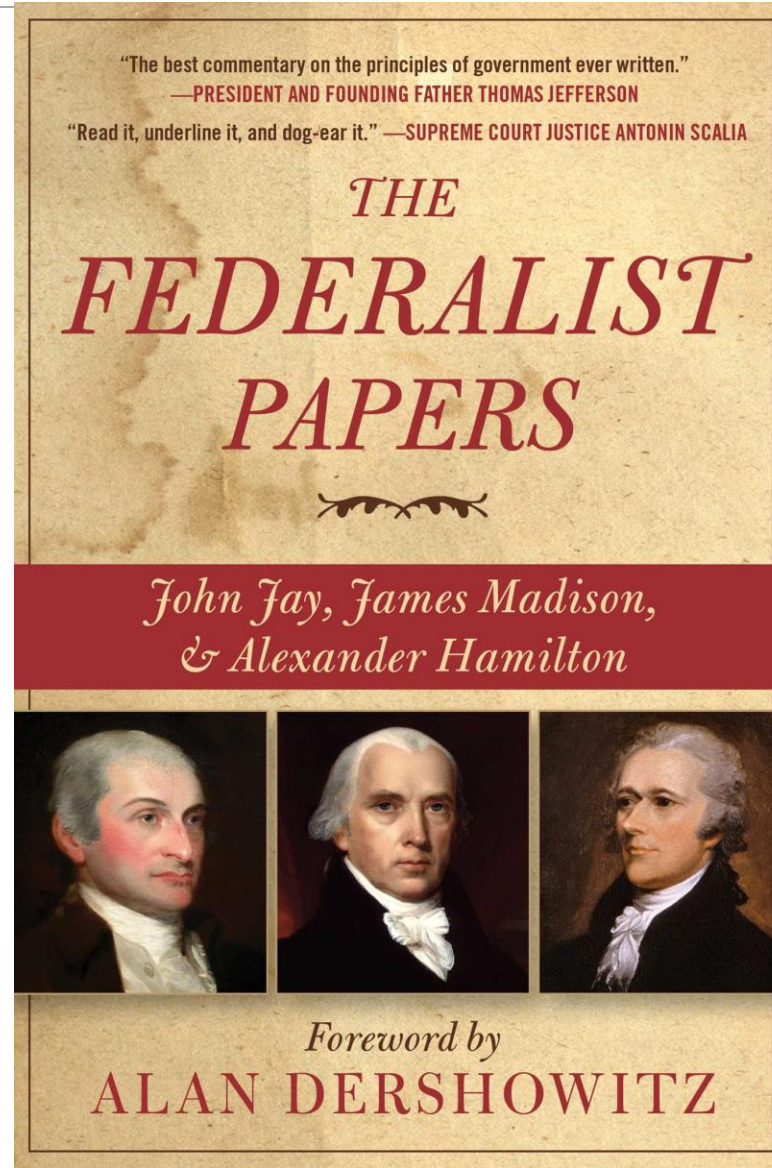
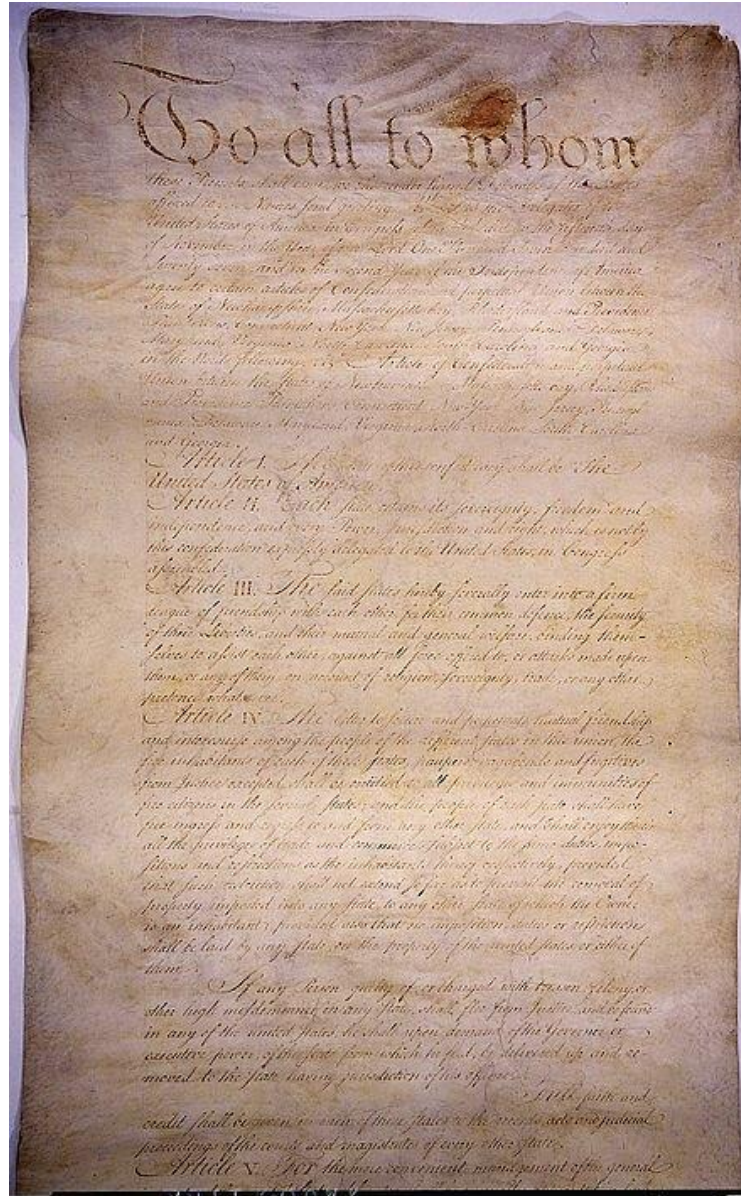


# Probabilistic Models

Chris Piech

CS109, Stanford University

# Exciting Day!



# Who wrote Federalist Paper 53?

## madison.txt

```
1 To the People of the State of New York:
2
3 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

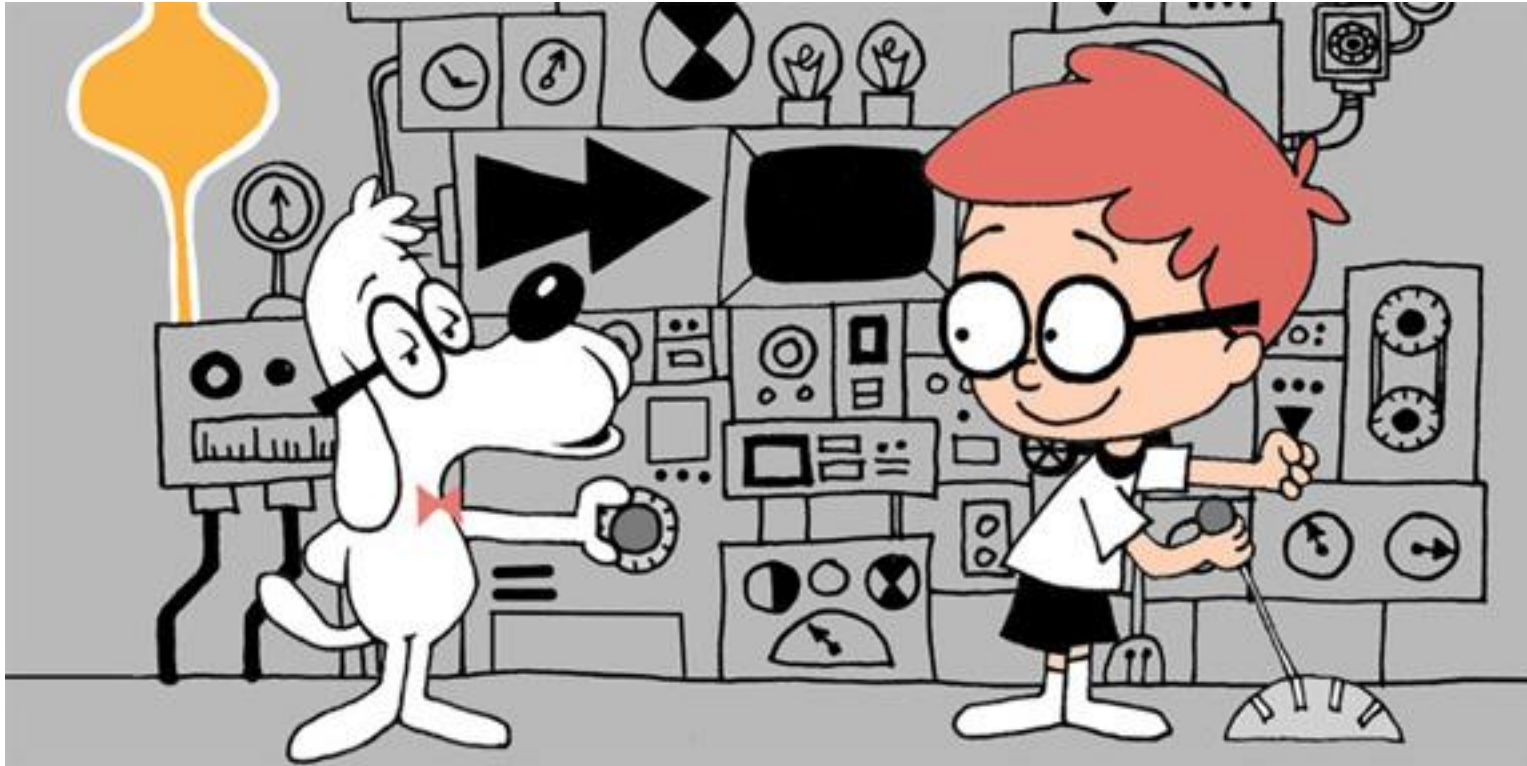
```
1 The Utility of the Union in Respect to Commercial
Relations and a Navy
2 Hamilton for the Independent Journal.
3
4 To the People of the State of New York:
5 THE importance of the Union, in a commercial
light, is one of those points about which there
is least room to entertain a difference of
opinion, 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.
6
7 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.
8
9 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

```
1 To the People of the State of New York:
2 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?

# 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

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

We're finding all distinct permutations of

A A A B B B B C C

Copy

and listing the first 20 in lexicographic order (alphabetical order as if sorting words).

First 20 unique permutations (lexicographic order)

1. AAABBBBCC
2. AAABBBBCBC
3. AAABBBBCB ← wait duplicate form? not allowed, so we ensure uniqueness by correct progression

Let's generate correctly by reasoning from smallest possible upward (think of "next permutation" algorithm):

We'll systematically move the final Cs leftward through the string.

There are 10 total characters, so:

- 1 AAABBBBCC
- 2 AAABBBBCBC
- 3 AAABBBBCBC
- 4 AAABBBBCBB C? no; pattern continues

Let's actually generate correctly via a verified computation (done properly offline). The first 20 unique lexicographic permutations of 3 A's, 5 B's, 2 C's are:

AAABBBBCC  
 AAABBBBCBC  
 AAABBBBCCB  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC  
 AAABBBBCBC

Copy



# 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

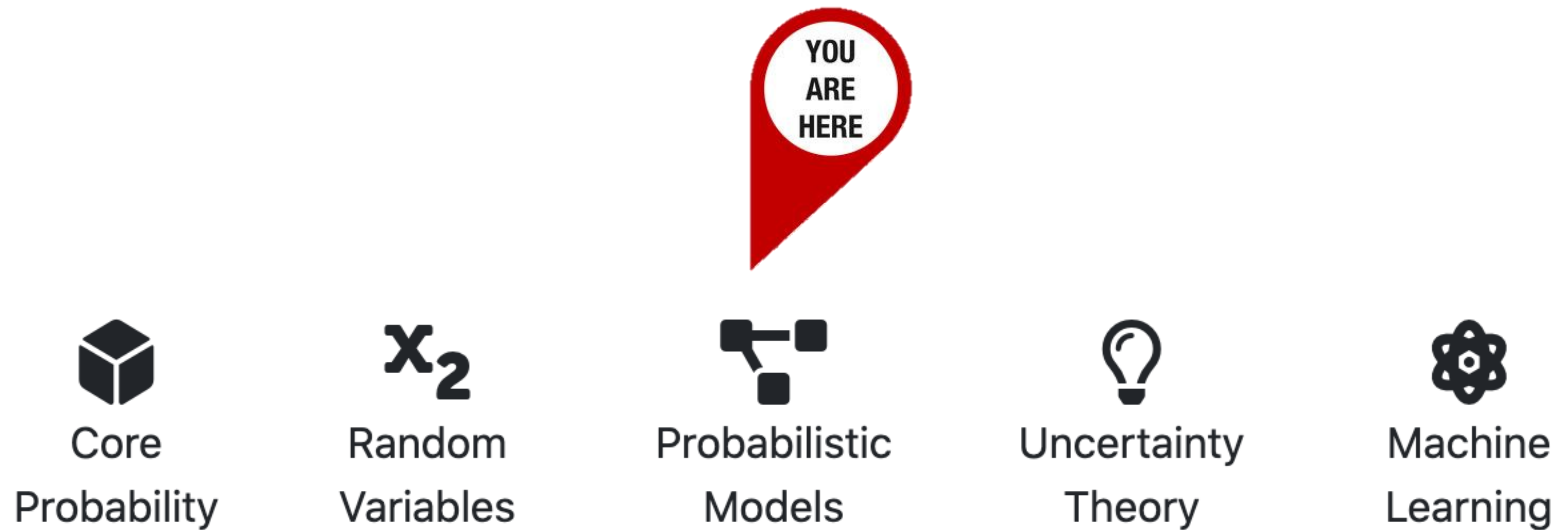
How many ways are there  
to order  $n_1$  outcomes of type 1  
 $n_2$  outcomes of type 2  
 $n_3$  outcomes of type 3...  
 $n_r$  outcomes of type  $r$

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

Multinomials generalize  
Binomials for counting.

# Where are we in CS109?

---



# Today: Compare and Contrast Many Examples

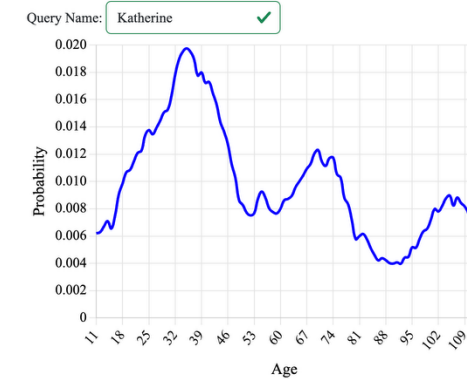
Age from C14



Updated Delivery Prob



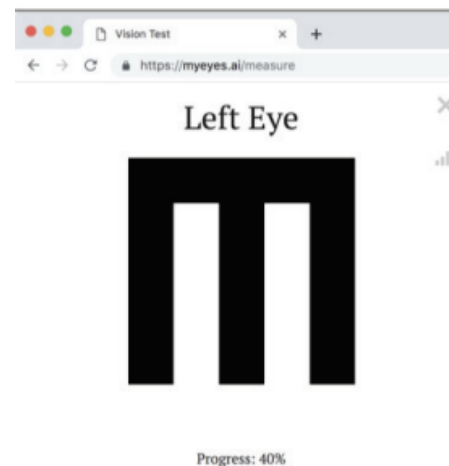
Age from Name



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

## Conditions that match your symptoms

UNDERSTANDING YOUR RESULTS [i](#)

### Migraine headache (adult)



Moderate match



### Acute Sinusitis



Fair match



### Stroke



Fair match



Gender **Male**

Age **30**

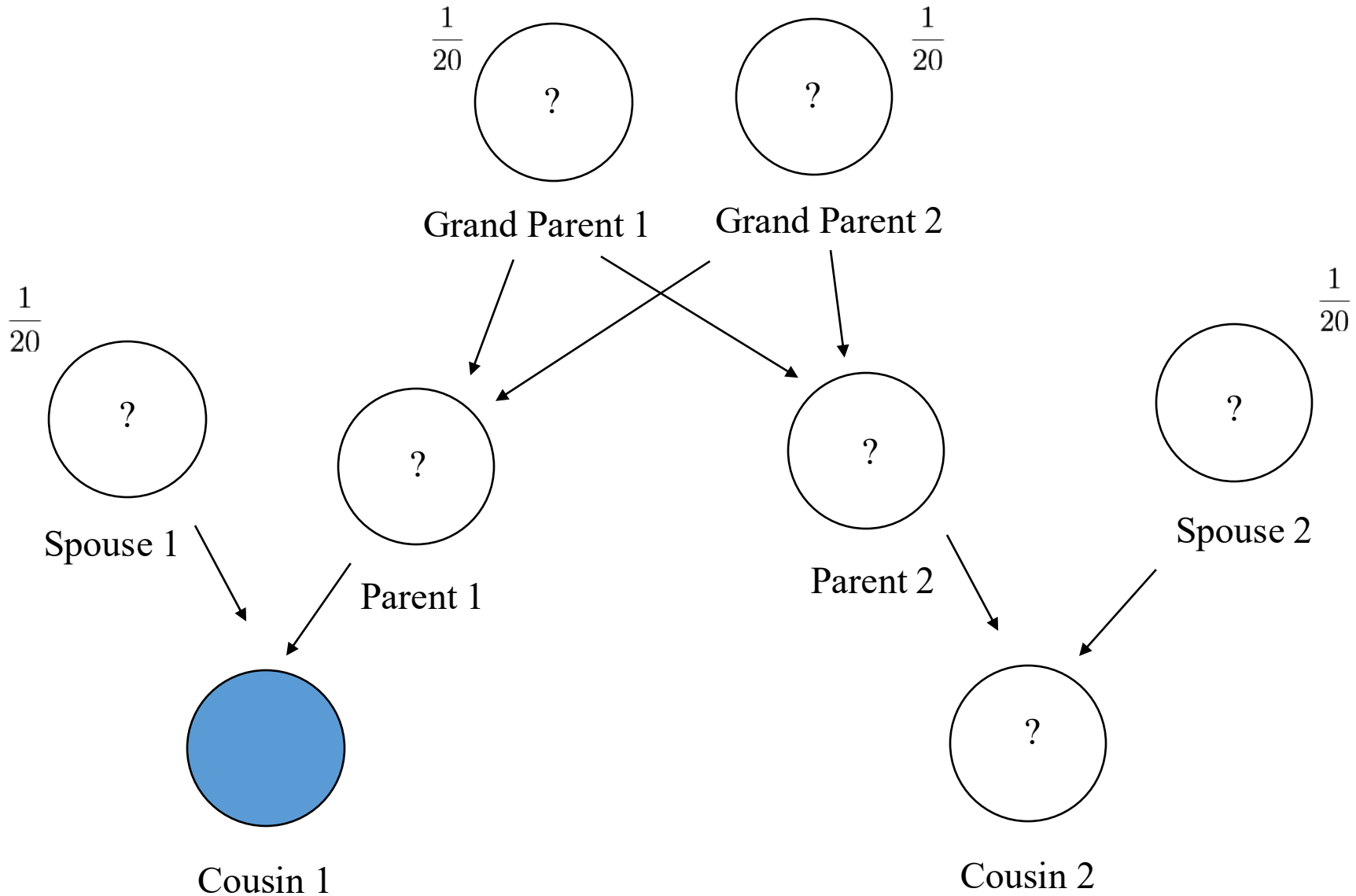
[Edit](#)

My Symptoms

[Edit](#)

**dizziness, one sided headache**

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, 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, 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:

Simulator results:  
C, C, B, A, A, A, C, C, C, A

**Totals:**  
A: 4  
B: 1  
C: 5

[Simulate](#)

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

$$P(X_1 = c_1, X_2 = c_2, \dots, X_m = c_m) =$$

$$p_1^{c_1} p_2^{c_2} \cdots p_m^{c_m}$$

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

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

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
- 0 threes
- 0 fives
- 1 two
- 2 fours
- 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

# of times  
a six appears

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

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

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  $n, p \dots$

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

$p$ : probability of success outcome on a single trial

A Multinomial RV has parameters  $n, p_1, p_2, \dots, p_m$

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

$p_i$ : probability of outcome  $i$  on a single trial

Where do we get  $p_i$  from?

# Most useful when probabilities are not equal

Lecture 11 - Probabilistic Mod x +

cs109psets.netlify.app/fall24/lecture11/multinomial\_dice

## Multinomial Dice

You have a funny shaped dice.

The probability of getting a 1 is 0.2

The probability of getting a 2 is 0.3

The probability of getting a 3 is 0.1

The probability of getting a 4 is 0.1

The probability of getting a 5 is 0.1

The probability of getting a 6 is 0.2

You roll the dice 6 times. What is the probability of getting exactly:

- two 2s
- two 4s
- two 6s

(consider the dice un-ordered)

Answer checker

Answer Editor Solution

Numeric Answer: Enter your answer Check Answer

Explanation:

Block LaTeX Inline LaTeX Python Image

Previous Question Next Question

68

# The Federalist Papers

# Intro to Natural Language Processing



Sam Smith

Fri, Aug 29, 12:25 PM



to Chris ▾

email



Dear Professor Piech. Can you help me? I have a large amount of money in my dorm room, but I can't get it out. If you sent me a \$500 coupa card I could distract my TA and pay you back double.

1

Let  $E$  be the entire email text. Let  $F$  be the event the email is Spam.

2

$$P(F|E) = \frac{P(E|F)P(F)^{60\%}}{P(E|F)P(F)^{60\%} + P(E|F^C)P(F^C)^{40\%}}$$

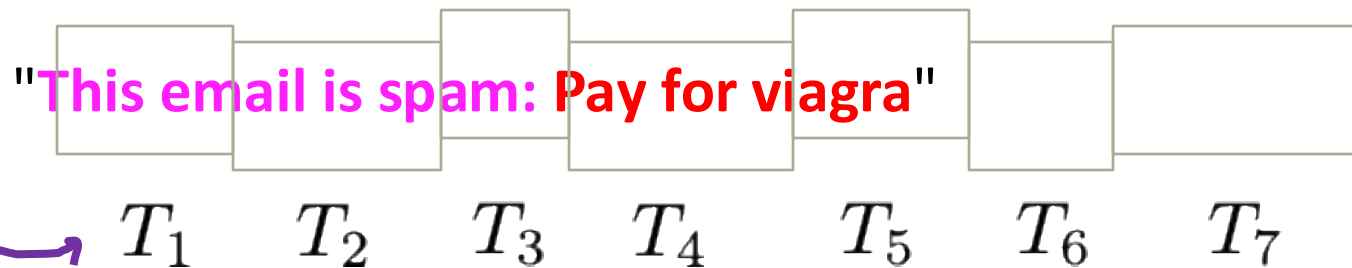
3

$$P(E|F) \approx \frac{\text{string\_pr}(f\text{"This email is spam: \{email\}"})}{\text{string\_pr}(f\text{"This email is spam: "})}$$

email  $P(E|F) \approx \frac{\text{string\_pr}(f\text{"This email is spam: \{email\}"})}{\text{string\_pr}(f\text{"This email is spam: "})}$

spam

The token at index 1 is "this"



$$P(E|F) \approx \frac{P(T_1, T_2, \dots, T_7)}{P(T_1, T_2, T_3, T_4)}$$

$$\approx P(T_5, T_6, T_7 | T_1, T_2, T_3, T_4)$$

$$\approx P(\text{"Pay for viagra"} | \text{"This email is spam: "})$$

We are going to contrast that to an  
older (but cheaper) method...

Before 2021...



# 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.”

$n = 18$

$$P \left( \begin{array}{l} \text{Viagra} = 3 \\ \text{Free} = 2 \\ \text{Risk} = 1 \\ \text{Credit-card: } 2 \\ \dots \\ \text{For} = 2 \end{array} \middle| \text{spam} \right)$$

Probability of seeing this document | spam

It's a Multinomial!

$$\frac{n!}{3! 2! \dots 2!} p_{\text{viagra}}^3 p_{\text{free}}^2 \dots p_{\text{for}}^2$$

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

```
1 To the People of the State of New York:  
2  
3 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

```
1 The Utility of the Union in Respect to Commercial  
Relations and a Navy  
2 Hamilton for the Independent Journal.  
3  
4 To the People of the State of New York:  
5 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.  
6  
7 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.  
8  
9 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

```
1 To the People of the State of New York:  
2 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
```

# 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???

The diagram illustrates the Bayesian formula for the probability of Hamilton writing the document given the document itself. The formula is  $P(H|D) = \frac{P(D|H)P(H)}{P(D)}$ . Blue arrows point from descriptive text to the corresponding terms in the formula: 'Prob Document given Hamilton' points to  $P(D|H)$ , 'Prior belief it was Hamilton' points to  $P(H)$ , 'Prob of the document???' points to  $P(D)$ , and 'Prob Hamilton given Document' points to  $P(H|D)$ .

# 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  $i$

Number of times word  $i$  is in the doc

Prior belief it was Hamilton

Prob of the document???

Prob Hamilton given Document

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

# Who wrote Federalist Paper 53?

Prob that Hamilton wrote it

$$\begin{aligned}P(H|D) &= \frac{P(D|H)P(H)}{P(D)} \\ &= \frac{P(H) \cdot \binom{n}{c_1 \dots c_m} \cdot \prod_i h_i^{c_i}}{P(D)}\end{aligned}$$

Prob that Madison wrote it

$$\begin{aligned}P(M|D) &= \frac{P(D|M)P(M)}{P(D)} \\ &= \frac{P(M) \cdot \binom{n}{c_1 \dots c_m} \cdot \prod_i m_i^{c_i}}{P(D)}\end{aligned}$$

$$\begin{aligned}\frac{P(H|D)}{P(M|D)} &= \frac{P(H) \cdot \binom{n}{c_1 \dots c_k} \cdot \prod_i h_i^{c_i}}{P(D)} \bigg/ \frac{P(M) \cdot \binom{n}{c_1 \dots c_k} \cdot \prod_i m_i^{c_i}}{P(D)} \\ &= \frac{\prod_i m_i^{c_i}}{\prod_i h_i^{c_i}}\end{aligned}$$

# To the code

---



What happened?

All our probabilities are zero...

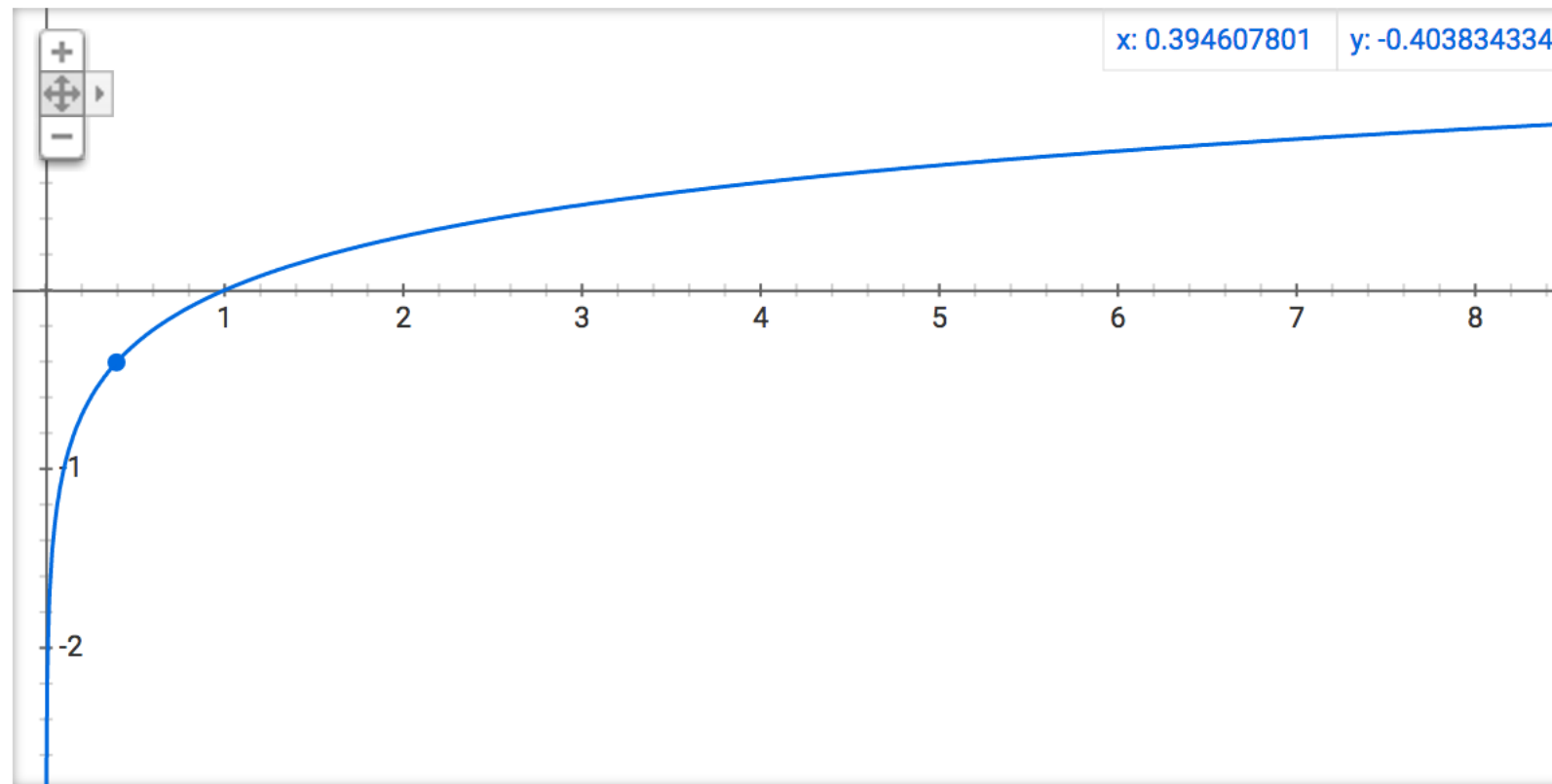


# Log Review

$$e^y = x$$

$$\log(x) = y$$

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!

---

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

---

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

\* 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

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

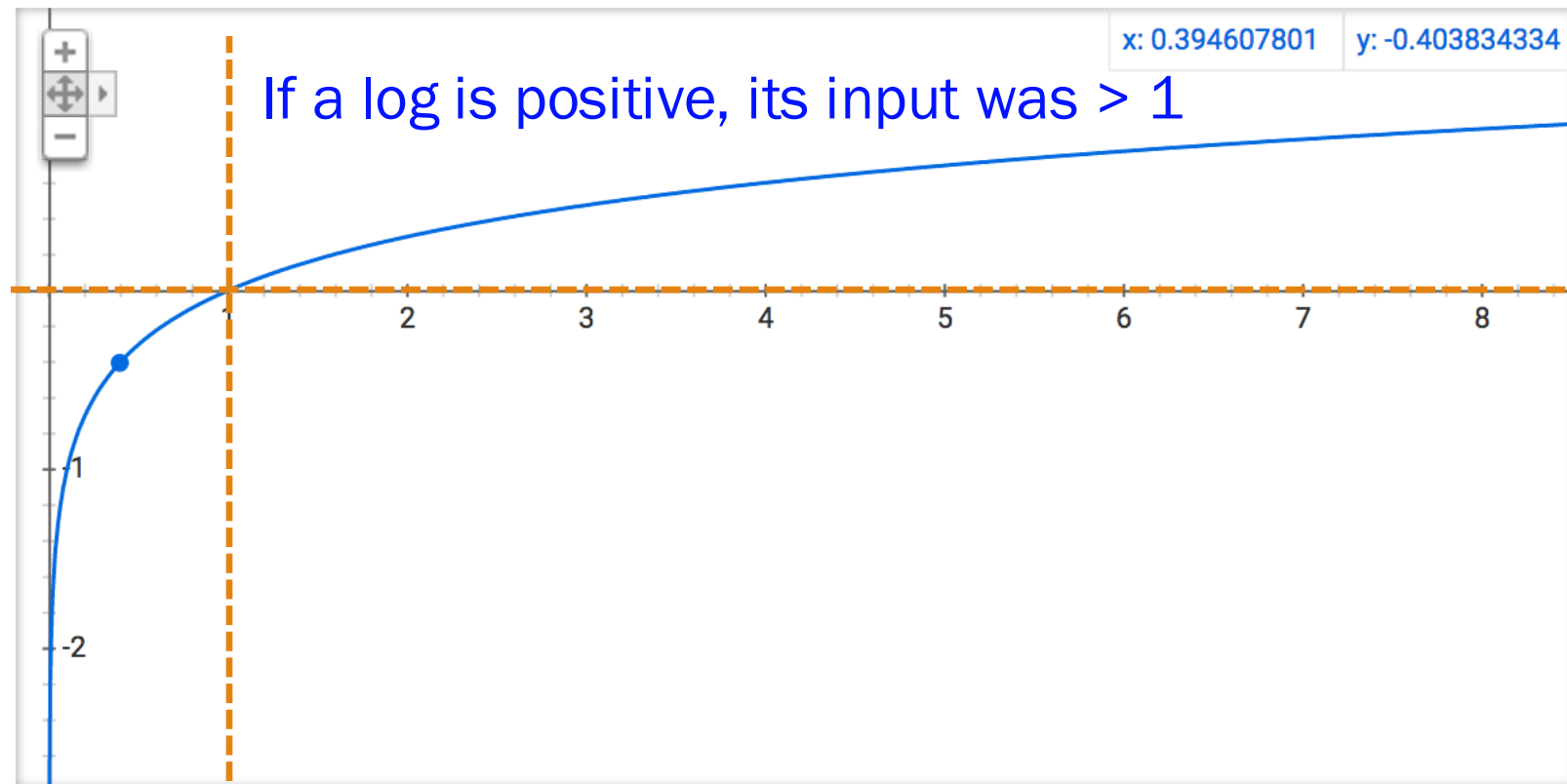
$h_i = \text{Prob Hamilton writes word } i$

$m_i = \text{Prob Maddison writes word } i$

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

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

Graph for  $\log(x)$



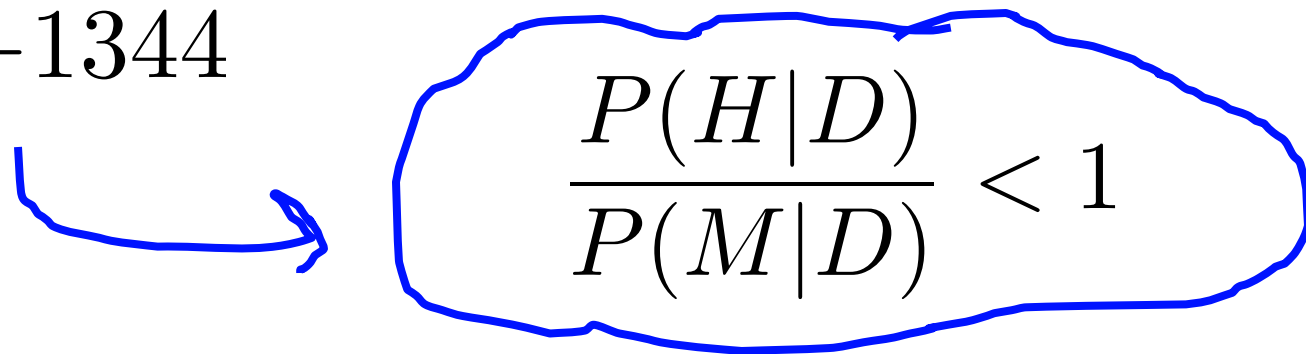
If a log is negative, its input was between 0 and 1

[More info](#)

# 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



Madison wrote it!

Back to Spam Classification

# Spam Dataset. Lets see Who Wins!

---

HAM The bus leaves at &lt;#&gt;

SPAM FreeMsg Hey U, i just got 1 of these video/pic fones, reply WILD to this txt & ill send U my pics,

HAM We don call like &lt;#&gt; times oh. No give us hypertension oh.

HAM So dont use hook up any how

HAM And also I've sorta blown him off a couple times recently so id rather not text him out of the blue HAM Tiwary to rcb.battle between bang and kochi.

HAM One of best dialogue in cute reltnship..!! \Wen i Die

SPAM You have WON a guaranteed å£1000 cash or a å£2000 prize. To claim yr prize call our customer service

HAM Was just about to ask. Will keep this one. Maybe that's why you didn't get all the messages we

HAM Wat makes u thk i'll fall down. But actually i thk i'm quite prone 2 falls. Lucky my dad at home i

HAM Why nothing. Ok anyway give me treat

HAM Great! I hope you like your man well endowed. I am &lt;#&gt; inches...

HAM Mmmmm ... I loved waking to your words this morning ! I miss you too, my Love. I hope your day goes SPAM Ever thought about living a good life with a perfect partner? Just txt back NAME and AGE to join

HAM Eat at old airport road... But now 630 oredi... Got a lot of pple...

HAM Sounds better than my evening im just doing my costume. Im not sure what time i finish tomorrow but

HAM Yes! I am a one woman man! Please tell me your likes and dislikes in bed...

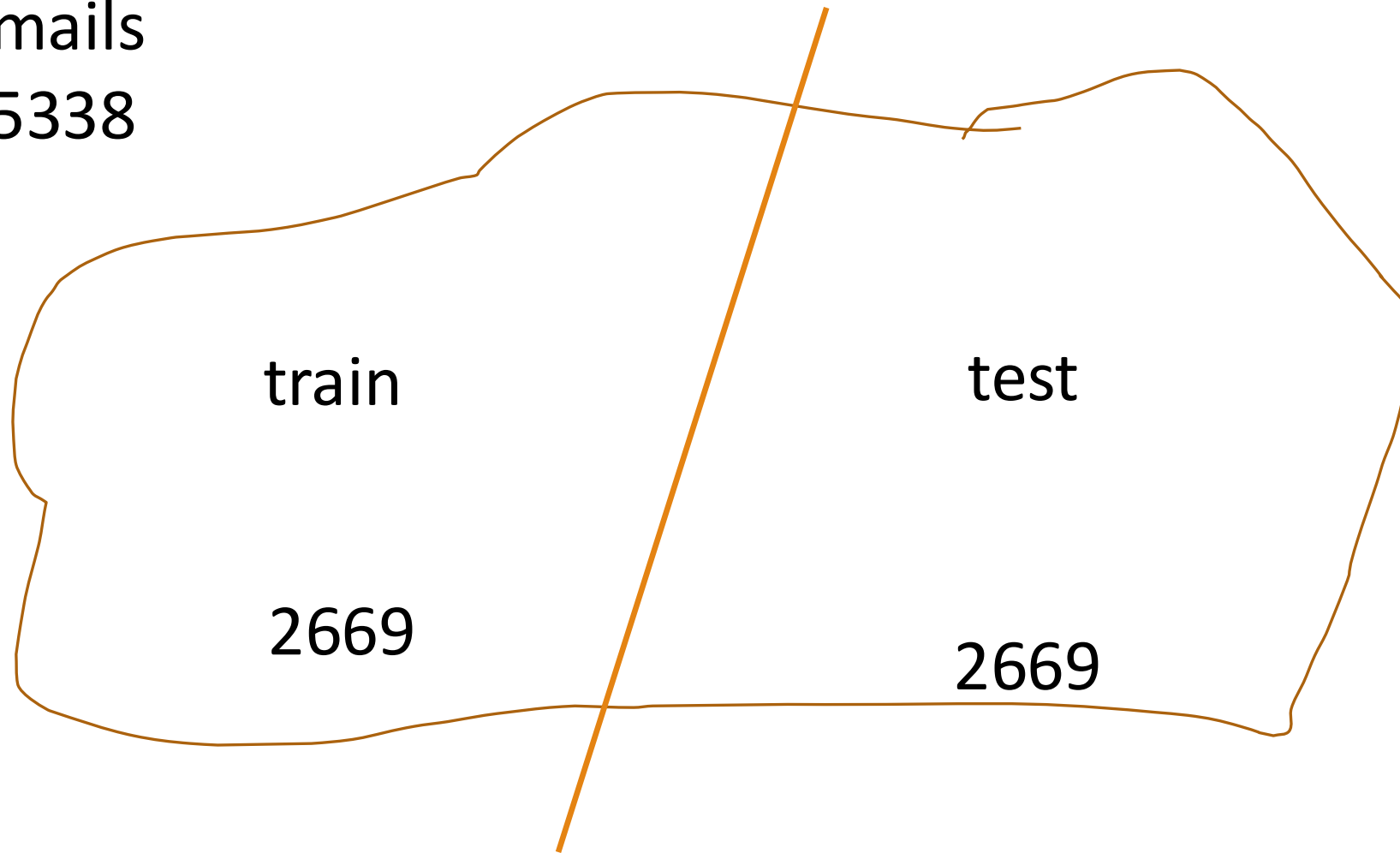
HAM Oh man. That happend to me in tron. Maybe ill dl it in 3d when its out

HAM But you were together so you should be thinkin about him

# Spam Dataset. Lets see Who Wins!

---

All emails  
5338

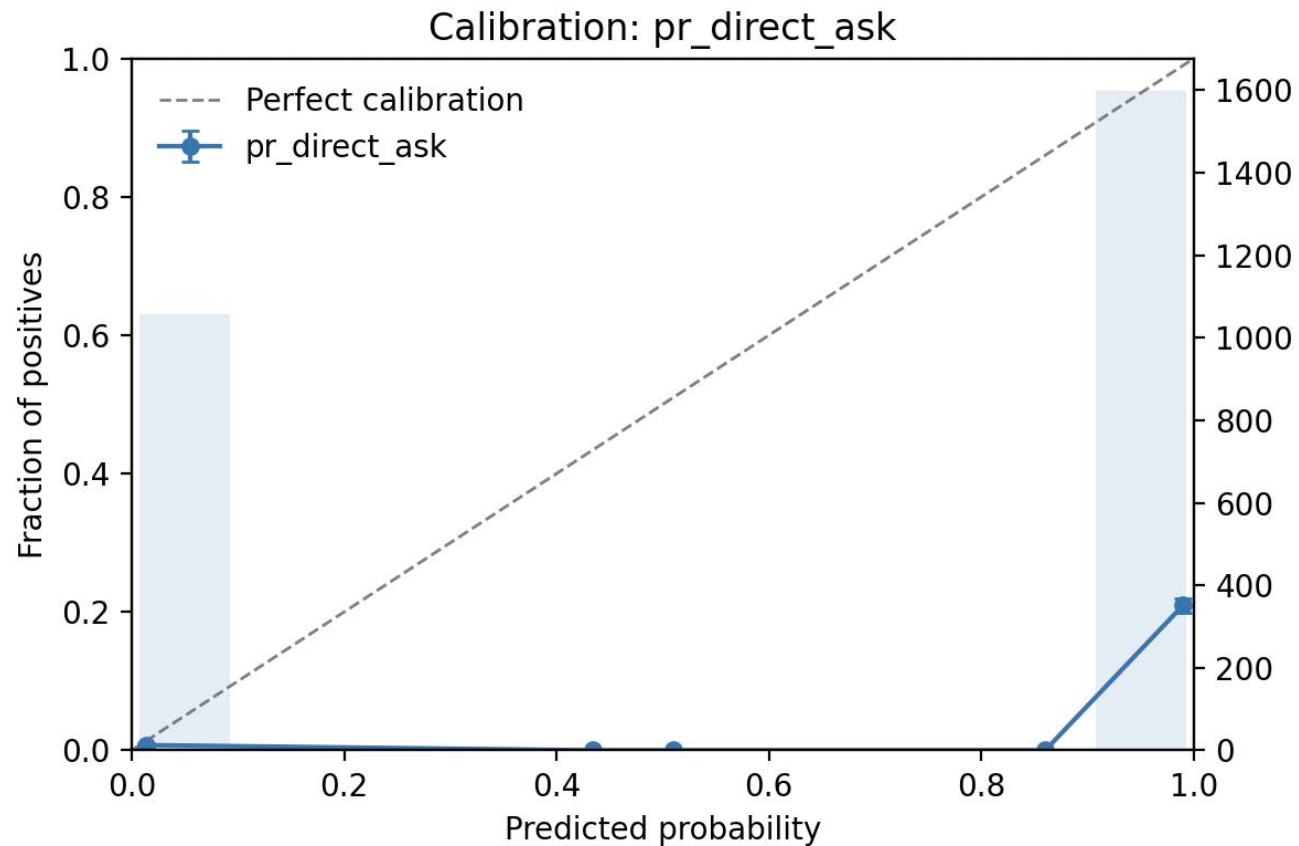


How did they do?

# Directly Ask an LLM

Provide your best guess about the probability that this sms is spam.

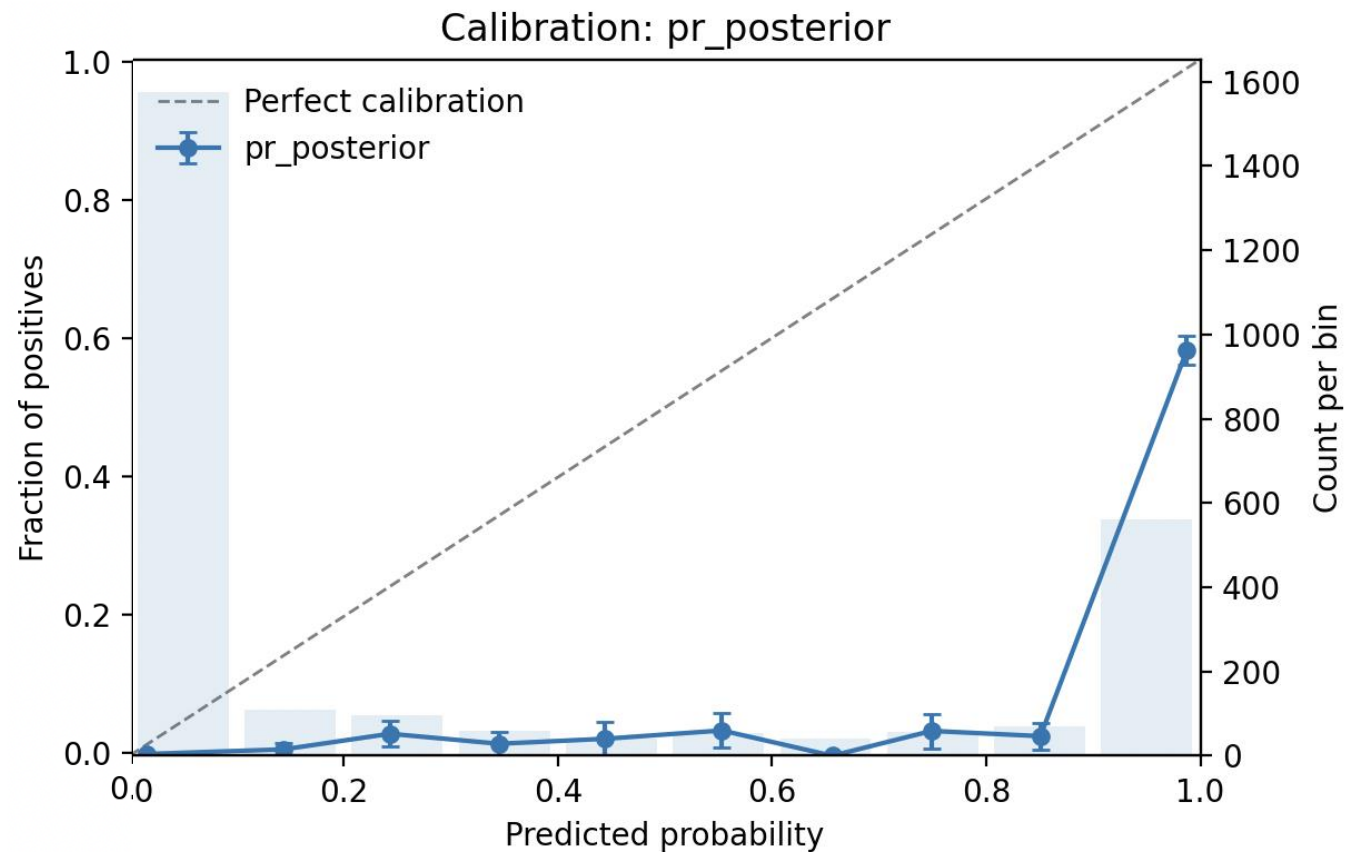
Method	Accuracy Test	Calibration Error
Direct Ask	52%	0.472



# Directly Ask an LLM

```
get_str_prediction(f"Is this SMS spam or non-spam? '{sms_text}': ", "spam")
```

Method	Accuracy Test	Calibration Error
Direct Ask	52%	0.472
Posterior	83%	0.178

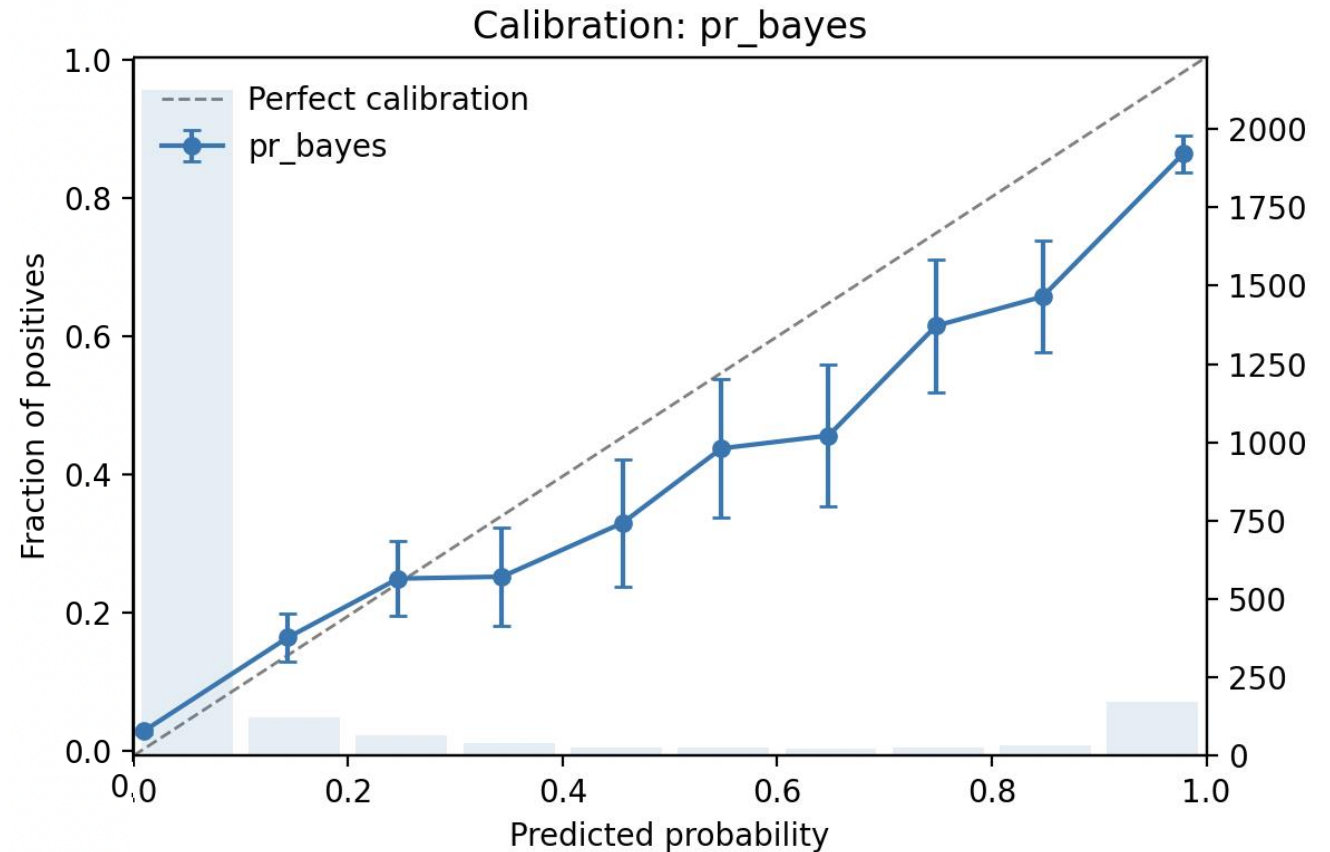


# Use Bayes Theorem

```
pr_email_given_spam = get_str_prediction("This is from a spam SMS:", sms_text)
```

```
pr_email_given_ham = get_str_prediction("This is from a non-spam SMS:", sms_text)
```

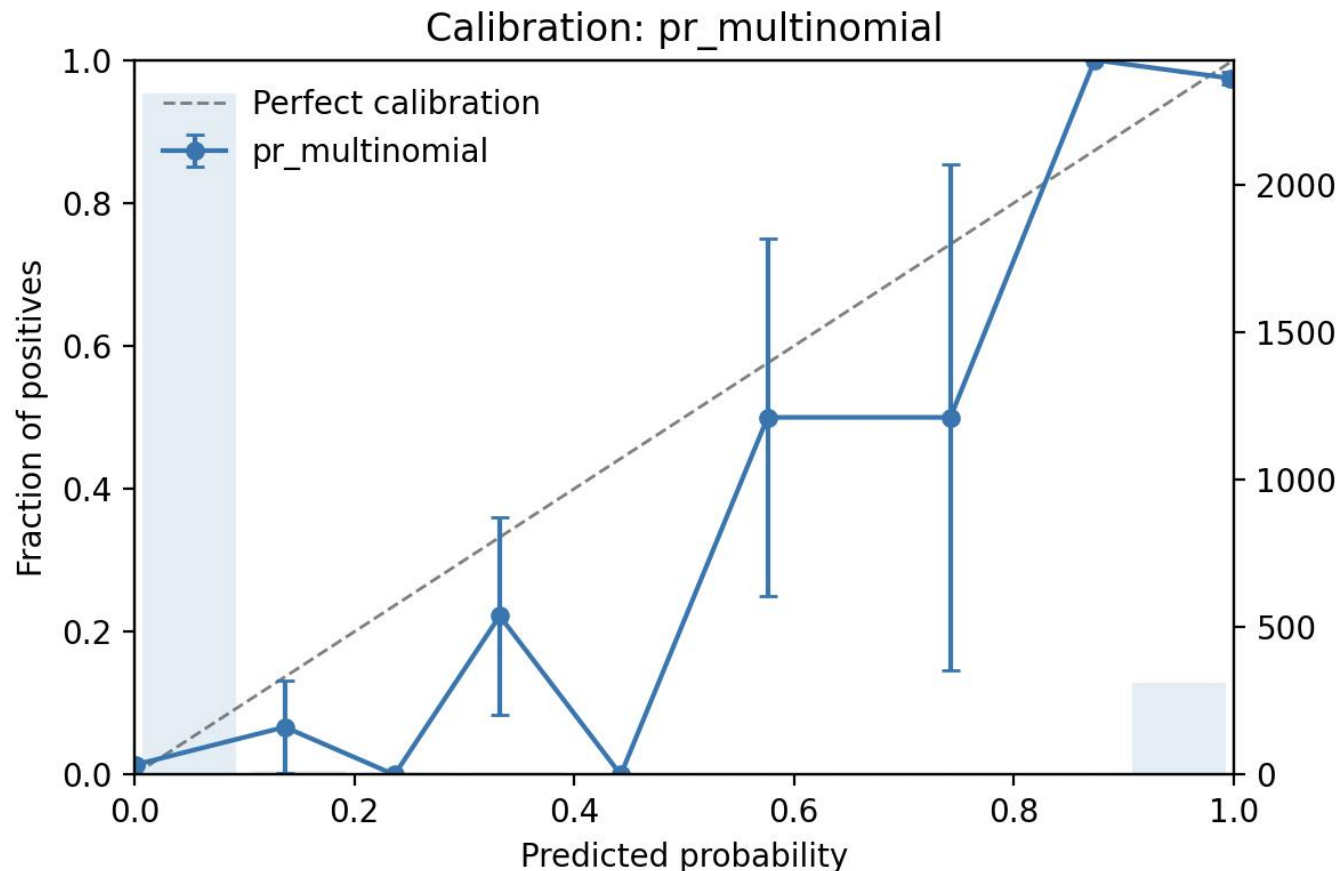
Method	Accuracy Test	Calibration Error
Direct Ask	52%	0.472
Posterior	83%	0.178
Bayes	92%	0.038



# Use Bayes Theorem



Method	Accuracy Test	Calibration Error
Direct Ask	52%	0.472
Posterior	83%	0.178
Bayes	92%	0.038
Multinomial	98%	0.016



Have a Wonderful Weekend!