## Brief Report

# Chi-Square and Student Bridge Distributions and the Behrens-Fisher Statistic 

Wolf-Dieter Richter (1)<br>Institute of Mathematics, University of Rostock, 18057 Rostock, Germany; wolf-dieter.richter@uni-rostock.de

Received: 23 June 2020; Accepted: 20 August 2020; Published: 25 August 2020


#### Abstract

We prove that the Behrens-Fisher statistic follows a Student bridge distribution, the mixing coefficient of which depends on the two sample variances only through their ratio. To this end, it is first shown that a weighted sum of two independent normalized chi-square distributed random variables is chi-square bridge distributed, and secondly that the Behrens-Fisher statistic is based on such a variable and a standard normally distributed one that is independent of the former. In case of a known variance ratio, exact standard statistical testing and confidence estimation methods apply without the need for any additional approximations. In addition, a three pillar bridges explanation is given for the choice of degrees of freedom in Welch's approximation to the exact distribution of the Behrens-Fisher statistic.


Keywords: heteroscedasticity; unbalancedness; sums of weighted chi-squares; variance ratio; Welch approximation; three pillar bridges property

MSC: 62 E 15; 62 F 03; 62 F 25; 28 A 50

## 1. Introduction

Generalizations and modifications of standard statistical distributions, such as chi-square and Student distributions, play a useful role because of their numerous possible applications in different areas of statistics. However, the modifications introduced here, chi-square and Student bridge distributions, are only considered from the subsequent application.

If the normalized chi-square distributed random variable from the denominator of the common ratio representation of a Student distributed random variable is replaced with a mixture of independent normalized chi-square distributed variables, then the resulting ratio follows a distribution that is called here a Student bridge distribution. The possibly most prominent example of this type of random variables is the Behrens-Fisher statistic. The mixing coefficient in the corresponding representation of this statistic depends on the variances of the underlying two Gaussian sample distributions only through their ratio. The variance ratio thus plays the role of a nuisance parameter when deriving the distribution of the Behrens-Fisher statistic. As described in [1], it may happen that the variance ratio is known although the individual variances are not when two instruments of equal precision average different numbers of replicates arriving at a response. Another situation where one of the two variances is known and the other one is not is dealt with in [2].

The well known Behrens-Fisher statistic was introduced already in [3,4]. Several authors provided approximations of its distribution. To mention only some of the earlier contributions, first of all we refer to the well known approximation in [5]. The approximative distribution whose percentage points are dealt with in [6] is often called the Behrens-Fisher distribution. Convolutions of weighted chi-squares are used for an evaluation of the Welch approximation to the distribution of the Behrens-Fisher statistic in [7]. In [8], the exact distribution of the Behrens-Fisher statistic is derived for the case of
two unknown variances, and depends on two unknown parameters, which brings with it the need for additional approximations for statistical applications.

The exact distribution of a modified Behrens-Fisher statistic considered in [9] is very closely related to the distribution derived in [8]. Authors of [9] emphasize that there are (at that time) not many computer programs for computing the special functions that appear as components of the exact distributions and replace these functions mostly with suitable elementary ones.

The alternative aim of the present brief report is to take up again and continue earlier structural considerations on weighted chi-square distributions and their convolutions and on accordingly generalized Student distributions. Knowing the symmetry properties of the generalized Student densities considered here, numerical results obtained in [9] can be taken over to dealing with asymmetric statistical problems in a common way. To be more specific, we prove that the distribution derived in [8] actually depends on the unknown variances only through their ratio, thus allowing to perform exact statistical decisions in case of known variance ratios without additional approximations. Our proof follows a different line than that presented in [8]. In particular, we make more visible the influence the mixture coefficient of the chi-square distributed variables from the denominator of the Behrens-Fisher statistic has on the resulting distribution of the Behrens-Fisher statistic itself. Although the densities of this statistic are visually quite close to each other with varying variance ratios, for many choices of the two sample sizes there are more or less exceptional situations of smaller closeness not mentioned in [8]. From a general structural point of view, our consideration makes the particularly high precision of known approximations to the exact distribution of the Behrens-Fisher statistic more understandable, but it also confirms their limitations, as pointed out in [9] for selected cases from a numerical point of view.

The more general problem of finding an optimal expectation test in the Gaussian two-sample scheme is called the Behrens-Fisher problem. It is dealt with in [10] as a problem in the presence of three nuisance parameters. Reviews on numerous papers dealing with the Behrens-Fisher problem and the distribution of the Behrens-Fisher statistic can be found, e.g., in [11] and in [12]. The connections between the different classical approaches to statistics and the Behrens-Fisher problem are emphasized in [11], while in [12] there is an emphasis on three procedures that are in a certain suitably defined sense exact solutions to the Behrens-Fisher problem. The multivariate Behrens-Fisher distribution is considered, e.g., in [13,14]; for the nonparametric approach to the Behrens-Fisher problem see [15] and the references given there.

The present paper does not deal with the general Behrens-Fisher problem but is devoted to the study of the probability density function of the Behrens-Fisher statistic with a focus on a function of the mixing parameter as a nuisance parameter. We explicitly describe the influence the single nuisance parameter has on the Student bridge distribution.

The two-sample $t$-test with a known ratio of variances where the pooled empirical variance is used instead of individual sample variances is dealt with in [1,2]. What these papers have in common is that, unlike here, Student distributions with estimated d.f. are used for performing statistical tests. A test statistic conditional on the value of the variance ratio is studied in [16].

We derive here exact representations of the pdf of the Behrens-Fisher statistic allowing heteroscedasticity and unbalancedness, i.e., different variances and sample sizes, respectively. These representations can be considered as heteroscedasticity-unbalancedness generalizations of Student's density.

The paper is organized as follows. The chi-square bridge distribution is introduced and its moments are described in Section 2. Section 3 deals with the Student bridge distribution and Section 4 with its application to the Behrens-Fisher statistic. A discussion including a three pillar bridges explanation for the choice of degrees of freedom in the Welch approximation to the exact distribution of the Behrens-Fisher statistic is presented in Section 5. Figures were drawn using Matlab.

## 2. Chi-Square Bridge Distribution

Let $C Q_{1, k}$ and $C Q_{2, m}$ be independent random variables, where $C Q_{i, d}$ is chi-square distributed with $d$ d.f., $d \in\{k, m\}$, and $i=1,2$. For $\gamma \in[0,1]$, we consider the mixture of normalized chi-squares or weighted sum of Chi-squares

$$
W=\operatorname{WSCS}(k, m ; \gamma)=\gamma \frac{C Q_{1, k}}{k}+(1-\gamma) \frac{C Q_{2, m}}{m} .
$$

The first and second order moments of $W$ are

$$
\mathbb{E}(W)=1 \text { and } V(W)=2\left(\frac{\gamma^{2}}{k}+\frac{(1-\gamma)^{2}}{m}\right)
$$

respectively. Minimal variance with respect to the mixing coefficient is attained for $\gamma=\gamma_{0}$, where

$$
\begin{equation*}
\gamma_{0}=\frac{k}{k+m} . \tag{1}
\end{equation*}
$$

Let $X \sim h$ indicate that the random variable $X$ follows the probability distribution $h$. If $A=\gamma / k$ and $B=(1-\gamma) / m$, then $A=B$ holds for $\gamma=\gamma_{0}$ and statistic $W / A$ follows in this case a chi-square distribution with $k+m$ d.f., $(k+m) \operatorname{WSCS}\left(k, m ; \gamma_{0}\right) \sim \chi_{k+m}^{2}$. Moreover, $k \cdot \operatorname{WSCS}(k, m ; 1) \sim \chi_{k}^{2}$ and $m \cdot \operatorname{WSCS}(k, m ; 0) \sim \chi_{m}^{2}$. That is why we say that the distribution of WSCS has a three pillar bridges property.

In what follows we assume that $\gamma \in(0,1)$. The density of $W$ can immediately be derived then from its convolution integral representation

$$
\begin{gathered}
f_{W}(x)=\int_{-\infty}^{\infty} f_{\frac{\gamma}{k}} C Q_{1, k}(x-z) f_{\frac{1-\gamma}{m}} C Q_{2, m}(z) d z \\
=\frac{\left(\frac{k}{\gamma}\right)^{\frac{k}{2}}\left(\frac{m}{1-\gamma}\right)^{\frac{m}{2}}}{2^{\frac{k+m}{2}} \Gamma\left(\frac{k}{2}\right) \Gamma\left(\frac{m}{2}\right)} \int_{0}^{x}(x-z)^{\frac{k}{2}-1} e^{-\frac{k(x-z)}{2 \gamma}} z^{\frac{m}{2}-1} e^{-\frac{m z}{2(1-\gamma)}} d z
\end{gathered}
$$

and allows according to commutativity of summands the following two representations:

$$
\begin{equation*}
f_{W}(x)=\left(\frac{k}{\gamma}\right)^{k / 2}\left(\frac{m}{1-\gamma}\right)^{m / 2} \frac{x^{(k+m) / 2-1}{ }_{1} F_{1}\left(\frac{m}{2}, \frac{k+m}{2} ; \frac{x}{2}\left(\frac{k}{\gamma}-\frac{m}{1-\gamma}\right)\right)}{2^{(k+m) / 2} \Gamma\left(\frac{k+m}{2}\right) \exp \left\{\frac{k x}{2 \gamma}\right\}}, x>0 \tag{2}
\end{equation*}
$$

and

$$
\begin{equation*}
f_{W}(x)=\left(\frac{k}{\gamma}\right)^{k / 2}\left(\frac{m}{1-\gamma}\right)^{m / 2} \frac{x^{(k+m) / 2-1}{ }_{1} F_{1}\left(\frac{k}{2}, \frac{k+m}{2} ; \frac{x}{2}\left(\frac{m}{1-\gamma}-\frac{k}{\gamma}\right)\right)}{2^{(k+m) / 2} \Gamma\left(\frac{k+m}{2}\right) \exp \left\{\frac{m x}{2(1-\gamma)}\right\}}, x>0 \tag{3}
\end{equation*}
$$

where

$$
{ }_{1} F_{1}(a, b ; z)=\frac{1}{B(a, b-a)} \int_{0}^{1} e^{z t} t^{a-1}(1-t)^{b-a-1} d t
$$

denotes the hypergeometric function of order (1,1), see, e.g., Formula 13.2.1 in [17] and Formula 9.210 in [18]. The Beta function can be expressed in terms of the Gamma function $x \rightarrow \Gamma(x)=\int_{0}^{\infty} t^{x-1} e^{-t} d t$ as $B(a, b)=\Gamma(a) \Gamma(b) / \Gamma(a+b)$. In the case that $\gamma=\gamma_{0}$, we have that ${ }_{1} F_{1}\left(a, b ; \frac{x}{2}\left(\frac{k}{\gamma}-\frac{m}{1-\gamma}\right)\right)=1$ and $f_{W}(x)=(1 / A) f_{\chi_{k+m}^{2}}(x / A)$. Choosing $\frac{k}{\gamma}<\frac{m}{1-\gamma}$ in (2) avoids unboundedness of $e^{z t}$ in the integrand of ${ }_{1} F_{1}(a, b ; z)$ and might motivate favoring Formula (2) over Formula (3), in this case.

Definition 1. The probability distribution having density (2) (or (3)) will be called chi-square bridge distribution with $(k, m)$ d.f. and mixing parameter $\gamma$, or $(k, m ; \gamma)$-chi-square distribution $\chi_{k, m ; \gamma}^{2}$, for short.

Figures 1 and 2 show the density $f_{W}$ of the distribution $\chi_{k, m ; \gamma}^{2}$ for four different pairs $(k, m)$, and $\gamma \in\{0.1,0.505,0.962\}$ or $\gamma \in\{0.01,0.505,0.7,0.99\}$, respectively.


Figure 1. (a) Example 1; (b) Example 2.


Figure 2. (a) Example 3; (b) Example 4.

## 3. Student Bridge Distribution

If $N$ denotes a standard Gaussian distributed random variable that is independent of $C Q_{1, k}$ and $C Q_{2, m}$, and $t_{l}$ Student's $t$-distribution with $l$ d.f., then the statistic

$$
T=T_{k, m ; \gamma}=\frac{N}{\sqrt{W S C S}(k, m ; \gamma)}
$$

satisfies

$$
\begin{equation*}
T_{k, m ; 1} \sim t_{k}, \quad T_{k, m ; \gamma_{0}} \sim t_{k+m} \text { and } T_{k, m ; 0} \sim t_{m} \tag{4}
\end{equation*}
$$

By the general integral representation of the density of the ratio of two independent continuous random variables,

$$
f_{T}(t)=\int_{-\infty}^{\infty} f_{N}(t x) f_{\sqrt{W S C S}(k, m ; \gamma)}(t)|t| d t
$$

where $f_{N}$ denotes the standard Gaussian density and $f_{\sqrt{W}}(t)=2 t f_{W}\left(t^{2}\right)$ can easily be derived from (2) or (3). Making use of (2) and changing the order of integration gives us:

$$
\begin{aligned}
& f_{T}(t)=\frac{2\left(\frac{k}{\alpha}\right)^{k / 2}\left(\frac{m}{1-\alpha}\right)^{m / 2}}{\sqrt{2 \pi} 2^{(k+m) / 2} \Gamma(k / 2) \Gamma(m / 2} \\
& \quad \cdot \int_{0}^{1} v^{m / 2-1}(1-v)^{k / 2-1} \int_{0}^{\infty} t^{k+m-1} \exp \left\{-\frac{t^{2} x^{2}}{2}-\frac{k t^{2}}{2 \alpha}-\frac{-t^{2} v\left(\frac{m}{1-\alpha}-\frac{k}{2}\right)}{2}\right\} t d t d v,
\end{aligned}
$$

and changing the variables $s=\frac{t^{2}}{2}\left(x^{2}+\frac{k}{\alpha}+v\left(\frac{m}{1-\alpha}-\frac{k}{2}\right)\right)$ leads to the first of the following two alternative representations of the density $f_{T}$,

$$
\begin{equation*}
f_{T}(x)=\left(\frac{m \gamma}{k(1-\gamma)}\right)^{m / 2} \frac{\Gamma\left(\frac{k+m+1}{2}\right){ }_{2} F_{1}\left(\frac{k+m+1}{2}, \frac{m}{2} ; \frac{k+m}{2} ; \frac{\frac{k}{\gamma}-\frac{m}{1-\gamma}}{\frac{k}{\gamma}+x^{2}}\right)}{\Gamma\left(\frac{k+m}{2}\right) \sqrt{\pi \frac{k}{\gamma}}\left(1+\frac{\gamma}{k} x^{2}\right)^{(k+m+1) / 2}}, x \in \mathbb{R} \tag{5}
\end{equation*}
$$

and

$$
\begin{equation*}
f_{T}(x)=\left(\frac{k(1-\gamma)}{m \gamma}\right)^{k / 2} \frac{\Gamma\left(\frac{k+m+1}{2}\right)_{2} F_{1}\left(\frac{k+m+1}{2}, \frac{k}{2} ; \frac{k+m}{2} ; \frac{\frac{m}{1-\gamma}-\frac{k}{\eta}}{1-\gamma}+x^{2}\right.}{\left.1-\frac{k+m}{2}\right) \sqrt{\pi \frac{m}{(1-\gamma)}}\left(1+\frac{1-\gamma}{m} x^{2}\right)^{(k+m+1) / 2}}, x \in \mathbb{R} \tag{6}
\end{equation*}
$$

Making use of (3) instead of (2) proves (6). Here, ${ }_{2} F_{1}$ denotes the hypergeometric function being defined for $\delta>\beta>0$ by

$$
{ }_{2} F_{1}(\gamma, \beta ; \delta ; y)=\frac{1}{B(\beta, \delta-\beta)} \int_{0}^{1}(1-z)^{\delta-\beta-1} z^{\beta-1}(1-y z)^{-\gamma} d z
$$

see 15.3 .1 in [17] and Formula 9.111 in [18]. Choosing $y<0$ avoids a zero of $1-y z$ within the range of integration and might motivate a favor of using Formulas (5) or (6) if $\frac{k}{\gamma}<\frac{m}{1-\gamma}$ or $\frac{k}{\gamma}>\frac{m}{1-\gamma}$, respectively. The following definition is motivated by the three pillar bridges property (4).

Definition 2. The probability distribution corresponding to density (5) (or (6)) will be called a Student bridge distribution with $(k, m)$ d.f. and mixing coefficient $\gamma$, or $(k, m ; \gamma)$-Student bridge distribution $t_{k, m ; \gamma}$, for short.

Figures 3 and 4 show the density $f_{T}$ of the distribution $t_{k, m ; \gamma}$ for the same choice of parameters as for $f_{W}$.

We remark that for $l=1,2, \ldots, \mathbb{E} T^{2 l-1}=0$ and

$$
\begin{aligned}
\mathbb{E} T^{2 l}= & \frac{(2 l-1)!!\left(\frac{k}{\gamma}\right)^{k / 2}\left(\frac{m}{1-\gamma}\right)^{m / 2}}{2^{(k+m) / 2} \Gamma((k+m) / 2)} \\
& \cdot \int_{0}^{\infty} x^{\frac{k+m}{2}-1-l} e^{-\frac{k x}{2 \gamma}}{ }_{1} F_{1}\left(\frac{m}{2}, \frac{k+m}{2}, \frac{x}{2}\left(\frac{k}{\gamma}-\frac{m}{1-\gamma}\right)\right) d x,
\end{aligned}
$$

that is

$$
\begin{aligned}
\mathbb{E} T^{2 l}= & \frac{\Gamma\left(\frac{k+m}{2}-l\right)}{\Gamma\left(\frac{k+m}{2}\right)}\left(\frac{k}{2 \gamma}\right)^{l}\left(\frac{\gamma m}{(1-\gamma) k}\right)^{m / 2} \\
& \quad \cdot{ }_{2} F_{1}\left(\frac{k+m}{2}-l, \frac{m}{2} ; \frac{k+m}{2} ; 1-\frac{\gamma m}{(1-\gamma) k}\right) .
\end{aligned}
$$


(a)

(b)

Figure 3. (a) Example 1; (b) Example 2.


Figure 4. (a) Example 3; (b) Example 4.

## 4. Behrens-Fisher Statistic

Let $X_{1}, \ldots, X_{n_{1}}$ and $Y_{1}, \ldots, Y_{n_{2}}$ be jointly independent Gaussian samples with expectations $\mu_{i}$ and variances $\sigma_{i}^{2}, i=1,2$. We consider the statistic

$$
T^{B F}=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_{X}^{2}}{n_{1}}+\frac{S_{Y}^{2}}{n_{2}}}}
$$

where $\bar{X}=\frac{1}{n_{1}} \sum_{i=1}^{n_{1}} X_{i}, \bar{Y}=\frac{1}{n_{2}} \sum_{i=1}^{n_{2}} Y_{i}$ and

$$
S_{X}^{2}=\frac{1}{n_{1}-1} \sum_{i=1}^{n_{1}}\left(X_{i}-\bar{X}\right)^{2}, S_{Y}^{2}=\frac{1}{n_{2}-1} \sum_{i=1}^{n_{2}}\left(Y_{i}-\bar{Y}\right)^{2}
$$

are common sample means and unbiased sample variances, respectively. By $Z \stackrel{d}{=} U$ we mean that two random variables $Z$ and $U$ have the same probability distribution.

Lemma 1. The Behrens-Fisher statistic allows the representation

$$
T^{B F} \stackrel{d}{=} \frac{Z_{n_{1}+n_{2}-1}}{\sqrt{S^{2}}}
$$

with

$$
S^{2}=A^{*} \sum_{i=1}^{n_{1}-1} Z_{i}^{2}+B^{*} \sum_{i=n_{1}}^{n_{1}+n_{2}-2} Z_{i}^{2}
$$

and where $\left(Z_{1}, \ldots, Z_{n_{1}+n_{2}-1}\right)^{T}$ is a standard Gaussian distributed random vector taking values in $\mathbb{R}^{n_{1}+n_{2}-1}$,

$$
A^{*}=\frac{1}{1+\frac{n_{1}}{n_{2}}\left(\frac{\sigma_{2}}{\sigma_{1}}\right)^{2}} \cdot \frac{1}{n_{1}-1}, B^{*}=\frac{1}{1+\left[\frac{n_{1}}{n_{2}}\left(\frac{\sigma_{2}}{\sigma_{1}}\right)^{2}\right]^{-1}} \cdot \frac{1}{n_{2}-1}
$$

and

$$
Z_{n_{1}+n_{2}-1} \stackrel{d}{=} \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_{1}^{2}}{n_{1}}+\frac{\sigma_{2}^{2}}{n_{2}}}} \text { and } S^{2} \stackrel{d}{=} \frac{\frac{S_{X}^{2}}{n_{1}}+\frac{S_{Y}^{2}}{n_{2}}}{\frac{\sigma_{1}^{2}}{n_{1}}+\frac{\sigma_{2}^{2}}{n_{2}}}
$$

are independent.
Proof. We put

$$
1_{k}=(1, \ldots, 1)^{T} \in \mathbb{R}^{k}, 0_{k}=(0, \ldots, 0)^{T} \in \mathbb{R}^{k}, 1^{+0}=\left(1_{n_{1}}^{T} 0_{n_{2}}^{T}\right)^{T} \text { and } 1^{0+}=\left(0_{n_{1}}^{T} 1_{n_{2}}^{T}\right)^{T}
$$

Let us further denote the orthogonal projection onto the linear space $\mathfrak{L}=\mathfrak{L}\left(1^{+0}, 1^{0+}\right)$ by $\Pi_{\mathfrak{L}}$ and let the matrix $P$ be defined such that $\Pi_{\mathfrak{L}} x=P x, \forall x \in \mathbb{R}^{n} ; n=n_{1}+n_{2}$, then

$$
P=\left(\begin{array}{ll}
\frac{1}{n_{1}} \mathbf{I}_{n_{1}} & \\
& \frac{1}{n_{2}} \mathbf{I}_{n_{2}}
\end{array}\right)
$$

where $\mathbf{I}_{k}=1_{k} 1_{k}^{T}$ is a $k \times k$-matrix. Here and below, missing off-diagonal matrix elements are zero. If $\mathfrak{L}^{\perp}$ is the subspace of $\mathbb{R}^{n}$ being orthogonally to $\mathfrak{L}$ then $I_{n}-P=\Pi_{\mathfrak{L}^{\perp}}$. The statistic $T^{B F}$ can be written as:

$$
\begin{equation*}
T^{B F}=\frac{\left(\pi_{\mathfrak{L}}\binom{X}{Y}, \frac{1}{n_{1}} 1^{+0}-\frac{1}{n_{2}} 1^{0+}\right)}{\left\|\binom{X}{Y}-\pi_{\mathfrak{L}}\binom{X}{Y}\right\|_{(n)}} \tag{7}
\end{equation*}
$$

where the functional $\|\cdot\|_{(n)}$ is defined for all $x=\left(x_{1}, \ldots, x_{n_{1}}\right)^{T} \in \mathbb{R}^{n_{1}}$ and $y=\left(y_{1}, \ldots, y_{n_{2}}\right)^{T} \in \mathbb{R}^{n_{2}}$ as:

$$
\left\|\binom{x}{y}\right\|_{(n)}=\left(\frac{1}{n_{1}\left(n_{1}-1\right)} \sum_{1}^{n_{1}} x_{i}^{2}+\frac{1}{n_{2}\left(n_{2}-1\right)} \sum_{1}^{n_{2}} y_{i}^{2}\right)^{1 / 2} .
$$

We note that $P \mu=\mu$ and $\left(I_{n}-P\right) \mu=0_{n}$, and put $\kappa=\left(\mu_{1} 1_{n_{1}}^{T} 0_{n_{2}}^{T}\right)^{T}$. The random vector

$$
\zeta=\binom{P\binom{X}{Y}}{\left(I_{n}-P\right)\binom{X}{Y}}
$$

takes its values in $\mathbb{R}^{2 n}$ and follows a singular Gaussian distribution of rank $n$,

$$
\zeta \sim \Phi_{\kappa, \Psi} \text { with } \Psi=\left(\begin{array}{cccc}
\frac{\sigma_{1}^{2}}{n_{1}} \mathbf{I}_{n_{1}} & & & \\
& \frac{\sigma_{2}^{2}}{n_{2}} \mathbf{I}_{n_{2}} & & \\
& & \sigma_{1}^{2}\left(I_{n_{1}}-\frac{1}{n_{1}} \mathbf{I}_{n_{1}}\right) & \\
& & & \sigma_{1}^{2}\left(I_{n_{2}}-\frac{1}{n_{2}} \mathbf{I}_{n_{2}}\right)
\end{array}\right)
$$

As a consequence, the nominator and denominator of the ratio statistic $T^{B F}$ are stochastically independent. Let

$$
B_{1}^{T}=\left(\begin{array}{cccc}
b_{1} & \cdots & b_{n_{1}-1} & \frac{1}{\sqrt{n_{1}}} 1_{n_{1}}
\end{array}\right), B_{2}^{T}=\left(\begin{array}{cccc}
c_{1} & \cdots & c_{n_{2}-1} & \frac{1}{\sqrt{n_{2}}} 1_{n_{2}}
\end{array}\right)
$$

be orthogonal $n_{1} \times n_{1}$ and $n_{2} \times n_{2}$ matrices, respectively. The random vector $\eta=B\left(I_{n}-P\right)\binom{X}{Y}$ with $B=\left(\begin{array}{cc}B_{1} & \\ & B_{2}\end{array}\right)$ follows a centered Gaussian distribution with the covariance matrix

$$
B\left(\begin{array}{cc}
\sigma_{1}^{2}\left(I_{n_{1}}-\frac{1}{n_{1}} \mathbf{I}_{n_{1}}\right) & \\
& \sigma_{1}^{2}\left(I_{n_{2}}-\frac{1}{n_{2}} \mathbf{I}_{n_{2}}\right)
\end{array}\right) B^{T}=\left(\begin{array}{cc}
\sigma_{1}^{2}\left(\begin{array}{cc}
I_{n_{1}-1} & 0 \\
0 & 0
\end{array}\right) & \\
& \\
& \sigma_{2}^{2}\left(\begin{array}{cc}
I_{n_{2}-1} & 0 \\
0 & 0
\end{array}\right)
\end{array}\right)
$$

The vector $\left(I_{n}-P\right)\binom{X}{Y}=B^{T} \eta$ allows almost surely the representation

$$
\begin{aligned}
B^{T} \eta & =B^{T}\left(N_{1}, \ldots, N_{n_{1}-1}, 0, N_{n_{1}+1}, \ldots, N_{n-1}, 0\right)^{T} \\
& =\sum_{i=1}^{n_{1}-1} N_{i}\binom{b_{i}}{0_{n_{2}}}+\sum_{j=1}^{n_{2}-1} N_{n_{1}+j}\binom{0_{n_{1}}}{c_{j}} \\
& =\left(\begin{array}{c}
\sum_{i=1}^{n_{1}-1} N_{i} b_{i} \\
n_{2}-1 \\
\sum_{j=1} N_{n_{1}+j} c_{j}
\end{array}\right)
\end{aligned}
$$

where the random variables $N_{i}$ are independent and centered normally distributed with variances $\sigma_{1}^{2}$ and $\sigma_{2}^{2}$ for $i=1, \ldots, n_{1}-1$ and $i=n_{1}+1, \ldots, n-1$, respectively. Let

$$
B_{1}^{* T}=\left(\begin{array}{lll}
b_{1} & \ldots & b_{n_{1}-1}
\end{array}\right) \text { and } B_{2}^{* T}=\left(\begin{array}{lll}
c_{1} & \ldots & c_{n_{2}-1}
\end{array}\right)
$$

The Kronecker product matrix $B^{* T}=B_{1}^{* T} \otimes B_{2}^{* T}=\left(\begin{array}{cc}B_{1}^{* T} & \\ & B_{2}^{* T}\end{array}\right)$ describes then a mapping from $\mathbb{R}^{2 n-2}$ to $\mathbb{R}^{2 n}$ and, a.s.,

$$
\begin{aligned}
\left\|B^{T} \eta\right\|_{(n)} & =\left\|\binom{B_{1}^{* T}\left(\begin{array}{c}
N_{1} \\
\vdots \\
N_{n_{1}-1}
\end{array}\right)}{B_{2}^{* T}\left(\begin{array}{c}
N_{n_{1}+1} \\
\vdots \\
N_{n-1}
\end{array}\right)}\right\|_{(n)}^{*} \\
& =\left(\frac{1}{n_{1}\left(n_{1}-1\right)}\left\|B_{1}^{* T}\left(\begin{array}{c}
N_{1} \\
\vdots \\
N_{n_{1}-1}
\end{array}\right)\right\|^{2}+\frac{1}{n_{2}\left(n_{2}-1\right)}\left\|B_{2}^{* T}\left(\begin{array}{c}
N_{n_{1}+j} \\
\vdots \\
N_{n-1}
\end{array}\right)\right\|^{2}\right)^{1 / 2} \\
& =\left(\frac{1}{n_{1}\left(n_{1}-1\right)}\left\|\sum_{i=1}^{n_{1}-1} N_{i} b_{i}\right\|^{2}+\frac{1}{n_{2}\left(n_{2}-1\right)}\left\|\sum_{j=1}^{n_{2}-1} N_{n_{1}+j} c_{j}\right\|^{2}\right)^{1 / 2} \\
& =\left(\frac{1}{n_{1}\left(n_{1}-1\right)} \sum_{i=1}^{n_{1}-1} N_{i}^{2}+\frac{1}{n_{2}\left(n_{2}-1\right)} \sum_{j=1}^{n_{2}-1} N_{n_{1}+j}^{2}\right)^{1 / 2}
\end{aligned}
$$

where the norm $\|.\|_{(n)}^{*}$ is defined in $\mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$. The variance of the nominator of the Behrens-Fisher statistic is

$$
\mathbb{V}(\bar{X}-\bar{Y})=\frac{\sigma_{1}^{2}}{n_{1}}+\frac{\sigma_{2}^{2}}{n_{2}} .
$$

Hence, $T^{B F}$ may be represented as

$$
\left.T^{B F} \stackrel{d}{=} \frac{\left(\frac{\sigma_{1}^{2}}{n_{1}}+\frac{\sigma_{2}^{2}}{n_{2}}\right)^{1 / 2} N_{n}}{\left(\frac{\sigma_{1}^{2}}{n_{1}\left(n_{1}-1\right)}\right.} \sum_{i=1}^{n_{1}-1} N_{i}^{2}+\frac{\sigma_{2}^{2}}{n_{2}\left(n_{2}-1\right)} \sum_{j=1}^{n_{2}-1} N_{n_{1}+j}^{2}\right)^{1 / 2}
$$

where the standard Gaussian distributed random variable $N_{n}$ is independent of $N_{1}, \ldots, N_{n_{1}-1}, N_{n_{1}+1}, \ldots, N_{n-1}$.

The constants $A^{*}$ and $B^{*}$ from Lemma 1 depend on $\sigma_{1}$ and $\sigma_{2}$ only through the variance ratio $V R=\sigma_{2}^{2} / \sigma_{1}^{2}$, which itself plays the role of a nuisance parameter.

If $\theta=V R / S R$ where $S R=\frac{n_{2}}{n_{1}}$ is the sample size ratio then the constants $A^{*}, B^{*}, V R$ and $n_{2}$ may be expressed in terms of the parameter triple $\left(\theta, n_{1}, S R\right)$. The constants $A^{*}$ and $B^{*}$ depend on the variance ratio $V R$, but in a different way for the different sample size ratio $S R$. For $n_{1}$ given, the inverse mapping $\left(A^{*}, B^{*}\right) \rightarrow(\theta, S R)$ is defined by:

$$
\theta=\frac{1}{A^{*}\left(n_{1}-1\right)}-1, S R=\frac{1+A^{*}+B^{*}-A^{*} n_{1}}{n_{1} B^{*}}
$$

Other parameter triples could be introduced, e.g., $(k, m, \theta)$, being closely related to but nevertheless different from the parameter triple in [6].

The first and second order moments of $S^{2}$ are

$$
\mathbb{E} S^{2}=1 \text { and } V\left(S^{2}\right)=\frac{2}{(1+\theta)^{2}}\left[\frac{1}{k}+\frac{\theta^{2}}{m}\right]
$$

Finally, it turns out that under the hypothesis

$$
H_{0}: \mu_{1}=\mu_{2}
$$

the statistic $T^{B F}$ allows the representation

$$
\begin{equation*}
\left.T^{B F}\right|_{H_{0}} \stackrel{d}{=} \frac{N}{\sqrt{\gamma \frac{C Q_{1, n_{1}-1}}{n_{1}-1}+(1-\gamma) \frac{C Q_{2, n_{2}-1}}{n_{2}-1}}} \tag{8}
\end{equation*}
$$

where the independent random variables $N$ and $C Q_{i, n_{i}-i}, i=1,2$ are as in Section $2, k=n_{1}-1, m=$ $n_{2}-1$, and the mixing coefficient is

$$
\begin{equation*}
\gamma=\left(1+\frac{V R}{S R}\right)^{-1} \tag{9}
\end{equation*}
$$

Thus, the Behrens-Fisher statistic follows the Student bridge distribution with d.f. $\left(n_{1}-1, n_{2}-1\right)$ and mixing coefficient $\gamma$, or $\left(n_{1}-1, n_{2}-1 ; \gamma\right)$-Student bridge distribution, for short,

$$
\begin{equation*}
\left.T^{B F}\right|_{H_{0}} \sim t_{n_{1}-1, n_{2}-1 ;\left(1+\frac{V R}{S R}\right)^{-1}} \tag{10}
\end{equation*}
$$

In case of a known variance ratio, standard statistical significance testing and confidence estimation methods are based therefore upon the $\left(n_{1}-1, n_{2}-1 ; \gamma\right)$-Student bridge distribution in the common way. Here, assumption $\frac{k}{\gamma}<\frac{m}{1-\gamma}$ from Section 2 means that

$$
\frac{n_{1}\left(n_{1}-1\right)}{n_{2}\left(n_{2}-1\right)}<\frac{\sigma_{1}^{2}}{\sigma_{2}^{2}}
$$

Without going here into technical details, the unrestricted distribution of $T^{B F}$ is a non-central Student bridge distribution in a suitably defined sense.

## 5. Discussion

### 5.1. Reflection of the Three Pillar Bridges Property

We now consider four examples from [8] for demonstrating the role the three pillar bridges property discussed in this paper may play in practical statistical work. In each example, we chose a pair of sample sizes $(k, m)$ from the set $\{(11,14),(14,14),(6,2),(3,9)\}$, and a variance ratio $V R$ from the positive real line that we assume to be known. In any case, we then determine the mixture coefficient $\gamma$ by a one-to-one calculation from $V R$. This way, Examples 1-4 are described (with some redundancy) as: $E 1(11,14|1.25| 0.505), E 2(14,14|0.04| 0.962), E 3(6,2|1.25| 0.505)$ and $E 4(3,9|0.04| 0.962)$. Figures $1-4$ show the densities $f_{W}=f_{W S C S}(k, m ; \gamma)$ and $f_{T}=f_{T_{k, m ; \gamma}}$ for more parameter combinations ( $k, m ; \gamma$ ) than required in Examples E1 to E4.

A value of $\gamma$ close to 1 corresponds to a value of $\theta=V R / S R$ close to zero, $V R \ll S R$, meaning that the sample size in the second population compared to that in the first is disproportionately large compared to the corresponding quotient of variances; in other words, the first population is under represented.

A value of $\gamma$ close to 0 corresponds to a very large value of $\theta=V R / S R, S R \ll V R$, meaning that the sample size in the first population compared to that in the second is disproportionately large compared to the corresponding quotient of variances; in other words, the second population is under represented.

Unlike these two cases of imbalance, a value of $\gamma$ in the order of $\gamma_{0}$ speaks for an approximately achieved balance. The latter can be observed close to the middle pillar of the three pillar Student
bridge distribution and is suitable to explain some effect when choosing the degree of freedom in the Welch approximation to the exact density of $T$.

The chi-square bridge densities shown in Figure 1 correspond tho those of the Student bridge densities in Figure 3. It is shown in [8] that for such cases the Welch approximation seems to be the best that were found so far. Welch's approximate degrees of freedom, see Formula (1.2) in [8] with $N_{1}-1=k, N_{2}-1=m$ and $\sigma_{i}^{2}$ replaced with $S_{i}^{2}, i=1,2$, are $f=25$ for Example 1 and $f=15$ for Example 2. This corresponds very well to the three pillar property of the Student bridge distribution.

If $\gamma=1 / 2$, as is approximately the case in Example 1, then the denominator $\mathcal{N}$ of $T$ can be written as $\mathcal{N}^{2}=\left(C Q_{1,11} / 11+C Q_{2,14} / 14\right) / 2$. Because the numbers 11 and 14 are of comparable size, a reasonable approximation is $\mathcal{N}^{2} \approx\left(C Q_{1,11}+C Q_{2,14}\right) /(2 \times 12.5) \sim C Q_{25} / 25$ finally leading for Example 1 to $T \approx t_{25}$.

In Example 2, the denominator of $T$ allows the representation $\mathcal{N}^{2}=0.962 C Q_{1,14} / 14+$ $0.038 C Q_{2,14} / 14$, which is reasonably approximated by $\mathcal{N}^{2} \approx C Q_{15} / 15$. Thus, $T \approx t_{15}$.

Figure 2 shows a broader variability between the densities when the mixing coefficient $\gamma$ is varied compared to Figure 1. This is reflected in more visible variation of the corresponding Student bridge densities in Figure 4, both in their distribution centers and their distribution tails. This should be taken into account if applications of the Student bridge distribution are required, in particular in the areas of the distributions just mentioned.

The Welch approximation is known to perform better when both $k$ and $m$ are sufficiently large. Our Figures show what may happen for small sample sizes.

Because the consideration in [8] is even for higher dimensions it might be of some interest to extend the present work to this case, too.

### 5.2. Examples Where the Student Bridge Distribution Should Be Preferred

The aim of this section is to give a complementary structural argumentation confirming the numerical discoveries in [9] with respect to the question of when Welch's approximation is not sufficiently precise. To this end, we present, for two cases of sample sizes and variance ratios, the exact Student bridge density and Welch's approximation to it in a joint figure.

Example 5. Assume that as in Figures 1-3 in [9], sample sizes $\left(n_{1}, n_{2}\right)$ are $(5,3),(3,4),(4,3),(2,4),(3,3)$ and $(2,2)$, and that the estimated variance ratio is always equal to 0.25 . If we assume that the exact variance ratio in (10) is equal to 0.25 , then the mixing coefficient $\gamma$ of the Student bridge distribution $t_{n_{1}-1, n_{2}-1 ; \gamma}$ is accordingly equal to $0.706,0.840,0.750,0.889,0.8$ and 0.8 . Figure 5 shows the density of $t_{4,2 ; 0.706}$ and the density of its Welch approximation $t_{6}$ that can hardly be visually distinguished from each other if considered on the whole line, but differ locally. Figure 6 shows the densities of the Student bridge distribution $t_{1,3 ; 0.889}$ and Welch's Student approximation to it, $t_{1}$. In this case, preference for the Student bridge density can even be seen globally.


Figure 5. Densities of $t_{4,2 ; 0.706}$ and $t_{6}$ around the origin and in the distribution tails.


Figure 6. Densities of $t_{1,3 ; 0.889}$ and $t_{1}$.

Funding: This research received no external funding.
Acknowledgments: The author is grateful to the reviewers and the academic editor for their valuable comments and suggestions.

Conflicts of Interest: The author declares no conflict of interest.

## References

1. Schechtman, E.; Sherman, M. The two-sample $t$-test with a known ratio of variances. Stat. Methodol. 2007, 4, 508-514. [CrossRef]
2. Maity, A.; Sherman, M. The two-sample test with one variance unknown. Am. Stat. 2006, 60, 163-166. [CrossRef]
3. Behrens, W. Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. Landwirtsch. Jahrbuecher 1929, 68, 807-837.
4. Fisher, R. The fiducial argument in statistical inference. Ann. Eugen. 1935, 6, 391-398. [CrossRef]
5. Welch, B. The generalization of 'Student's' problem when several different population variances are involved. Biometrika 1947, 34, 28-35. [CrossRef] [PubMed]
6. Fisher, R.; Yates, F. Statistical Tables for Biological, Acricultural and Medical Research; Oliver Boyd: Edinburgh, UK, 1957.
7. Feiveson, A.; Delaney, F. The Distribution and Properties of a Weighted Sum of Chi Squares; NASA Technical Note TN D-4575; National Aeronautics and Space Administration: Washington, DC, USA, 1968.
8. Nel, D.G.; van der Merwe, C.A.; Moser, B.K. The exact distributions of the univariate and multivariate Behrens-Fisher statistics with a comparison of several solutions in the univariate case. Commun. Statist. Theory Meth. 1990, 19, 279-298. [CrossRef]
9. Nadarajah, S.; Li, R. Exact distribution of a modified Behrens-Fisher statistic. Commun. Stat. Simul. Comput. 2017, 46, 6845-6864. [CrossRef]
10. Linnik, J. Statistical Problems with Nuisance Parameters; American Mathematical Society: Providence, RI, USA, 1968.
11. Kim, S.H.; Cohen, A.S. On the Behrens-Fisher problem: A review. J. Educ. Behav. Stat. 1998, 23, 356-377. [CrossRef]
12. Dudewicz, E.J.; Ma, Y.; Mai, E.S.; Su, H. Exact solutions to the Behrens-Fisher problem: Asymptotically optimal and finite sample efficient choice among. J. Stat. Plan. Inference 2007, 137, 1584-1605. [CrossRef]
13. Giron, F.J.; Castillo, C. The multivariate Behrens-Fisher distribution. J. Multiv. Anal. 2010, 101, 2091-2102. [CrossRef]
14. Anderson, M.J.; Walsh, D.C.I.; Clarke, K.R.; Gorley, R.N.; Guerra-Castro, E. Some solutions to the multivariate Behrens-Fisher problem for dissimilarity-based analyses. Aust. N. Z. J. Stat. 2017, 59, 57-79. [CrossRef]
15. Shirke, D.T.; Khorate, S.D. Two-sample nonparametric test for testing equality of locations based on data depth. J. Indian Soc. Probab. Stat. 2018, 19, 9-23. [CrossRef]
16. Sprott, D.A.; Farewell, V.T. The difference between two normal means. Am. Stat. 1993, 47, 126-128.
17. Abramowitz, M.; Stegun, I. Handbook of Mathematical Functions; US Government Printing Office: Dover, NY, USA, 1965.
18. Gradshteyn, I.; Ryzhik, I. Table of Integrals, Series and Products; Academic Press: New York, NY, USA, 1980.
(C) 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
