

On Analysis of the Behrens-Fisher Problem Based on Bayesian Evidence

Nengak Emmanuel Goltong^{id}, Sani Ibrahim Doguwa

Department of Statistics, Ahmadu Bello University, Zaria, Nigeria

Email: spomariae@gmail.com, sidoguwa@gmail.com

How to cite this paper: Goltong, N.E. and Doguwa, S.I. (2019) On Analysis of the Behrens-Fisher Problem Based on Bayesian Evidence. *Open Journal of Statistics*, 9, 1-14.

<https://doi.org/10.4236/ojs.2019.91001>

Received: December 17, 2018

Accepted: January 18, 2019

Published: January 21, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper we have demonstrated the ability of the new Bayesian measure of evidence of Yin (2012, *Computational Statistics*, 27: 237-249) to solve both the Behrens-Fisher problem and Lindley's paradox. We have provided a general proof that for any prior which yields a linear combination of two independent t random variables as posterior distribution of the difference of means, the new Bayesian measure of evidence given that prior will solve Lindley's paradox thereby serving as a general proof for the works of Yin and Li (2014, *Journal of Applied Mathematics*, 2014(978691)) and Goltong and Doguwa (2018, *Open Journal of Statistics*, 8: 902-914). Using the Pareto prior as an example, we have shown by the use of simulation results that the new Bayesian measure of evidence solves Lindley's paradox.

Keywords

Behrens-Fisher Problem, Lindley's Paradox, Metropolis-Hastings Algorithm, Pareto Prior, t Distribution

1. Introduction

The Behrens-Fisher problem simply refers to a hypothesis testing problem about the difference of two means based on two random samples that are assumed to be independent and normally distributed without assuming that the two variances are equal.

[1] noted that conclusions from a hypothesis test based on a Bayesian framework could contradict that from a frequentist framework. He showed that as $n \rightarrow \infty$, while a frequentist will reject H_0 , a Bayesian using the posterior probability under H_0 will fail to reject H_0 . Even though [2] argued that these two schools of statistical inference need not agree since the two evaluate different objects using different measures among other points, he also noted that it is not meaningful if a procedure would always conclude by picking the null whatever be the value of the test statistic. This points to the fact that even if one may

not be interested in having the Bayesian and Frequentist approaches reconciled, there is need to look out for test procedures that will not always favour the null hypothesis. More discussions and arguments concerning Lindley's paradox, can be found in [3], [4] and [2]. [5] developed a Bayesian hypothesis testing framework that solves Lindley's paradox in a one sample case. Using Jeffreys' independent prior, [6] extended the methodology of [5] to solve the Behrens-Fisher Problem so as to avoid Lindleys' Paradox. Also, [7] made an attempt to solve the Paradox in [1] by assigning Gamma priors to the precisions using the methodology of [5]. It is noteworthy that the posterior distribution of the difference of means derived by [6] and that derived by [7] are similar in the sense that both are a linear combination of two t distributed random variables. Also these two posterior distributions differ in the sense that, the coefficients of these linear combinations are not the same and also, the distributions of the t random variables differ in their degrees of freedom. In this paper, we propose to give a general result of the methodology of [5] in solving simultaneously, the Behrens-Fisher problem and Lindley's paradox when any given prior is assigned to the unknown variances which yields a posterior distribution that is a linear combination of two independent t random variables.

2. Literature Review

In an attempt to solve the Behrens-Fisher Problem, [8] derived generalized pivotal quantities and generalized p-values for hypothesis testing in the presence of nuisance parameters. The generalized p-value is given by

$$p(x) = Pr \left[F_{1, n_1+n_2-2} \geq (\bar{x}_1 - \bar{x}_2)^2 (n_1 + n_2 - 2) \left(\frac{\tilde{s}_1^2}{B} + \frac{\tilde{s}_2^2}{1-B} \right)^{-1} \right] \tag{1}$$

where $Pr[\cdot]$ means the probability of $[\cdot]$, s_i^2 is the variance of sample i of size n_i , $\tilde{s}_i^2 = (1 - n_i^{-1})s_i^2$, $i = 1, 2$ and $B \sim Beta((n_1 - 1)/2, (n_2 - 1)/2)$. [9] derived a general recipe for constructing generalized pivotal quantities and generalized confidence intervals using a systematic approach which was not provided by [8]. [10] used a generalized p-value approach to derive tests and confidence intervals for the common mean based on several well known estimates of the common mean. [11] used Stein's two-stage sampling scheme to solve the Behrens-Fisher problem so as to control the Type-II error and the length of the confidence interval given a specified Type-I error. [12] conducted a comparative study of the different methods of solving the Behrens-Fisher problem.

The idea of using informative priors is advocated for by [13], who disagrees that diffuseness of a prior reflects ignorance about the distribution of such a parameter. [14], [15], and [16] investigated the reconcilability of the Bayesian and frequentist approaches to hypothesis testing. [17] developed the posterior predictive p-value which takes care of nuisance parameters in hypothesis testing, given as

$$ppp(x) = P \left\{ F_{1, n_1+n_2} \geq \frac{(\bar{x}_1 - \bar{x}_2)^2 (n_1 + n_2)}{[\tilde{s}_1^2 + (\bar{x}_1 - \mu)^2] B_{n_1, n_2}^{-1} + [\tilde{s}_2^2 + (\bar{x}_2 - \mu)^2] (1 - B_{n_1, n_2})^{-1}} \right\} \tag{2}$$

where $P\{\cdot\}$ is the probability of $\{\cdot\}$, s_i^2 is the variance of sample i of size n_i , $\tilde{s}_i^2 = (1 - n_i^{-1})s_i^2$, $i = 1, 2$ and $B_{n_1, n_2} \sim \text{Beta}(n_1/2, n_2/2)$.

[5] developed a Bayesian testing procedure for testing a precise null hypothesis in the one sample case that avoids the dichotomy of the parameter space, thereby solving Lindley's paradox. This new Bayesian measure of evidence is given by

$$P^B(x) = P(|\theta - E(\theta|x)| \geq |\theta_0 - E(\theta|x)| | x) \tag{3}$$

where $P(\cdot)$ is the probability of (\cdot) , $E(\theta|x)$ is the posterior expectation of θ under the prior $\pi(\theta)$, where a smaller value of $P^B(x)$ corresponds to a bigger distance between θ_0 and the true θ and therefore, suggesting stronger evidence against the null hypothesis H_0 . [18] showed both analytically and by simulation results that the infimum of the posterior probability of H_0 , $Pr(H_0|x)$ under the conjugate class of priors is strictly less than the the p-value and the Generalized p-value in testing a normal mean and in testing two normal means respectively.

In order to solve the Behrens-Fisher problem, [6] extended the methodology of [5] assigning Jeffreys' objective prior to the unknown variances given as

$$\pi(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto \frac{1}{\sigma_1^2 \sigma_2^2} \tag{4}$$

and obtained the posterior distribution of the difference of means, $\theta|x$, as

$$\theta|x \sim \bar{x}_1 - \bar{x}_2 - \left(\frac{S_1 T_{n_1-1}}{\sqrt{n_1}} - \frac{S_2 T_{n_2-1}}{\sqrt{n_2}} \right) \tag{5}$$

thereby obtaining under Jeffreys independent prior, the Bayesian measure of evidence as

$$P_J^{BF}(x) = P \left(\left| \frac{S_1 T_{n_1-1}}{\sqrt{n_1}} - \frac{S_2 T_{n_2-1}}{\sqrt{n_2}} \right| \geq |\bar{x}_1 - \bar{x}_2| \right) \tag{6}$$

where T_{n_i-1} is a t random variable with $n_i - 1$ degrees of freedom. By the use of simulation results, this approach was shown to solve the two problems and also, it was shown mathematically to yield credible intervals that actually possess $1 - \alpha$ coverage probability.

[7] also extended the methodology of [5] by assigning Gamma priors to the precisions and showed mathematically that the paradox in [1] is avoided while solving the Behrens-Fisher problem. They also showed that results obtained by assigning Gamma priors to the precisions could coincide with results obtained by assigning Jeffrey's independent prior to the variances as is done in [6]. [19] proved that Lindley's Paradox could be reversed given that the probability density is persistently unbounded or where sufficient regularity of likelihood or prior is absent.

3. Main Result

In this section, we propose a more general result about the ability of the Bayesian measure of evidence of [5] to solve Lindley's paradox given any prior distribution. Before proceeding to this general result, let us consider a definition that is crucial to its postulation.

Definition 1. A rational function, B , is said to be asymptotically independent of n if B can be expressed in the following manner:

$$B = \frac{P(n)}{Q(n)} B^* \tag{7}$$

where $P(n)$ and $Q(n)$ are polynomials in n , and $Q(n)$ is a polynomial of degree at least equal to that of $P(n)$, and B^* is completely independent of n .

Lemma 1. Let the posterior distribution of θ given any prior distribution, say π , be given by

$$\theta|x \sim \bar{x}_1 - \bar{x}_2 - (A_1T_1 - A_2T_2) \tag{8}$$

where the coefficients A_1 and A_2 are functions of n_1^{-p} and n_2^{-q} respectively, such that these coefficients are of the form $A_1 = B_1n_1^{-p}$ and $A_2 = B_2n_2^{-q}$ for any B_1 either independent or asymptotically independent of n_1 , any B_2 either independent or asymptotically independent of n_2 and for any $p, q > 0$. Then, where T_1 and T_2 are independent t random variables, if the degrees of freedom of T_1 and T_2 are linear functions of n_1 and n_2 respectively, the Bayesian measure of evidence based on (8) solves Lindley's paradox, that is

$$P_\pi^{BF}(x) \rightarrow 0 \quad \text{as } (n_1, n_2) \rightarrow (\infty, \infty)$$

Proof. Recall that the methodology of [5] yields the Bayesian measure of evidence as

$$P^B(x) = P(|\theta - E(\theta|x)| \geq |\theta_0 - E(\theta|x)| | x) \tag{9}$$

Since the degrees of freedom of the two random variables are linear forms of the corresponding sample sizes, let $k_1n_1 + c_1$ be the degrees of freedom of T_1 , where k_1 is any constant and c_1 is independent of n_1 . Also, let $k_2n_2 + c_2$ be the degrees of freedom of T_2 , where k_2 is any constant and c_2 is independent of n_2 .

Then where B_1 and B_2 are completely independent of n_1 and n_2 respectively, we have that

$$\begin{aligned} \text{Var}(\theta|x) &= \frac{B_1^2}{n_1^{2p}} \frac{(k_1n_1 + c_1)}{(k_1n_1 + c_1 - 2)} + \frac{B_2^2}{n_2^{2q}} \frac{(k_2n_2 + c_2)}{(k_2n_2 + c_2 - 2)} \\ &= \frac{B_1^2 n_2^{2q} (k_1n_1 + c_1)(k_2n_2 + c_2^*) + B_2^2 n_1^{2p} (k_1n_1 + c_1^*)(k_2n_2 + c_2)}{n_1^{2p} n_2^{2q} (k_1n_1 + c_1^*)(k_2n_2 + c_2^*)} \end{aligned} \tag{10}$$

where $c_1^* = c_1 - 2$ and $c_2^* = c_2 - 2$. Note that (9) can be expressed as

$$P_\pi^{BF}(x) = 2 [1 - P(\chi_1^2 < Z_0^2)] \tag{11}$$

and we also have that

$$Z_0^2 = \frac{[\theta_0 - E(\theta|x)]^2}{\text{Var}(\theta|x)} \tag{12}$$

Now, let $a = [\theta_0 - E(\theta|x)]^2$, independent of n_1 and n_2 , then we have that

$$Z_0^2 = \frac{an_1^{2p} n_2^{2q} (k_1n_1 + c_1^*)(k_2n_2 + c_2^*)}{B_1^2 n_2^{2q} (k_1n_1 + c_1)(k_2n_2 + c_2^*) + B_2^2 n_1^{2p} (k_1n_1 + c_1^*)(k_2n_2 + c_2)} \tag{13}$$

$$\begin{aligned}
 & \lim_{n_1 \rightarrow \infty} Z_0^2 \\
 &= \lim_{n_1 \rightarrow \infty} \left\{ \frac{an_1^{2p}n_2^{2q}(k_1n_1 + c_1^*)(k_2n_2 + c_2^*)}{B_1^2n_2^{2q}(k_1n_1 + c_1)(k_2n_2 + c_2^*) + B_2^2n_1^{2p}(k_1n_1 + c_1^*)(k_2n_2 + c_2)} \right\} \\
 &= \lim_{n_1 \rightarrow \infty} \left\{ \frac{an_2^{2q} \left(k_1 + \frac{c_1^*}{n_1}\right) (k_2n_2 + c_2^*)}{B_1^2n_1^{-2p}n_2^{2q} \left(k_1 + \frac{c_1}{n_1}\right) (k_2n_2 + c_2^*) + B_2^2 \left(k_1 + \frac{c_1^*}{n_1}\right) (k_2n_2 + c_2)} \right\} \\
 &= \frac{an_2^{2q}(k_2n_2 + c_2^*)}{B_2^2(k_2n_2 + c_2)}
 \end{aligned} \tag{14}$$

Then we have that

$$\lim_{n_2 \rightarrow \infty} \left[\lim_{n_1 \rightarrow \infty} Z_0^2 \right] = \lim_{n_2 \rightarrow \infty} \left[\frac{an_2^{2q} \left(k_2 + \frac{c_2^*}{n_2}\right)}{B_2^2 \left(k_2 + \frac{c_2}{n_2}\right)} \right] = \infty \tag{15}$$

Similarly, we have that

$$\lim_{n_2 \rightarrow \infty} Z_0^2 = \frac{an_1^{2p}(k_1n_1 + c_1^*)}{B_1^2(k_1n_1 + c_1)} \tag{16}$$

Then we have that

$$\lim_{n_1 \rightarrow \infty} \left[\lim_{n_2 \rightarrow \infty} Z_0^2 \right] = \lim_{n_1 \rightarrow \infty} \left[\frac{an_1^{2p} \left(k_1 + \frac{c_1^*}{n_1}\right)}{B_1^2 \left(k_1 + \frac{c_1}{n_1}\right)} \right] = \infty \tag{17}$$

Since we have from (15) and (17) that

$$\lim_{n_2 \rightarrow \infty} \left[\lim_{n_1 \rightarrow \infty} Z_0^2 \right] = \lim_{n_1 \rightarrow \infty} \left[\lim_{n_2 \rightarrow \infty} Z_0^2 \right] = \infty \tag{18}$$

It implies that

$$\lim_{(n_1, n_2) \rightarrow (\infty, \infty)} Z_0^2 = \infty, \implies \lim_{(n_1, n_2) \rightarrow (\infty, \infty)} P_\pi^{BF}(x) = 0 \tag{19}$$

To show that this result still holds where B_1 and B_2 are only asymptotically independent of n_1 and n_2 respectively, it suffices to show that under mild assumptions, B_1 and B_2 can safely be assumed to be completely independent of n_1 and n_2 respectively. This result will be shown in two steps. First, let us consider the case where $Q(n)$ and $P(n)$ are polynomials of equal degree and recall that

$$\begin{aligned}
 B_i &= B_i^* \frac{P(n_i)}{Q(n_i)}, \quad \forall i = 1, 2. \\
 &= B_i^* \frac{P^*(n_i) + \alpha}{Q^*(n_i) + \beta}, \quad \forall i = 1, 2.
 \end{aligned} \tag{20}$$

where $P^*(n_i)$ is that part of the polynomial $P(n_i)$ consisting of terms that contain n_i only excluding the constant term, $Q^*(n_i)$ is that part of the polynomial $Q(n_i)$ consisting of terms that contain n_i only excluding the constant term, while α and β are the corresponding constant terms.

Without any loss of generality, assume each of the coefficients in $P^*(n_i)$ and $Q^*(n_i)$ to be equal to 1, and $\alpha = \beta = 0$. Then it is obvious that $P^*(n_i)$ and $Q^*(n_i)$ are equal and hence

$$B_i = B_i^*, \quad \forall i = 1, 2. \tag{21}$$

Thus, B_1 and B_2 are completely independent of n_1 and n_2 respectively.

On a general note, relaxing the assumptions on $P(n_i)$ and $Q(n_i)$ except the assumption that they are polynomials of the same degree, then it is easily seen that

$$\begin{aligned} \lim_{n_i \rightarrow \infty} B_i &= B_i^* \lim_{n_i \rightarrow \infty} \frac{P(n_i)}{Q(n_i)}, \quad \forall i = 1, 2. \\ &= B_i^* \kappa_i, \quad \forall i = 1, 2. \end{aligned} \tag{22}$$

where κ_i is a constant independent of n_i . Thus in general, where B_i is only asymptotically independent of n_i , and $Q(n)$ and $P(n)$ are polynomials of the same degree, then B_i can safely be assumed to be independent of n_i .

Secondly, where $Q(n_i)$ is a polynomial in n_i of degree r_1 and $P(n_i)$ of degree r_2 , and we have that $r_1 - r_2 = r > 0$, then we have that

$$\begin{aligned} B_i &= B_i^* \frac{P(n_i)}{Q(n_i)}, \quad \forall i = 1, 2. \\ &= B_i^* \frac{P(n_i)}{n_i^r \tilde{Q}(n_i)}, \quad \forall i = 1, 2. \\ &= \frac{\tilde{B}_i}{n_i^r}, \quad \forall i = 1, 2. \end{aligned} \tag{23}$$

where $\tilde{B}_i = B_i^* \frac{P(n_i)}{\tilde{Q}(n_i)}$. Then it is easily seen that $P(n_i)$ and $\tilde{Q}(n_i)$ are polynomials of the same degree in n_i . To make sense of the proof, recall that

$$A_i = \frac{B_i}{n_1^{p_i}} \tag{24}$$

where $p_1 = p$ and $p_2 = q$ and since $B_1 = \frac{\tilde{B}_1}{n_1^r}$, we have that

$$A_i = \frac{1}{n_i^{p_i}} \frac{\tilde{B}_i}{n_i^r} = \frac{\tilde{B}_i}{n_i^{p_i+r}} \tag{25}$$

Now since \tilde{B}_i is a rational function of n_i where the numerator and the denominator are polynomials of the same degree, it implies that \tilde{B}_i can safely be assumed to be independent of n_i . And this completes the proof. \square

Corollary 1. *Let the posterior distribution of θ given any prior distribution be a linear combination of two t distributed random variables such that the coefficients, say, A_1 and A_2 are functions of n_1^{-p} and n_2^{-q} respectively, and for any $p, q > 0$. Then, if the degrees of freedom of the variables are polynomials in n_1 and n_2 respectively, not necessarily of the same degree, the corresponding Bayesian measure of evidence solves Lindley's paradox.*

Proof. Recall that $A_1 = B_1 n_1^{-p}$ and $A_2 = B_2 n_2^{-q}$. For the purpose of this proof, it will suffice to use any polynomial function as an example.

Because the proof of the Lemma deals with polynomials of the same degree, we shall use polynomials of different degrees in this case. Also, it will suffice to assume that B_1 and B_2 are independent of n_1 and n_2 respectively.

Let the two t random variables be denoted by T_1 and T_2 , with $k_1n_1^2 + c_1n_1 + \gamma_1$ and $k_2n_2 + \gamma_2$ degrees of freedom respectively. Then we have that

$$\begin{aligned} Var(\theta|x) &= \frac{B_1^2(k_1n_1^2 + c_1n_1 + \gamma_1)}{n_1^{2p}(k_1n_1^2 + c_1n_1 + \gamma_1 - 2)} + \frac{B_2^2(k_2n_2 + \gamma_2)}{n_2^{2q}(k_2n_2 + \gamma_2 - 2)} \\ &= \frac{B_1^2n_2^{2q}(k_1n_1^2 + c_1n_1 + \gamma_1)(k_2n_2 + \gamma_2^*)}{n_1^{2p}n_2^{2q}(k_1n_1^2 + c_1n_1 + \gamma_1^*)(k_2n_2 + \gamma_2^*)} \\ &\quad + \frac{B_2^2n_1^{2p}(k_1n_1^2 + c_1n_1 + \gamma_1^*)(k_2n_2 + \gamma_2)}{n_1^{2p}n_2^{2q}(k_1n_1^2 + c_1n_1 + \gamma_1^*)(k_2n_2 + \gamma_2^*)} \end{aligned} \tag{26}$$

Now recall that $Z_0^2 = \frac{[\theta_0 - E(\theta|x)]^2}{Var(\theta|x)}$. Then, letting $A = [\theta_0 - E(\theta|x)]^2$, $\alpha = k_2n_2 + \gamma_2$, and $\alpha^* = k_2n_2 + \gamma_2^*$, we have that

$$\begin{aligned} Z_0^2 &= \frac{A}{Var(\theta|x)} \\ &= \frac{An_1^{2p}n_2^{2q}(k_1n_1^2 + c_1n_1 + \gamma_1^*)\alpha^*}{B_1^2n_2^{2q}(k_1n_1^2 + c_1n_1 + \gamma_1)\alpha^* + B_2^2n_1^{2p}(k_1n_1^2 + c_1n_1 + \gamma_1^*)\alpha} \\ &= \frac{An_2^{2q}(k_1 + c_1n_1^{-1} + \gamma_1^*n_1^{-2})\alpha^*}{B_1^2n_1^{-2p}n_2^{2q}(k_1 + c_1n_1^{-1} + \gamma_1n_1^{-2})\alpha^* + B_2^2(k_1 + c_1n_1^{-1} + \gamma_1^*n_1^{-2})\alpha} \end{aligned} \tag{27}$$

$$\lim_{n_1 \rightarrow \infty} Z_0^2 = \frac{An_2^{2q}(k_2n_2 + \gamma_2^*)}{B_2^2(k_2n_2 + \gamma_2)} \tag{28}$$

And so we have that

$$\lim_{n_2 \rightarrow \infty} \left[\lim_{n_1 \rightarrow \infty} Z_0^2 \right] = \infty \tag{29}$$

Similarly, it can easily be shown that $\lim_{n_1 \rightarrow \infty} \left[\lim_{n_2 \rightarrow \infty} Z_0^2 \right] = \infty$. And the rest of the arguments are same as those found in the proof of Lemma 1. \square

It follows from the Corollary that when the posterior distribution of the difference in means is a linear combination of two t random variables, it is not necessary for the degrees of freedoms to be linear functions of the corresponding sample sizes, but that Lindley's paradox will be solved even when the degrees of freedoms are polynomial functions of different degrees in the corresponding sample sizes.

It can easily be seen that the posterior distributions of the difference of means derived by [6] and [7] under the different circumstances satisfy the conditions for Lemma 1 to hold. As such, Lemma 1 then proves mathematically that using either the Bayesian measure of evidence under Jeffreys' independent prior as obtained by [6] or under the Gamma prior as obtained by [7], both of which are extensions of the methodology of [5], will solve Lindley's paradox. Note that [6] did not provide a mathematical proof of the ability of the new Bayesian measure of evidence of [5] to solve both the Behrens-Fisher problem and Lindley's paradox under Jeffreys' independent prior.

The Pareto Prior

As an example to demonstrate the implication of Lemma 1, we shall assign Pareto priors to the precisions. Consider samples of sizes n_1 and n_2 from $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ respectively. Under the assumption of independence, and letting $x = (x_1, x_2)$, the likelihood function is given by

$$f(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-n_1} \sigma_2^{-n_2} \exp \left[-\frac{1}{2} \sum_{i=1}^2 \frac{S_i(\mu_i)}{\sigma_i^2} \right] \quad (30)$$

where

$$S_i(\mu_i) = \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \quad (31)$$

Let $\tau_i \sim \text{Pareto}(\alpha_i, k_i)$, $i = 1, 2$ then we have that

$$\pi(\tau_i) = \alpha_i k_i^{\alpha_i} \tau_i^{-(\alpha_i+1)} \propto \tau_i^{-(\alpha_i+1)}, \forall i = 1, 2. \quad (32)$$

Let $\pi^{Pa}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2|x)$ denote the Posterior distribution of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ under the Pareto prior, then we have that

$$\begin{aligned} \pi^{Pa}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2|x) &\propto f(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \cdot \pi(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \\ &\propto f(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \cdot \pi(\mu_1, \mu_2|\sigma_1^2, \sigma_2^2) \cdot \pi(\tau_1) \cdot \pi(\tau_2) \\ &\propto \sigma_1^{-n_1} \sigma_2^{-n_2} e^{-\frac{1}{2} \sum_{i=1}^2 \frac{S_i(\mu_i)}{\sigma_i^2}} \tau_1^{-(\alpha_1+1)} \tau_2^{-(\alpha_2+1)} \end{aligned} \quad (33)$$

Then the corresponding Posterior distribution of (μ_1, μ_2) is obtained by integrating (33) with respect to σ_1^2 and σ_2^2 to obtain

$$\begin{aligned} \pi^{Pa}(\mu_1, \mu_2|x) &\propto \left[1 + \frac{n_1(\bar{x}_1 - \mu_1)^2}{(n_1 - 1)S_1^2} \right]^{-(n_1 - 2\alpha_1 - 4)/2} \\ &\quad \left[1 + \frac{n_2(\bar{x}_2 - \mu_2)^2}{(n_2 - 1)S_2^2} \right]^{-(n_2 - 2\alpha_2 - 4)/2} \end{aligned} \quad (34)$$

Let

$$t_i = \frac{\sqrt{n_i(n_i - 2\alpha_i - 5)}(\bar{x}_i - \mu_i)}{\sqrt{(n_i - 1)S_i}}, \quad i = 1, 2. \quad (35)$$

then we have that

$$\begin{aligned} \pi^{Pa}(\mu_1, \mu_2|x) &\propto \left[1 + \frac{t_1^2}{(n_1 - 2\alpha_1 - 5)} \right]^{-(n_1 - 2\alpha_1 - 4)/2} \\ &\quad \left[1 + \frac{t_2^2}{(n_2 - 2\alpha_2 - 5)} \right]^{-(n_2 - 2\alpha_2 - 4)/2} \end{aligned} \quad (36)$$

where (36) is the kernel of the joint distribution of two independent t random variables with $n_1 - 2\alpha_1 - 5$ and $n_2 - 2\alpha_2 - 5$ degrees of freedom respectively and we have the posterior distribution of the difference of means as

$$\theta|x \sim \bar{x}_1 - \bar{x}_2 - \left(\frac{S_1\sqrt{(n_1 - 1)}}{\sqrt{n_1(n_1 - 2\alpha_1 - 5)}}T_1 - \frac{S_2\sqrt{(n_2 - 1)}}{\sqrt{n_2(n_2 - 2\alpha_2 - 5)}}T_2 \right) \quad (37)$$

where T_i is a random variable that follows t distribution with $n_i - 2\alpha_i - 5$ degrees of freedom. The Bayesian measure of evidence under the Pareto prior is then given by

$$P_{Pa}^{BF}(x) = P\left(\left|\frac{S_1\sqrt{(n_1-1)}}{\sqrt{n_1(n_1-2\alpha_1-5)}}T_1 - \frac{S_2\sqrt{(n_2-1)}}{\sqrt{n_2(n_2-2\alpha_2-5)}}T_2\right| \geq |\bar{x}_1 - \bar{x}_2|\right) \quad (38)$$

Note that for the new Bayesian measure of evidence under the Pareto prior to be valid, sample sizes have to be at least 6 each since there are no negative degrees of freedom. In fact, even when sample sizes are as small as six, the choice of α_i has to ensure that $n_i - 2\alpha_i - 5$ results to a positive number. It can easily be seen that the Bayesian measure of evidence of [5] solves the paradox in [1] when a Pareto prior is assigned to the nuisance parameters by noting that equation (37) fulfils all the conditions for Lemma 1 to hold.

Also, since it can be easily seen from (37) that the posterior distribution of θ is symmetric about its expected value, $E(\theta|x) = \bar{x}_1 - \bar{x}_2$, then Theorem 2 of [6] applies here which implies that under the Pareto Prior, the Bayesian measure of evidence of [5] yields the $1 - \alpha$ credible intervals for $\theta = \mu_1 - \mu_2$ centered at $\bar{x}_1 - \bar{x}_2$.

4. Simulation Results and Discussion

For the simulation exercise, we used the Metropolis-Hastings algorithm with a thinning step of 12, while specifying the Normal distribution as the proposal distribution. The values in Table 1 were obtained by fixing the following values: $s_1^2 = 15$, and $s_2^2 = 12$. These values are chosen to demonstrate the case of moderate to large variances. It can be seen from the results here that the new Bayesian measure of evidence under the Pareto prior gives consistently, the highest probability value which corresponds to the weakest evidence against the null hypothesis. In fact, for small sample sizes, conclusion based on the new Bayesian measure of evidence under the Pareto prior has the tendency of failing to reject the null hypothesis while the other three measures of evidence can lead to the rejection of the null hypothesis. However, as the sample sizes increase, the conclusions from the four measures of evidence tend to move towards the same direction.

The values in Table 2 were obtained by fixing the following values: $s_1^2 = 2$, and $s_2^2 = 3$. Here on the other hand, the values chosen reflect the case of small variances. In a similar manner, the results in this table show clearly that the new Bayesian measure of evidence under the Pareto prior gives a consistently weak evidence against the null hypothesis compared to the other three measures of evidence. While for small sample sizes, conclusion based on the new Bayesian measure of evidence under the Pareto prior has the tendency of failing to reject the null hypothesis in contrast to the other three, as the sample sizes increase, the conclusions from the four measures of evidence tend to move towards the same direction.

It is then clear from the results presented in both Table 1 and Table 2 that the new Bayesian measure of evidence under the Pareto prior solves Lindley's paradox even though the evidence against the null provided under the Pareto prior is not as strong as that provided by the other three measures of evidence no matter the circumstances.

Table 1. The four different probability values for different values of $\alpha_1, \alpha_2, \beta_1$ and β_2 .

		$n_1 = n_2 = 6, \alpha_1 = 0.05, \alpha_2 = 0.02, s_1^2 = 15, s_2^2 = 12$							
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.2659	0.3241	0.3915	0.4681	0.5537	0.6475	0.7482	0.8542	0.9633
$ppp(x)$	0.2500	0.3019	0.3640	0.4374	0.5224	0.6188	0.7255	0.8401	0.9597
$P_J^{BF}(x)$	0.3148	0.3726	0.4384	0.5116	0.5921	0.6790	0.7710	0.8676	0.9668
$P_{Pa}^{BF}(x)$	0.7878	0.8116	0.8363	0.8612	0.8865	0.9122	0.9382	0.9645	0.9912
		$n_1 = n_2 = 20, \alpha_1 = 0.05, \alpha_2 = 0.02, s_1^2 = 15, s_2^2 = 12$							
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0378	0.0659	0.1103	0.1766	0.2702	0.3948	0.5504	0.7325	0.9319
$ppp(x)$	0.0437	0.0708	0.1129	0.1757	0.2655	0.3871	0.5419	0.7262	0.9301
$P_J^{BF}(x)$	0.0433	0.0732	0.1197	0.1880	0.2825	0.4069	0.5603	0.7391	0.9338
$P_{Pa}^{BF}(x)$	0.0759	0.1159	0.1722	0.2480	0.3456	0.4663	0.6084	0.7695	0.9415
		$n_1 = n_2 = 50, \alpha_1 = 0.05, \alpha_2 = 0.02, s_1^2 = 15, s_2^2 = 12$							
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0010	0.0035	0.0112	0.0319	0.0800	0.1767	0.3431	0.5875	0.8920
$ppp(x)$	0.0016	0.0047	0.0131	0.0342	0.0817	0.1764	0.3405	0.5842	0.8910
$P_J^{BF}(x)$	0.0011	0.0038	0.0122	0.0338	0.0830	0.1805	0.3482	0.5906	0.8929
$P_{Pa}^{BF}(x)$	0.0017	0.0056	0.0164	0.0422	0.0969	0.2002	0.3686	0.6073	0.8975
		$n_1 = 7, n_2 = 10, \alpha_1 = 0.003, \alpha_2 = 0.005, s_1^2 = 15, s_2^2 = 12$							
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.1917	0.2475	0.3152	0.3953	0.4880	0.5925	0.7072	0.8298	0.9571
$ppp(x)$	0.1907	0.2399	0.3012	0.3762	0.4659	0.5705	0.6888	0.8180	0.9540
$P_J^{BF}(x)$	0.2258	0.2829	0.3505	0.4288	0.5187	0.6180	0.7265	0.8416	0.9602
$P_{Pa}^{BF}(x)$	0.5057	0.5545	0.6064	0.6624	0.7208	0.7820	0.8463	0.9117	0.9779
		$n_1 = 25, n_2 = 20, \alpha_1 = 0.003, \alpha_2 = 0.005, s_1^2 = 15, s_2^2 = 12$							
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0275	0.0509	0.0900	0.1514	0.2418	0.3664	0.5262	0.7168	0.9277
$ppp(x)$	0.0327	0.0557	0.0932	0.1517	0.2385	0.3599	0.5185	0.7109	0.9260
$P_J^{BF}(x)$	0.0317	0.0568	0.0980	0.1613	0.2529	0.3776	0.5362	0.7232	0.9294
$P_{Pa}^{BF}(x)$	0.0548	0.0888	0.1392	0.2100	0.3068	0.4298	0.5792	0.7509	0.9368
		$n_1 = 55, n_2 = 60, \alpha_1 = 0.003, \alpha_2 = 0.005, s_1^2 = 15, s_2^2 = 12$							
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0004	0.0018	0.0067	0.0217	0.0612	0.1486	0.3108	0.5619	0.8846
$ppp(x)$	0.0007	0.0025	0.0081	0.0238	0.0631	0.1489	0.3088	0.5590	0.8836
$P_J^{BF}(x)$	0.0004	0.0020	0.0073	0.0231	0.0637	0.1522	0.3151	0.5650	0.8853
$P_{Pa}^{BF}(x)$	0.0008	0.0029	0.0096	0.0284	0.0736	0.1673	0.3326	0.5791	0.8897
		$n_1 = 55, n_2 = 60, \alpha_1 = 5.3, \alpha_2 = 3.9, s_1^2 = 15, s_2^2 = 12$							
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0004	0.0018	0.0067	0.0217	0.0612	0.1486	0.3108	0.5619	0.8846
$ppp(x)$	0.0007	0.0025	0.0081	0.0238	0.0631	0.1489	0.3088	0.5590	0.8836
$P_J^{BF}(x)$	0.0004	0.0020	0.0073	0.0231	0.0637	0.1522	0.3151	0.5650	0.8853
$P_{Pa}^{BF}(x)$	0.0008	0.0029	0.0096	0.0284	0.0736	0.1673	0.3326	0.5791	0.8897

Table 2. The four different probability values for different values of $\alpha_1, \alpha_2, \beta_1$ and β_2 .

$n_1 = n_2 = 7, \alpha_1 = 0.04, \alpha_2 = 0.07, s_1^2 = 2, s_2^2 = 3$									
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0120	0.0231	0.0441	0.0827	0.1499	0.2596	0.4237	0.6445	0.9078
$ppp(x)$	0.0303	0.0425	0.0624	0.0957	0.1524	0.2475	0.3994	0.6205	0.9002
$P_J^{BF}(x)$	0.0226	0.0387	0.0660	0.1120	0.1869	0.3007	0.4620	0.6715	0.9149
$P_{Pa}^{BF}(x)$	0.2991	0.3470	0.4042	0.4714	0.5495	0.6388	0.7385	0.8472	0.9611
$n_1 = n_2 = 25, \alpha_1 = 0.04, \alpha_2 = 0.07, s_1^2 = 2, s_2^2 = 3$									
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0000	0.0000	0.0001	0.0008	0.0055	0.0300	0.1241	0.3756	0.8240
$ppp(x)$	0.0001	0.0002	0.0006	0.0022	0.0088	0.0350	0.1258	0.3698	0.8205
$P_J^{BF}(x)$	0.0000	0.0000	0.0001	0.0010	0.0067	0.0338	0.1320	0.3849	0.8276
$P_{Pa}^{BF}(x)$	0.0000	0.0001	0.0006	0.0030	0.0141	0.0548	0.1724	0.4309	0.8433
$n_1 = n_2 = 50, \alpha_1 = 0.04, \alpha_2 = 0.07, s_1^2 = 2, s_2^2 = 3$									
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0021	0.0292	0.2089	0.7525
$ppp(x)$	0.0000	0.0000	0.0000	0.0000	0.0002	0.0030	0.0316	0.2081	0.7502
$P_J^{BF}(x)$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0023	0.0312	0.2131	0.7545
$P_{Pa}^{BF}(x)$	0.0000	0.0000	0.0000	0.0000	0.0002	0.0036	0.0391	0.2328	0.7647
$n_1 = 10, n_2 = 8, \alpha_1 = 0.5, \alpha_2 = 0.7, s_1^2 = 2, s_2^2 = 3$									
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0045	0.0104	0.0234	0.0510	0.1058	0.2058	0.3697	0.6051	0.8967
$ppp(x)$	0.0176	0.0261	0.0409	0.0673	0.1156	0.2032	0.3535	0.5854	0.8901
$P_J^{BF}(x)$	0.0097	0.0187	0.0363	0.0695	0.1310	0.2364	0.3995	0.6267	0.9025
$P_{Pa}^{BF}(x)$	0.2595	0.3045	0.3597	0.4267	0.5072	0.6012	0.7093	0.8289	0.9567
$n_1 = 20, n_2 = 25, \alpha_1 = 0.5, \alpha_2 = 0.7, s_1^2 = 2, s_2^2 = 3$									
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0000	0.0000	0.0002	0.0014	0.0082	0.0388	0.1429	0.3985	0.8322
$ppp(x)$	0.0001	0.0003	0.0010	0.0034	0.0122	0.0438	0.1435	0.3915	0.8284
$P_J^{BF}(x)$	0.0000	0.0000	0.0003	0.0019	0.0100	0.0436	0.1523	0.4088	0.8361
$P_{Pa}^{BF}(x)$	0.0001	0.0005	0.0019	0.0074	0.0265	0.0820	0.2175	0.4762	0.8574
$n_1 = 55, n_2 = 60, \alpha_1 = 0.5, \alpha_2 = 0.7, s_1^2 = 2, s_2^2 = 3$									
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	0.0189	0.1762	0.7343
$ppp(x)$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0014	0.0208	0.1759	0.7322
$P_J^{BF}(x)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0201	0.1799	0.7358
$P_{Pa}^{BF}(x)$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0017	0.0268	0.2010	0.7481
$n_1 = 55, n_2 = 60, \alpha_1 = 2.5, \alpha_2 = 3.7, s_1^2 = 2, s_2^2 = 3$									
$(\bar{x}_1 - \bar{x}_2)$	2.500	2.200	1.900	1.600	1.300	1.000	0.700	0.400	0.100
$p(x)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	0.0189	0.1762	0.7343
$ppp(x)$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0014	0.0208	0.1759	0.7322
$P_J^{BF}(x)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0201	0.1799	0.7358
$P_{Pa}^{BF}(x)$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0030	0.0360	0.2250	0.7608

5. Conclusion

In this paper, we have been able to investigate the ability of the new Bayesian measure of evidence of [5] to solve both the Behrens-Fisher problem and Lindley's paradox when a prior is assigned to the parameters which yields a posterior distribution that is a linear combination of two independent t random variables provided that the degrees of freedom of these random variables are linear functions of the corresponding sample sizes. Furthermore, we have shown that the ability of the methodology of [5] to solve Lindley's paradox is not restricted to the case where the degrees of freedoms of the posterior distribution are linear functions of the sample sizes but extend to the case where the degrees of freedom are polynomial functions in the corresponding sample sizes, not necessarily of the same degree. Lemma 1 provides a mathematical proof that the new Bayesian measure of evidence under Jeffreys' independent prior solves Lindley's paradox. More interestingly, the results we have obtained shows that the methodology of [5] does not only solve Lindley's paradox for the purpose reconciling the two approaches (*i.e.* Bayesian and Frequentist), but that this Bayesian procedure will not always favour the null hypothesis.

Using the Pareto prior as an example, we have demonstrated using simulation results that the new Bayesian measure of evidence solves Lindley's paradox even though the new Bayesian measure of evidence under the Pareto prior provides a consistently weak evidence against the null hypothesis compared to either the Generalized p-value, the Posterior Predictive p-value, or the new Bayesian measure of evidence under Jeffreys' independent prior. We were also able to proof that under the Pareto prior, a $1 - \alpha$ credible interval for the difference in means centered around the posterior expectation could be constructed by using the $\alpha/2$ and $1 - (\alpha/2)$ quantiles of the posterior distribution of θ .

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Lindley, D.-V. (1957) A Statistical Paradox. *Biometrika*, **44**, 187-192.
<https://doi.org/10.1093/biomet/44.1-2.187>
- [2] Robert, C.-P. (2014) On the Jeffreys-Lindley Paradox. *Philosophy of Science*, **81**, 216-232.
<https://doi.org/10.1086/675729>
- [3] Sprenger, J. (2013) Testing a Precise Null Hypothesis: The Case of Lindleys Paradox. *Philosophy of Science*, **80**, 733-744.
<https://doi.org/10.1086/673730>
- [4] Spanos, A. (2013) Who Should Be Afraid of the Jeffreys-Lindley Paradox? *Philosophy of Science*, **80**, 73-93.
<https://doi.org/10.1086/668875>
- [5] Yin, Y. (2012) A New Bayesian Procedure for Testing Point Null Hypotheses. *Computational Statistics*, **27**, 237-249.
<https://doi.org/10.1007/s00180-011-0252-6>

- [6] Yin, Y. and Li, B. (2014) Analysis of the Behrens-Fisher Problem Based on Bayesian Evidence. *Journal of Applied Mathematics*, **2014**, Article ID: 978691.
- [7] Goltong, N.-E. and Doguwa, S.-I. (2018) Bayesian Analysis of the Behrens-Fisher Problem under a Gamma Prior. *Open Journal of Statistics*, **8**, 902-914.
<https://doi.org/10.4236/ojs.2018.86060>
- [8] Tsui, K.-W. and Weerahandi, S. (1989) Generalized p-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*, **84**, 602-607.
<https://doi.org/10.2307/2289949>
- [9] Iyer, H.-K. and Paterson, P.-D. (2002) A Recipe for Constructing Generalized Pivotal Quantities and Generalized Confidence Intervals. *Technical Report 2002/10*, Department of Statistics, Colorado State University, Fort Collins.
- [10] Mitra, P.-K. and Sinha, B.-K. (2007) A Generalized p-Value Approach to Inference on Common Mean. *Technical Report*, Department of Mathematics and Statistics, University of Maryland, Baltimore.
- [11] Zheng, S., Shi, N.-Z. and Ma, W. (2010) Statistical Inference on Difference or Ratio of Means from Heteroscedastic Normal Populations. *Journal of Statistical Planning and Inference*, **140**, 1236-1242.
- [12] Ozkip, E., Yazici, B. and Sezer, A. (2014) A Simulation Study on Tests for the Behrens-Fisher Problem. *Turkiye Klinikleri Journal of Biostatistics*, **6**, 59-66.
- [13] Degroot, M.-H. (1982) Comment. *Journal of the American Statistical Association*, **77**, 336-339.
<https://doi.org/10.1080/01621459.1982.10477811>
- [14] Berger, J.-O. and Selke, T. (1987) Testing a Point Null Hypothesis: The Irreconcilability of p Values and Evidence. *Journal of the American Statistical Association*, **82**, 112-122.
<https://doi.org/10.2307/2289131>
- [15] Casella, G. and Berger, R.-L. (1987) Reconciling Bayesian and Frequentist Evidence in One-Sided Testing Problem. *Journal of the American Statistical Association*, **82**, 106-111.
<https://doi.org/10.1080/01621459.1987.10478396>
- [16] Berger, J.-O. and Delampady, M. (1987) Testing Precise Hypotheses. *Statistical Science*, **2**, 317-335.
<https://doi.org/10.1214/ss/1177013238>
- [17] Meng, X.-L. (1994) Posterior Predictive p-Values. *The Annals of Statistics*, **22**, 1142-1160.
<https://doi.org/10.1214/aos/1176325622>
- [18] Yin, Y. and Zhao, J. (2013) Testing Normal Means: The Reconcilability of the p Value and the Bayesian Evidence. *The Scientific World Journal*, **2013**, Article ID: 381539.

- [19] Pestman, W., Tuerlinckx, F. and Vanpaemel, W. (2018) A Reverse to the Jeffreys-Lindley Paradox. *Probability and Mathematical Statistics*, **38**, 243-247.