BASS XXIX

Charlotte, NC

24 October 2022

The Epistemological Superiority of Bayesian Inference over Frequentist Inference

Inferring What is Likely To Be True

Stephen J. Ruberg, PhD
President
Analytix Thinking, LLC
AnalytixThinking@gmail.com
AnalytixThinking.blog

Bringing data to life.

Background

Perspective

My journey into Bayesian thinking

- Never took a Bayesian class
- Never did a Bayesian analysis

Objectives

Tell stories

Give examples

Help with communication/teaching

Even a few new ideas

- Exploratory vs confirmatory
- Pr(false positive finding)

Epistemology

Bringing data to life.

Part 1

The First
Story on
My Journey

Bringing data to life.

Thought Experiment

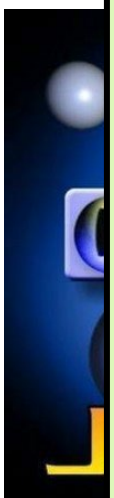
$300,000,000

IF I WIN

10% Probability

90% Probability

Thought Experiment



$

How likely are you to receive a share of the winnings?

A. 90% likely
B. 50% likely
C. Small likelihood
D. Very, very small likelihood

Conditional Probability

Example of Conditional Probability

- The key word is **IF**
- Very low probability of winning (odds: **1:292,301,338**)

Solution:

Unconditional

Conditional

Pr (**you receive a share**)

= Pr (**I choose to share IF I win**) * Pr (**I win**)

= **.90** * **0.0000000034211**

= **0.00000000307901**

Most decisions are made using **unconditional** probabilities.

Bringing data to life.

Conditional Probability

Power = Pr(reject H_0 | δ ≥ d)

- What is the Pr(δ ≥ d)? Conditional

Unconditional probability to reject H_0

- Pr(reject H_0 | δ ≥ d) * pr(δ ≥ d).

Power **pdf for δ**

$$\int_{-\infty}^{\infty} \Pr(\text{reject } H_0 \mid \delta) * \text{pdf}(\delta)\, d\delta$$

Bringing data to life.

Conditional Probability

$$\int_{-\infty}^{\infty} \Pr(\text{reject } H_0 \mid \delta) * \underbrace{\text{pdf}(\delta)}_{\text{“PRIOR”}} \, d\delta$$

Assurance

Pr(study success)

Average power

“Bayesian Power”

What about other Bayesian concepts?

Bringing data to life.

Part 2

My Thought Experiment on My Journey

Bringing data to life.

Another Thought Experiment

10,000 Coins

9,999 Fair Coins (H/T)

1 Biased Coin (H/H)

Problem

1. I draw out one coin.
2. I will flip it repeatedly and tell you the result.
3. You tell me when you decide whether I have the Biased Coin or not.

The Bet

Number of Flips	Result	Biased Coin?
1	H	Y or N
2	H	Y or N
3	H	Y or N
4	H	Y or N
5	H	Y or N
6	H	Y or N
7	H	Y or N
8	H	Y or N
9	H	Y or N
10	H	Y or N

Number of Flips	Result	Biased Coin?
11	H	Y or N
12	H	Y or N
13	H	Y or N
14	H	Y or N
15	H	Y or N
16	H	Y or N
17	H	Y or N
18	H	Y or N
19	H	Y or N
20	H	Y or N

A Problem of Inference

Decision Rule: See N consecutive H's

H_0: Coin is fair
pr(heads) = 0.50

H_a: Coin is
pr(h

pr[N consecutiv
(0.50)^N

N is v
your risk)

D
ce level of the test

Reject H_0 with a p-value < 0.001 !

Highly likely I have selected the biased coin !

–10 consecutive H's are observed

ecutive heads | fair coin] = $(0.50)^{10}$ = 0.0009766

Defines the *significance level of the data or the p-value*

Bringing data to life.

A Problem of Inference

NHST* $\cong$ proof by contradiction

We want H_a to be true**

or

We want to evaluate

pr(H_a is true | observed data) ≡

pr(H_0 is false | observed data)

*Null Hypothesis Significance Testing

** Except in equivalence testing

Bringing data to life.

A Problem of Inference

Question of Interest

How many consecutive H's are needed ***to bet*** that I selected the Biased Coin?

What is the pr(I pulled the biased coin)?

or

When is pr(biased coin | **n**) > 0.50?

Bringing data to life.

A Problem of Inference

pr(bias coin selected | n consecutive heads observed) = p

Bayes Theo

[as formul

$pr(H_a | n) = pr(n | H_a)$ ... + $pr(n | H_0)\ p(H_0)$]

= 1 ... ,000) + (.5^{n} * (9999/10,000)]

Highly **unlikely** I have selected the biased coin !

Don't make the bet !

d coin | **10** consecutive heads] = 0.093

Bringing data to life.

A Problem of Inference

How did we get into this mess?

Bringing data to life.

Two Perspectives

1. What is the probability of seeing N consecutive heads **IF** I have a fair coin?

   Frequentist Approach

2. What is the probability that I selected the biased coin **IF** I observe N consecutive heads … [from a coin randomly drawn from a bag of 9,999 fair coins and 1 biased coin]?

   Bayesian Approach

Frequentists Results

Number of Flips	Result	p-value
1	H	0.500000000
2	H	0.250000000
3	H	0.125000000
4	H	0.062500000
5	H	0.031250000
6	H	0.015625000
7	H	0.007812500
8	H	0.003906250
9	H	0.001953125
10	H	0.000976563

Number of Flips	Result	p-value
11	H	0.000488281
12	H	0.000244141
13	H	0.000122070
14	H	0.000061035
15	H	0.000030518
16	H	0.000015259
17	H	0.000007629
18	H	0.000003815
19	H	0.000001907
20	H	0.000000954

Frequentists Results

Pr (13 consecutive H's with a fair coin) = 0.000122070

≅ 1.2/10,000

More Likely ↑

Pr (Pull the 1 biased coin from the bag) = 1/10,000

Less Likely ↓

Pr (14 consecutive H's with a fair coin) = 0.000061035

≅ 0.6/10,000

Bringing data to life.

Frequentists Results

P-value is **conditional** on H_0 being true.

P-value = Pr(reject H_0 | H_0 is true)

<u>Recall the Lottery Example</u>

Pr (**you receive a share**)

= Pr (**I choose to share IF** I win) * Pr (**I win**)

What's Pr(H_0 is true)? 9,999/10,000

More on this later !!

Bringing data to life.

Two Perspectives

2. Pr (coin is biased | observed data)

If we have $P(A|B)$,

we want to obtain the conditional probability $P(B|A)$

Bayes Theorem (1763)*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

*As formulated by Laplace (1812)

Bayesian Results

Number of Flips	Result	Pr(Biased Coin)
1	H	0.000200

Some Observations

We started with 1/10,000 chance of pulling the biased coin.

With one small piece of evidence (i.e. a single H), we have a little greater probability that I have pulled the biased coin (i.e. 2/10,000).

Bayesian Results

Number of Flips	Result	Pr(Biased Coin)
1	H	0.000200
2	H	0.000400
3	H	0.000799
4	H	0.001598
5	H	0.003190
6	H	0.006360
7	H	0.012639
8	H	0.024968
9	H	0.048711
10	H	0.092897

Number of Flips	Result	Pr(Biased Coin)
11	H	0.170001
12	H	0.290600
13	H	0.450333
14	H	0.621006
15	H	0.766198
16	H	0.867624
17	H	0.929121
18	H	0.963258
19	H	0.981285
20	H	0.990554

Bayesian Results

Pr (Biased coin | 13 consecutive H’s) = **0.450333**

≅ 45%

Less Likely to Win Bet

Even odds for the bet = Pr(biased coin) = 50%

More Likely to Win Bet

Pr (Biased coin | 14 consecutive H’s) = **0.621006**

≅ 62%

Bringing data to life.

A Problem of Inference

100 Coins

99 Fair Coins (H/T)
1 Biased Coin (H/H)

Problem

1. **I draw out one coin.**
2. **I will flip it repeatedly and tell you the result.**
3. **You tell me when you decide whether I have the Biased Coin or not.**

The Results

Number of Flips	Result	Biased Coin?
1	H	
2	H	
3	H	
4	H	
5	H	
6	H	
7	H	
8	H	
9	H	
10	H	

Number of Flips	Result	Biased Coin?
11	H	
12	H	
13	H	
14	H	
15	H	
16	H	
17	H	
18	H	
19	H	
20	H	

The Results

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
1	0.000200	0.019802
2	0.000400	
3	0.000799	
4	0.001598	
5	0.003190	
6	0.006360	
7	0.012639	
8	0.024963	
9	0.048711	
10	0.092897	

~2/100

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
11	0.170001	
12	0.290600	
13	0.450333	
14	0.621006	
15	0.766198	
16	0.867624	
17	0.929121	
18	0.963258	
19	0.981285	
20	0.990554	

The Results

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
1	0.000200	0.019802
2	0.000400	0.038835
3	0.000799	0.074766
4	0.001598	0.139130
5	0.003190	0.244275
6	0.006360	0.392638
7	0.012639	0.563877
8	0.024963	0.721127
9	0.048711	0.837971
10	0.092897	0.911843

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
11	0.170001	
12	0.290600	
13	0.450333	
14	0.621006	
15	0.766198	
16	0.867624	
17	0.929121	
18	0.963258	
19	0.981285	
20	0.990554	

The Results

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
1	0.000200	0.019802
2	0.000400	0.038835
3	0.000799	0.074766
4	0.001598	0.139130
5	0.003190	0.244275
6	0.006360	0.392638
7	0.012639	0.563877
8	0.024963	0.721127
9	0.048711	0.837971
10	0.092897	0.911843

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
11	0.170001	0.953889
12	0.290600	0.976400
13	0.450333	0.988059
14	0.621006	0.993994
15	0.766198	0.996988
16	0.867624	0.998492
17	0.929121	0.999245
18	0.963258	0.999622
19	0.981285	0.999811
20	0.990554	0.999906

The Results

# of Flips	p-value	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
1	0.500000	0.000200	0.019802
2	0.250000	0.000400	0.038835
3	0.125000	0.000799	0.074766
4	0.062500	0.001598	0.139130
5	0.031250	0.003190	0.244275
6	0.015625	0.006360	0.392638
7	0.0078125	0.012639	0.563877
8	0.0039063	0.024963	0.721127
9	0.0019531	0.048711	0.837971
10	0.0009766	0.092897	0.911843

# of Flips	p-value	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
11	0.0004882	0.170001	0.953889
12	0.0002441	0.290600	0.976400
13	0.0001220	0.450333	0.988059
14	0.0000610	0.621006	0.993994
15	0.0000305	0.766198	0.996988
16	0.0000153	0.867624	0.998492
17	0.0000076	0.929121	0.999245
18	0.0000038	0.963258	0.999622
19	0.0000019	0.981285	0.999811
20	0.0000010	0.990554	0.999906

Note: The p-value never changes regardless of your prior knowledge!!!!

VERY Important Lesson

For the SAME DATA

(i.e., evidence),

you arrive at

DIFFERENT CONCLUSIONS

(i.e., decisions)

based on your

PRIOR KNOWLEDGE!

Bringing data to life.

Coin in Bag Summary

Cannot interpret a p-value in isolation

Need to know prior belief about H_0 (or H_a)

Conditional probability

$$\text{p-value} = pr(\mathbf{T} > \mathbf{c} \mid H_0 \text{ is true})$$

Test Statistic

critical value

How likely is this?

Bringing data to life.

Coin in Bag Summary

Frequentist ⇨ pr(Data | H_0)

Bayesian ⇨ pr(H_0 | Data)

as different as

Pr(cloudy | rain)

Pr (rain | cloudy)

Bringing data to life.

A Problem of Inference

Traditionally, statisticians have been “selling”

Pr(data|hypothesis) [i.e., the p-value]

The first great “bait and switch” that statisticians have pulled on scientists.

to quantify the likelihood of a hypothesis !!!!!!

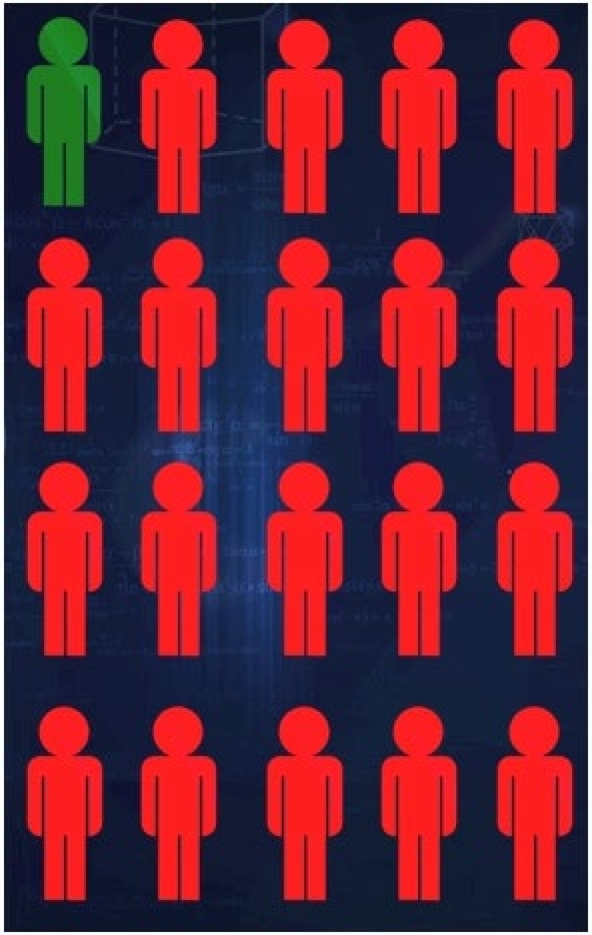
Bringing data to life.

Part 3

Another Story on My Journey

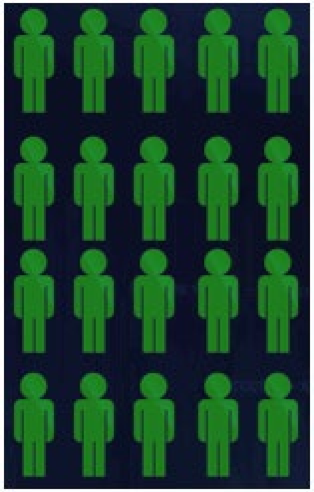
Bringing data to life.

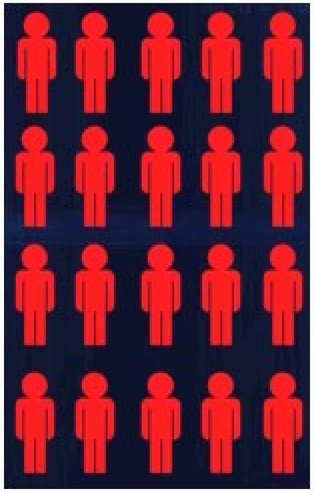
Another Thought Experiment



5% of Population have ALK gene

Patients





Diagnostic Test

95% Sensitivity

19 +'s
1 -

95% Specificity

1 +
19 -'s

Individual Patient

ALK(-)?

ALK(+)?

Diagnostic Test

Sample →

Result → **+**

Pr(Patient is ALK+) = ?

Diagnostic Decision-Making

Developing/Designing the "Assay"

Conditional Probability

		Patient Characteristic	
		Positive	**Negative**
Diagnostic Test	**Positive**	**True Positive 95% (Sensitivity)**	**False Positive 5%**
	Negative	**False Negative 5%**	**True Negative 95% (Specificity)**

Prob (diagnostic test is positive IF the patient has the characteristic)

Diagnostic Decision-Making

Interpreting an Observed Result

Conditional Probability

		Patient Characteristic (Unknown Truth)		
Observed ↓		**Positive**	**Negative**	
Diagnostic Test	**Positive**	True Positive 95%	False Positive 5%	Positive Predictive Value
	Negative	False Negative 5%	True Negative 95%	Negative Predictive Value

Prob (patient has the characteristic IF the diagnostic test is positive)

Diagnostic Decision-Making

Underlying Prevalence for ALK gene is 5%

		Have the ALK Gene		
		Positive (5%)	Negative (95%)	
Diagnostic Test	Positive	95	95	50%
	Negative	5	1805	99.7%
		100	1900	2000

$$\frac{TP}{TP + FP}$$

$$\frac{TN}{TN + FN}$$

Diagnostic Decision-Making

With a great diagnostic test, but a low prevalence,

There is a 50/50 chance you have the ALK gene!

But wait … what if we re-test?

Think of all the false positives with COVID

Think of diagnostic testing

- X-ray ⇨ CT Scan ⇨ Needle Biopsy
- Each step – more expensive, time-consuming, invasive
- But, identifying higher prevalence population!

Diagnostic Decision-Making

With a great diagnostic test, but a low prevalence

There is a 50/50 you have the ALK gene!

But wait … what if we re-test?

Repeat the ALK test on all patients who tested positive

- Prevalence is now 50%
- Let's rework the diagnostic test 2x2 table

Diagnostic Decision-Making

Prevalence of ALK in patients who tested positive is 50%

		Have the ALK Gene		
		Positive (50%)	**Negative (50%)**	
Diagnostic Test	**Positive**	**950**	**50**	**95%**
	Negative	**50**	**950**	**95%**
		1000	**1000**	**2000**

Diagnostic Decision-Making

<u>KEY MESSAGES</u>

Sensitivity and Specificity are the focus of ***assay design and development***

- Sensitivity ≡ Power;   1-Specificity ≡ α

The Positive (Negative) Predictive Values are the focus of ***interpreting results*** (assay outputs)

- Everyone knows this
- PPV is what matters to physicians and patients

Diagnostic Decision-Making

KEY MESSAGES

PPV (NPV) is dependent on the underlying PREVALENCE of the characteristic of interest (e.g., disease/marker status)

PREVALENCE is the "prior."

PPV ≡ Bayes Formula (slides 16, 22) !!!

The Clinical Trial Analogy

The diagnostic test is the clinical trial

The patient characteristic is whether the treatment meets its Critical Success Factors (unknown truth)

Sensitivity and (1-Specificity) are analogous to “power” and “significance level” of the hypothesis test for the CT

The PPV (NPV) is “Bayesian posterior probability” that the treatment meets (fails) the CSF

THE PPV (NPV) ARE DEPENDENT ON THE PRIOR PROBABILITY OF THE TREATMENT MEETING THE CSF

The Clinical Trial Analogy

Entering Ph 2 ⇒ Pr(drug meets CSFs) = 20%

"Prior"

Unknown →

Moderately Rigorous Ph 2 Trial Design

		Meets CSFs: Yes (20%)	*Meets CSFs*: No (80%)	
CT Result	Positive	True Positive 80% **320**	False Positive 10% **160**	**66.7%** (Positive Predictive Value)
	Negative	False Negative **80**	True Negative **1440**	**94.7%** (Negative Predictive Value)
		400	**1600**	**2000**

"Posterior"

↑ Observed

Conclusion on Inference

If we all understand PPV is the proper metric for evaluating the likelihood of a (unknown) condition to be present/true using a diagnostic test …

and …

A clinical trial is a direct analogy to a diagnostic test …

then …

Why do we not routinely use the Bayesian Posterior Probability to interpret a clinical trial result ?!?!?!?!

We Should !!!

Bringing data to life.

Part 4

How do we get out of this mess?

Bringing data to life.

A Path Forward

Three Inferential Questions*

What does the data say?

- A p-value is a partial/poor answer.

What do I believe?

- This requires incorporation of prior information.

What do I decide?

- This requires a utility function.

*Royall, R. M. (1997), Statistical Evidence: A Likelihood Paradigm, volume 71 of Monographs on Statistics and Applied Probability. London: Chapman & Hall.

Bringing data to life.

A Path Forward

Question 1 – What do the data say?

A p-value is only part of the story.

Scientists (everyone!) wants this.

Frequentist

PRIOR KNOWLEDGE + **NEW EXPERIMENT** → **UPDATED BELIEF**

Prior Probability
H_0 is False

New Evidence
(e.g., p-value)

Posterior Probability
H_0 is False

Think: pr(I get money from the lottery).

A Path Forward

Question 2 – What do I believe?

Let p_0 be the prior probability that **H_0 is false**.

Let **p=p-value** from the test of H_0 from the current experiment.

The Bayes Factor Bound is

$$\mathbf{BFB=1/[-}e * p * \ln(p)] \qquad (p < 1/e).$$

The ***upper bound*** on the posterior probability that H_0 is false (p_1) given the observed data is

$$p_1 \leq \{1 + [(1-p_0)/p_0] \,/\, BFB\}^{-1}.$$

posterior prior data

Thomas Sellke, M. J Bayarri & James O Berger (2001) Calibration of p Values for Testing Precise Null Hypotheses, The American Statistician, 55:1, 62-71.

Bringing data to life.

Another Thought Experiment

Suppose there are **100** potential predictive biomarkers that could be important for a new treatment.

- **100** hypothesis tests, one for each biomarker

Observed p-value = **0.0001** for one biomarker test

- Bonferroni adjusted p-value ≤ **100** * **0.0001** = **0.01**

EUREKA! We have discovered a novel biomarker-defined subgroup.

Bringing data to life.

Another Thought Experiment

ARE YOU SURE?

Suppose further our prior belief is

pr(finding a predictive biomarker)

= pr(at least one H_0 is false) = **0.20**

Prior all H_0 are true (none are predictive) = 0.80

Uniform prior per biomarker = **0.20**/**100** = **0.002**

Bringing data to life.

Another Thought Experiment

ARE YOU SURE?

p_0 = 0.002 (uniform prior across 100 biomarkers)
p = 0.0001 (from hypothesis test)

Recall Bonferroni adjusted p = 0.01

$$p_1 \leq \{1 + [(1-p_0)/p_0] \times [-e \times p \times \ln(p)]\}^{-1}$$

Bayesian posterior pr(H_0 is false) ≤ 0.44.

Berger J.O., Wang X., Shen L. (2014). A Bayesian approach to subgroup identification. *J Biopharm Stat,* 24(1), 110-29.

Bringing data to life.

Real Examples

Dalcetrapib

AnalytixThinking.Blog: Genetic Subgroups and CV Disease

There are a variety of other Bayesian clinical trial topics covered in my blog (e.g., fluvoxamine for COVID-9).

Blog 19: We Won't Get Fooled Again, Again
Blog 20: I Am (Probably) Wrong, Maybe

AnalytixThinking.Blog

Bringing data to life.

A Path Forward

A p-value is ***literally*** only part of the story!

BAYESIAN INFERENCE

Frequentist

PRIOR KNOWLEDGE + **NEW EXPERIMENT** → **UPDATED BELIEF**

PRIOR KNOWLEDGE	NEW EXPERIMENT	UPDATED BELIEF
Prior Probability H_0 is False	New Evidence (e.g., p-value)	Posterior Probability H_0 is False

$$BFB = 1/[-e * p * \ln(p)]$$

$$p_1 \leq \{1 + [(1-p_0)/p_0] / BFB\}^{-1}.$$

Bringing data to life.

A Path Forward

"Always use *Bayesian thinking* when interpreting clinical trial results so you can quantify how believable the results are."

Steve Ruberg
Your Run-of-the-Mill Bayesian Statistician

Bringing data to life.

Part 5

What Is a P-value Worth Anyway?

Bringing data to life.

What's a P-Value Worth?

Further investigation of P-values

AnalytixThinking.Blog
No. 7: What does p< 0.05 mean anyway?

$$p_1 \leq \{1 + [(1-p_0)/p_0] / \text{BFB} \}^{-1}.$$

posterior — prior — data

What's a P-Value Worth?

Prior	p-value	Posterior (upper bound)
0.3	0.05	0.513

A p-value = 0.05 is not very strong evidence against the null hypothesis!

Bringing data to life.

What's a P-Value Worth?

Prior	p-value	Posterior (upper bound)
0.3	0.05	0.513
0.7	0.05	0.851

A p-value = 0.05 might be enough evidence against the null hypothesis.

What's a P-Value Worth?

Prior	p-value	Posterior (upper bound)
0.1	0.05	0.214
0.2	0.05	0.380
0.3	0.05	0.513
0.4	0.05	0.621
0.5	0.05	0.711
0.6	0.05	0.787
0.7	0.05	0.851

A p-value = 0.05 does not move the "evidentiary needle" very much!

Bringing data to life.

Part 6

A False Dichotomy*

Confirmatory vs Exploratory

*Ruberg, S. J. (2020) Détente: A Practical Understanding of P-values and Bayesian Posterior Probabilities. *Clin Pharm Ther.*, 109(6): 1489-1498. doi.org/10.1002/cpt.2004.

Bringing data to life.

A False Dichotomy

Confirmatory

- Prespecified, control Type 1 Error, etc. etc.

Exploratory

- Prespecified, but with less statistical rigor (e.g., without control of Type 1 Error)
- Unspecified, go where the data leads you

Bringing data to life.

A False Dichotomy

Statistically significant results

- Confirmatory – credible, believable
- Exploratory – interesting, but need more data/another trial
  - Some journals (e.g., NEJM) prohibit reporting p-values
    - Implies no inference is possible or reasonable!

Researchers will ALWAYS evaluate/interpret exploratory analyses

- Why not help quantify what to believe about the results of an “exploratory” analysis?

Bringing data to life.

A False Dichotomy

With a stated prior in place,
the terms “confirmatory” and “exploratory”
lose their meaning!

All the ingredients are here.

$$p_1 \leq \{1 + [(1-p_0)/p_0] \,/\, BFB\}^{-1}.$$

posterior   prior   data

Bringing data to life.

A False Dichotomy

Thought Experiment

Treatment successful in Phase 2

- Prior probability that it works for Phase 3 is 0.70

Treatment effect more pronounced in a subgroup??

- Literature; mechanism of action; biology of disease
- Prior for exceptional response in subgroup is 0.20
- Pre-specified, but no formal statistical analysis plan

Results of Ph 3 study

- Overall treatment effect p-value = 0.03
- Subgroup treatment effect p-value = 0.001 ?

Bringing data to life.

A False Dichotomy

Thought Experiment (cont'd)

TEST	PRIOR H_0 IS FALSE	PHASE 3 P-VALUE
ALL PATIENTS	0.70	0.030
SUBGROUP	0.20	0.001

The "exploratory" result is more convincing than the "confirmatory" result!

The "exploratory" result is the primary finding of the trial!

*Upper bound using $p_1 \leq \{1 + [(1-p_0)/p_0] \,/\, \text{BFB}\}^{-1}$.

Bringing data to life.

A False Dichotomy

Thought Experiment – Summary

Why debate confirmatory or exploratory?

- Whether it be a trial or a hypothesis within a trial

Assign each hypothesis of interest a prior probability

- We must know something (informative prior)*
- We have implicit priors

Lessen post hoc debate about “credible” or “spurious”

Quantify level of belief ⇨ Better decision-making

*Wacholder, S. et al. Assessing the probability that a positive report is false: An approach for molecular epidemiology studies. *J. Nat. Canc. Inst.* **96**, 434-442 (2004).

Bringing data to life.

Part 7

Probability of a False Positive Finding

Bringing data to life.

Pr (False Positive)

P-value is **conditional** on H_0 being true.

P-value = Pr(reject H_0 | H_0 is true)

<u>Recall the Lottery Example</u>

Pr (**you receive a share**)

= Pr (**I choose to share IF** I win) * Pr (**I win**)

What's Pr(H_0 is true)? **9,999/10,000**

With this prior for H_0, a whole lot of evidence is needed to reject it (i.e., 14 consecutive Heads!!)

Bringing data to life.

Pr (False Positive)

$$Pr(\text{reject } H_0 \mid H_0 \text{ is true})$$

Designing experiment = significance level

- α-level, Type 1 Error, the size of the test

After data is collected = significance level

- Smallest p-value for which we would have rejected the null hypothesis

Bringing data to life.

Pr (False Positive)

P-value as evidence (Fisher, 1925, 1926)

- "The value for which p=0.05 … is to be considered significant or not."

P-value as decision-maker (Neyman-Pearson, 1933)

- Balance Type 1 and Type 2 errors using sample size

P-value as both (Lehman, 1986, p. 70).

- "It is then good practice to determine not only whether the hypothesis is accepted or rejected at the given significance level, but also to determine the smallest significance level $\hat{a} = \hat{a}(x)$, the significance probability or p-value, at which the hypothesis would be rejected for the given observation."

Bringing data to life.

A Problem of Inference

DESIGN

HYPOTHESIS, EXPERIMENT

TEST STAT

TEST SIZE

REJECTION REGION

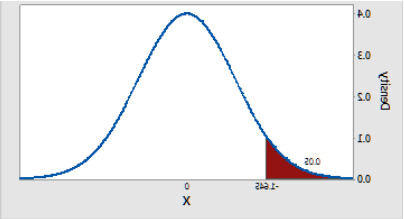
? ? ?

DATA

ANALYSIS

RUN MODEL

COMPUTE TEST STAT



P-VALUE

A 1-1 and onto mapping via an inverse probability function.*

*Kuffner, T. A., Walker, S. G. (2019) "Why are p-Values Controversial?", The American Statistician, 73(1), 1-3.

Pr (False Positive)

Conflating

the significance level

of the test (α)

with

the significance level

of the data (p-value)

The "silent hybrid solution" (Gigerenzer, 1989).

Pr (False Positive)

Philosophical Question

Design and experiment and accompanying suitable statistical test with a significance level of $\alpha=0.05$.

Conduct the experiment and observe $p=0.01$.

Reject the null hypothesis - "a positive finding"

What is the probability that this is a false positive finding?

Bringing data to life.

Pr (False Positive)

Pr(false positive finding) =

Pr(H_0 is true | p=0.01) =

1 – Pr(H_0 is false | p=0.01)

This is the REAL question of interest!

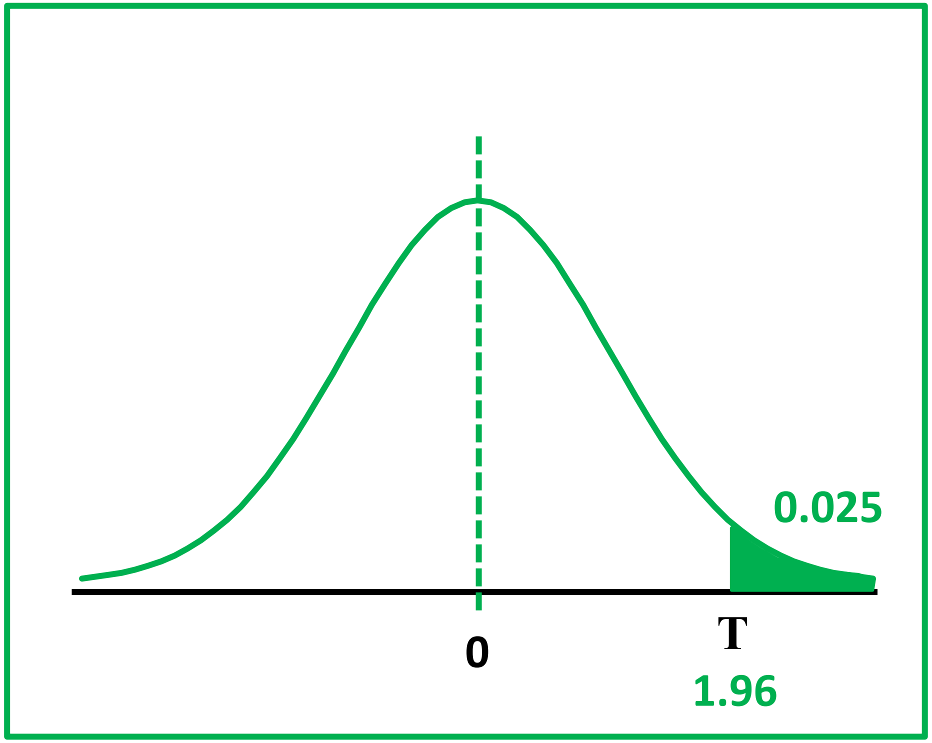
This is decidedly a Bayesian formulation.

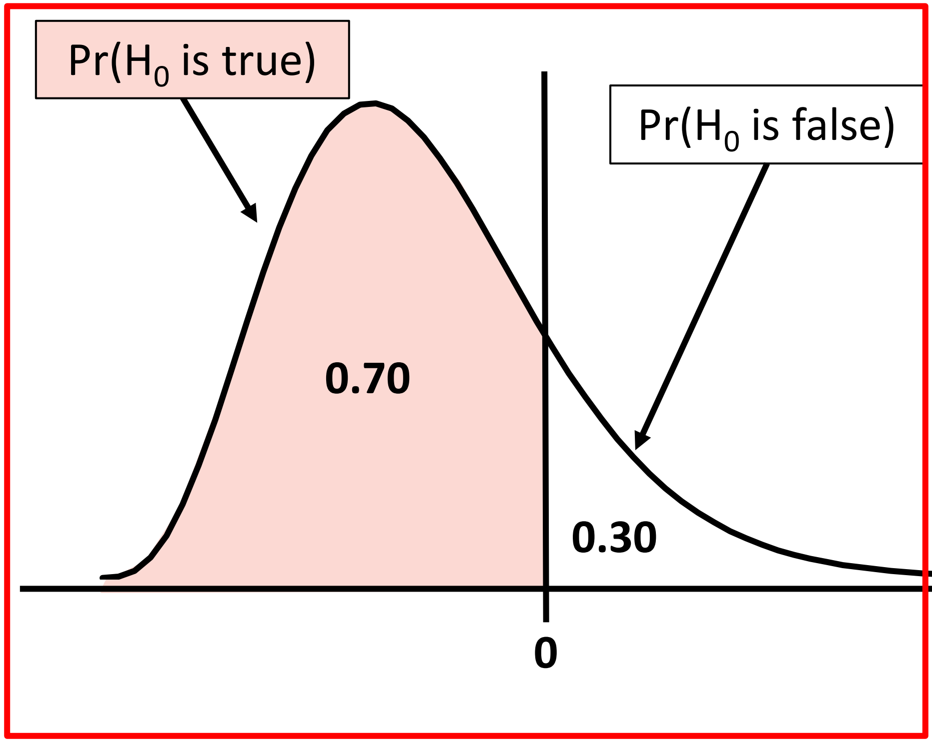
1 – Pr(**H_0 is false** | **p=0.01**)

hypothesis **data**
(test statistic)

Bringing data to life.

Pr (False Positive)





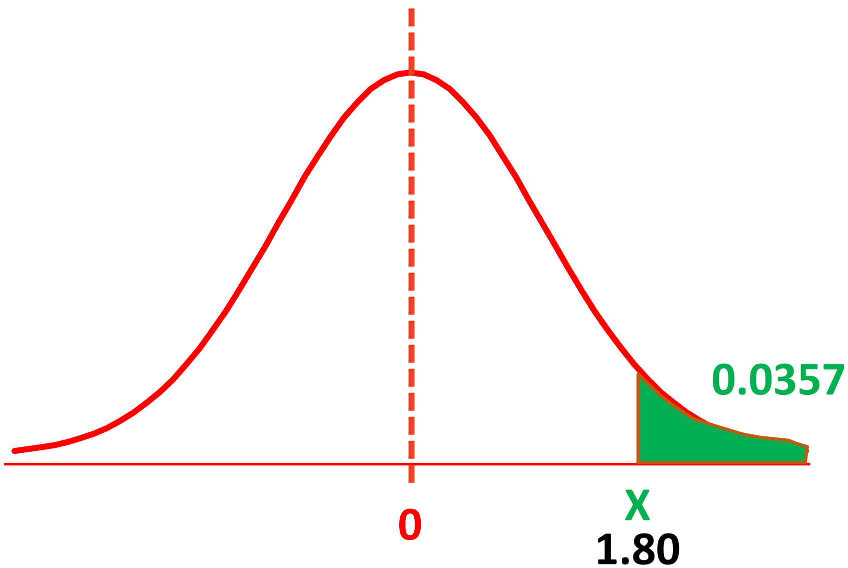
Pr(false positive finding) = Pr(Reject H_0 | H_0 is true) * Pr(H_0 is true)

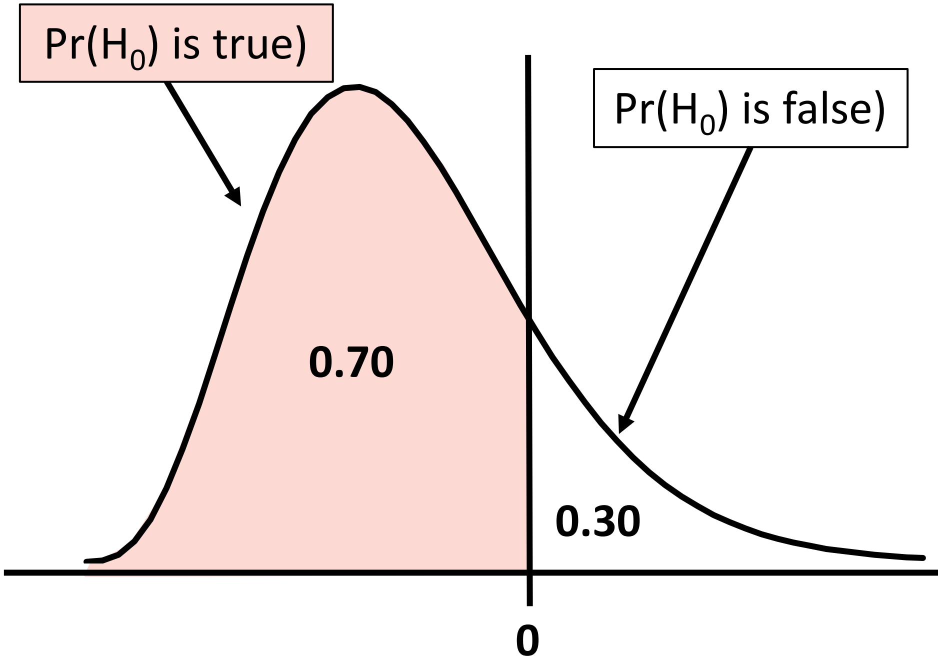
= 0.025 * 0.70

= 0.0175

Bringing data to life.

Pr (False Positive)





Pr(false positive finding) = Pr(Reject H_0 | H_0 is true) * Pr(H_0 is true)

= X * 0.70

= 0.025

Bringing data to life.

Pr (False Positive)

WOW!

Difficult to reconcile Frequentist approach and Bayesian approach.

- e.g., “frequentist properties of Bayesian methods”

Frequentist: pr(H_0 is true) = 1.

Bayesian: pr(H_0 is true) < 1.

Bringing data to life.

Part 8

Epistemology

How do we ~~know~~ ^believe what we ^believe ~~know~~?

Statistics is the science of discerning what is likely to be true.

Bringing data to life.

Epistemology

e·pis·te·mol·o·gy

/əˌpistəˈmäləjē/

noun **PHILOSOPHY**

the theory of knowledge, especially with regard to its methods, validity, and scope. Epistemology is the investigation of what distinguishes justified belief from opinion.

What is Probability?

Gerolamo Cardano

1501-1576

Pierre de Fermat

1607-1665

Blaise Pascal

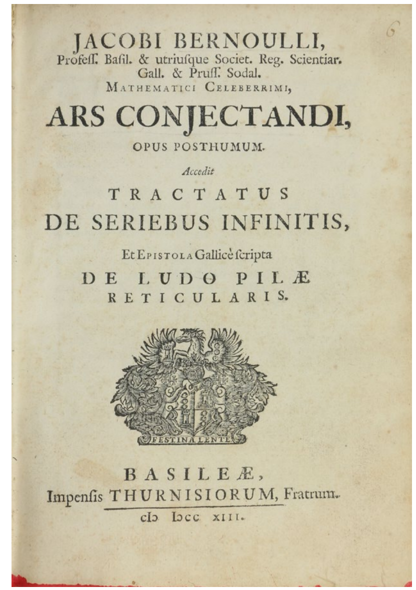
1623-1662

Christiaan Huygens

1629-1695

1713

(based on work from 1684-1689)



JACOBI BERNOULLI,
Profeff. Bafil. & utriufque Societ. Reg. Scientiar. Gall. & Pruff. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè fcripta
DE LUDO PILÆ RETICULARIS.
FESTINA LENTE
BASILEÆ,
Impenfis THURNISIORUM, Fratrum.
cIↄ Iↄcc XIII.

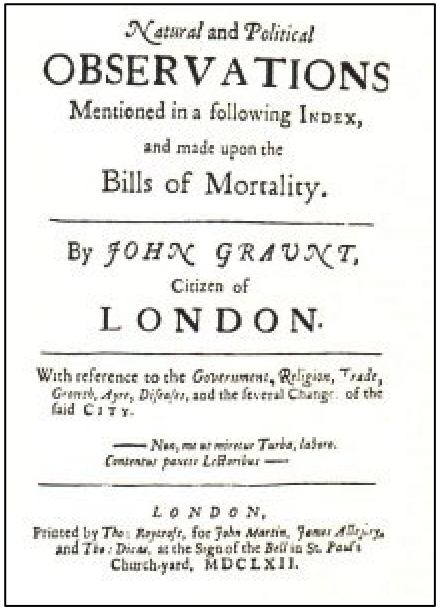
Jacob Bernoulli
1654-1705

Probability as **frequency** of events occurring

Bringing data to life.

What is Probability?

1662



Natural and Political
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

By JOHN GRAUNT,
Citizen of
LONDON.

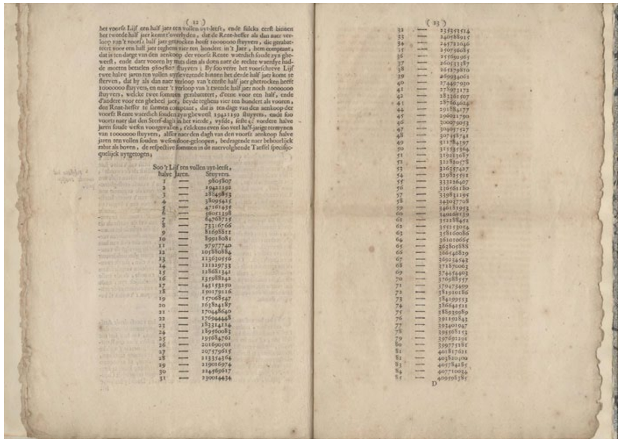
With reference to the Government, Religion, Trade, Growth, Ayre, Diseases, and the several Changes of the said CITY.

—— Non, me ut miretur Turba, laboro,
Contentus paucis Lectoribus ——

LONDON,
Printed by Tho: Roycroft, for John Martin, James Allestry, and Tho: Dicas, at the Sign of the Bell in St. Paul's Church-yard, MDCLXII.

1671

A Treatise on Life Annuities



Probability as a **concept**
(e.g., probability of dying at age X)

More than combinatorics

John Graunt

1620-1674

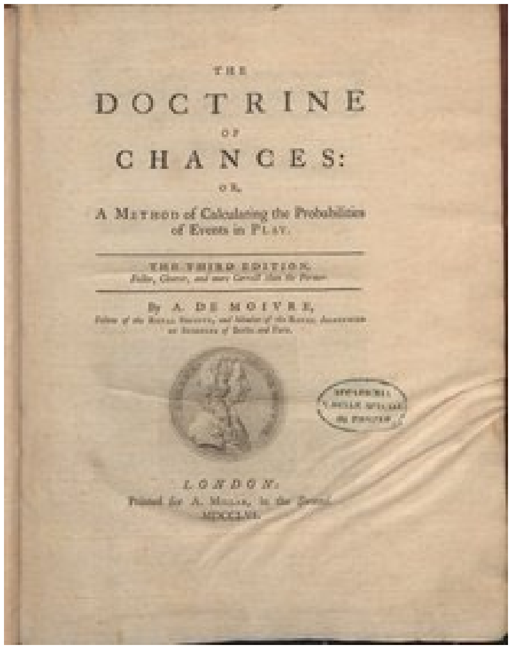
Johan de Witt

1625-1672

Bringing data to life.

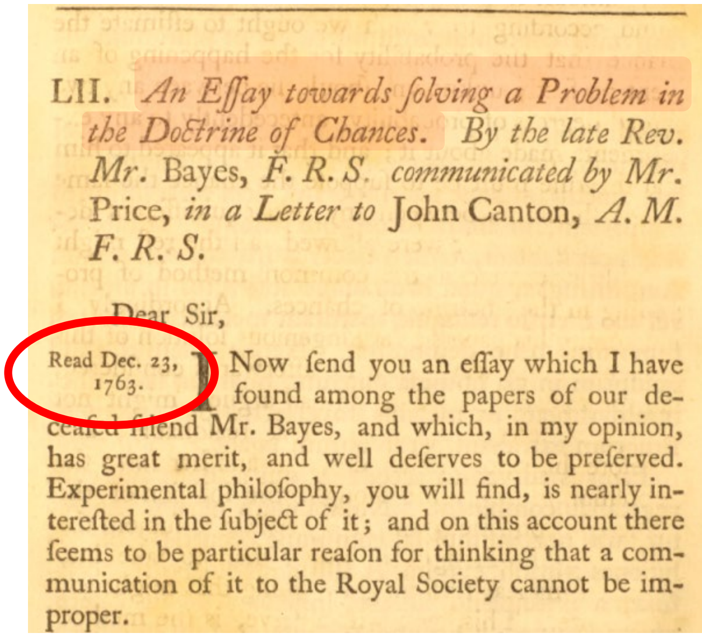
What is Probability?

1711, 1718, 1738, 1756



Given an observation, what am I to infer about the underlying phenomenon?

Inverse Probability



LII. *An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal Society cannot be improper.

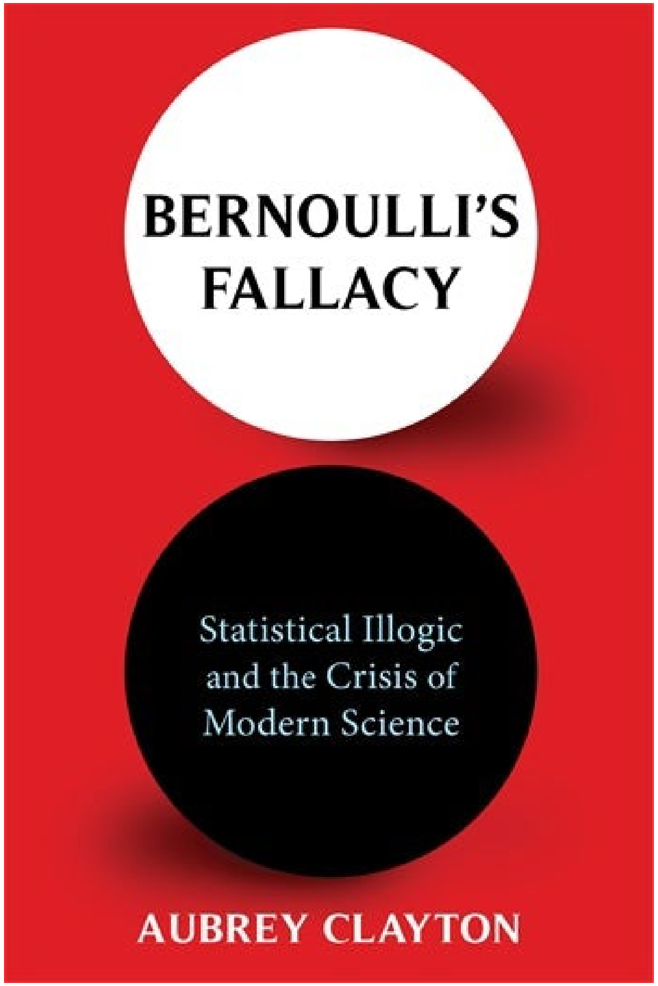
Abraham de Moivre

Thomas Bayes

Bringing data to life.

What is Probability?

October 26, 2021



Insights into …

- History
- Philosophy
- Epistemology

Argument for Bayesian approach

Examples

Bringing data to life.

Epistemology

Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise.”

John Tukey

Bringing data to life.

Epistemology

A p-value is no more than **the ultimate test statistic** scaled to the interval (0, 1).

A p-value is a “precise” answer* to the **wrong question** – pr(Data|Hypothesis).

A p-value is a **poor answer** to one of the three important questions of inference.

***Frequentists require models and assumptions.**

Bringing data to life.

Epistemology

A p-value is a statement about what happened **(post hoc)**

- The hypothesis test I wish I would have done now that I have seen the data

A p-value is **"indirect proof"**

- Proof by contradiction

Bringing data to life.

Epistemology

A Bayesian probability is a "vague" answer* to the **right question**.

A Bayesian probability is **what scientists** – indeed all of us – **want**: pr(Hypothesis|Data).

Report the **upper bound** on the posterior probability of the null hypothesis being false

- Using (at a minimum) a point prior for that hypothesis and the BFB.

$$p_1 \leq \{1 + [(1-p_0)/p_0] \,/\, \mathrm{BFB}\}^{-1}.$$

posterior — prior — data

*Vague in the sense of requiring a subjective prior.

Bringing data to life.

Epistemology

A **Bayesian posterior probability** is a statement about the state of Nature

- What do I believe about the hypothesis now that I have seen the data

A **Bayesian posterior probability** is **"direct proof"**

Epistemology

One cannot interpret a **p-value** in isolation.

One can interpret a **Bayesian posterior probability** directly.

Epistemology

Significance level and power are important elements of *study design* (***Frequentist***)

Positive and negative predictive value are the most appropriate measures for *interpretation of study outcomes* (***Bayesian***)

Bayesian perspective answers the question of interest.

(think diagnostic testing)

Bringing data to life.

Epistemology

Frequentist: compute p-value and then do ***post hoc assessment*** of how it fits into other evidence

- Is it consistent with previous/other findings?

Bayesian: Quantify belief ***a priori*** and build that into a **pre-specified analysis**

- Statisticians advocate pre-specification (ICH-E9)

Bringing data to life.

Epistemology

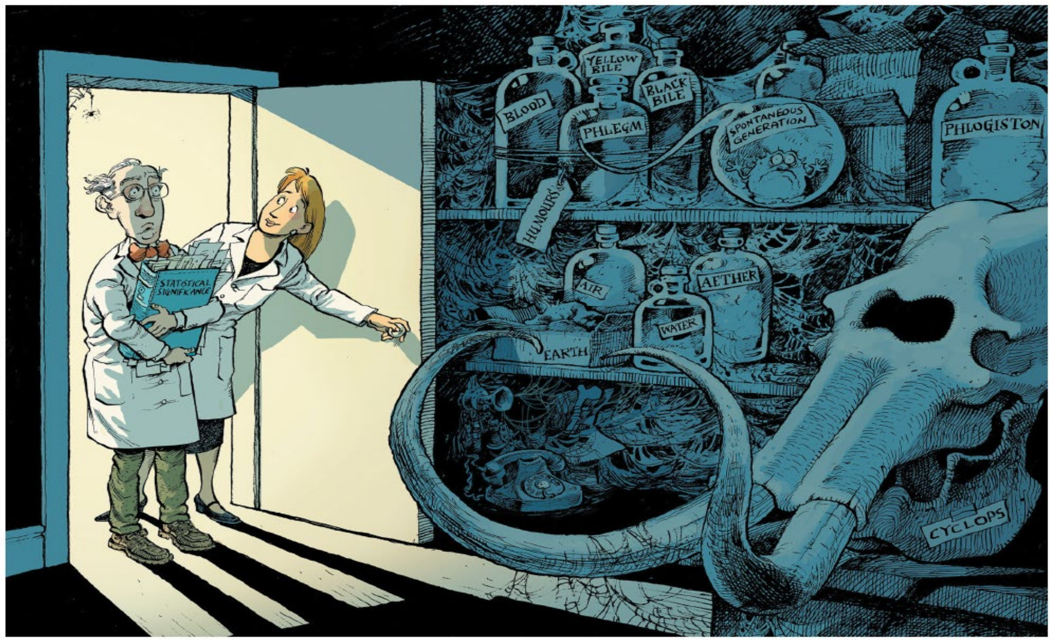
nature 20 March 2019

sh sis

the ds)

hyp ory)

ILLUSTRATION BY DAVID PARKINS



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Bringing data to life.

Epistemology - Summary

Frequentist	Bayesian
“Wrong” Question	Right Question
Indirect	Direct
Post hoc	A Priori
Not interpretable in isolation – need context	Context incorporated into interpretation
Past	Present / Future
Conditional	Unconditional
Exploratory/Confirmatory dichotomy	Hypotheses evaluated quantitatively by their prior

Extra Reading

Statistics in Biopharmaceutical Research

Taylor & Francis
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usbr20

A Bayesian Posterior Probability Is the Real Replication Probability

Stephen J. Ruberg

To cite this article: Stephen J. Ruberg (2020): A Bayesian Posterior Probability Is the Real Replication Probability, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2020.1831952

To link to this article: https://doi.org/10.1080/19466315.2020.1831952

Published online: 05 Nov 2020.

Clinical Pharmacology & Therapeutics

Tutorial | Open Access

Détente: A Practical Understanding of *P* values and Bayesian Posterior Probabilities

Stephen J. Ruberg

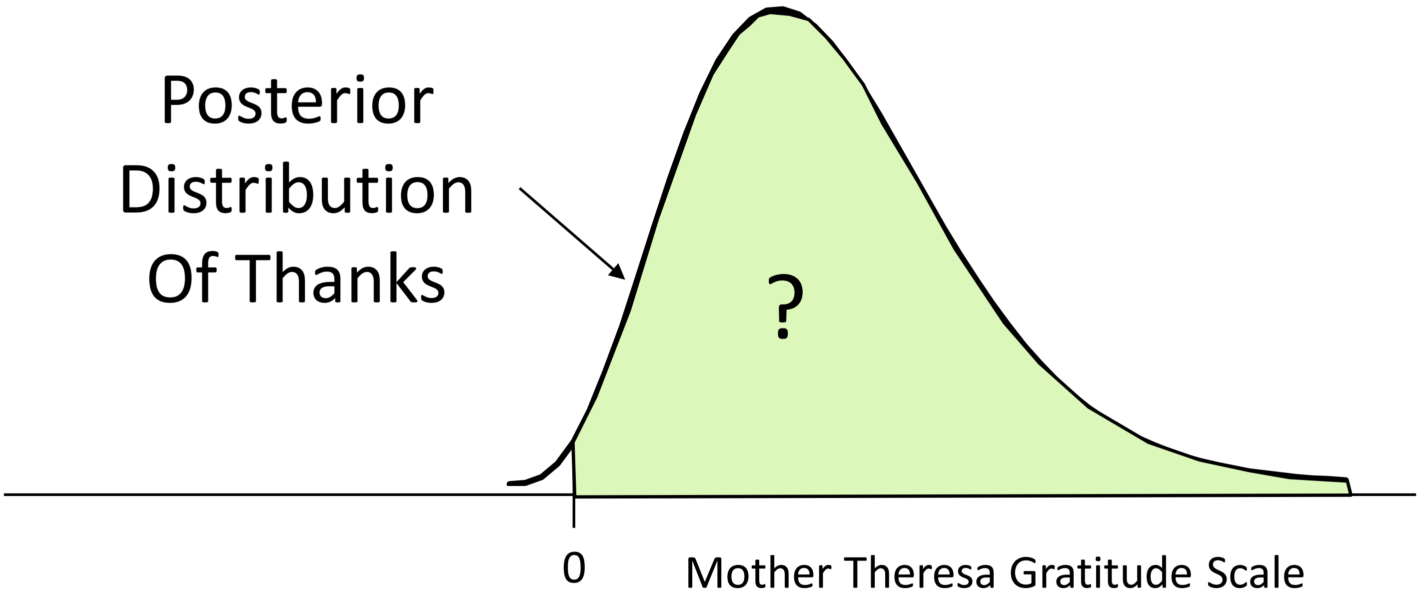
First published: 03 August 2020 | https://doi.org/10.1002/cpt.2004 | Citations: 1

Null hypothesis significance testing (NHST) with its benchmark *P* value < 0.05 has long been a stalwart of scientific reporting and such statistically significant findings have been used to imply scientifically or clinically significant findings. Challenges to this approach have arisen over the past 6 decades, but they have largely been unheeded. There is a growing movement for using Bayesian statistical inference to quantify the probability that a scientific finding is credible. There have been differences of opinion between the frequentist (i.e., NHST) and Bayesian schools of inference, and warnings about the use or misuse of *P* values have come from both schools of thought spanning many decades. Controversies in this arena have been heightened by the American Statistical Association statement on *P* values and the further denouncement of the term "statistical significance" by others. My experience has been that many scientists, including many statisticians, do not have a sound conceptual grasp of the fundamental differences in these approaches, thereby creating even greater confusion and acrimony. If we let A represent the observed data, and B represent the hypothesis of interest, then the fundamental distinction between these two approaches can be described as the frequentist approach using the conditional probability pr(A | B) (i.e., the *P* value), and the Bayesian approach using pr(B | A) (the posterior probability). This paper will further explain the fundamental differences in NHST and Bayesian approaches and demonstrate how they can co-exist harmoniously to guide clinical trial design and inference.

Thank You

Pr(I thank you) = 0.999

Pr(you thank me) = …



Bringing data to life.