

An opinionated view of Bayesian Statistics

A statistical pill from UEB VHIR

Alex Sanchez-Pla
Vall d'Hebron Research Institute and
Universitat de Barcelona
2022-02-26

Outline

- 1) Introduction and motivation
- 2) Probability and Bayes
- 3) BayesianInference
- 4) Further from the basics
- 5) Summary
- 6) References and Resources

Introduction and motivation

Bayesian Stats becoming popular

[nature](#) > [nature reviews drug discovery](#) > [review articles](#) > [article](#)

Published: 01 January 2006

Bayesian clinical trials

[Donald A. Berry](#)

[Front Psychol.](#) 2014; 5: 765.

Published online 2014 Aug 8. Prepublished online 2014 Apr 30.

doi: [10.3389/fpsyg.2014.00765](https://doi.org/10.3389/fpsyg.2014.00765)

The Bayesian boom: good thing or bad?

[Ulrike Hahn*](#)

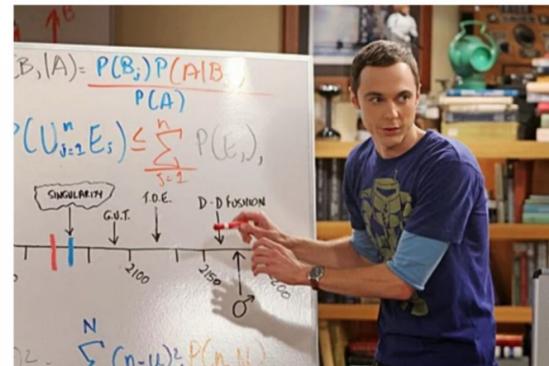
Time for Clinicians to Embrace Their Inner Bayesian?

Reanalysis of Results of a Clinical Trial of Extracorporeal Membrane Oxygenation

Roger J. Lewis, MD, PhD^{1,2,3}; Derek C. Angus, MD, MPH, FRCR^{4,5,6}

[Author Affiliations](#) | [Article Information](#)

JAMA. 2018;320(21):2208-2210. doi:10.1001/jama.2018.16916



Done

amp.theguardian.com

AA C

The obscure maths theorem that governs the reliability of Covid testing

There's been much debate about lateral flow tests - their accuracy depends on context and the theories of a 18th-century cleric

[Coronavirus - latest updates](#)

[See all our coronavirus coverage](#)

Tom Chivers

Sun 18 Apr 2021 02.00 EDT

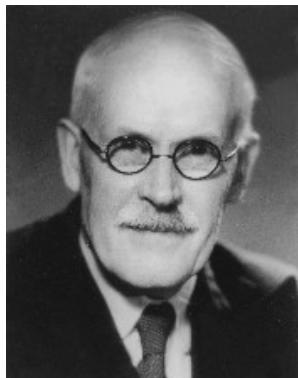


Bayesian statistics is not new



Reverend Thomas Bayes (1701 – 1761)

- Presbyterian minister
- Studied logic and theology in Edinburgh
- Bayes theorem (posthumous publication)



Sir Harold Jeffreys (1891 – 1989)

- Geophysicist, mathematician, astronomer
- Professor for astronomy at Cambridge
- Bayes factor

But what is it about?

- The main characteristic of the Bayesian approach is that it attempts to combine
 - Our previous beliefs about the topic being studied
 - With the current observations provided by an experiment
- This combination may change (for worse or for better) our beliefs and
 - We rely on this posterior beliefs to make inferences.
- This combination or "update" is based on a probability formula:

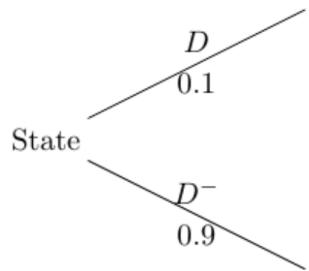
The Bayes Theorem

Probability and Bayes

Diagnostics and probabilities

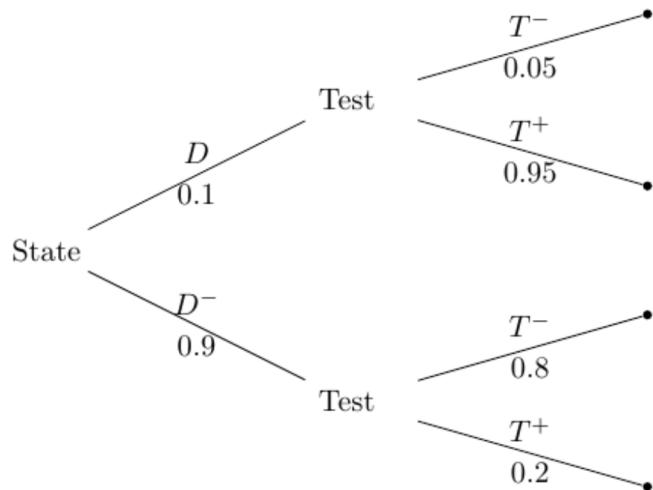
Consider a disease affecting **10%** of the population D .

That is, $1 - 0.1 = 90\%$ of the population is not affected, D^- .



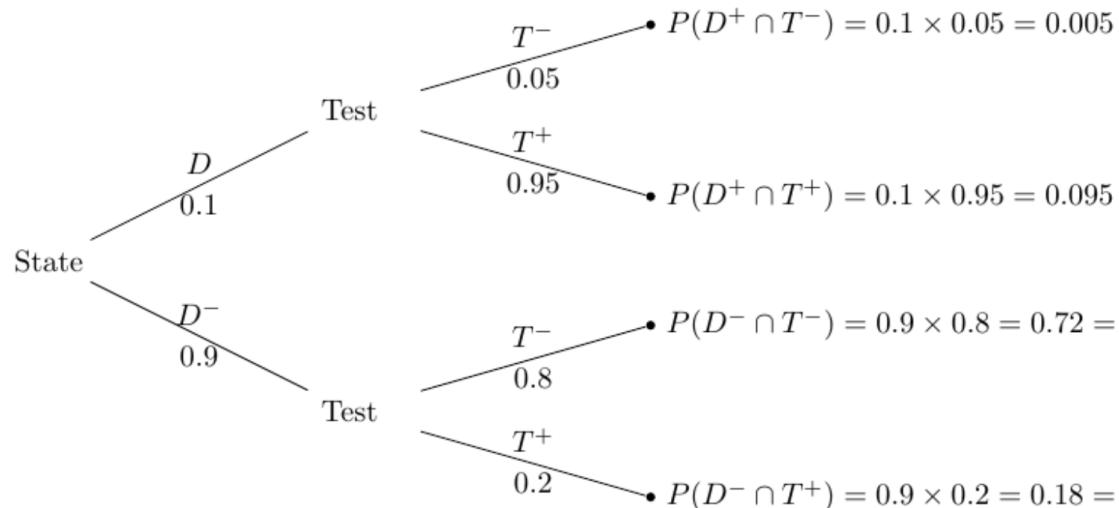
Diagnostics and probabilities

A diagnostic test can be positive, T^+ , or negative, T^- , with distinct probabilities for healthy or affected people.



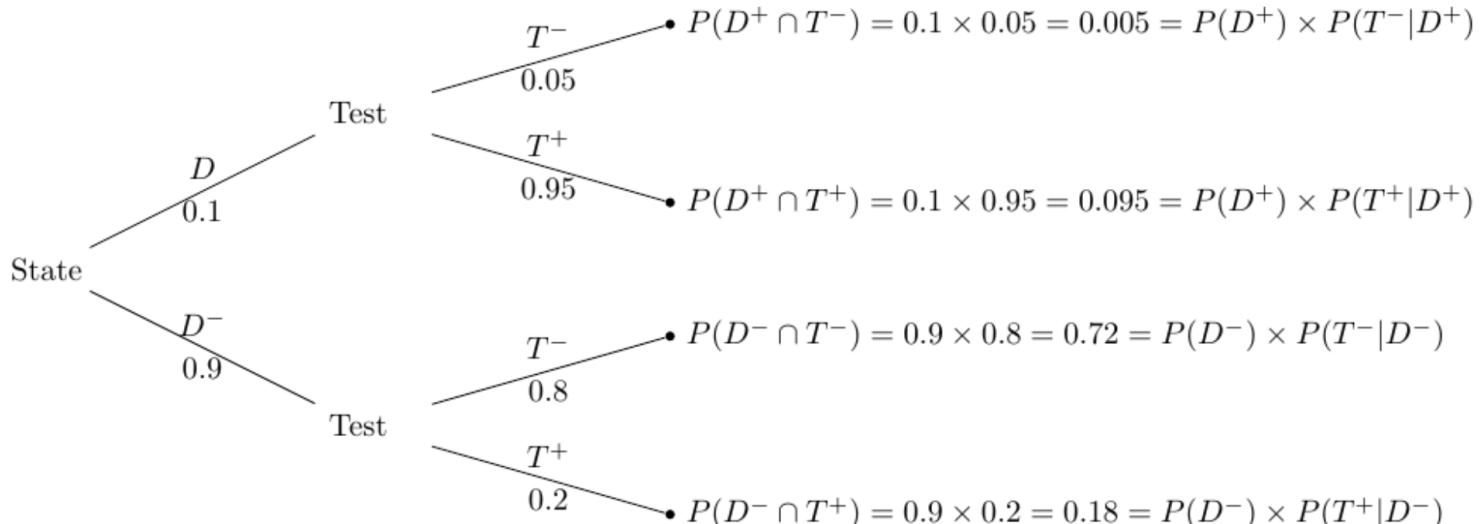
Diagnostics and probabilities

Each branch in the tree describes the probability of observing two events simultaneously



Diagnostics and probabilities

Probabilities in the second nodes are *conditional probabilities*



Conditional probabilities

- Conditional probabilities inform about the probability of an event when another is known to have happened.
- $P(T^+)$ is not the same as $P(T^+/D)$ nor as $P(T^+/D^-)$.
- It is straightforward to see that

$$P(T^+/D) = \frac{P(D \cap T^+)}{P(D)}, \quad P(D/T^+) = \frac{P(D \cap T^+)}{P(T^+)}, \quad \text{etc.}$$

- In general, given two events A and B , we will say that:
 - If $P(A/B) > P(A)$ (e.g Having the disease *increases* the probability of testing positive): B favours A .
 - If $P(A/B) < P(A)$ (e.g Not having the disease *decreases* the probability of testing positive): B disfavours A .
 - If $P(A/B) = P(A)$ or equivalently $P(B/A) = P(B)$ then A and B are said to be *independent*.

Back to diagnostics

Recall the usual table appearing in diagnostics:

		Patient's true state		Marginal totals	Diagnostic measures
		D^+	D^-		
Test result	T^+	<i>True positive</i> (TP)	<i>False positive</i> (FP)	Total T^+ (TP + FP)	Positive predictive value = $\frac{TP}{TP+FP}$
	T^-	<i>False negative</i> (FN)	<i>True negative</i> (TN)	Total T^- (FN + TN)	Negative predictive value = $\frac{TN}{FN+TN}$
Marginal totals		Total D^+ (TP + FN)	Total D^- (FP + TN)	Total Patients (TP + FP + FN + TN)	
Diagnostic measures		Sensitivity = $\frac{TP}{TP+FN}$	Specificity = $\frac{TN}{FP+TN}$		Accuracy = $\frac{TP+TN}{TP+FP+FN+TN}$

These terms can be related (expressed as) conditional probabilities,

Diagnostic and conditional probs.

		Patient's true state		Marginal probabilities	Diagnostic measures
		D^+	D^-		
Test result	T^+	$P(D^+ T^+) \Rightarrow$ $P(T^+ D^+) \Downarrow$	$P(D^- T^+) \Rightarrow$ $P(T^+ D^-) \Downarrow$	$P(T^+)$	Positive Predictive Value = $P(D^+ T^+)$
	T^-	$P(D^+ T^-) \Rightarrow$ $P(T^- D^+) \Downarrow$	$P(D^- T^-) \Rightarrow$ $P(T^- D^-) \Downarrow$	$P(T^-)$	Negative predictive value = $P(D^- T^-)$
Marginal probabilities		$P(D^+)$	$P(D^-)$	$P = 1$	
Diagnostic measures		Sensitivity = $P(T^+ D^+)$	Specificity = $P(T^- D^-)$		Accuracy $= P(T^+ D^+) \times P(D^+) + P(T^- D^-) \times P(D^-)$ $= P(D^+ T^+) \times P(T^+) + P(D^- T^-) \times P(T^-)$

Switching the focus

We are often given sensitivity or specificity, that is probability of testing positive if affected by the disease (equivalently ...)

$$Sens = P(T^+|D), \quad Spec = P(T^-|D^-)$$

But, in practice we may be more interested in the reverse concept, that is,

- probability of being affected if testing positive:

$$PPV = P(D|T^+),$$

or, similarly, the probability of not being affected if testing negative)

$$NPV = P(D^-|T^-).$$

Bayes formula allows expressing the later quantities in terms of the former ones and the prevalence of the disease.

Relating quantities

Notice that the probability of having the disease **and** testing positive can be expressed in terms of sensitivity but also of the PPV.

$$P(T^+|D^+) = \frac{P(T^+ \cap D^+)}{P(D^+)} \longrightarrow P(T^+ \cap D^+) = P(T^+|D^+) \times P(D^+)$$

$$P(D^+|T^+) = \frac{P(T^+ \cap D^+)}{P(T^+)} \longrightarrow P(T^+ \cap D^+) = P(D^+|T^+) \times P(T^+)$$

And also that the probability of testing positive can be computed as:

$$P(T^+) = P(T^+|D^+) \times P(D^+) + P(T^+ \cap D^-) \times P(D^-)$$

Now combining the expressions above gives us a formula for the probability of having the disease, given the data and the probability of testing positive.

Bayes theorem

From the previous equality, we have:

$$\begin{aligned} P(D^+|T^+) &= \frac{P(T^+ \cap D^+)}{P(T^+)} = \\ &= \frac{P(T^+|D^+) \times P(D^+)}{P(T^+|D^+) \times P(D^+) + P(T^+ \cap D^-) \times P(D^-)}. \end{aligned}$$

This is the so-called Bayes Theorem, which has deep implications, not only in diagnosis, but in many aspects of statistical inference.

Applying Bayes (# 1/2)

- Interested in diagnosing cancer in patients who visit a chest clinic
 - C represents the event "Person has cancer"
 - S represents the event "Person is a smoker".
- We *believe* that a good approach to the probability that a person entering the clinic turns out to have cancer is: $P(C) = 0.1$.
- Want to find probability that if the person is a smoker she has cancer.
 - Hard to find directly.
 - However, we are likely to know $P(S)$ by considering the percentage of patients who smoke.
 - Suppose $P(S) = 0.5$.
 - We are also likely to know $P(S|C)$ by checking from our record the proportion of smokers among those diagnosed.
 - Suppose $P(S|C) = 0.8$.

Applying Bayes (# 2/2)

- We can now use Bayes' rule to compute:

$$P(C|S) = \frac{P(S|C) \times P(S)}{P(C)} = \frac{0.1 \times 0.9}{0.5} = 0.16$$

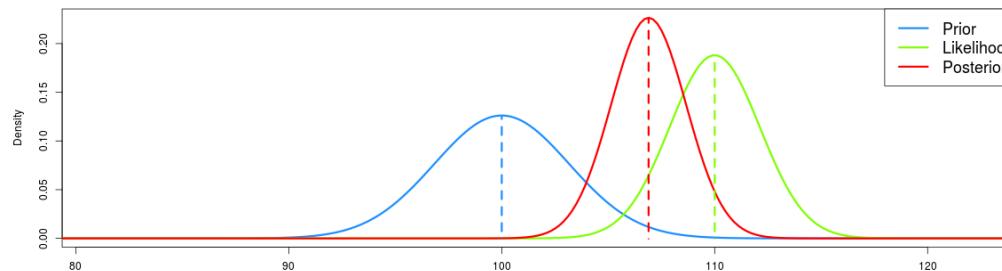
- Thus,
 - in the light of evidence, *that the person is a smoker*
 - we have updated our previous belief, or *prior probability*, from **0.1** to a *posterior probability* of **0.16**
- Leaving aside that it is still unlikely that the person has cancer, this is an example on how Bayesian reasoning works:
 - Previous beliefs are expressed as *priori* probabilities
 - Data are *combined* with beliefs and other informations to yield
 - *Posterior probabilities* or updated beliefs.

Bayes for probability distributions

- Bayes idea can be extended to distributions (random variables).

$$f(\theta|x) = \frac{f(\theta, x)}{f(x)} = \frac{f(\theta) \times f(x|\theta)}{f(x)}$$

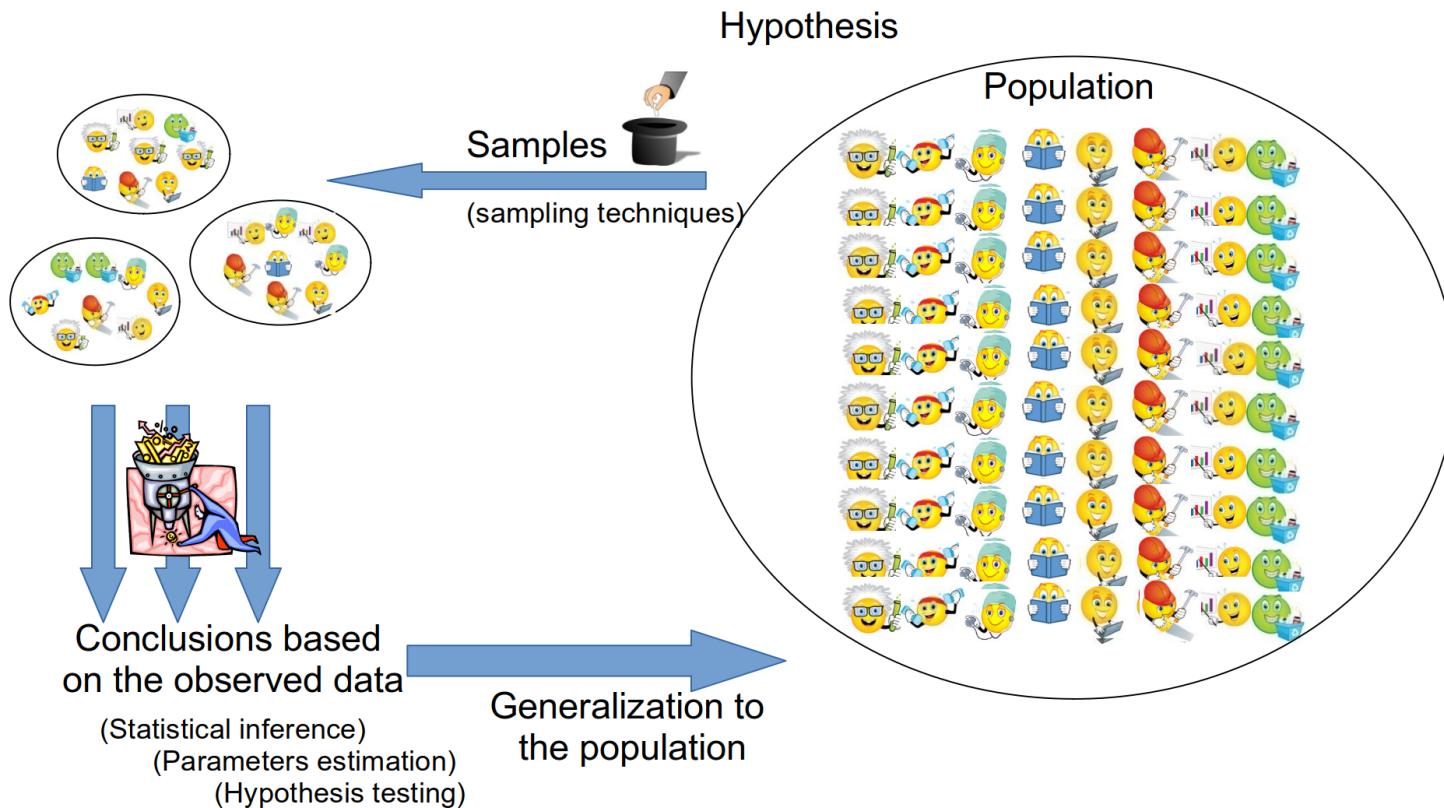
- *Prior* : probability distribution of the parameter \sim prior beliefs about which values it may take with which probabilities.
- *Likelihood*: \sim how likely it is to observe the data for each value of the parameter (data distribution with x fixed as θ varies).
- *Posterior*: Updated probability distribution of the parameter \sim How our beliefs are changed after having collected the data.



<https://www.rensvandeschoot.com/tutorials/fbi-the-app/>

Bayesian Inference

Goal of inferential statistics



Goal: make inferences on the parameters θ using the observed data x .

Two schools of thought

Frequentist

- No prior information
- Relies only (mainly) on collected data
- Data: repeatable random sample
- Parameters considered as fixed unknown values.
- Probability: long-term frequency
- *Easy* to compute

Bayesian

- Uses Prior information
- Combines collected data and prior information.
- Data fixed (assumed non repeatable)
- Parameters: unknown but probabilistically described
- Probability: subjective belief or based on experience
- Complex to compute.

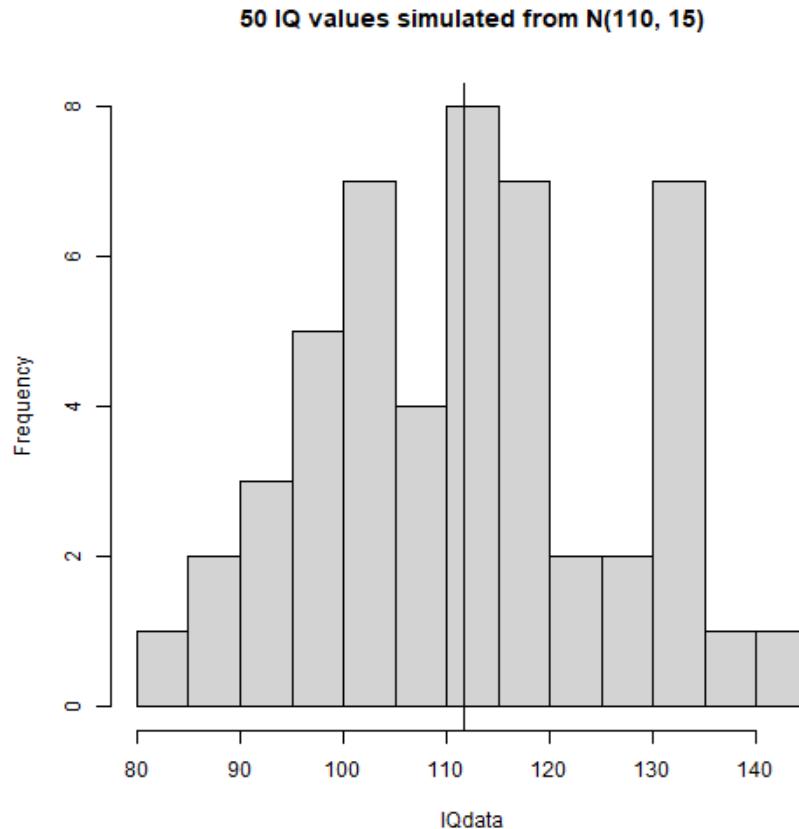
Random variables and inference

- Traditionally (*usually*) we assume a probability distribution for the data and rely on it to make inferences such as:
 - Improve knowledge of these parameters (estimation)
 - Take decision about their values (testing).
- Frequentist approach
 - Assumes probability distribution for the data, but parameters are fixed, unknown values.
 - Inference based solely on the data.
- Bayesian Statistics
 - Also assumes a prior distribution for the parameters
 - *Computes* a posterior distribution
 - Inference is based on the posterior distribution.

Example: IQ estimation

- We are interested in estimating the IQ from a population.
- A frequentist approach,
 - Assumes that data is normally distributed, with a fixed IQ mean.
 - To approach IQ:
 - Sample the population.
 - Estimate the mean based on the sample.
- A Bayesian approach
 - Also assumes data normality, but
 - Assumes a prior distribution for its parameters (e.g. $\mu \sim N()$)
 - Computes posterior distribution and uses it to do inference.

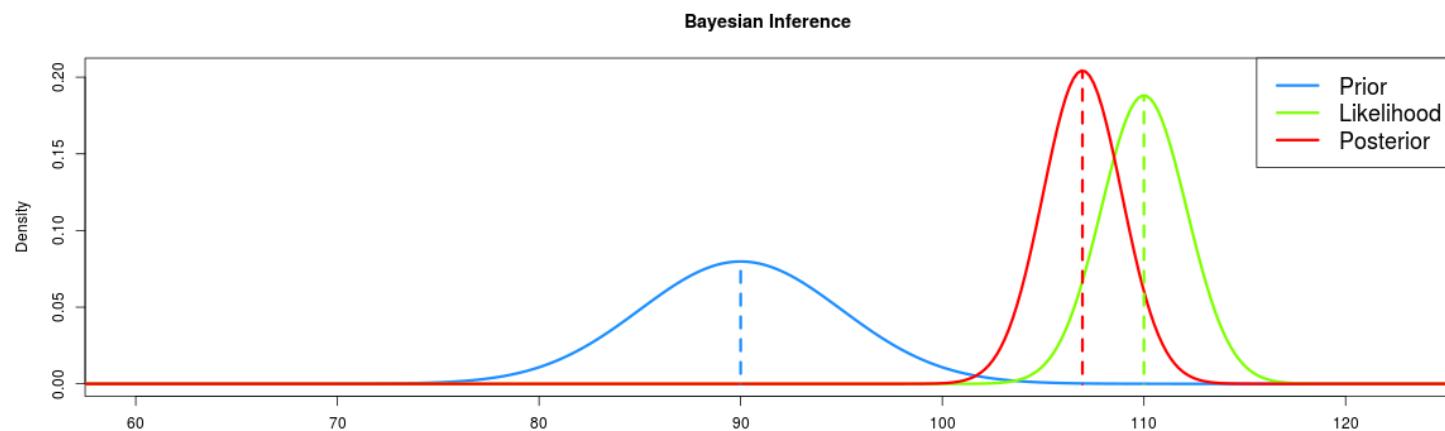
Ex. 1a: Frequentist IQ estimation



```
## Sample mean = 111.6832
```

```
## Confidence Intervals (95%) = 107.5518 115.8146
```

Ex. 1b: Bayesian IQ estimation



Summary

	Mean	Variance	95% Lower Bound	95% Upper Bound	Calculation
Prior	90	25	80.2	99.8	Based on input values
Data	110	225	80.6	139.4	Based on exact simulation (frequentist)
Likelihood	110	4.5	105.84	114.16	Based on exact simulation (frequentist)
Posterior	106.95	3.81	103.12	110.78	Based on analytical derivation (Bayesian)

<https://www.rensvandeschoot.com/tutorials/fbi-the-app/>

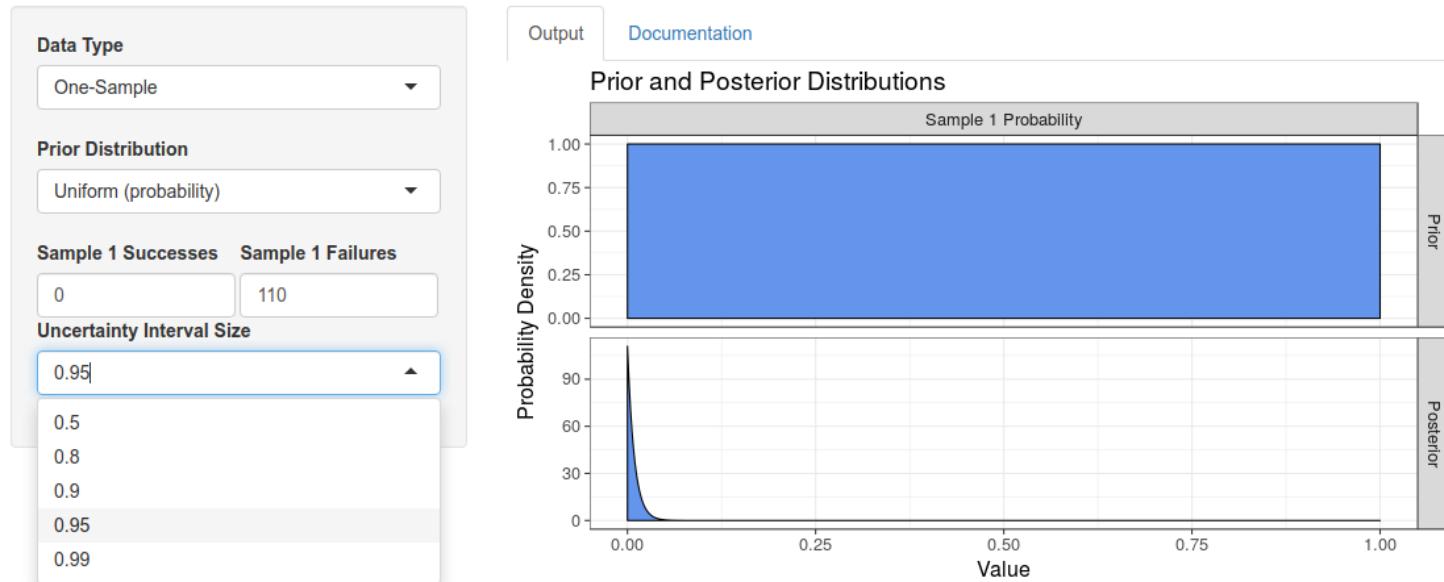
Ex. 2a: Frequentist prop. estimation

- Assume a binomial probability model with probability p for the number of affected individuals in a sample of size N .
- We collect data and estimate p using the relative frequency (% of individuals with the disease).
- Case (i):
 - A total of 4 affected people found in 300 individuals.
 - $\hat{p}_1 = 4/300 = 0.0133$
 - Confidence interval for p : $(0.00035, 0.026)$
- Case (ii):
 - In a sample of 110 individuals no affected person was found.
 - $\hat{p}_2 = 0$ even if we are strongly convinced that the disease is present in the population!

Ex. 2bii: Bayesian prop. estimation

- The fact that we believe there are some affected individuals, is expressed through the prior distribution.
- This will yield a non-zero estimate *even if we don't observe them*.

Bayesian Estimation with Binomial Data



95% Uncertainty Intervals

Estimand	Mean	Mode	Lower Bound	Upper Bound
Sample 1 Probability	0.009	0.000	0.000	0.033

Ex. 2bii: Changing the prior

Different prior distributions can yield different results

Bayesian inference

Example E of section 3.5 of Rice 3rd edition, page 94-95

beta prior, likelihood is binomial, posterior is beta

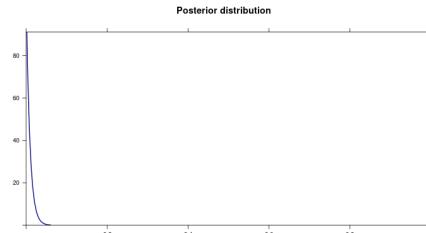
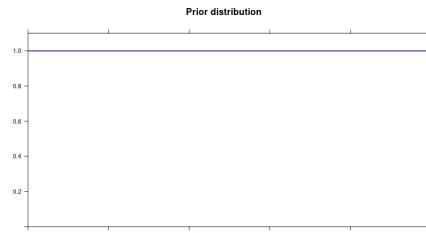
See Figure 3.16 for results from a $\text{beta}(1, 1)$ prior and 13 successes out of 20 attempts

alpha parameter for Beta prior:

beta parameter for Beta prior:

number of successes observed:

out of this many trials (must be larger than # successes):



Bayesian inference

Example E of section 3.5 of Rice 3rd edition, page 94-95

beta prior, likelihood is binomial, posterior is beta

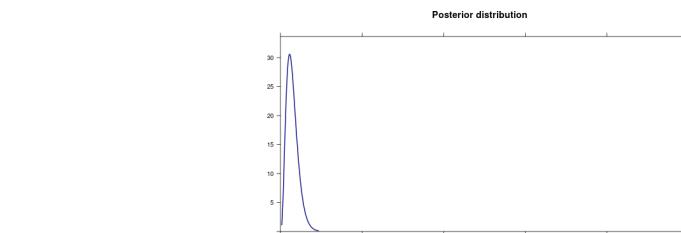
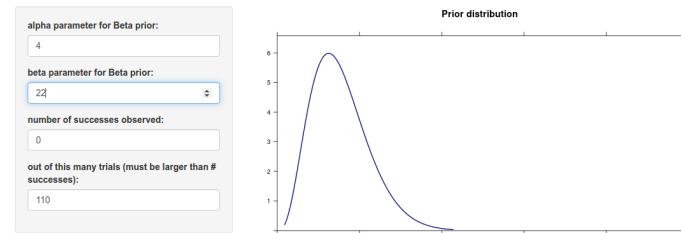
See Figure 3.16 for results from a $\text{beta}(1, 1)$ prior and 13 successes out of 20 attempts

alpha parameter for Beta prior:

beta parameter for Beta prior:

number of successes observed:

out of this many trials (must be larger than # successes):

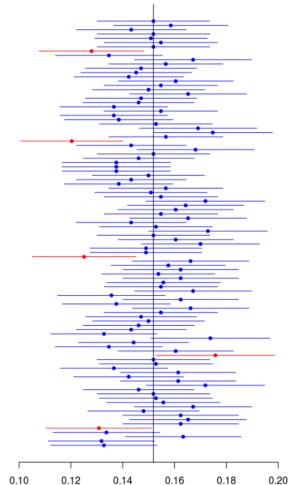


<https://r.amherst.edu/apps/nhorton/Shiny-Bayes/>

- This is good: flexibility to accommodate beliefs.
- But also one main source of criticism.
 - Prior choice said to be arbitrary, subjective or even "cheating"

Confidence vs Credible Intervals (1)

Confidence intervals are often misunderstood



- They are **erroneously** described as the values such that the probability that the parameter lies among them is, say $(1 - \alpha)$.
- But this is conceptually wrong because the parameter is assumed to be fixed, so it is either inside or outside the interval.
- The **right** definition is correct but hard to communicate: *The values s.t. if we sampled repeatedly from the same population we would expect that, on average, the interval would cover the true value of the parameter $(1 - \alpha)\%$ of the times*

Confidence vs Credible Intervals (2)

Credible intervals are what they seem they should be.

Bayesian Hypothesis testing

- Alternative to using p-values, which have been criticized "ad-nauseam".
- Intuitive idea:
 - p-values based on $P(Data|H_0)$ which is quantified through the sampling distribution of the test statistic.
 - What we are really interested in is $P(H|Data)$.
 - This can be addressed adopting a Bayesian approach.

Bayes factors

- The *prior odds* is defined as the ratio of the prior probabilities assigned to the hypotheses or models we're considering.

$$O[H_1 : H_2] = \frac{P(H_1)}{P(H_2)}$$

- The *posterior odds* is the ratio of the two posterior probabilities of these hypotheses.

$$POdds = PO[H_1 : H_2] = \frac{P(H_1 | Data)}{P(H_2 | Data)}$$

- This can be expanded using Bayes formula and reorganized.

The Bayes factor

$$\begin{aligned}POdds &= \frac{P(Data|H_1) \times P(H_1)/P(data)}{P(Data|H_2) \times P(H_2)/P(data)} \\&= \underbrace{\frac{P(Data|H_1)}{P(Data|H_2)}}_1 \times \underbrace{\frac{P(H_1)}{P(H_2)}}_2\end{aligned}$$

(1) is called *Bayes Factor* and (2) is the *Prior Odds* so that we can write:

$$POdds[H_1 : H_2] = BF[H_1 : H_2] \times O[H_1 : H_2]$$

The Bayes factor quantifies the evidence of data arising from hypothesis one versus hypothesis two.

- In a discrete case, this is simply the ratio of the likelihoods of the observed data under the two hypotheses or models

Bayesian tests rely on the use of Bayes factor.

Software for Bayesian Statistics



Example: Bayesian t-test (1)

Example: Turning the Hands of Time

In a series of four experiments, [Topolinski and Sparenberg \(2012\)](#) found support for the conjecture that clockwise movements induce psychological states of temporal progression and an orientation toward the future and novelty. Here we report the results of a preregistered replication attempt of Experiment 2 from [Topolinski and Sparenberg \(2012\)](#). Participants turned kitchen rolls either clockwise or counterclockwise while answering items from a questionnaire assessing openness to experience. Data from 102 participants



Wagenmakers et al. (2015)

Example: Bayesian t-test (2)

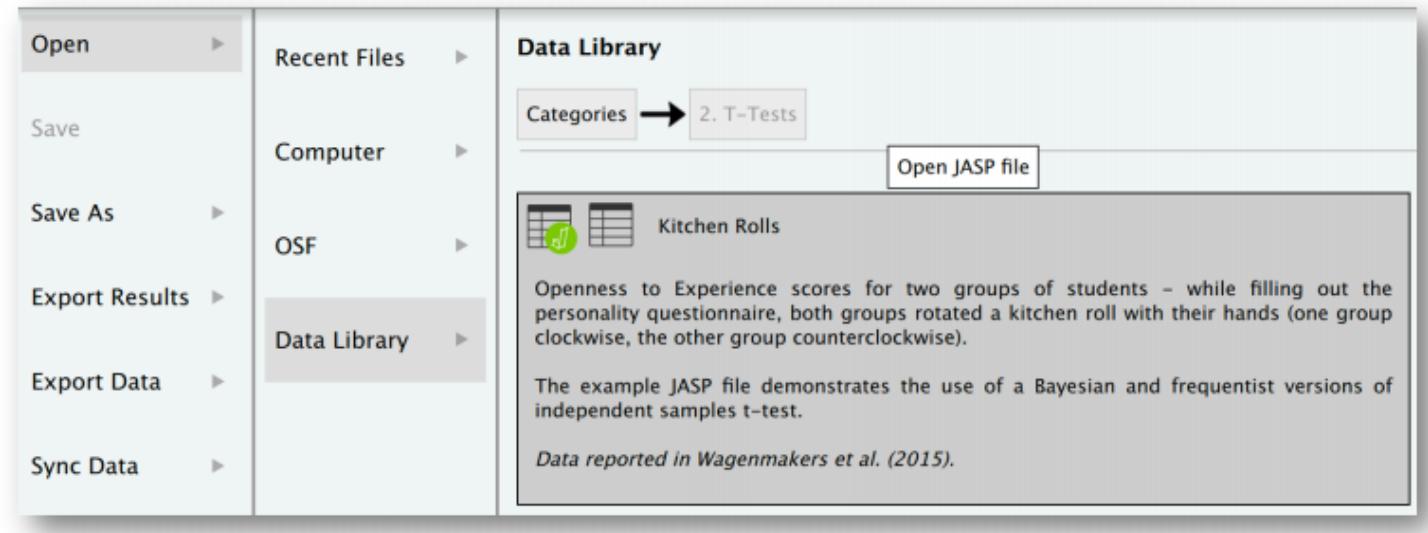
Step 1: Formulate hypotheses

- H₀: The direction in which the kitchen rolls are turned does not affect the reported openness.
- H₀: $\delta = 0$
- H₁: Participants who turn the kitchen rolls clock-wise report higher values of openness compared to participants who turn the kitchen rolls counter-clockwise.
- H₁: $\mu_{\text{clockwise}} - \mu_{\text{counterclockwise}} > 0 \rightarrow \delta > 0$

Example: Bayesian t-test (3)

Step 2: Load the data

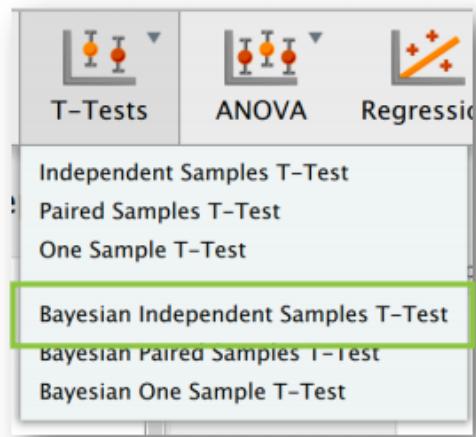
<https://osf.io/3wg6m/> → KitchenRolls.csv



Example: Bayesian t-test (4)

Step 3: Choose the hypothesis test

Bayesian independent samples t-test



Example: Bayesian t-test (5)

t-Test

Step 4: Settings

- Direction of the test
- Output content

Bayesian Independent Samples T-Test

ParticipantNumber
Condition
q1_check
q2_check
q1_NE0
q2_NE0
q3_NE0
q4_NE0
q5_NE0
q6_NFO

Dependent Variables
mean_NE0

Grouping Variable
Rotation

Alt. Hypothesis
 Group 1 ≠ Group 2
 Group 1 > Group 2
 Group 1 < Group 2

Plots
 Prior and posterior
 Additional info
 Bayes factor robustness check
 Additional info
 Sequential analysis
 Robustness check
 Descriptives
Credible interval 95 %

Bayes Factor
 BF_{10}
 BF_{01}
 Log(BF_{10})

Tests
 Student
 Mann-Whitney
No. samples 1000

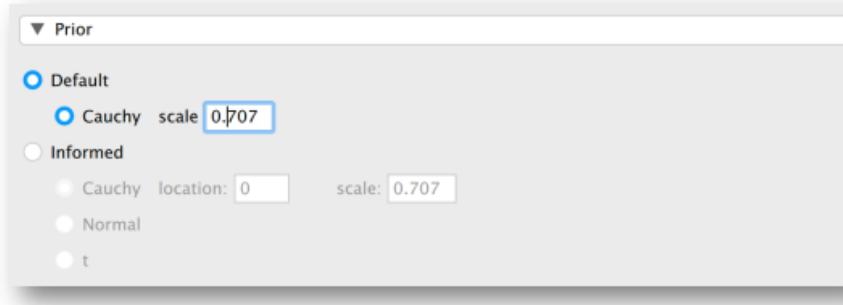
Missing Values
 Exclude cases analysis by analysis
 Exclude cases listwise

Additional Statistics
 Descriptives

Example: Bayesian t-test (6)

Step 4: Settings

- Prior on δ under H0: Spike Prior on 0
- Prior on δ under H1: truncated Cauchy distribution with location parameter 0 and scale parameter 0.707



t-distribution with 1 degree of freedom
Fatter tails than normal distribution

cc

Example: Bayesian t-test (7)

Default Priors

(= Uninformed Priors, Objective Priors)

- Based on mathematical desiderata
- Assumption: No prior knowledge
- Can be used independent of application domain
- JASP standard option

Informed Priors

(= Subjective Priors)

- Based on theoretical considerations
- Assumption: Prior knowledge exists
- Dependent on application domain
- Often part of advanced settings in JASP

Example: Bayesian t-test (8)

Step 5: Interpret results

Bayesian Independent Samples T-Test ▾		
	BF ₀₊	error %
mean_NEO	7.749	~ 0.009

Note. For all tests, the alternative hypothesis specifies that group *clock* is greater than group *counter*.

Bayes factor was determined through sampling methods.
error% = sampling error

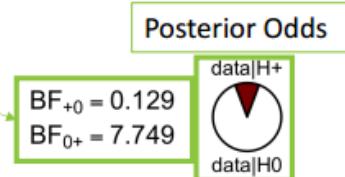
Bayes factor of 7.749 in favor of the null hypothesis:
The data are 7.749 times more likely under the null model than under the alternative model.

67

Example: Bayesian t-test (9)

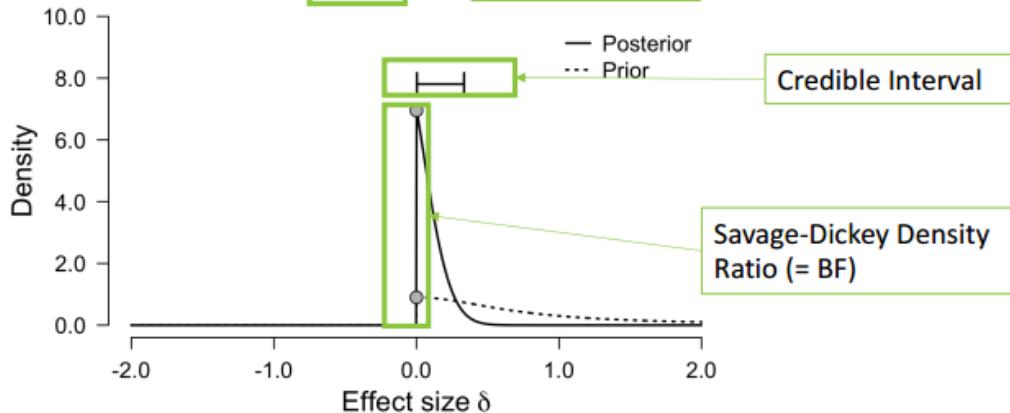
Step 5: Interpret results

Bayes Factor for:
 $\delta > 0$ vs. $\delta = 0$



median = 0.086
95% CI: [0.003, 0.334]

Median of the posterior distribution
and 95% credible interval



68

Further from the basics

Bayes in a Clinical Trial

Pfizer/BioNTechCOVID-19 vaccine trial uses a Bayesian design and analysis plan for the primary efficacy endpoint



Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Absalon, M.D.,
Alejandra Gurtman, M.D., Stephen Lockhart, D.M., John L. Perez, M.D., Gonzalo Pérez Marc, M.D.,
Edson D. Moreira, M.D., Cristiano Zerbini, M.D., Ruth Bailey, B.Sc., Kena A. Swanson, Ph.D.,
Satrajit Roychoudhury, Ph.D., Kenneth Koury, Ph.D., Ping Li, Ph.D., Warren V. Kalina, Ph.D., David Cooper, Ph.D.,
Robert W. Frenck, Jr., M.D., Laura L. Hammitt, M.D., Özlem Türeci, M.D., Haylene Nell, M.D., Axel Schaefer, M.D.,
Serhat Ünal, M.D., Dina B. Tresnan, D.V.M., Ph.D., Susan Mather, M.D., Philip R. Dormitzer, M.D., Ph.D.,
Uğur Şahin, M.D., Kathrin U. Jansen, Ph.D., and William C. Gruber, M.D., for the C4591001 Clinical Trial Group*

Reporting with Bayesian summaries

Table 2. Vaccine Efficacy against Covid-19 at Least 7 days after the Second Dose.*

Efficacy End Point	BNT162b2		Placebo		Vaccine Efficacy, % (95% Credible Interval)‡	Posterior Probability (Vaccine Efficacy >30%)§
	No. of Cases	Surveillance Time (n)†	No. of Cases	Surveillance Time (n)†		
	(N=18,198)				(N=18,325)	
Covid-19 occurrence at least 7 days after the second dose in participants without evidence of infection	8	2.214 (17,411)	162	2.222 (17,511)	95.0 (90.3–97.6)	>0.9999
	(N=19,965)				(N=20,172)	
Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection	9	2.332 (18,559)	169	2.345 (18,708)	94.6 (89.9–97.3)	>0.9999

* The total population without baseline infection was 36,523; total population including those with and those without prior evidence of infection was 40,137.

† The surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.

‡ The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

§ Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

More intuitive reporting

- Bayesian statistics give clear and interpretable results
 - *Given the trial data, the probability that the vaccine is at least 90% efficacious is 0.98*
- The credible intervals give another probabilistic interpretation
 - Given the trial data, with 95% probability that the vaccine efficacy is between 90.3% and 97.6%
- Both statements directly support regulatory decision making.

Summary

Summary

Bayesian statistics:

- Is useful in many settings, and you should know about it.
- Is often not very different in practice from frequentist statistics
- It is often helpful to think about analyses from both Bayesian and non-Bayesian points of view.
- Is not reserved for hard-core mathematicians, or computer scientists, or philosophers. If you find it helpful, use it.

To end with a quote ...



Using a spade for some jobs and shovel for others does *not* require you to sign up to a lifetime of using only Spadian or Shovelist philosophy, or to believing that *only* spades or *only* shovels represent the One True Path to garden neatness.



There are different ways of tackling statistical problems, too.

And there is no need to fight

Two old-timers slugging out the Bayes vs Frequentist battle;

If [Bayesians] would only do as [Bayes] did and publish posthumously we should all be saved a lot of trouble

Maurice Kendall (1907–1983), [JRSSA 1968](#)



The only good statistics is Bayesian Statistics

Dennis Lindley (1923–2013)

in [The Future of Statistics: A Bayesian 21st Century](#) (1975)

- For many years – until recently – Bayesian ideas in statistics* were widely dismissed, often without much thought
- Advocates of Bayes had to fight hard to be heard, leading to an ‘us against the world’ mentality – & predictable backlash
- Today, debates tend be less acrimonious, and more tolerant

* *and sometimes the statisticians who researched and used them*