Bayesian data analysis

175.746 2017 Lecture 12

Matt Williams

Connections

- Critical reading 7 (Gigerenzer, 2004)
- Critical reading 8 (Kruschke & Liddell, 2017)
- Week 12 (Bayesian data analysis)

- Kruschke, J. K. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan (2nd ed.). San Diego, CA: Academic Press.
- Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2013). A gentle introduction to Bayesian analysis: Applications to developmental research. Child Development, 85(3), 842–860. doi:10.1111/cdev.12169

NHST

NHST is ubiquitous in psychology, but it has many problems.

- A *p* value is a strange way to communicate uncertainty - it tells us the probability of a test statistic as large or larger than that observed, if the effect we're interested in is actually zero.
  - Rather than the probability that the hypothesis itself is true.
- NHST cannot deal with optional stopping (e.g., 5% Type 1 error rate does not hold)
- NHST can't really provide evidence for a null hypothesis – only against it.

NHST

$p <. 05$?

Bayesian vs. Frequentist probability

- In frequentist (conventional) probability, the probability of an event is the relative frequency of the event over multiple trials
  - This means that frequentists can make statements about events whose frequency can be tallied over multiple trials
    - E.g., how often a tossed coin turns up heads
    - Or how often a particular statistic will be observed, if a study was repeated many times

Frequentist interpretation of probability

- A frequentist *cannot* make a useful statement about the probability of something that is simply either true or false
  - E.g., a frequentist cannot say "There is 98% probability that this hypothesis is true"
- Frequentist statistics is the dominant mode of statistical inference
  - Null hypothesis significance tests, confidence intervals, OLS estimation etc. – all frequentist

Bayesian interpretation

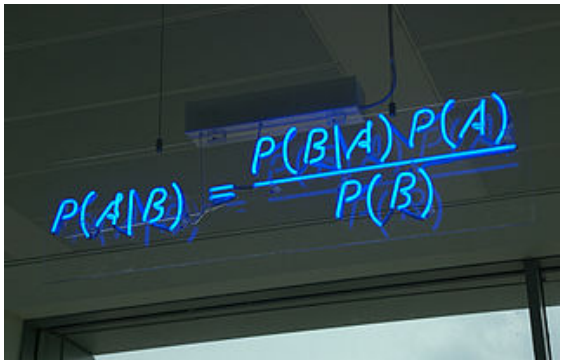
- To a Bayesian, probability is a measure of **certainty** or **belief**
- That belief *might* be based on observations of long run frequencies, but doesn't have to be
- Based on Bayes theorem (Rev. Thomas Bayes)

Bayes theorem

Bayes theorem shows us how to combine:

- Our *prior* beliefs – what we believed before collecting the data at hand, and
- The data collected

To produce a **posterior** probability distribution that represents our beliefs after observing the data at hand



Bayesian analysis advantages

- Allows us to make direct statements about probability - e.g., "There is a 97% probability that your hypothesis is false"
- Remain valid with optional stopping
- Can tell us whether the data supports a null hypothesis.
- Requires us to specify prior beliefs (the data analysis allows us to update those beliefs).

Bayes theorem in action

- Imagine a researcher is screening 10,000 women over 40 for breast cancer
- 100 (1%) actually have breast cancer
- The sensitivity of the test is 75%
  - So of the 100 women with cancer, 75 will receive a positive test result
- The specificity of the test is 96%
  - I.e., of the 9,900 women *without* cancer, 0.04*9,900=396 will receive a false positive test result
- **Mary** gets a positive test result. What's the probability that she has breast cancer?

Over in the frequentist world…

- On the other hand, imagine a significance testing approach to this problem…
- We specify a null hypothesis that Mary does not have cancer
- We observe a positive screening test for Mary
- If the null hypothesis of no cancer was true, this would only happen only 4% of the time
- $p < 0.05$! So we reject the null hypothesis and accept the alternative hypothesis that Mary has cancer.

Bayes theorem

Posterior probability — Likelihood — Prior probability

$$p(A|B) = \frac{p(B|A)\,p(A)}{p(B)}$$

- Bayes theorem formalises the solution to problems such as these, where we have both prior information (the base rate of breast cancer) and new data (the observation of a positive test result).
- In the theorem, *B* is the data we've observed, and A is the hypothesis that we're testing
- So for our example:
  - A = Hypothesis that Mary has cancer
  - B = Observation that Mary receives a positive test result

Bayesian analysis in psychology

Complications:

- In the cancer screening example, we had actual empirical data about the base rate (“prior probability”) of having cancer
  - We won’t usually have direct empirical info about the prior probability that an hypothesis is true
  - Have to formulate priors more indirectly
- Our hypotheses aren’t as simple as “Mary has cancer” (typically they’ll be about continuous parameters)

Bayesian analysis in psychology

- We need to set prior probability distributions on the parameters in our model. E.g., prior to seeing the data:
  - Which values of the intercept are most credible?
  - Which values of the slopes are most credible?
  - How much error variance is there likely to be?
  - Etc.

Where do priors come from?

- We can set prior probability distributions based on:
  - Our subjective beliefs; or
  - Empirical info from previous studies;
  - Known information about average effect sizes in psychology.
    - $r = 0.21$ or $d = 0.43$ is an average effect size in social psychology (Richard, Bond, & Stokes-Zoota, 2003).
  - Ignorance (non-informative prior – assume all possible values equally plausible)

Isn’t that horribly subjective?

- Incorporating prior information seems strange, but data analysts are *always* required to incorporate prior assumptions of some kind.
- E.g., in frequentist regression we rely on prior beliefs that:
  - Errors are normally distributed
  - Errors are independent
  - Errors have identical variance
  - We have included the “right” predictor variables
  - Etc.

Bayesian analysis steps

- Specify a prior (informative or non-informative)
  - E.g., a non-informative prior for a regression slope might be a uniform prior with bounds of $[-\infty, \infty]$
- Collect data and calculate *likelihood* of data for different parameter values
- Use Bayes theorem to combine prior and likelihood to calculate posterior probability distribution

Bayesian analysis - computation

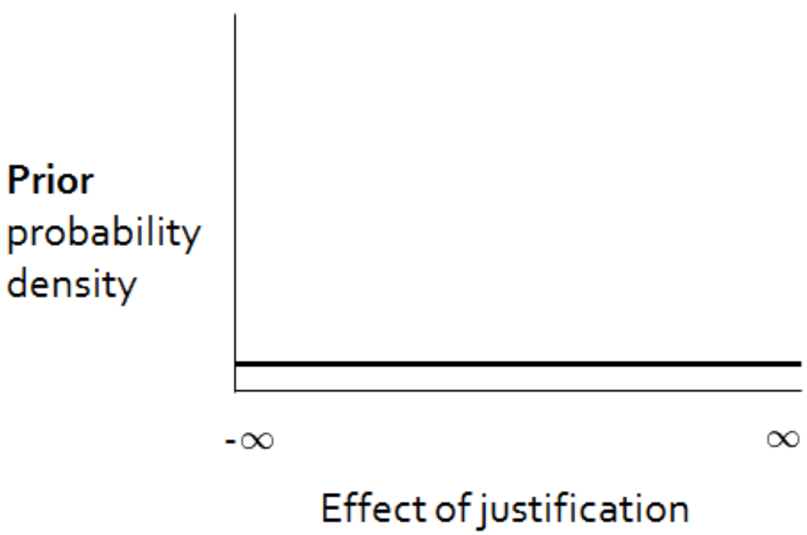
- Underlying mathematics more complex than for frequentist analysis
- Computation can be a little harder than frequentist analysis
  - Bayesian analysis not available in SPSS
- But more and more feasible thanks to easy-to-use computer packages (e.g., JASP, BEST, MCMCpack)

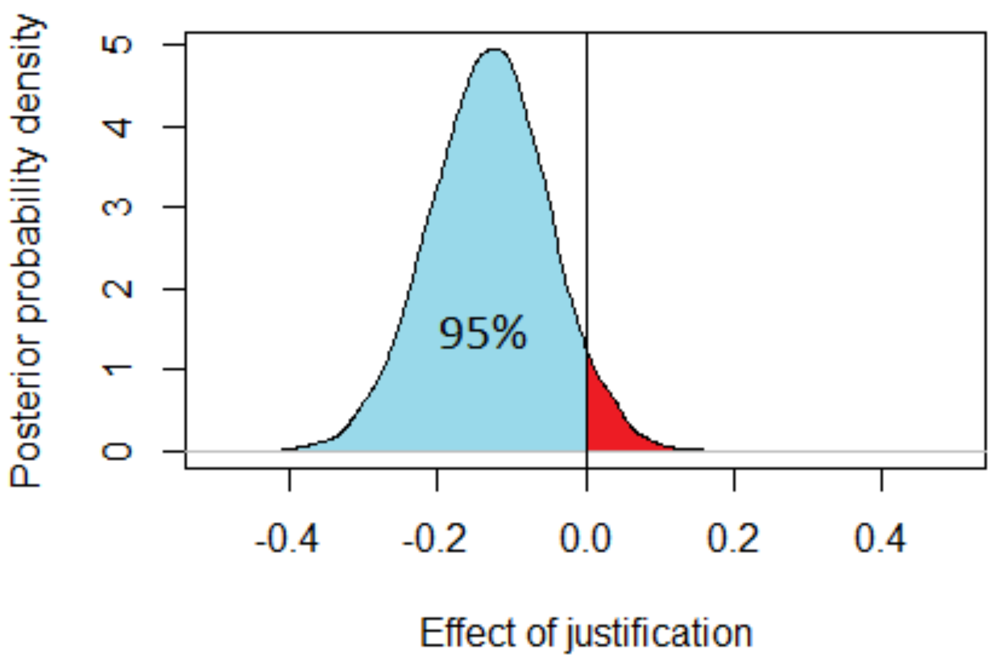
JASP

A Fresh Way to Do Statistics

Bayesian analysis example

- Images from Williams et al. (2014)
- Survey study of relationship between how justified people felt about decisions they regretted, and intensity of regret





Credible interval

- A key output from a Bayesian data analysis
- 95% CI for a correlation of [0.1, 0.3] would literally mean that **we are 95% certain that the true parameter lies in this interval**
- (Not the case for a traditional confidence interval!)

Bayesian interpretation of confidence intervals

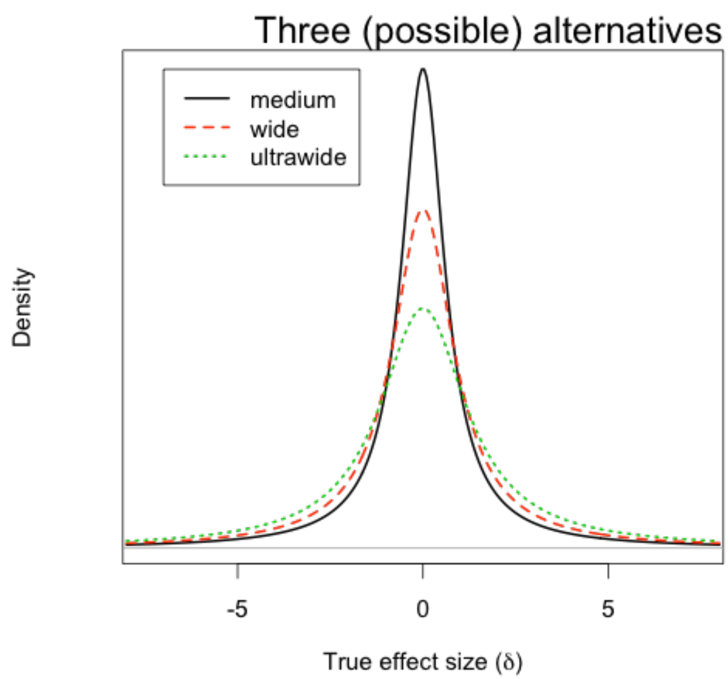
- We can achieve a more intuitive interpretation of a confidence interval by using a Bayesian interpretation
- If we assume a non-informative uniform prior, it **is** reasonable to say that there is a 95% probability that the true parameter falls within the calculated 95% confidence interval (see Greenland & Poole, 2013).
    - I.e., we're assuming that before the study we knew absolutely nothing about which parameter values were more likely – any effect size from negative infinity to positive infinity equally likely
    - Typically in psychology we know that small effects are more likely though – should ideally take this info into account.

Bayes factors

- An approach in between Bayesian analysis and frequentist statistics
- Focuses on comparing the likelihood of the *data* under two different models (null and alternative hypotheses)
- Typically involve a null hypothesis that a parameter is exactly zero, and alternative hypothesis that it is nonzero
  - But unlike the case in NHST, the alternative hypothesis specifically outlines which values of the parameter are most probable (if the true value isn't exactly zero)
  - We specify a prior for the parameter <u>under the alternative hypothesis</u>

Bayes factors

- To calculate the probability of the data under the alternative model, we must specify a prior on effect size. I.e., if the effect size is not exactly zero, which effect sizes are more and less credible?
  - E.g., some default options in the BayesFactor package shown in https://richarddmorey.org/2014/02/bayes-factor-t-tests-part-2-two-sample-tests/



Bayes factors

- We can then calculate the likelihood of the data if the null hypothesis was true, and the likelihood of the data if the alternative was true
- The Bayes Factor is the ratio of the two likelihood values
- It tells us about which hypothesis the data is more consistent with
  - Major advantage: Is capable of **supporting** a null hypothesis (not just failing to reject it)
- E.g., an ego depletion replication study: Bayes factor of 2.4 in favour of null (Lurquin et al., 2016)
- Easy online calculators available for Bayes Factor alternatives to common inferential tests, e.g., http://pcl.missouri.edu/bayesfactor

Conclusions

- Bayesian analysis allows us to make more direct and useful statements about uncertainty
- Avoids some limitations of NHST
  - Invalidity under optional stopping
  - Inability to tell us probability that hypothesis is true
  - Inability to provide evidence *for* a null hypothesis (although cf. equivalence testing)
- Challenging to specify priors and run computations but not impossible
- I encourage you to consider using this in your own data analyses (one day).

References

Greenland, S., & Poole, C. (2013). Living with *P* values: Resurrecting a Bayesian perspective on frequentist statistics. *Epidemiology*, *24*(1), 62–68. https://doi.org/10.1097/EDE.0b013e3182785741

Lurquin, J. H., Michaelson, L. E., Barker, J. E., Gustavson, D. E., Bastian, C. C. von, Carruth, N. P., & Miyake, A. (2016). No evidence of the ego-depletion effect across task characteristics and individual differences: A pre-registered study. *PLOS ONE*, *11*(2), e0147770. https://doi.org/10.1371/journal.pone.0147770

Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331–363. https://doi.org/10.1037/1089-2680.7.4.331

Williams, M. N., Towers, A., Hill, S., & Philipp, M. C. (2014). Predictably regretful: A comparison of the effects of time, domain, justification, and life rule contradiction on the intensity of regrets. In *New Zealand Psychological Society*. Nelson, New Zealand. Retrieved from http://hdl.handle.net/10179/5753