Revista Colombiana de Estadística - Theoretical Statistics
July 2025, volume 48, issue 2, pp. 1–18
http://doi.org/10.15446/rce.v48n2.117675

Improving the Welch-Satterthwaite Approximation

Mejorando la aproximación de Welch-Satterthwaite

Carlos A. Cardozo[1,a]

[1]Department of Mathematics, Pontificia Universidad Javeriana, Bogotá, Colombia

Abstract

The Welch-Satterthwaite (WS) methodology is typically used in medicine, biology and economic courses to make inferences about the difference between two population means. Despite his wide-spreading applications, it has been pointing out in many references the multiple limitations of the inferences based on it. In this work, we propose three simple ways to improve the classical WS approach. Under balanced samples scenarios, we give exact inference results of two of the proposed estimators. Additionally, under unbalanced samples scenarios, we offer first-order approximation results and through several Monte Carlo simulations, we assess the mean and variance of the proposed estimators under (very) small and moderate sample sizes. Nonetheless, the simplicity of the proposed approach we obtain a much better performance than the WS proposal. Lastly, one application is presented in which the proposed estimators potentially improve the performance of t-student interval estimation and hypothesis testing procedures.

***Key words*:** Approximated inference; t-test; Generalized Gama Distribution; Delta method; Maximum Likelihood; Monte Carlo Simulation.

Resumen

La metodología de Welch-Satterthwaite (WS) se utiliza típicamente en medicina, biología y economía para realizar inferencias sobre la diferencia entre dos medias poblacionales. A pesar de su amplia aplicación, se ha señalado en numerosas referencias las múltiples limitaciones de las inferencias basadas en esta metodología. En este trabajo, proponemos tres maneras sencillas de mejorar el enfoque clásico de WS. En escenarios de muestras balanceadas, proporcionamos resultados de inferencia exactos de dos de los estimadores propuestos. Además, en escenarios con muestras no balanceadas, ofrecemos resultados de aproximación de primer orden y mediante simulación Monte Carlo, evaluamos la media y la varianza de los estimadores propuestos con tamaños de muestra (muy) pequeños y moderados. No obstante, gracias a la

[a]Ph.D. E-mail: carlosacardozod@javeriana.edu.co

simplicidad del enfoque propuesto, obtenemos un rendimiento mucho mejor que la propuesta de WS. Finalmente, se presenta una aplicación en la que los estimadores propuestos mejoran potencialmente el rendimiento de la estimación del intervalo t-Student y los procedimientos de prueba de hipótesis.

***Palabras clave*:** Inferencia aproximada; Prueba t; Distribución gamma generalizada; Método delta; Máxima verosimilitud; Simulación Monte Carlo.

1. The Problem and the WS Approach

Let $a_1, a_2, \dots, a_n$ positive real numbers and $Y_1, Y_2, \dots, Y_n$ are independent random variables where Y_i is distributed $\chi^2_{(r_i)}$ and $r_i \in \mathbb{Z}^+$. We would like to study the statistical distribution of $W = \sum_{i=1}^n a_i Y_i$.

1.1. WS Approach

Probably the most popular approach to study the distribution of W is debt to Satterthwaite (1946) and Welch (1947). His proposal was to build a variable $\widehat{W} = \frac{\chi^2_{\hat{\nu}}}{\hat{\nu}}$ where $\hat{\nu} = \frac{(\sum_{i=1}^n a_i Y_i)^2}{\sum_{i=1}^n \frac{a_i^2}{r_i} Y_i^2}$ such that $\sum_{i=1}^n a_i Y_i \approx \widehat{W}$. According to Casella & Berger (2003) the method consists in the following

$$E[W^2] = E\Big[\sum_{i=1}^n a_i Y_i\Big]^2 = Var\Big[\sum_{i=1}^n a_i Y_i\Big] + \Big(E\Big[\sum_{i=1}^n a_i Y_i\Big]\Big)^2.$$

Equivalently, we can say that

$$E\Big[\sum_{i=1}^n a_i Y_i\Big]^2 = \Big(E\Big[\sum_{i=1}^n a_i Y_i\Big]\Big)^2 \Big[\frac{Var\Big[\sum_{i=1}^n a_i Y_i\Big]}{\Big(E\Big[\sum_{i=1}^n a_i Y_i\Big]\Big)^2} + 1\Big],$$

but,

$$E\Big[\sum_{i=1}^n a_i Y_i\Big] \approx E[\widehat{W}] = 1,$$

so we have that

$$E\Big[\sum_{i=1}^n a_i Y_i\Big]^2 = \Big[\frac{Var\Big[\sum_{i=1}^n a_i Y_i\Big]}{\Big(E\Big[\sum_{i=1}^n a_i Y_i\Big]\Big)^2} + 1\Big].$$

In this point, if we equates the second ordinary moments of W and $\widehat{W}$ then we obtain that

$$\nu = \frac{2\Big(E\Big[\sum_{i=1}^n a_i Y_i\Big]\Big)^2}{Var\Big[\sum_{i=1}^n a_i Y_i\Big]}$$

And, we know that $Var\Big[\sum_{i=1}^n a_i Y_i\Big] = 2\sum_{i=1}^n a_i^2 \frac{E[Y_i]^2}{r_i}$ because $Y_i \sim \chi^2_{(r_i)}$, $i = 1, 2, \ldots, n$ and $Y_i's$ are independent. Therefore, the degrees of freedom of $\widehat{W}$ after removing the expectations is

$$\hat{\nu} = \frac{(\sum_{i=1}^n a_i Y_i)^2}{\sum_{i=1}^n \frac{a_i^2}{r_i} Y_i^2}.$$

Now, the WS approximation is frequently used today when we want to make inferences about the difference of two population means from normal distributions with unknown population variances. Indeed, many traditional books of statistical inference, Casella & Berger (2003), Walpole et al. (2017), and Hogg et al. (2019), the WS approach is the only form to do it. However, there are multiple limitations of the inferences based on the WS approximation. For example, there is a poor performance of the WS approach when n_i are small in the two population problem. Also, if at the same time we have that the quotient of the two population variances is bigger then the performance of $\widehat{W}$ can be even worst and paradoxical, see for instance Hall & Willink (2001), Ballico (2000), and Crowder & Kupferman (2004). Miao & Chiou (2008) study and compare three different confidence interval procedures for the difference of two population means and give evidence that if both the underlying distributions are skewed and the homogeneity of variances assumption is also violated, an interval based on the WS procedure has much lower coverage probability than the nominal level. On the other hand, there are some works that claim exact inferences, for instance, Xiao (2018) is technically inaccessible and too long to an undergraduate audience (more than 110 pages!). All the previous reasons give us the motivation to seek a more accessible and accurate methodology than the WS. This work is organized as follows. In Section 2, we give three (simple) novel ways to approximate the distribution of W and some theoretical properties of them. A Monte Carlo simulation study is presented in Section 3. In Section 4, we offer one application of our methodologies. Finally, we give some conclusions and future lines in Section 5.

2. Improving the WS Approach

We define $X_1, \ldots, X_n$ as independent random variables where X_i is distributed $Gamma(\frac{r_i}{2}, \frac{\beta}{a_i})$. We approximate the statistical distribution of W by the distribution of $V = \sum_{i=1}^n a_i X_i$. In the following we present some theoretical implications of this approach.

Proposition 1. *The statistical distribution of V is $Gamma(\frac{1}{2}\sum_{i=1}^n r_i, \beta)$.*

This property is a consequence of elementary properties from Gamma variables. The complete specification of the distribution of V depends on the value β. But, this is an unknown parameter so we would like to establish it based on data. Now, we offer three different ways to do it.

2.1. A Maximum Likelihood Estimator of β

Using the maximum likelihood methodology on the random variables $X_1, \dots, X_n$ we will obtain our first estimator for β which we will denote as $\hat{\beta}_{ML}$,

$$\hat{\beta}_{ML} = 2\frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} r_i}. \tag{1}$$

We denote the associated V variable as

$$V_{ML} \sim Gamma\Big(\frac{\sum_{i=1}^{n} r_i}{2}, \hat{\beta}_{ML}\Big). \tag{2}$$

Also, we introduce two more estimators of β. We will see in the next sections that they are better options than the $\hat{\beta}_{ML}$ estimator under some scenarios.

Definition 1. We define $\hat{\beta}_1$ and $\hat{\beta}_2$ estimators of β and their associated V variables

1. $\hat{\beta}_1 = 2\frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} Y_i}$.
2. $\hat{\beta}_2 = 2\sqrt{\frac{\sum_{i=1}^{n} a_i^2 Y_i}{\sum_{i=1}^{n} Y_i}}$.
3. $V_1 \sim Gamma(\frac{\sum_{i=1}^{n} r_i}{2}, \hat{\beta}_1)$
4. $V_2 \sim Gamma(\frac{\sum_{i=1}^{n} r_i}{2}, \hat{\beta}_2)$.

We note that V_{ML}, V_1, and V_2 do not have necessary the same statistical distribution and that their distributions are conditional to the sample values of $y_1, \dots, y_n$. All the details about the deduction of $\hat{\beta}_{ML}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are in the Appendix A.

2.2. Balanced Case

Our next step is to study the distributions of V_{ML}, V_1, and V_2 in a particular situation of interest in real scenarios, for instance, balanced design experiments.

Property 3. If $a_1 = a_2 = \cdots = a_n = a$, we have that

1. $\hat{\nu} = \frac{(\sum_{i=1}^{n} Y_i)^2}{\sum_{i=1}^{n} \frac{Y_i}{r_i}}$.
2. $V_{ML} \sim Gamma(\frac{\sum_{i=1}^{n} r_i}{2}, \hat{\beta}_1)$ where $\hat{\beta}_{ML} = 2a\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} r_i}$.
3. $V_1 \sim Gamma(\frac{\sum_{i=1}^{n} r_i}{2}, \hat{\beta}_1)$ where $\hat{\beta}_1 = 2a$.
4. $V_2 \sim Gamma(\frac{\sum_{i=1}^{n} r_i}{2}, \hat{\beta}_2)$ where $\hat{\beta}_2 = 2a$.

Actually, in this case W is distributed as a $a\chi^2_{(\sum_{i=1}^{n} r_i)}$. Hence, W, V_1, and V_2 are all equally distributed. However, this is not necessarily the case of the $\widehat{W}$

and V_{ML}. Additionally, we obtain simple closed-form expressions for all the raw moments of W.

Property 4. Under the same assumptions of the Property 3, the expectation and variance of W are

1. $E[W] = a\sum_{i=1}^{n} r_i$.
2. $Var[W] = 2a^2\sum_{i=1}^{n} r_i$.

In fact, we knew them from the definition of W and properties of the expectation and variance. But, we could say more about the higher order moments of W through the generating function of V_1 or V_2,

$$M_W(t) = \left[\frac{1}{1-2at}\right]^{\frac{\sum_{i=1}^{n} r_i}{2}}, \qquad t < \frac{1}{2a}.$$

Hence, for $1 \leq m$

$$E[W^m] = \frac{(2a)^m\Gamma(\frac{\sum_{i=1}^{n} r_i}{2} + m)}{\Gamma(\frac{\sum_{i=1}^{n} r_i}{2})} = a^m \prod_{j=0}^{m-1}\Big(\sum_{i=1}^{n} r_i - 2j\Big)$$

2.3. Unbalanced Case

If $a_1, a_2, \ldots, a_n$ are not necessary equal we can compute the exact values of the expectation and variance of V_{ML}. On the other hand, we can give only approximate values for the expectation and variance of V_1, and V_2 based on the Delta method. First, we need to establish a few preliminary results.

Property 5.a The expectation and variance of $\hat{\beta}_{ML}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are

1. $E[\hat{\beta}_{ML}] = 2\frac{\sum_{i=1}^{n} a_i r_i}{\sum_{i=1}^{n} r_i}$
2. $Var[\hat{\beta}_{ML}] = 8\frac{\sum_{i=1}^{n} a_i^2 r_i}{(\sum_{i=1}^{n} r_i)^2}$
3. $E[\hat{\beta}_1] \approx 2\frac{\sum_{i=1}^{n} a_i r_i}{\sum_{i=1}^{n} r_i}$.
4. $Var[\hat{\beta}_1] \approx 8\frac{(\sum_{i=1}^{n} a_i r_i)^2}{(\sum_{i=1}^{n} r_i)^2}\Big(\frac{\sum_{i=1}^{n} a_i^2 r_i}{(\sum_{i=1}^{n} a_i r_i)^2} - \frac{1}{\sum_i r_i}\Big)$.
5. $E[\hat{\beta}_2] \approx \sqrt{2\frac{\sum_{i=1}^{n} a_i^2 r_i}{\sum_{i=1}^{n} r_i}}$.
6. $Var[\hat{\beta}_2] \approx 2\Big(\frac{\sum_{i=1}^{n} a_i^2 r_i}{\sum_{i=1}^{n} r_i}\Big)\Big(\frac{\sum_{i=1}^{n} a_i^4 r_i}{(\sum_{i=1}^{n} a_i^2 r_i)^2} - \frac{2}{\sum_{i=1}^{n} r_i}\Big)$.

The results 5.a.3-6 are based on first-order approximations. Now, using Property 5.a we can give approximate expressions about the conditional variances of V_1 and V_2.

Property 5.b The expectation and variance of V_{ML}, V_1 and V_2 are

1. $E[V_{ML}] = \sum_{i=1}^{n} a_i r_i$.
2. $Var[V_{ML}] = 2\frac{(\sum_{i=1}^{n} a_i Y_i)^2}{\sum_{i=1}^{n} r_i}$.
3. $E[V_1] \approx \sum_{i=1}^{n} a_i r_i$.
4. $Var[V_1|Y_1,\ldots,Y_n] \approx 2\Big(\sum_{i=1}^{n} r_i\Big)\Big(\frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} Y_i}\Big)^2$.
5. $E[V_2] \approx 2\sqrt{\frac{\sum_{i=1}^{n} a_i^2 r_i}{\sum_{i=1}^{n} r_i}}$.
6. $Var[V_2|Y_1,\ldots,Y_n] \approx 2\Big(\sum_{i=1}^{n} r_i\Big)\Big(\frac{\sum_{i=1}^{n} a_i^2 Y_i}{\sum_{i=1}^{n} Y_i}\Big)$.
7. $\frac{Var[V_1|Y_1,\ldots,Y_n]}{Var[V_2|Y_1,\ldots,Y_n]} \leq 1$.

Again, results 5.b. 3-7 are based on first-order approximations. The item 5.b.1 means that V_{ML} has exactly the same mean of W. The item 5.b.3 means that V_1 has approximately the same mean of W. And 5.b.7 means that V_1 can be more precise than V_2.

3. Monte Carlo Studies

We evaluate the performance of our proposals when $a_1, a_2, \ldots, a_n$ are not necessarily equal by doing a simulation study. The simulation plan consists of 16 scenarios. We take n_1 and $n_2 = n_1 + 2 \times j$ where $n_1 = 3, 5, 7, 9$ and $j = 1, 2, 3, 4$. For each fixed pair (n_1, n_2), we simulate $Y_1 \sim \chi^2_{(n_1-1)}$ and $Y_2 \sim \chi^2_{(n_2-1)}$ and compute $W = \frac{1}{n_1}Y_1 + \frac{1}{n_2}Y_2$. We take the number of replications $R = 15000$. Finally, for each replication, we calculate $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$. From Section 1, we have that the conditional expectation and variance of $\widehat{W}$ are

$$E[\widehat{W}|Y_1,\ldots,Y_n] = 1 \text{ and } Var[\widehat{W}|Y_1,\ldots,Y_n] = \frac{2}{\hat{\nu}}$$

In Tables 1 and 2, for each scenario, we present the mean and variance of target variable W, the mean and variance of $\widehat{W}$ in the mean value of $\hat{\nu}$, the mean and variance of $\hat{V}_{ML}$ in the mean value of $\hat{\beta}_{ML}$, the mean and variance of $\hat{V}_1$ in the mean value of $\hat{\beta}_1$, and the mean and variance of $\hat{V}_2$ in the mean value of $\hat{\beta}_2$. All calculations were performed using the statistical language and ambient R Development Core Team (2024) and all the codes are available on request.

3.1. Advantages of the Proposals Based on the Empirical Data

From the simulation results, we can say that

- The performance of $\widehat{W}$ was the poorest, in all the situations that we considered, as we see in the histograms given in Figures 1-8.
- The performance of $\hat{V}_{ML}$ and $\hat{V}_1$ was very similar, that is, their empirical distributions were quite close between them.
- On the other hand, the variance of $\hat{V}_2$ was clearly the closest to the variance of W.
- In all scenarios, the distribution of $\hat{V}_2$ was also very close to the expectation of W.

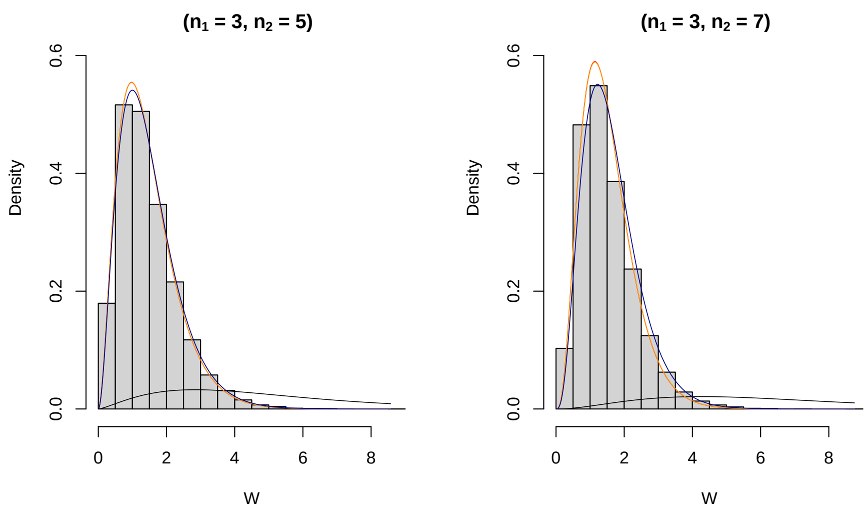

FIGURE 1: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 3, n_2 = 5, 7$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

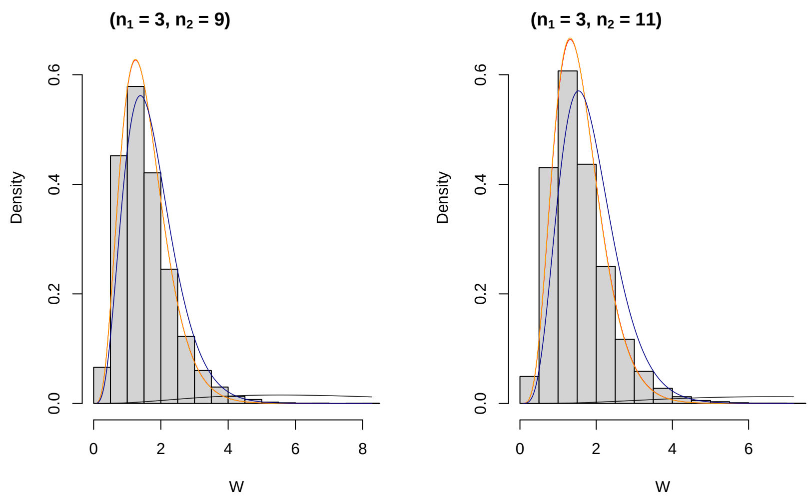

Figure 2: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 3, n_2 = 9, 11$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

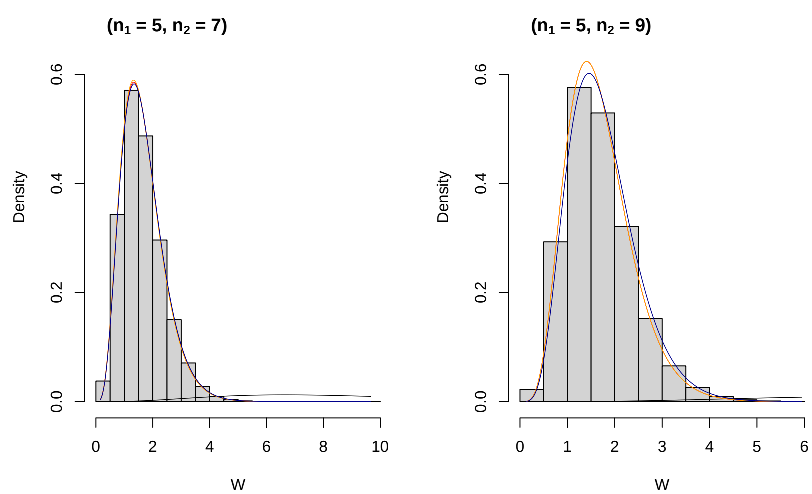

Figure 3: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 5, n_2 = 7, 9$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

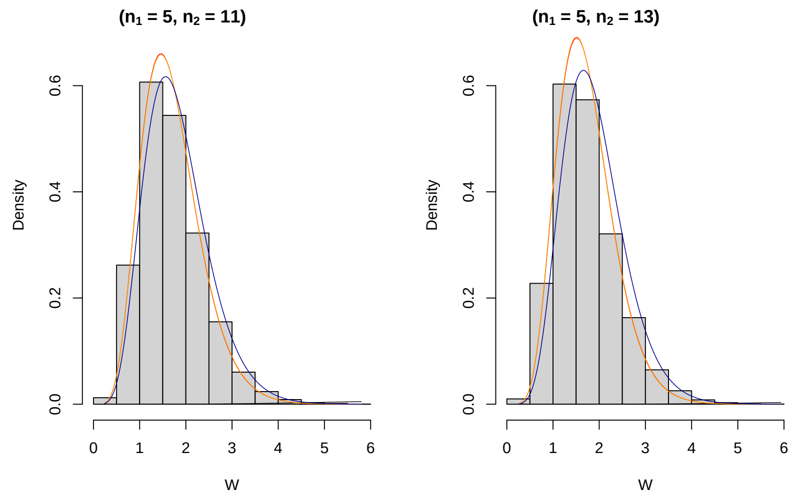

FIGURE 4: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 5, n_2 = 11, 13$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

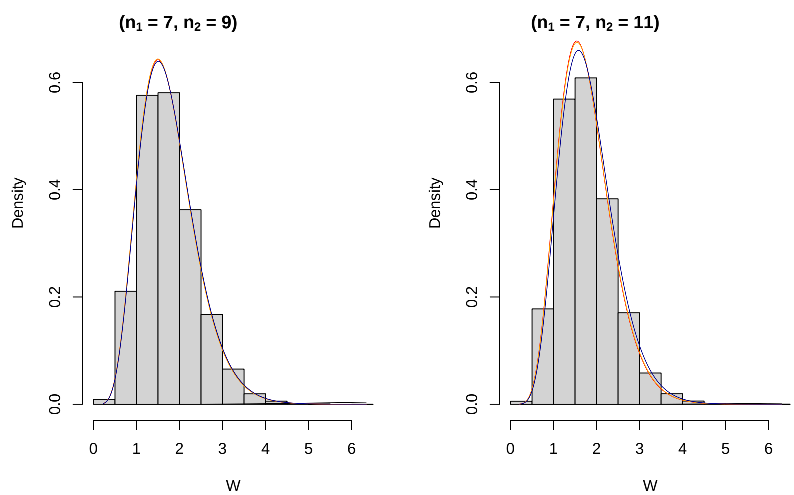

FIGURE 5: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 7, n_2 = 9, 11$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

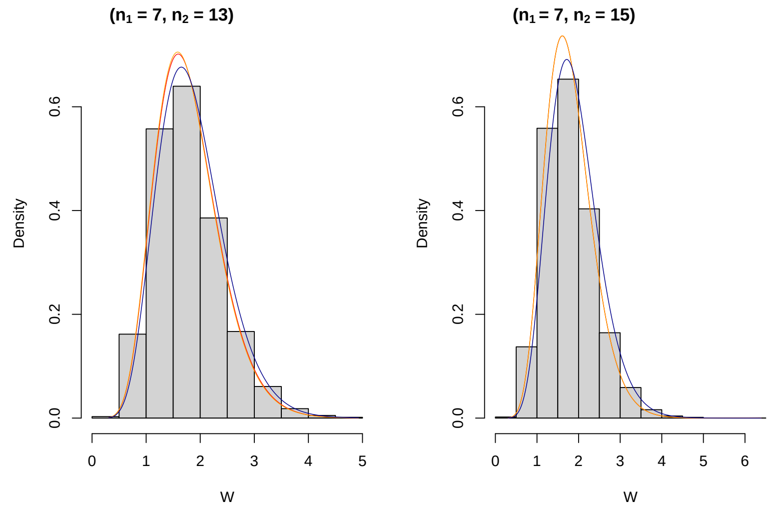

FIGURE 6: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 7, n_2 = 13, 15$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

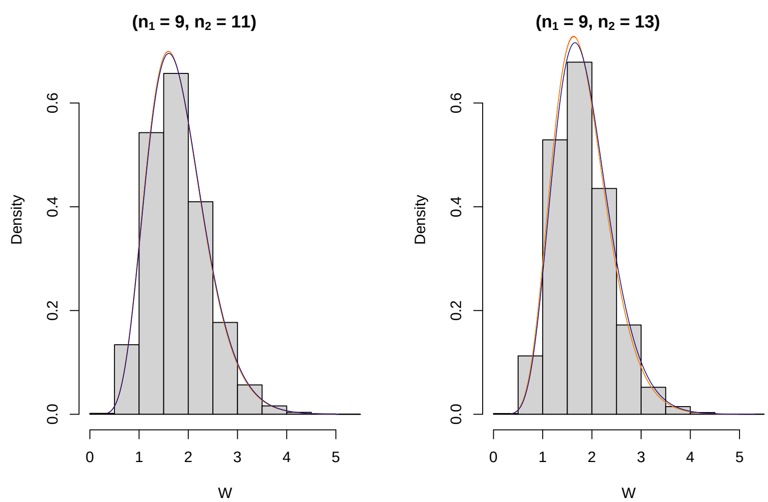

FIGURE 7: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 9, n_2 = 11, 13$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

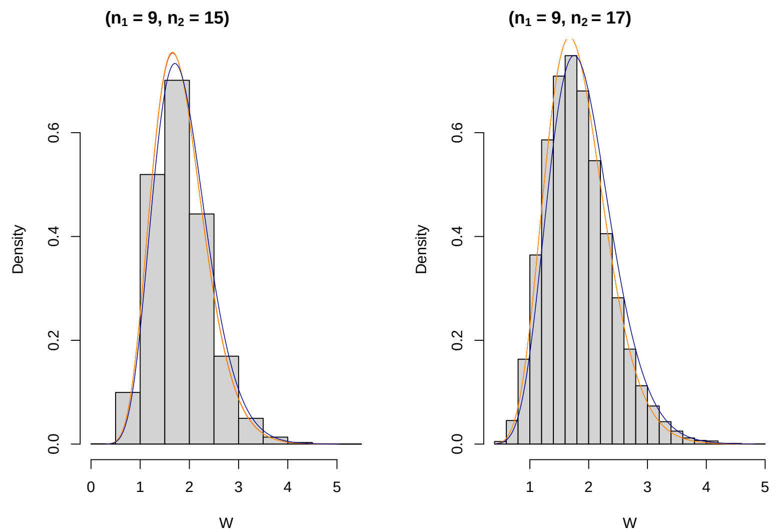

FIGURE 8: Histograms of the empirical values of W based on the pairs (n_1, n_2) where $n_1 = 9, n_2 = 15, 17$. The red, orange, blue and black densities curves are based on the probability density functions given by the Equation (2), Definition 1.3 and Definition 1.4, respectively. The scale parameters of these Gamma distributions are the mean values of $\hat{\beta}_{ML}, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\nu}$ estimations, respectively.

3.2. Limitations of the Proposals

However, based on empirical evidence from the Monte Carlo simulation we can say that for all values of n_1 when the value of n_2 increases $(1 < \frac{n_2}{n_1} < 4)$

- The ratio between the variance of $\hat{V}_1$ and the variance of the target variable W decreases. The same remark applies to the estimator $\hat{V}_{ML}$. So, we believe that this kind of behavior will be the same (or even worse than we had obtained) if we consider sample sizes where $4 \leq \frac{n_2}{n_1}$.

- The difference between the expectation of $\hat{V}_2$ and the expectation of W increases. Hence, our intuition is that this difference will be greater if we consider sample sizes where $4 \leq \frac{n_2}{n_1}$.

Table 1: The mean and variance of W, $\widehat{W}$, $\hat{\beta}_{ML}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ under the different simulation scenarios. The bold values correspond to the the closest values to the mean and variance of W.

		$n_2 = 5$		$n_2 = 7$		$n_2 = 9$		$n_2 = 11$	
		Mean	Var	Mean	Var	Mean	Var	Mean	Var
$n_1 = 3$	**W**	1.467	0.764	1.524	0.689	1.555	0.642	1.576	0.608
	$\widehat{W}$	1	0.416	1	0.321	1	0.266	1	0.231
	V_{ML}	1.473	0.724	**1.523**	0.580	1.566	0.491	1.567	0.413
	V_1	**1.466**	0.716	**1.525**	0.581	**1.558**	0.486	**1.573**	0.412
	V_2	1.501	**0.751**	1.628	**0.663**	1.742	**0.607**	1.839	**0.564**
		$n_2 = 7$		$n_2 = 9$		$n_2 = 11$		$n_2 = 13$	
$n_1 = 5$	**W**	1.657	0.574	1.688	0.518	1.709	0.485	1.723	0.462
	$\widehat{W}$	1	0.232	1	0.198	1	0.175	1	0.158
	V_{ML}	1.663	0.553	1.693	0.477	1.711	0.418	**1.723**	0.371
	V_1	**1.655**	0.548	**1.688**	0.475	**1.709**	0.417	**1.723**	0.371
	V_2	1.674	**0.561**	1.750	**0.511**	1.824	**0.475**	1.895	**0.448**

Table 2: The mean and variance of W, $\widehat{W}$, $\hat{\beta}_{ML}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ under the different simulation scenarios. The bold values correspond to the the closest values to the mean and variance of W.

		$n_2 = 9$		$n_2 = 11$		$n_2 = 13$		$n_2 = 15$	
		Mean	Var	Mean	Var	Mean	Var	Mean	Var
$n_1 = 7$	**W**	1.746	0.442	1.766	0.410	1.780	0.387	1.790	0.369
	$\widehat{W}$	1	0.160	1	0.142	1	0.129	1	0.119
	V_{ML}	1.751	0.438	1.769	0.391	**1.778**	0.351	1.775	0.394
	V_1	**1.747**	0.436	**1.767**	0.390	**1.778**	0.353	**1.787**	0.319
	V_2	1.759	**0.442**	1.806	**0.408**	1.854	**0.382**	1.905	**0.363**
		$n_2 = 11$		$n_2 = 13$		$n_2 = 15$		$n_2 = 17$	
$n_1 = 9$	**W**	1.798	0.363	1.812	0.340	1.822	0.322	1.830	0.308
	$\widehat{W}$	1	0.122	1	0.111	1	0.103	1	0.096
	V_{ML}	1.792	0.357	1.814	0.329	**1.823**	0.302	1.826	0.278
	V_1	**1.798**	0.359	**1.812**	0.328	**1.823**	0.302	**1.830**	0.279
	V_2	1.806	**0.362**	1.839	**0.338**	1.877	**0.320**	1.916	**0.306**

4. Applications

Now, we apply the previous properties to the problem of finding an interval estimate of the difference of two population means, $\mu_1 - \mu_2$, of two (approximately) normal distributions under unknown and unequal variances. For instance, in Walpole et al. (2017) we can find the traditional way to achieve this task which is considering the statistic

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{S_1^2}{n_1}\right) + \left(\frac{S_2^2}{n_2}\right)}}.$$

According to the WS approach, we can build an approximate confidence interval of $\mu_1 - \mu_2$, based on a t-student distribution, t_ν, where

$$\nu = \frac{\left[(\frac{S_1^2}{n_1}) + (\frac{S_1^2}{n_2})\right]^2}{\left[\left(\frac{S_1^2}{n_1}\right)^2/(n_1-1)\right] + \left[\left(\frac{S_1^2}{n_1}\right)^2/(n_2-1)\right]}.$$

Using the previous properties, we can approximate the distribution of the denominator of T in at least three forms.

4.1. Approximation of T using the $\hat{\beta}_{ML}$ Estimator

Property 6.a The denominator of T is approximately distributed as

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \approx \tilde{V}_{ML},$$

where

$$\tilde{V}_{ML} \sim \sqrt{Gamma\Big(\frac{1}{2}(n_1+n_2-2), \hat{\beta}_{ML}\Big)} \text{ and } \hat{\beta}_{ML} = \frac{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}{n_1+n_2-2}.$$

4.2. Approximation of T using the $\hat{\beta}_1$ Estimator

Property 6.b The denominator of T is approximately distributed as

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \approx \tilde{V}_1,$$

where

$$\tilde{V}_1 \sim \sqrt{Gamma\Big(\frac{1}{2}(n_1+n_2-2), \hat{\beta}_1\Big)} \text{ and } \hat{\beta}_1 = \frac{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}{S_1^2 + S_2^2}.$$

4.3. Approximation of T using $\hat{\beta}_2$ Estimator

Property 6.c The denominator of T is approximately distributed in the second form below.

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \approx \tilde{V}_2,$$

where

$$\tilde{V}_2 \sim \sqrt{Gamma\Big(\frac{1}{2}(n_1+n_2-2), \hat{\beta}_2\Big)} \text{ where } \hat{\beta}_2 = 2\sqrt{\frac{\left(\frac{S_1}{n_1}\right)^2 + \left(\frac{S_2}{n_2}\right)^2}{S_1^2 + S_2^2}}.$$

Based on Cardozo et al. (2022), we can say a little more about the statistical distribution of $\tilde{V}_{ML}$, $\tilde{V}_1$, and $\tilde{V}_2$.

Property 7.a The distribution of $\tilde{V}_{ML}$ is a Generalized Gamma, $GG(\eta_{ML}, \sigma_{ML}, \lambda_{ML})$ where

$$\eta_{ML} = \sqrt{\frac{1}{2}\hat{\beta}_{ML}(n_1+n_2-2)}, \sigma_{ML} = \frac{1}{\sqrt{2(n_1+n_2-2)}}, \lambda_{ML} = \frac{\sqrt{2}}{\sqrt{n_1+n_2-2}}.$$

Property 7.b The distribution of $\tilde{V}_1$ is a Generalized Gamma, $GG(\eta_1, \sigma_1, \lambda_1)$ where

$$\eta_1 = \sqrt{\frac{1}{2}\hat{\beta}_1(n_1+n_2-2)}, \sigma_1 = \frac{1}{\sqrt{2(n_1+n_2-2)}}, \lambda_1 = \frac{\sqrt{2}}{\sqrt{n_1+n_2-2}}.$$

Property 7.c The distribution of $\tilde{V}_2$ is a Generalized Gamma, $GG(\eta_2, \sigma_2, \lambda_2)$ where

$$\eta_2 = \sqrt{\frac{1}{2}\hat{\beta}_2(n_1+n_2-2)}, \sigma_2 = \frac{1}{\sqrt{2(n_1+n_2-2)}}, \lambda_2 = \frac{\sqrt{2}}{\sqrt{n_1+n_2-2}}.$$

If we put a lower bound on n_1+n_2, we can simplify the distributions of $\tilde{V}_1, \tilde{V}_2$ and $\tilde{V}_{ML}$ in the following way.

4.4. The LogNormal Distribution Approximation

Property 7.d Let assume that $4 \leq n_1, n_2$ and $10 \leq n_1+n_2$ then

$$\tilde{V}_{ML} \sim LN(\ln(\eta_{ML}), \sigma_{ML})$$

$$\tilde{V}_1 \sim LN(\ln(\eta_{ML}), \sigma_1)$$

$$\tilde{V}_2 \sim LN(\ln(\eta_{ML}), \sigma_2)$$

where $LN(\mu, \sigma)$ denotes the distribution of a positive variable Y such that $X = ln(Y)$ is distributed $N(\mu, \sigma^2)$.

The results in Property 7.d are approximate distributions, but very accurate; see Figure 9. The essential reason behind it is the following auxiliary result. Let Y_λ be a random variable with $GG(\eta, \sigma, \lambda)$ and Y a random variable with $LN(\eta, \sigma)$. Hence, Y_λ converges in distribution to Y when $\lambda \to 0$. As $0 < \lambda \leq 0.5$ the distance between the density of Y_λ and Y is quite small, that is, $|f_{Y_\lambda} - f_Y| = sup_{t\in R^+}|f_{Y_\lambda}(t) - f_Y(t)|$ is small. We can find a proof of this fact, for instance, in Lawless (1980). Therefore, if $10 \leq n_1+n_2$ then λ_{ML}, λ_1 and λ_2 are less or equal to 0.5. Additionally, σ_{ML}, σ_1 and σ_2 are less than 0.25 which implies a very symmetric lognormal densities.

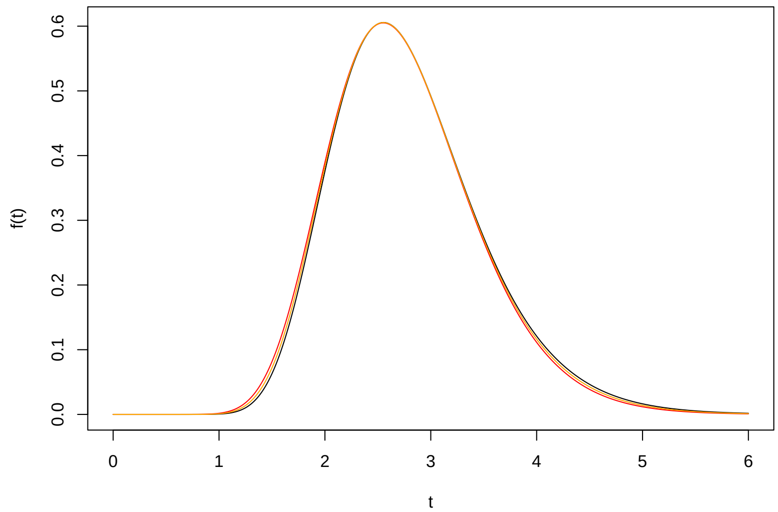

FIGURE 9: The black curve is the density of a lognormal distribution with parameters $\mu = 1, \sigma = 0.25$. The red and orange curves are based on the generalized gamma density functions with common $\eta = e, \sigma = 0.25$ and $\lambda = 0.5$ and $\lambda = 0.25$ respectively.

5. Conclusions

We show that there are important practical cases where two of our proposals obtain exact inference results no matter how many populations we are considering in the original problem. The offered approaches are relatively as simple of WS approach but with much better results. We also characterized approximately, in two ways, the distribution of the denominator of the $t-$student statistic using the generalized gamma distribution. Finally, we think that our proposals opens the possibility to a Bayesian approach.

Acknowledgment

The author would like to express their gratitude to the two anonymous referees for their comments. A special thanks to Luis Fransisco Gómez who pointed out the reference Xiao (2018).

[Received: November 2024 — Accepted: June 2025]

References

Ballico, M. (2000), 'Limitations of the welch–satterthwaite approximation for measurement uncertainty calculations', *Metrologia* **37**(1), 61–64.

Cardozo, C., Paula, G. & Vanegas, L. (2022), 'Generalized log-gamma additive partial linear models with P-spline smoothing', *Statistical Papers* **63**, 1953–1978.

Casella, G. & Berger, B. (2003), *Statistical Inference*, Duxbury-Thomson Learning.

Crowder, S. & Kupferman, S. (2004), 'Use of Welch–Satterthwaite approximation in calibration of voltage standards', *Journal of Quality Technology* **36**(1), 38–52.

Hall, B. D. & Willink, R. (2001), 'Does Welch–Satterthwaite make a good uncertainty estimate?', *Metrologia* **38**, 9–15.

Hogg, R., McKean, J. & Craig, A. (2019), *Introduction to Mathematical Statistics*, Pearson Education.

Lawless, J. (1980), 'Inference in the generalized gamma and log gamma distributions', *Technometrics* **22**, 409–419.

Miao, W. & Chiou, P. (2008), 'Confidence intervals for the difference between two means', *Computational Statistics & Data Analysis* **52**, 2238–2248.

R Development Core Team (2024), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Satterthwaite, F. E. (1946), 'An approximate distribution of estimates of variance components', *Biometrics Bulletin* **2**, 110–114.

Walpole, R. E., Myers, R. H., Myers, S. L. & Ye, K. (2017), *Probability and Statistics for Engineers and Scientists*, 9 edn, Pearson Education.

Welch, B. L. (1947), 'The generalization of Student's problem when several different population variances are involved', *Biometrika* **34**, 28–35.

Xiao, Y. (2018), 'On the solution of a generalized Behrens–Fisher problem', *Far East Journal of Theoretical Statistics* **54**, 21–140.

Appendix A.

In this section we give more details about some theoretical statements of the Section 2.

Appendix A.1. The maximum likelihood estimator of β

We assume that $X_1, X_2, \ldots, X_n$ is a set of independent random variables where X_i is distributed $Gamma(\frac{r_i}{2}, \frac{\beta}{a_i})$. Hence, we have that the likelihood function of β based on the scaled variables $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ where $\tilde{X}_i = a_i X_i$ for $i = 1, 2, \ldots, n$ is

$$L(\beta; \tilde{x}_1, \ldots, \tilde{x}_n) = \prod_{i=1}^{n} \Big[\frac{1}{\Gamma(\frac{r_i}{2}) \beta^{\frac{r_i}{2}}} \tilde{x}_i^{\frac{r_i}{2}-1} e^{-\frac{\tilde{x}_i}{\beta}} \Big] \tag{A1}$$

Thus, the log-likelihood function of β is

$$\log(L(\beta;\tilde{x}_1,\ldots,\tilde{x}_n)) = \sum_{i=1}^{n}\Big[-\log\Big(\Gamma(\frac{r_i}{2})\Big) - \frac{r_i}{2}log(\beta) + \Big(\frac{r_i}{2}-1\Big)\log(\tilde{x}_i) - \frac{\tilde{x}_i}{\beta}\Big] \quad \text{(A2)}$$

After some algebraic steps in the Equation (A2) we have that

$$\log(L(\beta;\tilde{x}_1,\ldots,\tilde{x}_n)) = -\sum_{i=1}^{n}\log\Big(\Gamma(\frac{r_i}{2})\Big) - \log(\beta)\sum_{i=1}^{n}\frac{r_i}{2} + \sum_{i=1}^{n}\Big(\frac{r_i}{2}-1\Big)\log(\tilde{x}_i) - \frac{1}{\beta}\sum_{i=1}^{n}\tilde{x}_i \quad \text{(A3)}$$

But, we note that in the Equation (A3) the terms

$$-\sum_{i=1}^{n}\log\Big(\Gamma(\frac{r_i}{2})\Big) \text{ and } \sum_{i=1}^{n}\Big(\frac{r_i}{2}-1\Big)\log(\tilde{x}_i)$$

do not depend of the parameter β. Hence, the log-likelihood function of β is proportional to

$$\log(L(\beta;\tilde{x}_1,\ldots,\tilde{x}_n)) \propto -\log(\beta)\sum_{i=1}^{n}\frac{r_i}{2} - \frac{1}{\beta}\sum_{i=1}^{n}\tilde{x}_i. \quad \text{(A4)}$$

We can equivalently maximize the right side of Equation (A4). So, we define the function

$$h(\beta) = -\log(\beta)\sum_{i=1}^{n}\frac{r_i}{2} - \frac{1}{\beta}\sum_{i=1}^{n}\tilde{x}_i. \quad \text{(A5)}$$

Now, we can see that $h(\beta)$ is a differentiable function in the interval $(0,+\infty)$ and

$$h'(\beta) = -\frac{1}{\beta}\sum_{i=1}^{n}\frac{r_i}{2} + \frac{1}{\beta^2}\sum_{i=1}^{n}\tilde{x}_i. \quad \text{(A6)}$$

The Equation (A6) is equal to zero when $\beta = 2\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} r_i}$. Therefore, the maximum likelihood estimator of β is

$$\hat{\beta}_{ML} = 2\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} r_i}. \quad \text{(A7)}$$

Appendix A.2. The Case of $\hat{\beta}_1$

In the case of the $\hat{\beta}_1$ and $\hat{\beta}_2$ we will use a moments estimation strategy. We know that $E[W] = \sum_{i=1}^{n} a_i r_i$ and $V \sim Gamma(\frac{1}{2}\sum_{i=1}^{n} r_i, \beta)$ (by the **Property**

1). So, we can make equal the first moments of V and W

$$E[V] = \Big(\frac{1}{2}\sum_{i=1}^{n} r_i\Big)\beta = E[W] = \sum_{i=1}^{n} a_i r_i. \tag{A8}$$

Therefore, using (A8) we have that

$$\hat{\beta}_1 = 2\frac{\sum_{i=1}^{n} a_i r_i}{\sum_{i=1}^{n} r_i}. \tag{A9}$$

But, we prefer to work with the non-constant version of $\hat{\beta}_1$

$$\hat{\beta}_1 = 2\frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} Y_i}, \tag{A10}$$

using in (A10) the fact that $E[Y_i] = r_i$ because $Y_i \sim \chi^2_{(r_i)}$ for $i = 1, \dots, n$.

Appendix A.3. The Case of $\hat{\beta}_2$

Lastly, we also know by the definition of the random variables W and V (**Section 1** and **Property 1**) that

$$Var[W] = 2\sum_{i=1}^{n} a_i^2 r_i,$$

and

$$Var[V] = \Big(\frac{1}{2}\sum_{i=1}^{n} r_i\Big)\beta^2.$$

Hence, if we make equal $Var[W]$ and $Var[V]$ then we get that

$$\Big(\frac{1}{2}\sum_{i=1}^{n} r_i\Big)\beta^2 = 2\sum_{i=1}^{n} a_i^2 r_i, \tag{A11}$$

and solving for β in (A11) we obtain that

$$\hat{\beta}_2 = 2\sqrt{\frac{\sum_{i=1}^{n} a_i^2 r_i}{\sum_{i=1}^{n} r_i}}.$$

But, we again prefer to work with the non-constant version of $\hat{\beta}_2$.

$$\hat{\beta}_2 = 2\sqrt{\frac{\sum_{i=1}^{n} a_i^2 Y_i}{\sum_{i=1}^{n} Y_i}}. \tag{A12}$$