ARTICLE IN PRESS
Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Contents lists available at ScienceDirect
Journal of Mathematical Psychology
journal homepage: www.elsevier.com/locate/jmp

Automatic Bayes factors for testing variances of two independent normal distributions

Florian Böing-Messing *, Joris Mulder

Department of Methodology and Statistics, Tilburg University, Postbus 90153, 5000 LE Tilburg, The Netherlands

HIGHLIGHTS

- We develop three automatic Bayes factors for testing two variances.
- We consider a fractional, a balanced, and an adjusted fractional Bayes approach.
- The Bayes factors do not require prior elicitation and are thus fully automatic.
- We evaluate the methods based on theoretical properties and numerical performance.
- The adjusted fractional Bayes factor performs best overall.

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Automatic Bayes factor
Automatic prior
Balancedness
Fractional Bayes factor
Homogeneity of variance
Inequality constrained hypothesis
Occam's razor

ABSTRACT

Researchers are frequently interested in testing variances of two independent populations. We often would like to know whether the population variances are equal, whether population 1 has smaller variance than population 2, or whether population 1 has larger variance than population 2. In this article we consider the Bayes factor, a Bayesian model selection and hypothesis testing criterion, for this multiple hypothesis test. Application of Bayes factors requires specification of prior distributions for the model parameters. Automatic Bayes factors circumvent the difficult task of prior elicitation by using data-driven mechanisms to specify priors in an automatic fashion. In this article we develop different automatic Bayes factors for testing two variances: first we apply the fractional Bayes factor (FBF) to the testing problem. It is shown that the FBF does not always function as Occam's razor. Second we develop a new automatic balanced Bayes factor with equal priors for the variances. Third we propose a Bayes factor based on an adjustment of the marginal likelihood in the FBF approach. The latter two methods always function as Occam's razor. Through theoretical considerations and numerical simulations it is shown that the third approach provides strongest evidence in favor of the true hypothesis.

1. Introduction

Researchers are frequently interested in comparing two independent populations on a continuous outcome measure. Traditionally, the focus has been on comparing means, whereas variances are mostly considered nuisance parameters. However, by regarding variances as mere nuisance parameters, one runs the risk of overlooking important information in the data. The variability of a population is a key characteristic which can be the core of a research question. For example, psychological research frequently investigates differences in variability between males and females (e.g. Arden & Plomin, 2006; Borkenau, Hřebíčková, Kuppens, Realo, & Allik, 2013; Feingold, 1992).

In this article we consider a Bayesian hypothesis test on the variances of two independent populations. The Bayes factor is a well-known Bayesian criterion for model selection and hypothesis testing (Jeffreys, 1961; Kass & Raftery, 1995). Unlike the p-value, which is often misinterpreted as an error probability (Hubbard & Armstrong, 2006), the Bayes factor has a straightforward interpretation as the relative evidence in the data in favor of a hypothesis as compared to another hypothesis. Moreover, contrary to p-values, the Bayes factor is able to quantify evidence in favor of a null hypothesis (Wagenmakers, 2007). Another useful property, which is not shared by p-values, is that the Bayes factor can straightforwardly be used for testing multiple hypotheses simultaneously (Berger & Mortera, 1999). These and other notions have resulted in a considerable development of Bayes factors for frequently

* Corresponding author.
E-mail address: florian.boeingmessing@gmail.com (F. Böing-Messing).

http://dx.doi.org/10.1016/j.jmp.2015.08.001
0022-2496/

Please cite this article in press as: Böing-Messing, F., & Mulder, J., Automatic Bayes factors for testing variances of two independent normal distributions. Journal of Mathematical Psychology (2015), http://dx.doi.org/10.1016/j.jmp.2015.08.001

encountered testing problems in the last decade. For example, Klugkist, Laudy, and Hoijtink (2005) proposed Bayes factors for testing analysis of variance models. Rouder, Speckman, Sun, Morey, and Iverson (2009) proposed a Bayesian t-test. Mulder, Hoijtink, and de Leeuw (2012) developed a software program for Bayesian testing of (in)equality constraints on means and regression coefficients in the multivariate normal linear model, and Wetzels and Wagenmakers (2012) proposed Bayesian tests for correlation coefficients. The goal of this article is to extend this literature by developing Bayes factors for testing variances. For more interesting references we also refer the reader to the special issue 'Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments' in the *Journal of Mathematical Psychology* in which this article appeared (Mulder & Wagenmakers, in preparation).

In applying Bayes factors for hypothesis testing, we need to specify a prior distribution of the model parameters under every hypothesis to be tested. A prior distribution is a probability distribution describing the probability of the possible parameter values before observing the data. In the case of testing two variances, we need to specify a prior for the common variance under the null hypothesis and for the two unique variances under the alternative hypothesis. Specifying priors is a difficult task from a practical point of view, and it is complicated by the fact that we cannot use noninformative improper priors for parameters to be tested because the Bayes factor would then be undefined (Jeffreys, 1961). This has stimulated researchers to develop Bayes factors which do not require prior elicitation using external prior information. Instead, these so-called automatic Bayes factors use information from the sample data to specify priors in an automatic fashion. So far, however, no automatic Bayes factors have been developed for testing variances.

In this article we develop three types of automatic Bayes factors for testing variances of two independent normal populations. We first consider the fractional Bayes factor (FBF) of O'Hagan (1995) and apply it for the first time to the problem of testing variances. In the FBF methodology the likelihood of the complete data is divided into two fractions: one for specifying the prior and one for testing the hypotheses. However, it has been shown (e.g. Mulder, 2014b) that the FBF may not be suitable for testing inequality constrained hypotheses (e.g. variance 1 is smaller than variance 2) because it may not function as Occam's razor. In other words, the FBF may not prefer the simpler hypothesis when two hypotheses fit the data equally well. This is a consequence of the fact that in the FBF the automatic prior is located at the likelihood of the data. We develop two novel solutions to this problem: the first is an automatic Bayes factor with equal automatic priors for both variances under the alternative hypothesis. This methodology is related to the constrained posterior priors approach of Mulder, Hoijtink, and Klugkist (2010). The second novel solution is an automatic Bayes factor based on adjusting the definition of the FBF such that the resulting automatic Bayes factor always functions as Occam's razor. This approach is related to the work of Mulder (2014b), with the difference that our method results in stronger evidence in favor of a true null hypothesis.

The remainder of this article is structured as follows. In the next section we provide details on the normal model to be used and introduce the hypotheses we shall be concerned with. We then discuss five theoretical properties which are used for evaluating the automatic Bayes factors. Following this, we develop the three automatic Bayes factors and evaluate them according to the theoretical properties. Subsequently, the performance of the Bayes factors is investigated by means of a small simulation study. We conclude the article with an application of the Bayes factors to two empirical data examples and a discussion of possible extensions and limitations of our approaches.

2. Model and hypotheses

We assume that the outcome variable of interest, X, is normally distributed in both populations:

$$X_j \sim \mathrm{N}\left(\mu_j, \sigma_j^2\right), \quad j = 1, 2, \tag{1}$$

where j is the population index and μ_j and σ_j^2 are the population-specific parameters. The unknown parameter in this model is $\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)' = \left(\left(\mu_1, \mu_2, \right)', \left(\sigma_1^2, \sigma_2^2\right)'\right)' \in \mathbb{R}^2 \times \Omega_u$, where $\Omega_u := \left(\mathbb{R}^+\right)^2$ is the unconstrained parameter space of $\boldsymbol{\sigma}^2$.

In this article we shall be concerned with testing the following nonnested (in)equality constrained hypotheses against one another:

$$\begin{array}{lcl} H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma^2, & & H_0\colon \sigma^2 \in \Omega_0 := \mathbb{R}^+, \\ H_1\colon \sigma_1^2 < \sigma_2^2, & \Leftrightarrow & H_1\colon \boldsymbol{\sigma}^2 \in \Omega_1 := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\right\}, \\ H_2\colon \sigma_1^2 > \sigma_2^2, & & H_2\colon \boldsymbol{\sigma}^2 \in \Omega_2 := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 > \sigma_2^2\right\}, \end{array} \tag{2}$$

where $\Omega_1, \Omega_2 \subset \Omega_u$ and Ω_0 denote the parameter spaces under the corresponding (in)equality constrained hypotheses.

We made two choices in formulating the hypotheses in Eq. (2). First, we do not test any constraints on the mean parameters μ_1 and μ_2. This is because the objective of this article is to provide a Bayesian alternative to the classical frequentist procedures for testing two variances. For a general framework for testing (in)equality constrained hypotheses on mean parameters, see, for example, Mulder et al. (2012). The second choice we made is to divide the classical alternative hypothesis $H_a\colon \sigma_1^2 \neq \sigma_2^2 \Leftrightarrow H_a\colon \sigma_1^2 < \sigma_2^2 \vee \sigma_1^2 > \sigma_2^2$ into two separate hypotheses, $H_1\colon \sigma_1^2 < \sigma_2^2$ and $H_2\colon \sigma_1^2 > \sigma_2^2$ ($\vee$ denotes logical disjunction and reads "or"). The advantage of this approach is that it allows us to quantify and compare the evidence in favor of a negative effect (H_1) and a positive effect (H_2). This is of great interest to applied researchers, who would often like to know not only whether there is an effect, but also in what direction.

Another hypothesis we will consider is the unconstrained hypothesis

$$H_u\colon \sigma_1^2, \sigma_2^2 > 0 \Leftrightarrow H_u\colon \boldsymbol{\sigma}^2 \in \Omega_u = \left(\mathbb{R}^+\right)^2. \tag{3}$$

This hypothesis is not of substantial interest to us because it is entirely covered by the hypotheses in Eq. (2). In other words, $\{H_0, H_1, H_2\}$ is a partition of H_u. The unconstrained hypothesis will be used to evaluate theoretical properties of the priors and Bayes factors such as balancedness and Occam's razor (discussed in the next section).

3. Properties for the automatic priors and Bayes factors

Based on the existing literature on automatic Bayes factors, we shall focus on the following theoretical properties when evaluating the automatic priors and Bayes factors:

1. *Proper priors: The priors must be proper probability distributions*. When using improper priors on parameters that are tested, the resulting Bayes factors depend on unspecified constants (see, for instance, O'Hagan, 1995). Improper priors may only be used on common nuisance parameters that are present under all hypotheses to be tested (Jeffreys, 1961).
2. *Minimal information: Priors under composite hypotheses should contain the information of a minimal study*. Using arbitrarily vague priors gives rise to the Jeffreys–Lindley paradox (Jeffreys, 1961; Lindley, 1957), whereas priors containing too much information about the parameters will dominate the data. Therefore it is often suggested to let the prior contain the