The distribution they should have told you about in high school

THE UNIVERSITY OF CHICAGO

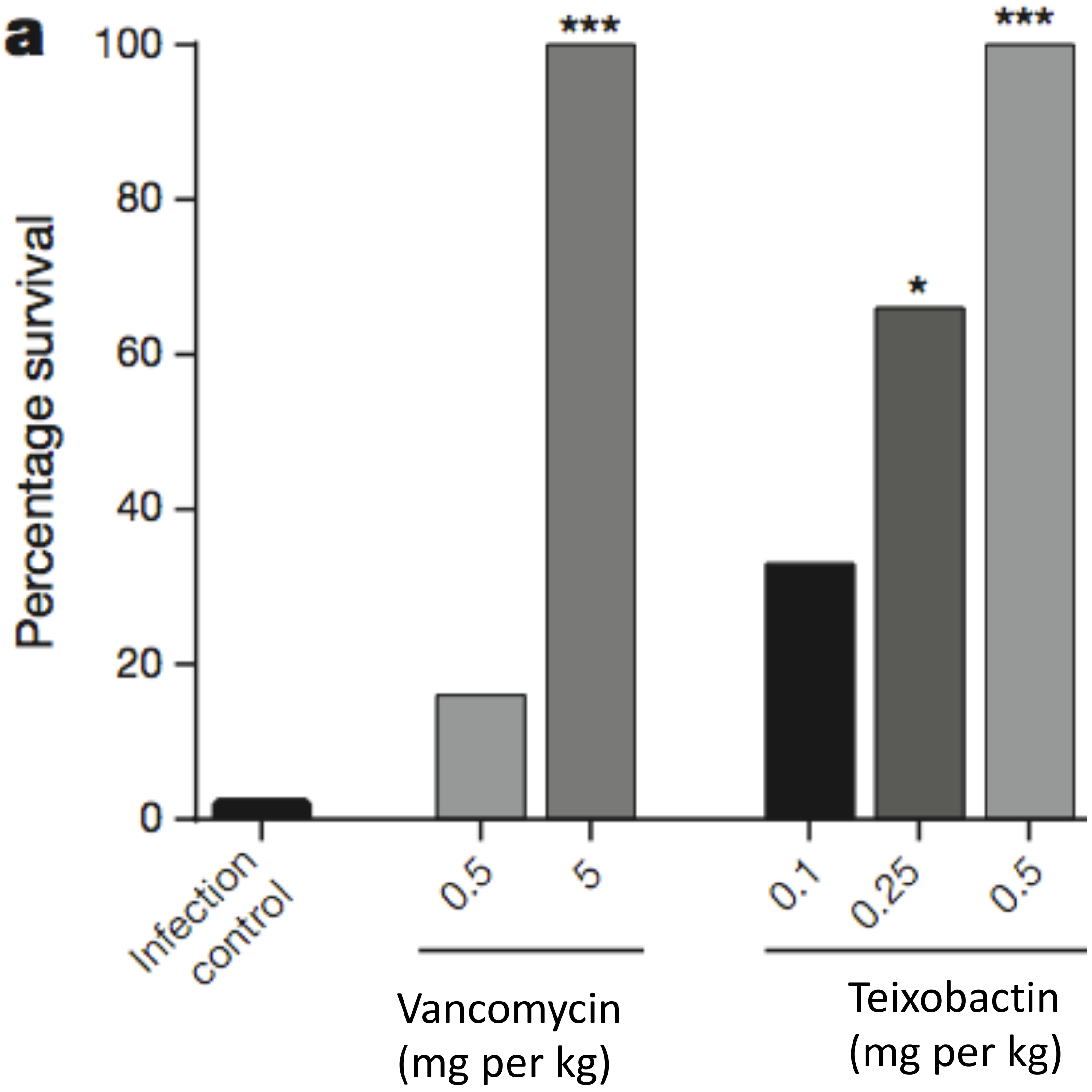
A new antibiotic

Ling et al. Nature 517:455 (2015)

The paper announced the discovery of a new polypeptide antibiotic (tiexobactin).

Mice were given a lethal infection, and then treated with two antibiotics at different doses.

Something about the graph is funny.



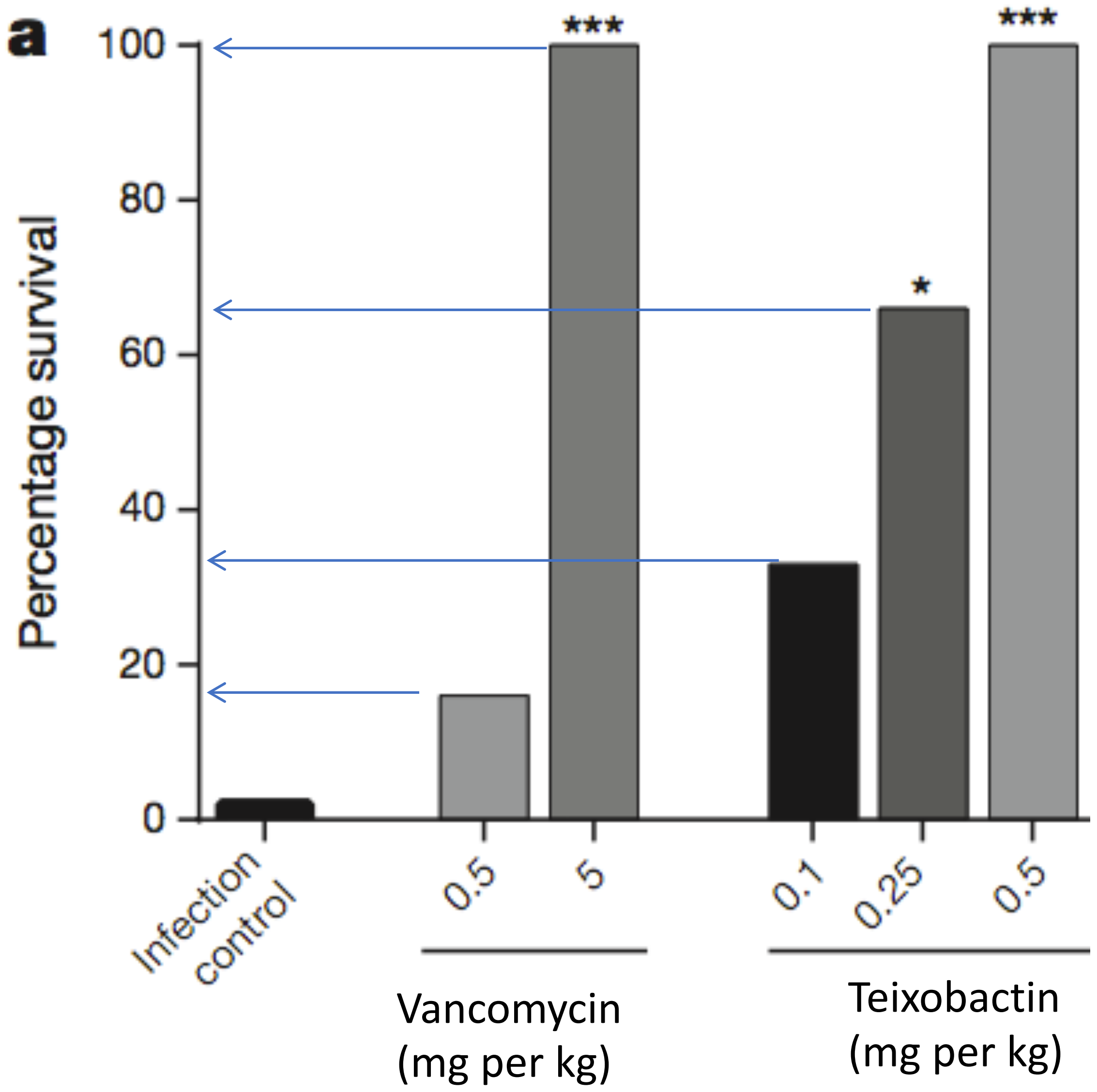
A new antibiotic

Ling et al. Nature 517:455 (2015)

The paper announced the discovery of a new polypeptide antibiotic (tiexobactin).

Mice were given a lethal infection, and then treated with two antibiotics at different doses.

Something about the graph is funny.



A new antibiotic

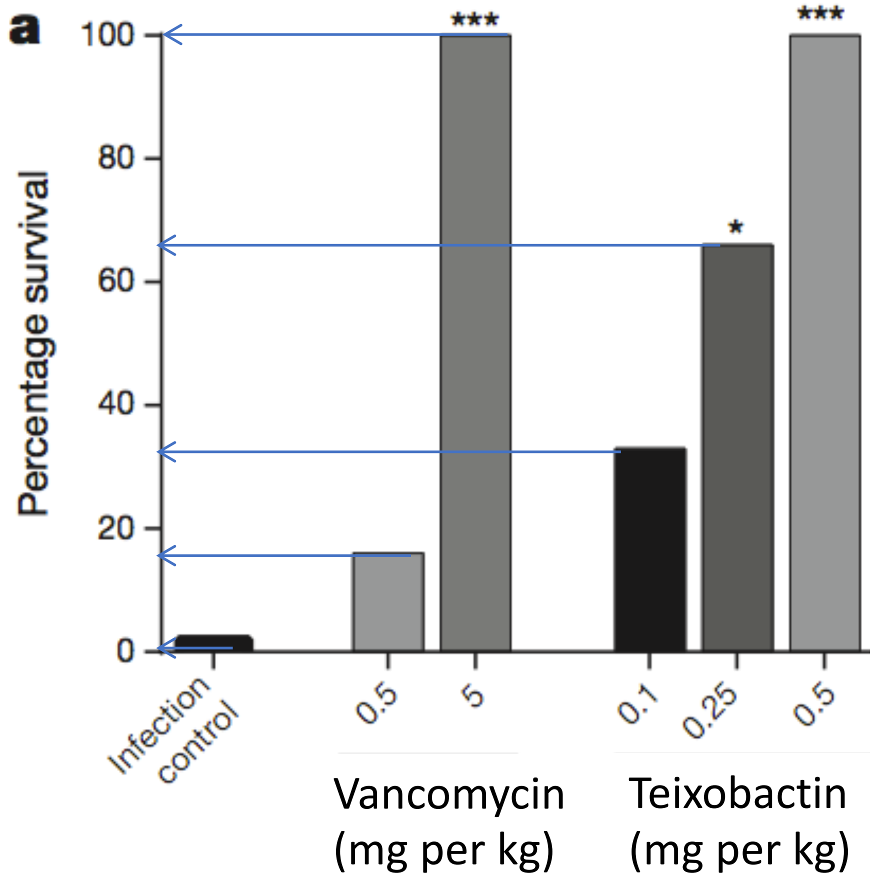
Ling et al. Nature 517:455 (2015)

The paper announced the discovery of a new polypeptide antibiotic (tiexobactin).

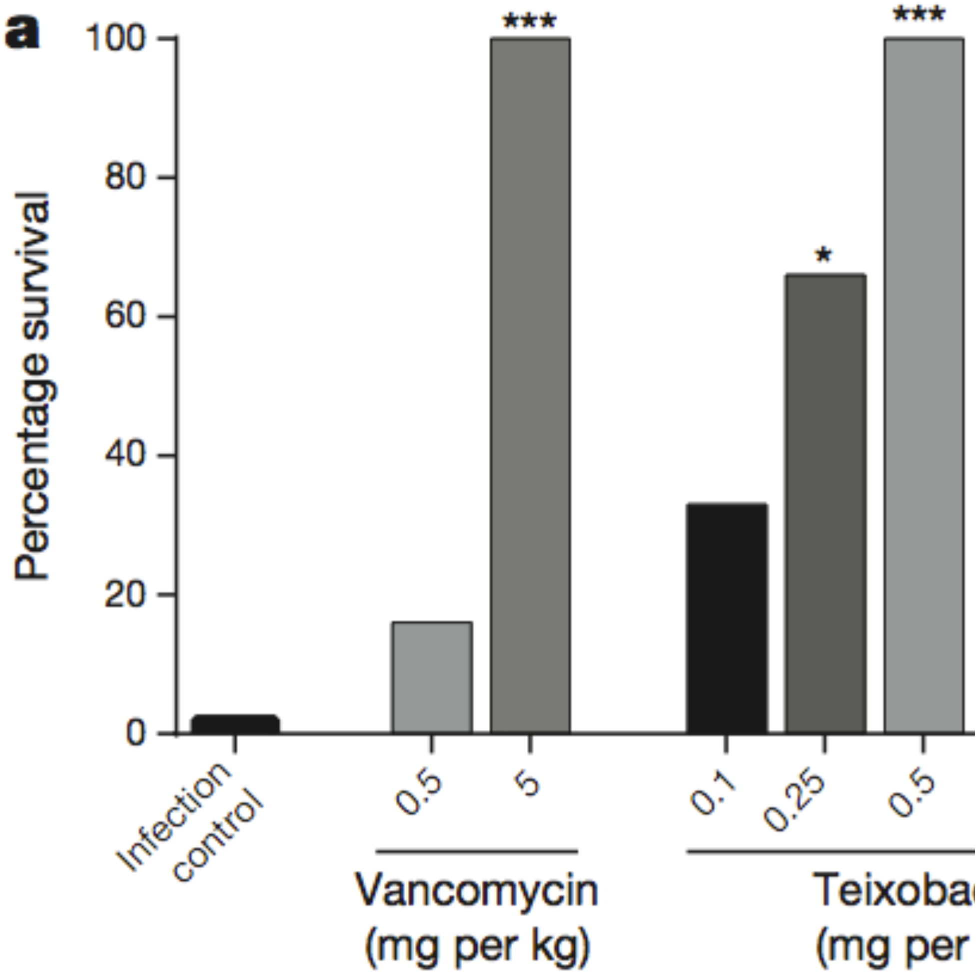
Mice were given a lethal infection, and then treated with two antibiotics at different doses.

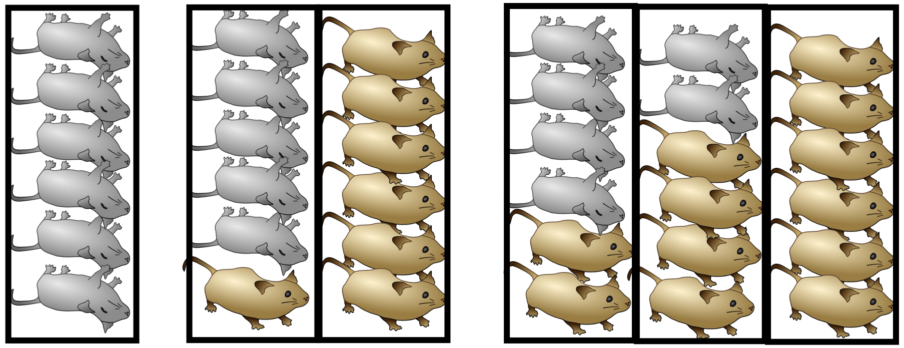
Something about the graph is funny.

All of the bars are multiples of 1/6.
Read the fine print..



This graph summarizes the fates of thirty-six mice, six cages of six mice each.





Binomial likelihood

$$P_{\mathrm{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \;{}_{n}C_{\mathrm{r}}\, p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\mathrm{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \;\binom{n}{k} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\mathrm{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \;\frac{n!}{k!(n-k!)} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

This is the number of outcomes
(these are rows from Pascal's triangle)

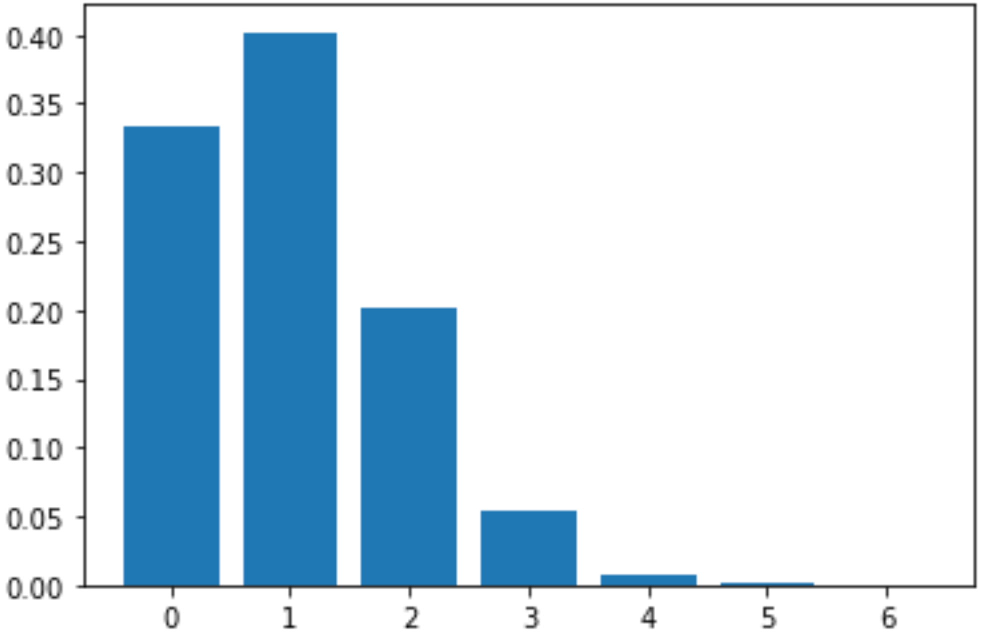
This is the probability of each outcome

Binomial likelihood

$$P_{\mathrm{Binomial}}(k;\, n, p) = {}_nC_r\, p^{k}(1-p)^{n-k}$$

$$P_{\mathrm{Binomial}}(k;\, n, p) = \binom{n}{k} p^{k}(1-p)^{n-k}$$

$$P_{\mathrm{Binomial}}(k;\, n, p) = \frac{n!}{k!(n-k!)} p^{k}(1-p)^{n-k}$$



We normally think about this as a function over k, which defines all the (n +1) outcomes in the sample space $k = \{0,1,2,3,4\ldots n\}$.

The sum of $P_{\mathrm{binomial}}(k;\, n, p)$ over k at constant n and p is always 1.

Binomial likelihood

$$P_{\text{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \;{}_{n}C_{\mathrm{r}}\;p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \;\binom{n}{k}p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \;\frac{n!}{k!(n-k!)}p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

Binomial likelihood

$$P_{\text{Binomial}}\,(\mathrm{k};\, \mathrm{n},\, \mathrm{p}) = \; {}_{n}C_{\mathrm{r}}\, p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\, \mathrm{n},\, \mathrm{p}) = \; \binom{n}{k} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\, \mathrm{n},\, \mathrm{p}) = \; \frac{n!}{k!(n-k!)} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

This is the number of outcomes

This is the probability of each outcome

Binomial likelihood

$$P_{\text{Binomial}}\,(p;\, n,\, k) = \quad f(n,k)\;\; p^{k}(1-p)^{n-k}$$

This is the number of outcomes

This is the probability of each outcome

Binomial log - likelihood

$$\log P_{\text{Binomial}}(p;\, n, k) = \log(\, f(n,k)\,) + k\log(p) + (n\text{-}k)\log\,(1-p)$$

Term that depends on n, k but independent of p

Terms that depend on p

Binomial log - likelihood

$$\log P_{\mathrm{Binomial}}(p; n, k) = \log(\, f(n,k)\,) \;+\; k \log(p) \;+ (n\text{-}k) \log\,(1-p)$$

Term that depends on n, k but independent of p

Terms that depend on p

Sometimes we use log-likelihood because it turns multiplication into addition and this makes some inferences easier.

Other times we use log-likelihood for numerical reasons; we may want to take the ratio of two likelihoods that are smaller than 10^{-16}

The Univariate Beta distribution

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1} (1 - x)^{b-1}$$

a and b are number
of successes and number
of failures (almost)
k, (n-k)

This is a number

Two terms with powers of
x and *1-x*
x is probability of success

The Univariate Beta distribution

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1}$$

The late David MacKay pointed out that probability distributions are often named for their normalization constants.

“Normalization constant” – jargon for “the number you have to multiply the math by to make the probability sum to 1”

The Univariate Beta distribution

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1}$$

You can think of this as the binomial probability reorganized to add up to 1 when integrated over x instead of summed over k.

The Univariate Beta distribution

$$f(x;a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

Expectation value

$$\mu = \frac{a}{a+b}$$

Standard deviation

$$\sigma = \frac{ab}{(a+b)^2(a+b+1)}$$

Mode

$\frac{a-1}{a+b-2}$ for a, b > 1

any value in [0,1] for α, β = 1

{0, 1} (bimodal) for α, β < 1

0 for $\alpha \leq 1$, $\beta > 1$

1 for $\alpha > 1$, $\beta \leq 1$

Bayesian inference

Event A: unknown state of nature

Event C: experiment

$$P(A \mid C) = P(C \mid A)\, P(A) \,/\, P(C)$$

Life's persistent question "What is death rate?"

posterior

Term that depends on experiment
in this case binomial

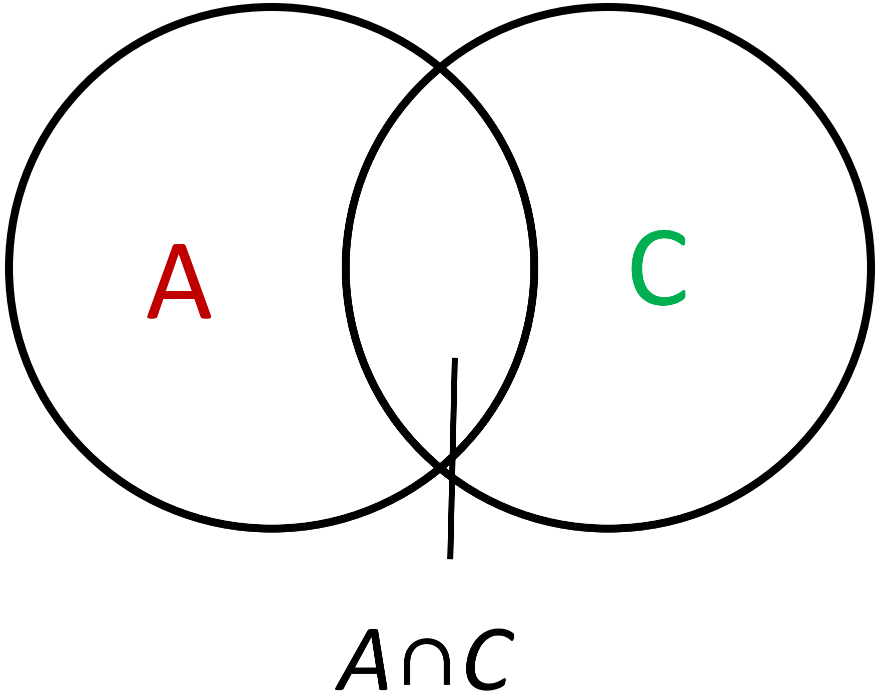
likelihood

Term with knowledge and bias about A

prior

Probability for event that is no longer uncertain

evidence



posterior density = prior density * likelihood (function of the data)

Beta is a conjugate prior for the binomial distribution:
Beta priors * binomial likelihoods = Beta posteriors

Beta (α, α) prior
x Binomial likelihood with m successes and n failures
= Beta$(\alpha + m, \alpha + n)$

As long as we write a paragraph justifying our choice of prior, we can easily get exact confidence intervals for parameters known by binomial sampling.

The Univariate Beta distribution

$$f(x;a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

$$p(x;0/3) = 4(1-x)^3$$
$$p(x;1/3) = 12x(1-x)^2$$
$$p(x;2/3) = 12x^2(1-x)$$
$$p(x;3/3) = 4x^3$$

The posterior distributions for 0, 1, 2, and 3 successes out of 3 are “polynomial functions” known as middle- and high-school torture instruments.

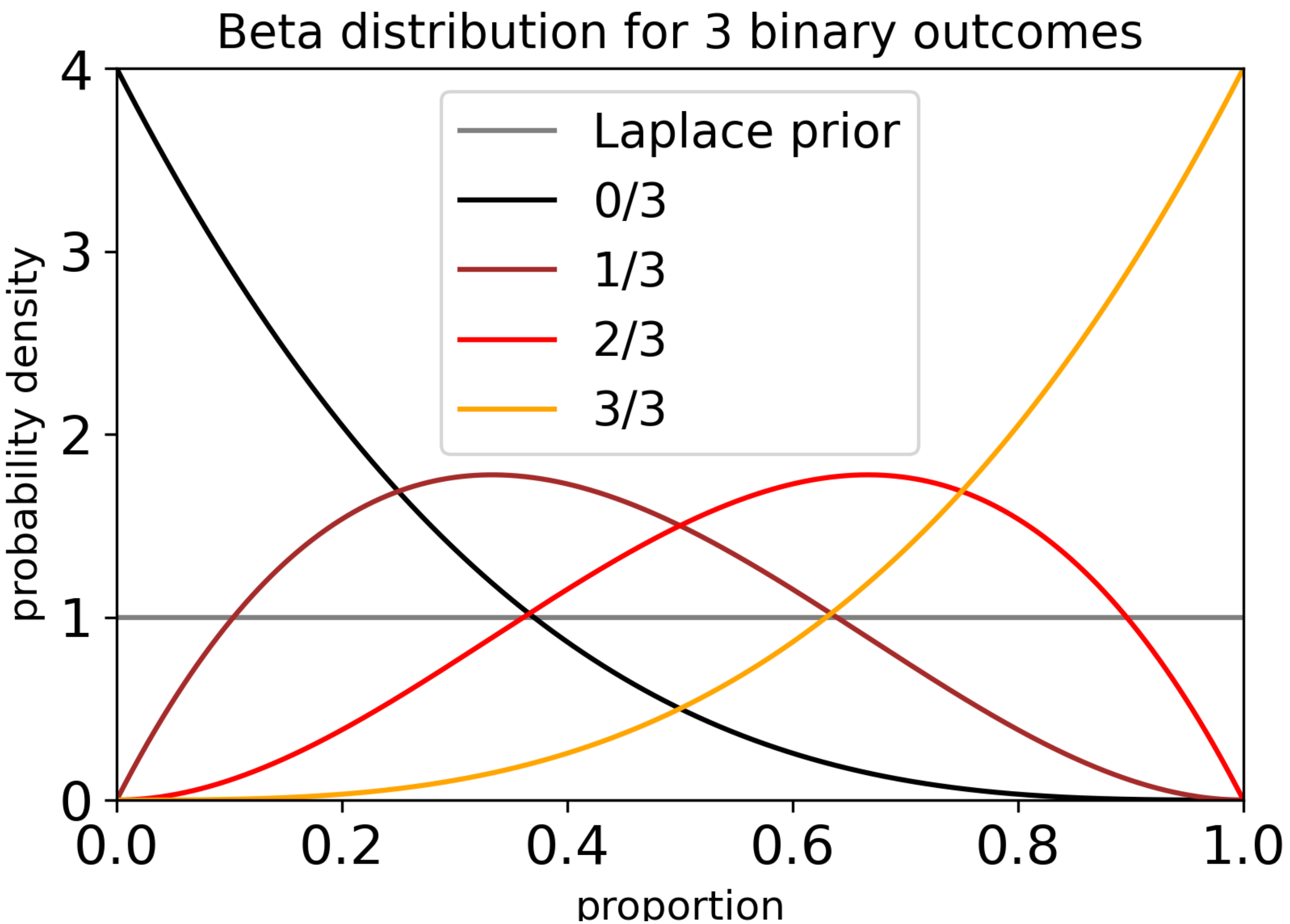
Beta distributions for all possible outcomes of an n=3 binomial experiment

$$p(x; 0/3) = 4(1 - x)^3$$

$$p(x; 1/3) = 12x(1 - x)^2$$

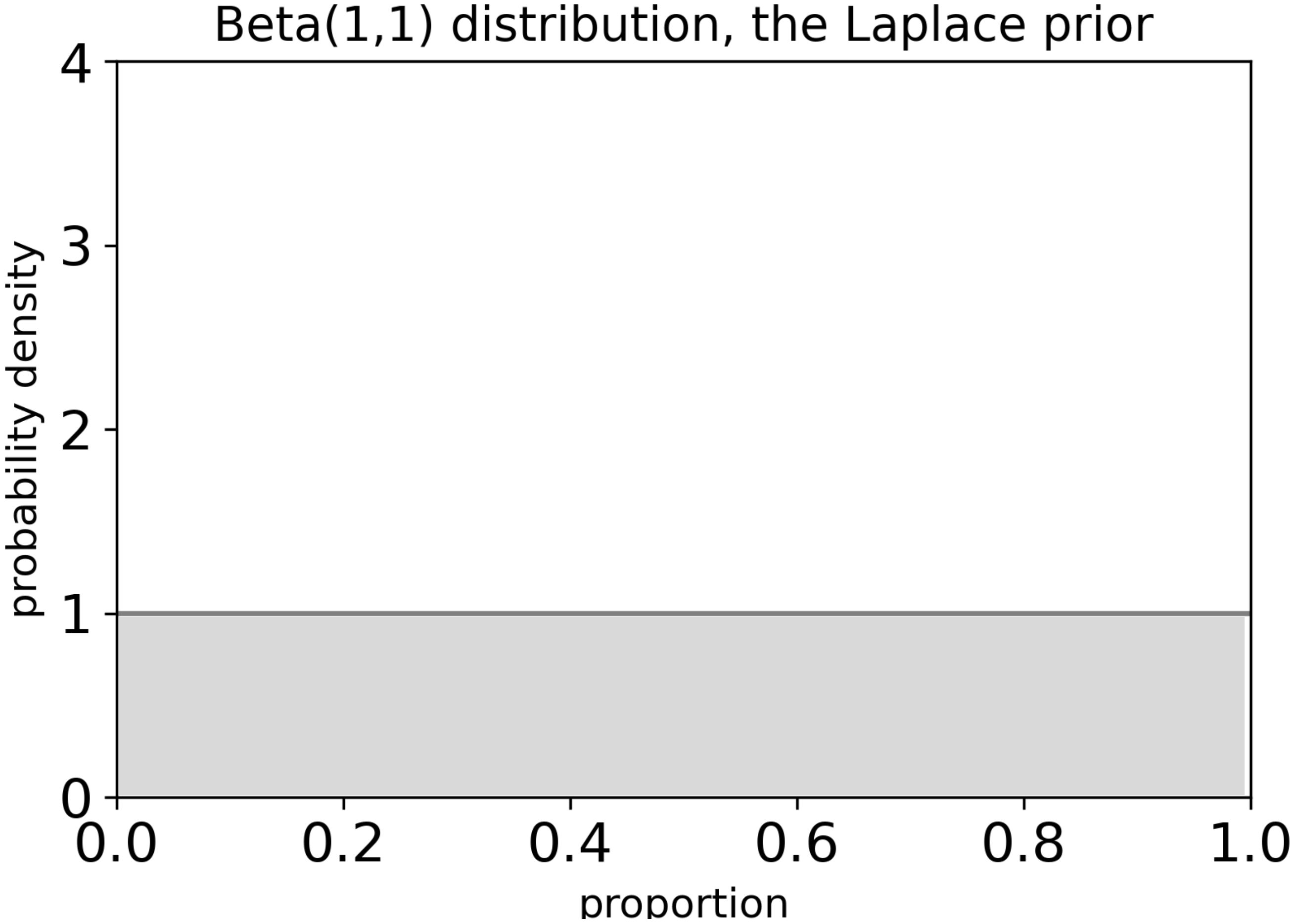
$$p(x; 2/3) = 12x^2(1 - x)$$

$$p(x; 3/3) = 4x^3$$



What does ignorance look like?

Before I do the experiment,
I don't know much about
the proportion.
But we're grownups, so I
have to put numbers on
my ignorance.



The oldest choice is the "flat" prior

$$\text{Beta}(a=1, b=1) = x^0 (1-x)^0 = 1$$

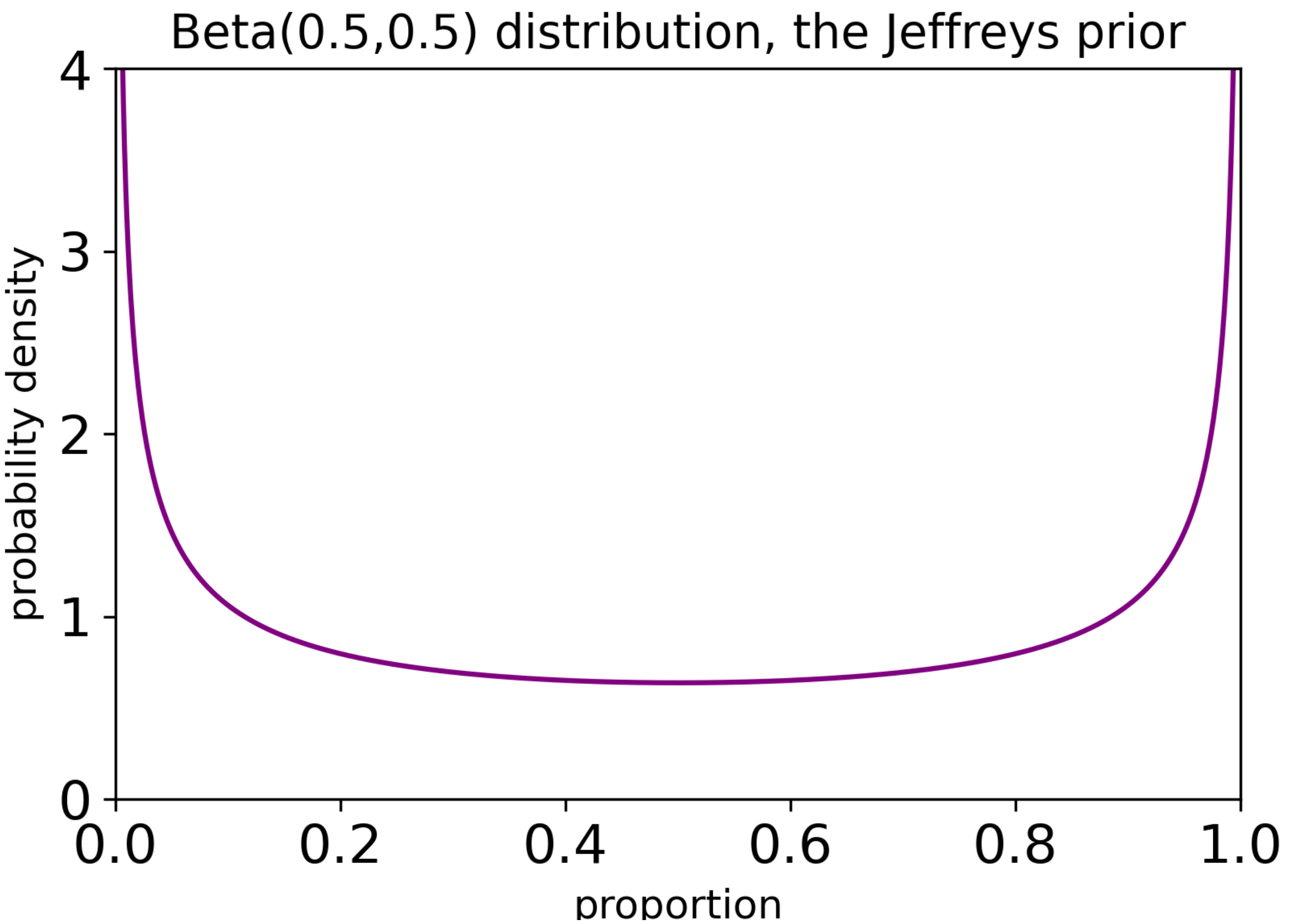
This prior introduces "bias" — it has an expectation value of 0.5 !!!

What does ignorance look like?

Perhaps there is reason to prefer the extremes?

Jeffreys prior does that.

Note the symmetry: if a = b,
your prior does not prefer successes or failures.

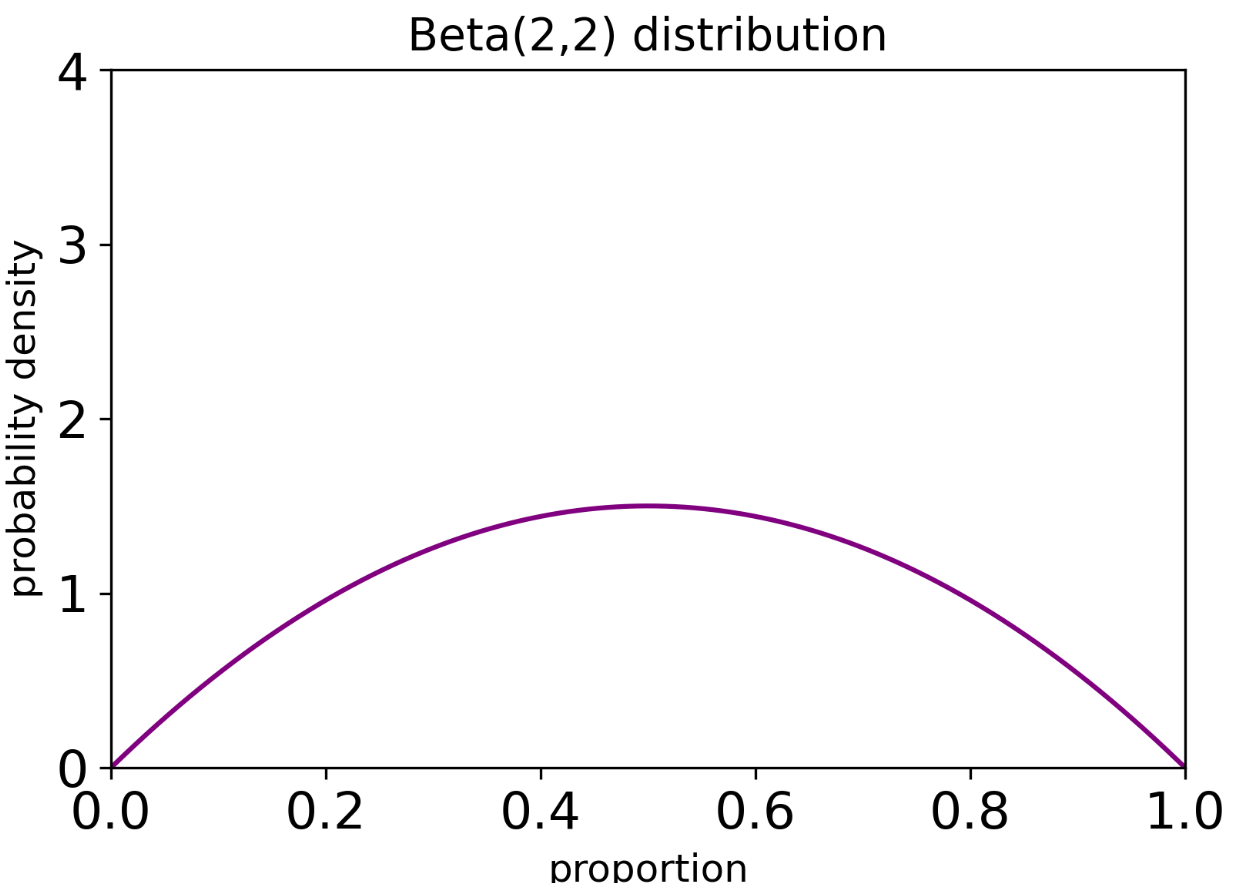


Beta(a=0.5, b=0.5) = c $x^{-0.5}$ $(1\text{-}x)^{-0.5} = \dfrac{c}{\sqrt{p(1-p)}}$

What does ignorance look like?

Note the symmetry: if a = b,
the prior does not prefer successes
or failures.

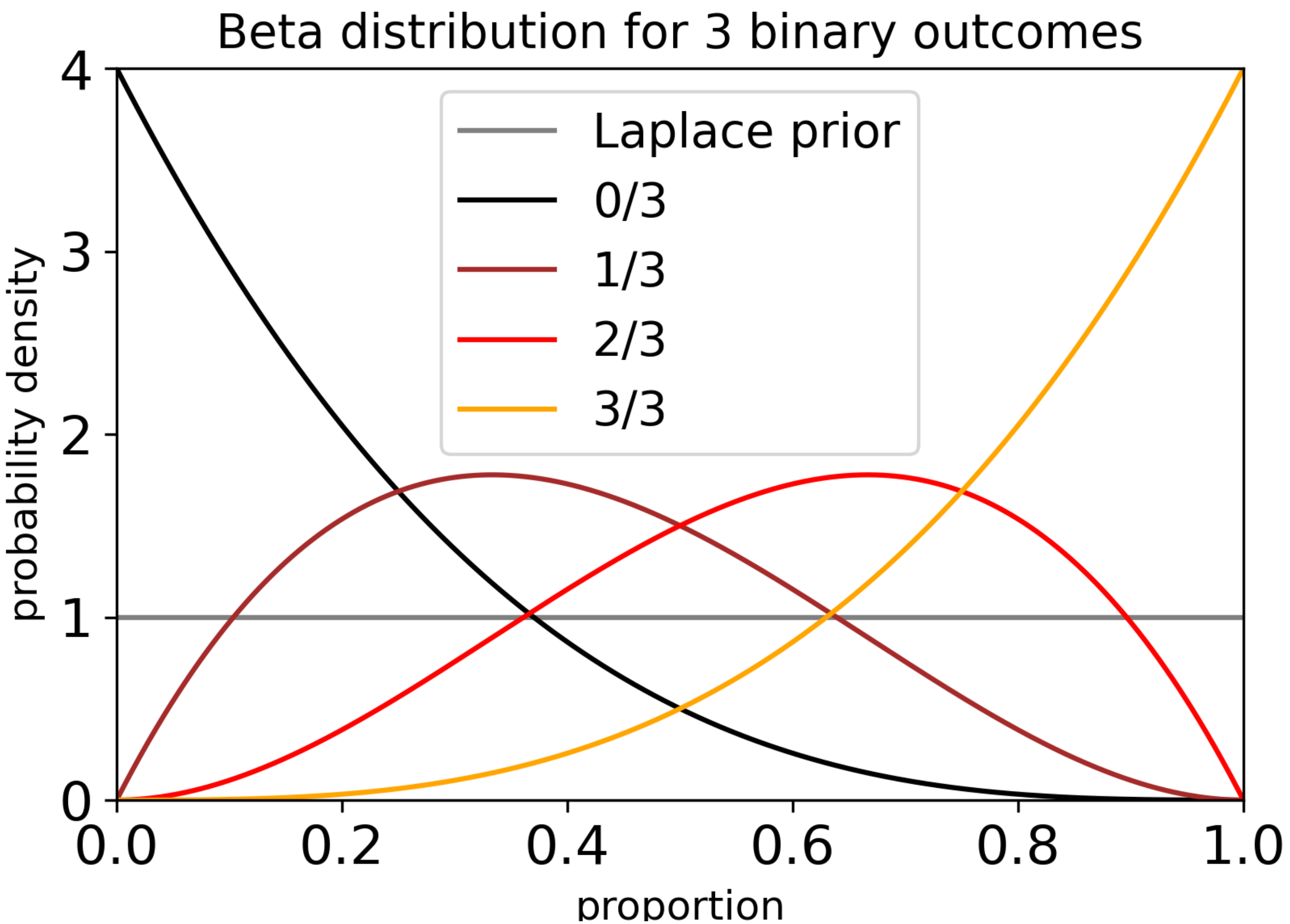
a = b = 2 is reasonable if we are certain
that both successes and failures
are possible (p cannot be exactly 0 or
exactly 1)

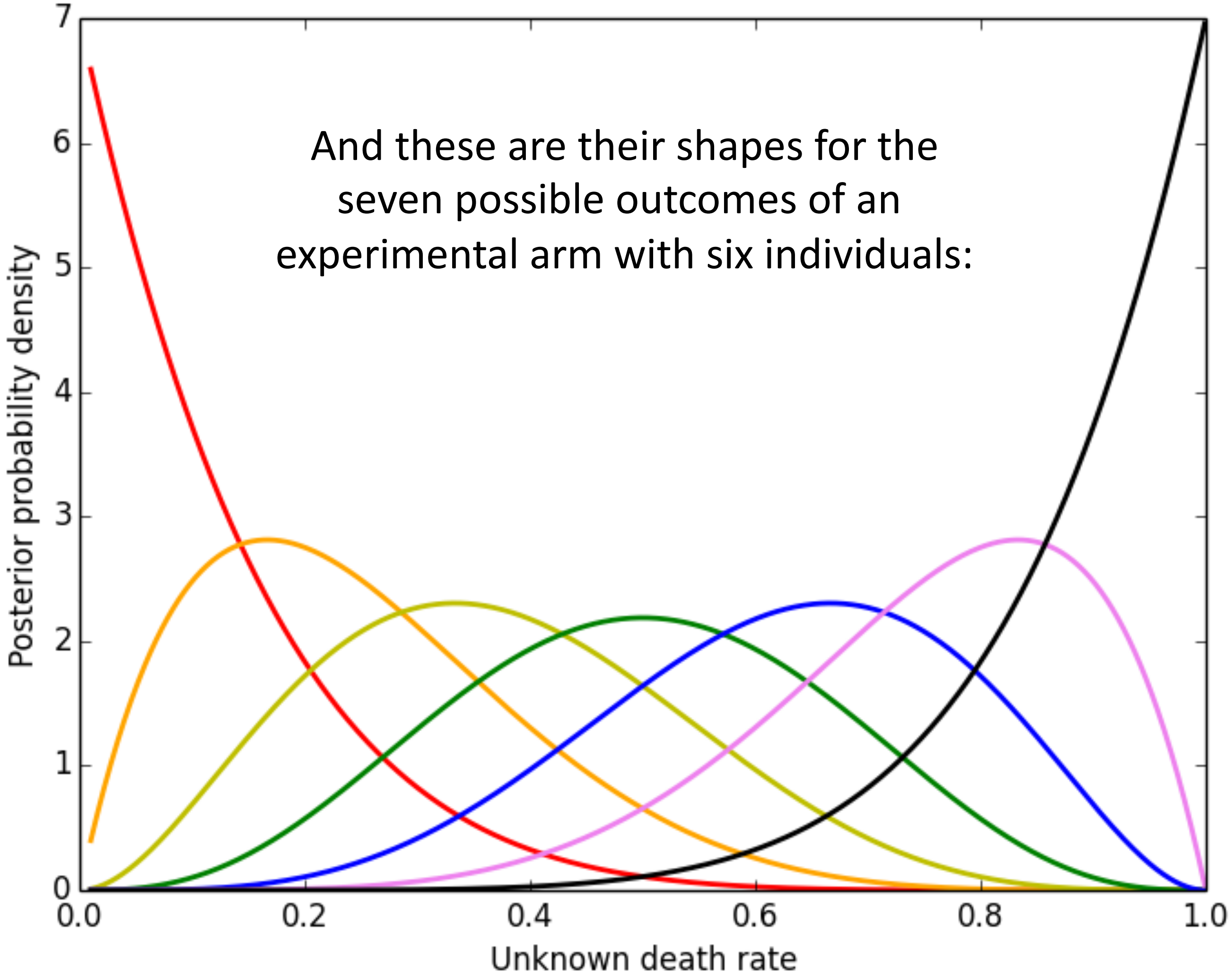


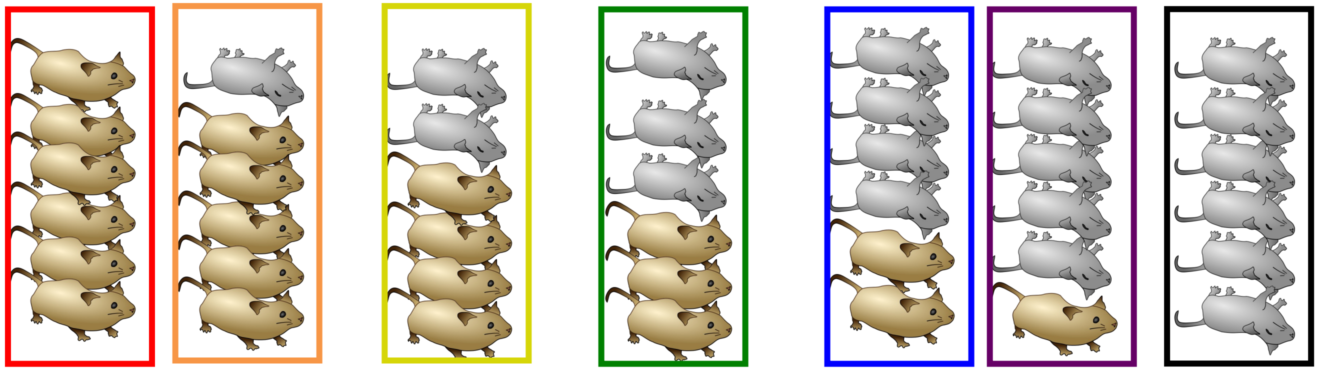
Beta(a=2, b=2) = c x (1-x)

Beta for n=3 binomial trial

$$p(x;0/3) = 4(1-x)^3$$
$$p(x;1/3) = 12x(1-x)^2$$
$$p(x;2/3) = 12x^2(1-x)$$
$$p(x;3/3) = 4x^3$$




And these are their shapes for the
seven possible outcomes of an
experimental arm with six individuals:
Posterior probability density
Unknown death rate
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0



The biased top

200 spins of a four-sided top.

Outcome A: 1
Outcome B: 34
Outcome C: 87
Outcome D: 78

The biased top

200 spins of a four-sided top.

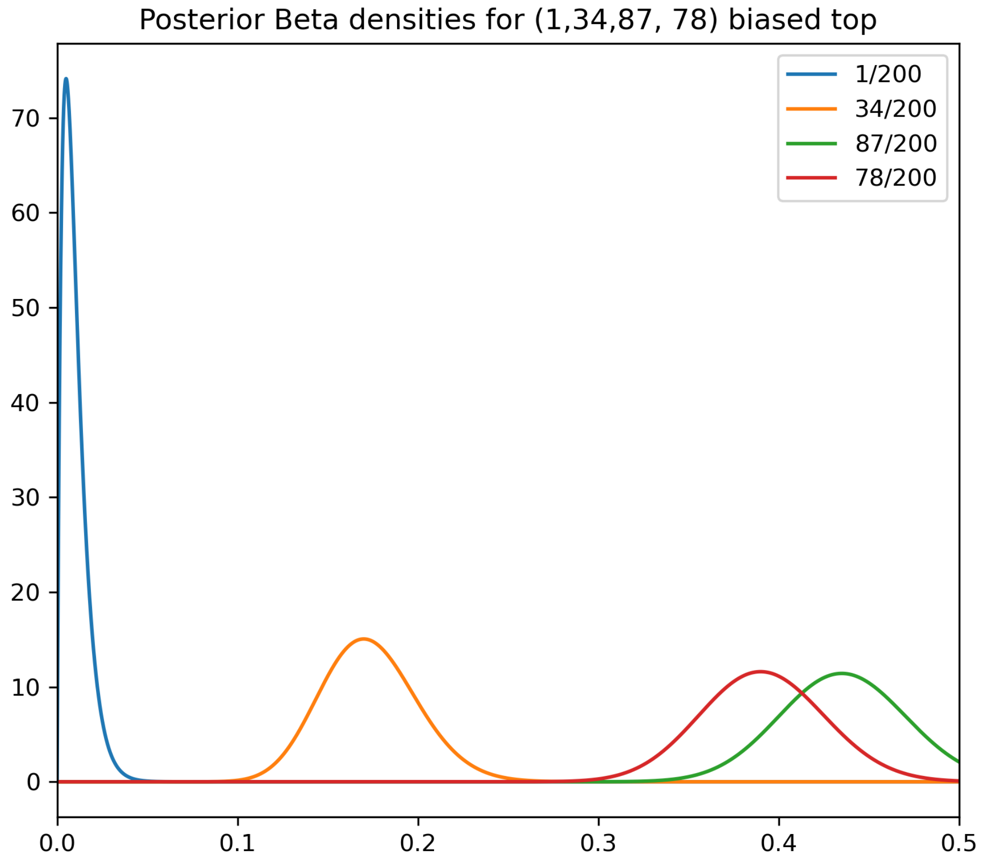
Outcome A:	1	Here be dragons
Outcome B:	34	0.17± 1.96 * S.E.M.
Outcome C:	87	0.435 ±1.96 *S.E.M.
Outcome D:	78	0.39 ± 1.96*S.E.M.
Outcome NOT A :	199/200	(can't use Standard Error on the Mean!)

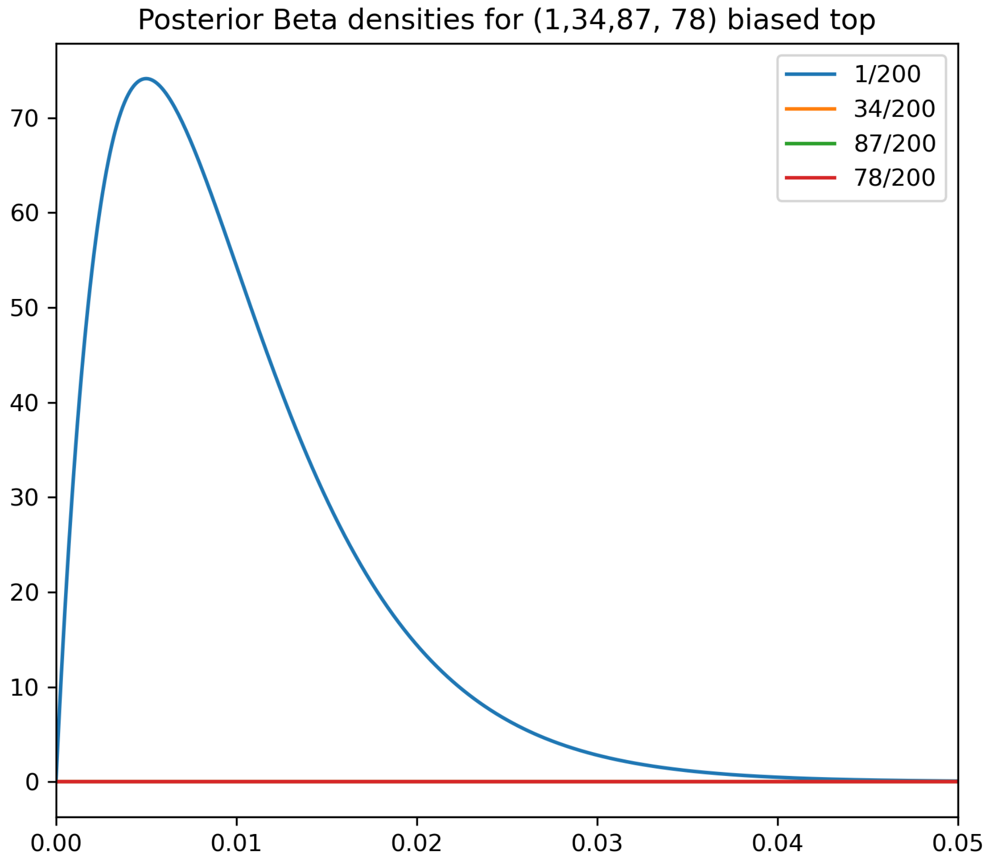
SEM only useful when min (Np, N(1-p)) > 30

The biased top

200 spins of a four-sided top.

1
34
87
78



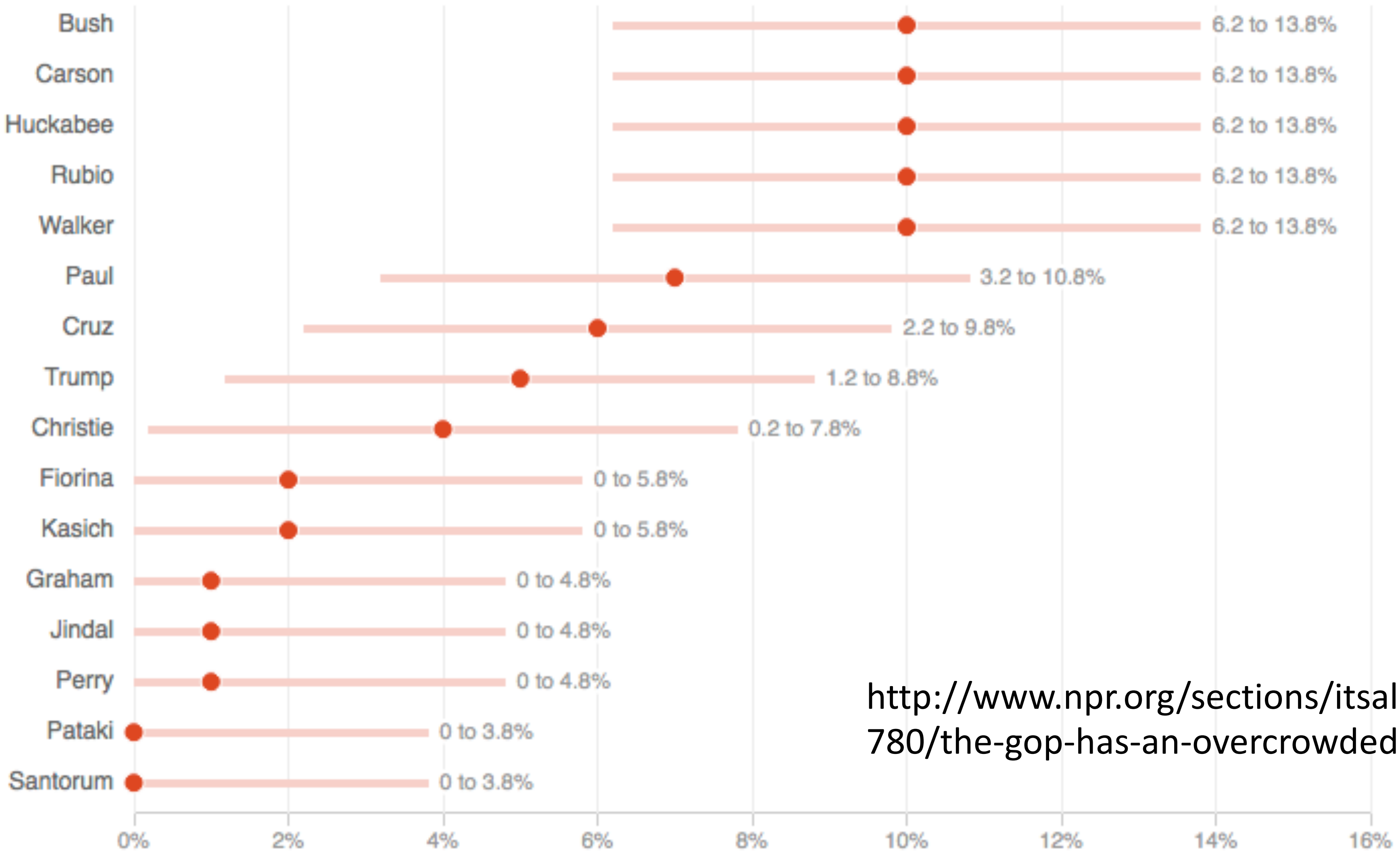


Where does probability come from?

- **Symmetry**
  - “All of the cards I haven’t seen face up are equally likely”
  - “The six faces of the die are equally likely”
- **Observations**
  - Sampling: Interpret (a small number of) observations as a sample from an infinite process.
  - Historical weather data, testing data, demographic data…
  - Scrambled versions of sampled data to explore contrary-to-fact symmetries
- **Modeling – when you can draw samples from the distribution of interest**
  - “Monte Carlo” methods – generate random samples
  - Weather, asteroids, pandemics
- **Expert opinion**
  - All you get for events like “North Korea will test another supersonic missile in Q1 2022”
  - Needs to map from “very likely” “highly unlikely” to real numbers
  - Numerical decisionmaking can extend into the wild and improbable (aliens, singularity..)

Republican Candidate Support, Factoring In The Margin Of Error

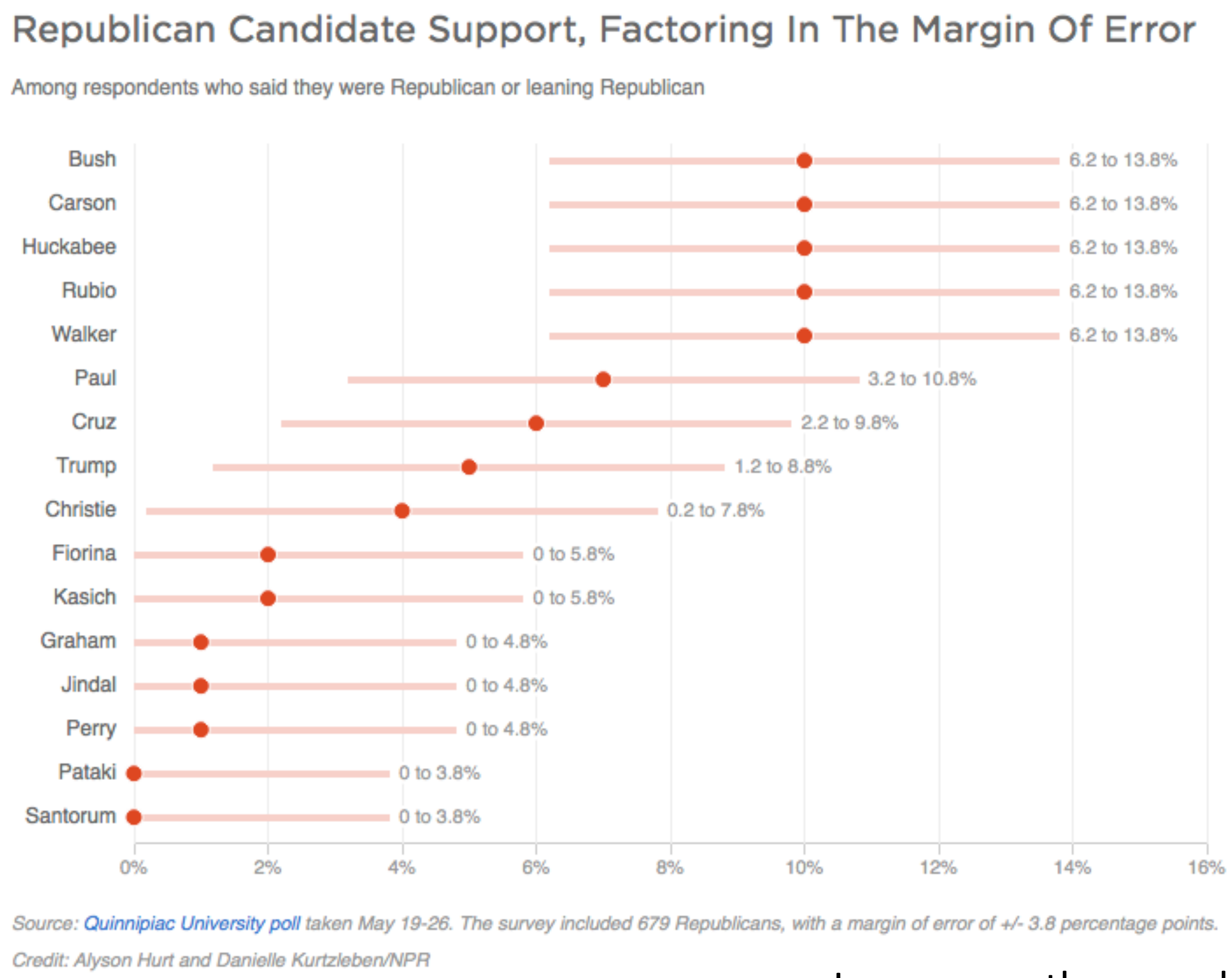
Among respondents who said they were Republican or leaning Republican



http://www.npr.org/sections/itsallpolitics/2015/05/29/410524780/the-gop-has-an-overcrowded-debate-problem

Source: Quinnipiac University poll taken May 19-26. The survey included 679 Republicans, with a margin of error of +/- 3.8 percentage points.

Credit: Alyson Hurt and Danielle Kurtzleben/NPR



- Proportions have been rounded.
- Error bars are symmetrical, clipped at zero
- Error bars are independent of point estimates!

In essence, the error bars are inappropriate.