Bayesian Statistics in a Nutshell

Intended for students of psychological and/or organizational sciences

Prepared by Andrew Jebb and Sang Eun Woo
August 27, 2014 (Updated August 1, 2020)

Bayesian statistics is an approach to statistical inference that is fundamentally different than the conventional frequentist approach. Bayesian methods derive their name from Bayes' Theorem, a mathematical equation built from simple probability axioms. Bayes' Theorem is important because allows us to calculate any *conditional probability* of interest. A conditional probability is defined as the probability of event A given that event B has occurred. It is a probability that is therefore "conditional" on another event. A paradigm of statistics can be built off of it because analyses are based on a collection of sample data. Given that the data have already occurred, , we can use Bayes' theorem to directly calculate the probability of different population values given ("conditional on") this already observed data. Once one becomes familiar with the underlying logic and terminology, Bayesian statistics is an intuitive approach to statistics that allows us to make direct probability statements regarding the population values we are interested in.

As we elaborate later, three primary advantages of Bayesian statistics are that it is:
1) A remarkably *rich source of information* on which to draw conclusions;
2) A natural framework to include *previous information*;
3) *Flexible*--e.g., accommodating complex models and small samples.

BAYESIAN VS. FREQUENTIST PARADIGMS

In general, there are two aspects of statistical inference: *estimation* and *hypothesis testing*. Estimation is concerned with finding a value that can accurately represent the population value (i.e., parameter). By contrast, hypothesis testing involves formally testing competing specific *statistical hypotheses*—i.e., hypotheses that the parameter is some specific (e.g., the population value is zero or not).

Both the Bayesian and frequentist statistical paradigms have their unique approaches to *estimation* and *hypothesis testing*. The basis of these differences lies in their philosophical differences about how probability should be conceived. However, **one does not have to subscribe to the Bayesian or frequentist notion of probability to use these statistics in practice** -- This is an important point because researchers who currently practice frequentist statistics are often unaware of its theoretical foundations.

Due to its view of probability, estimation in a frequentist paradigm tries to locate a *single* parameter estimate that best fits the data. It is possible (and strongly recommended by many methodologists and the Psychological Sciences journal) to then provide a range of plausible values around that point estimate through the use of a *confidence interval*, indicating the *precision* of the estimate. However, this is not done as often as recommended (Finch, Cumming & Thomason, 2001; Finch, Cumming, Williams, Palmer, Griffith, Alders, Anderson, & Goodman, 2004). Instead, researchers rely more on *null hypothesis significance testing* (NHST), a test of the hypothesis that the population value is zero (i.e., "null"). The practical and conceptual problems associated with NHST have been catalogued in a literature spanning decades (e.g., Anderson, Burnham, & Thompson, 2000; Cohen, 1994; Gigerenzer, 2004; Gigerenzer, Krauss, & Vitouch, 2004; Johnson, 1999; Kline, 2004; Kruschke, 2010; Meehl, 1978; Morrison & Henkel, 1970; Rozeboom, 1960; Schmidt, 1996; Simmons, Nelson, and Simonsohn, 2011; Wagenmakers, 2007). Although frequentist estimation is useful for research (see Cumming, 2014), NHST is problematic in that it reduces data analysis to a binary decision about whether the effect exists (Gigerenzer, 2004; Gigerenzer et al., 2004).

FUTHER REMARKS ON NHST

Many concerns have been articulated about the deficiencies and problems of NHST (as cited above), which we summarize into four points:

1) NHST can never provide evidence that the null hypothesis is true; the null can only be rejected, never accepted.

2) When one fails to reject the null hypothesis, nothing can be concluded from the results. This is an impressive waste of researcher time and resources (and sanity).

3) The null hypothesis that a population value (parameter) is precisely zero is perhaps never actually true; there is almost always an effect, even if very small, and indefinitely increasing the *statistical power* (i.e., sample size) will usually guarantee a "significant" result. Thus, even before conducting a significance test, the researcher basically already knows its outcome. This encourages researchers to try and inflate their ability to reject the null. (Bakker, van Dijk, & Wicherts, 2012; O'Boyle, Banks, & Gonzalez-Mule, 2017).

4) A result that is "statistically significant" (i.e., when the null hypothesis is rejected) does not entail that it is *practically significant*. In other words, statistical significance is just a way to formally assess whether an estimated effect is actually there or not; it does not actually bear on the research questions that impel scholars.

THE BAYESIAN ALTERNATIVE

In contrast, Bayesian inference avoids these problems. It is a logically sound way to perform statistical inference that is rapidly growing. Bayesian statistical inference is done primarily through estimation, with Bayesian hypothesis testing reserved for model selection (see Kass & Raftery, 1995 for a review). Bayesian statistics avoids the myriad problems associated with NHST. It also has significant intrinsic benefits.

The most important benefit is that Bayesian estimation provides a remarkably rich source of information on which to draw conclusions: Each estimated parameter is represented in a *probability distribution*, where each potential parameter value is probabilistically weighted, allowing the analyst to see how probable each potential parameter value is.

Another substantial advantage of the Bayesian paradigm is that it provides a natural framework to include previous information. This is accomplished through the use of prior distributions for each statistical parameter, which quantify the analyst's prior certainty (or uncertainty) regarding its possible values. The progression of science is based on the accumulation of research findings. That is, science builds off of itself, and our methods of data analysis should reflect this fact. Therefore, using Bayesian statistics allows researchers to formally integrate what is already known on the topic of interest within present analyses (Zyphur & Oswald, 2013).

Finally, Bayesian estimation is great for research where large samples are difficult or impossible to obtain, is more intuitive than frequentist methods, and can accommodate the increasingly complex models seen in contemporary research (for a more complete list of advantages, see Kruschke, Aguinis, & Joo, 2012, pp.730-739).

TECHNICAL & PRACTICAL ISSUES IN USING BAYESIAN STATISTICS

In spite of its advantages, Bayesian statistics requires that the analyst learn more about probability and statistical theory (although this can be considered an advantage). Also, intuitive software implementation of these analyses is still in development. Some popular programs for Bayesian estimation are WinBUGS/OpenBUGS and Stan. However, these require notable programming knowledge and a very in-depth understanding of the Bayesian modeling process. **BugsXLA is a software tool for Bayesian estimation that is much more intuitive for social scientists.** Instead of learning software coding, one sets up the model using a very simple GUI (graphical user interface) in Excel. The program then automatically uses the WinBUGS engine to conduct the analysis and imports the results directly back into Excel. Bayesian analyses have also started to become standard in STATA and SPSS.

We have used BugsXLA as the focal software in our paper, "A Bayesian Primer for the Organizational Sciences: The "Two Sources" and an Introduction to BugsXLA" (Jebb & Woo, 2015), which we hope will encourage our fellow scientists to begin exploring and using Bayesian methods. Click here for the Excel data for running the analyses. Of course, to run the analysis, you will need to install BugsXLA. You will also need to install WinBUGS.

RECOMMENDED READINGS

There are a lot of great resources for those interested in learning more about Bayesian statistics. In our opinion, the two most accessible texts for social scientists are **Scott Lynch's (2007)** ***Introduction to applied Bayesian statistics and estimation for social scientists*** and **John Kruschke's (2011)** ***Doing Bayesian data analysis: A tutorial with R and BUGS*****.** We recommend that these be read in complement, as some concepts are explained more intuitively in one than in the other.

For those with a slightly stronger background in statistics, we recommend **Gelman, Carlin, Stern, and Rubin (2004)**. For many, it remains the "classic" textbook on Bayesian modeling. It is our experience that other books are geared toward a statistics audience and may not be accessible to social scientists.

Last but not least, there are also many good journal articles delineating Bayesian methods, such as **Kruschke (2013), Kruschke et al. (2012),** and **Zyphur and Oswald (2013).**

In sum, we believe that research in every domain has much to gain from an increased use in Bayesian data analysis. We also believe that these methods have been made much more accessible in recent years and hope that their use will continue to spread for many years to come.

REFERENCES

Anderson, D.R., Burnham, K.P., & Thompson, W.L. (2000). Null hypothesis testing: Problems, prevalence and an alternative. *Journal of Wildlife Management, 64*, 912–923.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554.

Cohen, J. (1994). The earth is round (*p* < .05). *American Psychologist, 49*, 997–1003.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7-29. doi: 10.1177/0956797613504966

Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical inference in the *Journal of Applied Psychology*: Little evidence of reform. *Educational and Psychological Measurement*, *61*, 181-210.

Finch, S., Cumming, G., Williams, J., Palmer, L., Griffith, E., Alders, C…Goodman, O. (2004). Reform of statistical inference in psychology: The case of *Memory & Cognition. Behavior Research Methods, Instruments & Computers, 36*, 312-324.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, Florida: CRC Press.

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

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about null hypothesis testing but were afraid to ask. In D. Kaplan (Ed.), *Handbook on quantitative methods in the social sciences*, 389-406. Thousand Oaks, CA: Sage.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124. Retrieved from http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0020124

Jebb, A. T. & Woo, S. E. (2015). A Bayesian primer for the organizational sciences: The “two sources” and introduction to BugsXLA. Organizational Research Methods, 18(1), 92-132.

Johnson, D. H. (1999). The insignificance of statistical significance testing. *Journal of Wildlife Management, 63*, 763: 772.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*, 293-300.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Burlington, MA: Academic Press/Elsevier.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t-test. *Journal of Experimental Psychology: General*, *142*, 573-603.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organi-zational sciences.*Organizational Research Methods*, *15*, 722-752.

Lavine, M. (1999). What is Bayesian statistics and why everything else is wrong.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer.

Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834.

Morrison, D.E., & Henkel, R.E. (Eds.). (1970). *The significance test controversy*. Chicago: Aldine.

O'Boyle, E., Banks, G., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*, 43(2), pp. 376-399.

Orlitzky, M. (2012). How can significance tests be deinstitutionalized? *Organizational Research Methods, 15*, 199-228.

Rozeboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin, 57,* 416-428.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers.*Psychological Methods, 1*, 115-129.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779-804.

Zyphur, M. J., & Oswald, F. L. (2013). Bayesian estimation and inference: A user's guide. *Journal of Management.* doi: 10.1177/014920631350120

OTHER RESOURCES

- https://bayesmodels.com/
- https://sites.google.com/site/bayesianresearch/home
- https://drive.google.com/file/d/1bVnxCgKBFamCb2iT7TZWaZtMP9LHYkAE/view
- http://doingbayesiandataanalysis.blogspot.com/