Default Bayesian analysis of the Behrens–Fisher problem

Article *in* Journal of Statistical Planning and Inference · November 1999

Impact Factor: 0.68 · DOI: 10.1016/S0378-3758(99)00070-1

Walter Racugno
Università degli studi di Cagliari

Journal of Statistical Planning and Inference 81 (1999) 323–333

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Default Bayesian analysis of the Behrens–Fisher problem

Elías Moreno[a,*], Francesco Bertolino[b], Walter Racugno[b]

[a] *Department of Statistics, Faculty of Science, University of Granada, Fuentenueva s/n, 18071, Granada, Spain*

[b] *University of Cagliari, Cagliari, Italy*

Received 5 March 1998; accepted 8 March 1999

Abstract

In the Bayesian approach, the Behrens–Fisher problem has been posed as one of estimation for the difference of two means. No Bayesian solution to the Behrens–Fisher testing problem has yet been given due, perhaps, to the fact that the conventional priors used are improper. While default Bayesian analysis can be carried out for estimation purposes, it poses difficulties for testing problems. This paper generates sensible *intrinsic* and *fractional* prior distributions for the Behrens–Fisher testing problem from the improper priors commonly used for estimation. It allows us to compute the Bayes factor to compare the null and the alternative hypotheses. This default procedure of model selection is compared with a frequentist test and the Bayesian information criterion. We find discrepancy in the sense that frequentist and Bayesian information criterion reject the null hypothesis for data, that the Bayes factor for intrinsic or fractional priors do not.

MSC: 62C10; 62A15

Keywords: Bayes factors; Behrens–Fisher problem; Fractional priors; Intrinsic priors

1. Introduction

Let $\mathrm{N}(x|\mu_1,\sigma_1^2)$, $\mathrm{N}(y|\mu_2,\sigma_2^2)$ be two normal distributions where the means μ_1,μ_2 and variances σ_1^2, σ_2^2 are unknown. Samples of sizes n_1 and n_2, $\boldsymbol{x}=(x_1,x_2,\ldots,x_{n_1})$, $\boldsymbol{y}=(y_1,y_2,\ldots,y_{n_2})$, respectively, are taken, and the sample means and variances are $\bar{x}$ and s_x^2 and $\bar{y}$ and s_y^2.

The Behrens–Fisher problem consists in testing H_0: $\mu_1=\mu_2$ against one of the alternatives $\mu_1<\mu_2$, $\mu_1>\mu_2$ or $\mu_1\neq\mu_2$. In this paper we focus on the alternative H_1: $\mu_1\neq\mu_2$.

Under the frequentist point of view this problem poses the difficulty that the normal-theory linear model cannot be applied because of the presence of the two unrelated

* Corresponding author. Tel.: +34-58-243269; fax: +34-58-243267.

E-mail address: emoreno@goliat.ugr.es (E. Moreno)

0378-3758/99/$ - see front matter
PII: S0378-3758(99)00070-1

variances σ_1^2, σ_2^2. Some approximations to the solution have been given by Fisher (1936), Wald (1955), and Welch (1947), and their relative merits are discussed by Mehta and Srinivasan (1970) and Pfanzagl (1974).

Under the Bayesian viewpoint the Behrens–Fisher problem can be addressed as one of Bayesian model selection, for which we propose a fully default analysis. By doing so, we simply follow the recommendation of Jeffreys (1961) and distinguish between problems of estimation and problems of testing.

Consider the sampling models under the null and the alternative hypotheses, say

$$f_1(z|\theta_1) = \mathrm{N}(x|\mu,\tau_1^2)\mathrm{N}(y|\mu,\tau_2^2),$$

$$f_2(z|\theta_2) = \mathrm{N}(x|\mu_1,\sigma_1^2)\mathrm{N}(y|\mu_2,\sigma_2^2),$$

where $z=(x,y)$, $\theta_1=(\mu,\tau_1,\tau_2)$ and $\theta_2=(\mu_1,\mu_2,\sigma_1,\sigma_2)$. Paramaters θ_1 and θ_2 lie in different spaces, so that to avoid confusion in the subsequent development we are denoting them with different symbols. For the prior $\pi_i(\theta_i)$, the models to be compared are M_1: $\{f_1(z|\theta_1),\ \pi_1(\theta_1)\}$ and M_2: $\{f_2(z|\theta_2),\ \pi_2(\theta_2)\}$.

The main tool for model comparison is the Bayes factor that can be used to compute the posterior odds of the models. Indeed, given a suitable prior P on the set $\{M_1,M_2\}$ and the sample $\boldsymbol{z}=(\boldsymbol{x},\boldsymbol{y})$, the posterior odds is given by

$$\frac{P(M_2|\boldsymbol{z})}{P(M_1|\boldsymbol{z})} = B_{21}(\boldsymbol{z})\frac{P(M_2)}{P(M_1)},$$

where

$$B_{21}(\boldsymbol{z}) = \frac{\int_{\Theta_2} f_2(\boldsymbol{z}|\theta_2)\pi_2(\theta_2)\,\mathrm{d}\theta_2}{\int_{\Theta_1} f_1(\boldsymbol{z}|\theta_1)\pi_1(\theta_1)\,\mathrm{d}\theta_1}, \tag{1}$$

is the Bayes factor of M_2 against M_1. It contains all the information the data has to provide.

The usual priors of μ_1, μ_2, $\log\sigma_1$, $\log\sigma_2$ are independent uniform distributions over $(-\infty,\infty)$, that are improper. They are appropriate for estimation of the μ's and, in fact, the posterior distribution of $\mu_1-\mu_2$ is the so-called Behrens–Fisher distribution (Lindley, 1970). Note that the priors for θ_1 and θ_2,

$$\pi_1^{\mathrm{N}}(\theta_1) = \frac{c_1}{\tau_1\tau_2} 1_{R\times R^+\times R^+}(\mu,\tau_1,\tau_2),$$

$$\pi_2^{\mathrm{N}}(\theta_2) = \frac{c_2}{\sigma_1\sigma_2} 1_{R\times R\times R^+\times R^+}(\mu_1,\mu_2,\sigma_1,\sigma_2),$$

where 1_C denotes the indicator function of set C, depend on arbitrary positive constants c_1, c_2, that makes them unsuitable for model comparisons. In fact, the corresponding Bayes factor depends on the ratio c_2/c_1 of two unspecified constants. To avoid this difficulty the intrinsic Bayes factor (IBF) and the fractional Bayes factor (FBF) have been proposed by Berger and Pericchi (1996) and by O'Hagan (1995), respectively.

It is important to stress that for nested models the IBF and the FBF give consistent posterior inferences on the models (O'Hagan, 1997). Based on an asymptotic argument, these methods also provide sensible prior distributions for model comparison

(Berger and Pericchi, 1996; Moreno, 1997). The goal of this paper is to construct such priors, *intrinsic* and *fractional*, and to compute the Bayes factor in the Behrens–Fisher problem.

A thoughtful discussion on motivations for the use of intrinsic priors is given in Berger and Pericchi (1997). The justification is based on the following arguments. This approach is theoretically satisfactory as long as we are computing *actual* Bayes factors that do not depend on the arbitrary ratio c_2/c_1. Therefore, the method is a fully default Bayesian procedure for model comparison. Bayes factors with intrinsic or fractional priors depend on the data through the sufficient statistics. Furthermore, the equality $B_{21}(z)=1/B_{21}(z)$ holds and consequently the coherent condition $P(M_1|z)=1-P(M_2|z)$ is satisfied.

The paper is organized as follows. In Section 2 the intrinsic and fractional priors are derived. In Section 3 the Bayes factors are computed and their behavior is showed on specific examples. Comparisons with the Bayesian information criterion by Schwarz (1978) and the classical test by Welch (1947) are also carried out. Section 4 gives some concluding remarks.

2. The intrinsic and fractional priors

For clarity of exposition, before deriving the intrinsic and the fractional prior distributions, we give a brief summary of the intrinsic and fractional Bayes factors.

2.1. Intrinsic and fractional Bayes factors

The basic idea of a training sample is to split the sample z into two parts, $z=(z(\ell),z(n-\ell))$. Part $z(\ell)$, the training sample, is devoted to converting $\pi_i^{\mathrm{N}}(\theta_i)$ to a proper distribution

$$\pi_i(\theta_i|z(\ell))=\frac{f_i(z(\ell)|\theta_i)\pi_i^{\mathrm{N}}(\theta_i)}{m_i^{\mathrm{N}}(z(\ell))},$$

where $m_i^{\mathrm{N}}(z(\ell))=\int_{\Theta_i} f_i(z(\ell)|\theta_i)\pi_i^{\mathrm{N}}(\theta_i)\,\mathrm{d}\theta_i$, $i=1,2$. Note that $\pi_i(\theta_i|z(\ell))$ is well defined only if $z(\ell)$ is such that $0<m_i^{\mathrm{N}}(z(\ell))<\infty$, $i=1,2$. If there is no subsample of $z(\ell)$ for which the marginal is finite and greater than zero, $z(\ell)$ is then called a *minimal training sample.*

The other portion of the data, $z(n-\ell)$, is used to compute the Bayes factor, using $\pi_i(\theta_i|z(\ell))$ as the prior. The resulting partial Bayes factor is

$$B_{21}(z(n-\ell)|z(\ell))=B_{21}^{\mathrm{N}}(z)\cdot B_{12}^{\mathrm{N}}(z(\ell)),$$

where

$$B_{21}^{\mathrm{N}}(z)=\frac{m_2^{\mathrm{N}}(z)}{m_1^{\mathrm{N}}(z)},\qquad B_{12}^{\mathrm{N}}(z(\ell))=\frac{m_1^{\mathrm{N}}(z(\ell))}{m_2^{\mathrm{N}}(z(\ell))}.$$

Berger and Pericchi (1996) propose using training sample of minimal size and take an average of the partial Bayes factors over all the training samples contained in the sample z. This gives the *arithmetic intrinsic* Bayes factor (AIBF) of M_2 against M_1 as

$$B_{21}^{\mathrm{AI}}(z) = B_{21}^{\mathrm{N}}(z)\frac{1}{L}\sum_{\ell=1}^{L} B_{12}^{\mathrm{N}}(z(\ell)), \tag{2}$$

where L is the number of minimal training samples $z(\ell)$ contained in z.

O'Hagan (1995) notes, *the fractional Bayes factor* (FBF) *is defined by analogy with the partial Bayes factor to avoid the arbitrariness of choosing a particular training sample or having to consider all possible subsets of a given size.* He defines the FBF as

$$B_{21}(b_n, z) = B_{21}^{\mathrm{N}}(z)\frac{\int_{\Theta_1} f_1(z|\theta_1)^{b_n}\pi_1^{\mathrm{N}}(\theta_1)\,\mathrm{d}\theta_1}{\int_{\Theta_2} f_2(z|\theta_2)^{b_n}\pi_2^{\mathrm{N}}(\theta_2)\,\mathrm{d}\theta_2}, \tag{3}$$

where b_n is some suitable constant that depends on the sample size n. Among the values recommended by O'Hagan for b_n we will choose $b_n = m/n$, where m is the minimal training sample size (Berger and Mortera, 1995; Moreno, 1997).

Notice that $B_{21}^{\mathrm{AI}}(z)$ and $B_{21}(b_n, z)$ do not depend on the arbitrary constants involved in the improper priors. Both, the AIBF and the FBF, contain $B_{21}^{\mathrm{N}}(z)$ as a common factor. The other factors appearing on the right-hand side of (2) and (3) can be considered as the correction terms of $B_{21}^{\mathrm{N}}(z)$ to avoid the dependence of the unspecified constants.

2.2. Intrinsic priors for the Behrens–Fisher problem

Consider models M_1: $\{f_1(z|\theta_1), \pi_1(\theta_1)\}$ and M_2: $\{f_2(z|\theta_2), \pi_2^{\mathrm{N}}(\theta_2)\}$, where $f_1(z|\theta_1)$ is nested in $f_2(z|\theta_2)$, $\pi_1(\theta_1)$ is a probability density and $\pi_2^{\mathrm{N}}(\theta_2)$ is a default improper prior. Assume that, for some integer n, the likelihood function $f_2(z_1,\dots,z_n|\theta_2)$ is integrable with respect to $\pi_2^{\mathrm{N}}(\theta_2)\,\mathrm{d}\theta_2$. Then, it can be shown that the intrinsic methodology generates a unique probability density $\pi_2^{\mathrm{I}}(\theta_2)$ for θ_2, namely an intrinsic prior. That prior is given by Berger and Pericchi (1996) as

$$\pi_2^{\mathrm{I}}(\theta_2) = \pi_2^{\mathrm{N}}(\theta_2)E_{z(\ell)|\theta_2}^{M_2} B_{12}^{\mathrm{N}}(z(\ell)), \tag{4}$$

where the expectation is with respect to the density of the minimal theoretical training sample $z(\ell) = (z_1,\dots,z_\ell)$ under model M_2, and

$$B_{12}^{\mathrm{N}}(z(\ell)) = \frac{m_1(z(\ell))}{m_2^{\mathrm{N}}(z(\ell))}.$$

In the Behrens–Fisher problem the sampling model $f_1(z|\theta_1) = \mathrm{N}(x|\mu,\tau_1^2)\ \mathrm{N}(y|\mu,\tau_2^2)$ is nested in $f_2(z|\theta_2) = \mathrm{N}(x|\mu_1,\sigma_1^2)\ \mathrm{N}(y|\mu_2,\sigma_2^2)$. It is easy to prove that the training sample is the random vector $z(\ell) = (x_1, x_2, y_1, y_2)$, where under M_1, x_1, x_2 are i.i.d. $\mathrm{N}(x|\mu,\tau_1^2)$, and under M_2 they are i.i.d. $\mathrm{N}(x|\mu_1,\sigma_1^2)$. Similarly, y_1, y_2 are i.i.d. $\mathrm{N}(y|\mu,\tau_2^2)$ under M_1, and i.i.d. $\mathrm{N}(y|\mu_2,\sigma_2^2)$ under M_2.

Lemma 1. *For each point* μ, τ_1, τ_2, *the intrinsic prior of* θ_2 *is*

$$\pi_2^{\mathrm{I}}(\theta_2|\mu,\tau_1,\tau_2) = \prod_{i=1}^{2} \mathrm{N}\left(\mu_i|\mu, \frac{\tau_i^2+\sigma_i^2}{2}\right) \mathrm{HC}^+(\sigma_i|0,\tau_i),$$

where $\mathrm{HC}^+(\sigma_i|0,\tau_i)$ *denotes a half Cauchy density.*

Proof. See Appendix A.1.

Lemma 1 states that conditionally on (μ,σ_i,τ_i), the μ_i's are independent a priori and normal distributed and conditionally on τ_i's, the σ_i's are independent a priori and half Cauchy distributed. Note that, if the labels assigned to the populations are interchanged, the intrinsic prior does change. Furthermore, the intrinsic prior mean of μ_1 and μ_2 is the same. That is, if μ_i means the treatment effect, we are a priori given the same mean to both treatment. This is sensible.

Eliminating the point (μ,τ_1,τ_2) by integration with respect to the default prior $\pi_1^{\mathrm{N}}(\theta_1)$, we obtain the unconditional intrinsic prior of θ_2,

$$\pi_2^{\mathrm{I}}(\theta_2)=c_1\int_{R\times R^+\times R^+}\prod_{i=1}^{2}\mathrm{N}\left(\mu_i|\mu,\frac{\sigma_i^2+\tau_i^2}{2}\right)\mathrm{HC}^+(\sigma_i|0,\tau_i)\frac{1}{\tau_1\tau_2}\,\mathrm{d}\mu\,\mathrm{d}\tau_1\mathrm{d}\tau_2.$$

The integral in this expression cannot be solved in closed form. However, this is not a serious inconvenience.

2.3. Fractional priors for the Behrens–Fisher problem

Consider again the models M_1: $\{f_1(z|\theta_1),\pi_1(\theta_1)\}$ and M_2: $\{f_2(z|\theta_2),\ \pi_2^{\mathrm{N}}(\mathrm{d}\theta_2)\}$, where $f_1(z|\theta_1)$ is nested in $f_2(z|\theta_2)$, $\pi_1(\theta_1)$ is a probability density and $\pi_2^{\mathrm{N}}(\theta_2)$ is a default improper prior. Under mild conditions, the fractional Bayes factor asymptotically corresponds to an actual Bayes factor associated with $\pi_1(\theta_1)$ and the probability density $\pi_2^{\mathrm{F}}(\theta_2)$. The latter density is given by Moreno (1997) as

$$\pi_2^{\mathrm{F}}(\theta_2)=\pi_2^{\mathrm{N}}(\theta_2)F_{12}^{M_2}(\theta_2), \tag{5}$$

provided that

$$F_{12}^{M_2}(\theta_2)=\lim_{n\to\infty}[M_2]\frac{\int_{\Theta_1}f_1(z_1,\ldots,z_n|\theta_1)^{b_n}\pi_1(\theta_1)\,\mathrm{d}\theta_1}{\int_{\Theta_2}f_2(z_1,\ldots,z_n|\theta_2)^{b_n}\pi_2^{\mathrm{N}}(\theta_2)\,\mathrm{d}\theta_2},$$

is a degenerate random variable and provided that $\pi_2^{\mathrm{F}}(\theta_2)$ integrates to one.

In the Behrens–Fisher problem the density is a product of normals and the training sample for each normal consists of two random variables. Therefore, it seems natural to formulate the correction term $F_{12}^{M_2}(\theta_2)$ by taking the fraction $2/n_i$, $i=1,2$, of each normal. That is, the fractions of the likelihood considered are

$$f_1(z_1,\ldots,z_n|\theta_1)^{b_n}=\left\{\prod_{i=1}^{n_1}\mathrm{N}(x_i|\mu,\tau_1^2)\right\}^{2/n_1}\left\{\prod_{i=1}^{n_2}\mathrm{N}(y_i|\mu,\tau_2^2)\right\}^{2/n_2},$$

and

$$f_2(z_1,\ldots,z_n|\theta_2)^{b_n}=\left\{\prod_{i=1}^{n_1}\mathrm{N}(x_i|\mu_1,\sigma_1^2)\right\}^{2/n_1}\left\{\prod_{i=1}^{n_2}\mathrm{N}(y_i|\mu_2,\sigma_2^2)\right\}^{2/n_2}.$$

For these fractions we obtain the following result.

Lemma 2. *For each point μ, τ_1, τ_2, the fractional prior of θ_2 is given by*

$$\pi_2^{\mathrm{F}}(\theta_2|\mu,\tau_1,\tau_2) = \prod_{i=1}^{2} \mathrm{N}\left(\mu_i|\mu, \frac{\tau_i^2}{2}\right) \mathrm{HN}^+\left(\sigma_i|0, \frac{\tau_i^2}{2}\right),$$

where $\mathrm{HN}^+(\sigma_i|0, \tau_i^2/2)$ denotes a half-Normal density.

Proof. See Appendix A.2.

That is, conditionally on (μ, τ_i), the μ_i's are independent and normally distributed, and conditionally on τ_i, the σ_i's are independent and half normally distributed. The main differences between the intrinsic and fractional priors $\pi_2^{\mathrm{I}}(\theta_2|\ \mu,\tau_1,\tau_2)$ and $\pi_2^{\mathrm{F}}(\theta_2|\mu,\tau_1,\tau_2)$, is that the tails of the densities of the $\sigma_i s$ are thicker for the intrinsic than for the fractional, and that (μ_i, σ_i) are dependent in the intrinsic whereas they are independent in the fractional prior.

The unconditional fractional prior $\pi_2^{\mathrm{F}}(\theta_2)$ is obtained by integration with respect to $\pi_1^{\mathrm{N}}(\theta_1)$ in the above expression.

3. Bayes factors

For samples $(\boldsymbol{x},\boldsymbol{y})$, the Bayes factor with the intrinsic prior $\pi_2^{\mathrm{I}}(\theta_2)$ is given by

$$B_{21}^{\mathrm{I}}(\boldsymbol{x},\boldsymbol{y}) = \frac{\int \prod_{i=1}^{n_1} \mathrm{N}(x_i|\mu_1,\sigma_1^2) \prod_{i=1}^{n_2} \mathrm{N}(y_i|\mu_2,\sigma_2^2)\pi_2^{\mathrm{I}}(\theta_2)\,\mathrm{d}\theta_2}{\int \prod_{i=1}^{n_1} \mathrm{N}(x_i|\mu,\tau_1^2) \prod_{i=1}^{n_2} \mathrm{N}(y_i|\mu,\tau_2^2)\pi_1^{\mathrm{N}}(\theta_1)\,\mathrm{d}\theta_1}, \tag{6}$$

where the constant c_1 cancels out. Unfortunately, the above integrals cannot be given in closed form.

Something similar can be said of the Bayes factor with the fractional priors $B_{21}^{\mathrm{F}}(\boldsymbol{x},\boldsymbol{y})$ which is obtained by replacing $\pi_2^{\mathrm{I}}(\theta_2)$ with $\pi_2^{\mathrm{F}}(\theta_2)$ in (6). We remark that $B_{21}^{\mathrm{I}}(\boldsymbol{x},\boldsymbol{y})$ and $B_{21}^{\mathrm{F}}(\boldsymbol{x},\boldsymbol{y})$ are not exactly actual Bayes factors but they are respectively, a unique limit of actual Bayes factors (Moreno et al., 1998). In practice, they share all the general properties of an actual Bayes factor. For instance, the posterior inference is consistent and coherence is assured.

For comparison purposes we consider the Schwarz's approximation (BIC), that we denote as $B_{21}^{\mathrm{S}}(\boldsymbol{x},\boldsymbol{y})$. For the Behrens–Fisher problem it turns out to be (see Appendix A.3),

$$B_{21}^{\mathrm{S}}(\boldsymbol{x},\boldsymbol{y}) = \left\{1 + \frac{(\bar{x}-\hat{\mu})^2}{s_x^2}\right\}^{n_1/2} \left\{1 + \frac{\bar{y}-\hat{\mu})^2}{s_y^2}\right\}^{n_2/2} \exp\left\{-\frac{1}{2}\log\frac{n_1 n_2}{n}\right\},$$

where $\hat{\mu}$ is the maximum likelihood estimator of μ based on the samples $(\boldsymbol{x},\boldsymbol{y})$ and $n = n_1 + n_2$.

Let us illustrate the behavior of the above Bayes factors with the data from Box and Tiao (1973, example 2.5.4, p. 107),

$$n_1 = 20, \quad s_x^2 = 12, \quad n_2 = 12, \quad s_y^2 = 40$$

Table 1
P-values and values of B_{21}^{S}, B_{21}^{F}, and B_{21}^{I}

$\lvert\bar{x}-\bar{y}\rvert$	P-values (Welch)	B_{21}^{S}	B_{21}^{F}	B_{21}^{I}
0.00	1	0.36	0.15	0.20
2.20	0.29	0.69	0.29	0.35
4.22	0.05	3.10	1.04	1.32
5.00	0.02	6.33	1.96	2.49
10.0	0.0001	>100	>100	>100

and several possible values for $|\bar{x}-\bar{y}|$. In Table 1, the corresponding values of B_{21}^{S}, B_{21}^{F}, and B_{21}^{I} are displayed.

Table 1 shows that B_{21}^{F} and B_{21}^{I} exhibit a similar behavior. For small values of the difference $|\bar{x}-\bar{y}|$, they clearly favor M_1, and as the difference grows the model M_2 is strongly favored. This behavior is sensible.

The Schwarz's approximation B_{21}^{S} favors the complex model M_2 for small departures of $|\bar{x}-\bar{y}|$ from zero. In particular, for $|\bar{x}-\bar{y}|=5$ the evidence in favor of M_2 is larger than those given by B_{21}^{F} and B_{21}^{I}.

For $\alpha=0.05$, the Welch's approximation gives the critical region $W_{0.05}=\{\bar{x},\ \bar{y}\colon |\bar{x}-\bar{y}|\geqslant 4.22\}$. This conclusion is in agreement with the Schwarz's approach but is in disagreement with those given by either B_{21}^{I} or B_{21}^{F}.

4. Conclusions

In this paper we have developed a default Bayesian analysis of the Behrens–Fisher problem. First, our analysis converts the conventional improper priors, independent uniforms for the location parameters and the log of the scale parameters, into priors for which the Bayes factor is well defined.

The resulting priors, intrinsic and fractional, are reasonable on their own. Their associated Bayes factors B_{21}^{I} and B_{21}^{F}, give numerical answers very close each other and show a sensible discriminatory behavior in selecting the null and the alternative hypothesis.

We have seen that to approximate the Bayes factor avoiding the influence of the priors, as Schwarz information criterion does, yields an approximation with a clear tendency to favor the complex model . Hence, the conclusion arrived upon by Kass and Wasserman (1995) on the good behavior of the Bayesian Information Criterion cannot be supported by the problem considered here.

We have also seen that the default Bayesian analysis here does not convey the same evidence as that of the classical frequentist test by Welch (1947). The former favors the null hypothesis for data, that the latter does not.

In summary, among the solutions to the Behrens–Fisher problem we have discussed that given by either the Bayes factor with intrinsic priors or the Bayes factor with fractional priors is what we recommend.

Acknowledgements

This research was partially supported by the grant PB 96-1388 from the Ministry of Education, Spain, and by MURST 1998, Italy. We are grateful to the editor Professor Singpurwalla and an anonymous referee for their help in improving an earlier version.

Appendix A.

A.1. Proof of Lemma 1

Consider the models

$$M_1^0 : f_1(z|\theta_1) = \mathrm{N}(x|\mu,\tau_1^2)\mathrm{N}(y|\mu,\tau_2^2), \pi_1^0(\theta_1) = \delta(\mu,\tau_1,\tau_2),$$

$$M_2 : f_2(z|\theta_2) = \mathrm{N}(x|\mu_1,\sigma_1^2)\mathrm{N}(y|\mu_2,\sigma_2^2), \pi_2^{\mathrm{N}}(\theta_2) = \frac{c_2}{\sigma_1\sigma_2},$$

where $\delta(\mu,\tau_1,\tau_2)$ is the Dirac's delta at the point $\theta_1 = (\mu,\tau_1,\tau_2)$ and c_2 an arbitrary positive constant. The intrinsic prior of θ_2 stated in (4) becomes

$$\pi_2^{\mathrm{I}}(\theta_2|\theta_1) = \pi_2^{\mathrm{N}}(\theta_2) E^{M_2}_{z(\ell)|\theta_2} B_{12}^{\mathrm{N}}(z(\ell)),$$

where

$$B_{12}^{\mathrm{N}}(z(\ell)) = \frac{\Pi_{i=1}^2 \mathrm{N}(x_i|\mu,\tau_1^2)\mathrm{N}(y_i|\mu,\tau_2^2)}{m_2^{\mathrm{N}}(z(\ell))}.$$

It is easy to see that

$$m_2^{\mathrm{N}}(z(\ell)) = \frac{c_2}{4}\frac{1}{|x_1-x_2||y_1-y_2|}.$$

Therefore,

$$\begin{aligned}\pi_2^{\mathrm{I}}(\theta_2|\theta_1) &= \frac{1}{4\pi^4\sigma_1^3\sigma_2^3\tau_1^2\tau_2^2}\\ &\times \int\int |x_1-x_2| \exp\left\{-d_x^2(\tau_1^{-2}+\sigma_1^{-2}) - \frac{(m_x-\mu)^2}{\tau_1^2} - \frac{(m_x-\mu_1)^2}{\sigma_1^2}\right\} \mathrm{d}x_1\,\mathrm{d}x_2\\ &\times \int\int |y_1-y_2| \exp\left\{-d_y^2(\tau_2^{-2}+\sigma_2^{-2}) - \frac{(m_y-\mu)^2}{\tau_2^2} - \frac{(m_y-\mu_2)^2}{\sigma_2^2}\right\} \mathrm{d}y_1\,\mathrm{d}y_2,\end{aligned}$$

where each integration is made on the real line and

$$d_x^2 = \frac{(x_1-x_2)^2}{4}, \quad m_x = \frac{x_1+x_2}{2},$$

$$d_y^2 = \frac{(y_1-y_2)^2}{4}, \quad m_y = \frac{y_1+y_2}{2}.$$

By changing to the new variables (u_i, v_i), $i = 1,2$,

$$u_1 = x_1 - x_2, \quad v_1 = x_1 + x_2,$$

$$u_2 = y_1 - y_2, \quad v_2 = y_1 + y_2$$

it follows that

$$\pi_2^{\mathrm{I}}(\theta_2|\theta_1) = \frac{2}{\pi}\frac{\tau_1}{\tau_1^2+\sigma_1^2}\mathrm{N}\left(\mu_1|\mu, \frac{\tau_1^2+\sigma_1^2}{2}\right)\frac{2}{\pi}\frac{\tau_2}{\tau_2^2+\sigma_2^2}\mathrm{N}\left(\mu_2|\mu, \frac{\tau_2^2+\sigma_2^2}{2}\right),$$

and this proves Lemma 1. □

A.2. Proof of Lemma 2

For the models in A.1 the fractional prior of θ_2 is

$$\pi_2^{\mathrm{F}}(\theta_2|\theta_1) = \pi_2^{\mathrm{N}}(\theta_2)F_{12}^{M_2}(\theta_2|\theta_1),$$

where

$$F_{12}^{M_2}(\theta_2|\theta_1) = \lim_{n_1,n_2\to\infty}[P_{\theta_2}]\frac{m_1(z|\mu,\tau_1,\tau_2)}{m_2^{\mathrm{N}}(z)},$$

and

$$m_1(z|\mu,\tau_1,\tau_2) = \left\{\prod_{i=1}^{n_1}\mathrm{N}(x_i|\mu,\tau_1^2)\right\}^{2/n_1}\left\{\prod_{i=1}^{n_2}\mathrm{N}(y_i|\mu,\tau_2^2)\right\}^{2/n_2},$$

$$m_2^{\mathrm{N}}(z) = \int\left\{\prod_{i=1}^{n_1}\mathrm{N}(x_i|\mu_1,\sigma_1^2)\right\}^{2/n_1}\left\{\prod_{i=1}^{n_2}\mathrm{N}(y_i|\mu_2,\sigma_2^2)\right\}^{2/n_2}\pi_2^{\mathrm{N}}(\theta_2)\,\mathrm{d}\theta_2.$$

It is easy to see that

$$\begin{aligned}&\lim_{n_1,n_2\to\infty}[P_{\theta_2}]m_1(z|\mu,\tau_1,\tau_2)\\ &=\frac{1}{(2\pi)^2}\frac{1}{\tau_1^2}\frac{1}{\tau_2^2}\exp\left\{-\frac{\sigma_1^2}{\tau_1^2}-\frac{(\mu_1-\mu)^2}{\tau_1^2}\right\}\exp\left\{-\frac{\sigma_2^2}{\tau_2^2}-\frac{(\mu_2-\mu)^2}{\tau_2^2}\right\}\\ &=\frac{1}{16}\prod_{i=1}^{2}\mathrm{N}\left(\mu_i|\mu,\frac{\tau_i^2}{2}\right)\mathrm{HN}^+\left(\sigma_i|0,\frac{\tau_i^2}{2}\right),\end{aligned}$$

and that

$$\lim_{n_1,n_2\to\infty}[P_{\theta_2}]m_2^{\mathrm{N}}(z) = \frac{c_2}{16\sigma_1\sigma_2}.$$

Therefore, the ratio is

$$F_{12}^{M_2}(\theta_2|\theta_1) = \frac{\sigma_1\sigma_2}{c_2}\prod_{i=1}^{2}\mathrm{N}\left(\mu_i|\mu,\frac{\tau_i^2}{2}\right)\mathrm{HN}^+\left(\sigma_i|0,\frac{\tau_i^2}{2}\right),$$

and, consequently,

$$\pi_2^{\mathrm{F}}(\theta_2|\theta_1) = \prod_{i=1}^{2}\mathrm{N}\left(\mu_i|\mu,\frac{\tau_i^2}{2}\right)\mathrm{HN}^+\left(\sigma_i|0,\frac{\tau_i^2}{2}\right).$$

Hence, Lemma 2 is proved. □

A.3. Derivation of the Schwarz's approximation to the Bayes factor

The Laplace approximation to the Bayes factor with priors $\pi_1^{\mathrm{N}}(\theta_1)$, $\pi_2^{\mathrm{N}}(\theta_2)$ is easily seen to be

$$B_{21}^{L}(\boldsymbol{x},\boldsymbol{y}) = \left\{1+\frac{(\bar{x}-\hat{\mu})^2}{s_x^2}\right\}^{n_1/2}\left\{1+\frac{(\bar{y}-\hat{\mu})^2}{s_y^2}\right\}^{n_2/2}\left\{\frac{\det \hat{I}_1}{\det \hat{I}_2}\right\}^{1/2}$$
$$\times\frac{\pi_2^{\mathrm{N}}(\hat{\sigma}_1,\hat{\sigma}_2)}{\pi_1^{\mathrm{N}}(\hat{\tau}_1,\hat{\tau}_2)}(2\pi)^{1/2},$$

where $\hat{I}_1,\hat{I}_2$ denotes the observed information matrices.

The ratio $\det \hat{I}_1/\det \hat{I}_2$ can be written as

$$\frac{\det \hat{I}_1}{\det \hat{I}_2} = \frac{n}{n_1 n_2}\hat{\varphi},$$

where

$$\hat{\varphi} = \frac{(s_x s_y)^4}{(\hat{\tau}_1\hat{\tau}_2)^4}\left\{\frac{n_1}{n}\hat{\tau}_2^2\left(1-2\frac{(\bar{x}-\hat{\mu})^2}{\hat{\tau}_1^2}\right)+\frac{n_2}{n}\hat{\tau}_1^2\left(1-2\frac{(\bar{y}-\hat{\mu})^2}{\hat{\tau}_2^2}\right)\right\}.$$

Notice that $\hat{\varphi}$ tends to a constant as $n_1,n_2 \to \infty$. Setting

$$\log\left\{\hat{\varphi}^{1/2}\frac{\pi_2^{\mathrm{N}}(\hat{\sigma}_1,\hat{\sigma}_2)}{\pi_1^{\mathrm{N}}(\hat{\tau}_1,\hat{\tau}_2)}(2\pi)^{1/2}\right\} = 0,$$

the Schwarz's approximation is obtained.

References

Berger, J.O., Mortera, J., 1995. Discussion to A. O'Hagan, J. Roy. Statist. Soc. Ser. B 57, 99–138.

Berger, J.O., Pericchi, L.R., 1996. The intrinsic Bayes factor for model selection and prediction. J. Amer. Statist. Assoc. 91, 109–122.

Berger, J.O., Pericchi, L.R., 1997. On the justification of default and intrinsic Bayes factor. In: J.C. et al. (Eds.), Modelling and Prediction. New York, Springer, pp. 276–293.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, MA.

Fisher, R.A., 1936. The fiducial arguments in statistical inference. Ann. Eugen. 6, 391–422.

Jeffreys, H., 1961. Theory of Probability, Oxford University Press.

Kass, R.E., Wasserman, L., 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz Criterion. J. Amer. Statist. Assoc. 90, 928–934.

Lindley, D.V., 1970. Probability and Statistics. University Press. Cambridge.

Mehta, J.S., Srinivasan, R., 1970. On the Behrens–Fisher problem. Biometrika 57, 649–655.

Moreno, E., 1997. Bayes Factor for Intrinsic and Fractional Priors in Nested Models: Bayesian Robustness. IMS Lectures Notes-Monograph Series. vol. 31, 257–270.

Moreno, E., Bertolino, F., Racugno, W., 1998. An intrinsic limiting procedure for model selection and hypotheses testing. J. Amer. Statist. Assoc. 93, 1451–1460.

O'Hagan, A., 1995. Fractional Bayes factor for model comparison (with discussion). J. Roy. Statist. Soc. Ser. B 57, 99–138.

O'Hagan, A., 1997. Properties of intrinsic and fractional Bayes factors. Test 6, 101–119.

Pfanzagl, J., 1974. On the Behrens–Fisher problem. Biometrika 61, 39–47.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.
Wald, A., 1955. Testing the difference between the means of two normal populations with unknown standard deviations. In: T.W. Anderson et al. (Eds.), Selected papers in Statistics and Probability, Stanford University Press, pp. 669–695.
Welch, B.L., 1947. The generalization of Student's problem when several populations are involved. Biometrika 34, 28–35.