BAYESIAN STATISTICS IN PRACTICE

MARIËLLE ZONDERVAN-ZWIJNENBURG

NICE TO MEET YOU!

MARIËLLE ZONDERVAN-ZWIJNENBURG

10 THINGS YOU HAVE IN COMMON

MARIËLLE ZONDERVAN-ZWIJNENBURG

Outline

- Bayesian statistics
- Bayesian versus Frequentist
- Bayesian Estimation
  - Exercises
- Bayesian Hypothesis Testing
  - Exercises

- Q&A

BAYESIAN STATISTICS

MARIËLLE ZONDERVAN-ZWIJNENBURG

Bayes

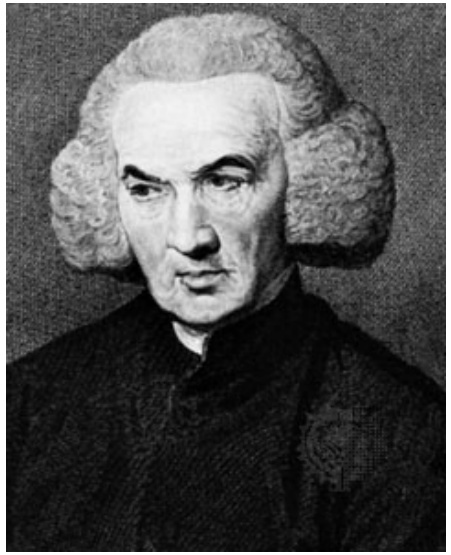
LII. *An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *communicated by Mr.* Price, *in a letter to* John Canton, *M. A. and F. R. S.*

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.

He had, you know, the honour of being a member of that illustrious Society, and was much esteemed by many as a very able mathematician. In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times. He adds, that he soon perceived that it would not be

It's about inverse probability. Assigning probability to something unobserved



Bayes - Laplace

- Laplace published what we call ‘Bayes formula’ (1812)

- $P(H|e) = \frac{P(e|H) \times P(H)}{P(e)}$

Bayes theorem

Likelihood
How probable is the evidence given that hypothesis *H* is true?

Prior
How probable was hypothesis *H* before observing evidence?

$$P(H|e) = \frac{P(e|H) \times P(H)}{P(e)}$$

Posterior
How probable is hypothesis *H* Given the observed evidence?

Marginal
How probable is the evidence (under all hypotheses)?

Bayes theorem

Likelihood
How probable is the evidence given that hypothesis *H* is true?

Prior
How probable was hypothesis *H* before observing evidence?

$$P(H|e) \approx P(e|H) \times P(H)$$

Posterior
How probable is hypothesis *H* Given the observed evidence?

Bayes theorem

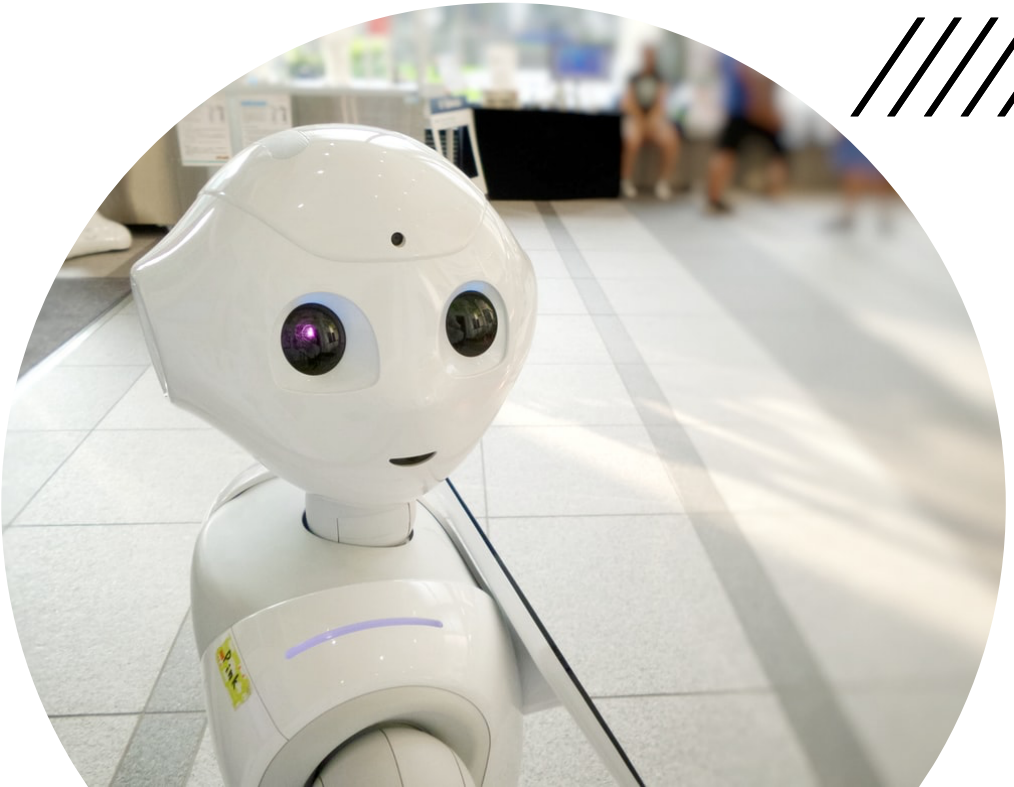
Likelihood
How probable is the evidence given that hypothesis *H* is true?

Prior
How probable was hypothesis *H* before observing evidence?

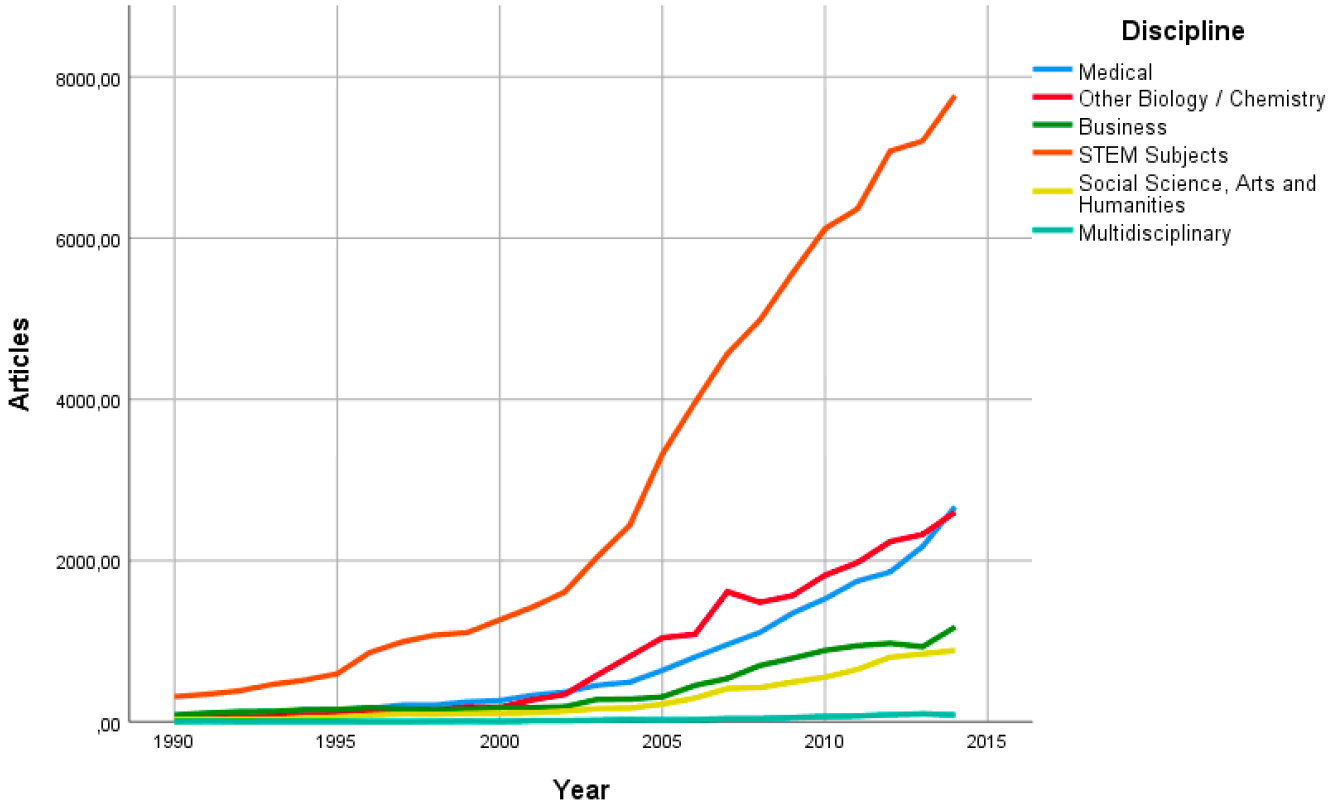
$$P(H|e) \approx P(e|H) \times P(H)$$ **= updating evidence!**

Posterior
How probable is hypothesis *H* Given the observed evidence?

BAYES AROUND US

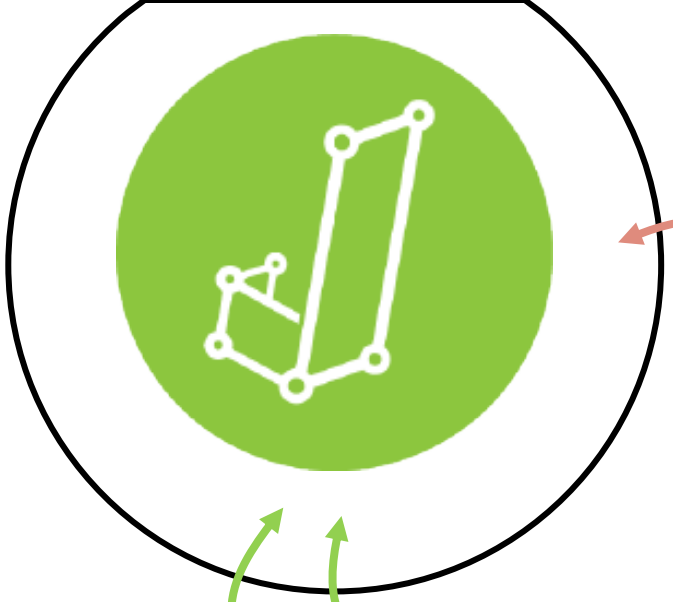


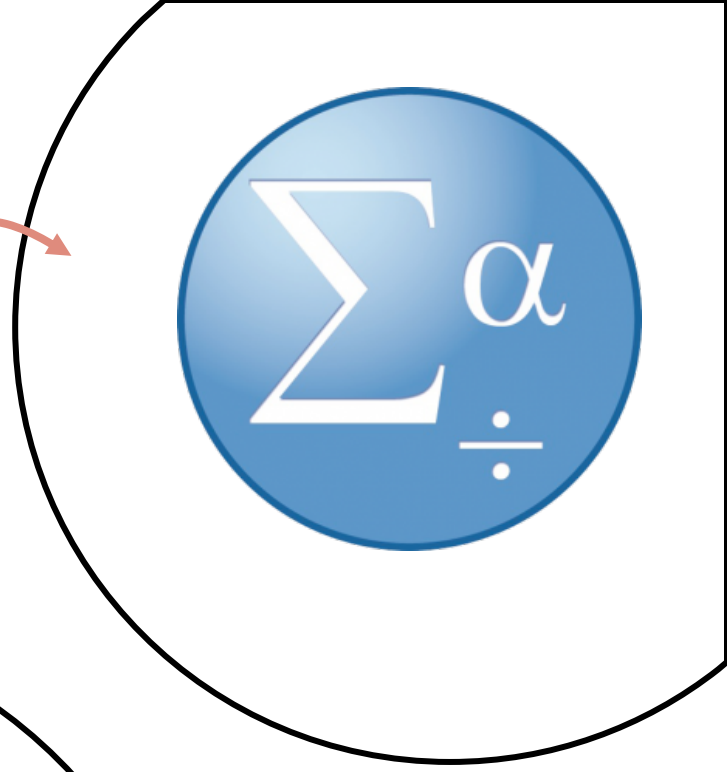
BAYES IN SCIENCE



Van De Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M.A.J., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217-239. https://doi.org/10.1037/met0000100

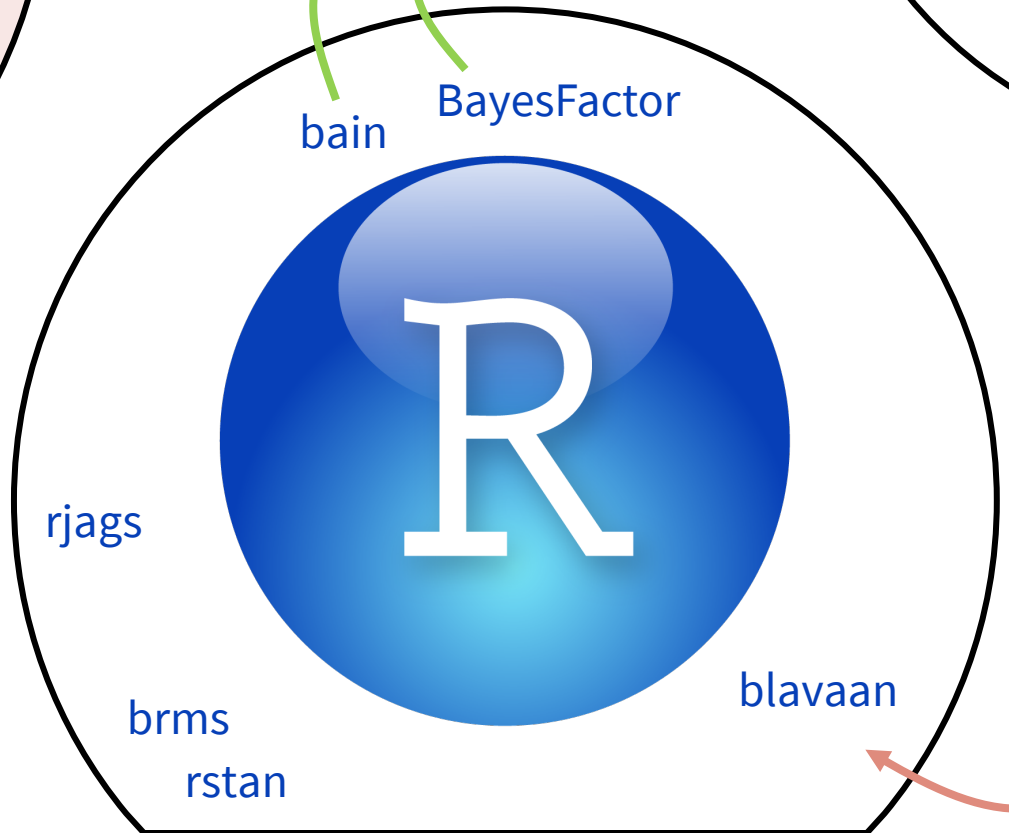
BAYES IN SCIENCE SOFTWARE





bain

BayesFactor



rjags

brms

rstan

blavaan

Mplus

BAYESIAN VS FREQUENTIST

MARIËLLE ZONDERVAN-ZWIJNENBURG

Estimation VS Hypothesis Testing

	Frequentist	**Bayesian**
Estimation	Maximum likelihood estimation (MLE)	Posterior distribution with highest density credibility interval (CI)
Hypothesis Testing	*p*-value	Bayes factor (BF) & posterior model probabilities (PMP)

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,* 178-206. doi: 10.3758/s13423-016-1221-4

Estimation VS Hypothesis Testing

	Frequentist	Bayesian
Estimation	Maximum likelihood estimation (MLE)	Posterior distribution with highest density credibility interval (CI)
Hypothesis Testing	*p*-value	Bayes factor (BF) & posterior model probabilities (PMP)

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,* 178-206. doi: 10.3758/s13423-016-1221-4

Bayesian vs Frequentist Probability

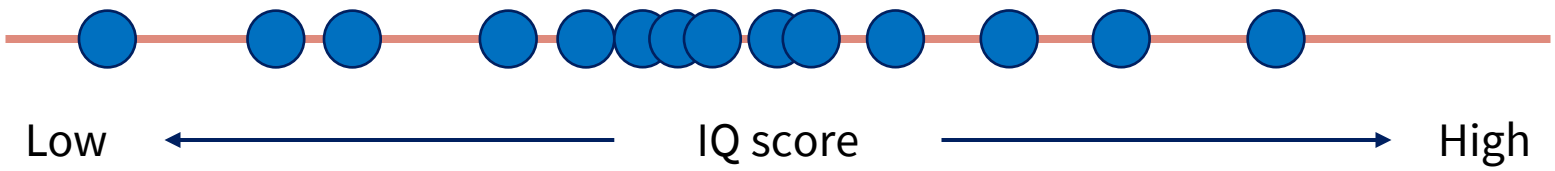
- Frequentist: relative frequency
- Bayesian: degree of belief

Bayesian vs Frequentist

- What is the probability of the data given the estimate / null hypothesis?
- What is the most probable.. parameter estimate / hypothesis given the prior & data?

Frequentist estimation

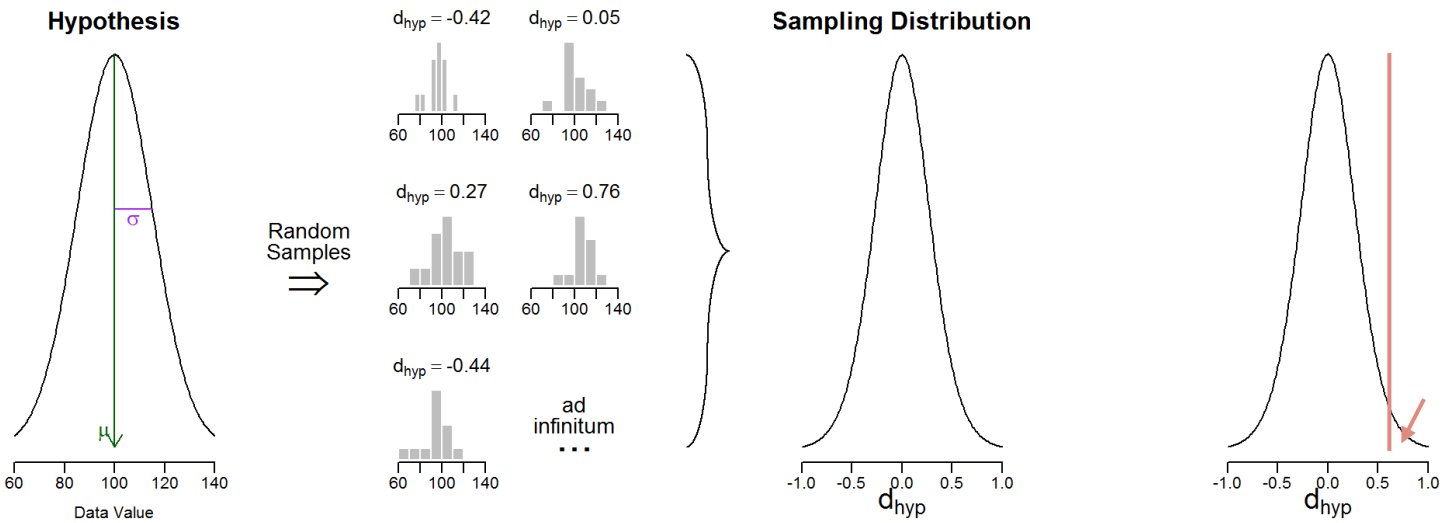
- MLE:
  find the estimates that make the data most likely under the assumed model
  - What is the model?
  - How well does it fit with different parameter settings?
  - What should be the parameter settings for maximum likelihood of observing the data? = maximum likelihood estimate



https://www.youtube.com/watch?v=XepXtl9YKwc

Frequentist Hypothesis Evaluation (NHST)

What is the probability of getting *different* data given the hypothesized parameter values (0, …, =)? → Sample data from the null distribution, evaluate % summary stats ≥ observed in data



Frequentist Hypothesis Evaluation (NHST)

- $p < .05$ → reject the null (with alpha .05)
- $p \geq .05$ → null not rejected (given current sample size)

Notes

- With a large enough sample size, any effect can be ‘significant’
- There is no such thing as ‘marginally’ or ‘very’ significant

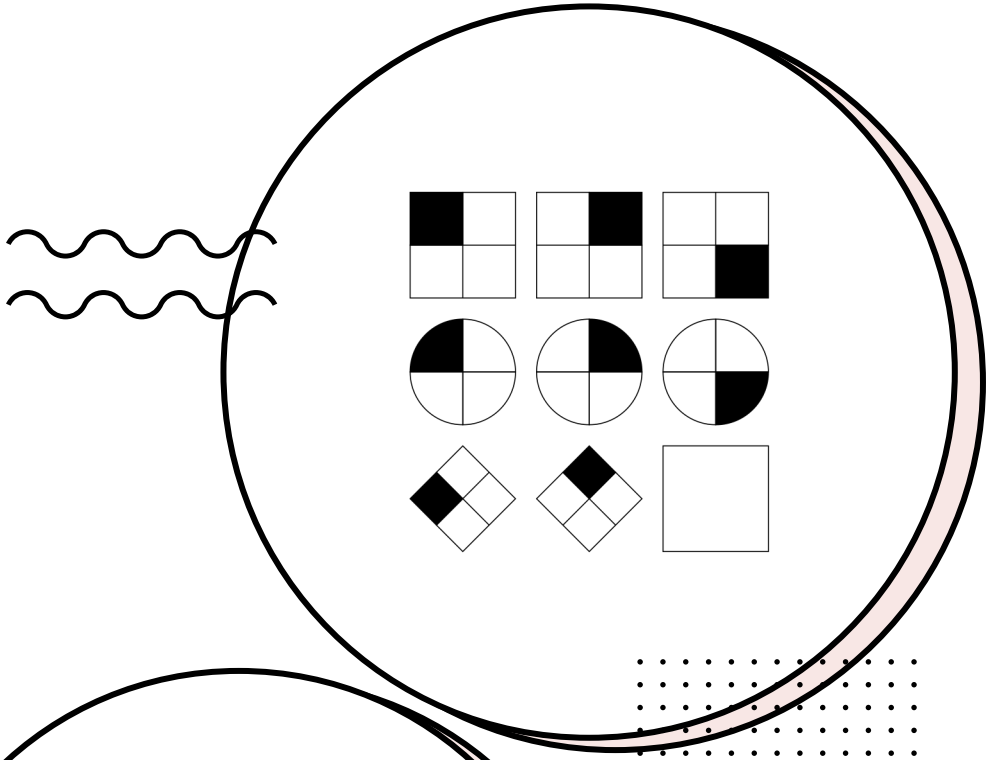
Confidence interval

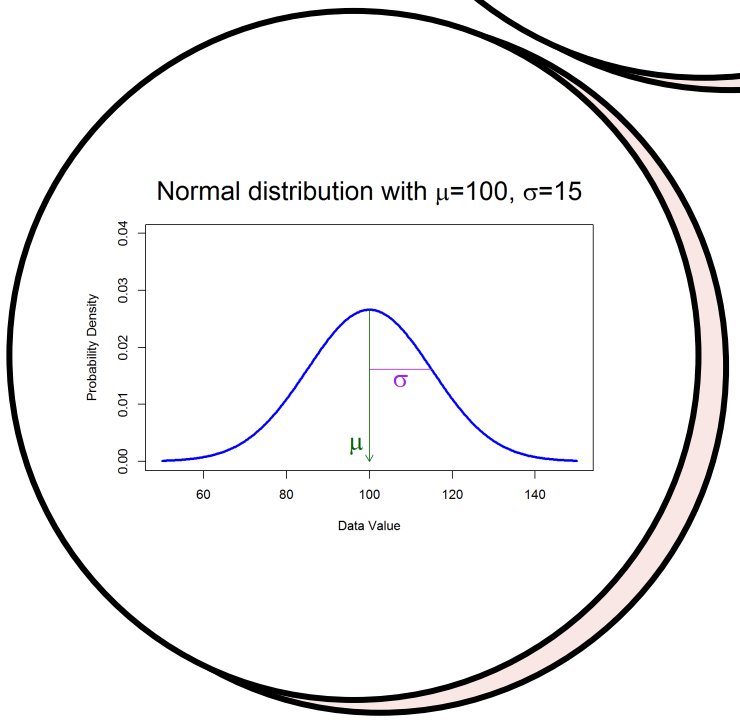
- If we were to repeat this experiment ∞ times, and calculate the confidence intervals, 95%* of the confidence intervals will include the true parameter value (5% not!)
- ≠ there is 95% probability that the true value is in the CI (this is Bayesian)
- *Given this observed data*, the true value is either in our confidence interval or it isn't

EXAMPLE

INTELLIGENCE

MARIËLLE ZONDERVAN-ZWIJNENBURG





IQ

- Data: IQ measurements
- Model: Normal distribution with mean (μ) and SD (σ) → one-sample *t*-test
- RQ: What settings of μ and σ are plausible descriptions of the data
- RQ: Is μ equal to 100?

BAYESIAN ESTIMATION

MARIËLLE ZONDERVAN-ZWIJNENBURG

Ingredients

Model

Pior distribution

Data

PRIOR DISTRIBUTION

MARIËLLE ZONDERVAN-ZWIJNENBURG

Prior Distributions - What

- A distribution with possible & plausibile values for the parameter

- NOT necessarily older than the data
- NOT necessarily unique → no such thing as correct prior
- NOT necessarily important → as the amount of data increases

Prior Distributions - Types

DEFAULT / REFERENCE / UNINFORMATIVE / *OBJECTIVE*

HISTORICAL

EXPERT ELICITATION

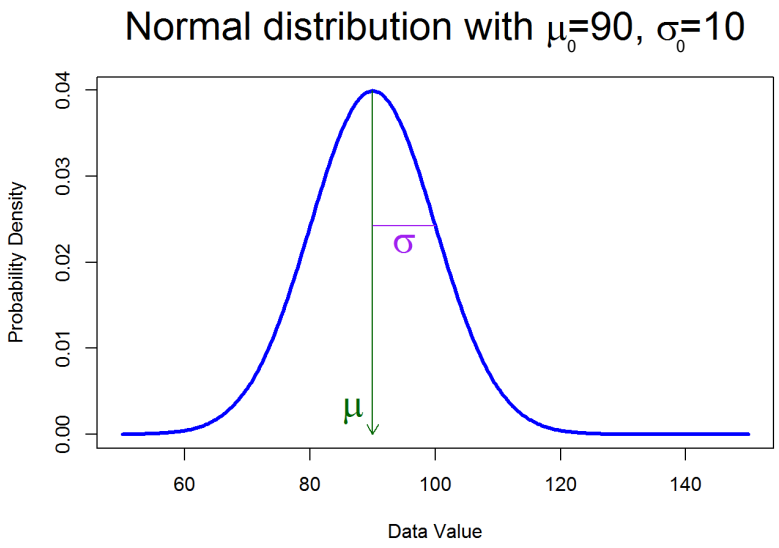
ROBUST

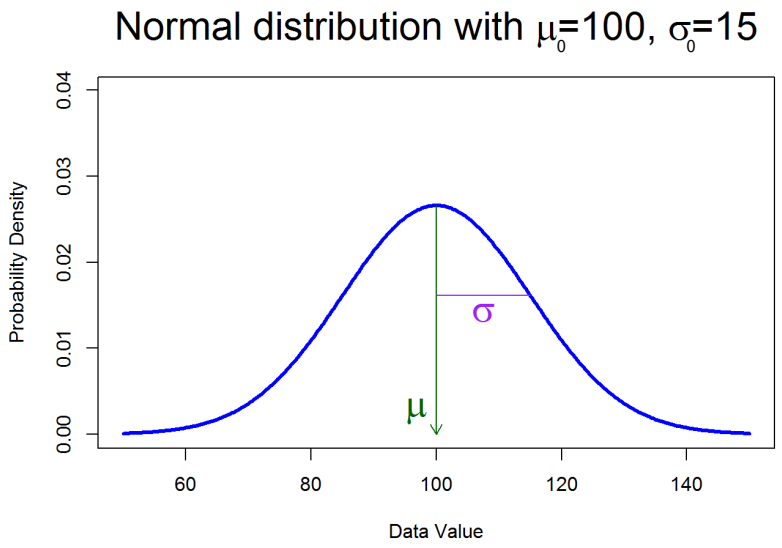
Example

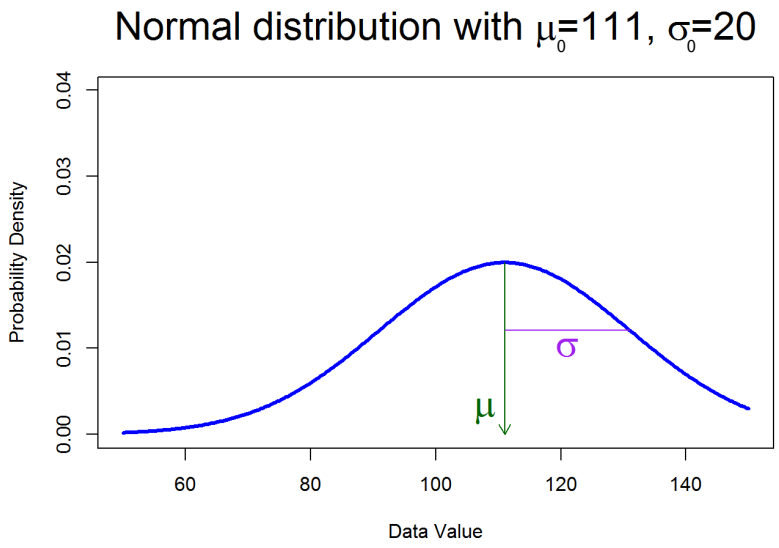
- Without expectations about the results, we still have information!
- What if we know a little more about IQ?

IQ Range	Classficiation
145-160	Highly advanced
130-144	Gifted
120-129	Superior
110-119	High average
90-109	Average
80-89	Low average
70-79	Borderline impaired
55-69	Mildly impaired
40-54	Moderately impaired

Example prior distributions for μ

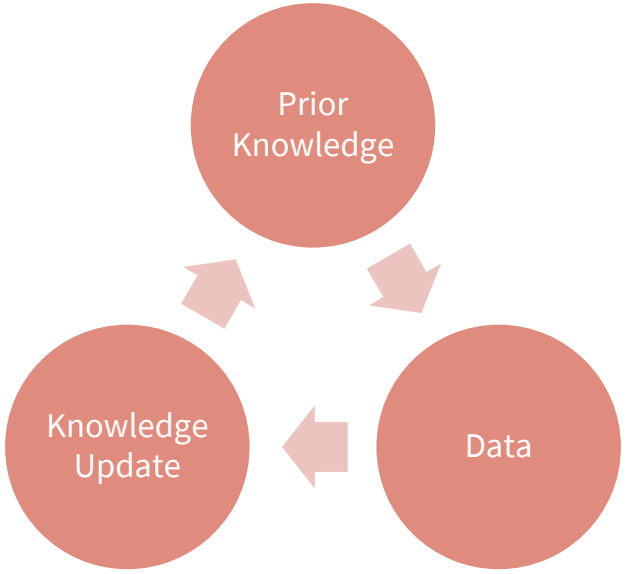






Why informative priors?

- Cumulative knowledge
- Small samples



Examples

LOGIN / REGISTER

Research Topic

Moving Beyond Non-Informative Prior Distributions: Achieving the Full Potential of Bayesian Methods for Psychological Research

EUROPEAN JOURNAL OF PSYCHOTRAUMATOLOGY

BASIC RESEARCH ARTICLE

Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors

Rens van de Schoot[1,2]*, Joris J. Broere[1], Koen H. Perryck[1], Mariëlle Zondervan-Zwijnenburg[1] and Nancy E. van Loey[3,4]

[1]Department of Methods and Statistics, Utrecht University, Utrecht, The Netherlands; [2]Optentia Research Program, Faculty of Humanities, North-West University, Vanderbijlpark, South Africa; [3]Department of Clinical & Health Psychology, Utrecht University, Utrecht, The Netherlands; [4]Department Behavioural Research, Association of Dutch Burn Centres, Beverwijk, The Netherlands

Research in Human Development, 00: 1–16, 2017

ISSN: 1542-7609 print / 1542-7617 online
DOI: https://doi.org/10.1080/15427609.2017.1370966

Routledge
Taylor & Francis Group

Check for updates

Where Do Priors Come From? Applying Guidelines to Construct Informative Priors in Small Sample Research

Mariëlle Zondervan-Zwijnenburg and Margot Peeters
Utrecht University

Sarah Depaoli
University of California–Merced

Rens Van de Schoot
Utrecht University and North-West University

INTERNATIONAL REVIEW OF PSYCHIATRY
https://doi.org/10.1080/09540261.2020.1809357

Taylor & Francis
Taylor & Francis Group

ARTICLE

Exploring meaning in life through a brief photo-ethnographic intervention using Instagram: a Bayesian growth modelling approach

Llewellyn E. van Zyl[a,b,c,d], Marielle A. J. Zondervan-Zwijnenburg[e], Leah R. Dickens[f] and Inge Hulshof[a,g]

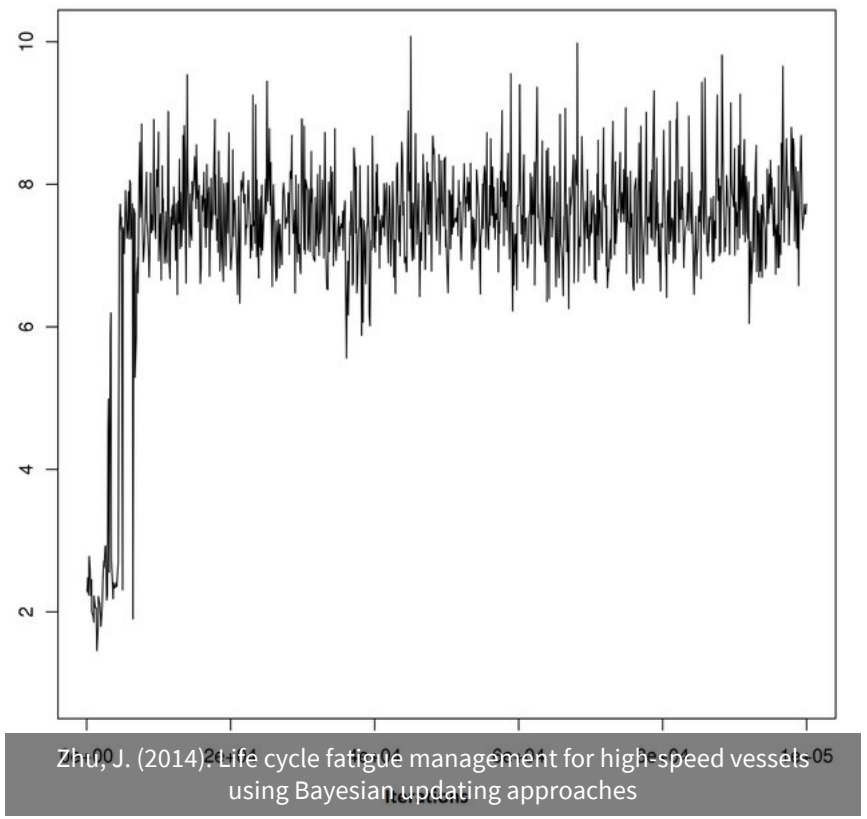
[a]Department of Industrial Engineering, University of Eindhoven, Eindhoven, the Netherlands; [b]Optentia Research Focus Area, North-West University (VTC), Vanderbijlpark, South Africa; [c]Department of Human Resource Management, University of Twente, Enschede, the Netherlands; [d]Department of Social Psychology, Institut für Psychologie, Goethe University, Frankfurt am Main, Germany; [e]Department of Methodology and Statistics, Utrecht University, Utrecht, the Netherlands; [f]Department of Psychology, Kenyon College, Gambier, OH, USA; [g]Department of Work and Organisational Psychology, Open University, Heerlen, the Netherlands

Why default/reference priors?

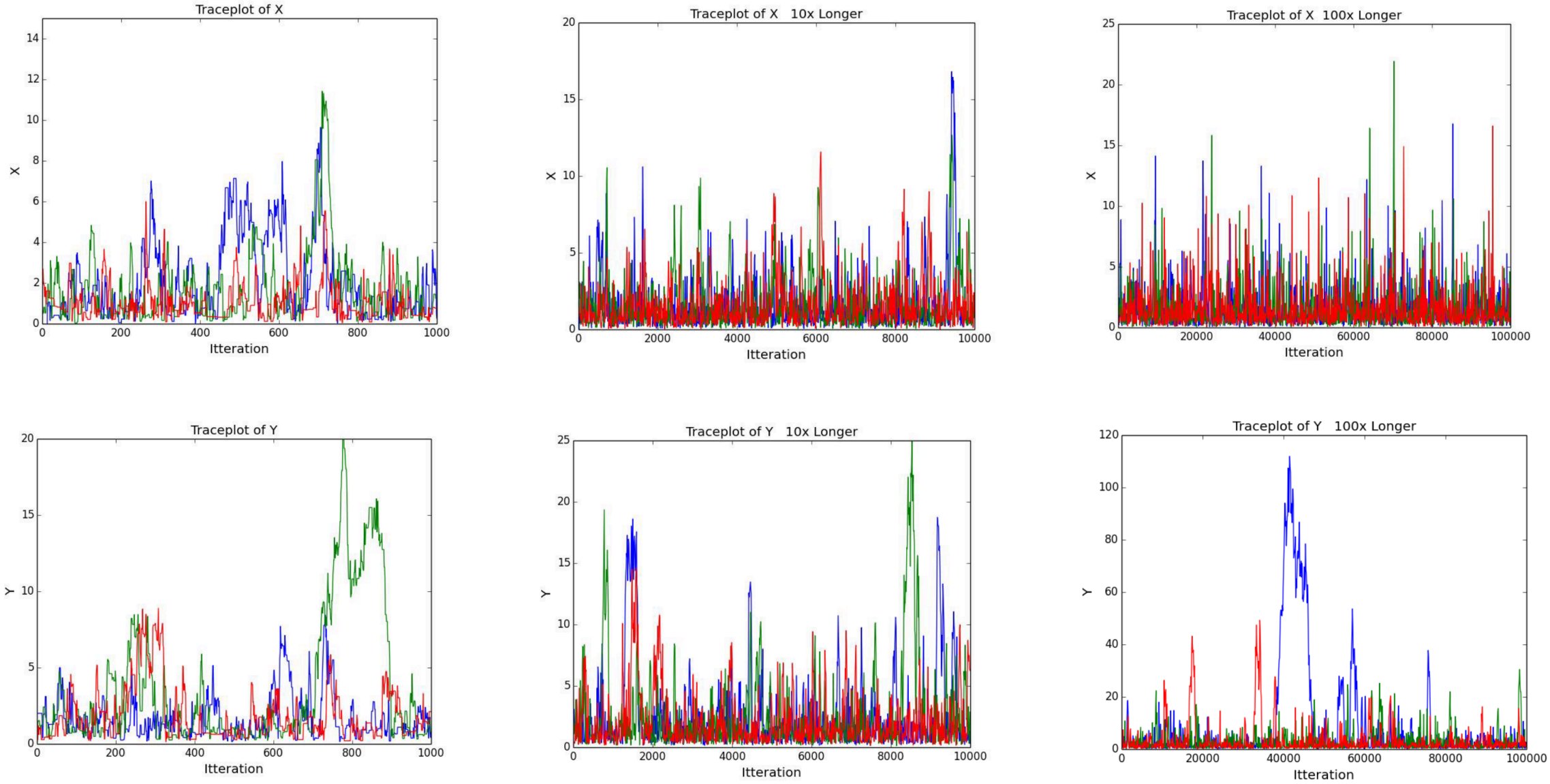
- If you want to do a frequentist analysis / don’t have prior information, but the analysis is too complex for frequentist methods (e.g., Dynamic structural equation modeling)

Estimation procedure

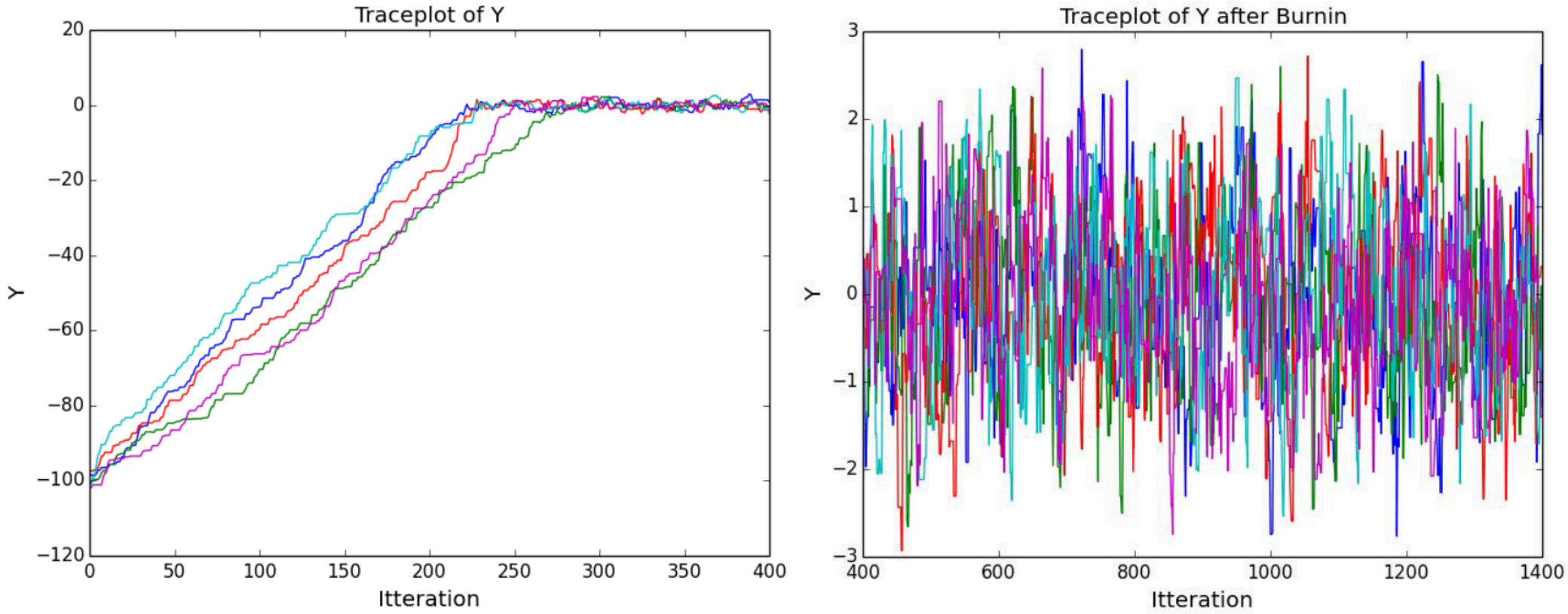
- Sampling from a multidimensional distribution
- (MCMC) Gibbs sampler: Starting values + Estimating one parameter value, while keeping all others constant
- Always a result, but how was the estimation process? Check trace plots (convergence) and posterior distribution (see e.g., Ex. 3)



Zhu, J. (2014). Life cycle fatigue management for high-speed vessels using Bayesian updating approaches



Ford, E.B. (2015). Convergence diagnostics for Markov chain Monte Carlo. Penn State University – Center for Astrostatistics. Presentation.



Ford, E.B. (2015). Convergence diagnostics for Markov chain Monte Carlo. Penn State University – Center for Astrostatistics. Presentation.

Convergence

Good signs

- Similar results with increased number of iterations
- Chains (with different starting values) mix well and give similar results

However

- You can’t prove convergence
- There may be another solution that is never hit by the chains

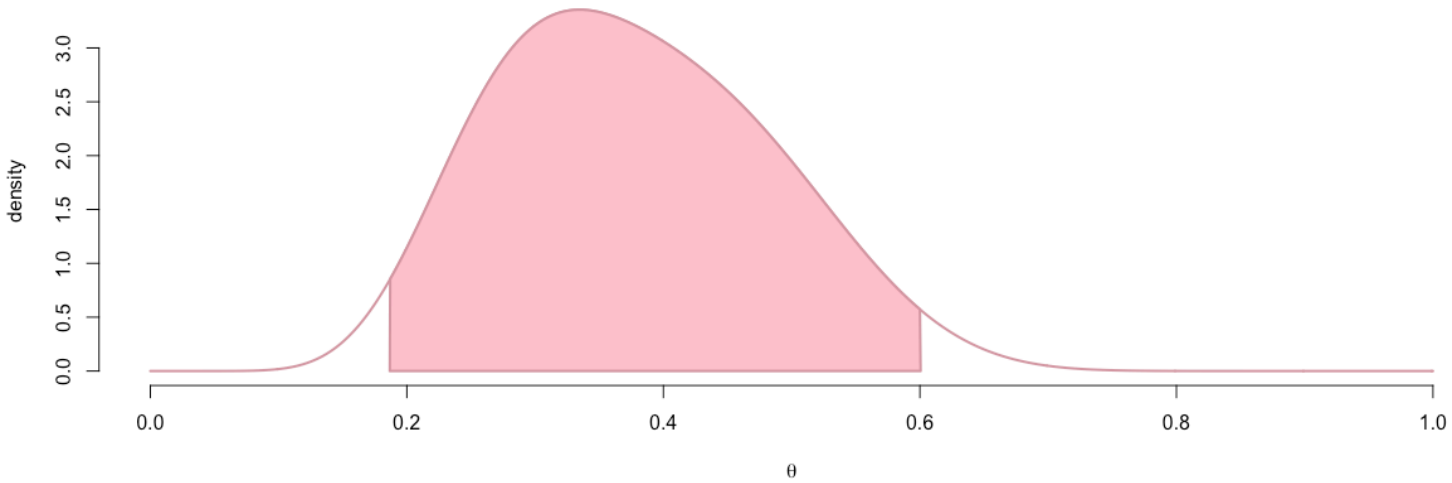
What to do

- Many iterations + multiple chains + look at plots + Potential Scale Reduction

Ford, E.B. (2015). Convergence diagnostics for Markov chain Monte Carlo. Penn State University – Center for Astrostatistics. Presentation.

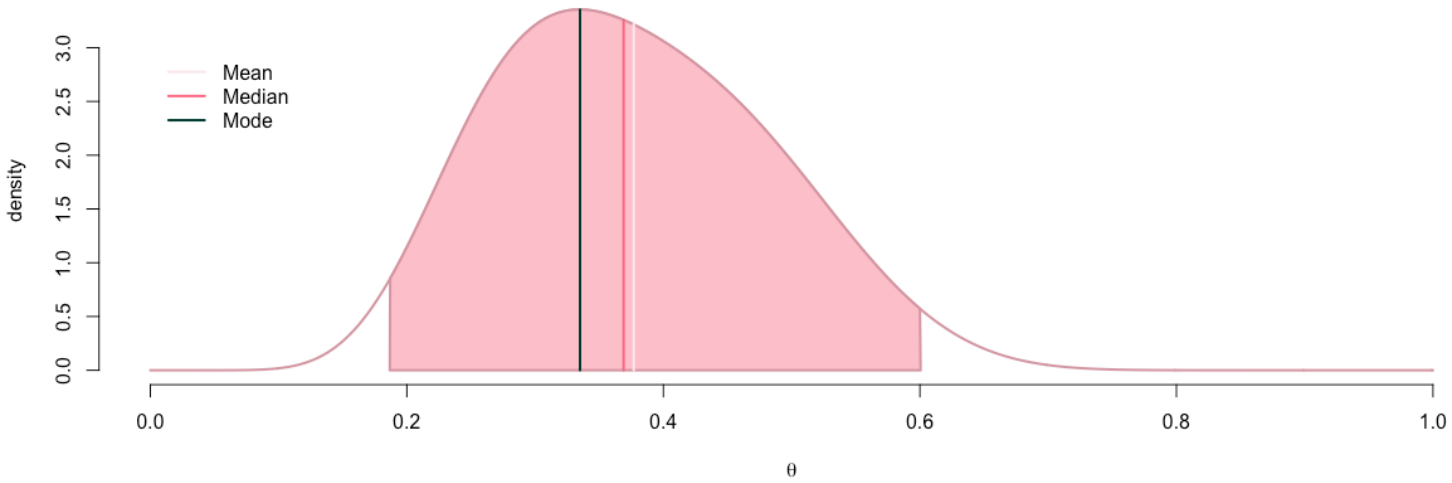
Result: Credible Interval

- Also: *credible region, highest posterior density interval (HDI)*
- Given our observed data, there is a 95%* probability that the true value falls within the credible interval



Result: point estimate

- Usually: the median (50 ↔ 50)



Exercises Part I

See our class notebook in Teams: (Web view)

Try it! Ex. 1a (Kruschke App)

- https://iupbsapps.shinyapps.io/KruschkeFreqAndBayesApp/
- Choose Bayesian estimation and adjust:
  - The data mean + watch the posterior mode of mu
  - The data sd + watch the posterior mode of sigma and posterior HDI of mu
  - The sample size + watch the HDI's of both parameters
- Set the data mean to 120, SD to 25 and sample size to 5, and adjust:
  - The prior SD for mu (with prior mode for mu still at 100) + watch the posterior distribution
  - The prior SD for sigma + watch the posterior mode for sigma
  - The sample size of the data, can you overcome the prior settings?

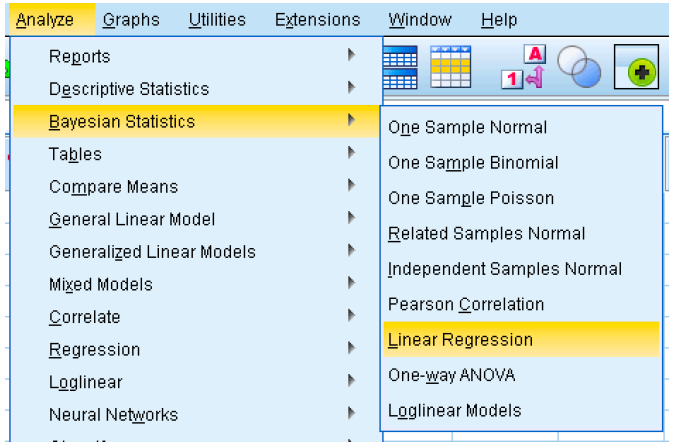
Try it! Ex. 1b (van de Schoot App)

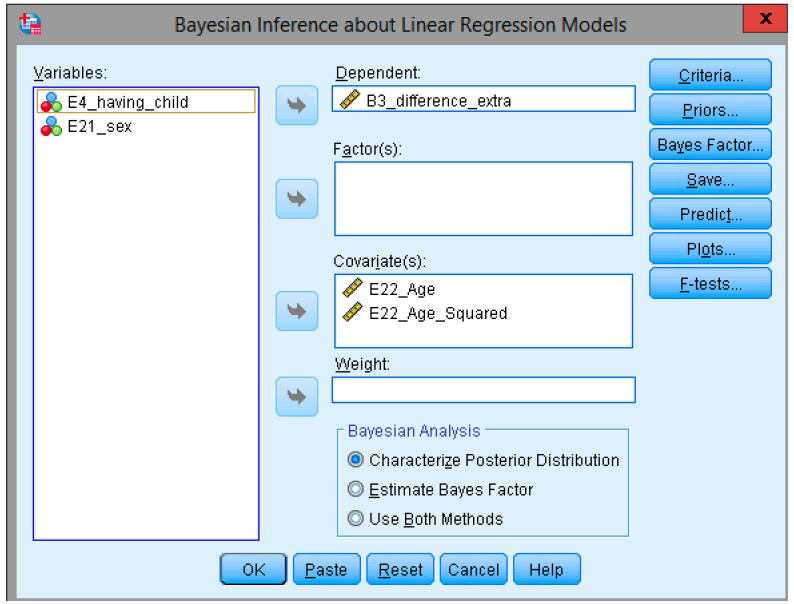
- https://www.rensvandeschoot.com/tutorials/fbi-the-app/
- **a.** Pretend you do not know nothing about IQ except that it cannot be smaller than 40 and that values larger than 180 are really impossible. Which prior will you choose?
- **b.** Generate data for 22 individuals and run the Bayesian model (default option). Copy-paste your model specifications and plot to this Word document
- **c.** Change the prior to a distribution which would make more sense for IQ: we know it cannot be smaller than 40 or larger than 180 AND it is expected to be normally distributed around 100 (=prior mean). But, how sure are you? Try values for the prior variance of 10 and 1. Notice the prior becomes more peaked the smaller your prior variance (higher certainty/precision). Run the the two models and write down the results. How would you describe the relation between your level of uncertainty and the posterior variance?
- **d.** Now, re-run the model with a larger sample size (n=100). Copy-paste the results. How are the current results different from the results under ‘c’?
- **e.** Repeat steps ‘c’ and ‘d’ but now for a different prior mean (assuming your prior knowledge conflicts with the data, let's say 90 and use n=22). Copy-paste your results. How did the new results differ when compared to the results with a ‘correct’ prior mean?
- **f.** What happens if your prior mean is really far away from the data, let's say 70 with n=22. Note that this situation is really extreme and often in practice the prior is much closer to the data.
- **g.** So far, we assumed the variance of IQ to be known, which is not realistic. Re-run models 'e' and 'f'' using the option 'Run with Sigma Unknown'. In the background the software JAGS estimates the mean of IQ and its variance using MCMC-sampling. The posterior is no longer analytically obtained but is approximated. As a result the posterior distribution looks wobbly and every time you hit the 'run'-button it will look a bit different (give it a try). Can you describe the difference between the previous results under 'f'' and the current results? If you run the model again but with n=50 the posterior is again a compromise between the prior and the data. Can you explain?

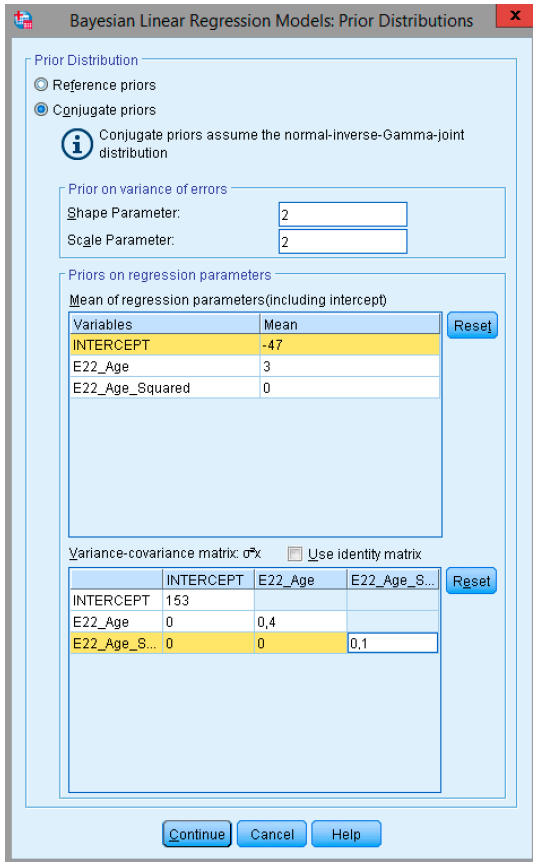
Try it! Ex. 2 (van de Schoot SPSS) (p1/2)

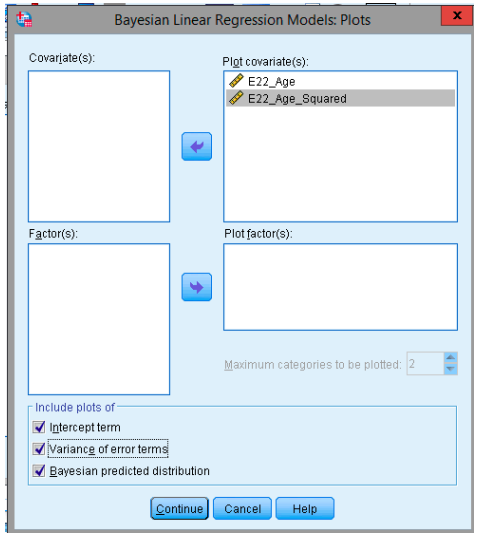
- https://www.rensvandeschoot.com/tutorials/bayesian-regression-spss/
- Data: https://www.rensvandeschoot.com/wp-content/uploads/2018/10/phd-delays.csv

Try it! Ex. 2 (SPSS) (p2/2)









Try it! Ex. 3 (R)

A. Exercises with blavaan (SEM – syntax similar to lavaan / Mplus) https://www.rensvandeschoot.com/tutorials/wambs-blavaan-tutorial-using-stan/

   OR

B. Exercises with brms (generalized (non-)linear multivariate multilevel models using Stan – syntax similar to lme4 package https://www.rensvandeschoot.com/brms-wambs/

Discuss: Informative Priors

WHOULD YOU USE THEM?

WHAT ARE THE ADVANTAGES?

WHAT ARE THE DANGERS?

CAN THE DANGERS BE TACKLED?

Extensions

- Expert priors
- Posterior predictive check
- SEM
- Prediction / machine learning
  - Which students will experience delay or drop-out?
  - How will COVID-19 develop (with or without face masks)

BAYESIAN HYPOTHESIS TESTING

MARIËLLE ZONDERVAN-ZWIJNENBURG

Two schools

Amsterdam: Bayes Factor

Kruschke et al. CI's, ROPE

BF definition

- Relative evidence of one model over another
- Compare two sets of posterior probabilities

Posterior odds = **BF** × **Prior odds**

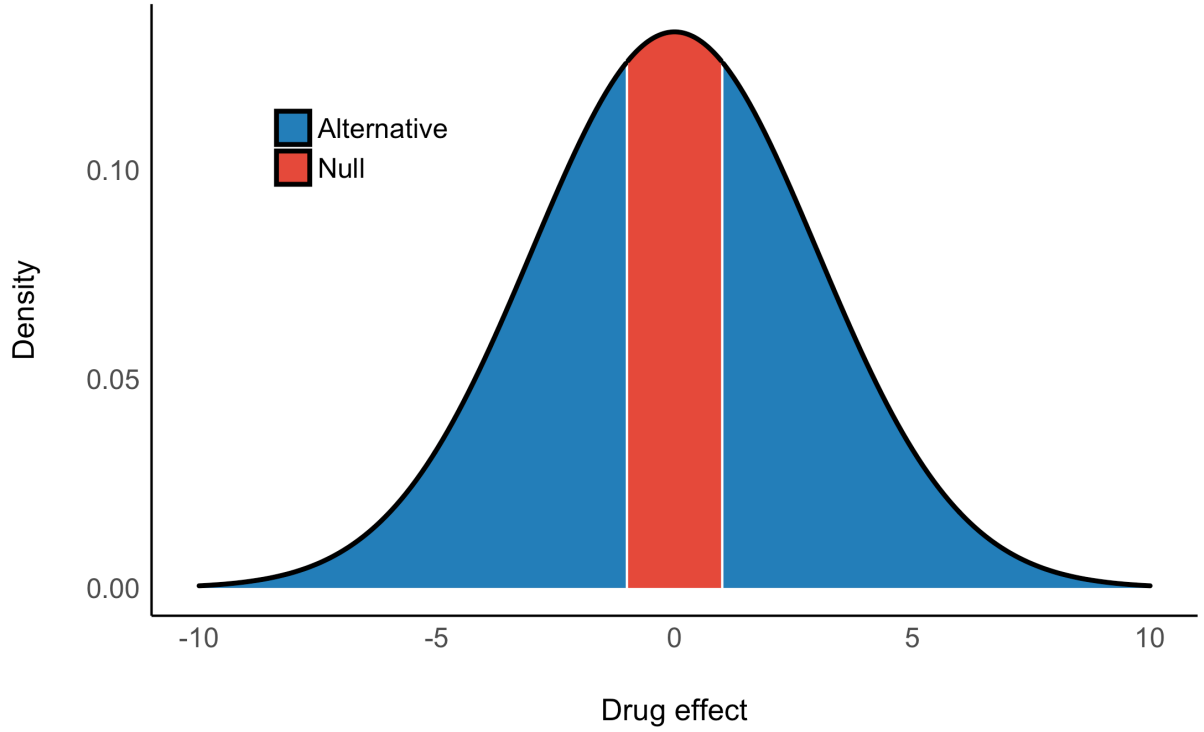
$$\frac{P(H_1|e)}{P(H_2|e)} = \boxed{\frac{P(e|H_1)}{P(e|H_2)}} \times \frac{P(H_1)}{P(H_2)}$$

BF = **Posterior odds** / **Prior odds**

$$\frac{P(e|H_1)}{P(e|H_2)} = \frac{P(H_1|e)}{P(H_2|e)} / \frac{P(H_1)}{P(H_2)}$$

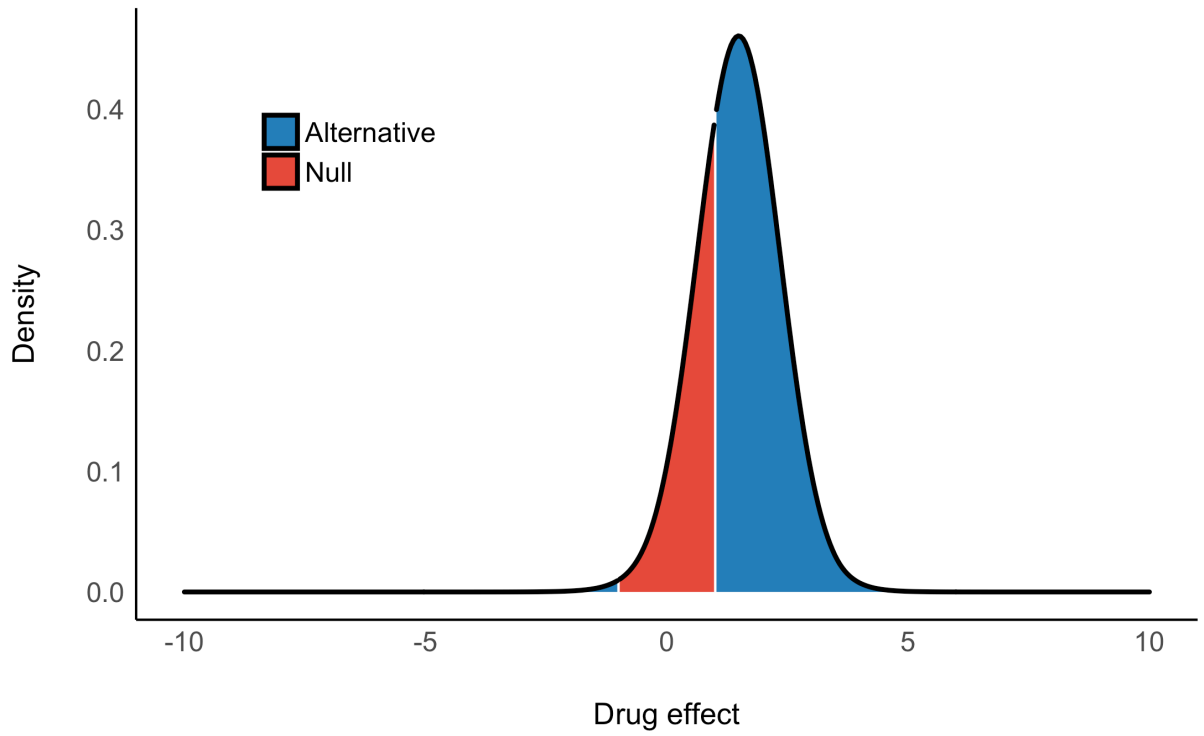
- BF = Relative probability of observed data by two models (LR)
- BF = Shift in (prior) evidence after observing the data
- The data provide evidence for/against a hypothesis, depending on whether the posterior odds are greater than the prior odds

Prior & prior odds



- Prior $b_{drug} \sim N(0,3)$
- Two models: $\frac{P(bdrug \in [-1,1]}{P(bdrug \notin [-1,1]}$
- Odds = $\frac{P(H_1)}{P(H_2)} = \frac{0.3125}{0.6875} = 2.2$

Posterior & posterior odds

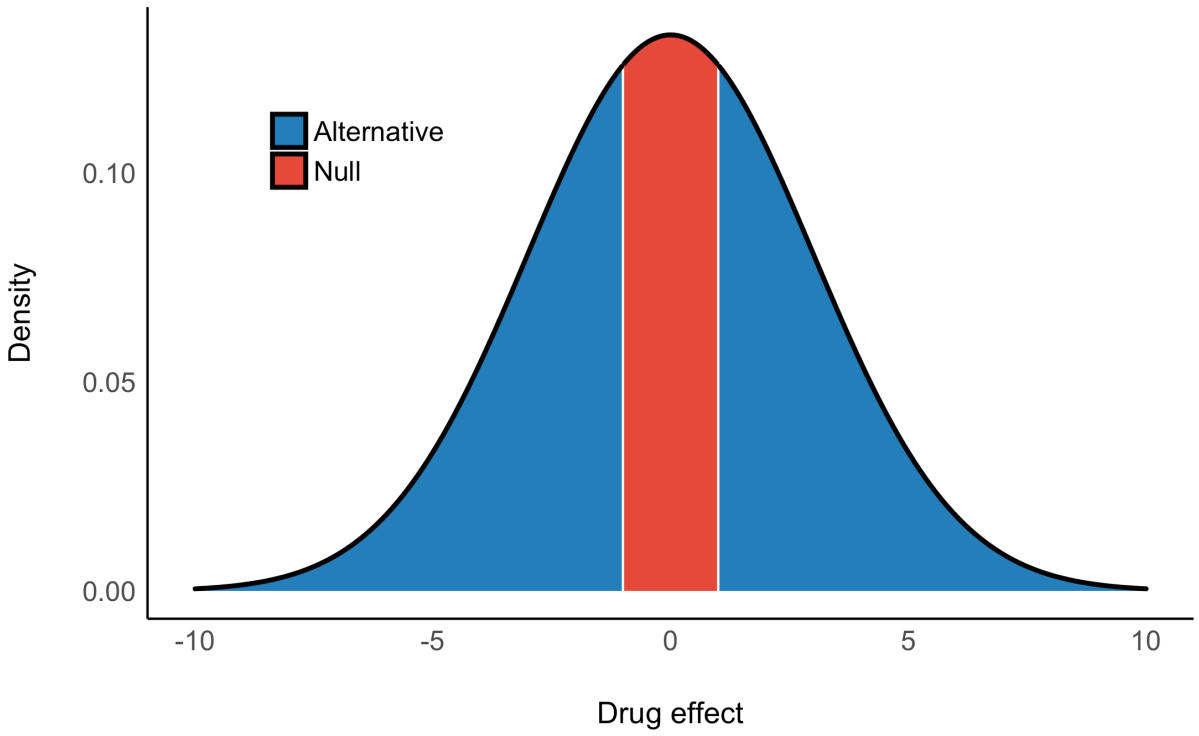


- Posterior
- Odds = $\frac{P(e|H_1)}{P(e|H_2)} = \frac{0.333}{0.666} = 2.0$

Prior & Posterior

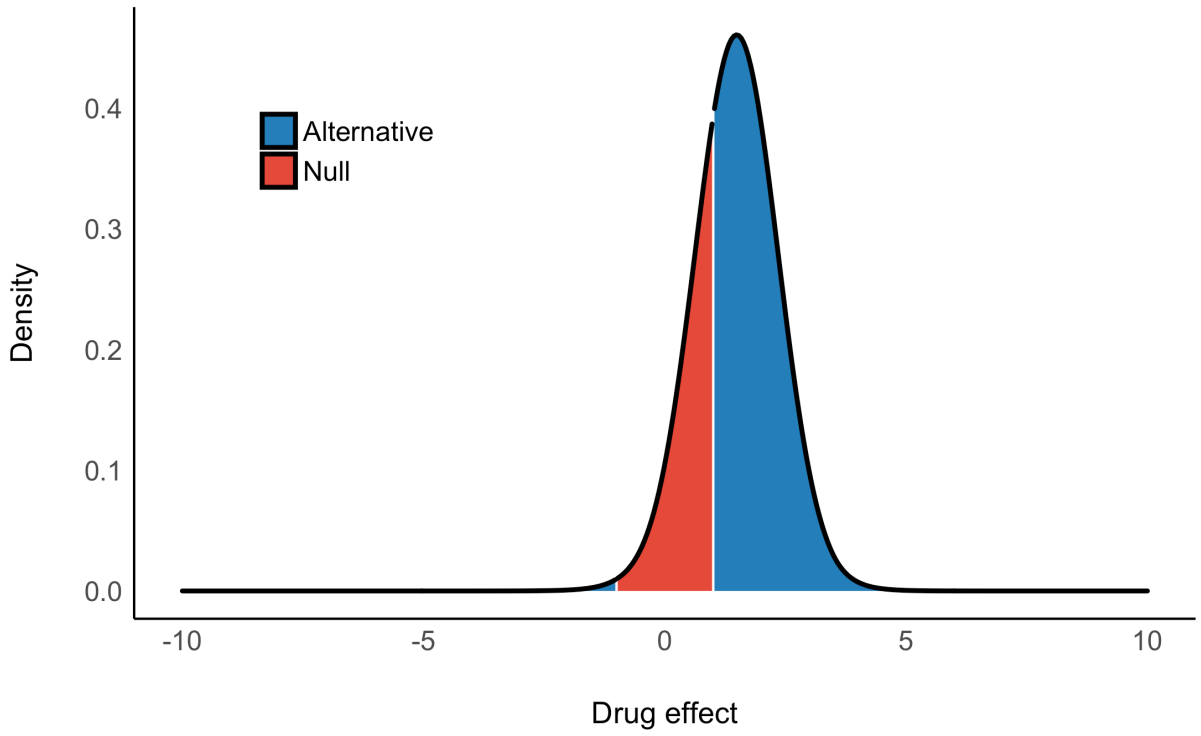
$\frac{P(H_1)}{P(H_2)} = 2.2$

Prior odds

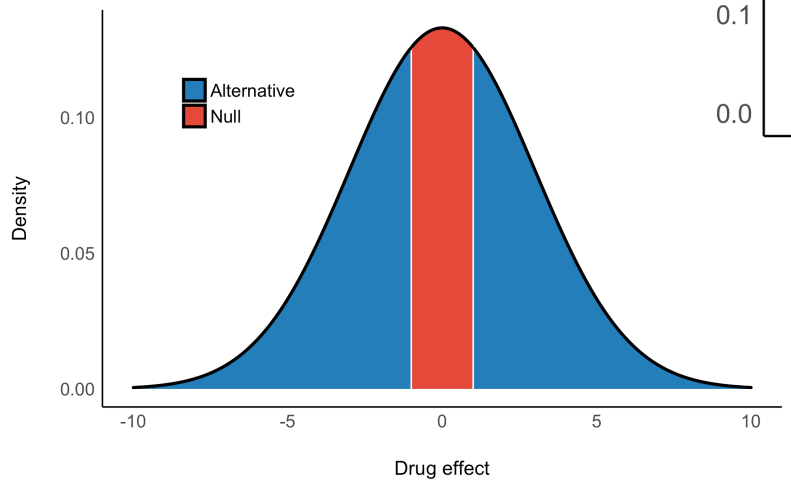


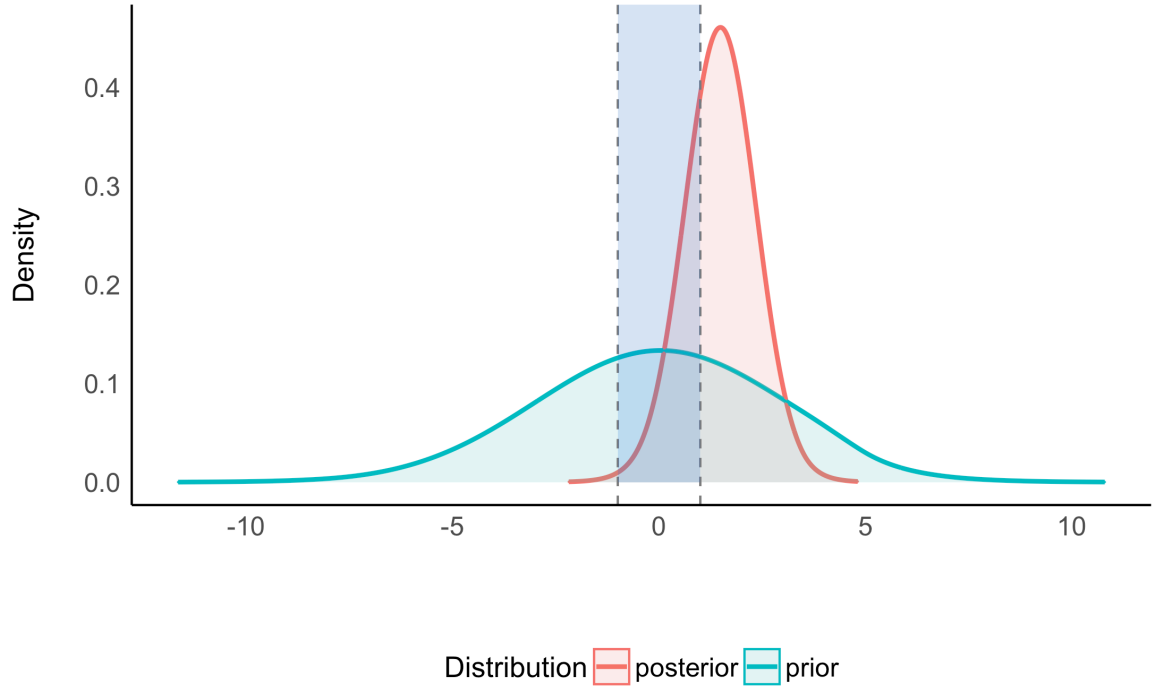
$\frac{P(H_1|e)}{P(H_2|e)} = 2.0$

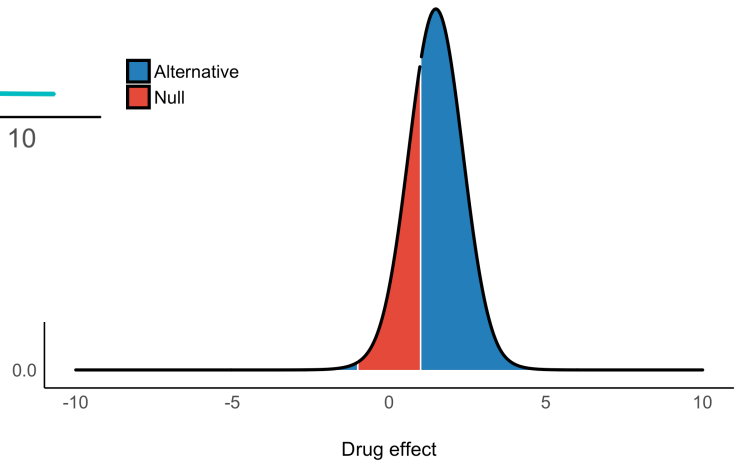
Posterior odds



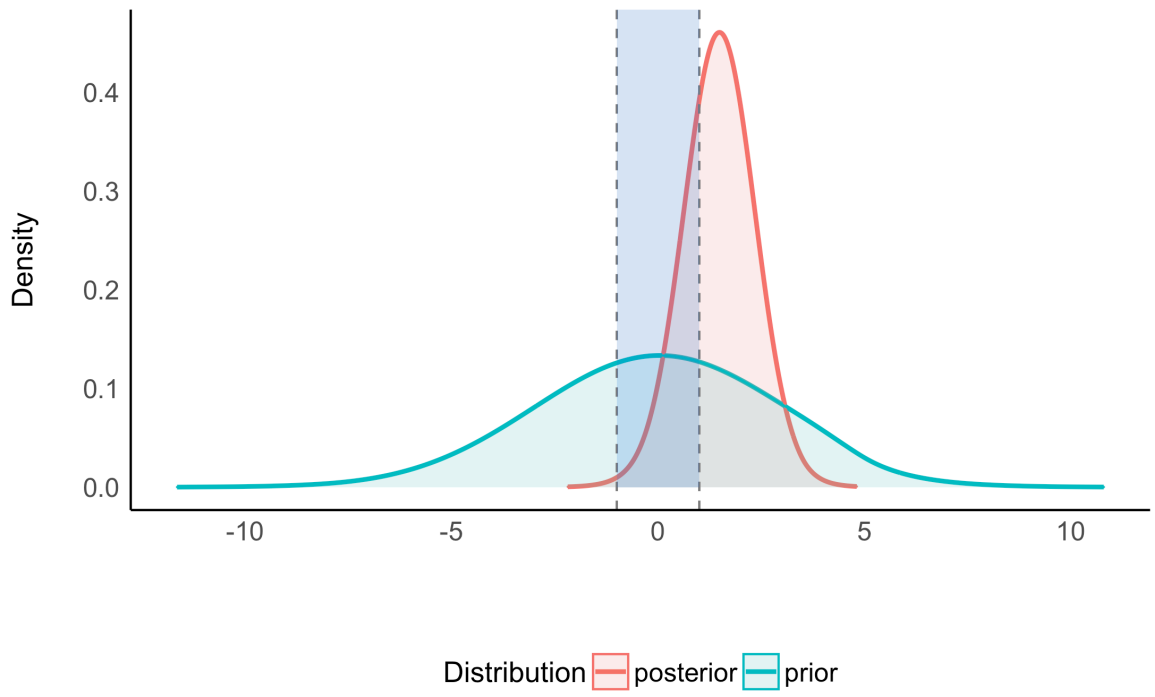
Bayes factor







Bayes factor



BF = **Posterior odds** / **Prior odds**

$$\frac{P(e|H_1)}{P(e|H_2)} = \frac{P(H_1|e)}{P(H_2|e)} / \frac{P(H_1)}{P(H_2)}$$

$$\frac{P(e|H_1)}{P(e|H_2)} = \frac{0.333}{0.666} / \frac{0.3125}{0.6875}$$

$$\frac{P(e|H_1)}{P(e|H_2)} = 2/2.2$$

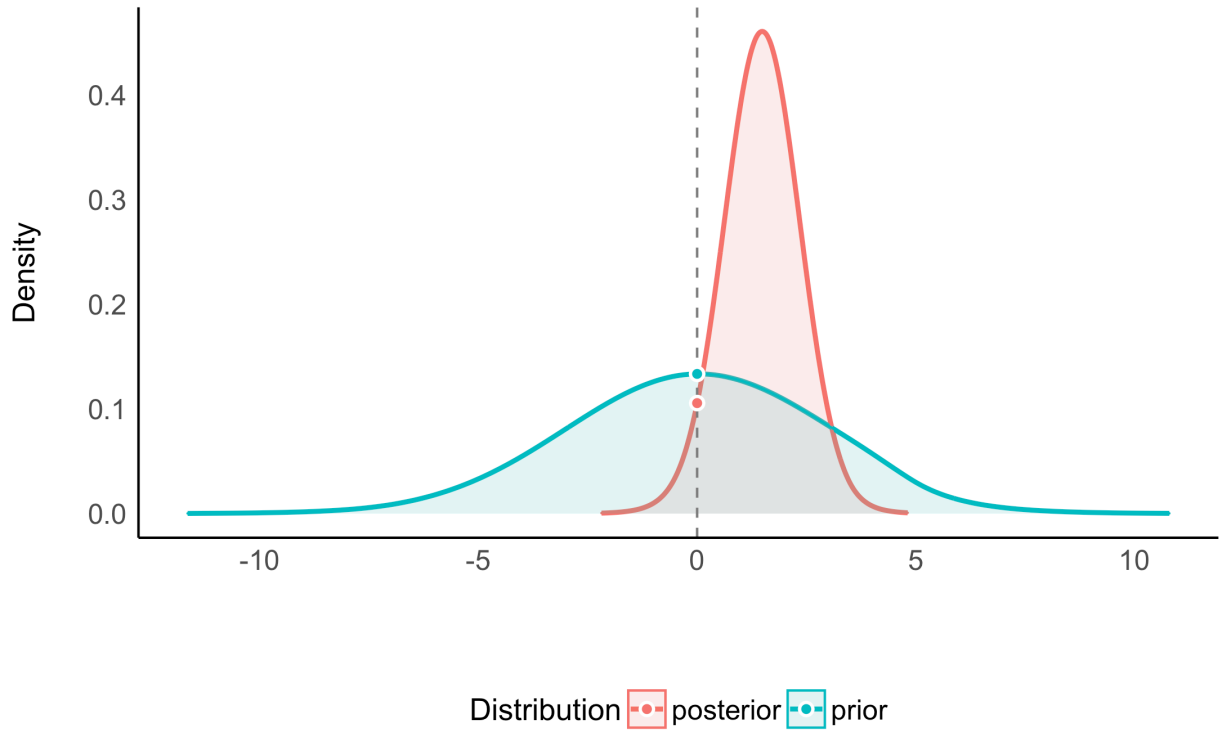
0.91 $BF_{10} = \frac{1}{0.91} = 1.1\ BF_{01}$

Posterior Model Probability:

$PMP_{10} = \frac{0.91}{1+0.91} = .48$

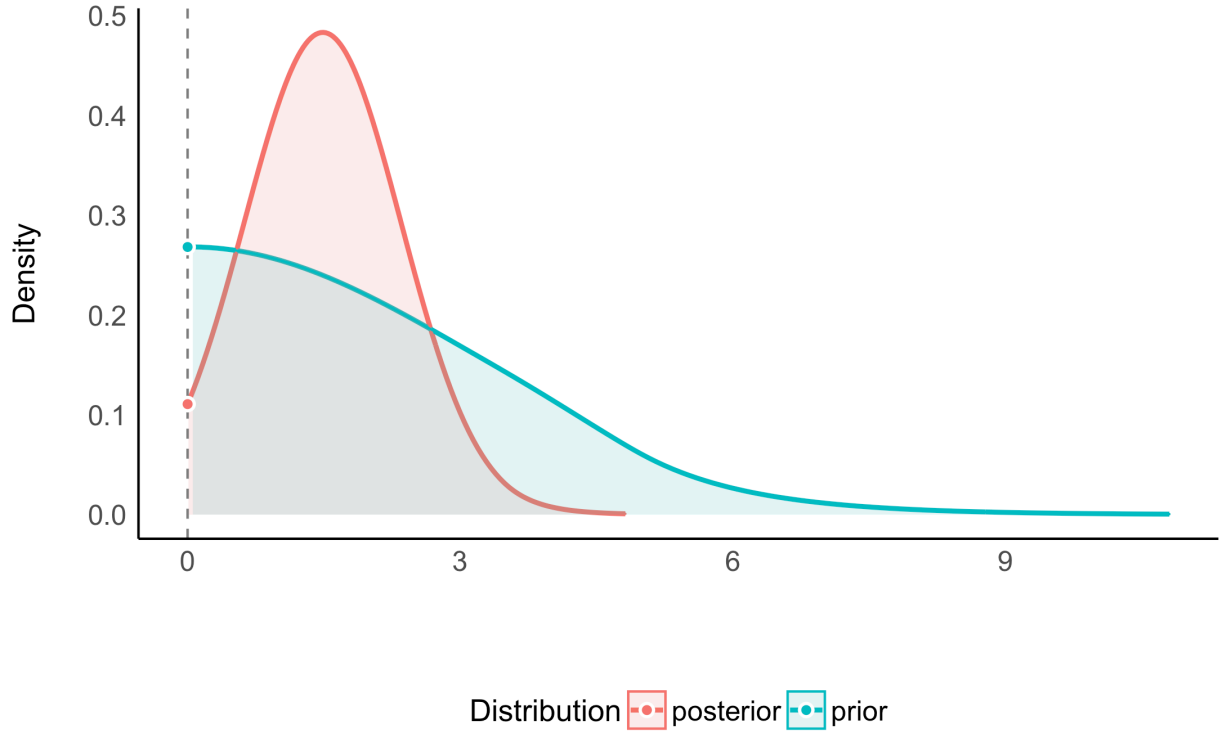
$PMP_{10} = 1-.48 = .52$

Prior and Posterior point-null



- Density of the null compared between prior and posterior = Savage-Dickey ratio
- Approximation of BF
- $\frac{0.11}{0.14} = 0.79$ $BF_{01} = 1.27$ BF_{10}
- $PMP_{01} = \frac{0.79}{1+0.79} = .44$

Prior and Posterior 1-sided point-null



- Prior truncated at 0
- $\frac{0.11}{0.28} = 0.39$ $BF_{01} = 2.55$ BF_{10}
- $PMP_{01} = \frac{0.39}{1+0.39} = .28$

BF notes

- Think about your prior → a diffuse prior functions as a penalty for Ha
- Sensitivity / robustness: show results for different prior settings
- BF ≠ effect size
- BF does not inform about uncertainty in parameter estimation
- BFs can be updated with new data!
- BFs can be multiplied with BFs from other studies!

BF & SEM

- More parameters = more hypotheses
  - H0: B1 = B2 = B3 = B4
  - H1: B1 > B2 > B3 > B4
  - H2: B1 > (B2, B3) > B4
  - H3: B1 < (B2, B3) < B4
  - HA: B1, B2, B3, B4
- bain → BF and posterior model probabilities for sets of (order-restricted) hypotheses

Try it! JASP

- https://www.rensvandeschoot.com/tutorials/jasp-for-bayesian-analyses-with-default-priors/
- In JASP, priors are not easily adjusted. The BayesFactor package on which the JASP Bayes factors are based includes more options

Try it! `bain` in R

- Download BFtutorial.pdf and BFtutorial.R from https://informative-hypotheses.sites.uu.nl/software/bain
- Execute steps 1-5 from BFtutorial.R (discussed in BFtutorial.pdf)
- Then choose what has your interest:
  Bayesian updating, Steps 6-7;
  Sensitivity analysis, Steps 8A+8B;
  the effect of outliers, Step 9;
  Evaluating Informative Hypotheses, Steps 10-11;
  Evaluating a replication study, Steps 12A, 12B, 12C

`bain` SEM (general description)

1. Run your analysis in R (lavaan/blavaan) or in Mplus / other software and import the results to R (e.g., MplusAutomation)
2. Obtain the estimates of interest *and* variance-covariance matrix <u>of the estimates</u> `vcov(fit)`
3. Specify hypotheses in the bain R-function with same estimate names
4. Get the result

See https://osf.io/r2tyk/ for an example → 3 datasets evaluated with lavaan, evaluated with bain + Bayesian evidence synthesis

Try it! bain with SEM in R

See http://developmentaldatascience.org/post/18-08-20_bain/

Follow the steps of Casper van Lissa (bain programmer)

1. Run the model in lavaan and save the output in an object (e.g., fit)
2. Evaluate the output and parameter names in the output
3. Use the bain function to extract the parameter estimates from the fit object and evaluate your specified hypotheses

Key References

Frequentist and Bayesian estimation

- Kruschke Tutorial for the Shiny App, Bayesian and Frequentist Side by Side https://jkkweb.sitehost.iu.edu/KruschkeFreqAndBayesAppTutorial.html

Key References

Bayes introduction

- Van De Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M.A.J., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217-239. https://doi.org/10.1037/met0000100
- Spiegelhalter, D.J., Abrams, K.R., Mules, J.P. (2004) Bayesian approaches to clinical trials and health-care evaluation.

Key References

Bayes factor

- Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). *bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework*. Journal of Open Source Software, 4(40), 1541. https://doi.org/10.21105/joss.01541
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). *Indices of Effect Existence and Significance in the Bayesian Framework*. Retrieved from 10.3389/fpsyg.2019.02767
- https://www.rensvandeschoot.com/tutorials/