Bernoulli
2014, Vol. 20, No. 4, pp. 2102-2130.
Supplementary Materials

Tail approximations for the Student $t-$, $F-$, and Welch statistics for non-normal and not necessarily i.i.d. random variables

Dmitrii Zholud*

This Appendix contains remarks on Theorem 1.1; an example of an application of Theorem 1.1 to real data; extended version of the literature review; comparison of the result of Theorem 3.1 with the exact distribution of the Welch statistic; proof of Theorem 5.1.

1.1 Remarks on Theorem 1.1

For the ease of presentation we repeat the statement of Theorem 1.1 below

Theorem 1.1. *For any fixed value of n and each of the three choices of T, there exists a functional $K:\mathcal{G}\to\mathbb{R}^+$, such that for all $g_0, g_1 \in \mathcal{G}$ the limit expression*

$$\frac{\boldsymbol{P}(T>u|H_1)}{\boldsymbol{P}(T>u|H_0)}=\frac{K_{g_1}}{K_{g_0}}+o(1) \quad as \quad u\to\infty \tag{1}$$

holds with constants $0<K_{g_0}=K(g_0)<\infty$ and $0<K_{g_1}=K(g_1)<\infty$. The exact expressions for $K(g)$ are given in Sections 2, 3, and 5 for the three choices of the test statistic T.

Remark 1. Standard assumption in the use of any of the test statistics described above is that $\mathbf{X}\sim MVN(\mathbf{0},\sigma^2\mathbf{1}_n)$, where $MVN(\boldsymbol{\mu},\boldsymbol{\Sigma})$ denote the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. It is easy to check that $MVN(\mathbf{0},\sigma^2\mathbf{1}_n)\in\mathcal{G}$ and that $K(MVN(\mathbf{0},\sigma^2\mathbf{1}_n))=1$.

Remark 2: Theorem 1.1 holds also for the Welch test. In this case $K(MVN(\mathbf{0},\sigma^2\mathbf{1}_n))$ is not necessarily equal to 1, see Table 2 in Section 4.

Remark 3: It might be tempting to think that if we start with $g_1(\mathbf{x})=g_0(\mathbf{x})\in\mathcal{G}$, then small perturbation of $g_1(\mathbf{x})$ will induce small change in the right-hand side of (1), i.e. that it will deviate "just slightly" from 1. This is, generally, not the case. For example, it can be shown that if T is a $Z-$statistic, $g_0\sim MVN(\mathbf{0},\mathbf{1}_n)$, and $g_1\sim MVN(\mu\times\mathbf{1},\mathbf{1}_n)$, then the right-hand side of (1) becomes either 0, 1, or ∞ for $\mu<0$, $\mu=0$, and $\mu>0$ respectively. Theorem 1.1 therefore does not hold for $Z-$statistic.

Remark 4: It is important to emphasize that large sample approximations, that is asymptotic behavior for the case when u is fixed and $n\to\infty$, or when both $n\to\infty$ and $u_n\to\infty$, are irrelevant to the topic of the present article. We always assume that $u\to\infty$ and n is fixed, and preferably small. In particular, the convergence speed in (1) is faster for smaller sample sizes.

*Department of Mathematical Statistics
Chalmers University of Technology and University of Göteborg, Sweden.
E-mail: dmitrii@chalmers.se

1.2 An example of an application of Theorem 1.1

Let $g_1(\mathbf{x})$ be the joint density of the generic pair $\mathbf{X}$ described in the HTS study in Section 1.2, and let $F_1(x)$ be the CDF of the p-value $p = t(T(\mathbf{X}))$ that corresponds to the right-sided $t-$test. Assume also that $g_1(\mathbf{x})$ is continuous and satisfies regularity constraints of Theorem 2.1 (which is most likely the case except for really "pathological" instances). Then, according to Theorem 1.1 with $g_0 \sim MVN(\mathbf{0}, \sigma^2 \mathbf{1}_2)$ and Remark 1 it follows that $F_1(0) = 0$ and the right-sided derivative of $F_1(x)$ at zero is equal to

$$\lim_{x \to 0+} \frac{F_1(x) - F_1(0)}{x} = \lim_{u \to \infty} \frac{F_1(t(u))}{t(u)} = \lim_{u \to \infty} \frac{\mathbf{P}\left(t(T(\mathbf{X})) \leq t(u)\right)}{t(u)}$$

$$= \lim_{u \to \infty} \frac{\mathbf{P}\left(T(\mathbf{X}) \geq u\right)}{t(u)} = \lim_{u \to \infty} \frac{\mathbf{P}\left(T(\mathbf{X}) > u | \mathbf{X} \sim g_1\right)}{\mathbf{P}\left(T(\mathbf{X}) > u | \mathbf{X} \sim g_0\right)} = K_{g_1}.$$

Here we use continuity of $g_1(\mathbf{x})$ to replace "greater or equal" by "strictly greater" in the probability in the numerator of the last fraction in the above equation. Finally, we conclude that K_{g_1} is the slope of the line tangent to the graph of $F_1(x)$ at zero, that is,

$$K_{g_1} = F_1'(0+).$$

Using this simple linear approximation of the left tail of the distribution of the $p-$values, as shown in Figure 1, one can, for example, correct theoretical $p-$values for non-normal distribution of the LSC. Further applications include quantifying the impact of non-normality, dependence, or non-homogeneity of data from HTS experiments on the number of false positives; a detailed description of the methods lies beyond the scope of the present article and a curious reader is referred to the Introduction Part of Zholud (2011) and Rootzén and Zholud (2014).

1.3 Extended version of the literature review

There is an extensive literature on the behavior of the Student $t-$ and $F-$ statistic under deviation from normality. Below we focus mostly on the Student one-sample $t-$statistic, with references to the Student two-sample $t-$, $F-$, and Welch statistics given as needed.

A brief and quite recent introduction to the Student one-sample $t-$test can be found in Zabell (2008), and Cressie (1980) is a review with emphasis on understanding of the behavior of the test statistic under non-normality.

A main theme in the literature is the *Normal Approximation*, which is commonly stated as follows: "if the sample size is large enough and the population distribution is in the domain of attraction of the normal law, then the Student one-sample $t-$statistic is approximately $N(0,1)$ distributed", see e.g. Giné et al. (1997). The non-central $t-$statistic is discussed in Bentkus et al. (2007). However, the Normal Approximation is inaccurate for small sample sizes, which are the center of interest in this paper.

Additional accuracy in the Normal approximation can be obtained by using the first few terms of a Gram-Charlier series, Geary (1936), Bartlett (1935), or *Edgeworth expansion*, see e.g. Field and Ronchetti (1990), Hall (1987) and Gaen (1949, 1950). The Edgeworth expansion improves the Normal approximation and performs better for smaller sample sizes, but still is inaccurate in the extreme tail area.

A different approach is to use *Saddlepoint approximations* to the distribution of the test statistics, see e.g. Zhou and Jing (2006), Jing et al. (2004) and Daniels and Young (1991). The simulation study of Jing et al. (2004) showed that the Saddlepoint approximation for Student's $t-$statistic is very accurate in the tail area, however it is not known whether the relative error of such approximation tends to zero as one goes far out in the tail. The formulas for the Saddlepoint approximation do not give explicit formulas for the tail behavior of the test statistic and are computationally complex. In particular, Saddlepoint approximations require knowledge of the population density, which limits their use in statistical inference. Details on Saddlepoint approximations can be found in Kolassa (2006), Jensen (1995), Reid (1988) and Lugannani and Rice (1980).

The Student one-sample $t-$test is closely related to the so-called self-normalized sum, see e.g. Shao (2004), Shao (1997) and Logan et al. (1973). The exact distribution of $t-$statistic for some special cases is discussed in Eden and Yates (1933), Rider (1929), Perlo (1933) and Laderman (1939).

Fundamental results on the Student two-sample $t-$test and $F-$test were formulated by Fisher (1924, 1925, 1935a,b) and the behavior of these tests under deviation from standard assumptions was thoroughly studied. One particular case is known as the two-means Behrens-Fisher problem, see e.g. Sawilowsky (2002) and Kim and Cohen (1998) and the list of references in these papers, and addresses the question of using the Student two-sample $t-$test when the variances of the two populations are unequal. A common approach to the Behrens-Fisher problem is to use the Welch-Satterthwaite approximation, see Aspin and Welch (1949), Welch (1937, 1947), and Satterthwaite (1941, 1946). Exact distribution for the Welch $t-$test for odd sample sizes is given in Ray and Pitman (1961). The formulas in the latter paper can be easily modified to hold for the Student two-sample t-test as well. The effect of non-normality on the $F-$test was considered by e.g. Box (1953, 1954), Gaen (1950) and David and Johnson (1951).

1.4 Welch statistic

In this section we consider the tails of the Welch statistic T and discuss the accuracy of our asymptotic approximations for $\mathbf{P}(T > u)$ as $u \to \infty$. For the rest of the section we assume that the data is a Gaussian zero-mean vector, and the components are independent, and identically distributed within the two samples $(X_1, X_2, .., X_{n_1})$ and $(X_{n_1+1}, X_{n_2+2}, .., X_n)$ having population variances σ_1^2 and σ_2^2 respectively.

Under the above assumptions Ray and Pitman (1961) showed that for odd sample sizes n_1 and n_2 the exact distribution of the Welch statistic T (also known as Fisher-Behrens-Welch or Welch-Aspin statistic) is given by the weighted sum of t-distributions. More precisely, formula (4.1) on page 380 of the cited reference (we use the notation k instead of n) says that

$$\rho_T(v) = (2\pi)^{-\frac{1}{2}} \left\{ \sum_{r=0}^{m-1} \alpha_r \Gamma\left(r+\frac{3}{2}\right)\left(\frac{1}{2a}+\frac{v^2}{2}\right)^{-(r+3/2)} + \right. \tag{2}$$

$$\left. + \sum_{r=0}^{k-1} \beta_r \Gamma\left(r+\frac{3}{2}\right)\left(\frac{1}{2b}+\frac{v^2}{2}\right)^{-(r+3/2)} \right\},$$

where $\rho_T(v)$ is the p.d.f. of the Welch statistic T with odd sample sizes $n_1 = 2m+1$ and $n_2 = 2k+1$. The constants a, b, α_r and β_r depend on m, k, σ_1 and σ_2, and the constants α_r and β_r depend also on r.

In the light of the asymptotic representation of Theorem 3.1, the expression (2) for the exact density of T for odd sample sizes may look suspicious, since $\rho_T(v)$ is a mixture of the Student $t-$densities that have tails heavier than $t_{n-2}(u)$, the tail of the Student $t-$density in Theorem 3.1. A more detailed investigation, however, revealed that $\alpha_0 + \beta_0 = 0$, and the main terms in the asymptotic expansion for the heaviest densities in (2) cancel out, bringing in summands of higher order, which, in turn, may cancel out the main terms of the asymptotic expansion of the next most heavy summands (i.e. the ones that correspond to $r = 1$), and so on.

We tested our theory for the case $n_1 = 3$, $n_2 = 5$ and $\sigma_1^2 = \sigma_2^2$ by computing the left-hand side of (11) using (2), see Figure 1, and also for other choices of n_1 and n_2 and for $\sigma_1 \neq \sigma_2$, see the attached Mathematica scripts in Supplementary Materials.

The figure below shows that the left-hand side of (11) converges to the constant K_g. Recall that the latter is obtained by substituting $\alpha = 1/n_1$ and $\beta = 1/n_2$ in (14). We also computed the first 10 terms in the Taylor series expansion for $\rho_T(u)$ and $K_g t_{n-2}(u)$ as $u \to \infty$, and found that the two expansions are in agreement with each other, see the corresponding Mathematica script in Supplementary Materials.

The asymptotic expressions for the Welch test under non-normality and dependence are obtained using the procedure similar to the one employed in the derivation of the asymptotic expressions for the Student two-sample $t-$ test, see Section 3. We also study the accuracy of the asymptotic approximations of Corollary 3.1.1 for the Student two-sample $t-$test and Welch test using simulations, see Section 7.

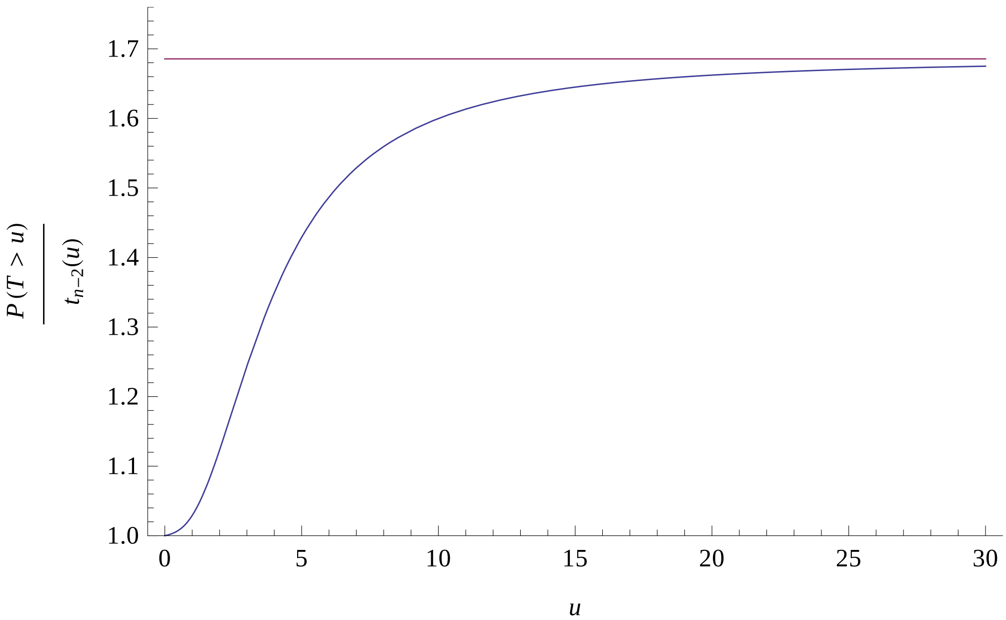

Figure 1. Convergence in (11) for the Welch $t-$statistic. Sample sizes: $n_1 = 3$ and $n_2 = 5$, variances equal. Blue curve is the left-hand side of (11) computed using the exact density of Welch's statistic, see (2). Horizontal line is the constant $K_g \approx 1.68$ computed from (14).

1.5 Proof of the main theorem for the F-statistic for comparison of variances

This section contains the proofs of Theorem 5.1 and Corollary 5.1.1. For convenience we state them here as well.

Theorem 1.2. *If g is continuous and satisfies regularity constraints of Theorem 5.1, then*

$$\frac{\boldsymbol{P}(T > u)}{F_{n_1-1,n_2-1}(u)} = K_g + o(1) \quad as \quad u \to \infty, \tag{3}$$

where $F_{n_1-1,n_2-1}(u)$ is the tail of the F-distribution with parameters $n_1 - 1$ and $n_2 - 1$ and $0 < K_g = K(g) < \infty$ is defined in (18) in Section 5.

Proof. Let A be an orthogonal linear operator defined on $\mathbb{R}^{n_2}$ such that $A\mathbf{e_{n_2}} = \mathbf{I}$. We have

$$\mathbf{P}(T > u) = \int_D g(\mathbf{x}, A\mathbf{y}) d\mathbf{x} d\mathbf{y},$$

where $D = \left\{ (\mathbf{x},\mathbf{y}) : \frac{s_1^2(\mathbf{x})}{\sum\limits_{i=1}^{n_2-1} y_i^2} > \frac{1}{n_2-1}u \right\}$.

Changing variables $y_i = (n_2-1)^{1/2} r s_1(\mathbf{x}) t_i$ for $1 \le i < n_2$ and $y_{n_2} = r$ (though, formally, we should have considered the case $r>0$ and $r<0$ separately), we write

$$\mathbf{P}\left(T > u\right) = \underset{\sum\limits_{i=1}^{n_2-1} t_i^2 < u^{-1}}{\int \cdots \int} G(\mathbf{t}) d\mathbf{t}, \tag{4}$$

where

$$G(\mathbf{t}) = (n_2-1)^{\frac{n_2-1}{2}} \underset{\mathbb{R}^{n_1}}{\int \cdots \int} s_1(\mathbf{x})^{n_2-1} \int\limits_{-\infty}^{\infty} g\left(\mathbf{x}, r\mathbf{I} + s_1(\mathbf{x}) A \mathbf{v}(\mathbf{t})\right) dr d\mathbf{x}$$

and

$$\mathbf{v}(\mathbf{t}) = (n_2-1)^{1/2}\left(t_1, t_2, .., t_{n_2-1}, 0\right).$$

Continuity of G at zero follows from the finiteness of integral (17) and continuity of g by the dominated convergence theorem, and Lemma 8.1 (A) in Section 1 of Appendix A implies that $\mathbf{P}\left(T > u\right)$ is asymptotically proportional to $t_{n_2-1}(\sqrt{u})$. It can be shown that

$$\frac{t_{n_2-1}(\sqrt{u})}{F_{n_1-1,n_2-1}(u)} = \frac{\Gamma\left(\frac{n_2}{2}\right)\Gamma\left(\frac{n_1-1}{2}\right)(n_1-1)^{\frac{n_2-1}{2}}}{2\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)} + o(1) \quad \text{as} \quad u \to \infty,$$

and (3) and the expression for K_g in (18) follow. □

Corollary 1.2.1 (Gaussian zero-mean case, independent samples). *If X and Y are independent zero-mean Gaussian random vectors with strictly non-degenerate covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, then (3) holds with*

$$K_g = C \underset{\mathbb{R}^{n_1}}{\int \cdots \int} \frac{s_1(\mathbf{x})^{n_2-1}}{\left(1 + \mathbf{x}\mathbf{\Sigma}_1^{-1}\mathbf{x}^T\right)^{n/2}} d\mathbf{x}, \tag{5}$$

where the constant C is given by

$$C = \frac{(n-2)(n_1-1)^{\frac{n_2-1}{2}}\Gamma\left(\frac{n_1-1}{2}\right)|\mathbf{I}\mathbf{\Sigma}_2\mathbf{I}^T|^{1/2}}{2\pi^{\frac{n_1+1}{2}}|\mathbf{\Sigma}_1|^{1/2}|\mathbf{\Sigma}_2|^{1/2}}.$$

Proof. The regularity assumption (17) follows from Lemma 9.2 and the derivation of the constant K_g is a calculus exercise. □

References

A.A. Aspin and B.L. Welch. Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika*, 36(3):290–296, 1949.

M.S. Bartlett. The effects of non-normality on the $t-$distribution. *Proc. Cambridge Philos. Soc.*, 31: 223–231, 1935.

V. Bentkus, B.-Y. Jing, Q.M. Shao, and W. Zhou. Limiting distributions of the non-central $t-$statistic and their applications to the power of $t-$tests under non-normality. *Bernoulli*, 13(2):346–364, 2007.

G.E.P. Box. Non-normality and tests on variances. *Biometrika*, 40:318–335, 1953.

G.E.P. Box. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, 25(2):290–302, 1954.

Noel Cressie. Relaxing assumptions in the one sample $t-$test. *Australian Journal of Statistics*, 22(2): 143–153, 1980.

H.E. Daniels and G.A. Young. Saddlepoint approximation for the Studentized mean, with an application to the bootstrap. *Biometrika*, 78:169–179, 1991.

F.N. David and N.L. Johnson. The effect of non-normality on the power function of the $F-$test in the analysis of variance. *Biometrika*, 38:43–57, 1951.

T. Eden and F. Yates. On the validity of Fisher's z test when applied to an actual example of non-normal data. (with five text-figures.). *The Journal of Agricultural Science*, 23(1):6–17, 1933.

C. Field and E. Ronchetti. *Small Sample Asymptotics.* IMS, Hayward, CA, 1990.

R.A. Fisher. On a distribution yielding the error functions of several well known statistics. *Proceeding of the International Mathematical Congress, Toronto*, 2:805–813, 1924.

R.A. Fisher. Applications of "Student's" distribution. *Metron*, 5:90–104, 1925.

R.A. Fisher. The mathematical distributions used in the common tests of significance. *Econometrica*, 3(4):353–365, 1935a.

R.A. Fisher. The fiducial argument in statistical inference. *Annals of Eugenics*, 8:391–398, 1935b.

A.K. Gaen. The distribution of "Student's" t in random samples of any size drawn from non-normal universes. *Biometrika*, 36:353–369, 1949.

A.K. Gaen. The distribution of the variance ratio in random samples of any size drawn from non-normal universes. *Biometrika*, 37:236–255, 1950.

R.C. Geary. The distribution of Student's ratio for non-normal samples. *J. Roy. Statist. Soc., Suppl.*, 3:178–184, 1936.

E. Giné, F. Götze, and D.M. Mason. When is the Student $t-$statistic asymptotically standard normal? *Ann. Probab.*, 25:1514–1531, 1997.

P. Hall. Edgeworth expansion for Student's t statistic under minimal moment conditions. *Ann. Probab.*, 15:920–931, 1987.

J.L. Jensen. *Saddlepoint Approximations.* Oxford Univ. Press., 1995.

B.-Y. Jing, Q.M. Shao, and W. Zhou. Saddlepoint approximation for Student's $t-$statistic with no moment conditions. *Ann. Statist.*, 32(6):2679–2711, 2004.

S.-H. Kim and A.S. Cohen. On the Behrens-Fisher problem: A review. *Journal of Educational and Behavioral Statistics*, 23(4):356–377, 1998.

J.E. Kolassa. *Series Approximation Methods in Statistics*, volume 88 of *Lecture Notes in Statistics.* Springer, New York., 3 edition, 2006.

J. Laderman. The distribution of "Student's" ratio for samples of two items drawn from non-normal universes. *The Annals of Mathematical Statistics*, 10(4):376–379, 1939.

B.F. Logan, C.L. Mallows, S.O. Rice, and L.A. Shepp. Limit distributions of self-normalized sums. *Ann. Probab.*, 1:788–809, 1973.

R. Lugannani and S. Rice. Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. in Appl. Probab.*, 12:475–490, 1980.

V. Perlo. On the distribution of the "Student's" ratio for samples of three drawn from rectangular distribution. *Biometrika*, 25:203–204, 1933.

W.D. Ray and A.E.N.T. Pitman. An exact distribution of the Fisher-Behrens-Welch statistic for testing the difference between the means of two normal populations with unknown variance. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(2):377–384, 1961.

N. Reid. Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.*, 3:213–238, 1988.

P.R. Rider. On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes. *Biometrika*, 21(1):124–143, 1929.

H. Rootzén and D.S. Zholud. Efficient estimation of the number of false positives in high-throughput screening. *Submitted to Biometrika*, 2014.

F.E. Satterthwaite. Synthesis of variance. *Psychometrika*, 6(5):309–316, 1941.

F.E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114, 1946.

S.S. Sawilowsky. Fermat, Schubert, Einstein, and Behrens-Fisher: The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$. *Journal of Modern Applied Statistical Methods*, 1(2):461–472, 2002.

Q.-M. Shao. Self-normalized large deviations. *Ann. Probab.*, 25:285–328, 1997.

Q.-M. Shao. Recent progress on self-normalized limit theorems. *Probability, Finance and Insurance*, pages 50–68, 2004.

B.L. Welch. The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3-4):350–362, 1937.

B.L. Welch. The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34(1-2):28–35, 1947.

S.L. Zabell. On Student's 1908 article. *Journal of the American Statistical Association*, 103:1–7, 2008.

D.S. Zholud. *Extreme Value Analysis of Huge Datasets: Tail Estimation Methods in High-Throughput Screening and Bioinformatics*. PhD thesis, Göteborg University. ISBN: 978-91-628-8354-6, 2011.

W. Zhou and B.-Y. Jing. Tail probability approximations for Student's $t-$statistics. *Probability Theory and Related Fields*, 136(4):541–559, 2006.