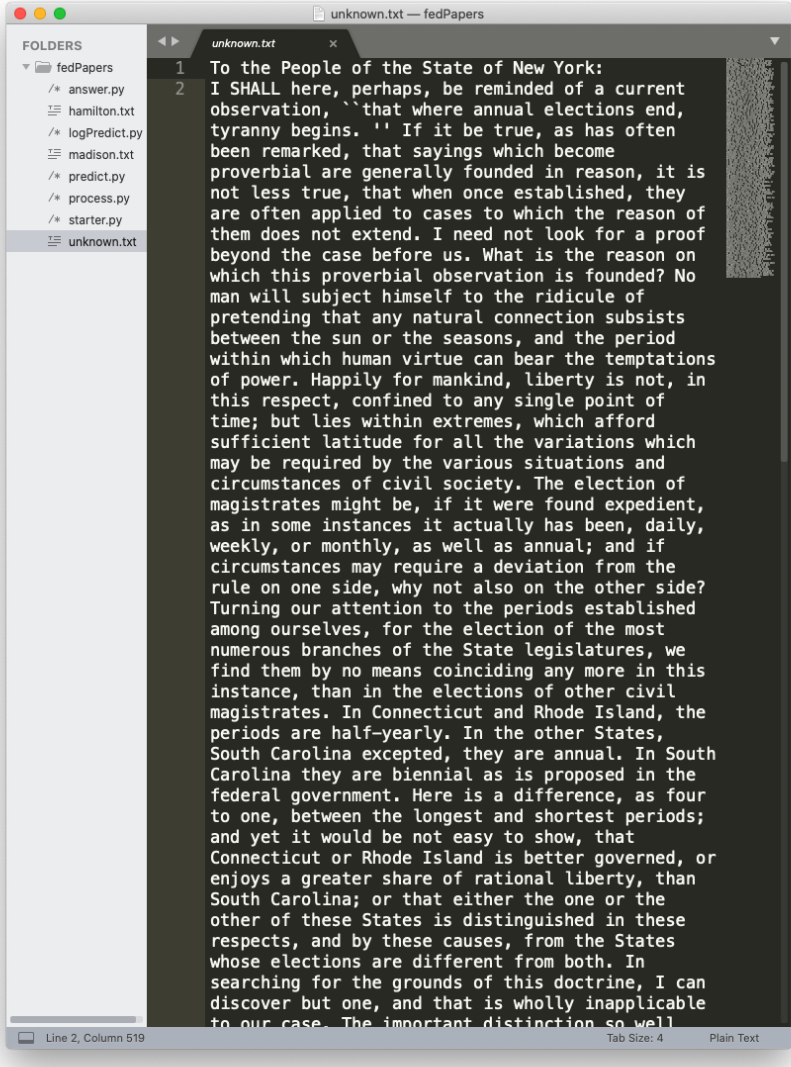
Who wrote Federalist Paper 53?

madison.txt

hamilton.txt

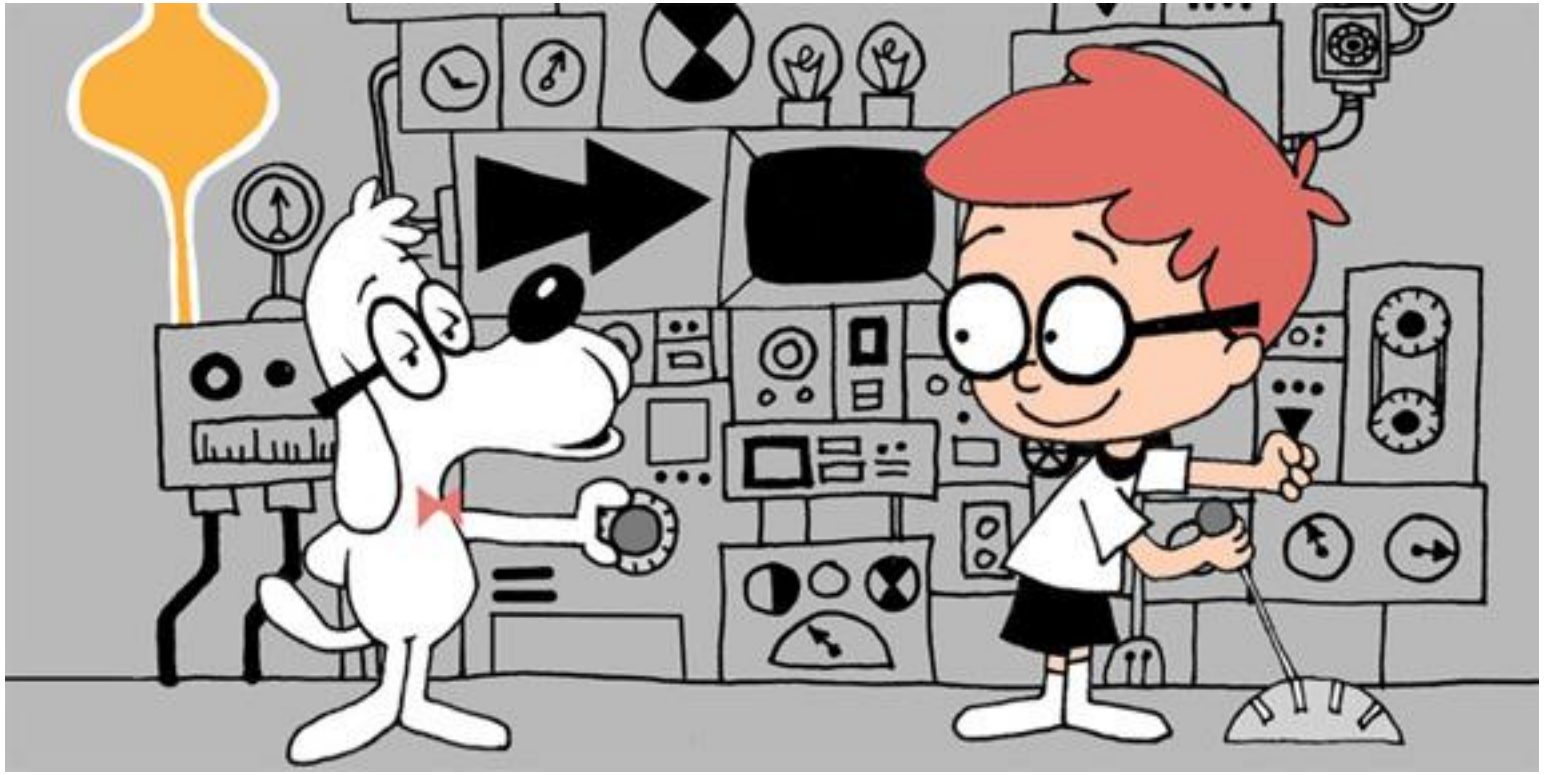


unknown.txt



First, some review

Recall the good times



Permutations

$n!$

How many ways are there to order n objects?

How Many Ways Are There to Order…

(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)

4 heads and 6 tails?

$$\frac{10!}{4! \cdot 6!} = \binom{10}{4}$$

How Many Ways Are There to Order...

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

3 As, 5 Bs and 2Cs

$$\frac{10!}{3! \cdot 5! \cdot 2!} = \binom{10}{3, 5, 2}$$

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.

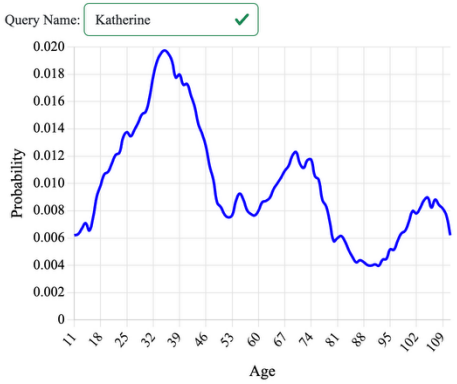
Where are we in CS109?

Compare and Contrast Many Examples

Age from C14

Updated Delivery Prob

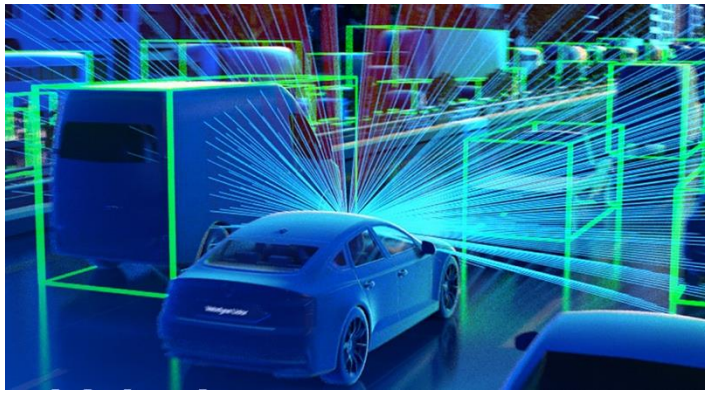
Age from Name



Hidden Chambers

Stanford Eye Test

Updating Lidar Belief



Joint Table for 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

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

Joint Table for 10 Random Variables?

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

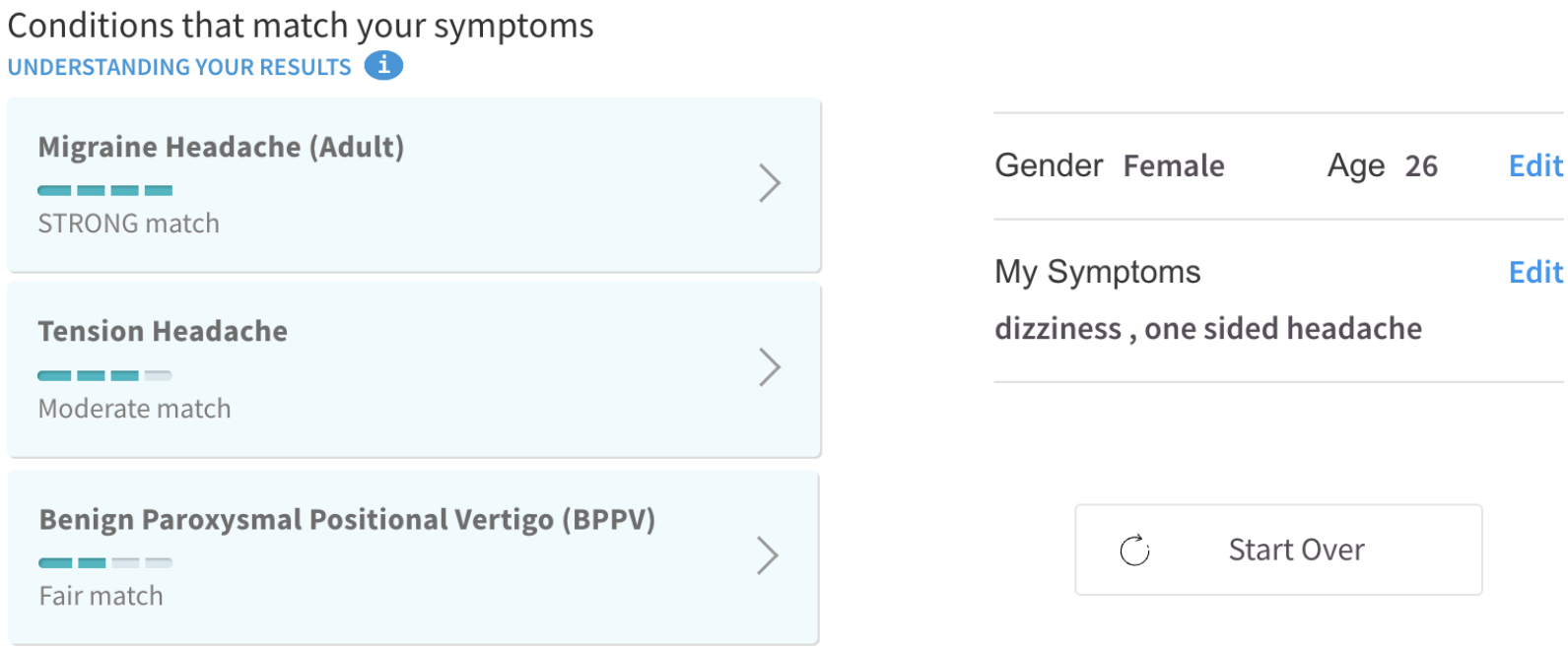
$$\text{\# unique assignments} = 5^{10}$$

10 million entries in your joint table.

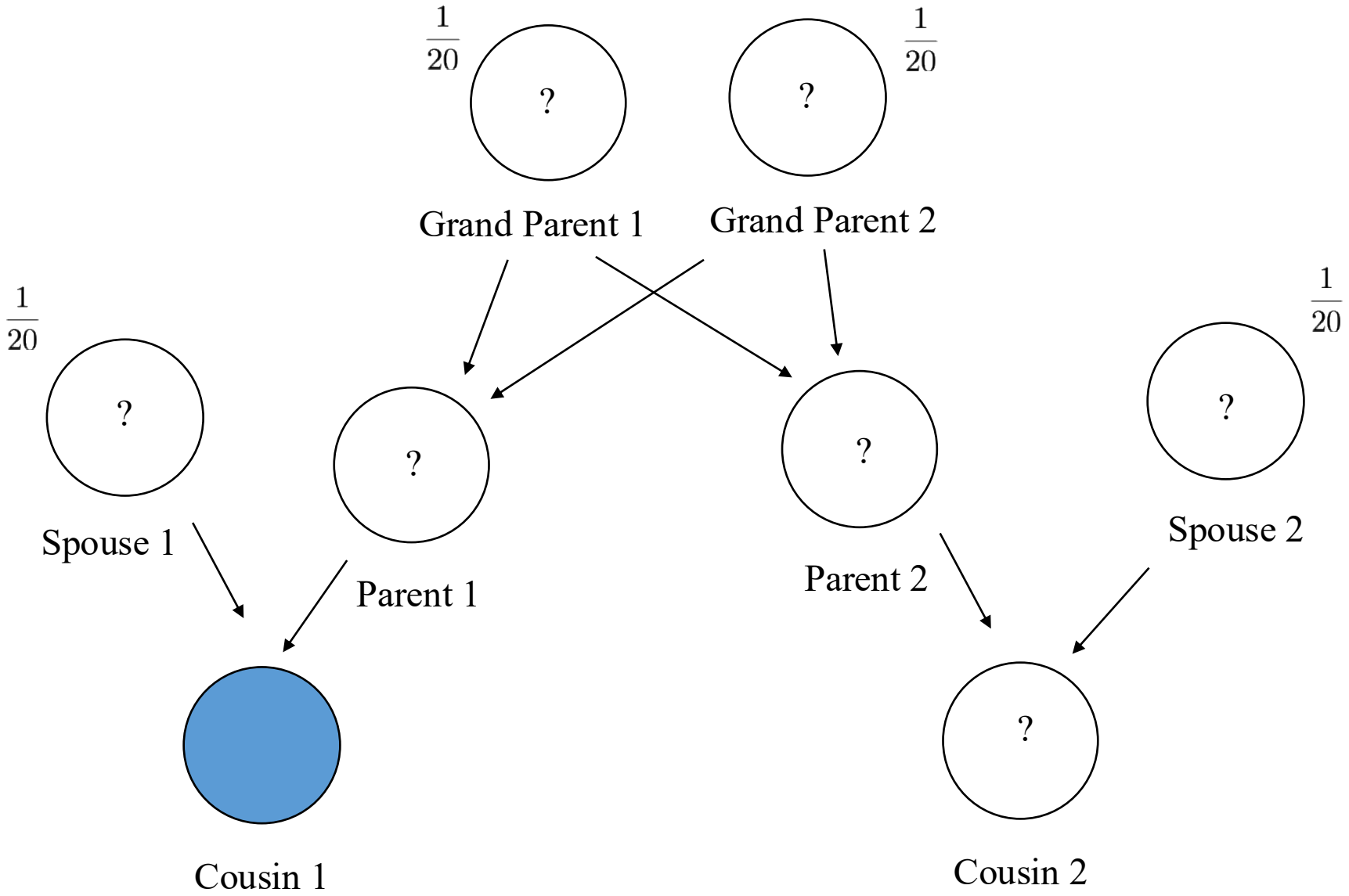
So, we are going to need models …

… **probabilistic models** …

Multiple Random Variables. Start of Digital Revolution



You observe that someone has a **recessive** gene.
What is the probability that their **cousin** has the same recessive gene?



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)

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

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

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

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

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

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

# of times a six appears

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

choose where the sixes appear

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 $n, p_1, p_2, \ldots, p_m$

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

p_i: probability of outcome i on a single trial

Where do we get p_i from?

Most useful when probabilities are not equal

You have a funny shaped six-sided die with the following probabilities:

- $P(1) = 0.2, \quad P(2) = 0.3$
- $P(3) = 0.1, \quad P(4) = 0.1$
- $P(5) = 0.1, \quad P(6) = 0.2$

You roll the dice 6 times.
What is the probability of getting exactly: **two 2s, two 4s, and two 6s?**

The Federalist Papers

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 Drugs with a credit-card. Drugs are great.
So are credit-cards. Risk free Drugs! Click for free."
n = 18

$$P\left(\begin{array}{l}\text{Drugs = 3}\\ \text{Free = 2}\\ \text{Risk = 1}\\ \text{Credit-card: 2}\\ \ldots\\ \text{For = 2}\end{array} \middle| \text{spam}\right) = \frac{n!}{3!2!\ldots2!} p_{\text{drugs}}^3 p_{\text{free}}^2 \cdots p_{\text{for}}^2$$

It's a Multinomial!

Probability of seeing this document | spam

The probability of a word in spam email being Drugs

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)

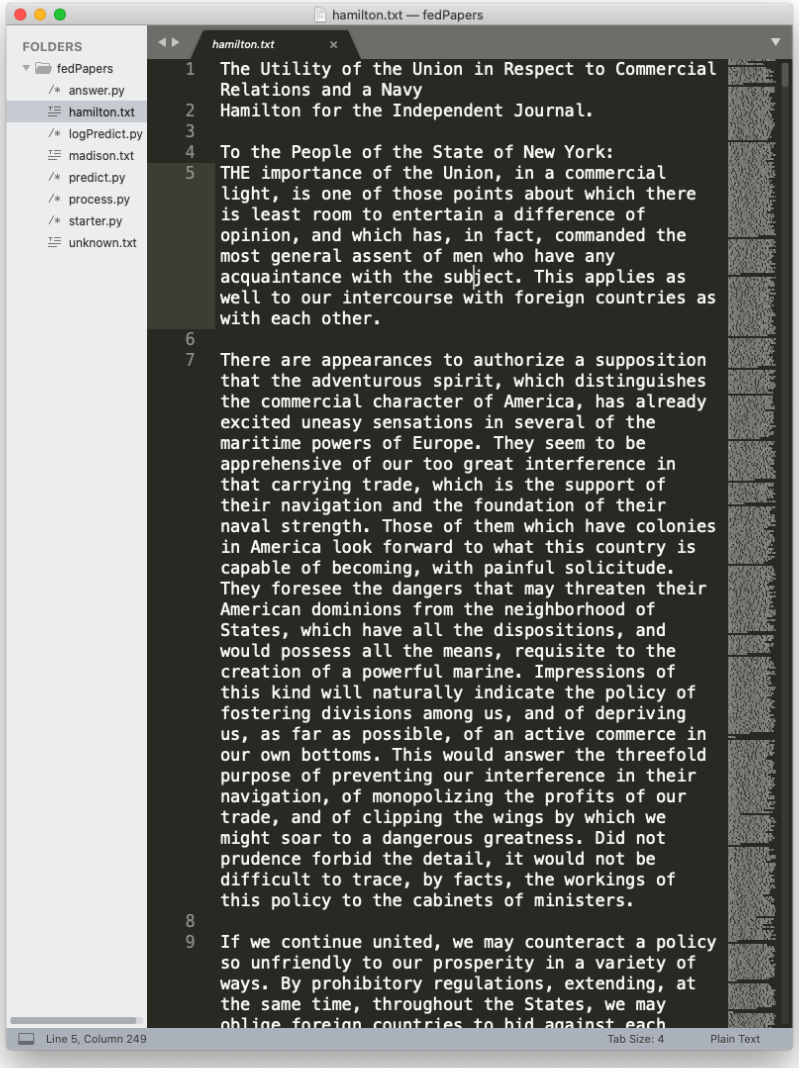
Who wrote which essays?

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

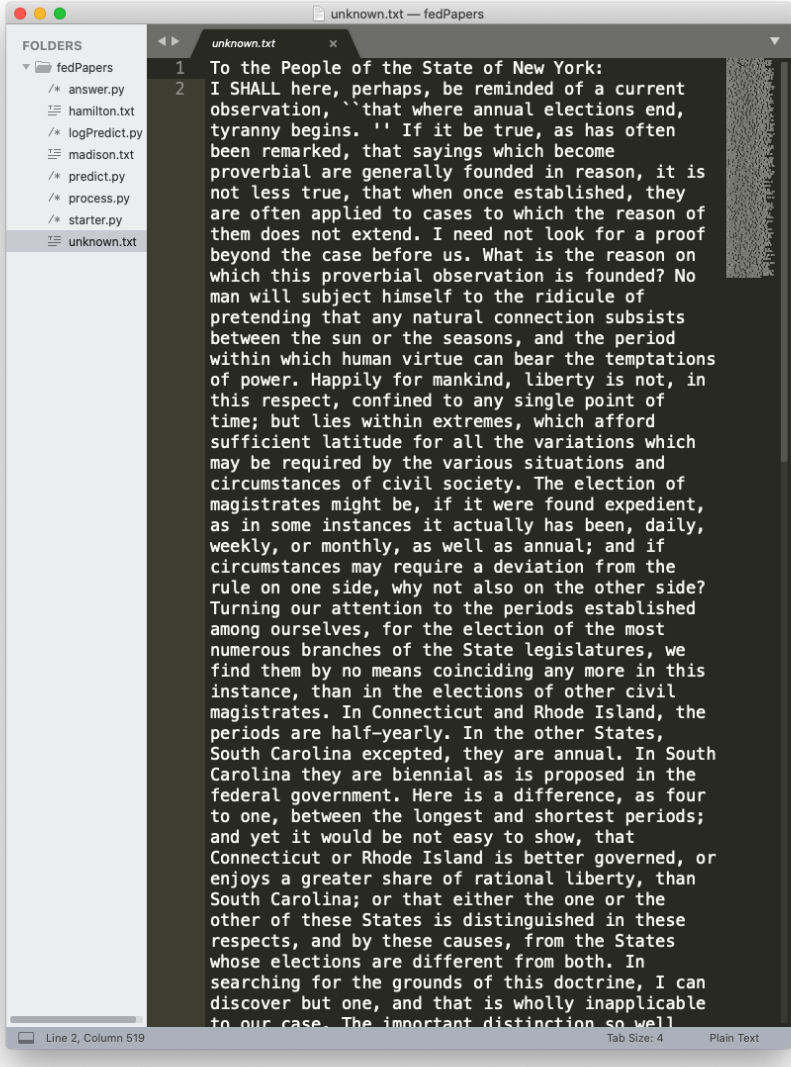
Who wrote Federalist Paper 53?

madison.txt

hamilton.txt



unknown.txt

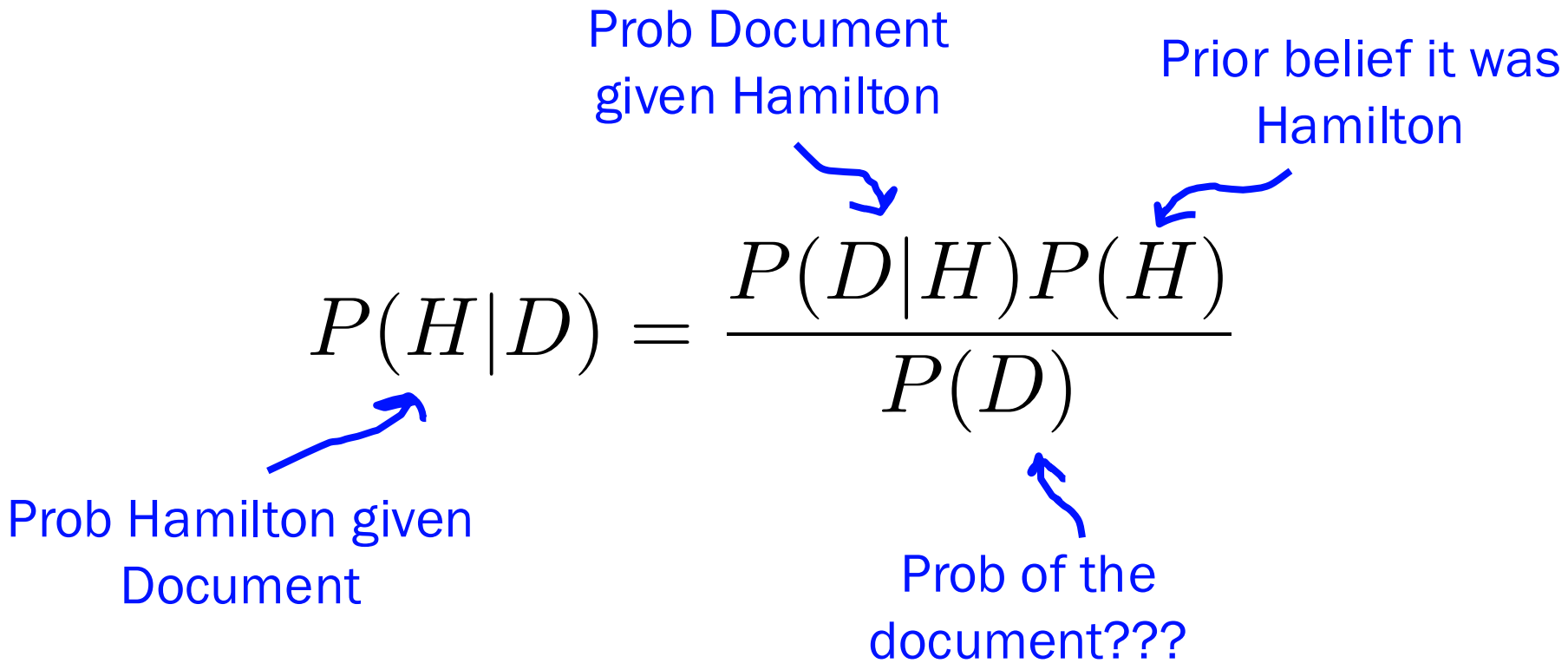


Where to start?

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?



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?

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

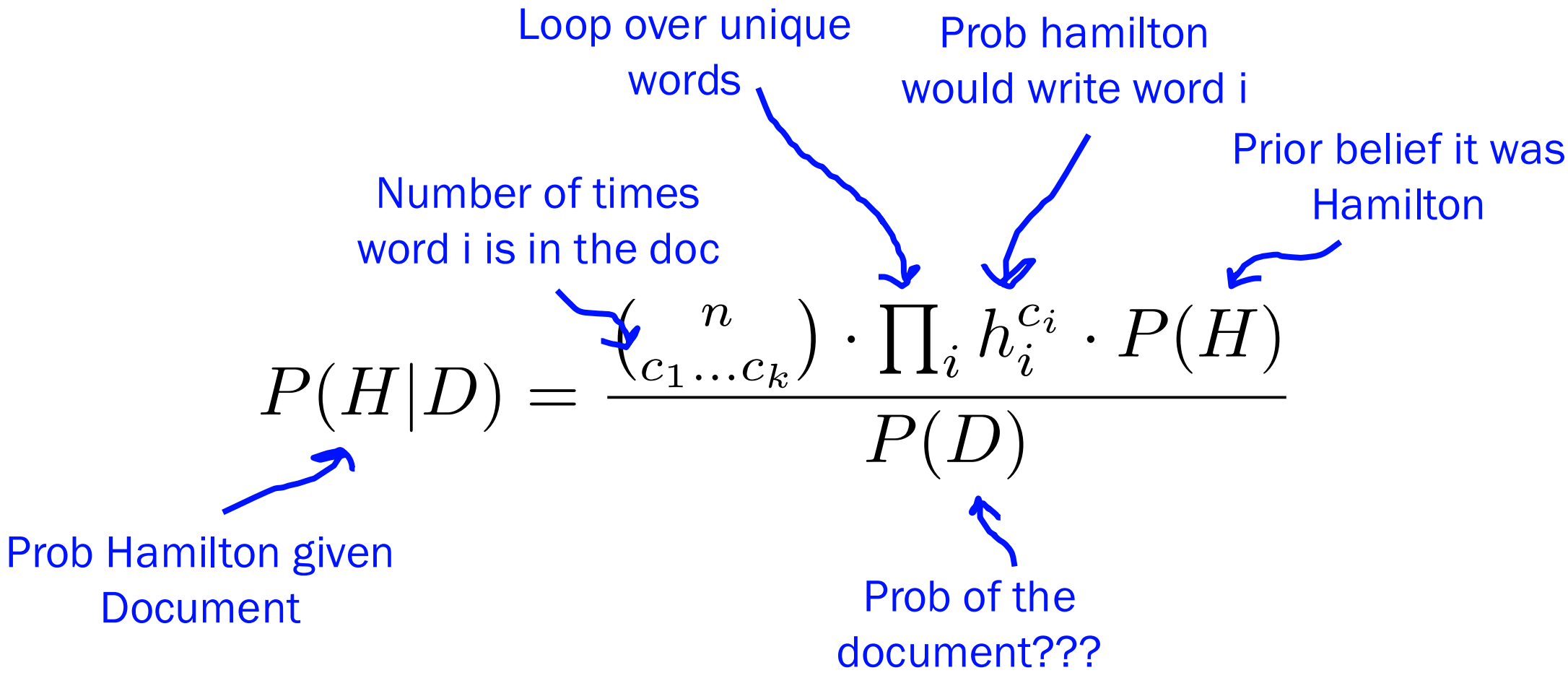
Number of words in the document

$$P(D|H) = \binom{n}{c_1, \ldots, c_k} \cdot \prod_{i=1}^{k} h_i^{c_i}$$

Number of times word i is in the doc

Prob Hamilton would write word i

Who wrote Federalist Paper 53?



Who wrote Federalist Paper 53?

Prob that Hamilton wrote it

$$P(H \mid D) = \frac{P(D \mid H)\,P(H)}{P(D)} = \frac{P(H)\binom{n}{c_1,\ldots,c_k}\prod_{i=1}^{k} h_i^{c_i}}{P(D)}$$

Prob that Madison wrote it

$$P(M \mid D) = \frac{P(D \mid M)\,P(M)}{P(D)} = \frac{P(M)\binom{n}{c_1,\ldots,c_k}\prod_{i=1}^{k} m_i^{c_i}}{P(D)}$$

$$\frac{P(H \mid D)}{P(M \mid D)} = \frac{P(H)\binom{n}{c_1,\ldots,c_k}\prod_{i=1}^{k} h_i^{c_i}}{P(D)} \Bigg/ \frac{P(M)\binom{n}{c_1,\ldots,c_k}\prod_{i=1}^{k} m_i^{c_i}}{P(D)}$$

$$\frac{P(H \mid D)}{P(M \mid D)} = \frac{\prod_{i=1}^{k} h_i^{c_i}}{\prod_{i=1}^{k} m_i^{c_i}}$$

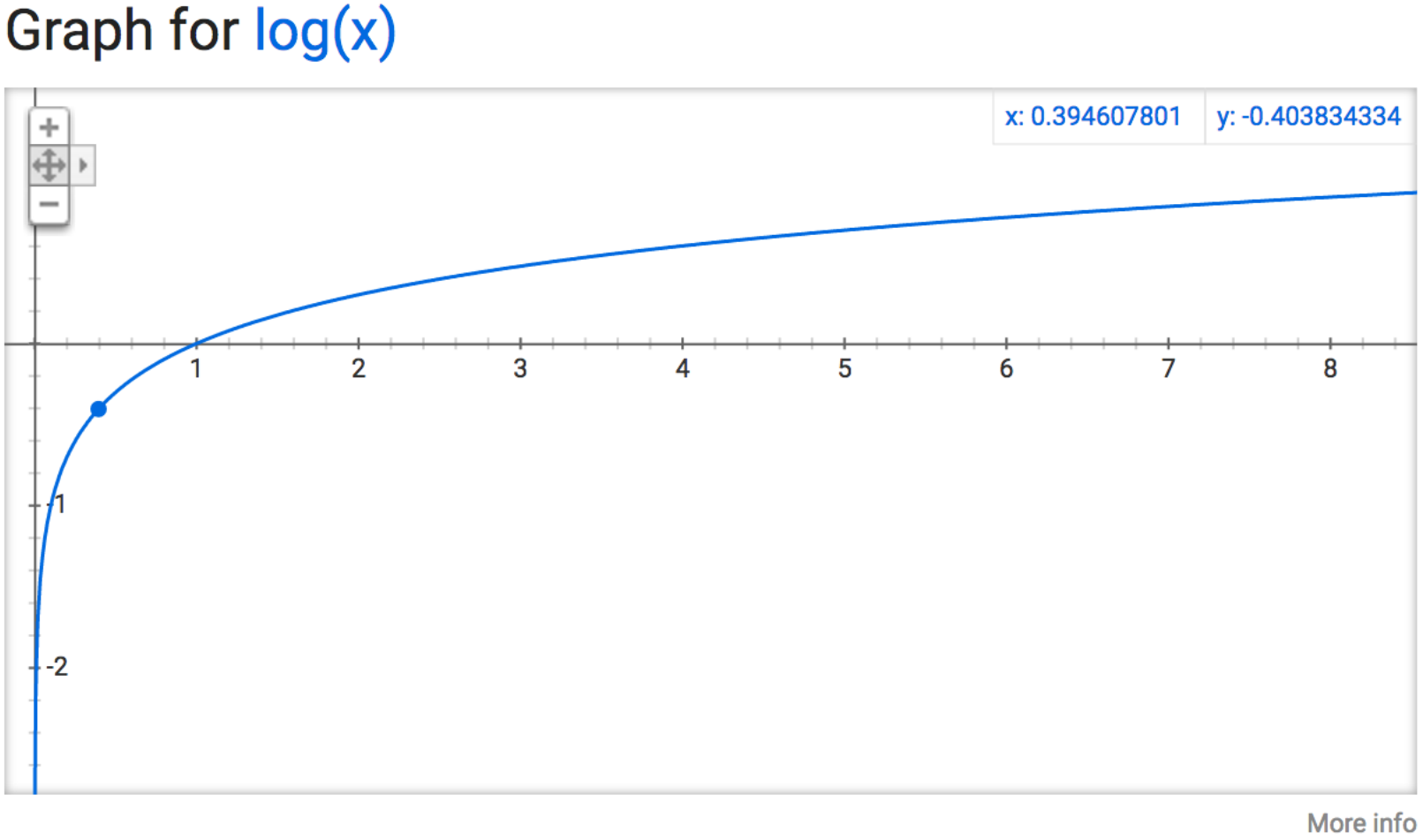
To the code

What happened?

All our probabilities are zero…

Log Review

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



Log Identities

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

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

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

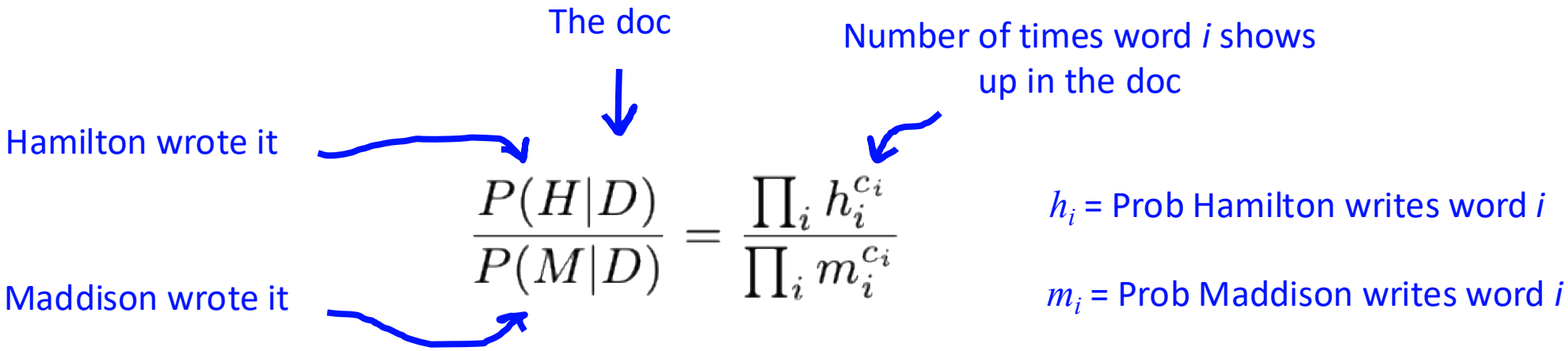
Products become sums!

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

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

* 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!

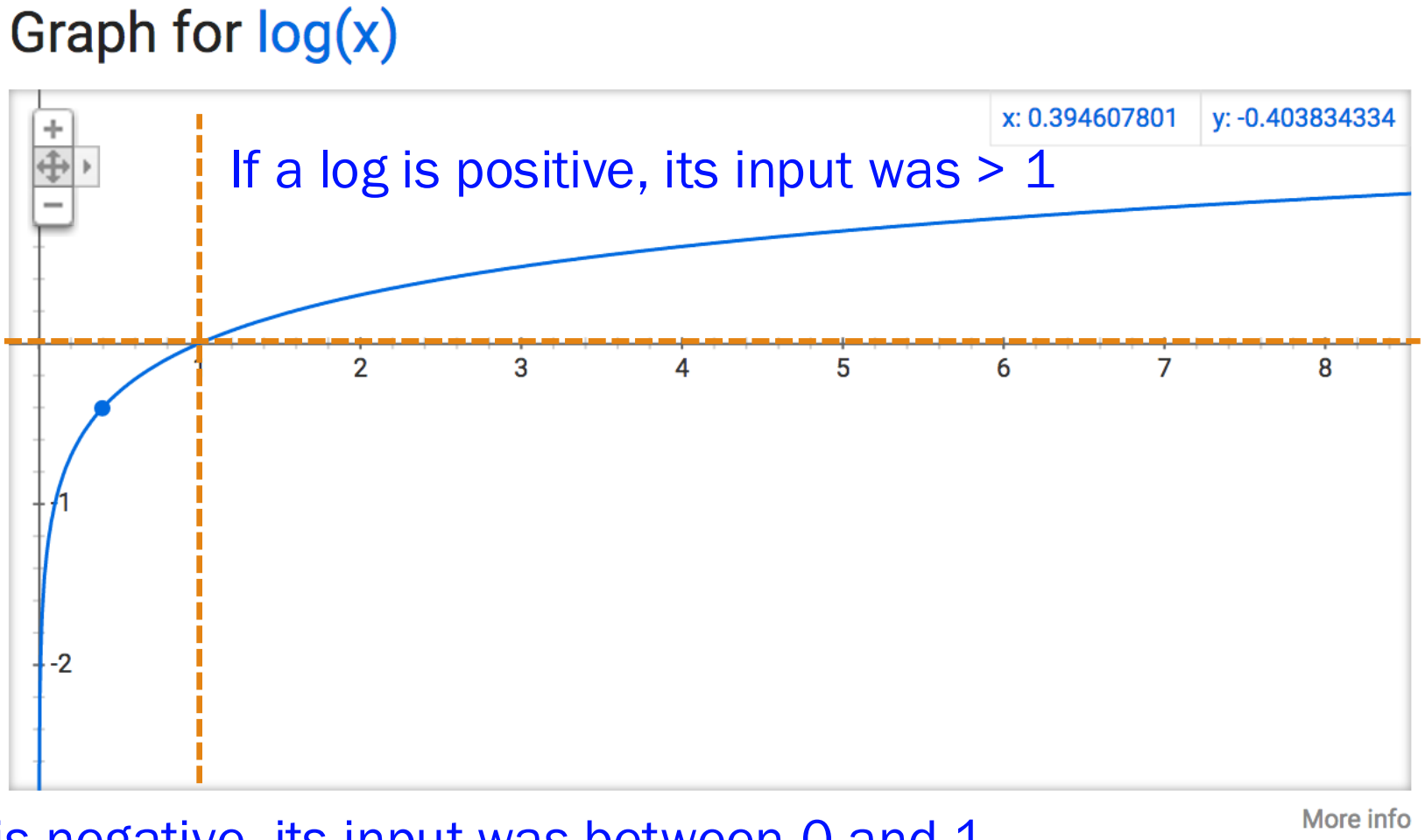


$$\log \frac{P(H|D)}{P(M|D)} = \log \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}}$$

$$= \sum_i \log h_i^{c_i} - \sum_i \log m_i^{c_i}$$

$$= \sum_i c_i \cdot \log h_i - \sum_i c_i \log m_i$$

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



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

Use logs when probabilities become too small!

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

$$= \sum_i \log h_i^{c_i} - \sum_i \log m_i^{c_i}$$

Hamilton Term	-12925
Madison Term	-11581

$$= \sum_i c_i \cdot \log h_i - \sum_i c_i \log m_i$$

$$= -1344$$

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

Madison wrote it!

If time: Practice Midterm Question

Have a Wonderful Weekend!