🤔

You want fair results from a biased coin. What can you do?

- ❑ Toss it twice. If both results are the same, toss it again. If they’re different, take the first result.
- ❑ Alternate between putting heads or tails on top in the starting position.
- ❑ Place the coin on its side and give it a side spin.

Bayesian Statistics

a different take on probability

👨‍🏫 Statistical Reasoning Lecture #7
Alexander Savi, 2025
📚 Mehmetoglu & Mittner Ch. 15

Dynamic Textures #382 by Claus O. Wilke 🎨

📢 Announcements

Assignment 7

- ❏ Deadline Friday 23:55
- ❏ Graded Tuesday

Practice mode

- ❏ Assignment 1 to 6 open in practice mode on Monday after lecture latest (practice with answers, not able to submit)
- ❏ Assignment 7 opens Saturday 10AM latest

🔮 Today

Topics

1 | Statistical reasoning with GLM
2 | Multiple linear regression
3 | Dummy-variable regression
4 | Logistic regression
5 | Multilevel and longitudinal analysis
6 | Statistics superpowers
7 | Bayesian statistics (a brief introduction)
   7.1 | Bayes theorem
   7.2 | Bayes factor
   7.3 | Bayesian parameter estimation
   7.4 | Frequentist vs. Bayesian inference

Lecture

Conceptual understanding using a simple example

Assignment

Conceptual understanding & Bayesian regression in practice (requires book)

Bayes' Theorem

the probability of a hypothesis given the data

Dynamic Textures #382 by Claus O. Wilke 🎨

Mental disorder diagnosis

> **We found that many young children are being prescribed medications very soon after their diagnosis of ADHD is documented. That's concerning, because we know starting ADHD treatment with a behavioral approach is beneficial; it has a big positive effect on the child as well as on the family."**

— Stanford Medicine (Aug. 29, 2025)

Mental disorder diagnosis

Data: 📋 positive test result
Hypothesis: 😵‍💫 ADHD

Then, what would P(😵‍💫 | 📋) be? 🦸

🦹 *High, I hope?*

Sure, but how can you tell? 🦸

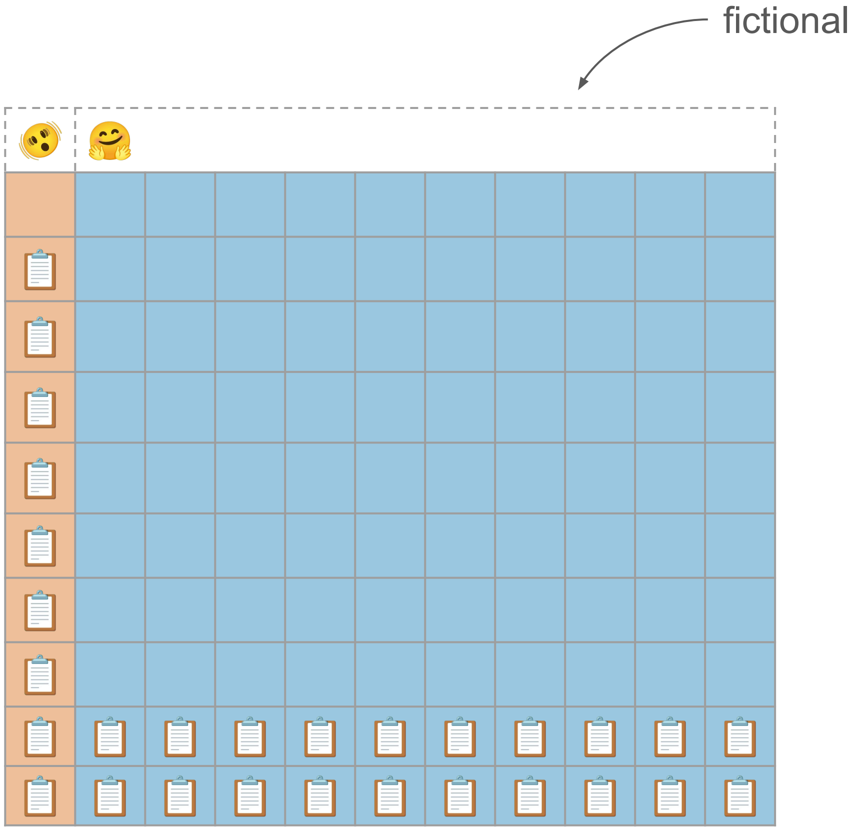
Mental disorder diagnosis

Data: 📋 positive test result
Hypothesis: 😵‍💫 ADHD

Then, what would P(😵‍💫 | 📋) be? 🦸

🦹 *High, I hope?*

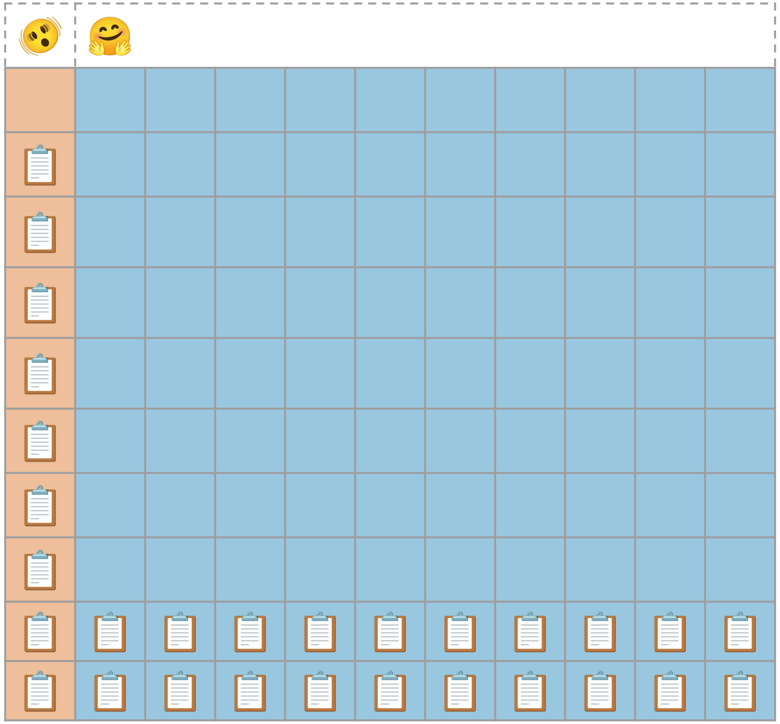
Sure, but how can you tell? 🦸



Base rate

“It is the proportion of individuals in a population who have a certain characteristic or trait.”

— Wikipedia



Base rate fallacy

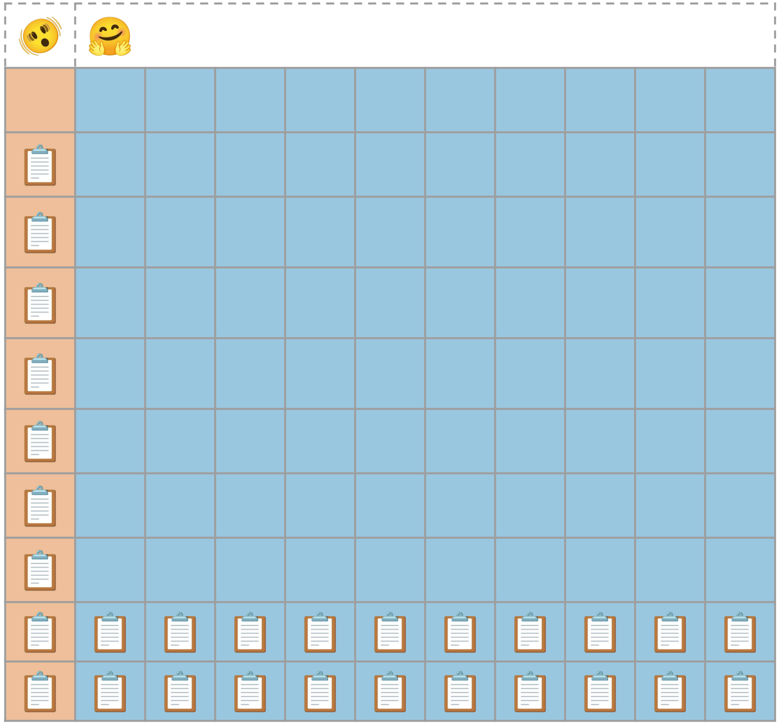
"[A] type of fallacy in which people tend to ignore the base rate (e.g., general prevalence) in favor of the individuating information (i.e., information pertaining only to a specific case)."

— [Wikipedia]

Interesting, let me ask it differently. Out of the number of people who test positive, how many have ADHD? 🦸

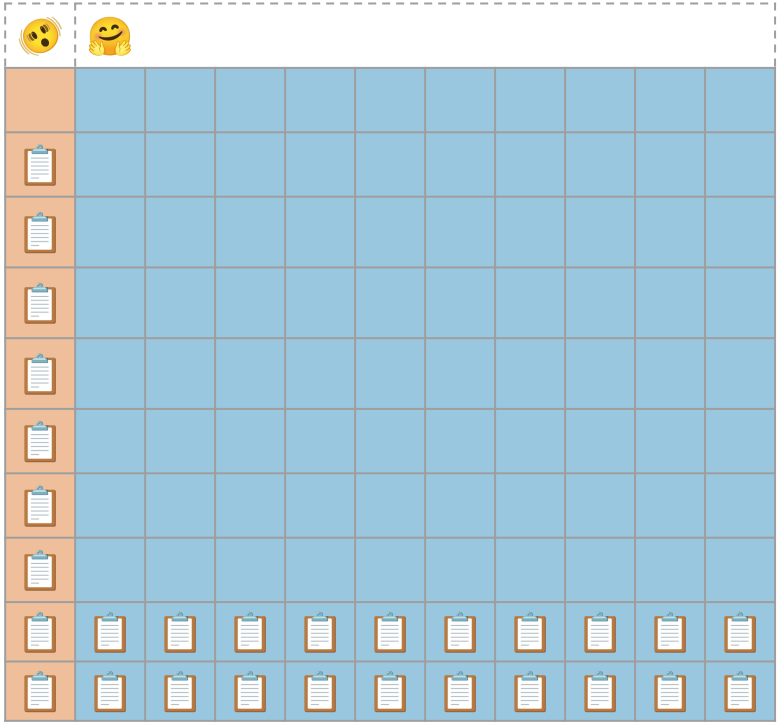
🦹 *That's easy, it's 9/29 ≈ .3*

That's not high at all... 🦸



Bayes’ theorem

P(H | D) = (P(H) × P(D | H)) / P(D)

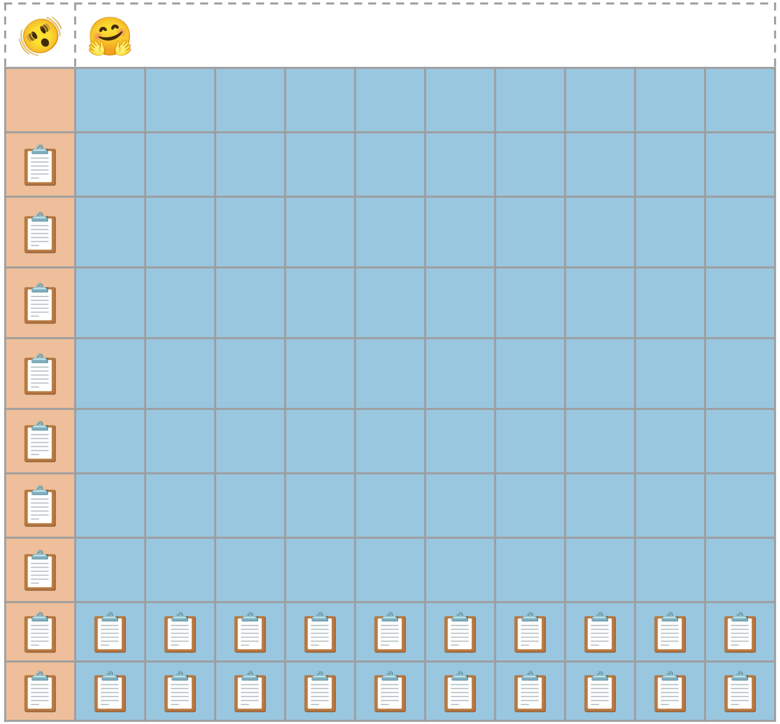


Prior

P(H | D) = (**P(H)** × P(D | H)) / P(D)

P(H) = ▇ / (▇ + ▇) = 10 / (10 + 100) = .09

🦹 *This would be the probability of ADHD*



Likelihood

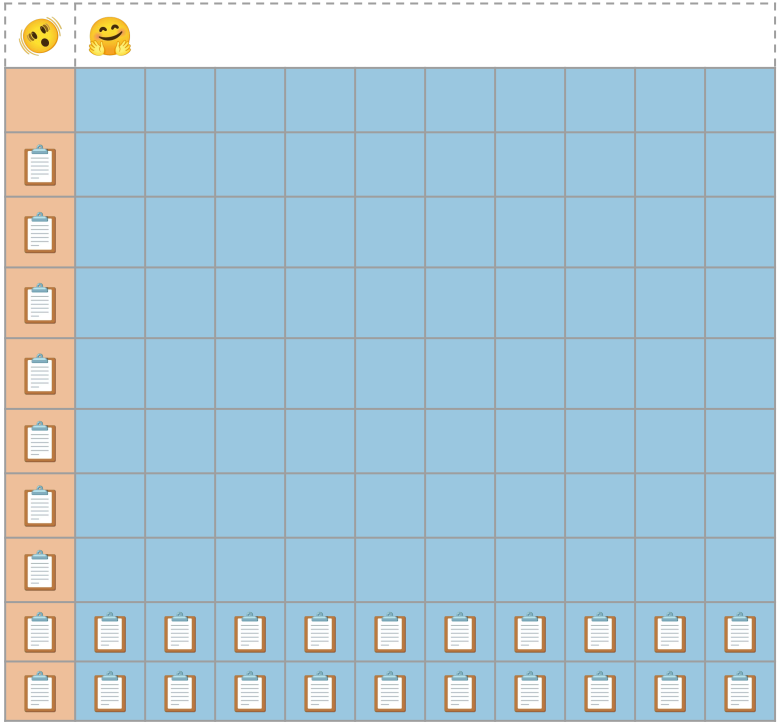
P(H | D) = (P(H) × **P(D | H)**) / P(D)

P(H) = 🟧 / (🟧 + 🟦) = 10 / (10 + 100) = .09

P(D | H) = 📋 / 🟧 = 9 / 10 = .9

🦹 *The probability of a positive test result, given ADHD*

Weird, it somehow reminds me of frequentist statistics 🦸



Prior × likelihood

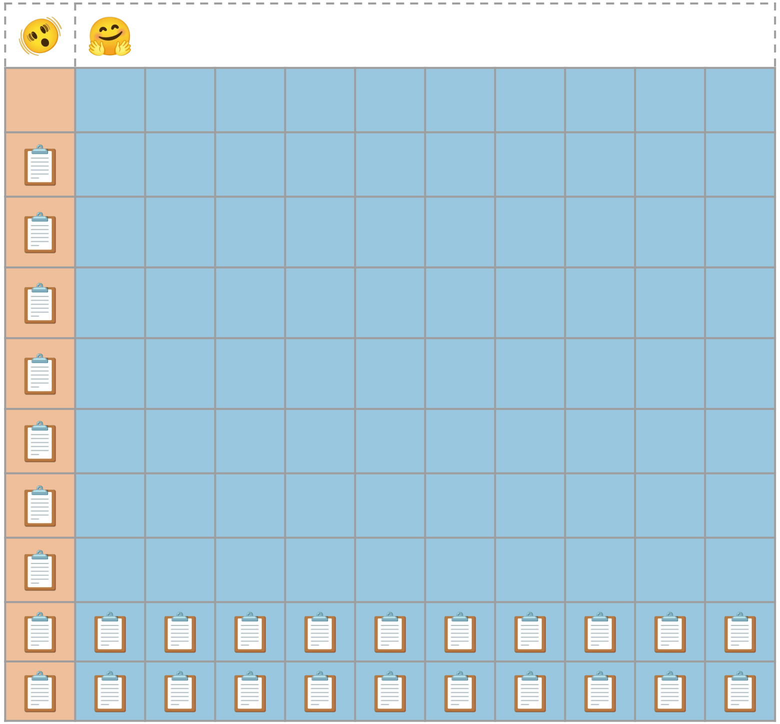
P(H | D) = (**P(H) × P(D | H)**) / P(D)

P(H) = 🟧 / (🟧 + 🟦) = 10 / (10 + 100) = .09

P(D | H) = 📋 / 🟧 = 9 / 10 = .9

P(H) × P(D | H) = 📋 / (🟧 + 🟦) = .09 × .9 = .08

🦹 *The probability of a positive test result **and** ADHD*



Remember the product rule?

Marginal likelihood

P(H | D) = (P(H) × P(D | H)) / **P(D)**

P(H) = 🟧 / (🟧 + 🟦) = 10 / (10 + 100) = .09

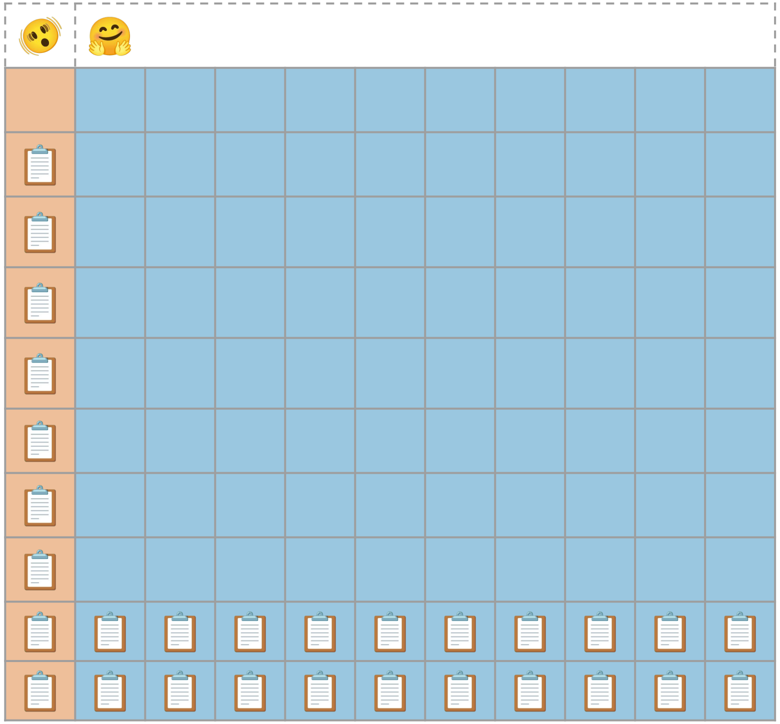
P(D | H) = 📋 / 🟧 = 9 / 10 = .9

P(H) × P(D | H) = 📋 / (🟧 + 🟦) = .09 × .9 = .08

P(D) = (P(H) × P(D | H)) + (P(¬H) × P(D | ¬H)) =

.08 + .18 = .26

🦹 *The probability of a positive test result*



Posterior

P(H | D) = (P(H) × P(D | H)) / P(D) = .08/.26 ≈ .3

P(H) = 🟧 / (🟧 + 🟦) = 10 / (10 + 100) = .09

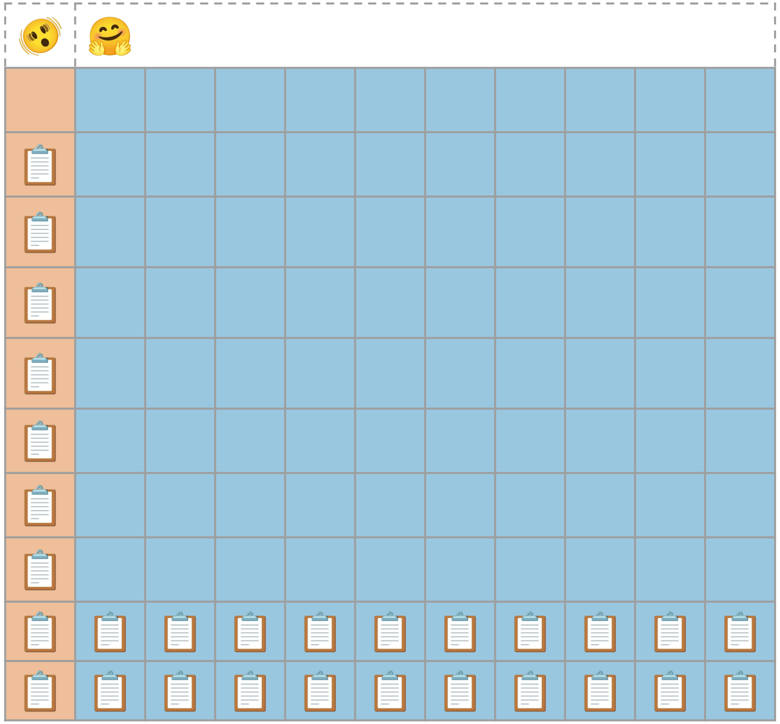
P(D | H) = 📋 / 🟧 = 9 / 10 = .9

P(H) × P(D | H) = 📋 / (🟧 + 🟦) = .09 × .9 = .08

P(D) = (P(H) × P(D | H)) + (P(¬H) × P(D | ¬H)) =

.08 + .18 = .26

🦹 *The probability of someone with ADHD, given a positive test result.*



Posterior

P(H | D) = (P(H) × P(D | H)) / P(D) = .08/.26 ≈ .3

P(H) = / (+) = 10 / (10 + 100) = .09

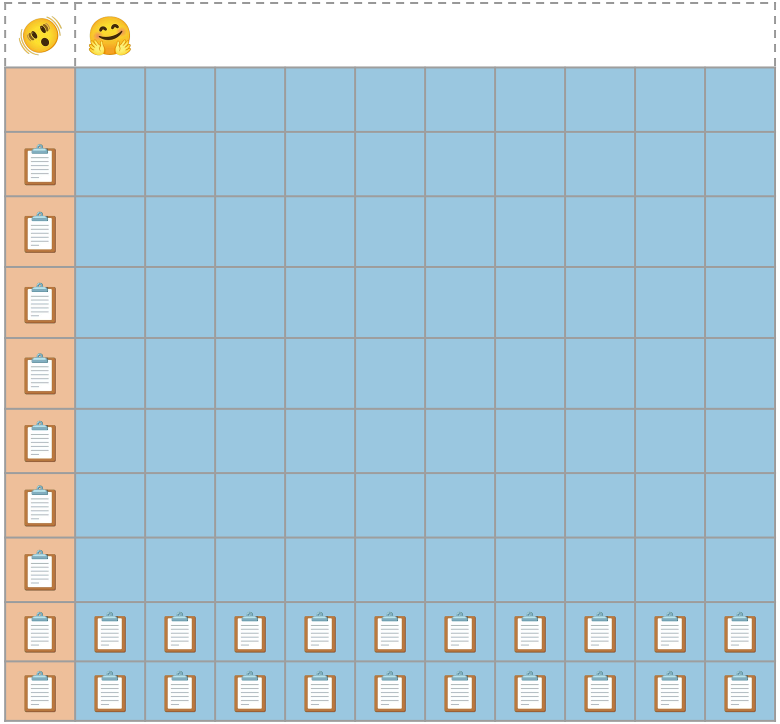
P(D | H) = / = 9 / 10 = .9

P(H) × P(D | H) = / (+) = .09 × .9 = .08

P(D) = (P(H) × P(D | H)) + (P(¬H) × P(D | ¬H)) =

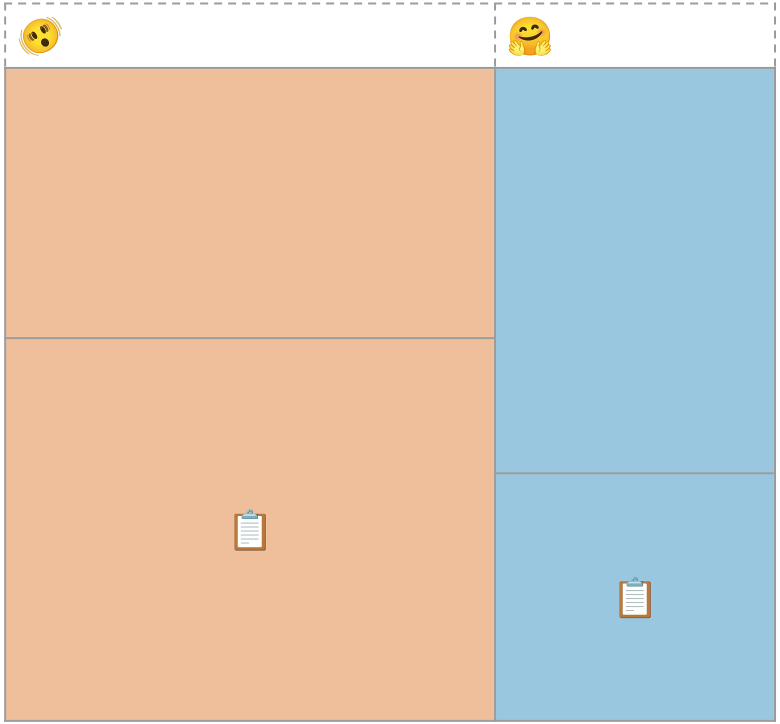
.08 + .18 = .26

Remember the rephrased question "Out of the number of people who test positive, how many have ADHD?", it's the exact same!



Bayes’ theorem

$P(H_{🫨} \mid D) = P(H_{🫨}) \times P(D \mid H_{🫨}) / P(D) \approx .3$
$P(H_{🤗} \mid D) = P(H_{🤗}) \times P(D \mid H_{🤗}) / P(D) \approx .7$ (‘null’ hypothesis)



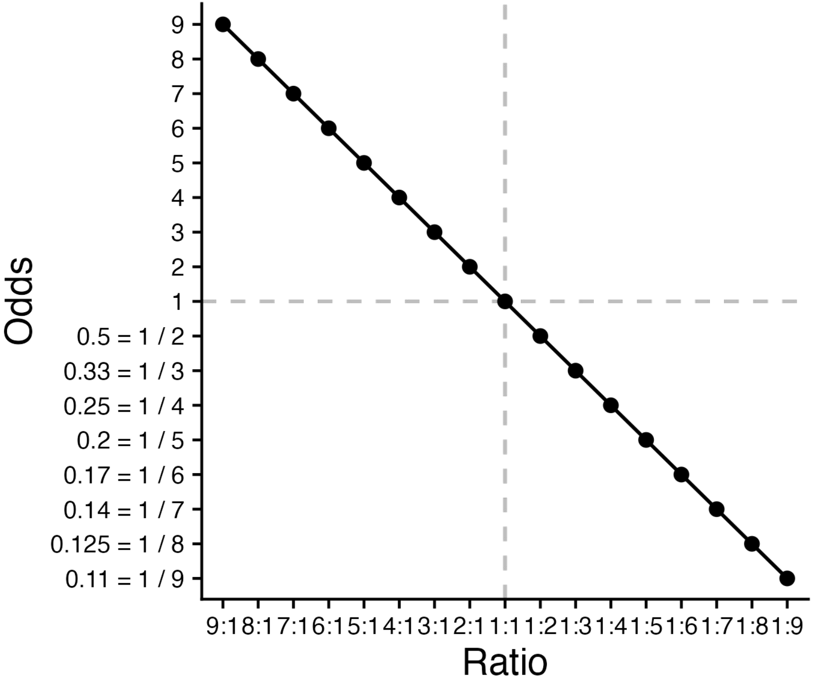
Bayes Factor

comparing the evidence for two hypotheses

Dynamic Textures #382 by Claus O. Wilke

Bayes factor

> "The Bayes factor is a ratio of two competing statistical models represented by their evidence, and is used to quantify the support for one model over the other."
>
> — [Wikipedia]



That sounds a lot like an odds ratio 🦸‍♀️

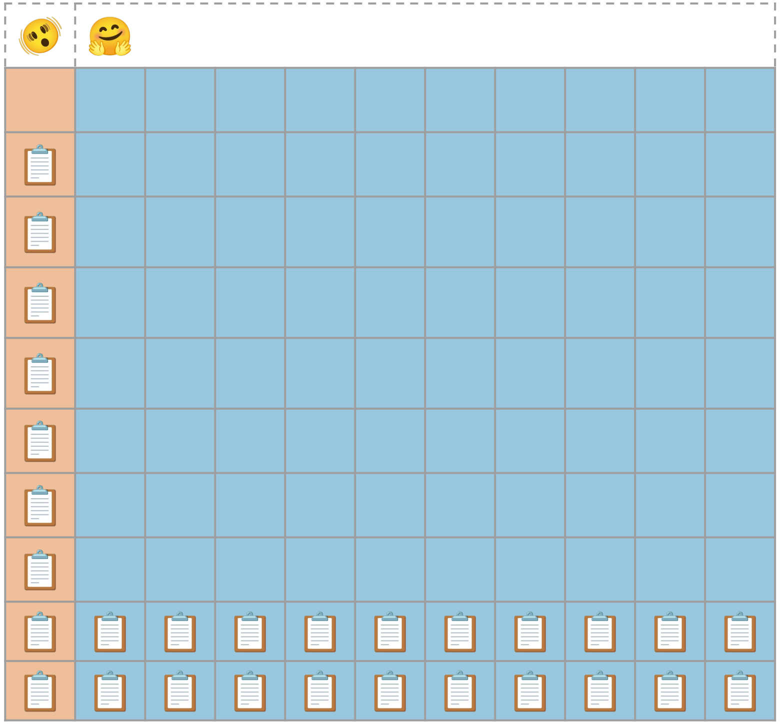
Bayes factor

K = P(D | H🫨) / P(D | H🤗) = ?

P(H | D) = (P(H) × **P(D | H)**) / P(D)
P(D | H🫨) = 📋 / 🟧 = 9 / 10 = .9
P(D | H🤗) = 📋 / 🟦 = 20 / 100 = .2

K = P(D | H🫨) / P(D | H🤗) = 4.5

- Continuous degree of evidence (vs. all-or-none)
- Monitor evidence during data collection
- Evidence of absence (data support a null effect) and absence of evidence (data are not informative)



Bayes factor interpretation

Bayes factor	Interpretation
>100	Decisive evidence for H_A
30–100	Very strong evidence for H_A
10–30	Strong evidence for H_A
3–10	Substantial evidence for H_A
1–3	Anecdotal evidence for H_A
1	No evidence
1/3–1	Anecdotal evidence for H_0
1/10–1/3	Substantial evidence for H_0
1/30–1/10	Strong evidence for H_0
1/100–1/30	Very strong evidence for H_0
<1/100	Decisive evidence for H_0

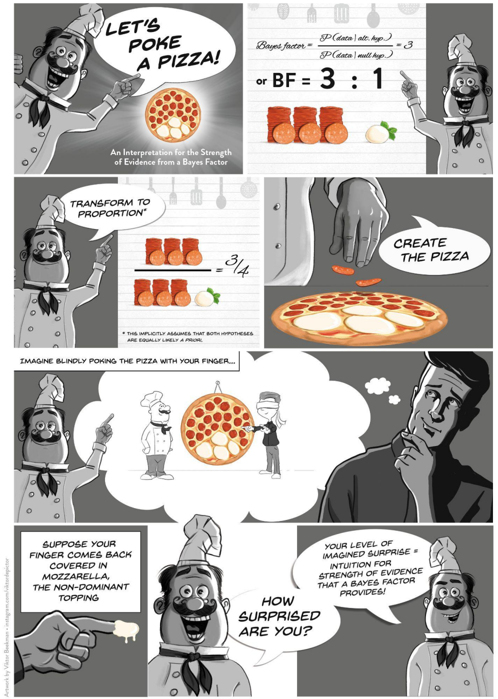


Illustration by [Eric-Jan Wagenmakers & Viktor Beekman]

Bayesian Parameter Estimation

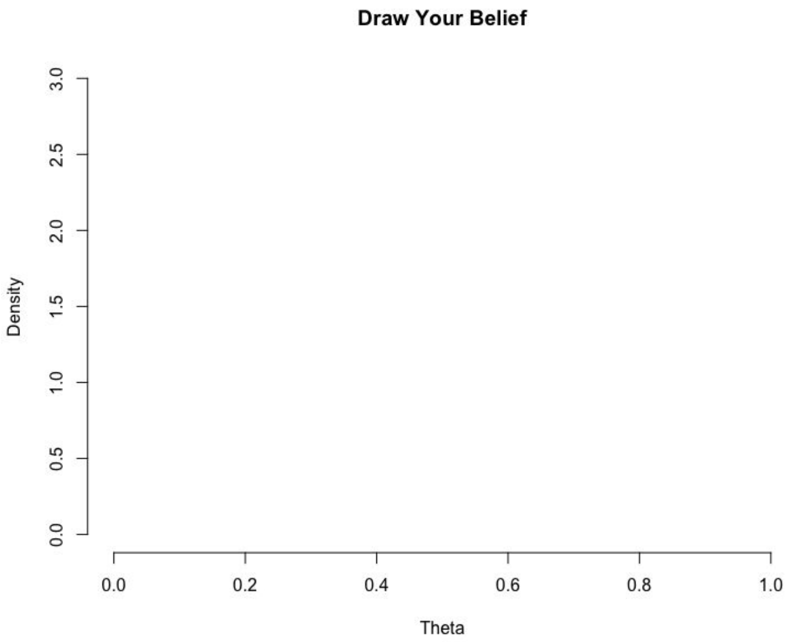
Dynamic Textures #382 by Claus O. Wilke 🎨

🎲 Prior belief

🪙: theta = P(Heads)

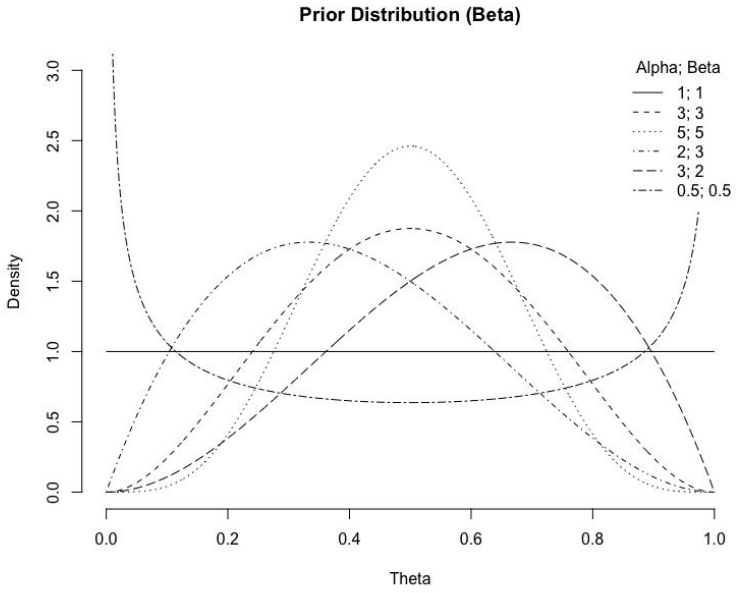
“Which values of theta *are good estimates (H), given my data (D)?”*

P(H | D) = (P(H) × P(D | H)) / P(D)



💡 How to get [fair results from a biased coin]?

Prior distribution P(H)



💡 Beta distribution

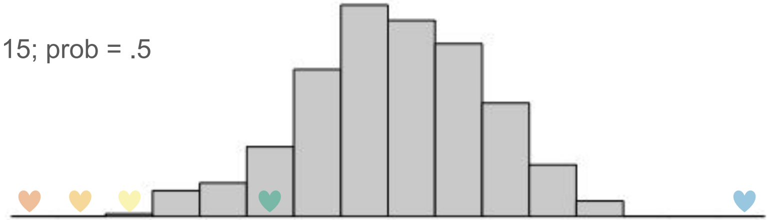
- Prior belief/information
- Uninformative/informative
- Weakly/strongly informative
- Skeptical
- Point-valued

Quote of the week

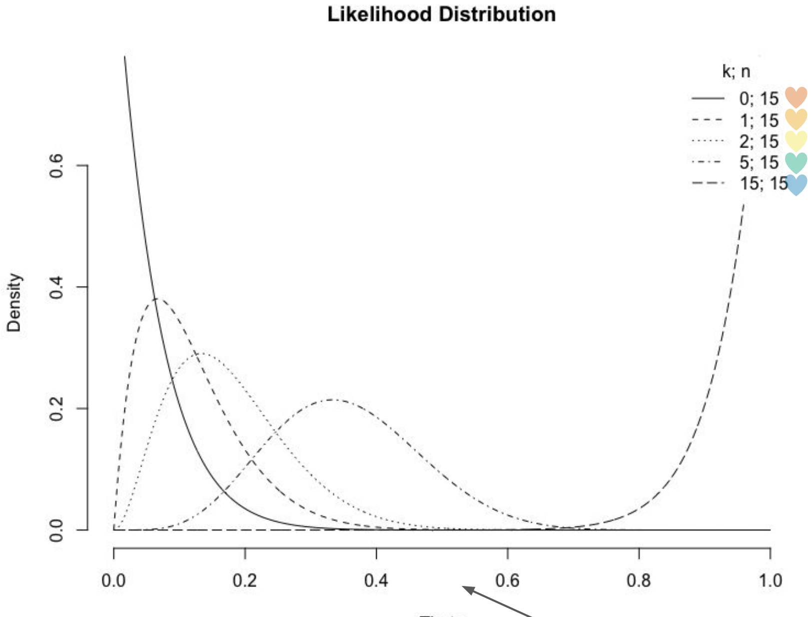
Illustration by [Eric-Jan Wagenmakers & Viktor Beekman]

Likelihood distribution P(D | H)

n = 15; prob = .5



```
dbinom(x = k, size = n, prob = seq(0, 1,
.01))
```



prob = .5

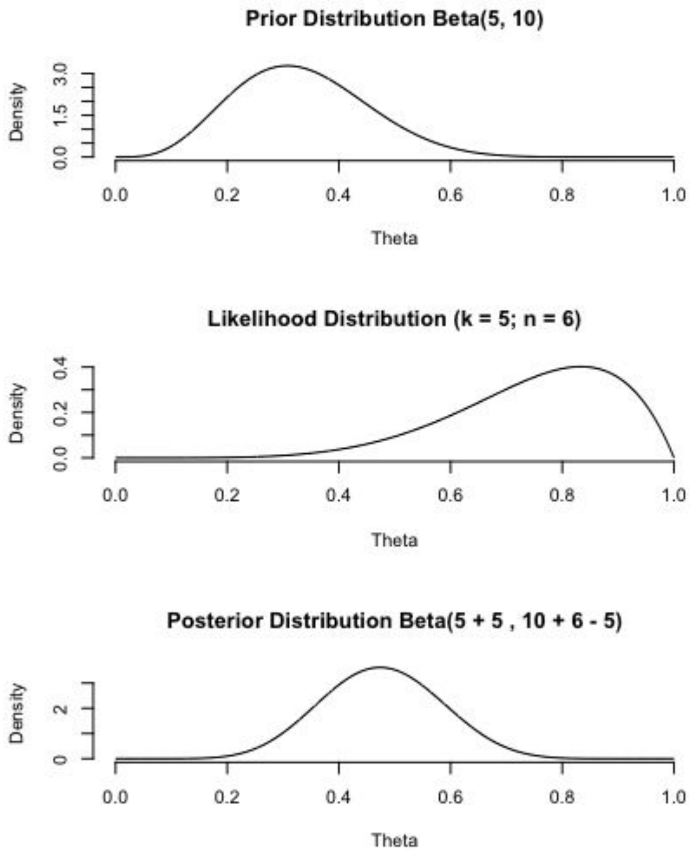
Posterior distribution P(H | D)

P(H | D) = (P(H) × P(D | H)) / P(D)

P(D) (marginal likelihood distribution): “integral of doom” 💀

But, beta distribution is *conjugate prior* of binomial distribution:

For a prior theta~Beta(a,b) and data *k* and *N*, the posterior is theta~Beta(a+k,b+N−k).



Quote of the week

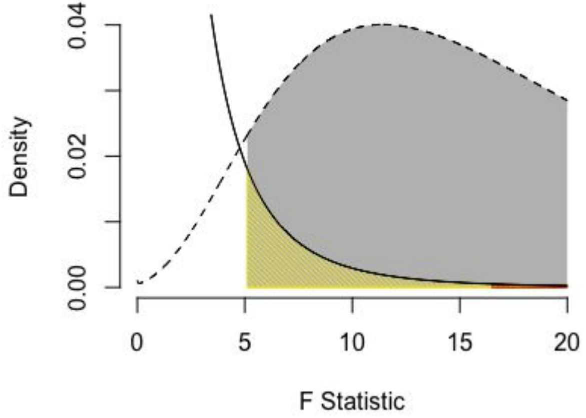
Illustration by Eric-Jan Wagenmakers & Viktor Beekman

Frequentist ⚔️ Bayesian

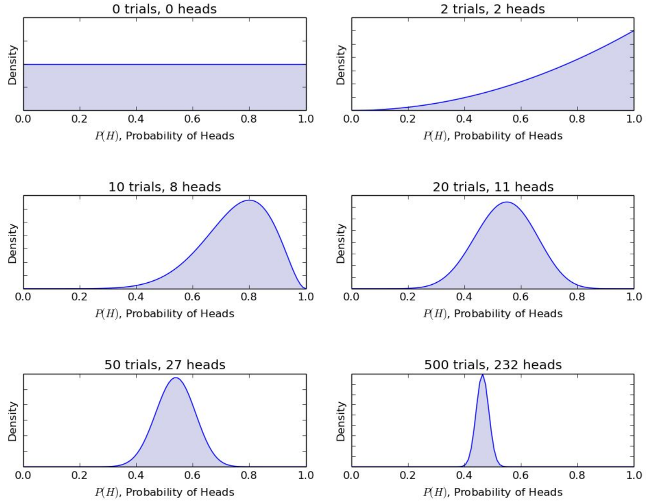
Dynamic Textures #382 by Claus O. Wilke 🎨

🔁 Recap

Frequentist inference: What's the probability of the data, or more extreme data, given the null hypothesis?



Bayesian inference: What's the probability of a hypothesis, given the data?



Some key differences

Probability

- the long-term frequency of events

Parameters

- fixed and unknown, point estimate

Data

- random, subject to sampling variability

Hypothesis

- reject or do not reject H_0

Probability

- a degree of belief

Parameters

- random, probability distribution

Data

- fixed

Hypothesis

- probability of H

Pros & cons

Computation

- ❏ less intensive (often analytical solutions)

Repeated testing

- ❏ increases type I error (or use e-value)

Evidence for H0

- ❏ no (but model comparison with likelihood ratio tests)

Support

- ❏ well-supported

Computation

- ❏ intensive techniques to approximate posterior distributions

Repeated testing

- ❏ no type I error (unless Bayes factor cutoff): fishing, stopping when criterion is used, etc.

Evidence for H0

- ❏ yes

Support

- ❏ increasingly supported (brms, JASP)

Similar results, but

Both give highly similar results, but can diverge with:

- informative priors
- small samples
- complex models
- optional stopping

Cooling Down

Dynamic Textures #382 by Claus O. Wilke 🎨

🥡 Takeaways

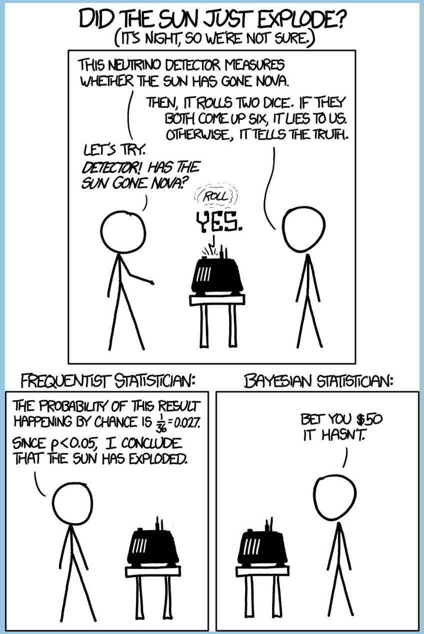


Illustration by [Randall Munroe](#) ([wtf](#) / [🔥](#))

Takeaways

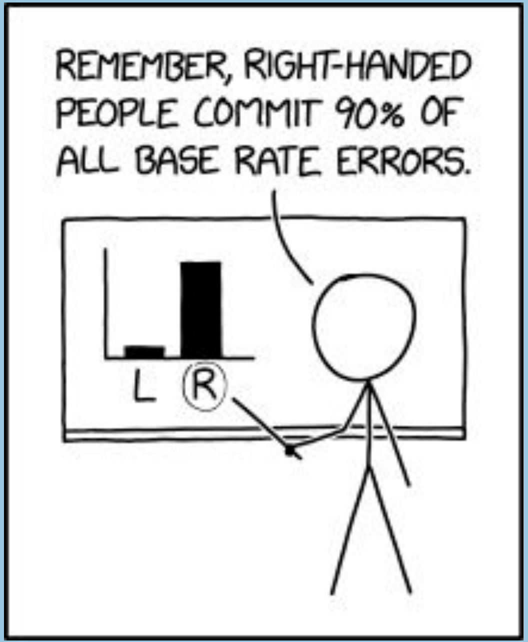


Illustration by Randall Munroe (wtf)

Takeaways

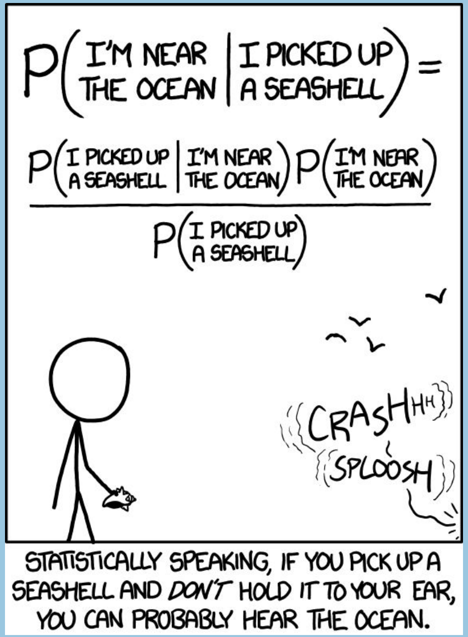


Illustration by Randall Munroe (wtf)

Takeaways

Perspectives on Psychological Science
Volume 6, Issue 3, May 2011, Page 313

https://doi.org/10.1177/1745691611406928

SAGE journals

Bayesians Caught Smuggling Priors Into Rotterdam Harbor

Rogier A. Kievit

Department of Psychology, University of Amsterdam, The Netherlands

Corresponding author(s):
Roger A. Kievit, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018WB Amsterdam, The Netherlands E-mail: r.a.kievit@uva.nl

Amsterdam, April 13, 2011. A group of international Bayesians was arrested today in the Rotterdam harbor. According to Dutch customs, they were attempting to smuggle over 1.5 million priors into the country, hidden between electronic equipment. The arrest represents the largest capture of priors in history.

"This is our biggest catch yet. Uniform priors, Gaussian priors, Dirichlet priors, even informative priors, it's all here," says customs officers Benjamin Roosken, responsible for the arrest. "There are priors for memory experiments, intelligence tests, flanker tasks, meta-analyses, political preference, everything! God only knows what would have happened if this had gotten through. We're pretty lucky to catch them too. The chance of being in the right place, given the right time, if you take into account the number of arrests, divided by the number of successful arrests every year, it's pretty slim. We're very glad indeed."

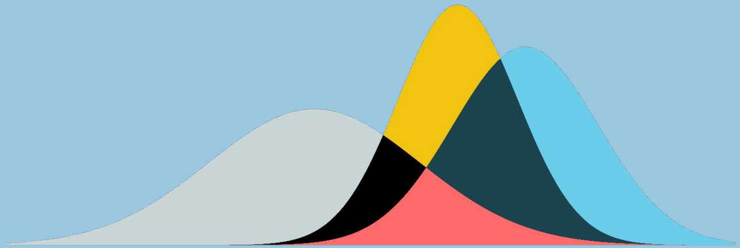
Sources suggest that the shipment of priors was going to be introduced into the Dutch scientific community by "white-washing" them. "They are getting very good at it. They found ghost-journals with fake articles, refer to the papers where the priors are allegedly based on empirical data, and before you know it, they're out in the open. Of course, when you look up the reference, everything is long gone," says Roosken.

Until recently, the Dutch government adopted a lenient, pragmatic approach toward priors. As an anonymous source states, "It was quite simple. Scientists were allowed to *use* priors, but not to create them at home. It may sound a bit counterintuitive, but it worked quite well, for a while at least." However, according to critics, this policy created an uncontrollable backdoor industry.

The discovery of international smuggling rings has caused the government to revise its strategy and crack down hard on illegal trade. The capture of the smuggling ring symbolizes a new, tough stance on priors. "We will not stand for this unjustified and illegal use of priors any longer," says Roosken. If found guilty, the defendants may face 12 years in prison (95% CI [10.2, 13.8], $p < .01$).

Kievit, 2011 🔒

Takeaways



Illustrations by [Kristoffer Magnusson]

Frequentist
Significance Testing

Bayesian
Inference

🥇 Exam(ple) question

Sir Ronald Fisher test positief op Covid en vraagt zich nu af hoe hoog de kans is dat hij ook daadwerkelijk besmet is met het virus. Hij heeft een aantal feiten opgezocht.

- Van alle mensen test 1 op de 5 positief.
- Van de mensen die besmet zijn met Covid test 9 op de 10 positief.
- 1 op de 1000 mensen heeft Covid.

Wat is nu de kans dat Sir Ronald Fisher besmet is met het Covid virus?

A. .001
B. .0045
C. .2
D. .9

🗃️ Look here!

Bayes theorem

- Explanation video (3Blue1Brown)
- Product rule (prior × likelihood) (3Blue1Brown)
- Interactive visualization (Seeing Theory)
- Redefining Bayes rule (3Blue1Brown)

Prior and posterior

- Interactive visualization (Seeing Theory).

Likelihood

- Interactive visualization (Seeing Theory).

Bayesian inference

- Web simulation (Kristoffer Magnusson)
- Brief introduction (Johnny van Doorn)

Bayesian meets frequentist

- Side-by-side Shiny app (John Kruschke)
- Video tutorial for Shiny app (Eero Liski)

👶 Bayesian thinking for toddlers (Eric-Jan Wagenmakers)

✍️ Colophon

Slides

alexandersavi.nl/teaching/

License