Cognitive Psychology 60 (2010) 158–189

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method

Eric-Jan Wagenmakers [a,*], Tom Lodewyckx [b], Himanshu Kuriyal [c], Raoul Grasman [a]

[a] *University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands*
[b] *Leuven University, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium*
[c] *Indian Institute of Technology, Kharagpur, India*

ARTICLE INFO

Article history:
Accepted 14 December 2009
Available online 12 January 2010

Keywords:
Statistical evidence
Model selection
Bayes factor
Hierarchical modeling
Random effects
Order-restrictions

ABSTRACT

In the field of cognitive psychology, the *p*-value hypothesis test has established a stranglehold on statistical reporting. This is unfortunate, as the *p*-value provides at best a rough estimate of the evidence that the data provide for the presence of an experimental effect. An alternative and arguably more appropriate measure of evidence is conveyed by a Bayesian hypothesis test, which prefers the model with the highest average likelihood. One of the main problems with this Bayesian hypothesis test, however, is that it often requires relatively sophisticated numerical methods for its computation. Here we draw attention to the *Savage–Dickey density ratio* method, a method that can be used to compute the result of a Bayesian hypothesis test for nested models and under certain plausible restrictions on the parameter priors. Practical examples demonstrate the method's validity, generality, and flexibility.

1. Introduction

Inside every Non-Bayesian, there is a Bayesian struggling to get out – Dennis Lindley, as cited in Jaynes (2003).

How do cognitive psychologists analyze their data? Gert Gigerenzer answered this question by invoking the Freudian concept of unconscious conflict between the Superego, the Ego, and the Id (Gigerenzer, 1993, 2004; Gigerenzer, Krauss, & Vitouch, 2004). In Gigerenzer's analogy, the cognitive

* Corresponding author. Fax: +31 20 639 0279.
E-mail address: EJ.Wagenmakers@gmail.com (E.-J. Wagenmakers).

0010-0285/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2009.12.001

psychologist's Superego wants to follow the Neyman–Pearson tradition; it seeks to contrast two well-defined hypotheses (i.e., the null hypothesis and an alternative hypothesis), it operates using concepts of α-level and power, and it is generally concerned with procedures that will work well in the long run. In contrast, the cognitive psychologist's Ego follows the Fisherian tradition; it does not posit a specific alternative hypothesis, it ignores power, and it computes a p-value that is supposed to indicate the statistical evidence against the null hypothesis. Finally, the cognitive psychologist's Id is *Bayesian*, and it desperately wants to attach probabilities to hypotheses. However, this wish is suppressed by the Superego and Ego. In its continual struggle to obtain what it desires, the Id—although unable to change the statistical analysis procedures that are used—wields its influence to change and distort the interpretations that these analysis procedures afford.[1]

The unconscious Freudian conflict has arguably resulted in widespread confusion. Researchers often assume that a small p-value means that the null hypothesis is likely to be false, that a large p-value means that the null hypothesis is likely to be true, and that a 95% confidence interval for a parameter μ means that there is a 95% chance that μ lies in the specified interval. All of these conclusions are false (Haller & Krauss, 2002)—this is because the conclusions are Bayesian, but the methodology that is used is not.

To resolve the unconscious Freudian conflict and bring the statistical procedures in line with their interpretation, two courses of action present themselves. First, one can try to suppress the Id even more strongly, perhaps by rigorous statistical education and repeated warnings such as "Never use the unfortunate expression 'accept the null-hypothesis'." (Wilkinson & the Task Force on Statistical Inference., 1999, p. 599). Second, one can explore Bayesian statistical procedures that provide exactly what the Id wants—probabilities for hypotheses. Using Bayesian procedures, one can quantify support both in favor of and against the null hypothesis (Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wetzels, Raaijmakers, Jakab, & Wagenmakers, 2009), and one can state that the probability that a parameter μ lies in a 95% "credible interval" is, indeed, .95. In this article, we promote the second course of action.

In order to keep this article self-contained, we first provide a brief overview of the Bayesian paradigm, with special emphasis on the difference between parameter estimation and hypothesis testing. We then describe a method, known as the Savage–Dickey density ratio, to carry out a Bayesian hypothesis test with relative ease. Next we illustrate the practical value of the Savage–Dickey method by applying it to three data sets. The first data set is used to test the hypothesis that the sexual behavior of so-called virginity pledgers differs from that of non-pledgers (i.e., a hypothesis test for the equality of two rates, Brückner & Bearman, 2005); the second data set is used to test the hypothesis that prior study of both choice alternatives improves later performance in a two-choice perceptual identification task (i.e., a hypothesis test in a hierarchical within-subjects design, Zeelenberg, Wagenmakers, & Raaijmakers, 2002); and the third data set is used to test the hypothesis that typically developing children outperform children with ADHD on the Wisconsin card sorting test (i.e., a hypothesis test in a hierarchical between-subjects design, Geurts, Verté, Oosterlaan, Roeyers, & Sergeant, 2004).

In these examples, we show how the Bayesian hypothesis test can be adjusted to deal with random effects and order-restrictions, both for within-subjects and between-subjects designs. WinBUGS code is presented in Appendix B and R code is available online.[2]

2. Bayesian background

Before outlining the Savage–Dickey method, it is important to introduce some key concepts of Bayesian inference. More detailed information can be found in Bayesian articles and books that discuss philosophical foundations (Lindley, 2000; O'Hagan & Forster, 2004), computational innovations (Gamerman & Lopes, 2006), and practical contributions (Congdon, 2003; Ntzoufras, 2009). An in-depth discussion on the advantages of Bayesian inference, especially when compared to p-value

[1] For more information about the difference between the three statistical paradigms, see for instance Christensen (2005), Hubbard and Bayarri (2003) and Royall (1997).

[2] All computer code is available from the first author's website, http://users.fmg.uva.nl/ewagenmakers/papers.html.