A fully Bayesian solution to k-sample tests for comparison and the Behrens-Fisher problem based on the Henstock-Kurzweil integral

Fabrice Pautot, general engineer ECL96, independent researcher, Paris, France

Draft version 1.1 to be LaTeXified, the 24th of September 2019

Email to: fabrice.pautot@laposte.net

Abstract: We present a simple, fully probabilistic, Bayesian solution to k-sample omnibus tests for comparison, with the Behrens-Fisher problem as a special case, which is free from the many defects found in the standard, classical, frequentist, likelihoodist and Bayesian approaches to those problems. We solve the main measure-theoretic difficulty for degenerate problems with continuous parameters of interest and Lebesgue-negligible point null hypothesis by approximating the corresponding continuous random variables by sequences of discrete ones defined on partitions of the parameter spaces and by taking the limit of the prior-to-posterior ratios of the probability of the null hypothesis for the corresponding discrete problems. Those limits are well defined under proper technicalities thanks to the Henstock-Kurzweil integral that is as powerful as the Lebesgue integral but still relies on Riemann sums, which are essential in the present approach. The solutions to the relative continuous problems take the form of Bayes-Poincaré factors that are new objects in Bayesian probability theory and should play a key role in the general theory of point null hypothesis testing, including other important problems such as the Jeffreys-Lindley paradox.

Keywords: Behrens-Fisher problem, k-sample tests for comparison, point null hypothesis testing, well-posed problem, Bayes factors, discrete and continuous random variables and problems, absolute and relative solutions, Riemann sums, Lebesgue measure, Henstock-Kurzweil integral, Bayes-Poincaré factors.

Introduction

The Behrens-Fisher problem(s) [1][2][3][4][5][6][8][9][10][11][12][13][14] and, more generally, k-sample hypothesis tests for comparison, is one of the most famous open problem in statistics and applied probability theory since 1929.

Let us first recall that, according to the numerous authors over the last ninety years, the Behrens-Fisher problems come in two main variants: the **hypothesis testing problems** (i.e. test the equality of the numerical values of the parameters of interest) and the **estimation problems** (estimate the difference between the numerical values of those parameters). Hereafter, we deal only with the hypothesis testing problems, which are supposed to be the most important and difficult ones [14][17], even if they are often formulated as interval estimation problems and confidence tests on the difference between both parameters of interest for 2-sample problems.

To the best of our knowledge, there is no universally accepted solution to this problem in any statistical framework, frequentist, likelihoodist, fiducial or Bayesian. This is due to the interplay of several issues and difficulties including the proper treatment of the nuisance parameters and the non-existence of an uniformly most powerful (UMP) test proved by Linnik [5] (in the frequentist and likelihoodist frameworks), the proper assignment of prior probabilities and probability distributions for the hypotheses and the parameters, in particular, the fact that the null hypothesis has null prior and posterior probabilities if the parameter of interest is continuous (in the Bayesian framework), the logical (in)dependence of the experiments, etc.

The original, historical Behrens-Fisher problem runs as follows. Given two mutually independent and conditionally independently and identically distributed samples $x_1 = \left(x_1^1, \ldots, x_1^{n_1}\right)$ and $x_2 = \left(x_2^1, \ldots, x_2^{n_2}\right)$ of sizes n_1 and n_2 from two

Gaussian random variables $\mathcal{N}\left(\mu_1,\sigma_1^2\right)$ and $\mathcal{N}\left(\mu_2,\sigma_2^2\right)$ with probability density functions $f\left(x\middle|\mu_1,\sigma_1\right)$ and $f\left(x\middle|\mu_2,\sigma_2\right)$ respectively, test the null hypothesis H_0

$$H_0 : \mu_1 = \mu_2$$

against the omnibus alternative hypothesis H_1

$$H_1 : \mu_1 \neq \mu_2$$

when σ_1 and σ_2 are unknown and not necessarily equal to each other. Some authors (but not Fisher) further assume that $\sigma_1 \neq \sigma_2$. We do not, but the solution to come can be generalized to this case…

To be short, let us just recall that within the frequentist, likelihoodist and fiducial statistical frameworks, the well-known, classical solutions typically rely on adaptations of the independent/unpaired Student t statistic for the 2-sample Gaussian test for comparison with equal/homoscedastic variances $\sigma_1 = \sigma_2$ and unequal sample size

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

where $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$, $s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(x_1^i-\bar{x}_1\right)$, $s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}\left(x_2^i-\bar{x}_2\right)$, $\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}x_1^i$, $\bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_1}x_2^i$

Such modifications and adaptations include the Behrens-Fisher statistic itself [10][14][15][16], the Welch-Alpin-Satterthwaite t statistics [11][15], the Wald-Romanovskaja satistic [15] and many other variations [15].

On the Bayesian side, Jeffreys (1940) [17] derived the same statistic as Fisher's [16] with a different interpretation, of course. Jeffreys' original derivation was quite *ad hoc* and complex [9] but it can be recast in his own, modern and general, Bayes factor framework for testing binary hypotheses. Let us briefly recall this Bayes factor framework in the case of the historical 2-sample Behrens-Fisher problem.

Under hypothesis H_1, the likelihood or model M_1 is

$$p\left(x_1,x_2\middle|\mu_1,\mu_2,\sigma_1,\sigma_2,M_1\right) = p\left(x_1\middle|\mu_1,\sigma_1,M_1\right)p\left(x_2\middle|\mu_2,\sigma_2,M_1\right) = \prod_{i=1}^{n_1} f\left(x_1^i\middle|\mu_1,\sigma_1\right)\prod_{i=1}^{n_2} f\left(x_2^i\middle|\mu_2,\sigma_2\right)$$

by conditional independence, so that the probability for the data $\left(x_1,x_2\right)$ given a joint prior probability distribution for the parameters $p\left(\mu_1,\sigma_1,\mu_2,\sigma_2\right)$ is

$$p\left(x_1,x_2\middle|M_1\right) = \iint\iint \prod_{i=1}^{n_1} f\left(x_1^i\middle|\mu_1,\sigma_1\right)\prod_{i=1}^{n_2} f\left(x_2^i\middle|\mu_2,\sigma_2\right)p\left(\mu_1,\sigma_1,\mu_2,\sigma_2\right)\mathrm{d}\mu_1\mathrm{d}\sigma_1\mathrm{d}\mu_2\mathrm{d}\sigma_2$$

Under hypothesis H_0, by definition there is a common parameter of interest μ_0

$$\mu_0 = \mu_1 = \mu_2$$

so that the likelihood or model M_0 is

$$p\left(x_1,x_2\middle|\mu_0,\sigma_1,\sigma_2,M_0\right)=p\left(x_1\middle|\mu_0,\sigma_1,M_0\right)p\left(x_2\middle|\mu_0,\sigma_2,M_0\right)=\prod_{i=1}^{n_1} f\left(x_1^i\middle|\mu_0,\sigma_1\right)\prod_{i=1}^{n_2} f\left(x_2^i\middle|\mu_0,\sigma_2\right)$$

and the probability of the data given a joint prior probability distribution for the parameters $p\left(\mu_0,\sigma_1,\sigma_2\right)$ is

$$p\left(x_1,x_2\middle|M_0\right)=\iiint\prod_{i=1}^{n_1} f\left(x_1^i\middle|\mu_0,\sigma_1\right)\prod_{i=1}^{n_2} f\left(x_2^j\middle|\mu_0,\sigma_2\right)p\left(\mu_0,\sigma_1,\sigma_2\right)\mathrm{d}\mu_0\mathrm{d}\sigma_1\mathrm{d}\sigma_2$$

Therefore, the classical Bayesian solution relies on the “Bayes factor”

$$B_{01}\triangleq\frac{p\left(x_1,x_2\middle|M_0\right)}{p\left(x_1,x_2\middle|M_1\right)}=\frac{\iiint\prod_{i=1}^{n_1} f\left(x_1^i\middle|\theta_0,\sigma_1\right)\prod_{i=1}^{n_2} f\left(x_2^i\middle|\mu_0,\sigma_2\right)p\left(\mu_0,\sigma_1,\sigma_2\right)\mathrm{d}\mu_0\mathrm{d}\sigma_1\mathrm{d}\sigma_2}{\iint\iint\prod_{i=1}^{n_1} f\left(x_1^i\middle|\mu_1,\sigma_1\right)\prod_{i=1}^{n_2} f\left(x_2^i\middle|\mu_2,\sigma_2\right)p\left(\mu_1,\sigma_1,\mu_2,\sigma_2\right)\mathrm{d}\mu_1\mathrm{d}\sigma_1\mathrm{d}\mu_2\mathrm{d}\sigma_2}$$

that is equal to the genuine Bayes factor of interest, the prior-to-posterior odds ratio

$$B_{01}=\left[\frac{p\left(M_0\middle|x_1,x_2\right)}{p\left(M_1\middle|x_1,x_2\right)}\right]/\left[\frac{p\left(M_0\right)}{p\left(M_1\right)}\right]$$

if and only if $p\left(M_0\right)\neq 0$ and $p\left(M_0\right)\neq 1$.

For instance, for the homoscedastic 2-sample problem with $\sigma_1=\sigma_2=\sigma$, Gönen *et al.* [9] show that, for certain priors and under certain conditions, the Jeffreys’ Bayes factor reduces to

$$B_{01}=\left[\frac{1+t^2/\nu}{1+t^2/\left(\nu\left(1+n_\delta\sigma_\delta^2\right)\right)}\right]^{-(\nu+1)/2}\left(1+n_\delta\sigma_\delta^2\right)^{1/2}$$

where t is the statistic defined above, $\nu=n_1+n_2-2$, $n_\delta=\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}$ and σ_δ^2 is the prior variance on $\frac{\mu_1-\mu_2}{\sigma}$.

See also [8]. As already found by Jeffreys [17], this kind of results shows that the classical approaches within the frequentist, likelihoodist and fiducial statistical frameworks can be derived and interpreted in Jeffreys’ Bayesian Bayes factor framework for binary hypothesis testing.

Unfortunately, we hold that this classical Bayes factor approach is not correct for at least one simple and sufficient reason. Indeed, under the null model M_0, we cannot say that there is a common, a single parameter θ_0 or μ_0: this is not because two parameters θ_1 and θ_2 have the same numerical values that there is only one parameter!

There is a confusion between two different concepts: the identity of two parameters (so that there is only one of them) and the equality of their numerical values. But **equality is not, does not imply identity**, as pointed out especially by Henri Poincaré. See for instance this paper (in French) [7], page 30 of the .pdf version:

The impossibility of understanding the application of the mathematical continuum to experimental data *(…)* ***as soon as one interprets the equality of two rational or real numbers as the identity of two entities****, allows us to conclude…*

For instance, we find this mistake in equation 1 of [6]

$$H_0 : \mu_c = \mu_t = \mu$$

and around equation 4 of [2]

The hypothesis sv assumes the [two] means and the [two] standard deviations are the same, so ***two parameters*** *(a constant A and a standard deviation sigma1) [instead of 4] have been removed by marginalization.*

Eq. 4

where $B \to A$ *and* $\sigma_1 \to \sigma_2$.

See also [9] for another variation. This mistake alone, the *principle of the identity of equality and identity* rejected by Poincaré, was our initial stimulus for developing the new and completely different Bayesian solution that we describe below. But, *a posteriori*, it shall appear that the standard Bayes factor and likelihoodist approaches described above suffer other, even more cogent, defects and criticisms.

Formal statement of the well-posed problems and their absolute and relative solutions

Now, let us make a formal statement of the informal and the well-defined, well-posed and formal problems that we address together with their absolute and relative solutions.

According the Lehmann [12], any of the following problems can be called (generalized) Behrens-Fisher problems, but we reserve this terminology for the historical 2-sample Gaussian problem.

Let $k \in \mathbb{N}, k > 1$. Let

$$\mathcal{L}(\theta,\eta), \theta = \left(\theta^1,\ldots,\theta^{d_\theta}\right) \in \Theta \subset \mathbb{R}^{d_\theta}, \eta \in \mathrm{H} \subset \mathbb{R}^{d_\eta}, d_\theta \in \mathbb{N}^*, d_\eta \in \mathbb{N}^*$$

be a parametric family of probability distributions on some probability space $(\Omega,\mathcal{F},\mathbb{P}), \Omega \subset \mathbb{R}$ (i.e. the data can be discrete or continuous) with bounded and positive probability mass function or probability density function $p\left(x \middle| \theta,\eta\right)$.

θ is the parameter of interest and η is the nuisance parameter.

Remark: we can consider more general problem with different parametric families of distributions $\mathcal{L}_i\left(\theta_i,\eta_i\right), i = 1,k$

Let

$$X_i \sim \mathcal{L}\left(\theta_i,\eta_i\right), \theta_i \in \Theta_i, \eta_i \in \mathrm{H}_i, i = 1,k$$

be k mutually independent random variables.

Let

$$x_i = \left(x_i^1,\ldots,x_i^{n_i}\right), i = 1,k, n_i \in \mathbb{N}^*$$

be k conditionally independent and identically distributed samples of size n_i drawn from random variables X_i.

Let $\Theta_0 = \bigcap_{i=1}^{k} \Theta_i$. Let also

$$p(\theta_i,\eta_i), i = 1,k$$

be the joint prior (proper) probability distributions for the parameters of each experiment.

Let

$$p(\theta_i) = \int_{\mathrm{H}_i} p(\theta_i, \eta_i)\,\mathrm{d}\eta_i,\ i = 1, k$$

be the marginal prior probability distributions for the parameters of interest. Remark: the integrals must be understood as generalized sums: they are sums if η_i is discrete.

Let

$$p(\theta_i, \eta_i | x_i) = \frac{p(\theta_i, \eta_i)\, p(x_i | \theta_i, \eta_i)}{\int_{\Theta_i}\int_{\mathrm{H}_i} p(\theta_i, \eta_i)\, p(x_i | \theta_i, \eta_i)\,\mathrm{d}\eta_i \mathrm{d}\theta_i},\ i = 1, k \$$$

be the joint posterior probability distributions for the parameters.

Last, let

$$p(\theta_i | x_i) = \int_{\mathrm{H}_i} p(\theta_i, \eta_i | x_i)\,\mathrm{d}\eta_i,\ i = 1, k$$

be the marginal posterior probability distributions for the parameters of interest.

We first define the informal, ill-defined k -sample testing problems for comparison that we address.

Definition: the **informal** k -sample (multivariate) hypothesis test for comparison problem is the problem $\mathcal{P}_0$

$$\mathcal{P}_0 : \begin{cases} \text{Test the null hypothesis } H_0 \\ H_0 : \theta_1 = \theta_2 = \ldots = \theta_k \\ \text{against the omnibus alternative hypothesis } H_1 \\ H_1 : \neg H_0 \end{cases}$$

Definition: if $d_\theta = 1$, then $\mathcal{P}_0$ is said to be **univariate**.

Definition: if for $i = 1, k, \theta_i$ is discrete, then $\mathcal{P}_0$ is said to be **discrete**.

Definition: if $\exists i = 1, k, \exists j = 1, d_\theta \diagdown p(\theta_i)$ is continuous in θ_i^j (i.e. no atoms), then $\mathcal{P}_0$ is said to be **continuous**.

Definition: **the historical Behrens-Fisher problem** is the informal 2 -sample, univariate and continuous problem $\mathcal{P}_0$ with $\mathcal{L}(\theta,\eta) := \mathcal{N}(\mu,\sigma^2)$, $\mathcal{N}(\mu,\sigma^2)$ is the parametric family of Gaussian distributions, $\theta := \mu$ and $\eta := \sigma$.

Now, what remains to be well-defined is this: what does it mean to test H_0 against H_1? How to take the final decision that is case-by-case and may depend on subjective and non-quantitative, qualitative considerations? This is simply not part of the well-defined, well-posed problem. We do not mean we cannot put some mathematics on the decision-theoretic part of the whole problem. We mean that, in any case, this decision will rely, will be based on the prior and the posterior probabilities of the hypotheses H_0 and H_1, that is on the prior and posterior probabilities of the null hypothesis H_0 for an omnibus test. Moreover, there is nothing specific to the Behrens-Fisher problem and k -sample tests for comparison in the decision-theoretic parts of the whole problems. It follows that the decision-theoretic part of the problem can and has to be decoupled from the probabilistic part. Therefore,

Definition: the **formal**, **absolute** k -sample (multivariate) hypothesis test for comparison problem is the problem $\mathcal{P}_0'$

$$\mathcal{P}_0': \begin{cases} \text{Compute the prior and posterior probabilities } p(H_0) \text{ and } p(H_0 | x_1,...,x_k) \\ \text{of the null hypothesis } H_0 : \theta_1 = \theta_2 = ... = \theta_k \end{cases}$$

Lemma: if $\mathcal{P}_0'$ is discrete, then $p(H_0) = \sum_{\theta\in\Theta_0} \prod_{i=1}^{k} p(\theta_i = \theta)$.

Proof:

$$\begin{aligned} & p(H_0) = p(\theta_1 = \theta_2 = ... = \theta_k) = \\ & \sum_{\theta\in\Theta_0} p(\theta_1 = \theta_2 = ... = \theta_k = \theta) = && \text{by total probability} \\ & \sum_{\theta\in\Theta_0} p(\theta_1 = \theta, \theta_2 = \theta, ..., \theta_k = \theta) = \\ & \sum_{\theta\in\Theta_0} \prod_{i=1}^{k} p(\theta_i = \theta) && \text{by independence} \end{aligned}$$

Corollary: if $\mathcal{P}_0'$ is discrete, then $p(H_0 | x_1,...,x_k) = \sum_{\theta\in\Theta_0} \prod_{i=1}^{k} p(\theta_i = \theta | x_i)$

Proof: replace the marginal priors by the marginal posteriors in previous lemma.

Corollary: the solution to the formal discrete absolute problem $\mathcal{P}_0'$ is given by

$$p(H_0) = \sum_{\theta\in\Theta_0} \prod_{i=1}^{k} p(\theta_i = \theta) \quad \text{and} \quad p(H_0 | x_1,...,x_k) = \sum_{\theta\in\Theta_0} \prod_{i=1}^{k} p(\theta_i = \theta | x_i)$$

Lemma: the solution to the formal continuous absolute problem $\mathcal{P}_0'$ is given by

$$p(H_0) = p(H_0 | x_1,...,x_k) = 0$$

Proof: the set $\{(\theta_1,\theta_2,...,\theta_k) \setminus (\theta_1,\theta_2,...,\theta_k) \in \Theta_0^k, \theta_1 = \theta_2 = ... = \theta_k\}$ has Lebesgue probability measure 0 because, by definition, $\forall\theta_0 \in \Theta_i^j, p(\theta_i^j = \theta_0) = 0$ since θ_i^j is a continuous random variable without atoms.

Hence, the formal continuous absolute problem $\mathcal{P}_0'$ admits a **trivial and totally useless absolute solution**. Therefore we need to introduce formal, **weaker**, **relative** problems with non-trivial and useful solutions.

For sake of simplicity, we present those relative problems and solutions in the univariate and compact case. But it should be clear that those problems and solutions can be easily generalized to the non-compact and multivariate case, modulo some integration-theoretic technicalities that are beyond the scope of the present paper. In the same way, those definitions can be easily generalized to joint and marginal improper prior probability distributions, defined as limits of sequences of proper ones, especially because we do not need to go through improper integrals with the Henstock-Kurzweil integral, unlike the Riemann integral.

Suppose $\forall i = 1,k, \Theta_i = [L_i, U_i]$ is compact. Let $\Delta\theta^l = \dfrac{\mu(\Theta_0)}{l}$ and let

$$\Theta_i^l = \left\{\theta_{i,j}^l, j = 1,\dots,\left|\Theta_i^l\right|\right\}$$

be $\Delta\theta^l$-fine partitions of Θ_i. Generally speaking, Θ_0 is a finite set of intervals. Suppose that the partitions $\Theta_i, i = 1,k$ coincide on each interval and form a partition of it. Let Θ_0^l be the set of those common partitions.

$\forall i = 1,k$, let $\left(\theta_i^l\right)_{l\in\mathbb{N}^*}$ be sequences of discrete random variables on partitions Θ_i^l with prior probability mass functions

$$\forall \theta \in \Theta_i^l,\ p_{\theta_i^l}(\theta) = \frac{p_{\theta_i}(\theta)}{\sum\limits_{\theta\in\Theta_i^l} p_{\theta_i}(\theta)}$$

respectively.

Lemma: $\forall l \in \mathbb{N}^*$, the solution to the formal discrete absolute problem $\mathcal{P}_0'$ for random variables $\theta_i^l, i = 1,k$ is given by

$$p\left(H_0 \middle| l, \Theta_0^l\right) = \sum_{\theta\in\Theta_0^l} \prod_{i=1}^{k} p_{\theta_i^l}(\theta) \quad \text{and} \quad p\left(H_0 \middle| x_1,\dots,x_k, l, \Theta_0^l\right) = \sum_{\theta\in\Theta_0^l} \prod_{i=1}^{k} p_{\theta_i^l | x_i}(\theta)$$

Corollary: if the function $\theta \mapsto \prod\limits_{i=1}^{k} p_{\theta_i}(\theta), \theta \in \Theta_0$ is Lebesgue-integrable, then

$$\lim_{l\to+\infty} p\left(H_0 \middle| l, \Theta_0^l\right) = \lim_{l\to+\infty} p\left(H_0 \middle| x_1,\dots,x_k, l, \Theta_0^l\right) = 0$$

Proof:

$$p\left(H_0 \middle| l, \Theta_0^l\right) = \sum_{\theta\in\Theta_0^l} \prod_{i=1}^{k} p_{\theta_i^l}(\theta) = \sum_{\theta\in\Theta_0^l} \prod_{i=1}^{k} \frac{p_{\theta_i}(\theta)}{\sum\limits_{\theta\in\Theta_i^l} p_{\theta_i}(\theta)} = \frac{\sum\limits_{\theta\in\Theta_0^l} \prod\limits_{i=1}^{k} p_{\theta_i}(\theta)}{\prod\limits_{i=1}^{k} \sum\limits_{\theta\in\Theta_i^l} p_{\theta_i}(\theta)} = \left(\Delta\theta^l\right)^{k-1} \frac{\Delta\theta^l \sum\limits_{\theta\in\Theta_0^l} \prod\limits_{i=1}^{k} p_{\theta_i}(\theta)}{\prod\limits_{i=1}^{k} \Delta\theta^l \sum\limits_{\theta\in\Theta_i^l} p_{\theta_i}(\theta)}$$

$$\underset{l\to+\infty}{\sim} \left(\Delta\theta^l\right)^{k-1} \frac{\int\limits_{\Theta_0} \prod\limits_{i=1}^{k} p_{\theta_i}(\theta) \mathrm{d}\theta}{\prod\limits_{i=1}^{k} \int\limits_{\Theta_i} p_{\theta_i}(\theta) \mathrm{d}\theta} \underset{l\to+\infty}{\to} 0$$

because i) all functions are Henstock-Kurzweil-integrable [13] if they are Lebesgue-integral and ii) all integrals are finite and non-zero on compact $\Theta_i, i = 1,k$ and Θ_0 if the functions are bounded and positive. Same for $\lim\limits_{l\to+\infty} p\left(H_0 \middle| x_1,\dots,x_k, l, \Theta_0^l\right) = 0$.

Remark: at this point, we should certainly add something to say that the sequences of discrete random variables $\left(\theta_i^l\right)_{l\in\mathbb{N}^*}$ converge towards the continuous random variables $\theta_i, i = 1,k$ respectively. But what should we add exactly, we do not know: theorems or… definitions???

Now, our main theorem.

Theorem: if the functions $\theta \mapsto \prod_{i=1}^{k} p_{\theta_i}(\theta)$ and $\theta \mapsto \prod_{i=1}^{k} p_{\theta_i|x_i}(\theta)$, $\theta \in \Theta_0$ are Lebesgue-integrable, then

$$\lim_{l\to+\infty} \frac{p\left(H_0 \middle| x_1,\dots,x_k,l,\Theta_0^l\right)}{p\left(H_0 \middle| l,\Theta_0^l\right)} \triangleq B_{01} = \left[\frac{\int_{\Theta_0} \prod_{i=1}^{k} p_{\theta_i|x_i}(\theta)\mathrm{d}\theta}{\prod_{i=1}^{k} \int_{\Theta_i} p_{\theta_i|x_i}(\theta)\mathrm{d}\theta}\right] / \left[\frac{\int_{\Theta_0} \prod_{i=1}^{k} p_{\theta_i}(\theta)\mathrm{d}\theta}{\prod_{i=1}^{k} \int_{\Theta_i} p_{\theta_i}(\theta)\mathrm{d}\theta}\right]$$

Proof:

$$\frac{p\left(H_0 \middle| x_1,\dots,x_k,l,\Theta_0^l\right)}{p\left(H_0 \middle| l,\Theta_0^l\right)} = \left[\left(\Delta\theta^l\right)^{k-1} \frac{\Delta\theta^l \sum_{\theta\in\Theta_0^l} \prod_{i=1}^{k} p_{\theta_i|x_i}\left(\theta_i^l\right)}{\prod_{i=1}^{k} \Delta\theta^l \sum_{\theta\in\Theta_i^l} p_{\theta_i|x_i}\left(\theta_i^l\right)}\right] / \left[\left(\Delta\theta^l\right)^{k-1} \frac{\Delta\theta^l \sum_{\theta\in\Theta_0^l} \prod_{i=1}^{k} p_{\theta_i}\left(\theta_i^l\right)}{\prod_{i=1}^{k} \Delta\theta^l \sum_{\theta\in\Theta_i^l} p_{\theta_i}\left(\theta_i^l\right)}\right]$$

$$\xrightarrow[l\to+\infty]{} \left[\frac{\int_{\Theta_0} \prod_{i=1}^{k} p_{\theta_i|x_i}(\theta)\mathrm{d}\theta}{\prod_{i=1}^{k} \int_{\Theta_i} p_{\theta_i|x_i}(\theta)\mathrm{d}\theta}\right] / \left[\frac{\int_{\Theta_0} \prod_{i=1}^{k} p_{\theta_i}(\theta)\mathrm{d}\theta}{\prod_{i=1}^{k} \int_{\Theta_i} p_{\theta_i}(\theta)\mathrm{d}\theta}\right]$$

because, as before, i) all functions are Henstock-Kurzweil-integrable if they are Lebesgue-integral and ii) the integrals are finite and non-zero on compact $\Theta_i, i = 1,k$ and Θ_0 if the functions are bounded and positive.

Remark: in the multivariate case $d_\theta > 1$, we need to check that the Henstock-Kurzweil integral still is equivalent to the Lebesgue integral. If it is not, we can assume all functions to be Henstock-Kurzweil-integrable.

Remark: the conditions on the probability distributions and the different integrals are sufficient but not necessary. Our purpose is just to avoid technicalities that are beyond the scope of the present paper.

Definition: the **formal**, **relative**, continuous k-sample (multivariate) hypothesis test for comparison problem is the problem $\mathcal{P}_0$ "

$$\mathcal{P}_0'' : \left\{ \text{Compute } B_{01} = \lim_{l\to+\infty} \frac{p\left(H_0 \middle| x_1,\dots,x_k,l,\Theta_0^l\right)}{p\left(H_0 \middle| l,\Theta_0^l\right)} \right.$$

whenever the limit exists.

Lemma:

$$B_{01} = \lim_{l\to+\infty} \left[\frac{p\left(H_0 \middle| x_1,\dots,x_k,l,\Theta_0^l\right)}{p\left(H_1 \middle| x_1,\dots,x_k,l,\Theta_0^l\right)}\right] / \left[\frac{p\left(H_0 \middle| l,\Theta_0^l\right)}{p\left(H_1 \middle| l,\Theta_0^l\right)}\right]$$

Proof:

$$\lim_{l\to+\infty}\left[\frac{p\left(H_0|x_1,\ldots,x_k,l,\Theta_0^l\right)}{p\left(H_1|x_1,\ldots,x_k,l,\Theta_0^l\right)}\right]/\left[\frac{p\left(H_0|l,\Theta_0^l\right)}{p\left(H_1|l,\Theta_0^l\right)}\right]=\lim_{l\to+\infty}\left[\frac{p\left(H_0|x_1,\ldots,x_k,l,\Theta_0^l\right)}{p\left(H_0|l,\Theta_0^l\right)}\right]/\left[\frac{p\left(H_1|x_1,\ldots,x_k,l,\Theta_0^l\right)}{p\left(H_1|l,\Theta_0^l\right)}\right]=$$

$$\lim_{l\to+\infty}\left[\frac{p\left(H_0|x_1,\ldots,x_k,l,\Theta_0^l\right)}{p\left(H_0|l,\Theta_0^l\right)}\right]/\left[\frac{1-p\left(H_0|x_1,\ldots,x_k,l,\Theta_0^l\right)}{1-p\left(H_0|l,\Theta_0^l\right)}\right]=\lim_{l\to+\infty}\left[\frac{p\left(H_0|x_1,\ldots,x_k,l,\Theta_0^l\right)}{p\left(H_0|l,\Theta_0^l\right)}\right]=B_{01}$$

because $\lim_{l\to+\infty}\frac{1-p\left(H_0|x_1,\ldots,x_k,l,\Theta_0^l\right)}{1-p\left(H_0|l,\Theta_0^l\right)}=1$.

It follows that B_{01} is the limit of the sequence of Bayes factors for the discrete problems. Therefore,

Definition: B_{01} is called a **limit**, a **relative**, a **generalized Bayes factor**, or a **Bayes-Poincaré** factor.

Remark: it is worth observing, perhaps for the first time, that those Bayes factors

$$\left[\frac{p\left(H_0|x_1,\ldots,x_k,l,\Theta_0^l\right)}{p\left(H_1|x_1,\ldots,x_k,l,\Theta_0^l\right)}\right]/\left[\frac{p\left(H_0|l,\Theta_0^l\right)}{p\left(H_1|l,\Theta_0^l\right)}\right]$$

are not the ratios of any (marginal, supremum…) likelihoods. It follows that the Bayes-Poincaré factors are of a different nature than the classical, well-known likelihood ratio-based Bayes factors. In particular, the Bayes-Poincaré factors are not exponential.

Corollary: as an example, the solution to the formal, relative, univariate, continuous, 2-sample historical Behrens-Fisher problem with Jeffreys' joint proper-improper prior

$$p\left(\mu_i,\sigma_i|L,U\right)=p\left(\mu_i|L,U\right)p\left(\sigma_i\right)\propto\left(U-L\right)^{-1}\sigma_i^{-1}$$

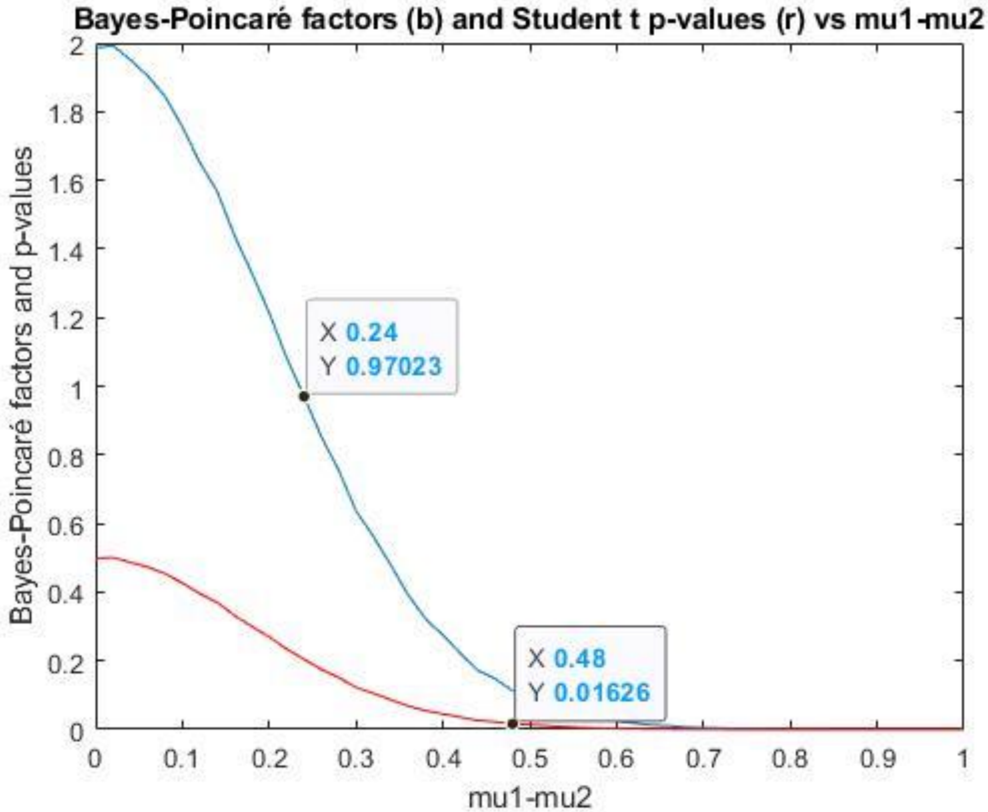
over $\left[L,U\right]\times\mathbb{R}^{+*}$, $L<U$, $i=1,2$, is given by

$$B_{01}|L,U=\left(U-L\right)\frac{\int_L^U\left[\sum_{i=1}^{n_1}\left(x_1^i-\mu\right)^2\right]^{-\frac{n_1}{2}}\left[\sum_{i=1}^{n_2}\left(x_2^i-\mu\right)^2\right]^{-\frac{n_2}{2}}\mathrm{d}\mu}{\int_L^U\left[\sum_{i=1}^{n_1}\left(x_1^i-\mu\right)^2\right]^{-\frac{n_1}{2}}\mathrm{d}\mu\int_L^U\left[\sum_{i=1}^{n_2}\left(x_2^i-\mu\right)^2\right]^{-\frac{n_2}{2}}\mathrm{d}\mu}$$

Proof: see **Appendix 1**

Simulation results

Finally, we present some early simulation results, just to check that the Bayes-Poincaré factors work as expected. We compare them to the p-values of the standard two-sample unpaired/independent Student t test for unequal variances with Satterthwaite's approximation. See **Appendix 2** for the corresponding Matlab code and the parameters of the simulations. Those results shall be completed with the standard Bayes factors for likelihoods M_0 and M_1.



It appears that the Bayes-Poincaré factors are actually meaningful: they decrease as $|\mu_1 - \mu_2|$ increases. However, in this particular example, they yield completely different decisions from the decisions based on the 2-sample unpaired/independent Student t test p-values. For instance, if the decision rule for the Bayes-Poincaré factor is simply

$$\begin{cases} \text{Do not reject } H_0 \text{ if } B_{01} \geq 1 \\ \text{Reject } H_0 \text{ if } B_{01} < 1 \end{cases}$$

and the decision rule for the p-values of the Student t test is

$$\begin{cases} \text{Do not reject } H_0 \text{ if } p \geq \alpha \\ \text{Reject } H_0 \text{ if } p < \alpha \end{cases}$$

for significance level $\alpha \in [0,1]$, then both decision procedures are strongly inconsistent for standard values of $\alpha \ll 0.1$ over a large range of values of $|\mu_1 - \mu_2|$. Conversely, both procedures can match only for unusually large values of $\alpha \simeq 0.2$

Moreover, it is quite clear that, depending on $U - L$, the Bayes-Poincaré factors can be greater than one or smaller than one for almost all values of $|\mu_1 - \mu_2|$! This issue comes from the Bayes-Laplace uniform prior whose normalization constant $U - L$ is unbounded over $\mathbb{R}$ and should disappear for location priors having bounded normalization constants. This shows that priors that are standard in estimation theory are not suitable for testing point null, continuous hypotheses because . It will take some time to tame the Bayes-Poincaré factors and their suitable priors.

Conclusion

The likelihood or model M_0 under the null hypothesis H_0, within the classical, standard likelihoodist and Bayes factor frameworks, succumbs to the *principle of the identity of equality and identity*. This was our main stimulus for developing a new and different solution to the Behrens-Fisher problem that is completely internal to Bayesian probability theory. *A posteriori*, upon inspection, those classical likelihoodist and Bayesian solutions suffer many fatal defects and criticisms, including

- Two new and fictitious models/likelihoods M_0 and M_1 for the pooled data (x_1, x_2) are introduced under both hypotheses H_0 and H_1 on top of both original models $\mathcal{L}(\theta_1, \eta_1)$ and $\mathcal{L}(\theta_2, \eta_2)$, in violation of Ockam's razor;
- Model M_0 requires a call to the *principle of the identity of equality and identity*, which is external to probability theory and false according to Henri Poincaré. This is not because two parameters have the same numerical values that they are identical;
- Conversely, under model M_1, this is not because there are two different parameters that their numerical values are necessarily different. They are different almost surely with probability $p=1$ if the parameters are continuous and different with probability $p<1$ if they are discrete;
- It follows that M_1 is not the logical negation of M_0, even in the continuous case, in contradiction with the definition of the original hypotheses H_0 and H_1;
- The prior probabilities for models M_0 and M_1 are assigned, quite arbitrarily, and decorrelated from the prior probabilities for the hypotheses H_0 and H_1 that must be computed from the prior probability distributions of the parameters of the original models and are equal to 0 respectively 1 for continuous parameters;
- For a continuous parameter of interest with continuous marginal prior probability distributions under both experiments, the prior and posterior probabilities for the null hypothesis H_0 are equal to zero. It follows that the Bayes factor is undefined. Therefore, the solution cannot rely on a Bayes factor;
- The classical solution remains the same, regardless of whether the parameter of interest is discrete or continuous, while the situation is completely different since the Bayes factor is well defined in the discrete case and undefined in the continuous one.

The first step towards our new, more rigorous solution was to distinguish discrete problems from continuous ones. The absolute solutions to the discrete problems are straightforward but become degenerate, trivial and totally useless for continuous ones for which the Bayes factors (defined as prior-to-posterior odds ratios, not likelihood ratios) are undefined because the point null hypothesis is Lebesgue-negligible. This measure-theoretic issue was easily addressed, in a second step, by approximating the continuous problems by sequences of discrete ones. This yields the introduction of relative solutions to continuous problems and Bayes-Poincaré factors, which are generically well defined thanks to the Henstock-Kurzweil integral that is as powerful as the Lebesgue one but still relies on Riemann sums that are essential in the present solution. This "discretization" technique is of general interest for all point null hypothesis tests and might finally provide us with the solutions to other important and long-standing problems, such as the Jeffreys-Lindley paradox [18]. However, it will take some time to tame the Bayes-Poincaré factors and their suitable priors because they are very much different from the traditional Bayes factors. While the present work is fairly elementary, it nevertheless goes deep into the mathematical foundations of probability theory because it appears that the standard Borel-Lebesgue-Kolmogorov measure-theoretic setting is not always the most natural and suitable one.

References

[1] Wikipedia, *List of unsolved problems in statistics*, https://en.wikipedia.org/wiki/List_of_unsolved_problems_in_statistics

[2] Bretthorst G. L., *On the difference in means*, in Physics and Probability, Essays in honor of Edwin T. Jaynes, pp. 177-194, W. T. Gandy and P. W. Milonni (eds), Cambridge University Press, England (1993)

[3] Ghosh M., Kim Y.-H., *The Behrens-Fisher problem revisited: a Bayes-frequentist synthesis*, The Canadian Journal of Statistics, Vol. 29, No.1, 2001

[4] Paul S., Wang Y.-G., Ullah I., *A review of the Behrens-Fisher problem and some of its analogs: does the same size fit all?*, Revstat statistical journal, 2018

[5] Kim S.-H., Cohen A. S., *On the Behrens-Fisher problem: a review*, paper presented at the Annual Meeting of the Psychometric Society (Minneapolis MN, June 1995)

[6] Barbieri A., Marin J.-M., Florin K., *A fully objective Bayesian approach for the Behrens-Fisher problem using historical studies*, arXiv:1611.06873v1, November 2016

[7] Ly I., *Identité et égalité: le criticisme de Poincaré*, Philosophiques, volume 31, numéro 1, printemps 2004, p. 179-212, https://www.erudit.org/fr/revues/philoso/2004-v31-n1-philoso741/008939ar/

[8] Wang M., *A simple two-sample Bayesian t-test for hypothesis testing*, the American Statistician, September 2015

[9] Gönen M., Johnson W. O., Lu Y., Westfall P. H, *The Bayesian two-sample test*, the American Statistician, 59(3), 252-257

[10] Wikipedia, *Behrens-Fisher distribution*, https://en.wikipedia.org/wiki/Behrens–Fisher_distribution

[11] Wikipedia, *Welch's t-test*, https://en.wikipedia.org/wiki/Welch%27s_t-test

[12] Lehmann, E. L., *Nonparametrics: Statistical Methods Based on Ranks*, McGraw-Hill ISBN 0-07-037073-7

[13] Wikipedia, *Henstock–Kurzweil integral*, https://en.wikipedia.org/wiki/Henstock%E2%80%93Kurzweil_integral

[14] Wikipedia, *The Behrens-Fisher problem,* https://en.wikipedia.org/wiki/Behrens%E2%80%93Fisher_problem#cite_ref-1

[15] Pfanzagl J., *On the Behrens-Fisher problem*, Biometrika 61, 1, p. 39, 1974

[16] Fisher R. A., *The fiducial argument in statistical inference*, Ann. Eugen. 6, pp. 391-422, 1939

[17] Jeffreys H., *Note on the Behrens-Fisher formula*, Annals of Human Genetics 10(1):48-51, 1940

[18] Berger J. O., *Could Fisher, Jeffreys and Neyman Have Agreed on Testing?*, Statistical Science, vol. 18, Issue 1, pp. 1-32, 2003

Appendix 1: Calculations for the formal, relative, univariate, continuous, 2-sample historical Behrens-Fisher problem with Jeffreys' joint proper-improper prior for the Gaussian probability distribution

Let

$$p(\mu_i,\sigma_i)=p(\mu_i)\,p(\sigma_i)\propto\sigma_i^{-1},\, i=1,2$$

be the classical Jeffreys' improper prior over $\mathbb{R}\times\mathbb{R}^{+*}$ for the Gaussian distribution.

In order to keep full control, we start with proper priors with compact support

$$p(\mu_i,\sigma_i|N,a,b)=p(\mu_i|N)\,p(\sigma_i|a,b)=(2N)^{-1}\frac{\sigma_i^{-1}}{\log(b)-\log(a)}$$

over $[-N,N]\times[a,b]$, $0<N$, $0<a<b$, $i=1,2$.

We introduce two identical sequences of discrete uniform random variables $(\mu_i^l)_{l\in\mathbb{N}^*}$, $i=1,2$ defined on a partition of $[-N,N]$ such as

$$\Omega^l=\{-N,-N+\Delta\mu,-N+2\Delta\mu,\ldots,N\},\,\Delta\mu=\frac{N}{l}$$

of cardinal $|\Omega^l|=2l+1$. The prior probability for the null hypothesis H_0 and the discrete parameters μ_1^l and μ_2^l is

$$p(H_0|l,N)=\sum_{\Omega^l}p(\mu^l)^2=\sum_{\Omega^l}(2l+1)^{-2}=(2l+1)(2l+1)^{-2}=(2l+1)^{-1}$$

But it is more convenient to write it like this

$$p(H_0|l,N)=\frac{\sum_{\Omega_l}1\times1}{\sum_{\Omega_l}1\sum_{\Omega_l}1}=\Delta\mu\frac{\Delta\mu\sum_{\Omega_l}1\times1}{\Delta\mu\sum_{\Omega_l}1\Delta\mu\sum_{\Omega_l}1}\underset{\Delta\mu\to0^+}{\sim}\Delta\mu\frac{\int_{-N}^{N}\mathrm{d}\mu}{\int_{-N}^{N}\mathrm{d}\mu\int_{-N}^{N}\mathrm{d}\mu}=\frac{N}{l}\frac{2N}{(2N)^2}=\frac{1}{2l}$$

Dropping index i for clarity, both joint posteriors write

$$p(\mu^l,\sigma|x,l,N,a,b)=\frac{p(\mu^l|l,N)\,p(\sigma|a,b)\,p(x|\mu^l,l,N,\sigma,a,b)}{\sum_{\Omega^l}p(\mu^{l'}|l,N)\int_a^b p(\sigma|a,b)\,p(x|\mu^{l'},l,N,\sigma,a,b)\mathrm{d}\sigma}=\frac{p(\sigma|a,b)\,p(x|\mu^l,l,N,\sigma,a,b)}{\sum_{\Omega^l}\int_a^b p(\sigma|a,b)\,p(x|\mu^{l'},l,N,\sigma,a,b)\mathrm{d}\sigma}$$

with

$$p(x|\mu^l,l,N,\sigma,a,b)=\left(\sqrt{2\pi}\right)^{-m}\sigma^{-m}e^{-\frac{\sigma^{-2}}{2}\sum_{i=1}^{m}(x^i-\mu^l)^2}$$

We need to evaluate the integral

$$\int_a^b p(\sigma|a,b)\,p(x|\mu^l,l,N,\sigma,a,b)\,\mathrm{d}\sigma = \frac{\left(\sqrt{2\pi}\right)^{-m}}{\log(b)-\log(a)}\int_a^b \sigma^{-m-1} e^{-\frac{\sigma^{-2}}{2}\sum_{i=1}^{m}\left(x^i-\mu^l\right)^2}\mathrm{d}\sigma$$

Let

$$A(\mu^l) = \frac{1}{2}\sum_{i=1}^{m}\left(x^i-\mu^l\right)^2,\ y = A(\mu^l)\sigma^{-2} \Leftrightarrow \sigma = A(\mu^l)^{\frac{1}{2}} y^{-\frac{1}{2}},\ \mathrm{d}\sigma = -\frac{1}{2}A(\mu^l)^{\frac{1}{2}} y^{-\frac{3}{2}}\mathrm{d}y$$

Then

$$\int_a^b \sigma^{-m-1} e^{-A(\mu^l)\sigma^{-2}}\mathrm{d}\sigma = -\frac{A(\mu^l)^{\frac{1}{2}}}{2}\int_{A(\mu^l)a^{-2}}^{A(\mu^l)b^{-2}}\left(A(\mu^l)^{\frac{1}{2}} y^{-\frac{1}{2}}\right)^{-m-1} y^{-\frac{3}{2}} e^{-y}\mathrm{d}y =$$

$

$$\frac{A(\mu^l)^{-\frac{m}{2}}}{2}\int_{A(\mu^l)b^{-2}}^{A(\mu^l)a^{-2}} y^{\frac{m}{2}-1} e^{-y}\mathrm{d}y = \frac{A(\mu^l)^{-\frac{m}{2}}}{2}\left[\Gamma\left(\frac{m}{2},A(\mu^l)b^{-2}\right)-\Gamma\left(\frac{m}{2},A(\mu^l)a^{-2}\right)\right]$$

It follows that

$$p(\mu^l|x,l,N,a,b) = \frac{\int_a^b p(\sigma|a,b)\,p(x|\mu^l,l,N,\sigma,a,b)\,\mathrm{d}\sigma}{\sum_{\Omega^l}\int_a^b p(\sigma|a,b)\,p(x|\mu^{l'},l,N,\sigma,a,b)\,\mathrm{d}\sigma} = \frac{A(\mu^l)^{-\frac{m}{2}}\left[\Gamma\left(\frac{m}{2},A(\mu^l)b^{-2}\right)-\Gamma\left(\frac{m}{2},A(\mu^l)a^{-2}\right)\right]}{\sum_{\Omega^l}A(\mu^{l'})^{-\frac{m}{2}}\left[\Gamma\left(\frac{m}{2},A(\mu^{l'})b^{-2}\right)-\Gamma\left(\frac{m}{2},A(\mu^{l'})a^{-2}\right)\right]}$$

Now that the normalization constant $\log(b)-\log(a)$ has cancelled out, we can take the limits $a\to 0^+$ and $b\to+\infty$ to get, if and only if $A(\mu_l)>0$

$$p(\mu^l|x,l,N) = \frac{A(\mu^l)^{-\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}{\sum_{\Omega_l}A(\mu^{l'})^{-\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} = \frac{A(\mu^l)^{-\frac{m}{2}}}{\sum_{\Omega_l}A(\mu^{l'})^{-\frac{m}{2}}}$$

Therefore, the null hypothesis H_0 has posterior probability

$$p(H_0|x_1,x_2,l,N) = \sum_{\Omega^l} p(\mu_1^l=\mu^l|x_1,l,N)\,p(\mu_2^l=\mu^l|x_2,l,N) = \frac{\sum_{\Omega^l}\mathrm{SSE1}(\mu^l)^{-\frac{m}{2}}\mathrm{SSE2}(\mu^l)^{-\frac{n}{2}}}{\sum_{\Omega^l}\mathrm{SSE1}(\mu^l)^{-\frac{m}{2}}\sum_{\Omega^l}\mathrm{SSE2}(\mu^l)^{-\frac{n}{2}}}$$

if $\mathrm{SSE1}(\mu^l)=\sum_{i=1}^{m}\left(x_1^i-\mu^l\right)^2$ and $\mathrm{SSE2}(\mu^l)=\sum_{j=1}^{n}\left(x_2^j-\mu^l\right)^2$. As expected, the ratio

$$\frac{p(H_0|x_1,x_2,l,N)}{p(H_0|l,N)}$$

now has a well-defined limit when $l \to +\infty$, equivalently $\Delta\mu \to 0^+$

$$\frac{p\left(H_0 \middle| x_1, x_2, l, N\right)}{p\left(H_0 \middle| l, N\right)} = \Delta\mu \frac{\Delta\mu \sum_{\Omega^l} \mathrm{SSE1}\left(\mu^l\right)^{-\frac{m}{2}} \mathrm{SSE2}\left(\mu^l\right)^{-\frac{n}{2}}}{\Delta\mu \sum_{\Omega^l} \mathrm{SSE1}\left(\mu^l\right)^{-\frac{m}{2}} \Delta\mu \sum_{\Omega^l} \mathrm{SSE2}\left(\mu^l\right)^{-\frac{n}{2}}} / \Delta\mu \frac{\Delta\mu \sum_{\Omega_l} 1 \times 1}{\Delta\mu \sum_{\Omega_l} 1 \Delta\mu \sum_{\Omega_l} 1}$$

$$\underset{\Delta\mu \to 0^+}{\to} \frac{p\left(H_0 \middle| x_1, x_2, N\right)}{p\left(H_0 \middle| N\right)} = 2N \frac{\int_{-N}^{N} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu}{\int_{-N}^{N} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{d}\mu \int_{-N}^{N} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu} \triangleq B_{01} \middle| N$$

because all functions are Riemann-integrable.

For $m > 2$, $n > 2$ and non pathological data, the improper integrals converge

$$\lim_{N \to +\infty} \int_{-N}^{N} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{d}\mu = \int_{-\infty}^{+\infty} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{d}\mu < +\infty$$

$$\lim_{N \to +\infty} \int_{-N}^{N} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu = \int_{-\infty}^{+\infty} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu < +\infty$$

$$\lim_{N \to +\infty} \int_{-N}^{N} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu = \int_{-\infty}^{+\infty} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu < +\infty$$

It follows that we have the undesirable but perfectly normal result

$$B_{01} \middle| N \underset{N \to +\infty}{\sim} 2N \frac{\int_{-\infty}^{+\infty} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu}{\int_{-\infty}^{+\infty} \mathrm{SSE1}(\mu)^{-\frac{m}{2}} \mathrm{d}\mu \int_{-\infty}^{+\infty} \mathrm{SSE2}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu} \underset{N \to +\infty}{\to} B_{01} = +\infty$$

This is not a defect of the present method but of the uniform prior, due to the fact that

$$\lim_{N \to +\infty} \int_{-N}^{N} \mathrm{d}\mu = +\infty \qquad \text{while} \qquad \lim_{N \to +\infty} \int_{-N}^{N} \mathrm{SSE}(\mu)^{-\frac{n}{2}} \mathrm{d}\mu < +\infty \text{ for } n > 2$$

This is a very intuitive result, the larger N, the smaller $p\left(H_0 \middle| N\right)$ but almost not $p\left(H_0 \middle| x_1, x_2, N\right)$ because the posterior distributions concentrate their mass around the sample means.

Hence, this issue should disappear for any location prior whose normalization constant remains bounded over $\mathbb{R}$ such as a Gaussian prior $\mathcal{N}\left(0, \tau^2\right)$.

Appendix 2: Matlab simulation code

```
clear all
close all
clc
format long

% Standard deviations of both Gaussian distributions
sigma1 = 1;
sigma2 = 1;

% Sample sizes
N1 = 100;
N2 = 100;
N12 = N1/2;
N22 = N1/2;

L = 0;
U = 1;

delta_delta_mu = 0.01;
% Values of mu1 - mu2
delta_mu = 0:delta_delta_mu:0.5; delta_mu2 = 2 * delta_mu;
nb_delta_mu = length(delta_mu);

mu1 = zeros(1,nb_delta_mu);
mu2 = zeros(1,nb_delta_mu);

% Bayes-Poincaré factors
BPF = zeros(1,nb_delta_mu);
% p-values for the 2-sample independent/unpaired Student t test with unequal variances
ans Satterthwaite's approximation
pvalues = zeros(1,nb_delta_mu);

delta_mus = 0.001;
% Delta-fine partition for the Riemann sums/rectangle method for numerical
% evaluation of the integrals
mus = L:delta_mus:U;
mus2 = mus.^2;
nb_mus = length(mus);

% Number of simulations
nb_simu = 1000;

for k = 1 : nb_delta_mu

    k

    % Mathematical expectations for both Gaussian distributions
    mu1(k) = 0.5 + delta_mu(k);
    mu2(k) = 0.5 - delta_mu(k);

    for simu = 1 : nb_simu

        % Samples
        x1 = normrnd(mu1(k),sigma1,1,N1);
        x2 = normrnd(mu2(k),sigma2,1,N2);
```

```
        % 2-sample independent/unpaires Student t test with unequal variances ans
Satterthwaite's approximation
        %[h,p,ci,stats] = ttest2(x1,x2,'Vartype','unequal');
        [h,p] = ttest2(x1,x2,'Vartype','unequal');

        x1m = sum(x1);
        x2m = sum(x2);

        x1m2 = sum(x1.^2);
        x2m2 = sum(x2.^2);

        SSE1 = x1m2 - 2*x1m * mus + N1 * mus2;
        SSE2 = x2m2 - 2*x2m * mus + N2 * mus2;

        % Mean of the Bayes-Poincaré factors over the simulations
        BPF(k) = BPF(k) + sum((SSE1).^(-N12).*(SSE2).^(-N22)) / (sum((SSE1).^(-N12)) *
sum((SSE2).^(-N22)));

        % Mean of the p-values over the simulations
        pvalues(k) = pvalues(k) + p;

    end

    BPF(k) = BPF(k) * (U - L) / delta_mus / nb_simu;
    pvalues(k) = pvalues(k) / nb_simu;

end

figure(1)
plot(delta_mu2,BPF)
hold
plot(delta_mu2,pvalues,'r')
xlabel('mu1-mu2')
ylabel('Bayes-Poincaré factors and p-values')
title('Bayes-Poincaré factors (b) and Student t p-values (r) vs mu1-mu2')
```

End of the document