Short communication

An encompassing prior generalization of the Savage–Dickey density ratio

Ruud Wetzels *, Raoul P.P.P. Grasman, Eric-Jan Wagenmakers

University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

An encompassing prior (EP) approach to facilitate Bayesian model selection for nested models with inequality constraints has been previously proposed. In this approach, samples are drawn from the prior and posterior distributions of an encompassing model that contains an inequality restricted version as a special case. The Bayes factor in favor of the inequality restriction then simplifies to the ratio of the proportions of posterior and prior samples consistent with the inequality restriction. This formalism has been applied almost exclusively to models with inequality or "about equality" constraints. It is shown that the EP approach naturally extends to exact equality constraints by considering the ratio of the heights for the posterior and prior distributions at the point that is subject to test (i.e., the Savage–Dickey density ratio). The EP approach generalizes the Savage–Dickey ratio method, and can accommodate both inequality and exact equality constraints. The general EP approach is found to be a computationally efficient procedure to calculate Bayes factors for nested models. However, the EP approach to exact equality constraints is vulnerable to the Borel–Kolmogorov paradox, the consequences of which warrant careful consideration.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In this article we focus on Bayesian model selection for nested models. Consider, for instance, a parameter vector $\theta = (\psi, \phi) \in \Theta \subseteq \Psi \times \Phi$ and suppose we want to compare an encompassing model M_e to a restricted version $M_1 : \psi = \psi_0$. Then, after observing the data D, the Bayes factor in favor of M_1 is

$$BF_{1e} = \frac{p(D|M_1)}{p(D|M_e)} = \frac{\int p(D|\psi = \psi_0, \phi)p(\psi = \psi_0, \phi)\mathrm{d}\phi}{\iint p(D|\psi, \phi)p(\psi, \phi)\mathrm{d}\psi\,\mathrm{d}\phi}.$$

Thus, the Bayes factor is the ratio of the marginal likelihoods of two competing models; alternatively, the Bayes factor can be conceptualized as the change from prior model odds $p(M_1)/p(M_e)$ to posterior model odds $p(M_1|D)/p(M_e|D)$ (Kass and Raftery, 1995). The Bayes factor quantifies the evidence that the data provide for one model versus another, and as such it represents "the standard Bayesian solution to the hypothesis testing and model selection problems" (Lewis and Raftery, 1997, p. 648).

Unfortunately, for most models the Bayes factor cannot be obtained in analytic form. Several methods have been proposed to estimate the Bayes factor numerically (see Gamerman and Lopes (2006, Chap. 7) for a description of 11 such methods). Nevertheless, calculation of the Bayes factor often remains a computationally complicated task.

Here we first describe an encompassing prior (EP) approach that was recently proposed by Hoijtink, Klugkist, and colleagues (Klugkist et al., 2005a,b; Hoijtink et al., 2008). The EP approach applies to nested models and virtually eliminates the computational complications inherent in most other methods. Next we show that the EP approach is a generalization

---

* Corresponding address: University of Amsterdam, Roetersstraat 15, room 5.29, 1018WB, Amsterdam, The Netherlands. Tel.: +31 20 5258871; fax: +31 20 6390026.

E-mail address: wetzels.ruud@gmail.com (R. Wetzels).

of the Savage–Dickey density ratio. Finally, we discuss the Borel–Kolmogorov paradox and examine the implications of this paradox for the EP approach.

2. Bayes factors from the encompassing prior approach

For concreteness, consider two Normally distributed random variables with means μ_1 and μ_2, and common standard deviation σ. We focus on the following hypotheses:

$$M_e : \mu_1, \mu_2; \sigma,$$
$$M_1 : \mu_1 > \mu_2; \sigma,$$
$$M_2 : \mu_1 \approx \mu_2; \sigma,$$
$$M_3 : \mu_1 = \mu_2; \sigma.$$

In the encompassing model M_e, all parameters are free to vary. Models M_1, M_2, and M_3 are nested in M_e and stipulate particular restrictions on the means; specifically, M_1 features an inequality constraint, M_2 features an "about equality" constraint, and M_3 features an exact equality constraint. We now deal with these in turn.

2.1. Computing Bayes factors for inequality constraints

Suppose we compare two models, an encompassing model M_e and an inequality constrained model M_1. We denote the prior distributions under M_e by $p(\psi, \phi | M_e)$, where ψ is the parameter vector of interest (e.g., μ_1 and μ_2 in the earlier example) and ϕ is the parameter vector of nuisance parameters (e.g., σ in the earlier example).

Then, the prior distribution of the parameters under model M_1 can be obtained from $p(\psi, \phi | M_e)$ by restricting the parameter space of ψ:

$$p(\psi, \phi | M_1) = \frac{p(\psi, \phi | M_e) I_{M_1}(\psi, \phi)}{\iint p(\psi, \phi | M_e) I_{M_1}(\psi, \phi) \mathrm{d}\psi \mathrm{d}\phi}. \tag{1}$$

In Eq. (1), $I_{M_1}(\psi, \phi)$ is the indicator function of model M_1. This means that $I_{M_1}(\psi, \phi) = 1$ if the parameter values are in accordance with the constraints imposed by model M_1, and $I_{M_1}(\psi, \phi) = 0$, otherwise. Note that this specification of priors is only valid under the assumption that the nuisance parameters in M_e and M_1 fulfill exactly the same role (for a debate see Consonni and Veronese (2008) and Del Negro and Schorfheide (2008)).

Under the above specification of priors, Klugkist and Hoijtink (2007) showed that the Bayes factor BF_{1e} can be easily obtained by drawing values from the posterior and prior distribution for M_e:

$$BF_{1e} = \frac{\frac{1}{m} \sum_{i=1}^{m} I_{M_1}(\psi^{(i)}, \phi^{(i)} | D, M_e)}{\frac{1}{n} \sum_{j=1}^{n} I_{M_1}(\psi^{(j)}, \phi^{(j)} | M_e)}, \tag{2}$$

where m represents the total number of MCMC samples for the posterior of ψ, and n represents the total number of MCMC samples for the prior of ψ. The numerator represents the proportion of M_e's *posterior* samples for ψ that obey the constraint imposed by M_1, and the denominator represents the proportion of M_e's *prior* samples for ψ that obey the constraint imposed by M_1.

To illustrate, consider again our initial example in which $M_e : \mu_1, \mu_2; \sigma$ and $M_1 : \mu_1 > \mu_2; \sigma$. Fig. 1(a) shows the joint parameter space for μ_1 and μ_2; for illustrative purposes, we assume that the joint prior is uniform across the parameter space. In Fig. 1(a), half of the prior samples are in accordance with the constraints imposed by M_1. Fig. 1(a) also shows three possible encompassing posterior distributions: A, B, and C. In case A, half of the posterior samples are in accordance with the constraint, and this yields $BF_{1e} = 1$. In case B, very few samples are in accordance with the constraint, and this yields a Bayes factor BF_{1e} that is close to zero (i.e., very large support against M_1). In case C, almost all samples are in accordance with the constraint, and this yields a Bayes factor BF_{1e} that is close to 2.

2.2. Bayes factors for about equality constraints

In the EP approach, the Bayes factor for about equality constraints can be calculated in the same manner as for inequality constraints. To illustrate, consider our example in which $M_e : \mu_1, \mu_2; \sigma$ and $M_2 : \mu_1 \approx \mu_2; \sigma$. Fig. 1(b) shows as a gray area the proportion of prior samples that are in accordance with the constraints imposed by M_2, which in this case equals about .20. Note that $\mu_1 \approx \mu_2$ means $|\mu_1 - \mu_2| < \varepsilon$. The choice for ε defines the size of the parameter space that is allowed by the constraint.

Now consider the three possible encompassing posterior distributions shown in Fig. 1(b). In case A, about 80% of the posterior samples are in accordance with the constraint, and this yields a Bayes factor $BF_{2e} = .8/.2 = 4$. In case B and C, slightly less than half of the samples, about 40%, are in accordance with the constraint, and this yields a Bayes factor $BF_{2e} = .4/.2 = 2$.
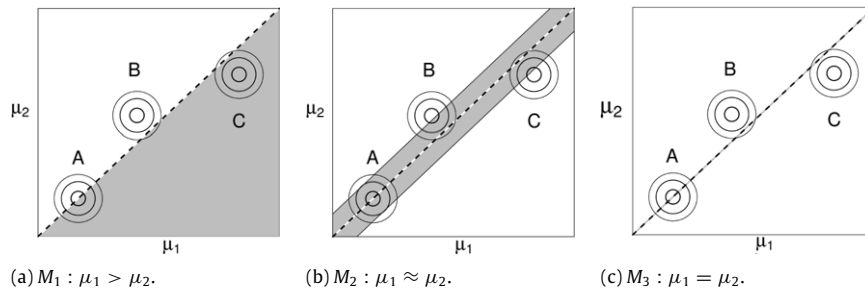
(a) $M_1 : \mu_1 > \mu_2$.   (b) $M_2 : \mu_1 \approx \mu_2$.   (c) $M_3 : \mu_1 = \mu_2$.

Fig. 1. The encompassing prior approach for inequality, about equality, and exact equality constraints. For illustrative purposes, we assume that the encompassing prior is uniform over the parameter space. The gray area represents the part of the encompassing parameter space that is in accordance with the constraints imposed by the nested model. The circles A, B and C represent three different encompassing posterior distributions. Note that the lower and upper bound for μ_1 and μ_2 are the same.

As before, the Bayes factors are calculated with relative ease—all that is required are prior and posterior samples from the encompassing model M_e.

2.3. Bayes factors for exact equality constraints

In some situations, any difference between μ_1 and μ_2 is deemed relevant, and this requires a test for exact equality. For instance, one may wish to test whether a chemical compound adds to the effectiveness of a particular medicine. In such experimental studies, an exact null effect is a priori plausible. However, it may appear that the EP approach does not extend to exact equality constraints in a straightforward fashion.

To illustrate, consider our example in which $M_e : \mu_1, \mu_2; \sigma$ and now $M_3 : \mu_1 = \mu_2; \sigma$. Fig. 1(c) shows that the only values allowed by the constrained model M_3 are those that fall exactly on the diagonal. As μ_1 and μ_2 are continuous variables, the proportion of prior and posterior samples that obey this constraint is zero. Therefore, the EP Bayes factor is 0/0, which has led several researchers to conclude that the EP Bayes factor is not defined for exact equality constraints (Rossell et al., 2008, pp. 111–112; Myung et al., 2008, p. 317; Klugkist, 2008, p. 71). The next two sections investigate in what sense the EP Bayes factor can be defined for exact equality constraints, and its relation to the Savage–Dickey density ratio. Difficulties that arise because of the Borel–Kolmogorov paradox are discussed in the subsequent sections.

2.3.1. Bayes factors for exact equality constraints: an iterative method

In order to estimate the EP Bayes factor for exact equality constrained models, Laudy (2006, p. 115) and Klugkist (2008) proposed an iterative procedure. In the context of a test between $M_e : \mu_1, \mu_2; \sigma$ and $M_3 : \mu_1 = \mu_2; \sigma$, the procedure comprises the following steps:

Step 1: Choose a small value ε_1 and define $M_{3.1} : |\mu_1 - \mu_2| < \varepsilon_1$;
Step 2: Compute the Bayes factor $BF_{(3.1)e}$ using Eq. (2);
Step 3: Define $\varepsilon_2 < \varepsilon_1$ and $M_{3.2} : |\mu_1 - \mu_2| < \varepsilon_2$;
Step 4: Sample from the constrained ($|\mu_1 - \mu_2| < \varepsilon_1$) prior and posterior and compute the Bayes factor $BF_{(3.2)(3.1)}$.

Repeat steps 3 and 4, with each $\varepsilon_{n+1} < \varepsilon_n$, until $BF_{n+1,n} \approx 1$. Then the required Bayes factor BF_{3e} can be calculated by multiplication:

$$BF_{3e} = BF_{(3.1)e} \times BF_{(3.2)(3.1)} \times \cdots \times BF_{n(n-1)}. \tag{3}$$

In the limit (i.e., when $\varepsilon_n \to 0$), this method yields the Bayes factor for the exact equality model M_3 versus the encompassing model M_e. Although this iterative method solves the problem of having no samples that obey an exact equality constraint, the method is only approximate and potentially time consuming.

2.3.2. Bayes factors for exact equality constraints: a one-step method—equivalence to the Savage–Dickey density ratio

The iterative procedure turns out to be identical to the Savage–Dickey density ratio method, a one-step method that is both principled and fast. In order to understand this intuitively, Fig. 2 shows a fictitious prior and posterior distribution for $\mu_1 - \mu_2$, obtained under the encompassing model M_e. The surface of the dashed areas equals the proportion of the prior and posterior distribution that is consistent with the constraint $|\mu_1 - \mu_2| < \varepsilon$. In the EP approach, the Bayes factor is obtained by integrating the posterior and prior distribution over the area defined by the constraint. However, it is clear that as $\varepsilon \to 0$, the area of both regions equals 0.

The Bayes factor is given by the ratio of the two integrals. Hence, the Bayes factor for the equality constraint in the EP approach is the limit

$$BF_{3e} = \lim_{\epsilon \to 0} \frac{\int_{-\epsilon/2}^{\epsilon/2} p(\psi_0 + \psi \mid D, M_e)\mathrm{d}\psi}{\int_{-\epsilon/2}^{\epsilon/2} p(\psi_0 + \psi \mid M_e)\mathrm{d}\psi}.$$
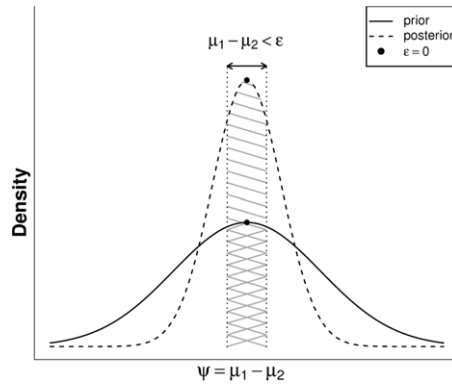
Fig. 2. The encompassing prior approach for exact equality constraints is the Savage–Dickey density ratio. The top dot represents the value of the posterior distribution at $\mu_1 = \mu_2$ and the bottom dot represents the value of the prior distribution at $\mu_1 = \mu_2$. The ratio of the heights of both densities equals the Bayes factor. Note that the posterior of ψ does not have to be centered around zero.

Here we generically formulated the hypothesis in terms of the parameter ψ. In the example hypothesis $H_0 : \mu_1 = \mu_2$ this corresponds to defining $\psi = \mu_1 - \mu_2$ and $\psi_0 = 0$. We also marginalized over any nuisance parameters not of interest; σ in our example, i.e., $p(\mu_1, \mu_2 \mid D, M_e) = \int_{-\infty}^{\infty} p(\mu_1, \mu_2, \sigma \mid D, M_e)d\sigma$. Then, in the example, $\psi = \mu_1 - \mu_2$ has marginal posterior density $p(\psi|D, M_e) = \int p(\mu_1, \mu_1 - \psi |D, M_e)d\mu_1$ (see e.g., Miller and Miller, 2004, pp. 246). These integrals can be evaluated analytically, with quadratures, or can be approximated using MCMC sampling (Gamerman and Lopes, 2006). To calculate the Bayes factor, only the marginal posterior density of interest needs to be considered.

Clearly the limit above approaches the form 0/0 and so l'Hôpital's 0/0 rule can be employed to obtain

$$BF_{3e} = \lim_{\epsilon \to 0} \frac{p(\psi_0 + \epsilon/2 \mid D, M_e)/2 + p(\psi_0 - \epsilon/2 \mid D, M_e)/2}{p(\psi_0 + \epsilon/2 \mid M_e)/2 + p(\psi_0 - \epsilon/2 \mid M_e)/2} = \frac{p(\psi_0 \mid D, M_e)}{p(\psi_0 \mid M_e)}, \tag{4}$$

where ψ_0 represents the point of exact equality specified by the constrained model; in our example, $M_3 : \psi_0$ means $\mu_1 - \mu_2 = 0$.

Eq. (4) shows that the Bayes factor BF_{3e} simplifies to the ratio of the height of the marginal posterior and the height of the marginal prior at the point of interest, if the limiting processes in the numerator and the denominator are chosen to be equal. This result is known as the Savage–Dickey density ratio (Dickey and Lientz, 1970; O'Hagan and Forster, 2004; Dickey, 1971; Verdinelli and Wasserman, 1995). For the example shown in Fig. 2, the Bayes factor in favor of the exact equality model, BF_{3e}, is approximately 2.

For completeness, we now sketch the proof that the Savage–Dickey density ratio equals the Bayes factor (cf. O'Hagan and Forster, 2004). As before, let ψ be the parameter of interest and ϕ the nuisance parameter; let M_e be the encompassing model, a restricted version of which is defined as $M_3 : \psi = \psi_0$. The Savage–Dickey density ratio is equal to the Bayes factor if the prior of the nuisance parameter in the restricted model M_3 is defined by conditioning, that is, if $p(\phi|M_3) = p(\phi|\psi = \psi_0, M_e)$ (cf. Eq. (1)).

The foregoing allows us to rewrite the marginal likelihood for M_3:

$$p(D|M_3) = \int p(D|\phi, M_3)p(\phi|M_3)d\phi,$$
$$= \int p(D|\phi, \psi = \psi_0, M_e)p(\phi|\psi = \psi_0, M_e)d\phi,$$
$$= p(D|\psi = \psi_0, M_e). \tag{5}$$

We now apply Bayes' rule to the end result of Eq. (5) and obtain

$$p(D|M_3) = \frac{p(\psi = \psi_0|D, M_e)p(D|M_e)}{p(\psi = \psi_0|M_e)}. \tag{6}$$

Dividing both sides of Eq. (6) by $p(D|M_e)$ results in

$$BF_{3e} = \frac{p(D|M_3)}{p(D|M_e)} = \frac{p(\psi = \psi_0|D, M_e)}{p(\psi = \psi_0|M_e)}, \tag{7}$$

which shows that the Bayes factor equals the ratio of the posterior and prior ordinate under M_e at the point of interest (i.e., $\psi = \psi_0$).
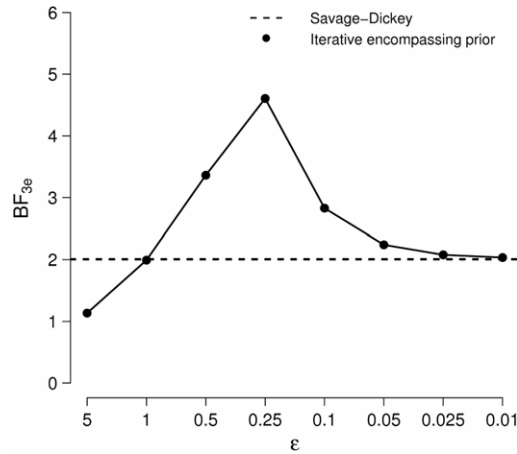
Fig. 3. A comparison of the Savage–Dickey density ratio and the iterative encompassing prior approach for simulated data. The Bayes factor favoring the null hypothesis, BF_{3e}, is 2.011. The dashed line shows the Savage–Dickey Bayes factor. The dots represent the iterative Bayes factor calculated by systematically decreasing the interval surrounding the exact equality of interest (Eq. (3)).

2.3.3. An example comparing the iterative EP approach to the Savage–Dickey method

In order to illustrate how the results from the two methods converge, we randomly drew 100 samples from a Normal distribution with mean 0.2 and variance 1, and found a corresponding one-sample t-statistic of 1.945. We then used a Bayesian t-test with a Cauchy(0,1) prior on effect size $\delta = \mu/\sigma$ and a folded Cauchy(0,1) on σ (for details see Rouder et al., 2009) to compute the Bayes factor in favor of $H_0 : \delta = 0$ relative to $H_1 : \delta \sim$ Cauchy(0,1), which yielded $BF_{3e} = 2.011$.

Fig. 3 compares the behavior of the iterative encompassing prior approach to that of the Savage–Dickey density ratio. The dashed horizontal line shows the result from the Savage–Dickey implementation of the Bayesian t-test (Wetzels et al., 2009). The dots show the result from the iterative encompassing prior approach (Eq. (3)), as a function of the size of the smallest interval ε. When $\varepsilon = 0.01$, the iterative EP Bayes factor has converged to the correct Bayes factor.

Note that the iterative EP approach involves the product of multiple Bayes factors (cf. Eq. (3)). In contrast, the Savage–Dickey procedure involves only one Bayes factor. Because the computation of each Bayes factor requires many MCMC samples, the computational demands are likely to be much higher in the iterative EP approach than in the Savage–Dickey approach.

3. The Borel–Kolmogorov paradox

The main drawback of the EP approach to exact equalities is that it is subject to the Borel–Kolmogorov paradox (DeGroot and Schervish, 2002; Jaynes, 2003; Lindley, 1997; Proschan and Presnell, 1998; Rao, 1988; Singpurwalla and Swift, 2001). This paradox arises when one conditions on events of probability zero. In the case of exact equality constraints, priors for the constrained model are constructed by conditioning on a null-set, and this gives rise to the Borel–Kolmogorov paradox.

3.1. The Borel–Kolmogorov paradox: an example

Consider the following situation, inspired by an example from Lindley (1997). Suppose that a point P is described by its Cartesian coordinates X and Y. Furthermore, suppose that $0 \le X \le 1$ and $0 \le Y \le 1$, and that P has a uniform distribution on the unit square. Suppose you are told that P lies on the diagonal through the origin, event B. What is your probability that X, associated with that P, and hence also Y, is less then $1/2$ (i.e., event A)?

The paradox lies in the fact that the answer to this question depends on how we parameterize the diagonal. We examine two situations: $Z_1 = X - Y = 0$ (see Fig. 4(a)) and $Z_2 = X/Y = 1$ (see Fig. 4(b)). Note that because X and Y are continuous, the probability that $X = Y$ is zero. Because conditioning on an event with probability zero is problematic, we consider values of X and Y that lie in the proximity of the line $X = Y$.

From the geometry of the problem, the associated probability in the first case is

$$P(Y - \epsilon \le X \le Y + \epsilon) = 2(1/2 - 1/2(1 - \epsilon)^2) = (2 - \epsilon)\epsilon,$$

while the associated probability for the second case is

$$P(Y - \epsilon Y \le X \le Y - \epsilon Y) = 2(1/2 - 1/2 \cdot 1 \cdot (1 - \epsilon)) = \epsilon.$$

Now consider the probabilities that (X, Y) lies in the left lower quadrant of the square (i.e., $X, Y \le 1/2$) and either that $|X - Y| < \epsilon$ and or that $|X/Y - 1| < \epsilon$. Again from geometry, the probability of the first case is

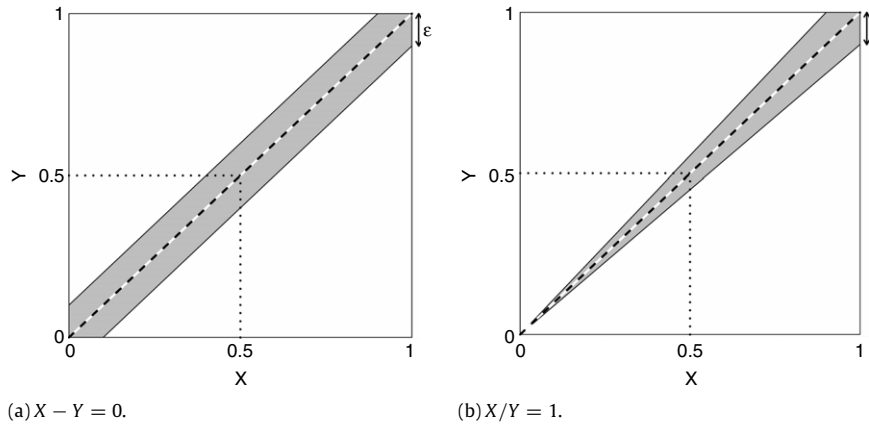$$P(|X - Y| < \epsilon \cap X, Y \le 1/2) = (1 - \epsilon)\epsilon,$$

(a) $X - Y = 0$.   (b) $X/Y = 1$.

Fig. 4. Example of the Borel–Kolmogorov paradox. The shaded areas indicate the acceptable values of X and Y for the two parameterizations. In panel (a), all values of X are equally likely while in panel (b), the wedge shaped area implies that values close to 1 are more likely than those close to 0.

and for the second case

$$P(|X/Y - 1| < \epsilon \cap X, Y \le 1/2) = \frac{1}{4}\epsilon.$$

Hence, the corresponding conditional probabilities of the events that (X, Y) lies in the left lower quadrant of the square, given that either the event $|X - Y| \le \epsilon$ or that the event $|X/Y - 1| \le \epsilon$ occurred, are respectively

$$P(X, Y \le 1/2 \mid |X - Y| < \epsilon) = \frac{1 - \epsilon}{2 - \epsilon},$$

and

$$P(X, Y \le 1/2 \mid |X/Y - 1| < \epsilon) = \frac{1}{4}.$$

If we now take to the limit $\epsilon \to 0$, both the events $|X - Y| \le \epsilon$ and $|X/Y - 1| \le \epsilon$ coincide with the lower left half of the diagonal of the square. However, although the first probability coincides with our intuition that $P(X, Y \le 1/2 \mid X = Y) = 1/2$, the second has it that the probability for this seemingly equal event should be 1/4!

This example shows that the probability of an event conditioned on a limiting event of zero probability depends on the way in which the limiting event was generated, that is, on the parameterization that was chosen to generate the zero probability event. In effect, conditional probability is not invariant under coordinate transformations of the conditioning variable. This paradox is resolved if one accepts that conditional probability cannot be unambiguously defined with respect to events of zero probability without specifying the limiting process from which it should result (Jaynes, 2003). It is on random variables, not on singular events, that conditioning is unambiguous (see Kolmogorov (1956), Billingsley (2008) and Wolpert (1995)).

3.2. Implications for the limiting encompassing prior and the Savage–Dickey density ratio

The foregoing implies that in the EP approach to exact equalities, the resulting Bayes factor may depend on the choice of the parameterization, a feature that is clearly undesirable (Dawid and Lauritzen, 2001; Schweder and Hjort, 1996; Wolpert, 1995). Note that the Borel–Kolmogorov paradox does not occur in the case of inequality constraints, where one conditions on an interval, rather than on a single point.

Eq. (4) shows that the EP Bayes factor for exact equality constraints is equal to the Savage–Dickey ratio. However, the rectangular regions of integration and the use of the same limiting processes in both the numerator and the denominator are arbitrary choices. Different choices of the limiting process can lead to different Bayes factors, as shown next. To this end, let $\gamma_i(\epsilon) \ge 0$ and $\delta_i(\epsilon) > 0$, differentiable in a neighborhood $(0, \epsilon)$, such that $\lim_{\epsilon \to 0} \gamma_i(\epsilon) = \lim_{\epsilon \to 0} \delta_i(\epsilon) = 0$ and $\gamma_i'(0) + \delta_i'(0) \ne 0$, for $i = 1, 2$. Here prime $'$ indicates derivative. Then, without loss of generality, these functions can be chosen to suit any form of the limiting process in the EP process

$$BF_{3e} = \lim_{\epsilon \to 0} \frac{\int_{-\gamma_1(\epsilon)}^{\delta_1(\epsilon)} p(\psi_0 + \psi \mid D, M_e) \mathrm{d}\psi}{\int_{-\gamma_2(\epsilon)}^{\delta_2(\epsilon)} p(\psi_0 + \psi \mid M_e) \mathrm{d}\psi}.$$

Intuitively this would seem to go to the same limit as earlier, but in fact it does not, as l'Hôpital's rule shows:

$$BF_{3e} = \lim_{\epsilon \to 0} \frac{p(\psi_0 + \delta_1(\epsilon) \mid D, M_e)\delta_1'(\epsilon) + p(\psi_0 - \gamma_1(\epsilon) \mid D, M_e)\gamma_1'(\epsilon)}{p(\psi_0 + \delta_2(\epsilon) \mid M_e)\delta_2'(\epsilon) + p(\psi_0 - \gamma_2(\epsilon) \mid M_e)\gamma_2'(\epsilon)}$$

$$= \frac{p(\psi_0 \mid D, M_e)\,[\delta_1'(0) + \gamma_1'(0)]}{p(\psi_0 \mid M_e)\,[\delta_2'(0) + \gamma_2'(0)]}. \tag{8}$$

This is the above Savage–Dickey ratio if and only if $\delta_1'(0) + \gamma_1'(0) = \delta_2'(0) + \gamma_2'(0)$. As both $\delta_1'(0) + \gamma_1'(0)$ and $\delta_2'(0) + \gamma_2'(0)$ measure the rate at which the numerator and denominator approach zero, the limit of the EP approach equals the Savage–Dickey ratio if and only if both numerator and denominator approach 0 at the same rate. If the rate at which the numerator and the denominator approach zero is not the same, any desired value of the Bayes factor can be obtained.

In light of the Borel–Kolmogorov paradox, it is important to understand when the Savage–Dickey ratio method is invariant under smooth transformations of the chosen parameterization, especially when nuisance parameters are present. To this end, suppose the chosen set of (absolute continuous) parameters is θ with prior $p(\theta|M_e)$ and posterior $p(\theta|D, M_e)$. Let g be a differentiable invertible transform (a diffeomorphism) with inverse h so that

$$\chi = g(\theta) \quad \text{and} \quad h(\chi) = \theta.$$

The implied prior is denoted $\widetilde{p}(\chi|M_e)$ and the implied posterior is denoted $\widetilde{p}(\chi|D, M_e)$. In general, the parameter vector can be partitioned as $\theta = (\psi, \phi)$, where ϕ contains nuisance parameters that are not involved in the evaluation of the null hypothesis. We are interested in evaluating the evidence for the simple hypothesis

$$M_3 : \psi = \psi_0,$$

which, in terms of $\chi = (\nu, \xi)$ can often be cast equivalently as $M_3 : \nu = \nu_0$. We wish to know under what circumstances the Savage–Dickey ratios are equal. That is, we want to determine conditions on g under which the desired equality

$$BF_{3e} = \frac{p(\psi_0|D, M_e)}{p(\psi_0|M_e)} = \frac{\widetilde{p}(\nu_0|D, M_e)}{\widetilde{p}(\nu_0|M_e)}, \tag{9}$$

is true. It turns out that this equality holds, as long as the transformation g does not depend on the data D, and as long as the parameters on which M_3 imposes a simple hypothesis transform independently of the nuisance parameters. This follows from the following considerations.

By the "change of variables" rule,

$$\widetilde{p}(\chi|D, M_e) = p(h(\chi)|D, M_e)|h'(\chi)|_+,$$
$$\widetilde{p}(\chi|M_e) = p(h(\chi)|M_e)|h'(\chi)|_+,$$

where $|h'(\chi)|_+$ denotes the absolute value of the determinant of the Jacobian matrix $h'(\chi) = \partial_\chi h(\chi)$ of the transformation h. Partition $h(\chi)$ as $\theta = h(\chi) = (\psi, \phi) = (\psi(\nu, \xi), \phi(\nu, \xi))$. In terms of hypothesis and nuisance parameters these can be expressed as

$$\widetilde{p}(\nu, \xi|M_e) = p(\psi(\nu, \xi), \phi(\nu, \xi)|M_e)\,|\phi_\xi(\nu, \xi)|_+|\psi_\nu(\nu, \xi) - \phi_\nu(\nu, \xi)\phi_\xi(\nu, \xi)^{-1}\psi_\xi(\nu, \xi)|_+, \tag{10}$$

and similarly for $\widetilde{p}(\nu, \xi|D, M_e)$. Here $\phi_\xi(\nu, \xi)$ denotes the matrix of partial derivatives of ϕ with respect to ξ, ψ_ν of ψ with respect to ν, etc.

The implicit function theorem ensures the existence of a function $\nu(\psi, \xi)$, such that $\psi_0 = \psi(\nu(\psi_0, \xi), \xi)$, for all ξ. Then, by the chain rule,

$$\int p(\psi_0, \phi|M_e)\mathrm{d}\phi = \int p(\psi(\nu(\psi_0, \xi), \xi), \phi(\nu(\psi_0, \xi), \xi)|M_e)\,|\partial_\xi \phi(\nu(\psi_0, \xi), \xi)|_+\,\mathrm{d}\xi.$$

The Jacobian in the last integral can be expressed as

$$|\partial_\xi \phi(\nu(\psi_0, \xi), \xi)|_+ = |\phi_\nu(\nu(\psi_0, \xi), \xi)\nu_\xi(\psi_0, \xi) + \phi_\xi(\nu(\psi_0, \xi), \xi)|_+.$$

Therefore, if $\nu_\xi(\psi, \xi) \equiv 0$, implying that $\nu(\psi, \xi) = \nu(\psi)$ does not depend on ξ, then

$$\int p(\psi_0, \phi|M_e)\mathrm{d}\phi = \int p(\psi(\nu(\psi_0), \xi|M_e), \phi(\nu(\psi_0), \xi))\,|\phi_\xi(\nu(\psi_0), \xi)|_+\,\mathrm{d}\xi,$$

which, by Eq. (10) can be expressed as

$$\int p(\psi_0, \phi|M_e)\mathrm{d}\phi = \int \widetilde{p}(\nu_0, \xi|M_e)\,|\psi_\nu(\nu_0, \xi) - \phi_\nu(\nu_0, \xi)\phi_\xi(\nu_0, \xi)^{-1}\psi_\xi(\nu_0, \xi)|_+^{-1}\,\mathrm{d}\xi,$$

$$= \int \widetilde{p}(\nu_0, \xi|M_e)\,|\psi_\nu(\nu_0)|_+^{-1}\,\mathrm{d}\xi.$$

Here we used the fact that $\psi(\nu(\psi_0), \xi) = \psi(\nu(\psi_0, \xi), \xi) = \psi_0$. We also used the fact that $\nu_0 = \nu(\psi_0)$, which is warranted by the above assumption that $\nu_\xi(\psi, \xi) = 0$. Specifically, because $\partial_\xi \psi(\nu(\psi_0, \xi), \xi) = \psi_\nu(\nu(\psi_0, \xi), \xi)\nu_\xi(\psi_0, \xi) + \psi_\xi(\nu(\psi_0, \xi), \xi) = \partial_\xi \psi_0 = 0$, it follows that $\psi_\xi(\nu(\psi_0, \xi), \xi) \equiv 0$ for all ψ_0. This implies that ψ does not depend on ξ,

and therefore that $\psi_0 = \psi(\nu(\psi_0, \xi), \xi) = \psi(\nu(\psi_0)) = \psi(\nu_0)$. The latter conclusion that $\psi_\xi \equiv 0$ also yields the second equality.

Consequently, the evidence for M_3 is obtained from the Savage–Dickey ratio

$$BF_{3e} = \frac{p(\psi_0|D, M_e)}{p(\psi_0|M_e)} = \frac{\int p(\psi_0, \phi|D, M_e)\mathrm{d}\phi}{\int p(\psi_0, \phi|M_e)\mathrm{d}\phi} = \frac{\int \widetilde{p}(\psi_0, \xi|D, M_e)|\psi_\nu(\nu_0)|_+^{-1}\mathrm{d}\xi}{\int \widetilde{p}(\psi_0, \xi|M_e)|\psi_\nu(\nu_0)|_+^{-1}\mathrm{d}\xi} = \frac{\widetilde{p}(\psi_0|D, M_e)}{\widetilde{p}(\psi_0|M_e)},$$

which is (9).

In sum, computing the Bayes factor for exact equality constraints is a delicate matter. The iterative EP approach and the Savage–Dickey density ratio can lead to different Bayes factors if the limiting process in the iterative EP approach is not carefully chosen (i.e., the numerator and denominator of Eq. (8) should approach 0 at the same rate). Moreover, both methods suffer from the Borel–Kolmogorov paradox. However, the Savage–Dickey density ratio is invariant under smooth transformations of the chosen parameterization, as long as the transformation does not depend on the data, and as long as the parameters transform independently of the nuisance parameters.

4. Concluding remarks

Here we have shown that the Savage–Dickey density ratio method is a special case of the encompassing prior (EP) approach proposed by Hoijtink, Klugkist, and colleagues. The EP approach was developed to account for models with inequality constraints; as it turns out, the approach naturally extends to models with exact equality constraints. Consequently, the EP approach offers a unified, elegant, and simple method to compute Bayes factors in nested models.

The main drawback of the EP/Savage–Dickey method for exact equalities is its susceptibility to the Borel–Kolmogorov paradox. We have shown that the SD-ratio yields the same value under different transformations, as long as the parameters, on which M_1 imposes a simple hypothesis, transform independently of the nuisance parameters. It should be noted that in order to avoid the Borel–Kolmogorov paradox, alternative procedures seek to construct priors not by the usual conditioning, but by method such as marginalization (Kass and Raftery, 1995), Jeffreys conditioning (Dawid and Lauritzen, 2001), reference conditioning (Roverato and Consonni, 2004), Kullback–Leibler projection (Consonni and Veronese, 2008; Dawid and Lauritzen, 2001), and Hausdorff integrals (Kleibergen, 2004).

Unfortunately, these alternative procedures give rise to paradoxes and problems of their own (see Consonni and Veronese (2008), for a review and a comparison). Presently, there does not appear to be a universally agreed-on method for specifying priors in nested models that is clearly superior to the conditioning procedure inherent in the Hoijtink and Klugkist EP approach.

Acknowledgements

References

Billingsley, P., 2008. Probability and Measure. Wiley, New York.

Consonni, G., Veronese, P., 2008. Compatibility of prior specifications across linear models. Statistical Science 23, 332–353.

Dawid, A.P., Lauritzen, S.L., 2001, Compatible prior distributions. In: George, E. (Ed.), Bayesian Methods with Applications to Science, Policy, and Official Statistics. Monographs of Official Statistics, Luxembourg, pp. 109–118.

DeGroot, M., Schervish, M., 2002. Probability and Statistics. Addison-Wesley, Boston.

Del Negro, M., Schorfheide, F., 2008. Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). Journal of Monetary Economics 55 (7), 1191–1208.

Dickey, J.M., 1971. The weighted likelihood ratio, linear hypotheses on normal location parameters. The Annals of Mathematical Statistics 42, 204–223.

Dickey, J.M., Lientz, B.P., 1970. The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. The Annals of Mathematical Statistics 41, 214–226.

Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman & Hall/CRC, Boca Raton, FL.

Hoijtink, H., Klugkist, I., Boelen, P., 2008. Bayesian Evaluation of Informative Hypotheses. Springer, New York.

Jaynes, E.T., 2003. Probability Theory: The Logic of Science. Cambridge University Press, Cambridge, UK.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 377–395.

Kleibergen, F., 2004. Invariant Bayesian inference in regression models that is robust against the Jeffreys–Lindley's paradox. Journal of Econometrics 123, 227–258.

Klugkist, I., 2008. Encompassing prior based model selection for inequality constrained analysis of variance. In: Hoijtink, H., Klugkist, I., Boelen, P. (Eds.), Bayesian Evaluation of Informative Hypotheses. Springer, New York, pp. 53–83.

Klugkist, I., Hoijtink, H., 2007. The Bayes factor for inequality and about equality constrained models. Computational Statistics and Data Analysis 51, 6367–6379.

Klugkist, I., Kato, B., Hoijtink, H., 2005a. Bayesian model selection using encompassing priors. Statistica Neerlandica 59, 57–69.

Klugkist, I., Laudy, O., Hoijtink, H., 2005b. Inequality constrained analysis of variance: a Bayesian approach. Psychological Methods 10, 477–493.

Kolmogorov, A., 1956. Foundations of the Theory of Probability. Chelsea Publishing Company, New York.

Laudy, O., 2006. Bayesian inequality constrained models for categorical data. Ph.D. Thesis, Utrecht University.

Lewis, S.M., Raftery, A.E., 1997. Estimating Bayes factors via posterior simulation with the Laplace–Metropolis estimator. Journal of the American Statistical Association 92, 648–655.

Lindley, D., 1997. Some comments on Bayes factors. Journal of Statistical Planning and Inference 61, 181–189.

Miller, I., Miller, M., 2004. John E. Freund's Mathematical Statistics with Applications. Prentice Hall, New Jersey.

Myung, J.I., Karabatsos, G., Iverson, G.J., 2008. A statistician's view on Bayesian evaluation of informative hypotheses. In: Hoijtink, H., Klugkist, I., Boelen, P. (Eds.), Bayesian Evaluation of Informative Hypotheses. Springer, New York, pp. 309–327.

O'Hagan, A., Forster, J., 2004. Kendall's advanced theory of statistics vol. 2B: Bayesian inference, 2nd ed. Arnold, London.

Proschan, M.A., Presnell, B., 1998. Expect the unexpected from conditional expectation. The American Statistician 52, 248–252.

Rao, M., 1988. Paradoxes in conditional probability. Journal of Multivariate Analysis 27, 434–446.

Rossell, D., Baladandayuthapani, V., Johnson, V.E., 2008. Bayes factors based on test statistics under order restrictions. In: Hoijtink, H., Klugkist, I., Boelen, P.A. (Eds.), Bayesian Evaluation of Informative Hypotheses. Springer, New York, pp. 111–129.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t-tests for accepting and rejecting the null hypothesis. Psychonomic Bulletin & Review 16, 225–237.

Roverato, A., Consonni, G., 2004. Compatible prior distributions for DAG models. Journal of the Royal Statistical Society B 66, 47–61.

Schweder, T., Hjort, N.L., 1996. Bayesian synthesis or likelihood synthesis–what does Borel's paradox say? Report International Whaling Commission 46, pp. 475–479.

Singpurwalla, N.D., Swift, A., 2001. Network reliability and Borel's paradox. The American Statistician 55, 213–218.

Verdinelli, I., Wasserman, L., 1995. Computing Bayes factors using a generalization of the Savage–Dickey density ratio. Journal of the American Statistical Association 90, 614–618.

Wetzels, R., Raaijmakers, J., Jakab, E., Wagenmakers, E.-J., 2009. How to quantify support for and against the null hypothesis: a flexible WinBUGS implementation of a default Bayesian t-test. Psychonomic Bulletin & Review 16, 752–760.

Wolpert, R., 1995. Comment–inference from a deterministic population dynamics model for bowhead whales. Journal of the American Statistical Association 90, 426–427.