Some simple approaches to improve the Welch- Satterthwaite approximation

Carlos Alberto Cardozo Delgado*
Departamento de Matemáticas, Pontificia Universidad Javeriana
and
Luis Francisco Gómez López
Administración de Empresas, Universidad Militar Nueva Granada

June 26, 2024

Abstract

In this work we propose three simple ways to improve the classical Welch-Satterthwaite (WS) approximation to the effective degree of freedom of a non-negative linear combination of χ^2 distributions. The WS option is typically used in design of experiments and metrology. However, it has been pointing out in many references the multiple limitations of the inferences based on the WS approximation. Three novel estimators of the effective degree of freedom of a non-negative linear combination of χ^2 distributions are given. We also study some theoretical properties of the proposed estimators. Additionally, through several Monte Carlo simulations, we assess the bias, variance, and mean square error of the proposed estimators under (very) small and moderate sample sizes. The proposed estimators have a much better performance than the WS proposal and his implications. Finally, two applications are presented in which the proposed estimators help to improve the performance of some interval estimation and hypothesis testing procedures.

Keywords: Asymmetric distributions, Unbiased estimation, Consistent estimation, Monte Carlo simulation

*The authors gratefully acknowledge *please remember to list all relevant funding sources in the non-anonymized (unblinded) version.*

1 Introduction

2 Critical aspects of the WS approach

Lloyd (2013) conducted three numerical experiments highlighting potential limitations of the two-sample t-test when using the Welch-Satterthwaite (WS) approximation. Notably, in their first experiment, Lloyd (2013) assumed that the ratio of population variances was known. However, in practice, this information is rarely available. Therefore, we will replicate this simulation without assuming a known variance ratio, providing a more realistic assessment of the two-sample t-test's performance under the WS approximation.

We now outline the initial simulation plan associated with the two-sample t-test statistic, $T = \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\widehat{\sigma}_1^2}{n_1}+\frac{\widehat{\sigma}_2^2}{n_2}}}$. Let $X_1, X_2, \ldots, X_{n_1} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_1, Y_2, \ldots, Y_{n_2} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$ where $\mathbf{X} = (X_1, X_2, \ldots, X_{n_1})$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n_2})$ are 2 independent random vectors. We assume the following parameters:

- $\mu_1 = 1$ and $\mu_2 = 2$.
- $\sigma_1^2 = 1$ and varying variance ratios, $\frac{\sigma_2^2}{\sigma_1^2} \equiv \rho = 2, 4, 6, 8, 10, 12$.
- n_1 takes values $n_1 = 2, 3, \ldots, 11$, with sample size ratios, $\frac{n_2}{n_1} \equiv k = 1, 2, 4$.

For each combination of (ρ, n_1, k), we construct the sampling distribution of the two-sample t-test statistic based on $R = 10000$ replications. In contrast to Lloyd (2013), we do not pre-calculate and fix the degrees of freedom, ν, using the WS approach. Instead, for each replication, we estimate the degrees of freedom, denoted by $\hat{\nu}$, using the WS approach (see Section 4.1 for details). Next, we determine the lower and upper critical values, $t_{\alpha,\hat{\nu}}$ and $t_{1-\alpha,\hat{\nu}}$, respectively for $\alpha = 0.05, 0.10$ to calculate the corresponding type I error rates based on the WS approach as[1]:

[1]The index j takes on values from 1 to 360 because ρ takes 6, n_1 10, k 3 and α 2 values. Therefore

$$\hat{\alpha}_j = F_T\left(t_{\alpha,\hat{\nu}}\right) + \left(1 - F_T\left(t_{1-\alpha,\hat{\nu}}\right)\right) \text{ for } j = 1, 2, \dots, 360 \tag{1}$$

where F_T represents the empirical cumulative distribution function of the two-sample t-test statistic.

Finally, to assess the accuracy of the estimated type I error rates, we calculate the relative difference between each $\hat{\alpha}_j$ and the corresponding true significance level α and compare it to a threshold of $d = 0.10$:

$$\frac{|\hat{\alpha}_j - \alpha|}{\alpha} > d = 0.10 \tag{2}$$

Table 1 and Table 2 present the results for the combination (ρ, n_1, k) for $\alpha = 0.05$ and $\alpha = 0.10$. Also Figure 1 visually represents these results, demonstrating that the accuracy of the estimated type I error rates generally improves as n_1 increases and k decreases. However, the effect of the variance ratio, ρ, on accuracy is not as clear-cut and warrants further investigation.

Table 1: Proportion of replicates where the relative absolute difference between a nominal α and the α-value given by WS approximation test is bigger than $d = 0.10$ where $\alpha = 0.05$

			n_1									
α	k	ρ	2	3	4	5	6	7	8	9	10	11
		2	0.861	0.620	0.500	0.468	0.428	0.315	0.086	0.269	0.086	0.120
		4	0.916	0.846	0.744	0.618	0.544	0.483	0.070	0.167	0.250	0.143
		6	0.925	0.869	0.785	0.666	0.596	0.394	0.458	0.252	0.185	0.097
		8	0.929	0.890	0.792	0.699	0.592	0.290	0.202	0.156	0.132	0.197
		10	0.949	0.879	0.762	0.579	0.477	0.693	0.177	0.114	0.042	0.111

$\widehat{\alpha}_j$ can take 360 values.

(continued)

α	k	ρ	2	3	4	5	6	7	8	9	10	11
	1	12	0.955	0.891	0.796	0.511	0.464	0.468	0.169	0.067	0.081	0.135
0.05	2	2	0.790	0.484	0.489	0.464	0.385	0.268	0.109	0.068	0.145	0.051
		4	0.707	0.648	0.358	0.202	0.278	0.043	0.041	0.001	0.003	0.004
		6	0.747	0.455	0.294	0.148	0.077	0.082	0.002	0.001	0.001	0.000
		8	0.761	0.612	0.263	0.108	0.099	0.001	0.005	0.000	0.005	0.000
		10	0.685	0.396	0.279	0.108	0.002	0.000	0.000	0.000	0.000	0.000
		12	0.678	0.452	0.272	0.388	0.002	0.000	0.000	0.000	0.000	0.000
	4	2	0.955	0.915	0.806	0.730	0.598	0.550	0.360	0.481	0.582	0.156
		4	0.951	0.842	0.737	0.637	0.326	0.276	0.170	0.257	0.025	0.109
		6	0.917	0.684	0.295	0.274	0.106	0.329	0.216	0.704	0.565	0.006
		8	0.639	0.406	0.279	0.157	0.085	0.454	0.163	0.017	0.030	0.001
		10	0.372	0.257	0.205	0.095	0.083	0.040	0.048	0.012	0.008	0.001
		12	0.307	0.256	0.126	0.099	0.022	0.017	0.022	0.012	0.006	0.000

Table 2: Proportion of replicates where the relative absolute difference between a nominal α and the α-value given by WS approximation test is bigger than $d = 0.10$ where $\alpha = 0.10$

			n_1									
α	k	ρ	2	3	4	5	6	7	8	9	10	11
	1	2	0.863	0.575	0.440	0.354	0.295	0.162	0.025	0.085	0.020	0.015
		4	0.871	0.742	0.542	0.376	0.205	0.153	0.155	0.004	0.000	0.000
		6	0.898	0.765	0.615	0.456	0.278	0.269	0.018	0.000	0.000	0.009
		8	0.918	0.783	0.522	0.329	0.148	0.075	0.000	0.000	0.028	0.063
		10	0.903	0.784	0.536	0.326	0.106	0.077	0.011	0.018	0.000	0.002
		12	0.919	0.770	0.650	0.251	0.269	0.114	0.013	0.012	0.026	0.000
		2	0.504	0.433	0.208	0.120	0.170	0.136	0.023	0.004	0.057	0.000
		4	0.645	0.454	0.149	0.053	0.094	0.022	0.030	0.000	0.000	0.002

(continued)

α	k	ρ	2	3	4	5	6	7	8	9	10	11
0.1	2	6	0.674	0.248	0.044	0.069	0.053	0.004	0.000	0.000	0.000	0.000
		8	0.634	0.417	0.020	0.005	0.004	0.000	0.000	0.000	0.000	0.000
		10	0.571	0.305	0.124	0.144	0.000	0.000	0.000	0.000	0.000	0.000
		12	0.566	0.303	0.055	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	4	2	0.934	0.824	0.681	0.552	0.372	0.174	0.174	0.007	0.054	0.002
		4	0.909	0.714	0.512	0.407	0.161	0.196	0.018	0.018	0.019	0.000
		6	0.726	0.383	0.197	0.226	0.032	0.062	0.040	0.204	0.000	0.000
		8	0.330	0.327	0.136	0.101	0.013	0.007	0.040	0.000	0.003	0.000
		10	0.303	0.193	0.126	0.043	0.016	0.017	0.006	0.000	0.001	0.000
		12	0.257	0.217	0.045	0.044	0.005	0.005	0.002	0.000	0.000	0.000

3 Conclusion

4 Appendix

4.1 Degrees of freedom estimated by the WS approach

In our case, following Satterthwaite (1946) and Welch (1947), the degrees of freedom $\hat{\nu}$ can be expressed as:

$$
\begin{aligned}
\hat{\nu} &= \frac{\left(\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{\widehat{\sigma}_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{\widehat{\sigma}_2^2}{n_2}\right)^2} \\
&= \frac{\left(\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\rho}\widehat{\sigma}_1^2}{kn_1}\right)^2}{\frac{1}{n_1-1}\left(\frac{\widehat{\sigma}_1^2}{n_1}\right)^2 + \frac{1}{kn_1-1}\left(\frac{\widehat{\rho}\widehat{\sigma}_1^2}{kn_1}\right)^2} \text{ where } \frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_1^2} \equiv \widehat{\rho} \\
&= \frac{(n_1-1)(kn_1-1)(k+\widehat{\rho})^2}{(kn_1-1)k^2 + (n_1-1)\widehat{\rho}^2}
\end{aligned} \tag{3}
$$

SUPPLEMENTARY MATERIAL

001_simulation_script.R R script used to produce the simulation results presented in Table 1, Table 2 and Figure 1

simulation_tbl_stacked.csv Simulated dataset Table 1 and Table 2

References

Lloyd, M. (2013), ‘2-Sample t-Distribution Approximation’, *Academic Forum* **31**, 23–30.

Satterthwaite, F. E. (1946), ‘An Approximate Distribution of Estimates of Variance Components’, *Biometrics Bulletin* **2**(6), 110.

Welch, B. L. (1947), ‘The Generalization of ‘Student’s’ Problem when Several Different Population Variances are Involved’, *Biometrika* **34**(1/2), 28.

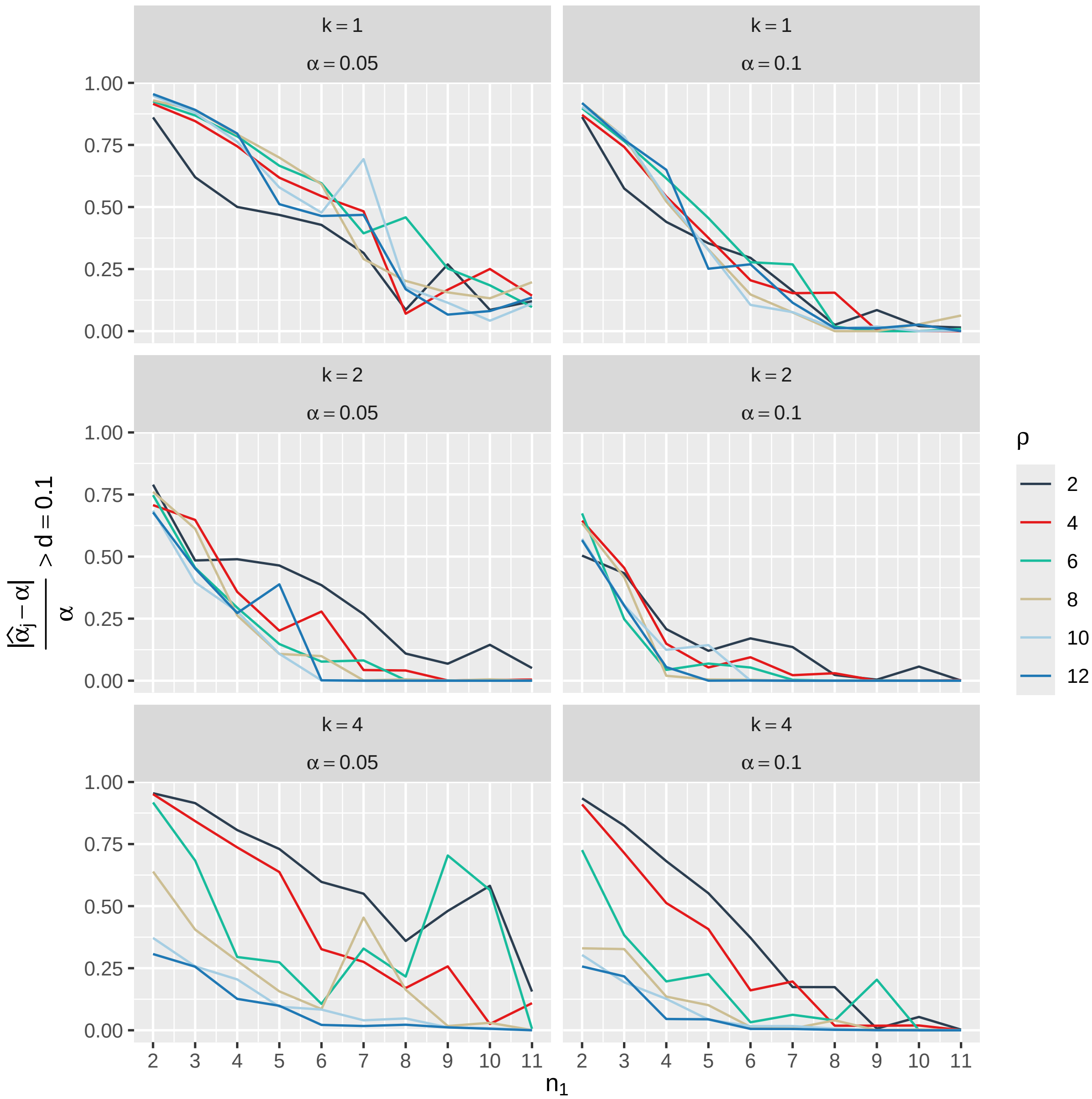

Figure 1: Graphical representation of Table 1 and Table 2