Tutorial: Bayesian statistics (part 1)

Filip Melinscak

17.4.2019

University of Zurich[UZH]

Psychiatrische Universitätsklinik Zürich

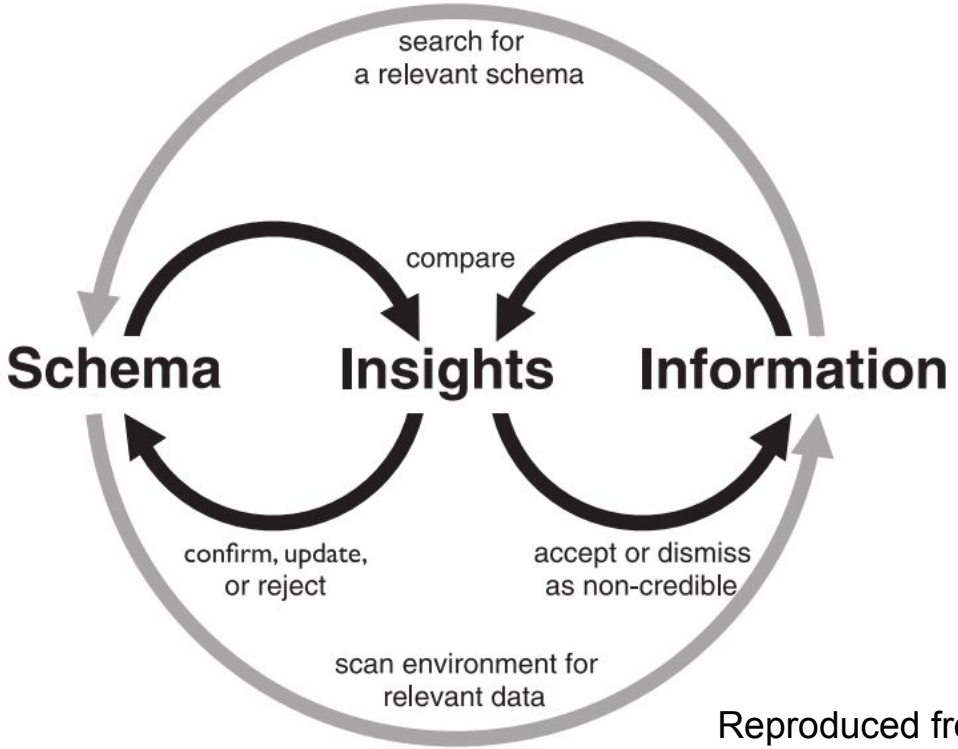
Outline

We will use the bottom-up approach, starting from principles, and ending on practical applications.

- Why we need statistics?
- Which statistics we need?
  - Frequentist and Bayesian interpretation of probability
  - Common misconceptions about frequentist statistics
  - Argument against frequentist and in favor of Bayesian statistics
- Bayesian inference
  - Conceptual foundation: probability theory
  - Bayes' rule in the discrete case
  - Bayes' rule in the continuous case (parameter estimation)
  - Bayesian model selection

Why we need statistics?

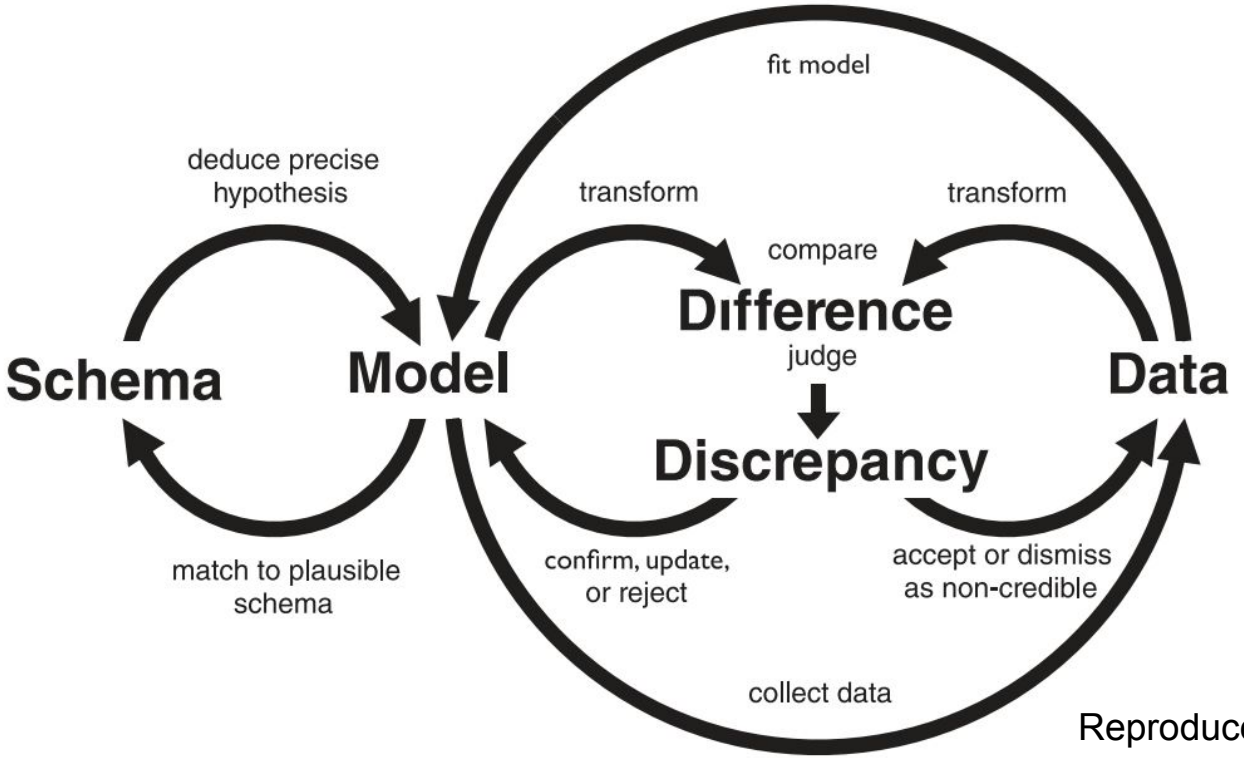
Sensemaking process



Reproduced from [Gro14]

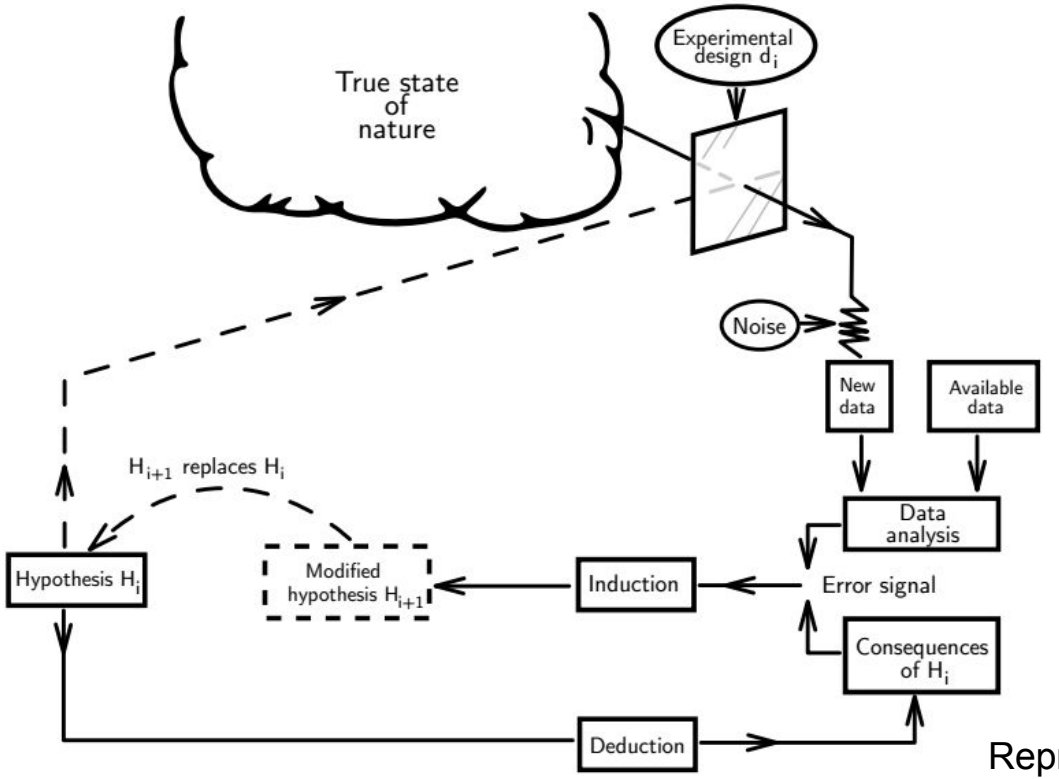
Why is this insufficient for quantitative data analysis?
Why do we need statistics?

From sensemaking to data analysis (statistics)



Reproduced from [Gro14]

From statistics to scientific inquiry



Reproduced from [Box76]

Role of statistics in science

- To borrow from John Maynard Keynes:

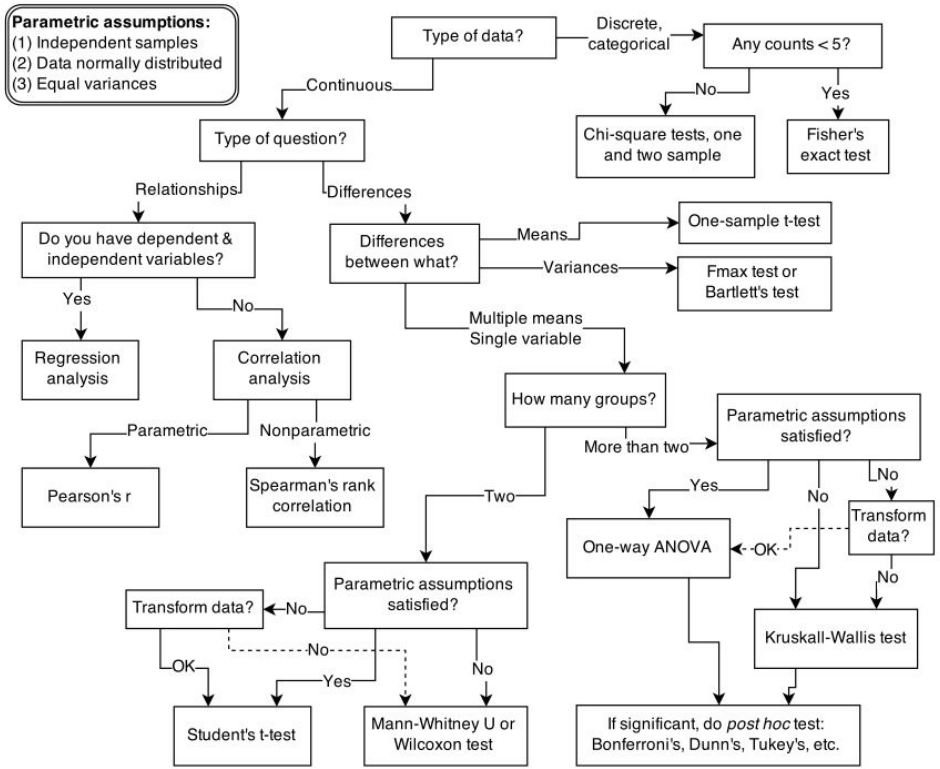
> “The ideas of ~~economists~~ statisticians and ~~political philosophers~~ philosophers of science, both when they are right and when they are wrong, are more powerful than is commonly understood. [...] Practical men, who believe themselves to be quite exempt from any ~~intellectual influence~~ statistical philosophy, are usually the slaves of some defunct ~~economist~~ methodologist.”

- Statistics as the “grammar of science”:

> “The unity of all science consists alone in its method, not in its material.” - Karl Pearson

- Distinct roles of applied statistics and applied mathematics (cf. the relationship of chemistry and cooking)

“Role of statistics in science”



Reproduced from [McE18]

- The misconceptions about the role of statistics:
  - Provides objective rules for analyzing data
  - Allows us to determine the truth of scientific claims
- This “decision tree” view of statistics obscures the unity of all statistics and makes it more difficult to learn; inherently inflexible

How does statistical philosophy influence our work?

- IMO, your statistical philosophy (knowingly or not) largely influences **all** parts of your scientific workflow:
  - Types of theories and questions you test (if you only know NHST, you will only ask questions that can be answered by NHST)
  - Experimental design
  - Data analyses and model checking
  - Interpretation of results (both yours and from literature)
  - Scientific communication (data and model visualization, framing of results)
- Bayesian inference provides you with a **principled way of thinking** of all the components above

Which statistics we need?

Two interpretations of probability

- **Probability theory**: mathematical rules for manipulating probabilities
- Probability theory largely uncontroversial, but its correspondence to the real world is; **two interpretations of probability**:
  - **Aleatory/frequentist probability**: expected frequency over many repetitions of a procedure
  - **Epistemic/Bayesian probability**: degree of belief agent should assign to event or proposition (inherently dependent on the agent's state of knowledge)
- **Why this matters**?
  - Aleatory probability does not apply to singular events or propositions (e.g. **this** hypothesis is true, **this** effect exists)
  - Epistemic probability applies both to singular and repetitive events
  - **Correct interpretation of probability statements is crucial for making sound inferences**

Testing our intuitions about p-values

Suppose you have a treatment that you suspect may alter performance on a certain task. You compare the means of your control and experimental groups (say 20 subjects in each sample). Further, suppose you use a simple independent means t-test and your result is (t = 2.7, d.f. = 38, p = 0.01). Please mark each of the statements below as “true” or “false”. “False” means that the statement does not follow logically from the above premises. Also note that several or none of the statements may be correct.

1) You have absolutely disproved the null hypothesis (that is, there is no difference between the population means). [ ] true / false [ ]

2) You have found the probability of the null hypothesis being true. [ ] true / false [ ]

3) You have absolutely proved your experimental hypothesis (that there is a difference between the population means). [ ] true / false [ ]

4) You can deduce the probability of the experimental hypothesis being true. [ ] true / false [ ]

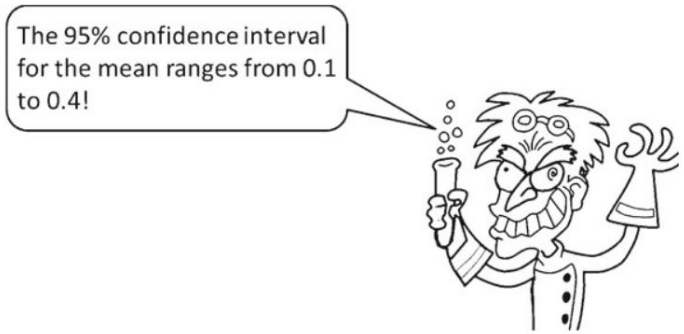
5) You know, if you decide to reject the null hypothesis, the probability that you are making the wrong decision. [ ] true / false [ ]

6) You have a reliable experimental finding in the sense that if, hypothetically, the experiment were repeated a great number of times, you would obtain a significant result on 99% of occasions. [ ] true / false [ ]

Reproduced from [Gig04]

Testing our intuitions about confidence intervals

Professor Bumbledorf conducts an experiment, analyzes the data, and reports:



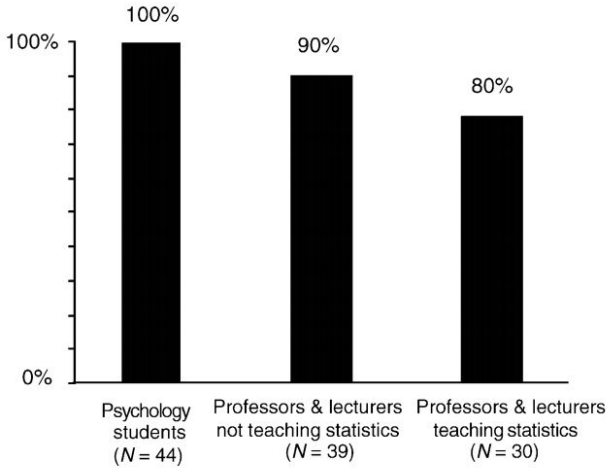
Please mark each of the statements below as "true" or "false". False means that the statement does not follow logically from Bumbledorf's result. Also note that all, several, or none of the statements may be correct:

1.	The probability that the true mean is greater than 0 is at least 95%.	☐True	☐False
2.	The probability that the true mean equals 0 is smaller than 5%.	☐True	☐False
3.	The "null hypothesis" that the true mean equals 0 is likely to be incorrect.	☐True	☐False
4.	There is a 95% probability that the true mean lies between 0.1 and 0.4.	☐True	☐False
5.	We can be 95% confident that the true mean lies between 0.1 and 0.4.	☐True	☐False
6.	If we were to repeat the experiment over and over, then 95% of the time the true mean falls between 0.1 and 0.4.	☐True	☐False

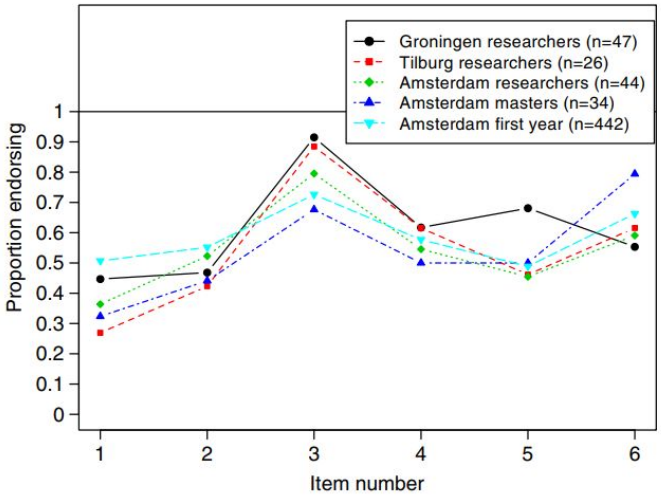
Reproduced from [Hoe14]

Testing our intuitions about frequentist results

- If you endorsed any of the previous statements, **the good news** is that you are in good company!



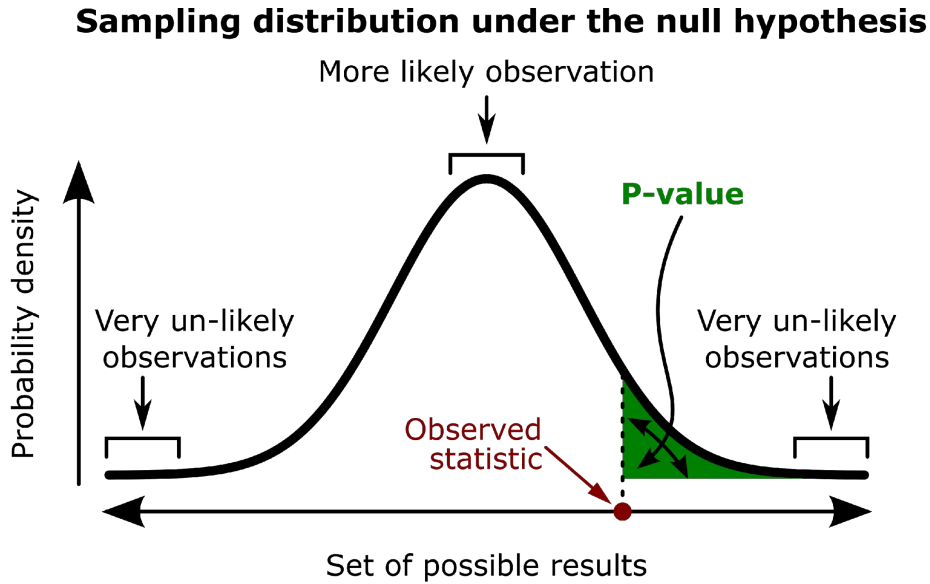
Reproduced from [Gig04]



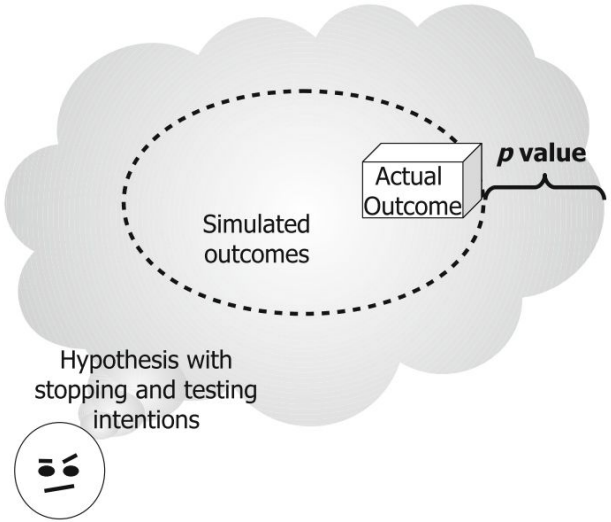
Reproduced from [Hoe14]

- **The bad news** is that none of the statements are correct!

Reminder on p-values



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.



Reproduced from [Kru18]

Argument for Bayesian statistics

The philosophical argument in favor of Bayesian statistics is straightforward [Lin00]:

1. Statistics is the **study of uncertainty**
2. Uncertainty should be measured by **probabilities**, which are manipulated using probability calculus (sum and product rules)
3. Probabilities can be used to describe the **uncertainty of data**
4. Probabilities can also be used to describe the **uncertainty of (hidden) parameters**
5. Statistical inference should be performed according to the rules of probability calculus (i.e. the **Bayes' rule**)

Advantages of Bayesian inference

	Bayesian Inference	Classical Inference
Desiderata for Parameter Estimation		
1. To incorporate prior knowledge	✔	✗
2. To quantify confidence that θ lies in a specific interval	✔	✗
3. To condition on what is known (i.e., the data)	✔	✗
4. To be coherent (i.e., not internally inconsistent)	✔	✗
5. To extend naturally to complicated models	✔	✗
Desiderata for Hypothesis Testing		
1. To quantify evidence that the data provide for $\mathcal{H}_0$ vs. $\mathcal{H}_1$	✔	✗
2. To quantify evidence in favor of $\mathcal{H}_0$	✔	✗
3. To allow evidence to be monitored as data accumulate	✔	✗
4. To not depend on unknown or absent sampling plans	✔	✗
5. To not be "violently biased" against $\mathcal{H}_0$	✔	✗

Reproduced from [Wag18]

- Common misconception: Bayesian inference is modern/advanced/difficult to understand, whereas frequentist inference is established/easy (Bayesian *computation* can be difficult, but there is software to help here)
- IMO, framing problems in Bayesian terms is conceptually simple, and the interpretation of results is straightforward

Bayesian inference

Foundation of Bayesian inference: probability theory

- We need only a **few rules from probability theory**:
  - **Product** (multiplication/chain) **rule**: what is the probability of A **and** B?

$$P(A, B) = P(A)P(B|A)$$
$$= P(B)P(A|B)$$

  - **Sum** (addition) **rule**: what is the probability of A **or** B, if A and B are mutually exclusive?

$$P(A \cup B) = P(A) + P(B)$$

  - **Total probability rule / "extending the conversation":** assume we have a **disjoint set** {A_1, A_2, ... A_K} (set of mutually exclusive events, of which one is true, e.g. {A, not A}), we can express of probability of B as a sum of joint probabilities with A_k events

$$P(B) = P(B, A) + P(B, \neg A)$$
$$P(B) = \sum_{k=1}^{K} P(B, A_k) = \sum_{k=1}^{K} P(B|A_k)P(A_k)$$

product rule

Conditioning on data (discrete case)

- Assume we are interested in the credibility of some hypothesis/model ‘M’ (and its negation ‘not M’), after observing data X
- Further, assume we know the probability of the data given M and notM (P(X|M) and P(X|notM), and the probability of M and notM before observing the data (P(M), P(notM))
- Given:

$$P(M), P(\neg M), P(X|M), P(X|\neg M)$$

- Desired:

$$P(M|X)$$

- We will now derive the Bayes’ rule, which will tell us how to go from the **prior** P(M) to the **posterior** P(M|X)

Bayes' rule in the discrete case (1)

- Let's write the product rule for X and M: $P(M, X) = P(M)P(X|M)$
- By symmetry, it is also true: $P(M, X) = P(X)P(M|X)$
- We can equate right-hand sides: $P(X)P(M|X) = P(M)P(X|M)$
- Finally, we rearrange so that we have what we want on LHS:

$$P(M|X) = \frac{P(X|M)P(M)}{P(X)}$$

- Rejoice, this is the **Bayes' rule**! But how do we compute the **prior predictive probability** of the data P(X)?

Bayes' rule in the discrete case (2)

- We can use the total probability rule ("extend the conversation") to get P(X):

$$\begin{aligned} P(X) &= P(X, M) + P(X, \neg M) \\ &= P(X|M)P(M) + P(X|\neg M)P(\neg M) \end{aligned}$$

- Finally, we can reformulate Bayes' rule in terms of probabilities we know:

$$P(M|X) = \frac{P(X|M)P(M)}{P(X|M)P(M) + P(X|\neg M)P(\neg M)}$$

- For more than two hypotheses:

$$P(M_i|X) = \frac{P(X|M_i)P(M_i)}{\sum_{k=1}^{K} P(X|M_k)P(M_k)}$$

Example: Bayes' rule for a truth-detecting-machine

- Imagine we have a machine that magically detects the truth of hypotheses we input: when the hypothesis is true, the machine is 80% accurate, when the hypothesis is false, the machine is 95% accurate
- Moreover, imagine that we are not very good at coming up with hypotheses that are actually true: only 10% of the time we input true hypotheses
- If we input a hypothesis, and we get a "TRUTH" reading from the machine, what is the probability that the hypothesis is **actually** true?

Example: Bayes' rule for a truth-detecting-machine

- Given:

$$P(T|H) = 0.80$$
$$P(F|\neg H) = 0.95 \Rightarrow P(T|\neg H) = 0.05$$
$$P(H) = 0.1$$
$$P(\neg H) = 1 - P(H) = 0.9$$
$$P(H|T) = ?$$

- We apply the Bayes' theorem:

$$P(H|T) = \frac{P(T|H)P(H)}{P(T|H)P(H) + P(T|\neg H)P(\neg H)} = \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.05 \times 0.9} = 0.64$$

After applying our truth-telling-machine we are only 64% sure of the truth of our hypothesis!

Example: Bayes' rule for a truth-detecting-machine

- What is the relative belief (i.e. **posterior odds**) of H vs. notH after observing T?

$$\frac{P(H|T)}{P(\neg H|T)} = \frac{\frac{P(T|H)P(H)}{P(T)}}{\frac{P(T|\neg H)P(\neg H)}{P(T)}}$$

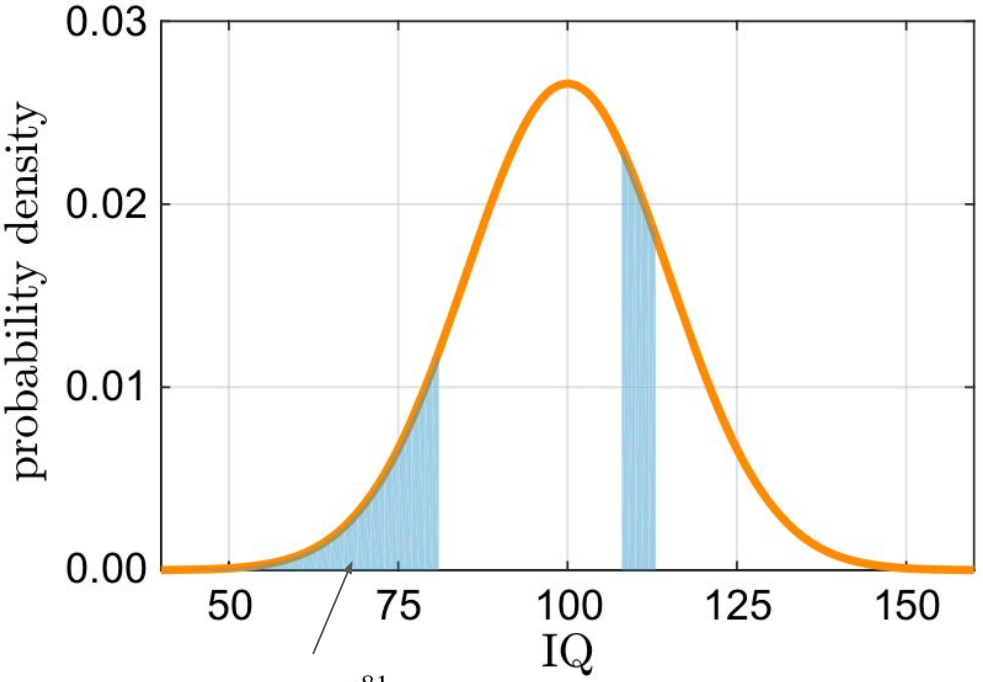
$$\underbrace{\frac{P(H|T)}{P(\neg H|T)}}_{\text{posterior odds}} = \underbrace{\frac{P(H)}{P(\neg H)}}_{\text{prior odds}} \times \underbrace{\frac{P(T|H)}{P(T|\neg H)}}_{\text{Bayes factor}}$$

$$= \frac{0.1}{0.9} \times \frac{0.8}{0.05} = 16/9$$

Bayes' rule in the continuous case (1)

- We go from probabilities to **probability density functions** (PDFs)
- By definition:

$$1 = \int_A p(a)da$$



$$P(a < 81) = \int_{-\infty}^{81} p(a)da$$

Reproduced from [Etz18]

Bayes' rule in the continuous case (2)

- To derive Bayes' rule, we first need the continuous **product rule**:

$$\begin{aligned} p(a,b) &= p(a)p(b|a) \\ &= p(b)p(a|b) \end{aligned}$$

- And the continuous **total probability rule** (i.e. **marginalization**, when reading right to left):

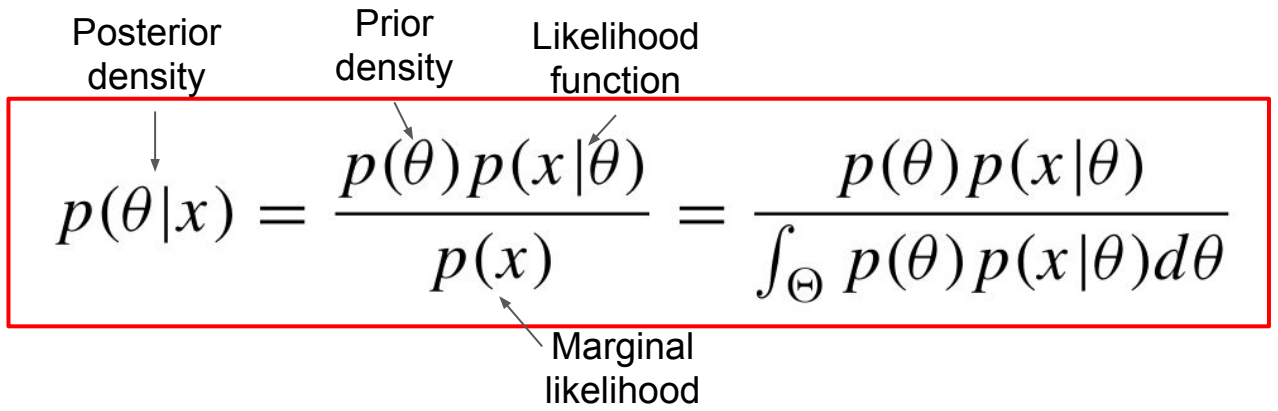
$$p(a) = \int_B p(a,b)db$$

- Bayes' rule is then:

$$\begin{aligned} p(a|b) &= \frac{p(a,b)}{p(b)} = \frac{p(b|a)p(a)}{p(b)} \\ &= \frac{p(b|a)p(a)}{\int_A p(b|a)p(a)da} \end{aligned}$$

Bayes' rule for parameter estimation

- Suppose we observed some data x produced by a stochastic process we are modeling as p(x| θ), where θ represents parameters of the process (e.g. p(x|θ) = N(x; μ, σ), where μ and σ are the parameters)
- **How can we calculate the credibility of parameter values given the data** (i.e. p(θ | x)? The answer is again Bayes' rule:



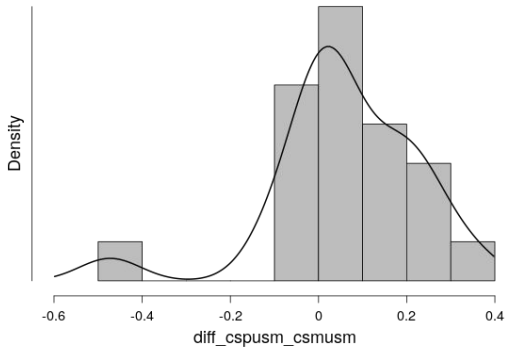
- The denominator usually does not have an analytic solution, so we have to use approximations

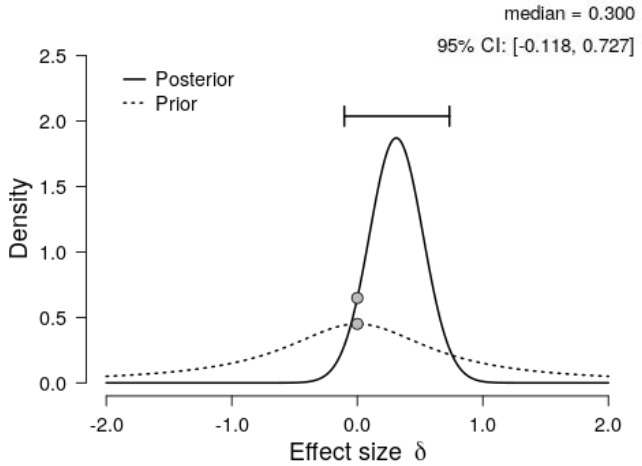
Example: Bayesian parameter estimation of CS+/CS- difference with JASP

- JASP is a free, open-source alternative to SPSS that supports both classical and Bayesian analyses
- We will analyze the SCR CS+/CS- difference for 21 subjects using the Bayesian one sample t-test

Descriptive Statistics

	diff_cspusm_csmusm
Valid	21
Missing	5
Mean	0.05930
Std. Deviation	0.1712
Minimum	−0.4727
Maximum	0.3522





We used the Cauchy prior with the default scale parameter of 0.707.

Bayesian model selection

- If we have two or more models under consideration we can do two types of inference: **continuous within-model** (parameter estimation), and **discrete between-model** (model selection) inference
- Parameter selection can be done independently for each model:

$$p(\theta|X, \mathcal{M}_1) = \frac{p(\theta|\mathcal{M}_1)P(X|\theta, \mathcal{M}_1)}{\int_\Theta p(\theta|\mathcal{M}_1)P(X|\theta, \mathcal{M}_1)d\theta}$$

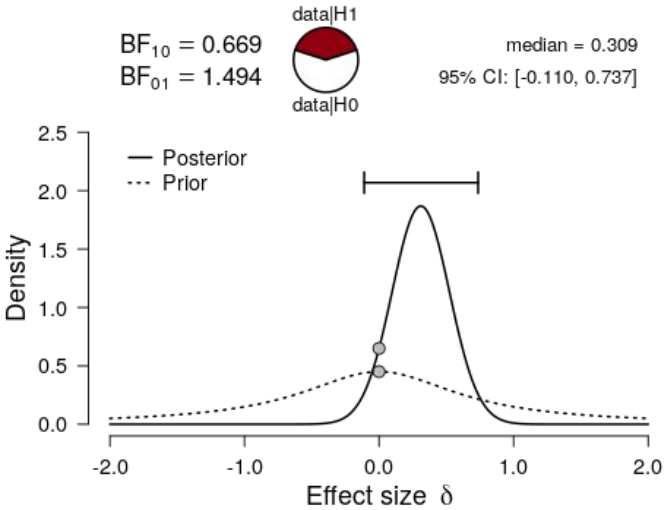
- Models can be compared using posterior odds and Bayes factors:

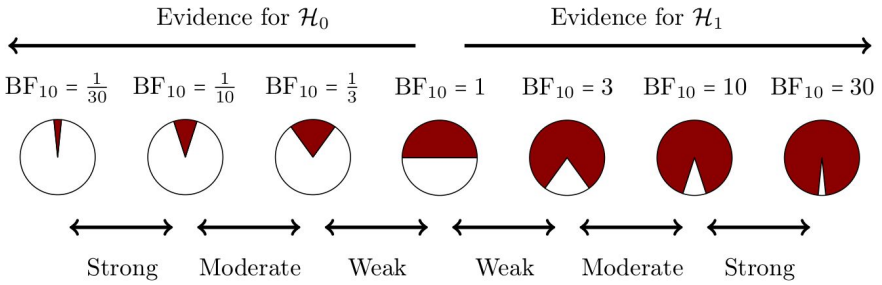
Posterior odds | Prior odds | Bayes factor

$$\frac{P(\mathcal{M}_1|X)}{P(\mathcal{M}_0|X)} = \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_0)} \times \frac{P(X|\mathcal{M}_1)}{P(X|\mathcal{M}_0)}$$

Example: hypothesis testing of CS+/CS- difference

- Again we use JASP with the same data as before; but now we do not want to also compare the model of no effect existing vs. a model predicting the effect exists

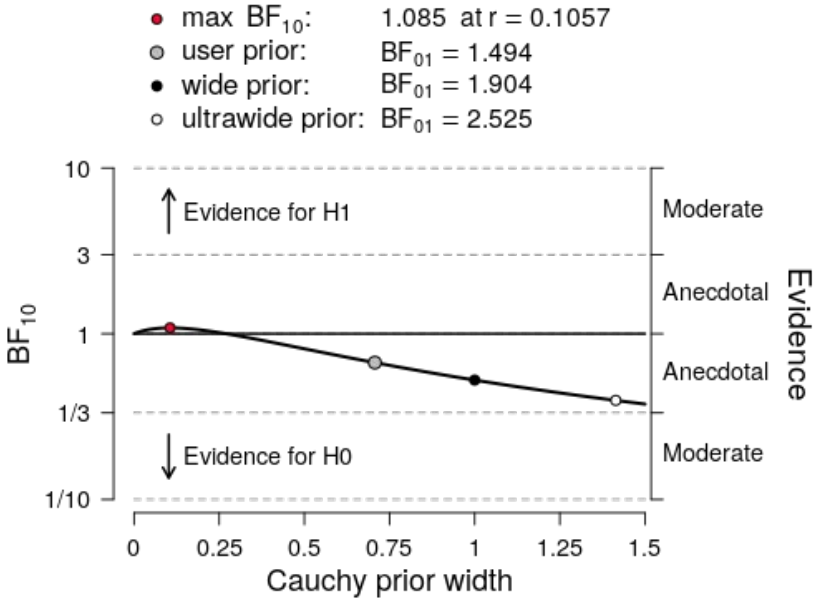




Reproduced from [vDo19]

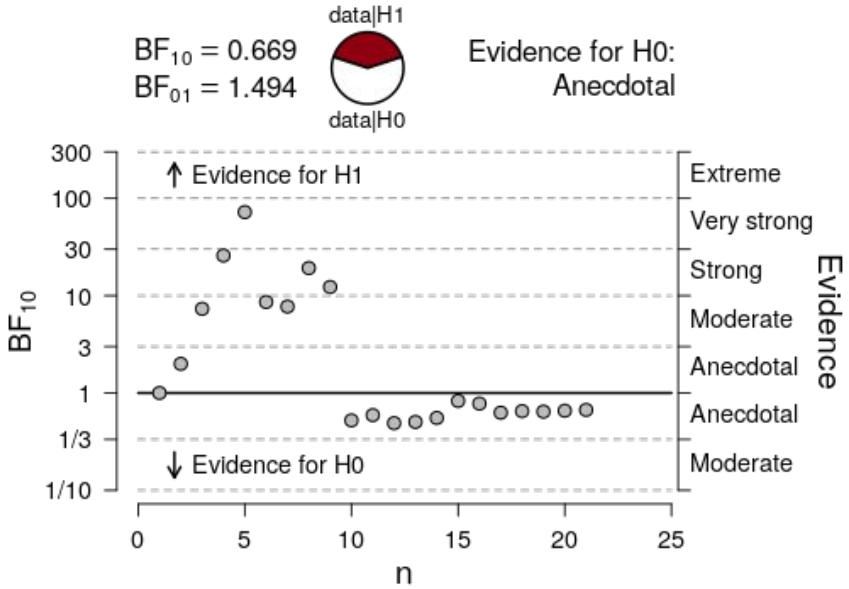
Example: hypothesis testing of CS+/CS- difference

- How do our conclusions depend on the prior? We can answer using a **robustness** (or sensitivity) **check**



Example: hypothesis testing of CS+/CS- difference

- We could also collect data until we reach a certain level of certainty



Going further

- Most of the presentation was based on the paper of **Etz & Vandekerckhove (2018)**, but the paper has a slower pace and goes into more details
- The recent special issue of Psychonomic Bulletin & Review on Bayesian methods features many excellent papers: http://bit.ly/BayesInPsych
- For a slightly different approach, check Richard McElreath's "Statistical Rethinking" course: https://github.com/rmcelreath/statrethinking_winter2019

References

[Box76] Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791-799.

[Etz18] Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, 25(1), 5-34.

[Gig04] Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5), 587-606.

[Gro14] Grolemund, Garrett, and Hadley Wickham. "A cognitive interpretation of data analysis." *International Statistical Review* 82.2 (2014): 184-204.

[Hoe14] Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164.

[Kru18] Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178-206.

[Lin00] Lindley, Dennis V. "The philosophy of statistics." *Journal of the Royal Statistical Society: Series D (The Statistician)* 49.3 (2000): 293-337.

[McE18] McElreath, R. (2018). Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC.

[vDo19] van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., ... & Ly, A. (2019). The JASP Guidelines for Conducting and Reporting a Bayesian Analysis.

[Wag18] Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Matzke, D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review*, 25(1), 35-57.