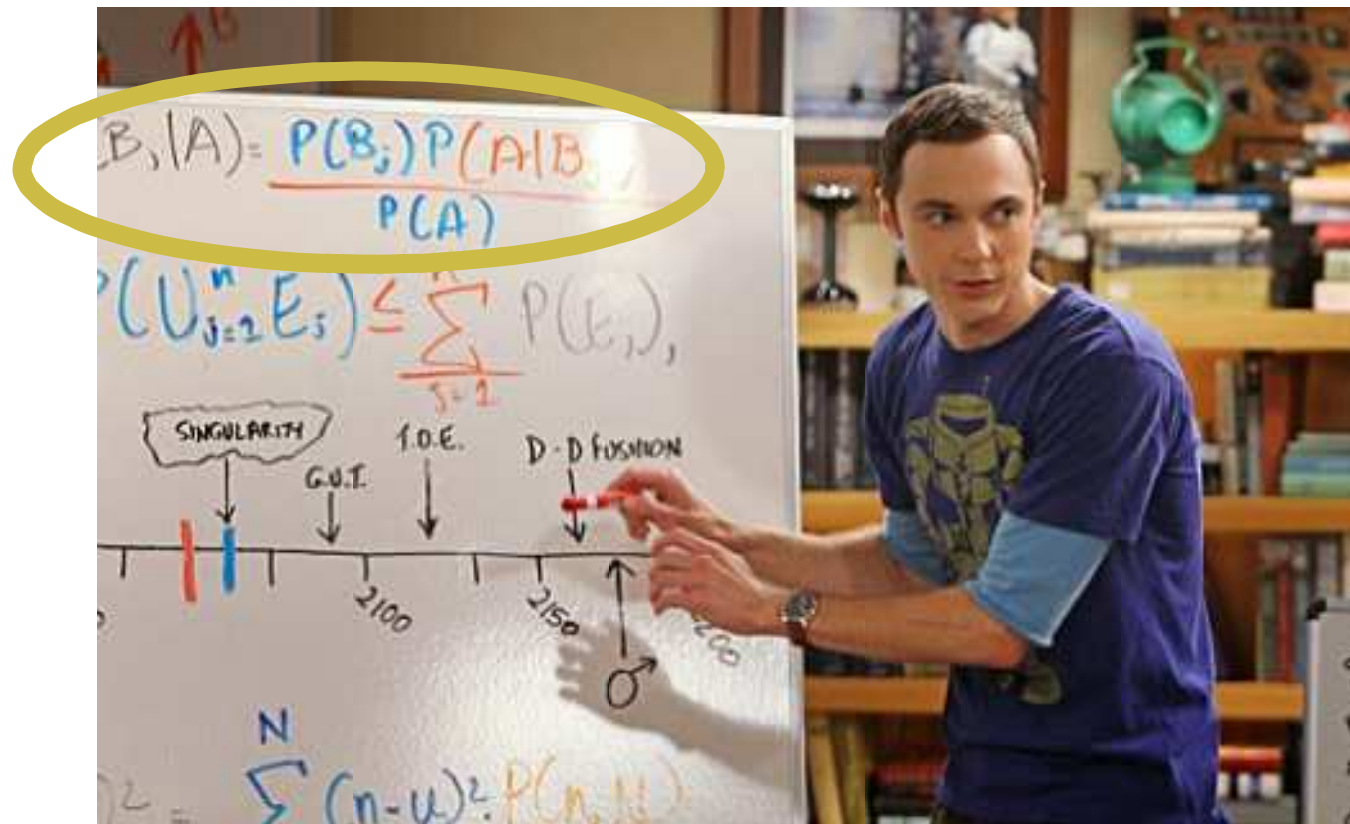


Bayesian probability theory







H



statement

conjectures

hypotheses



H

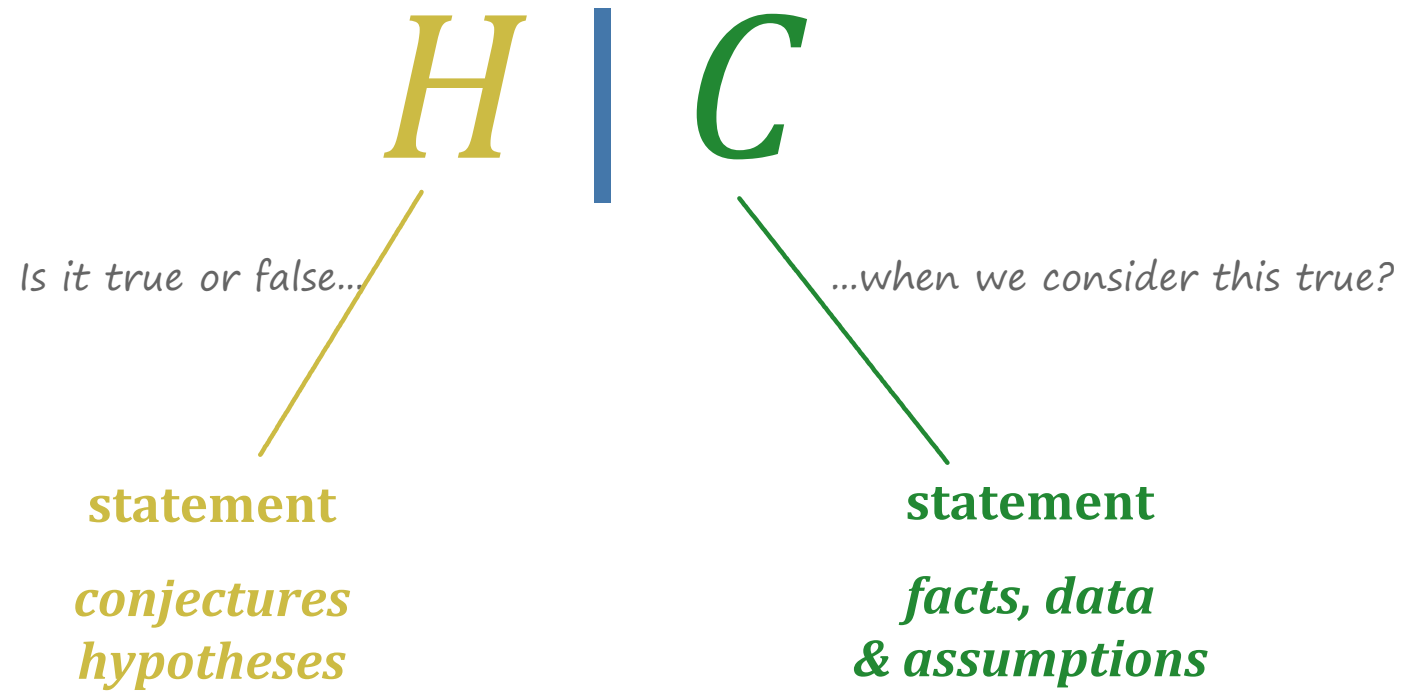
Is it true or false?

statement

conjectures

hypotheses

Probability of a statement given another statement



'Probability of statement H given statement C '

probability
credibility
plausibility
degree of belief

$P(H | C)$

Is it true or false...

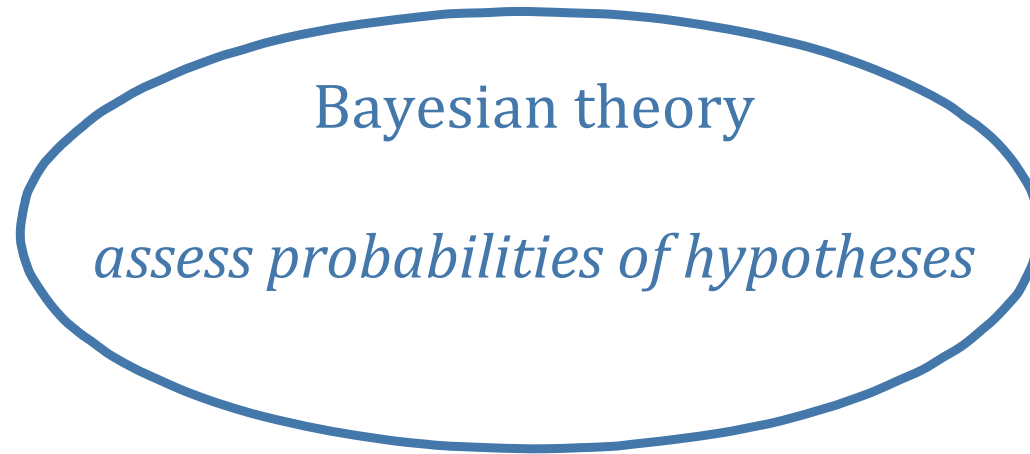
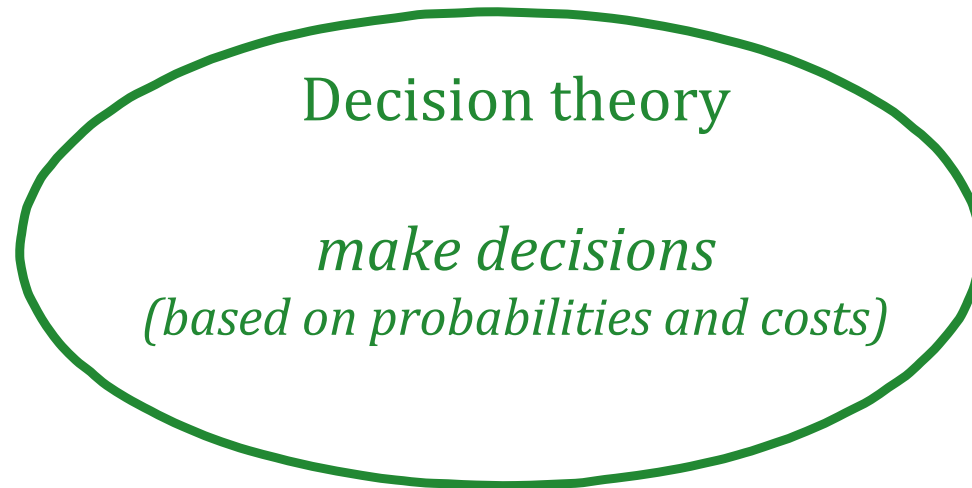
statement

*conjectures
hypotheses*

...when we consider this true?

statement

*facts, data
& assumptions*



probability
credibility
plausibility
degree of belief

$$P(H | C)$$

Is it true or false...

statement

conjectures
hypotheses

...when we consider this true?

statement

facts, data
& assumptions



probability
credibility
plausibility
degree of belief

$$C = D \text{ \& } A$$

facts, data *assumptions*

$$P(H | C)$$

statement
conjectures
hypotheses

statement
facts, data
& assumptions



probability
credibility
plausibility
degree of belief

$$C = D \ \& \ A$$

facts, data *assumptions*

$$P(H \mid D \ \& \ A)$$

statement
conjectures
hypotheses

statement
facts, data

statement
assumptions



The three basic rules of the probability calculus

$$P(\text{not-}Q \mid S) = 1 - P(Q \mid S)$$

$$P(Q \ \& \ R \mid S) = P(Q \mid R \ \& \ S) \cdot P(R \mid S)$$

$$P(Q \text{ or } R \mid S) = P(Q \mid S) + P(R \mid S) - P(Q \ \& \ R \mid S)$$

*All probability calculations and results, however complicated they might look,
are just the application of the three rules above, over and over and over again*

The three basic rules of the probability calculus

$$P(\text{not-}Q \mid S) = 1 - P(Q \mid S)$$

$$P(Q \ \& \ R \mid S) = P(Q \mid R \ \& \ S) \cdot P(R \mid S)$$

$$P(Q \text{ or } R \mid S) = P(Q \mid S) + P(R \mid S) - P(Q \ \& \ R \mid S)$$

Rules of Inference: Elementary Valid Argument Forms

Name	Abbreviation	Form
1. Modus Ponens	M.P.	$p \supset q$ p $\therefore q$
2. Modus Tollens	M.T.	$p \supset q$ $\neg q$ $\therefore \neg p$
3. Hypothetical Syllogism	H.S.	$p \supset q$ $q \supset r$ $\therefore p \supset r$
4. Disjunctive Syllogism	D.S.	$p \vee q$ $\neg p$ $\therefore q$
5. Constructive Dilemma	C.D.	$(p \supset q) \vee (r \supset s)$ $p \vee r$ $\therefore q \vee s$
6. Absorption	Abs.	$p \supset q$ $\therefore p \supset (p \wedge q)$
7. Simplification	Simp.	$p \wedge q$ $\therefore p$
8. Conjunction	Conj.	p q $\therefore p \wedge q$
9. Addition	Add.	p $\therefore p \vee q$

Introduction to LOGIC



Irving M. Copi
Carl Cohen
Victor Rodych

10th edition





Bayes's theorem

$$P(H_{\#1} \mid D \ \& \ A) = \frac{P(D \mid H_{\#1} \ \& \ A) \cdot P(H_{\#1} \mid A)}{P(D \mid H_{\#1} \ \& \ A) \cdot P(H_{\#1} \mid A) + P(D \mid H_{\#2} \ \& \ A) \cdot P(H_{\#2} \mid A) + \dots}$$

(sum over all possible hypotheses)



Bayes's theorem

$$P(H_{\#1} \mid D \ \& \ A) = \frac{P(D \mid H_{\#1} \ \& \ A) \cdot P(H_{\#1} \mid A)}{P(D \mid H_{\#1} \ \& \ A) \cdot P(H_{\#1} \mid A) + P(D \mid H_{\#2} \ \& \ A) \cdot P(H_{\#2} \mid A) + \dots}$$

(sum over all possible hypotheses)

$$P(\textit{hypothesis} \mid \textit{data} \ \& \ \textit{assumptions}) \propto P(\textit{data} \mid \textit{hypothesis} \ \& \ \textit{assumptions}) \times P(\textit{hypothesis} \mid \textit{assumptions})$$

Probability of some hypotheses, given data

$$P(H \mid D \ \& \ A) \propto P(D \mid H \ \& \ A) \cdot P(H \mid A)$$

Typical elements of Bayesian analysis

Probability of some hypotheses, given data

$$P(H \mid D \ \& \ A) \propto P(D \mid H \ \& \ A) \cdot P(H \mid A)$$

- Formulate precise hypotheses
- Assess which data we have or need
- Examine which assumptions we need to make
- Assess the probability of the data given each hypothesis
- Assess the pre-data probability of each hypothesis



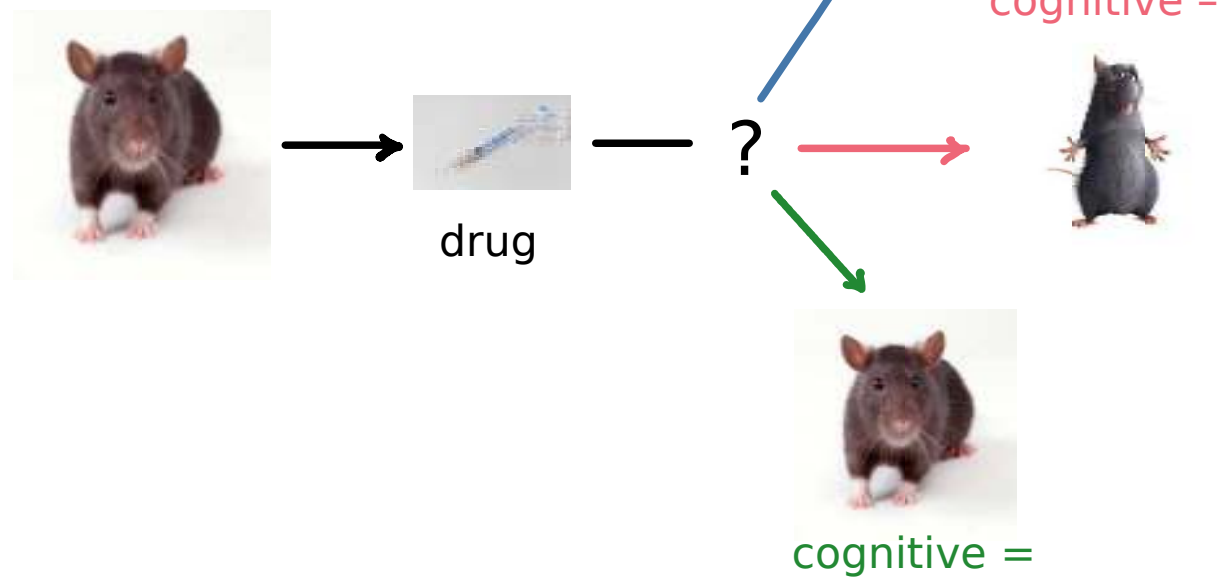
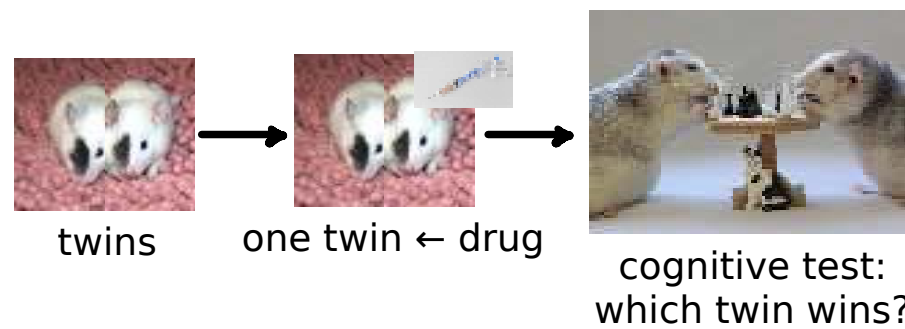
Bayesian probability theory forces us
to state *clearly* and *precisely*:

- What are our conjectures/hypotheses?
 - What are our facts?
 - What are our assumptions?



The rats & drug investigation:

approach via Bayesian probability theory



- Formulate precise hypotheses
- Assess which data we have or need
- Examine which assumptions we need to make
- Assess the probability of the data given each hypothesis
- Assess the pre-data probability of each hypothesis

Probability of each hypothesis, given the data

$$P(H \mid D \ \& \ A) \propto P(D \mid H \ \& \ A) \cdot P(H \mid A)$$



► Formulate precise hypotheses

- *What is our question?*
- *What is the purpose of this study?*

- Assess which data we have or need
- Examine which assumptions we need to make
- Assess the probability of the data given each hypothesis
- Assess the pre-data probability of each hypothesis



What are our hypotheses? What are our question & purpose?

Does the drug enhance cognitive abilities?

→ *Yes, No, Sometimes, It depends, ...*

What are our hypotheses? What are our question & purpose?

~~*Does the drug enhance cognitive abilities?*~~

~~→ *Yes, No, Sometimes, It depends, ...*~~

Ultimate question, but too complex

(we'll return to it later)



What are our hypotheses? What are our question & purpose?

Is there a systematic effect?

→ *Yes, No?*

What are our hypotheses? What are our question & purpose?

Is there a systematic effect?

→ *Yes, No?*

Less complex, but too vague

What do we mean by 'systematic'?

If 'systematic' = 'every time' then the answer is No.

We want a question that is

- less categorical
- quantifiable



What are our hypotheses? What are our question & purpose?

How many of the tested rats show increased cognitive abilities?

→ 0, 1, 2, ..., 17

What are our hypotheses? What are our question & purpose?

How many of the tested rats show increased cognitive abilities?

→ 0, 1, 2, ..., 17

Better! But no probabilities here:
after the experiment, we know the exact answer with 100% certainty.

Are we interested in *these specific* 17 rat twins *only*?



What are our hypotheses? What are our question & purpose?

*Would the drug lead to a positive result, if tested on a **new** twin pair?*

→ *Yes, No*

What are our hypotheses? What are our question & purpose?

✓ *Would the drug lead to a positive result, if tested on a **new** twin pair?*

→ *Yes, No*

Good!

We are concretely asking if our study *extrapolates*.
This question can be answered also in practice.



What are our hypotheses? What are our question & purpose?

*In a **much larger number** of tests, how many positive results?*

→ $n = 0, 1, 2, \dots$, full population N

What are our hypotheses? What are our question & purpose?

✓ *In a **much larger number** of tests, how many a positive results?*
→ *$n = 0, 1, 2, \dots$, full population N*

Also good!

Although practically impossible to answer experimentally,
this question has clear and unequivocal answers.
It quantifies ‘how systematic’ the effect is.

NB: We need to specify what’s the “full population”

The probabilities for the two good questions are often connected:

$$P(\text{New} \mid \text{data}) = \sum n/N P(n \mid \text{data})$$

What are our hypotheses? What are our question & purpose?

✓ *In a **much larger number** of tests, how many a positive results?*
→ *$n = 0, 1, 2, \dots$, full population N*

Also good!

Although practically impossible to answer experimentally,
this question has clear and unequivocal answers.
It quantifies ‘how systematic’ the effect is

NB: We need to specify what’s the “full population”

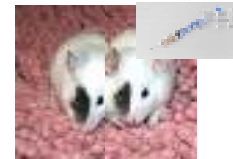
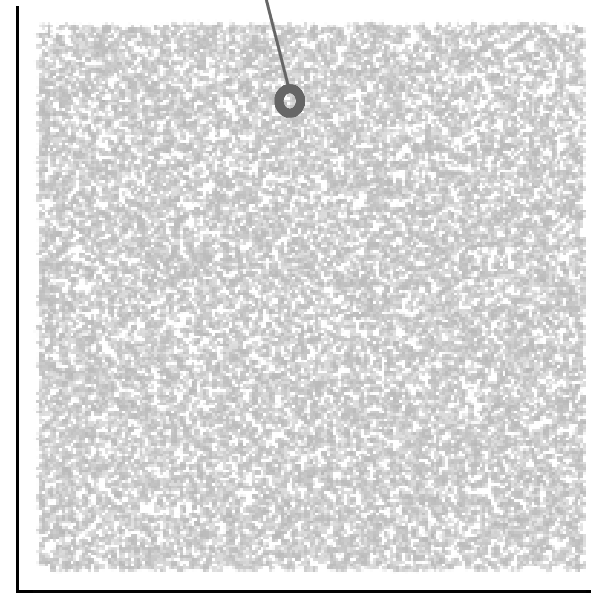
The probabilities for the two good questions are often connected:

$$P(\text{New} \mid \text{data}) = \sum n/N P(n \mid \text{data})$$



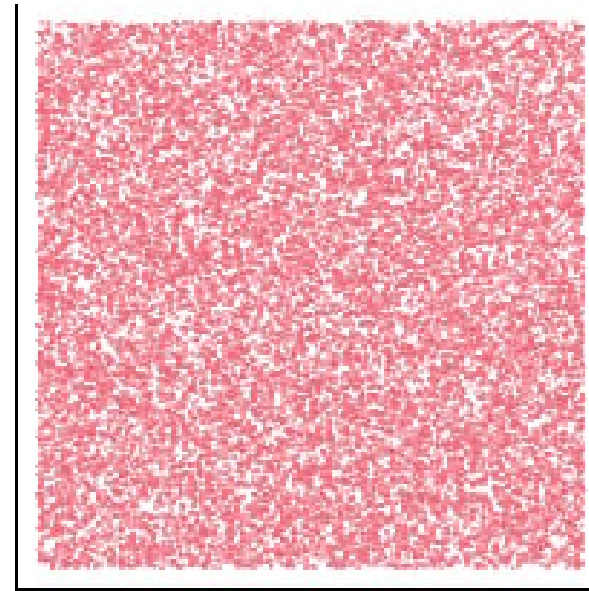
Let's compare 1 000 000 001 concrete hypotheses:

Let's compare 1 000 000 001 concrete hypotheses:



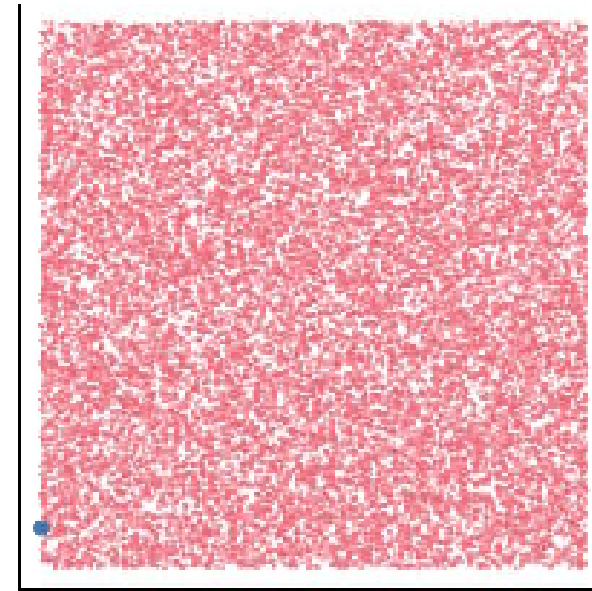
Let's compare 1 000 000 001 concrete hypotheses:

H_0 = 'In 1 billion tests, **no** test yields cognitive+ (all cognitive-).'



Let's compare 1 000 000 001 concrete hypotheses:

H_0 = 'In 1 billion tests, **no** test yields cognitive+ (all cognitive-).'

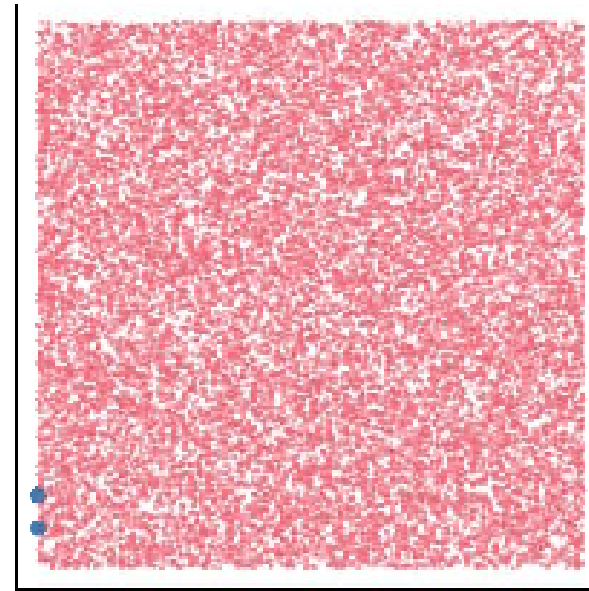


Let's compare 1 000 000 001 concrete hypotheses:

H_0 = 'In 1 billion tests, **no** test yields cognitive+ (all cognitive-).'

$H_{0.000\,000\,001}$ = 'In 1 billion tests, **1** test yields cognitive+'

$H_{0.000\,000\,002}$ = 'In 1 billion tests, **2** tests yield cognitive+'



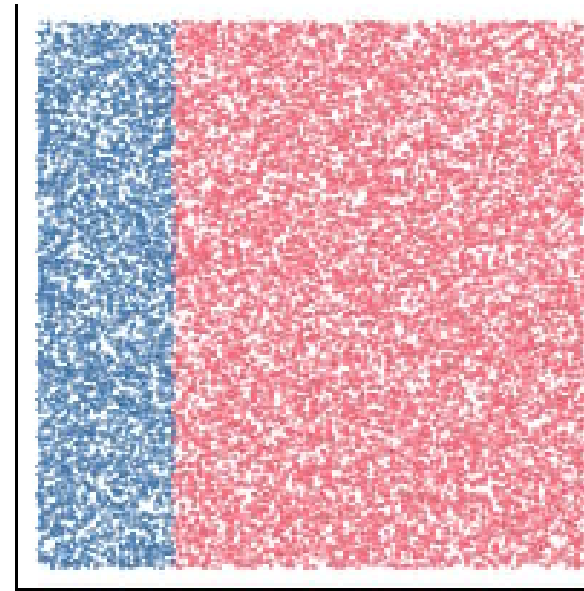
Let's compare 1 000 000 001 concrete hypotheses:

H_0 = 'In 1 billion tests, **no** test yields cognitive+ (all cognitive-).'

$H_{0.000\,000\,001}$ = 'In 1 billion tests, **1** test yields cognitive+.'

$H_{0.000\,000\,002}$ = 'In 1 billion tests, **2** tests yield cognitive+.'

...



Let's compare 1 000 000 001 concrete hypotheses:

H_0 = 'In 1 billion tests, **no** test yields cognitive+ (all cognitive-).'

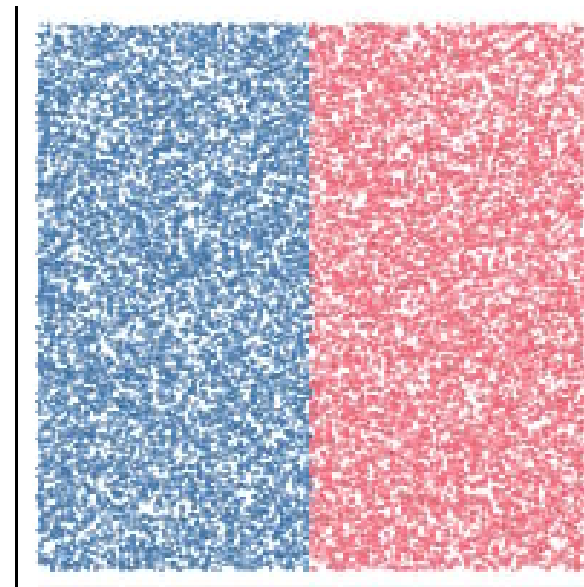
$H_{0.000\,000\,001}$ = 'In 1 billion tests, **1** test yields cognitive+'

$H_{0.000\,000\,002}$ = 'In 1 billion tests, **2** tests yield cognitive+'

...

$H_{0.25}$ = 'In 1 billion tests, **250 000 000** yield cognitive+'

...



Let's compare 1 000 000 001 concrete hypotheses:

H_0 = 'In 1 billion tests, **no** test yields cognitive+ (all cognitive-).'

$H_{0.000\,000\,001}$ = 'In 1 billion tests, **1** test yields cognitive+'

$H_{0.000\,000\,002}$ = 'In 1 billion tests, **2** tests yield cognitive+'

...

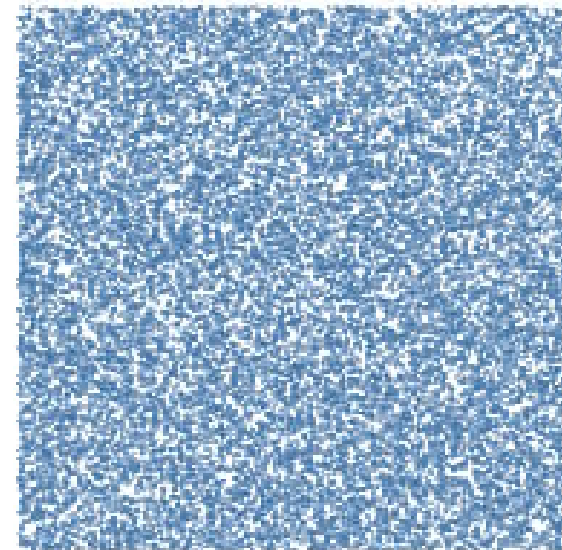
$H_{0.25}$ = 'In 1 billion tests, **250 000 000** yield cognitive+'

...

$H_{0.5}$ = 'In 1 billion tests, **500 000 000** yield cognitive+'

...

H_1 = 'In 1 billion tests, **all** tests yield cognitive+'



Let's compare 1 000 000 001 concrete hypotheses:

H_0 = 'In 1 billion tests, **no** test yields cognitive+ (all cognitive-)'

$H_{0.000\,000\,001}$ = 'In 1 billion tests, **1** test yields cognitive+'

...

H_1 = 'In 1 billion tests, **all** tests yield cognitive+'



H_f = 'In 1 000 000 000 tests, a fraction **f** yield cognitive+'

f = 0, 1/1 000 000 000, 2/1 000 000 000, ..., 999 999 999/1 000 000 000, 1

(all

-)

(all +)

- 
- Formulate precise hypotheses

► Assess which data we have or need

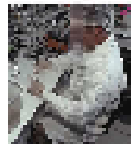
- Examine which assumptions we need to make
- Assess the probability of the data given each hypothesis
- Assess the pre-data probability of each hypothesis

Which data do we have?

17 twins tested: 13 drug → cognitive+, 4 drug → cognitive-



- Lab#1's stopping rule: test 17*



Lab#1's stopping rule:
test 17

- Lab#2's stopping rule: test until at least four "+" and "-"*



Lab#2's stopping rule:
test until at least four "+" and "-"

- 
- Formulate precise hypotheses
 - Assess which data we have or need
 - ▶ Examine which assumptions we need to make
 - Assess the probability of the data given each hypothesis
 - Assess the pre-data probability of each hypothesis



What do we need to assume?

Assumptions are necessary for two purposes:

- to assess the probability of the data, given each hypothesis
- to assess the pre-data probability of each hypothesis

What do we need to assume?

- to assess the probability of the data, given each hypothesis

$P(\text{'In 17 tests, 13 cognitive+'} \mid \text{'In 1 billion tests, 100 cognitive+'} \ \& \ A)$

What do we need to assume?

- to assess the probability of the data, given each hypothesis

$$P(\text{'In 17 tests, 13 cognitive+'} \mid \text{'In 1 billion tests, 100 cognitive+'} \ \& \ A)$$

Assumptions:

What do we need to assume?

- to assess the probability of the data, given each hypothesis

$$P(\text{'In 17 tests, 13 cognitive+'} \mid \text{'In 1 billion tests, 100 cognitive+'} \ \& \ A)$$

Assumptions:

- *The tested rats are part of the larger set of 1 billion tests*
- *The tested rats are **not specially** chosen from the larger population*

What do we need to assume?

- to assess the probability of the data, given each hypothesis

$$P(\text{'In 17 tests, 13 cognitive+'} \mid \text{'In 1 billion tests, 100 cognitive+'} \ \& \ A)$$

Assumptions:

- *The tested rats are part of the larger set of 1 billion tests*
- *The tested rats are **not specially** chosen from the larger population*

→ If some tested rats were unsystematically exchanged with some in the remaining population, our results would still be valid

We say that the tested rats are exchangeable with the full population

What do we need to assume?

- to assess the probability of the data, given each hypothesis

$$P(\text{'In 17 tests, 13 cognitive+'} \mid \text{'In 1 billion tests, 100 cognitive+'} \ \& \ A)$$

Assumptions:

- *The tested rats are part of the larger set of 1 billion tests*
- *The tested rats are **not specially** chosen from the larger population*

→ If some tested rats were unsystematically exchanged with some in the remaining population, our results would still be valid

The Journal of Philosophy
1981, Vol. 78, No. 2, 45-58

THE ROLE OF EXCHANGEABILITY IN INFERENCE¹

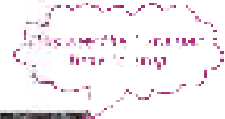
BY D. V. LINDLEY AND MELVIN R. NOVICK

University College London and The University of Iowa

This paper is concerned with basic problems of statistical inference. The thesis is in three parts: (1) that inference is a procedure whereby one passes from a population (or sample) to a new individual; (2) that this connection can be established using de Finetti's idea of exchangeability or Fisher's concept of a subpopulation; (3) in making the connection use must be made of the appropriate probability. These three principles are used in a variety of situations and the topics discussed include analysis of variance and covariance, contingency tables, and calibration. Some comments on randomization are also included.

- 
- Formulate precise hypotheses
 - Assess which data we have or need
 - Examine which assumptions we need to make
 - ▶ Assess the probability of the data given each hypothesis
 - Assess the pre-data probability of each hypothesis

Probability of data given hypotheses: Lab#1



$$P(13+ \ 4- \mid H_{0.7} \ \& \ A) = \binom{17}{4} \times 0.7^{13} \times (1 - 0.7)^4 = 0.1868$$

*(This is an approximation to 8 significant digits:
the correct distribution is a hypergeometric one)*

Probability of data given hypotheses: Lab#2



$$P(13+ \ 4- \mid H_{0.7} \ \& \ A) = \binom{16}{3} \times 0.7^{13} \times (1 - 0.7)^4 = 0.04395$$

↑
because of the stopping rule
we couldn't shuffle the last -

Probability of sequence given hypotheses (same for both labs)

$$P(+ + + + - + + + - + + + + - + + - | H_{0.7} \& A) = 0.7^{13} \times (1 - 0.7)^4 = 0.00007848$$

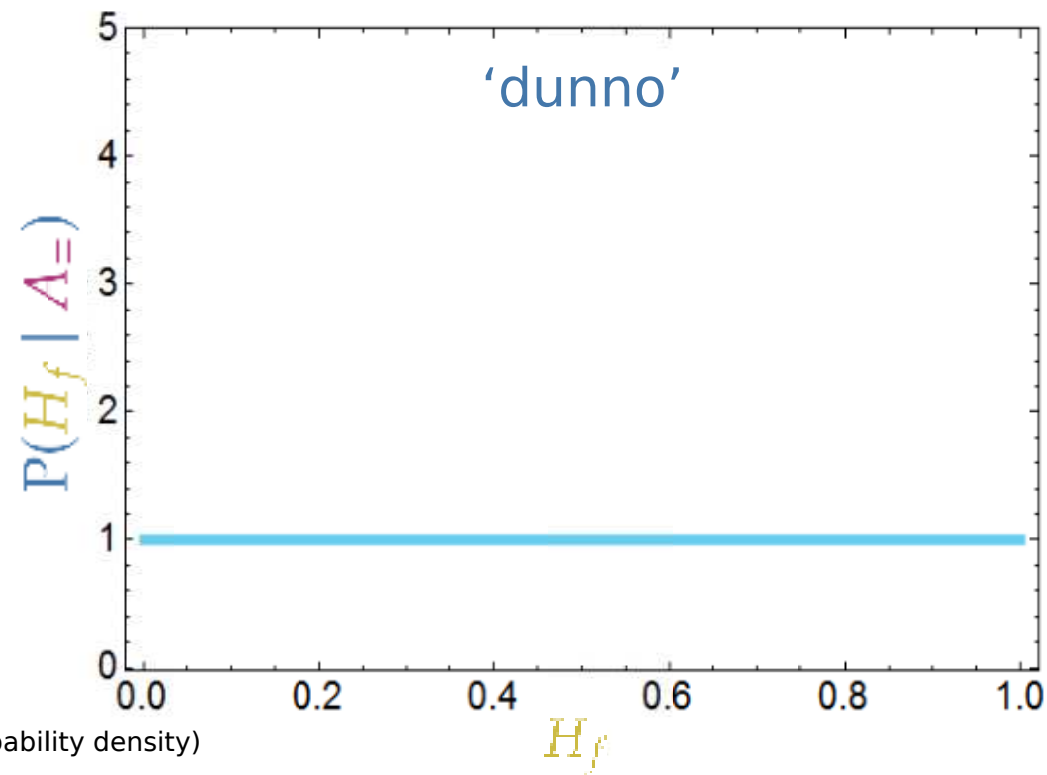
(Considering the sequences as outcomes would lead to a p -value = 1)

- 
- Formulate precise hypotheses
 - Assess which data we have or need
 - Examine which assumptions we need to make
 - Assess the probability of the data given each hypothesis
 - ▶ Assess the pre-data probability of each hypothesis

Pre-data probabilities of hypotheses

Let's consider three possible assumptions as examples:

A_+ : '*Dunno*', all frequencies equally plausible



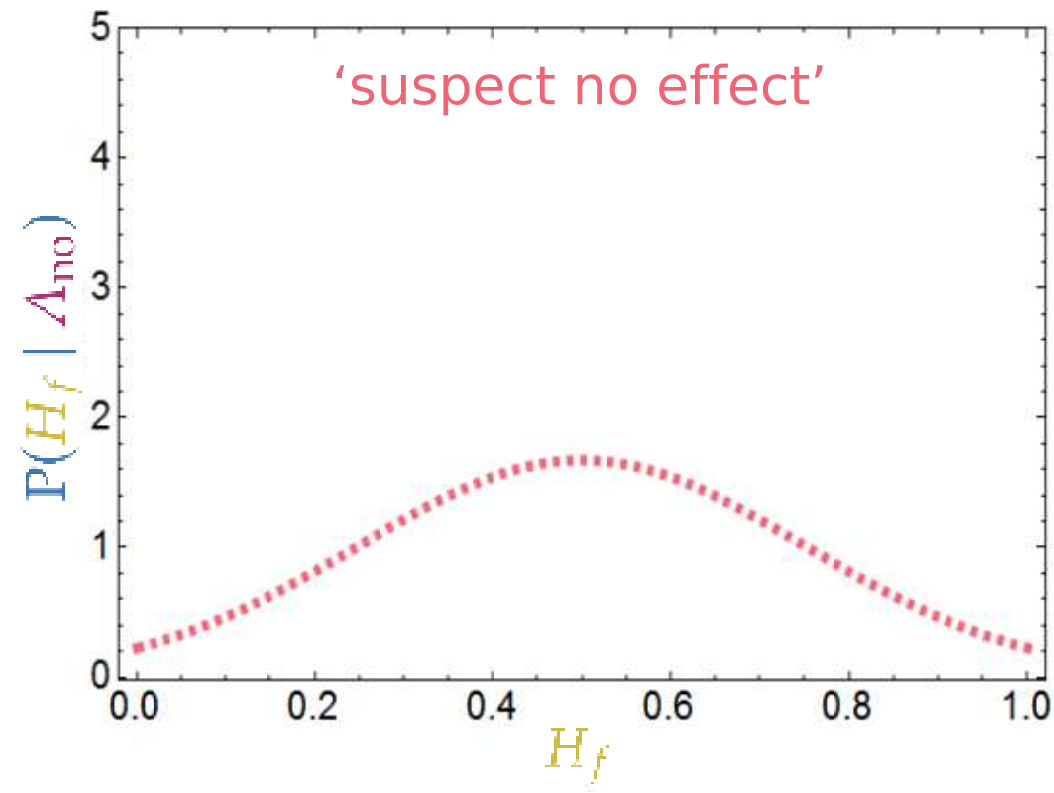
y-axis multiplied by 100 000 000 (probability density)

Pre-data probabilities of hypotheses

Let's consider three possible assumptions as examples:

$A_{=}$: '*Dunno*', all frequencies equally plausible

A_{no} : '*Suspect **no** effect*', frequencies $f \sim 0.5$ slightly more plausible
(equal number of $+$ and $-$)



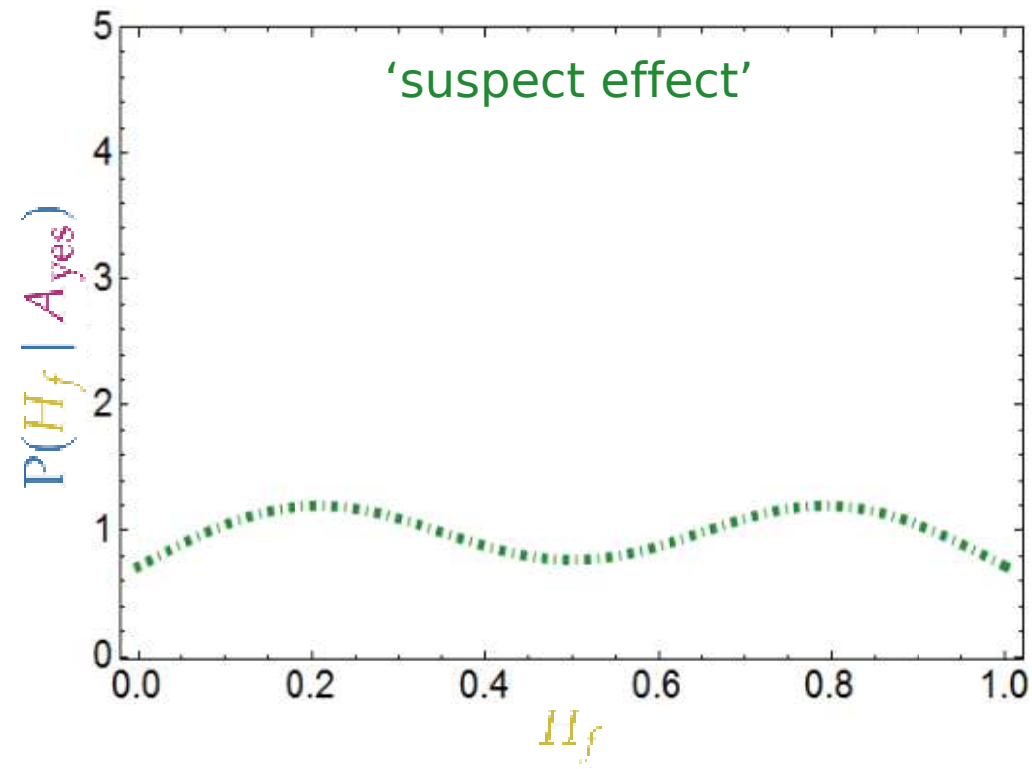
Pre-data probabilities of hypotheses

Let's consider three possible assumptions as examples:

$A_{=}$: '*Dunno*', all frequencies equally plausible

A_{no} : '*Suspect **no** effect*', frequencies ~ 0.5 slightly more plausible
(equal number of + and -)

A_{yes} : '*Suspect effect*', frequencies $f \sim 0.25, 0.75$ slightly more plausible
(fewer + than - or vice versa)



- Formulate precise hypotheses
- Assess which data we have or need
- Examine which assumptions we need to make
- Assess the probability of the data given each hypothesis
- Assess the pre-data probability of each hypothesis

► Probability of each hypothesis, given the data

$$P(H \mid D \ \& \ A) \propto P(D \mid H \ \& \ A) \cdot P(H \mid A)$$



$$P(\text{hypothesis} \mid \text{data} \ \& \ \text{assumptions}) \propto P(\text{data} \mid \text{hypothesis} \ \& \ \text{assumptions}) \cdot P(\text{hypothesis} \mid \text{assumptions})$$

$$P(H_f \mid D \ \& \ A) = \frac{P(D \mid H_f \ \& \ A) \cdot P(H_f \mid A)}{P(D \mid H_0 \ \& \ A) \cdot P(H_0 \mid A) + \dots + P(D \mid H_1 \ \& \ A) \cdot P(H_1 \mid A)}$$

(1 000 000 001 terms)

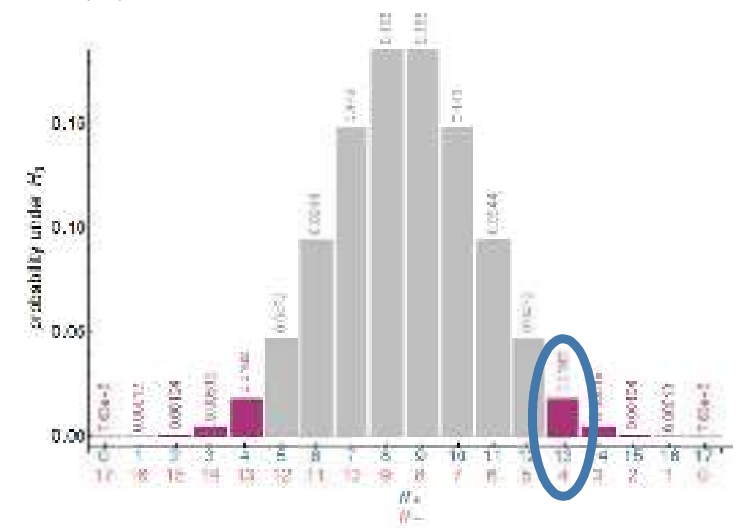
$$P(\text{hypothesis} \mid \text{data} \ \& \ \text{assumptions}) \propto P(\text{data} \mid \text{hypothesis} \ \& \ \text{assumptions}) \cdot P(\text{hypothesis} \mid \text{assumptions})$$

$$P(H_f \mid D \ \& \ A) = \frac{P(D \mid H_f \ \& \ A) \cdot P(H_f \mid A)}{P(D \mid H_0 \ \& \ A) \cdot P(H_0 \mid A) + \dots + P(D \mid H_1 \ \& \ A) \cdot P(H_1 \mid A)}$$

(1 000 000 001 terms)

Bayes's formula:

- is *not* listing outcomes that *could* have happened (but didn't)
- is listing *alternative* hypotheses



Probability of hypotheses given data: Lab#1

$$P(H_f \mid \text{data lab\#1} \ \& \ A) =$$

$$P(D \mid H_f \ \& \ A) \cdot P(H_f \mid A)$$

$$P(D \mid H_0 \ \& \ A) \cdot P(H_0 \mid A) + \cdots + P(D \mid H_{0.5} \ \& \ A) \cdot P(H_{0.5} \mid A) + \cdots + P(D \mid H_1 \ \& \ A) \cdot P(H_1 \mid A)$$

Probability of hypotheses given data: Lab#1

$$P(H_f \mid \text{data lab\#1} \ \& \ A) =$$

$$\binom{17}{4} \cdot f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)$$

$$\binom{17}{4} \cdot 0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \dots + \binom{17}{4} \cdot 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \dots + \binom{17}{4} \cdot 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)$$

Probability of hypotheses given data: Lab#1

$$P(H_f \mid \text{data lab\#1} \ \& \ A) =$$

$$\binom{17}{4} \cdot f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)$$

$$\binom{17}{4} \cdot 0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \dots + \binom{17}{4} \cdot 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \dots + \binom{17}{4} \cdot 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)$$

Probability of hypotheses given data: Lab#1

$$P(H_f \mid \text{data lab\#1} \ \& \ A) =$$

$$\frac{\binom{17}{4} \cdot f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{\binom{17}{4} \cdot 0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + \binom{17}{4} \cdot 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + \binom{17}{4} \cdot 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$
$$= \frac{f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$

Probability of hypotheses given data: Lab#2

$$P(H_f \mid \text{data lab\#2} \ \& \ A) =$$

$$\frac{\binom{16}{3} \cdot f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{\binom{16}{3} \cdot 0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + \binom{16}{3} \cdot 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + \binom{16}{3} \cdot 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$

↑
we couldn't shuffle the last —
because of the stopping rule

Probability of hypotheses given data: Lab#2

$$P(H_f \mid \text{data lab\#2} \ \& \ A) =$$

$$\frac{\binom{16}{3} \cdot f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{\binom{16}{3} \cdot 0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + \binom{16}{3} \cdot 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + \binom{16}{3} \cdot 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$

Probability of hypotheses given data: Lab#2

$$P(H_f \mid \text{data lab\#2} \ \& \ A) =$$

$$\frac{\binom{16}{3} \cdot f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{\binom{16}{3} \cdot 0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + \binom{16}{3} \cdot 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + \binom{16}{3} \cdot 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$
$$= \frac{f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$

Probability of hypotheses given data *sequence*

$$P(H_f \mid \text{data sequence} \ \& \ A) =$$

$$\frac{f^{13} \cdot (1 - f)^4 \cdot P(H_f \mid A)}{0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$



$$P(H_f \mid \text{data lab\#1} \ \& \ A) = P(H_f \mid \text{data lab\#2} \ \& \ A) =$$

$$P(H_f \mid \text{data sequence} \ \& \ A) =$$

$$\frac{f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$

stopping rules don't affect the final probability!



$$P(H_f \mid \text{data lab\#1} \ \& \ A) = P(H_f \mid \text{data lab\#2} \ \& \ A) =$$

$$P(H_f \mid \text{data sequence} \ \& \ A) =$$

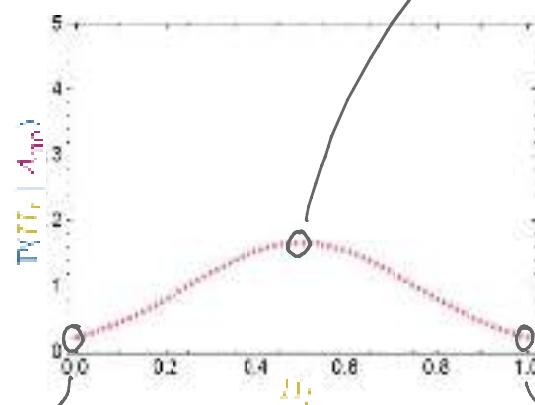
$$\frac{f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \cdots + 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \cdots + 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$

Now let's substitute the pre-data probabilities

$$P(H_f \mid \text{data lab\#1} \ \& \ A) = P(H_f \mid \text{data lab\#2} \ \& \ A) =$$

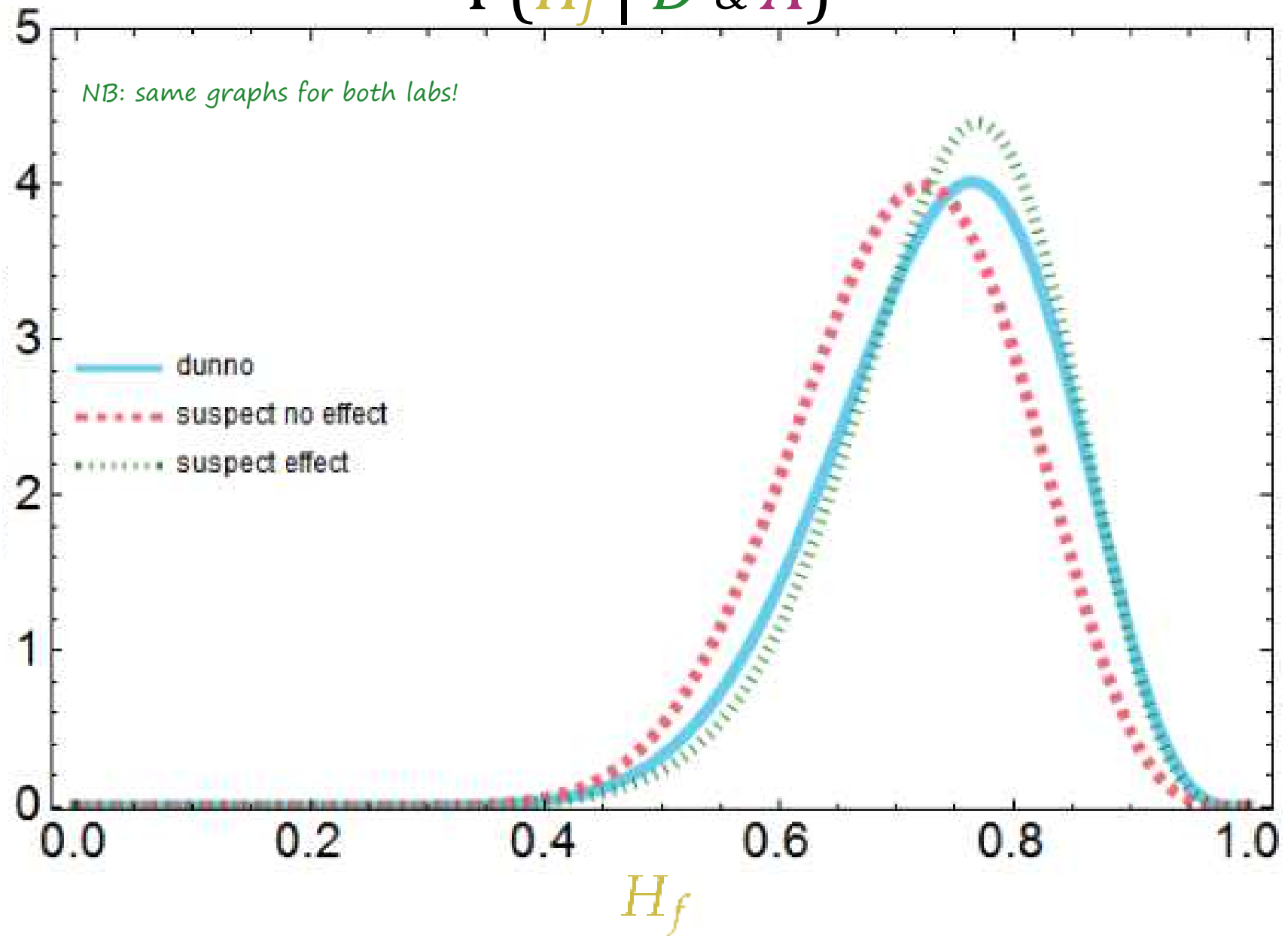
$$P(H_f \mid \text{data sequence} \ \& \ A) =$$

$$\frac{f^{13} \cdot (1-f)^4 \cdot P(H_f \mid A)}{0^{13} \cdot 1^4 \cdot P(H_0 \mid A) + \dots + 0.5^{13} \cdot 0.5^4 \cdot P(H_{0.5} \mid A) + \dots + 1^{13} \cdot 0^4 \cdot P(H_1 \mid A)}$$





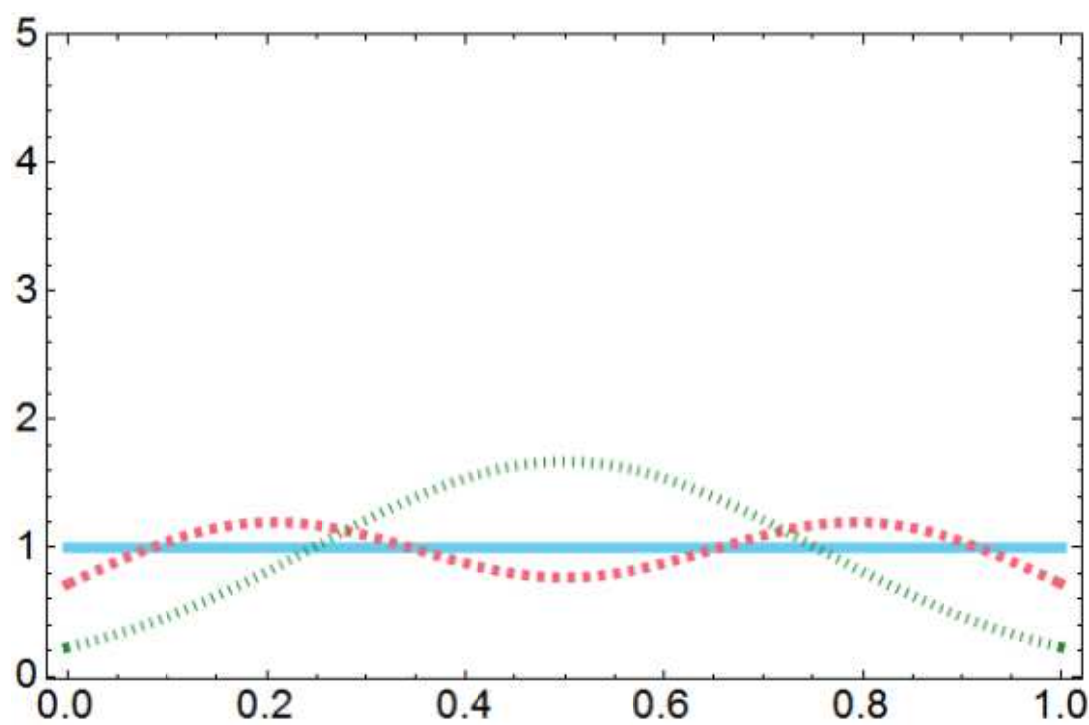
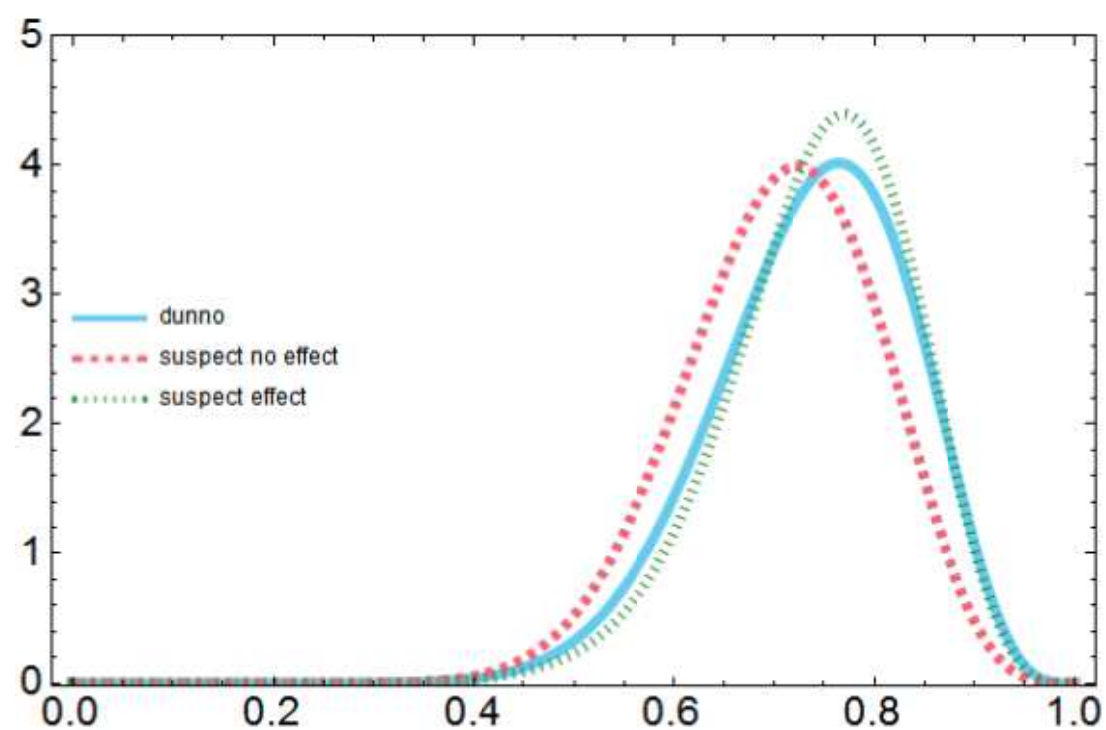
$$P(H_f \mid D \text{ \& } A)$$



$$P(H_f | A)$$



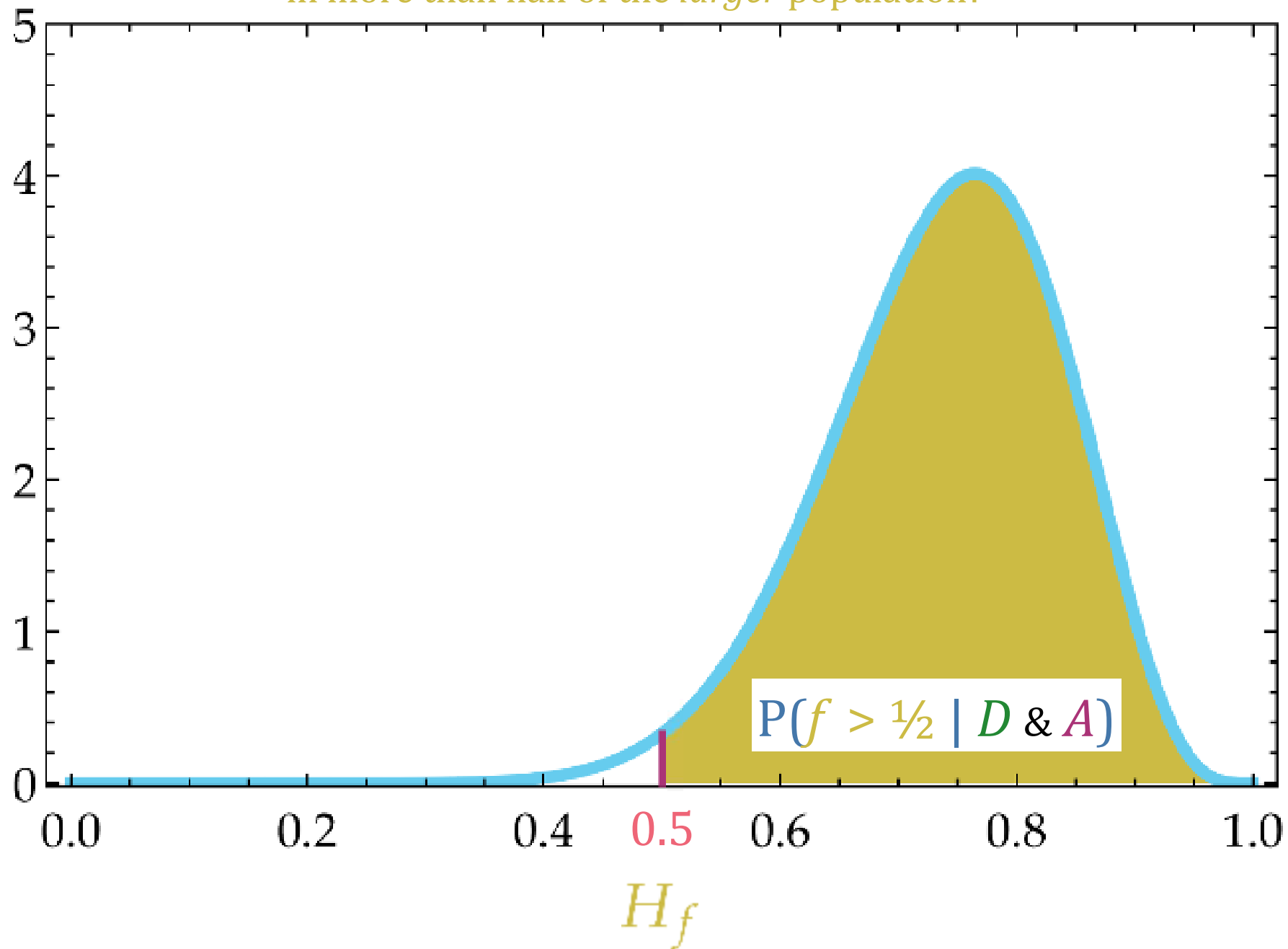
$$P(H_f | D \text{ \& } A)$$





What is the probability that the drug leads to "cognitive+"
in more than half of the *larger* population?

What is the probability that the drug leads to "cognitive+"
in more than half of the *larger* population?



What is the probability that the drug leads to "cognitive+"
in more than half of the *larger* population?

‘dunno’:

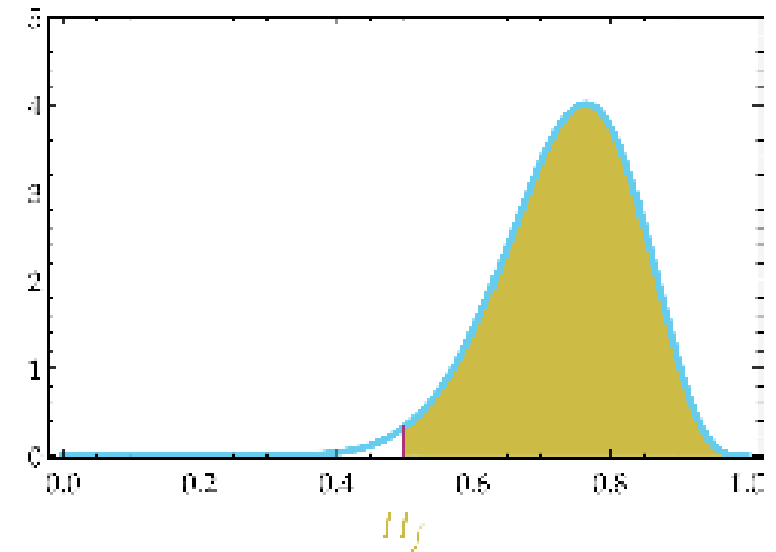
$$P(f > \tfrac{1}{2} \mid D \& A_{=}) = 98.5\%$$

‘suspect no effect’:

$$P(f > \tfrac{1}{2} \mid D \& A_{\text{no}}) = 97.6\%$$

‘suspect effect’:

$$P(f > \tfrac{1}{2} \mid D \& A_{\text{yes}}) = 98.9\%$$



All three scientists agree that almost surely there is *some* effect

They are more uncertain about *how strong* the effect is (as measured by f)





frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.



frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

(alright, but what does this mumbo-jumbo *concretely* mean?)



frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

Bayesian paper

“Given the data, the assumption that our sample is exchangeable in a much larger population, and an initial assumption of uniform ignorance about the future frequency of positive test results, we predict:

- With 98% credibility, more than 1/2 of future tests will respond positively to the drug.
- With 90% credibility, the average of future positive responses lies between 0.56 and 0.88.

frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

Bayesian paper

“Given the data, the assumption that our sample is exchangeable in a much larger population, and an initial assumption of uniform ignorance about the future frequency of positive test results, we predict

Assumptions behind the conclusions are plainly stated

- With 98% credibility, more than 1/2 of future tests will respond positively to the drug.
- With 90% credibility, the average of future positive responses lies between 0.56 and 0.88.

frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

Bayesian paper

“Given the data, the assumption that our sample is exchangeable in a much larger population, and an initial assumption of uniform ignorance about the future frequency of positive test results, we predict:

- With 98% credibility, more than 1/2 of future tests will respond positively to the drug.

The results are concrete predictions (no vague “significant” bullshit)

- With 90% credibility, the average of future positive responses lies between 0.56 and 0.88.

frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

Bayesian paper

“Given the data, the assumption that our sample is exchangeable in a much larger population, and an initial assumption of uniform ignorance about the future frequency of positive test results, we predict:

- With **98% credibility, more than 1/2 of future tests** will respond **positively** to the drug.
This says that we're almost certain that there is some positive effect
- With 90% credibility, the average of future positive responses lies between 0.56 and 0.88.

frequentist paper

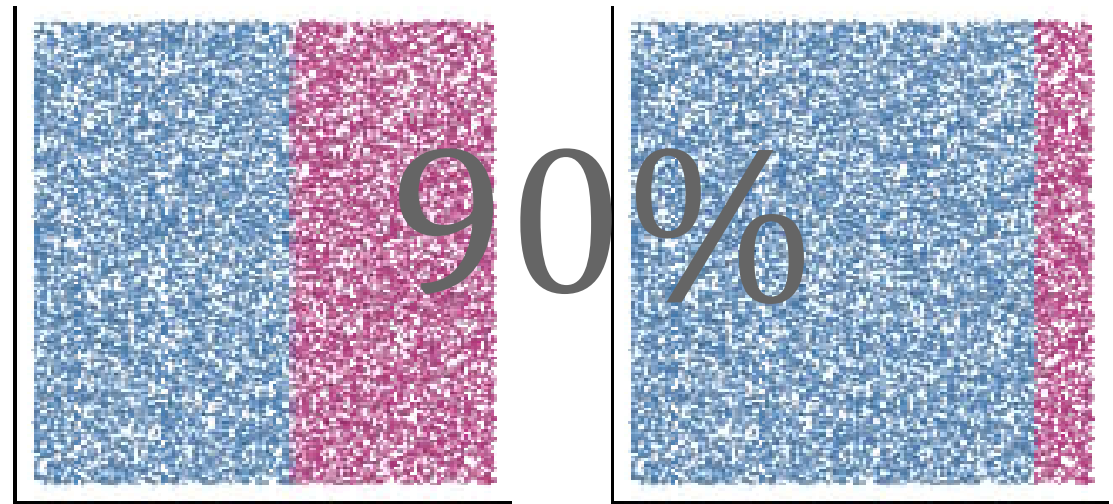
“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

Bayesian paper

“Given the data, the assumption that our sample is exchangeable in a much larger population, and an initial assumption of uniform ignorance about the future frequency of positive test results, we predict:

- With 98% credibility, more than 1/2 of future tests will respond positively to the drug.
- With **90% credibility**, the **average** of future positive responses lies **between 0.56 and 0.88**.

This reports our predictions about the population percentage of the effect





frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

Bayesian paper

“Given the data, the assumption that our sample is exchangeable in a much larger population, and an initial assumption of uniform ignorance about the future frequency of positive test results, we predict:

- With 98% credibility, more than 1/2 of future tests will respond positively to the drug.
- With 90% credibility, the average of future positive responses lies between 0.56 and 0.88.

frequentist paper

“According to a two-tailed test, the data are significant ($p = 0.049$) at the 0.05 level.

Bayesian paper

“Given the data, the assumption that our sample is exchangeable in a much larger population, and an initial assumption of uniform ignorance about the future frequency of positive test results, we predict:

- With 98% credibility, more than 1/2 of future tests will respond positively to the drug.
- With 90% credibility, the average of future positive responses lies between 0.56 and 0.88.

The paper could also add the results from different assumptions:

A strongly sceptical pre-data probability leads to:

- 92%: more than 1/2 of future tests will be positive
- 90%: average will be between 0.48 and 0.74

First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole

The Event Horizon Telescope Collaboration

(See the end matter for the full list of authors.)

Received 2019 March 1; revised 2019 March 12; accepted 2019 March 12; published 2019 April 10

We used two distinct Bayesian-inference algorithms and demonstrate that such crescent models are statistically preferred over other comparably complex geometric models that we have explored.

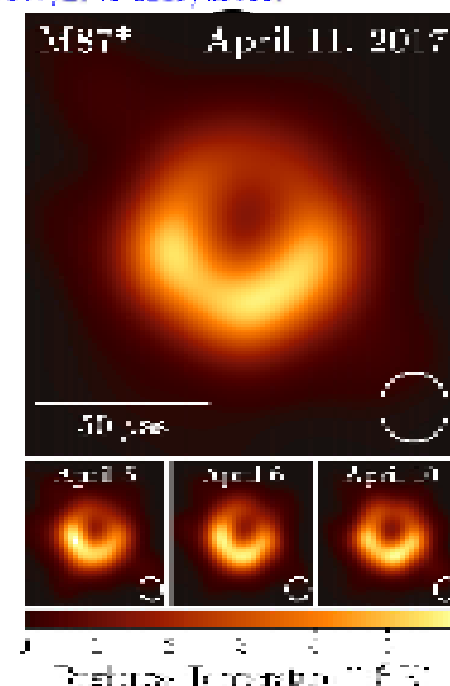
Our quantitative modeling approach seeks to estimate the posterior distribution $P(\Theta|\mathcal{D})$ of some parameters Θ within the context of a model and conditioned on some data $\mathcal{D}$,

$$P(\Theta|\mathcal{D}) = \frac{P(\mathcal{D}|\Theta)P(\Theta)}{P(\mathcal{D})} \equiv \frac{\mathcal{L}(\Theta)\pi(\Theta)}{\mathcal{Z}}. \quad (7)$$

Here, $\mathcal{L}(\Theta) = P(\mathcal{D}|\Theta)$ is the likelihood of the data given the model parameters, $\pi(\Theta) = P(\Theta)$ is the prior probability of the model parameters, and

$$\mathcal{Z} = P(\mathcal{D}) = \int \mathcal{L}(\Theta)\pi(\Theta)d\Theta \quad (8)$$

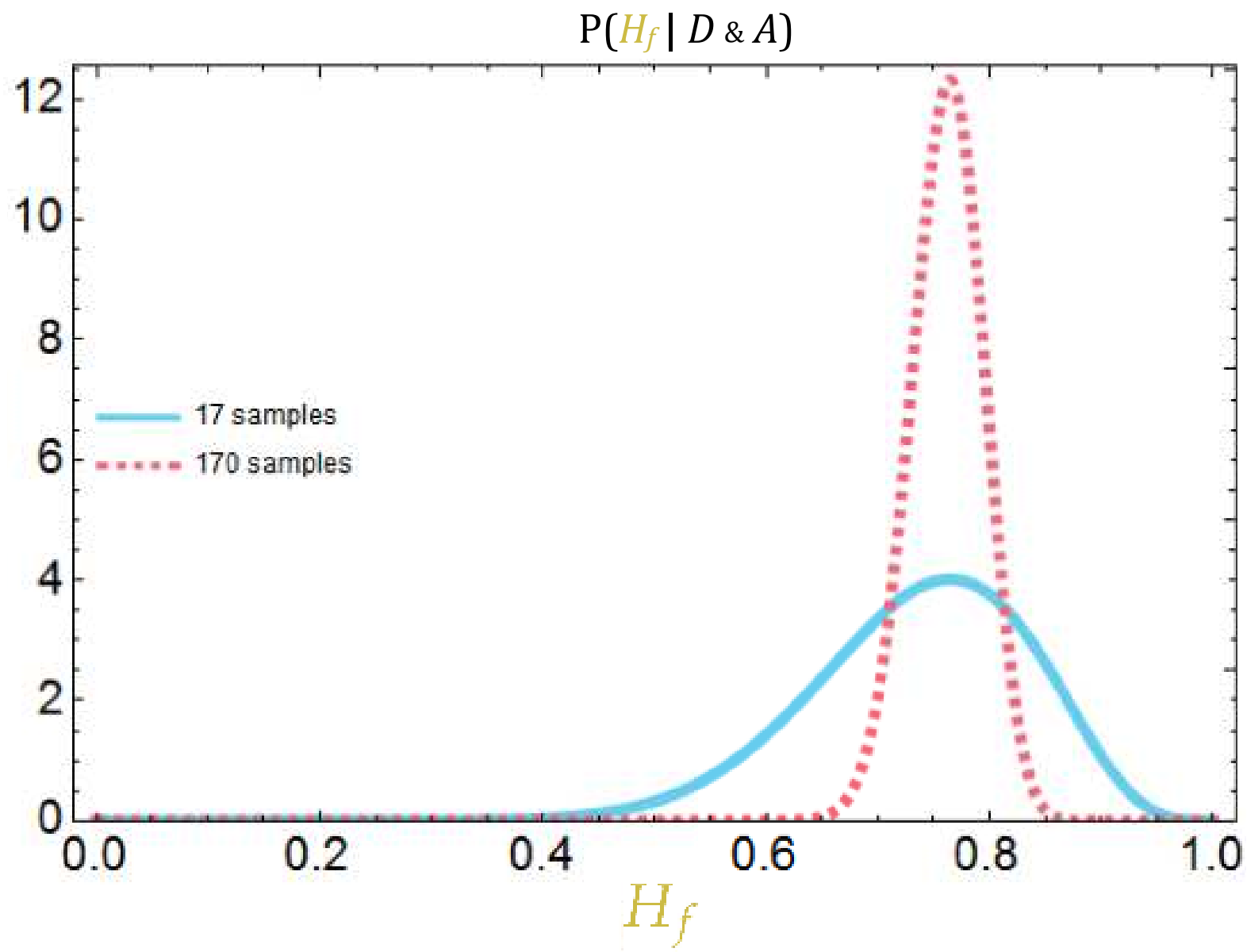
is the Bayesian evidence,





How many samples, if we want a pre-established credibility?

How many samples, if we want a pre-established credibility?





```
clear
```

```
%% Data:
```

```
positive = 13;
```

```
negative = 4;
```

```
%% Parameters for pre-data distribution (mean and standard deviation):
```

```
mean = 0.5;
```

```
sd = 0.2;
```

```
betaShape1 = ((1 - mean) * mean/sd^2 - 1) * mean; % shape-parameters of beta distribution
```

```
betaShape2 = betaShape1 * (1 - mean)/mean;
```

```
%% Pre-data distribution (represented by a beta distribution, https://mathworld.wolfram.com/BetaDistribution.html):
```

```
predata = @(f) betapdf(f, betaShape1, betaShape2);
```

```
%% Final distribution, numerator and denominator of Bayes's formula:
```

```
numerator = @(f) nchoosek(positive+negative, positive) .* f.^positive .* (1-f).^negative .* predata(f);
```

```
denominator = integral(numerator, 0, 1); % integral approximates sum
```

```
%% Plot the two distributions:
```

```
fgrid = 0:(1/1000):1; % create a grid of f-coordinates
```

```
plot(fgrid, numerator(fgrid)/denominator);
```

```
hold on
```

```
plot(fgrid, predata(fgrid), '--');
```

```
hold off
```

```
grid on
```

```
set(gca, 'XAxisLocation', 'origin');
```

```
set(gca, 'YAxisLocation', 'origin');
```

```
xlabel('f');
```

```
ylabel('probability');
```

```
legend('given data', 'initial assumption', 'Location', 'northwest')
```

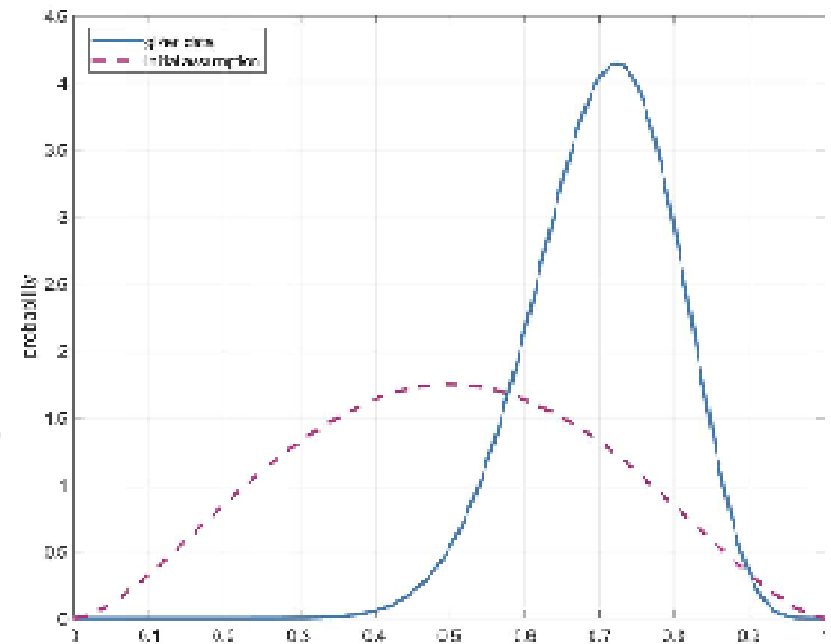
```
%% Print probability for  $f > 0.5$ , given the data:
```

```
disp('probability for  $f > 0.5$ :');
```

```
disp(integral(numerator, 0.5, 1)/denominator);
```

```
% gives 0.9758658
```

*Example Matlab script to calculate the post-data distributions
and to output the final probability that $f > 0.5$*



```
library('ggplot2')
```

```
## Data:  
positive <- 13  
negative <- 4
```

```
## Parameters for pre-data distribution (mean and standard deviation):  
mean <- 0.5  
sd <- 0.2
```

```
betaShape1 <- ((1 - mean) * mean/sd^2 - 1) * mean # shape-parameters of beta distribution  
betaShape2 <- betaShape1 * (1 - mean)/mean
```

```
## Pre-data distribution (represented by a beta distribution, https://mathworld.wolfram.com/BetaDistribution.html):  
predata <- function(f) dbeta(f, betaShape1, betaShape2)
```

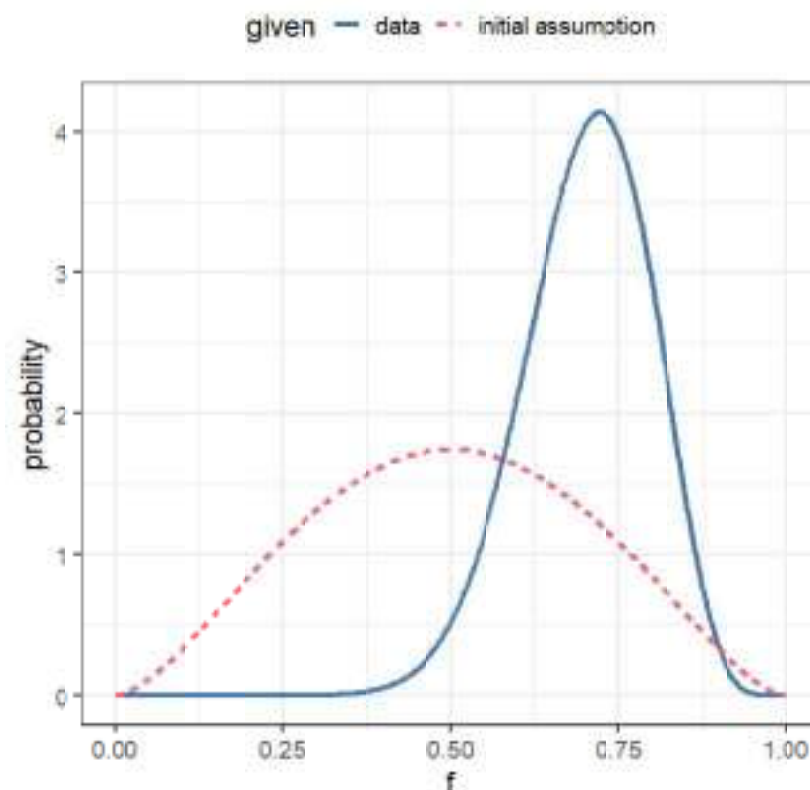
```
## Final distribution, numerator and denominator of Bayes's formula:  
numerator <- function(f) choose(positive+negative, negative) * f^positive * (1-f)^negative * predata(f)  
denominator <- integrate(numerator, 0, 1)$value # integral approximates sum
```

```
## Plot the two distributions:  
fgrid <- seq(0, 1, length.out=1000) # create a grid of f-coordinates  
toPlot <- rbind(data.table(f=fgrid,  
                           probability=predata(fgrid),  
                           given='initial assumption'),  
               data.table(f=fgrid,  
                           probability=numerator(fgrid)/denominator,  
                           given='data'))
```

```
qplot(x=f, y=probability, data=toPlot, geom='line',  
       color=given, lty=given, lwd=l(1.5)) + theme(legend.pos='top')
```

```
## Print probability for  $f > 0.5$ , given the data:  
print('probability for  $f > 0.5$ :')  
print(integrate(numerator, 0.5, 1)$value/denominator)  
# gives 0.9758658
```

*Example R script to calculate the post-data distributions
and to output the final probability that $f > 0.5$*





What are our hypotheses? What are our question & purpose?

Does the drug enhance cognitive abilities?

→ *Yes, No, Sometimes, It depends, ...*

What are our hypotheses? What are our question & purpose?

Does the drug enhance cognitive abilities?

→ *Yes, No, Sometimes, It depends, ...*

$P(\text{'In 17 tests, 13 cognitive+'} \mid \text{'The drug enhances cognitive abilities' \& } A)$

What are our hypotheses? What are our question & purpose?

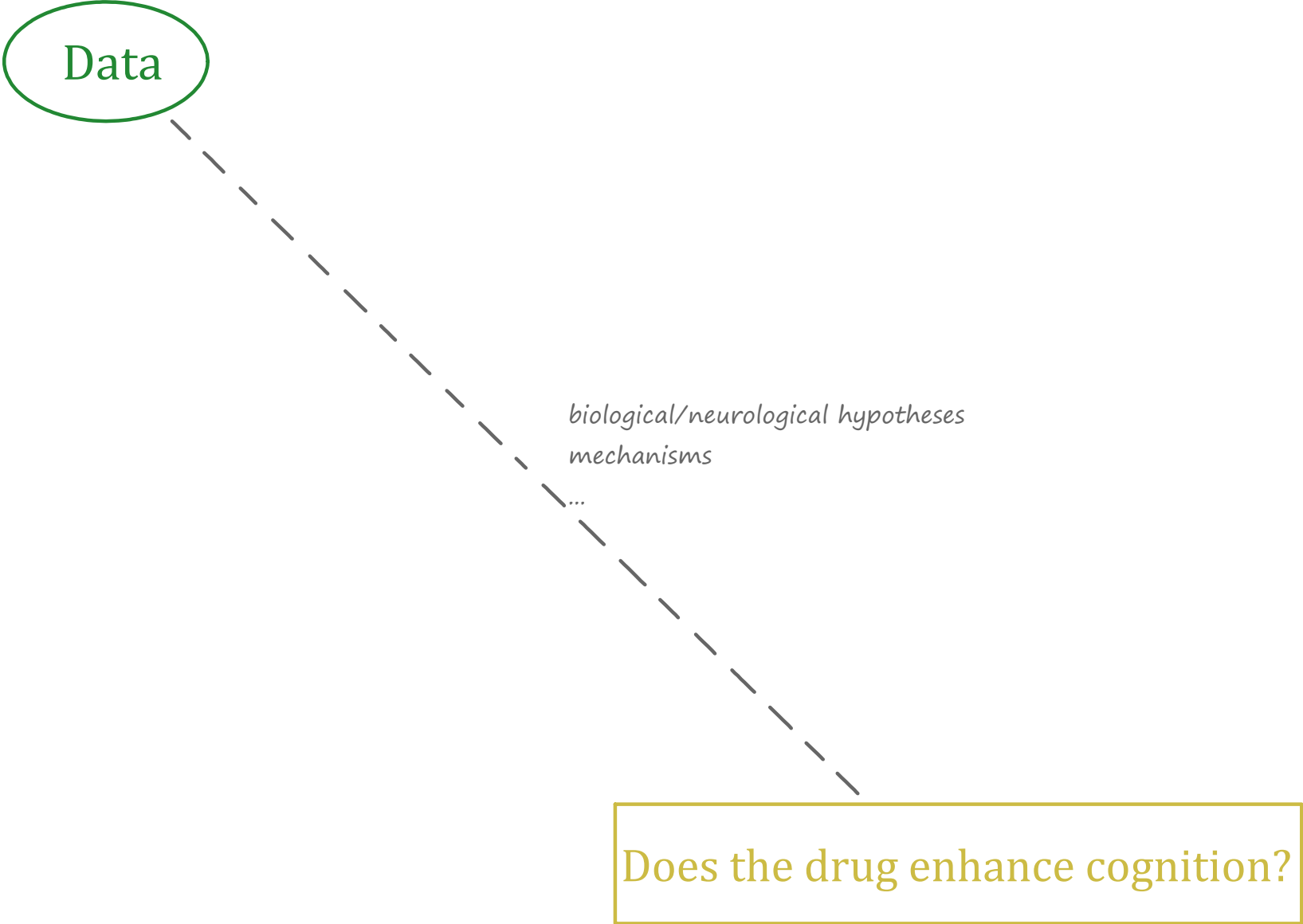
Does the drug enhance cognitive abilities?

→ *Yes, No, Sometimes, It depends, ...*

What do we need to assume?

$P(\text{'In 17 tests, 13 cognitive+'} \mid \text{'The drug enhances cognitive abilities' \& } A)$





Data

*biological/neurological hypotheses
mechanisms*

...

Does the drug enhance cognition?

```
graph TD; Data([Data]) -- solid line --> Q1[How systematic is the effect on rats in general?]; Q1 -. "biological/neurological hypotheses<br/>mechanisms<br/>..." .-> Q2[Does the drug enhance cognition?];
```

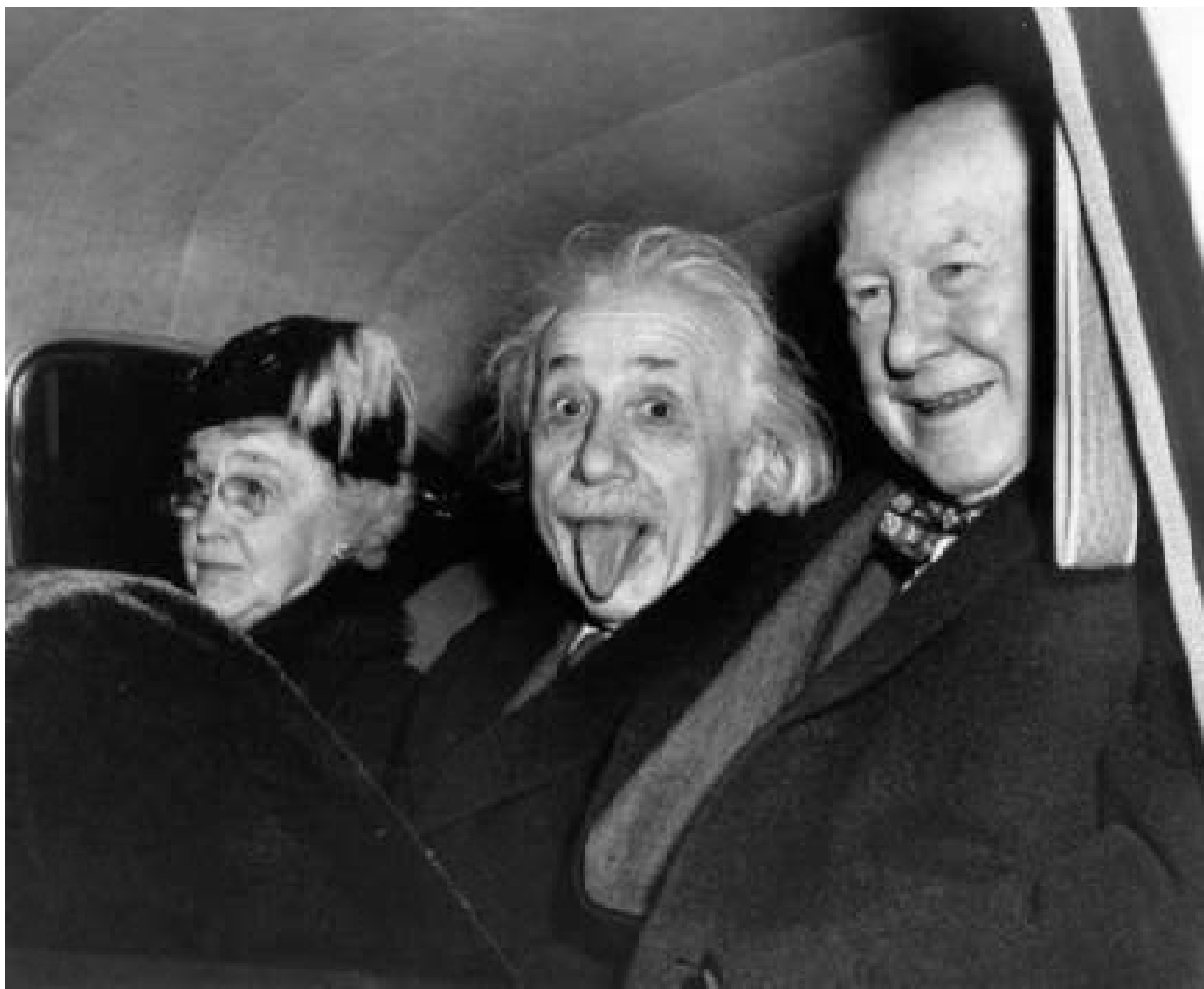
Data

How systematic is the effect on rats in general?

*biological/neurological hypotheses
mechanisms*

...

Does the drug enhance cognition?



Thank you!