Better Confidence Intervals for the Variance in a Random Sample

Ruth Hummel*, Senin Banga†, Thomas P. Hettmansperger‡

Abstract

Classical inferential procedures about the variance in the one-sample and two-sample random normal models are well-known to be sensitive to the normality assumption. While the two-sample procedures have been studied extensively and a variety of robust procedures have been developed (by Levene, Brown and Forsythe, Boos, Shoemaker, and using bootstrapping or jackknifing, etc.), not much has been done in the one-sample case. In this paper we propose two alternative methods for finding a confidence interval for the variance, and we investigate improvements in robustness measured by the achieved simulated coverage of each of these methods compared to the classical Chi-square method, the large-sample procedure based on the central limit theorem, and an approximate bootstrap.

Key words and phrases. variance, bias correction, Pearson curves, matching moments, ABC bootstrap, kurtosis, robustness

1 Motivation: Sensitivity of the Chi-square test to the Normal Assumption

It is well-known that the classical Chi-square procedure for performing tests and constructing confidence intervals for the variance is extremely sensitive to normality. When the parent distribution of the sample data is far from normal, the achieved confidence level of the confidence interval can be extremely far from the nominal level. As is illustrated in Figure 1, the simulated coverage probablity of a nominally 90% confidence inteval for the variance can be as low as 60% for extremely non-normal populations. It is tempting to hope that these disastrous results only occur for seriously non-normal populations, but we can see in Figure 1 that the coverage is frequently underestimated by 10 to 20% even for reasonably large samples from symmetric distributions such as the t with 7 degrees of freedom and slightly contaminated normal distributions.

It might be imagined that these results will improve with a large enough sample size. In reality, the size of the test and the coverage probability of the confidence interval can get even farther from the assumed level as the sample size gets larger. This is illustrated in Figure 1.

*Department of Statistics, The Pennsylvania State University, State College, PA 16802.

†Minitab, Inc.

‡Department of Statistics, The Pennsylvania State University, State College, PA 16802

The research of all three authors was supported in part by Minitab, Inc.

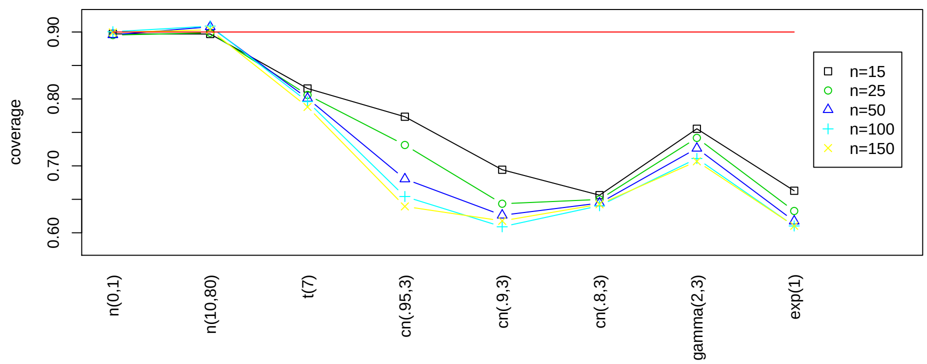


Figure 1: Chi-square achieved coverage levels for a variety of distributions, for varying sample sizes

Although much work has been done on improving the robustness of the two-sample procedures about variances ([3], [9]), improving robustness for the one-sample variance is still largely unexplored.

The goal of this paper is to compare, through simulation studies, the improvements in robustness that can be realized from several alternative methods versus the normality-sensitive classical Chi-Square procedure.

2 Several Alternative Methods

2.1 Large-sample normal approximations:

It is well-known that, provided the fourth moment of the parent distribution is finite, the sample variance is asymptotically normally distributed with expected value σ^2 and variance $(\gamma-1)\sigma^4/n$ [2], where γ is the population kurtosis. This asymptotic distribution gives the following nominal $100(1-\alpha)$ percent confidence interval for the variance:

$$\frac{S^2}{1-z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\gamma}-1}{n}}} \leq \sigma^2 \leq \frac{S^2}{1+z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\gamma}-1}{n}}} \tag{1}$$

where $\hat{\gamma}$ is the consistent estimator of the kurtosis that is unbiased for normal samples, [1] and $\hat{\gamma_e}$ is the corresponding estimate of the kurtosis excess (such that $\gamma = 3$ and $\gamma_e = 0$ for the normal distribution),

$$\hat{\gamma_e} = \hat{\gamma} - 3 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \frac{(x_i-\bar{x})^4}{S^4} - \frac{3(n-1)^2}{(n-2)(n-3)}.$$

Because of the high skewness of the distribution of S^2 for smaller samples, we follow the usual method of applying the natural log to the sample variance, which helps adjust for this skewness. By the Cramer δ method, $\ln S^2$ will be assymptotically approximately normally distributed with mean $\ln\sigma^2$ and variance $(\gamma-1)/n$. This transformation leads to the following confidence interval for the population variance:

$$S^2 \exp\Big(z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\gamma}-1}{n}}\Big) \leq \sigma^2 \leq S^2 \exp\Big(-z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\gamma}-1}{n}}\Big) \tag{2}$$

2.2 Adjusted Degrees of Freedom

This approach relies on the fact that the sample variance is a sum of squares, and, for samples sufficiently large, can be approximated as a Chi-square with an appropriate estimate for the degrees of freedom. Using the method of matching moments ([9], [6], [8], [7]), we can find such an estimate for the degrees of freedom. (Shoemaker uses a similar argument when approximating the distribution of a ratio of independent sample variances in [9].)

We match the first two moments of the distribution of S^2 with that of a random variable Y distributed as $c\chi^2_r$. That is, we must solve for r and c in the following system of equations: (1) $\sigma^2 = cr$, and (2) $\frac{\sigma^4}{n}(\gamma - \frac{n-3}{n-1}) = 2rc^2$.

(We use the fact that, as mentioned in [6], $\mathrm{Var}(S^2) = \frac{\sigma^4}{n}(\gamma - \frac{n-3}{n-1})$ when randomly sampling from any distribution with the first four moments finite.)

The unique solution is given as

$$r = \frac{2n(n-1)}{(n-1)\gamma - n + 3} = \frac{2n(n-1)}{(n-1)\gamma_e + 2n}$$

$$c = \frac{\sigma^2}{2n}\Big(\gamma - \frac{n-3}{n-1}\Big) = \frac{\sigma^2}{2n}\Big(\gamma_e + \frac{2n}{n-1}\Big).$$

That is, $\frac{rS^2}{\sigma^2}$ is approximately distributed as χ^2_r.

Consequently, an approximate two-sided $(1-\alpha)100$ percent confidence interval for the variance may be given as

$$\frac{\hat{r}S^2}{\chi^2_{\hat{r},\alpha/2}} \leq \sigma^2 \leq \frac{\hat{r}S^2}{\chi^2_{\hat{r},1-\alpha/2}} \tag{3}$$

where

$$\hat{r} = \frac{2n}{\hat{\gamma_e} + 2n/(n-1)}.$$

If the sample is known to come from the normal population, then $r = n - 1$ and (3) reduces to the classical Chi-square interval.

2.3 ABC Bootstrap Approximation

Since the bootstrap method is a nonparametric method, we are interested in including a bootstrap confidence interval for the variance as part of our simulated comparison. Rather than include a typical bootstrap for comparison, we compare the above methods to the approximate bootstrap confidence (ABC) interval. This is a numerical approximation to the bootstrap that performs very similarly but much faster. The ABC method requires less computational time because it uses Taylor series expansions to approximate the bootsrap results rather than actually resampling. Details about this method and evidence of its comparability to the bootstrap can be found in [4].

3 Simulation Results

Validity (size achieved) of the classical Chi-square method (ChSq), the Ln Asymptotic approximation (Ln Asympt), the Chi-square adjustment (AdjDF), and the ABC bootstrap approximation (ABC) are assessed in comparison to the nominal expected coverage level of $1 - \alpha = 0.90$.

Samples are randomly generated from Normal distributions N(0,1) and N(80,10); Student t distributions, T(r), with r = 5 and 10; the continuous uniform distribution on (0,1), Unif(0,1); Beta distributions with (α, β) parameters (3,3), (.4,.7), and (8,1); Chi-square distributions, χ^2(r), with r = 10 and 20; Gamma distributions with (α=shape, β=scale) parameters (2,3) and (.5,6); the exponential distibution with rate=scale=1, Exp(1); the Laplace distribution with $\mu = 0, \lambda = 1$, Lapl(0,1); and two contaminated standard normal distributions, CN(λ, σ), where λ is the proportion of the sample that is from a standard normal distribution (λ is also called the mixing parameter), while the remaining $(1 - \lambda)$ proportion is from a normal distribution with the same mean ($\mu = 0$) but a different variance, specified by the parameter σ^2, CN($\lambda = .9, \sigma = 3$) and CN($\lambda = .8, \sigma = 3$). A summary of relevant distributional features is described in Table 1.

Table 1:

Distributional Features					
Distribution	*Description*	*Mean*	*Variance*	*Skewness*	*Kurtosis Excess*
N(0,1)	symmetric	0	1	0	0
N(80,10)	symmetric	80	10	0	0
T(10)	heavy-tailed, symmetric	0	1.25	0	1
T(5)	heavy-tailed, symmetric	0	1.6667	0	6
Unif(0,1)	light-tailed, symmetric	0.5	0.0833	0	-1.2
Beta(3,3)	light-tailed, symmetric	0.5	0.0357	0	-0.6667
Gamma(2,3)	heavy-tailed, skewed	6	18	1.4142	3
Gamma(.5,6)	heavy-tailed, skewed	3	18	2.8284	12
Exp(1)	light-tailed, skewed	1	1	2	6
Lapl(0,1)	heavy-tailed, symmetric	0	2	0	3
CN(.9,3)	heavy-tailed, symmetric	0	1.8	0	5.3333
CN(.8,3)	heavy-tailed, symmetric	0	2.6	0	4.5444
Beta(.4,.7)	U-shaped	3.6364	0.1102	0.5301	-1.1448
Beta(8,1)	heavy-tailed, skewed	0.8889	0.0099	1.4230	2.2841

In each study, 5,000 samples of sizes 15, 25, 50, and 100 were drawn from each of the distributions described above, and two-sided 90% upper and lower confidence limits on the variance are calculated using the different methods described in the previous section.

Note that a two-sided 90% confidence interval can also be considered as upper and lower one-sided 95% confidence intervals (since each of the methods we have described is symmetric in coverage).

Since the targeted coverage probability for each side is 95%, the simulation error is

$$\sqrt{\frac{(1 - 0.95)(.95)}{5000}} = 0.00308 \approx 0.31\%$$

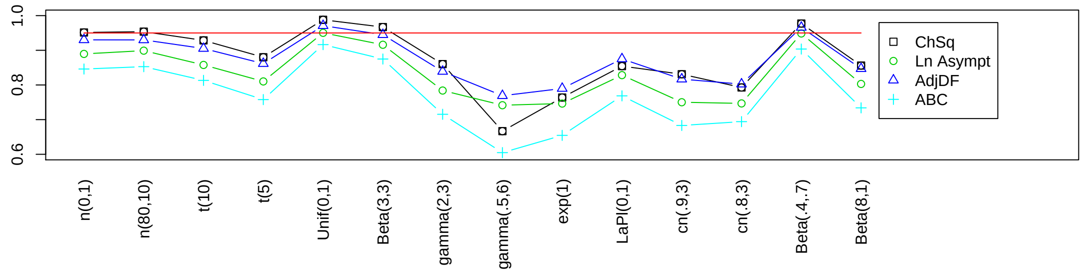
for the one-sided confidence levels and

$$\sqrt{\frac{(1 - 0.90)(.90)}{5000}} = 0.00424 \approx 0.42\%$$

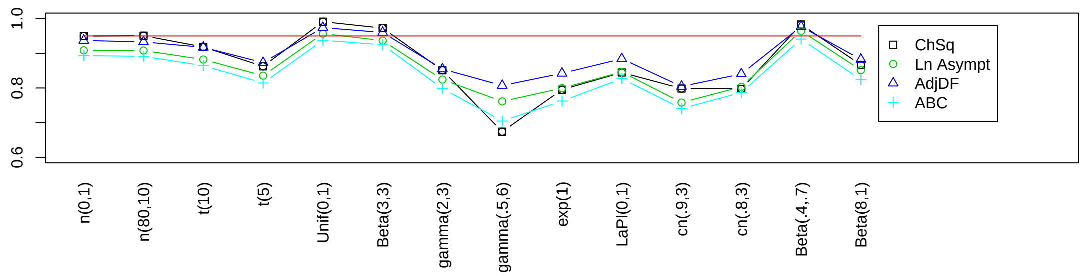
for the two-sided confidence level.

The results are presented in Figures 2 through 4 and Tables 2 through 4. In addition, the mean lengths of the confidence intervals obtained by each method in each tested setting are compared in Table 5.

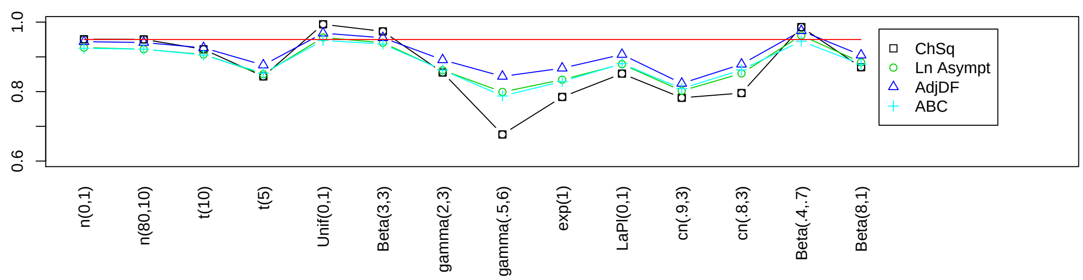
(2a) Sample size = 15



(2b) Sample size = 25



(2c) Sample size = 50



(2d) Sample size = 100

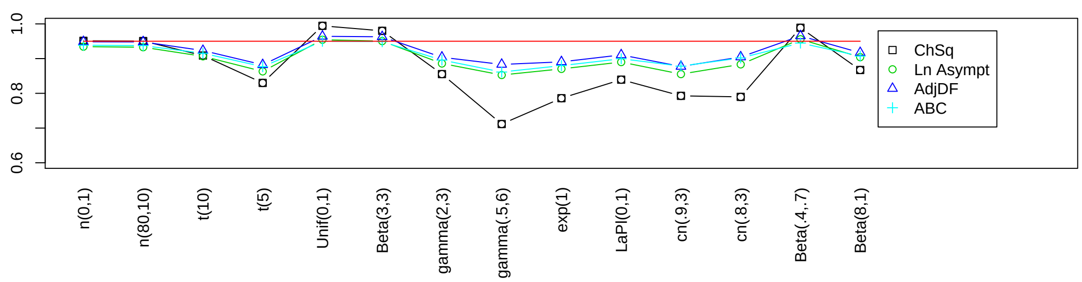
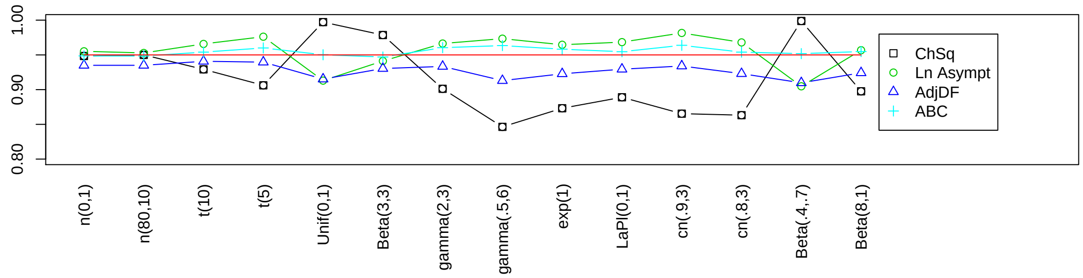
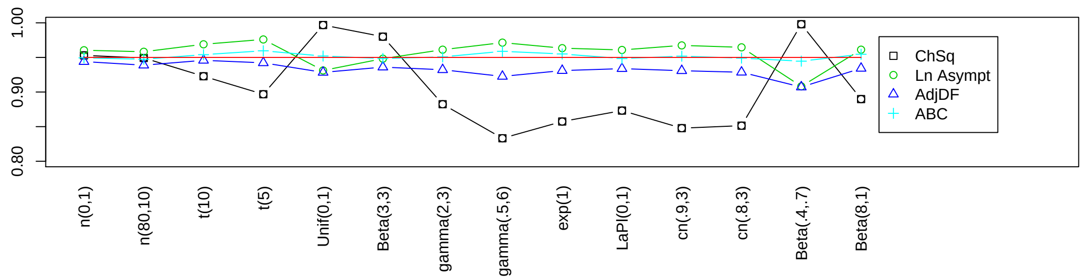


Figure 2: Achieved one-sided upper coverage probability for varying distributions; nominal confidence level = 0.95.

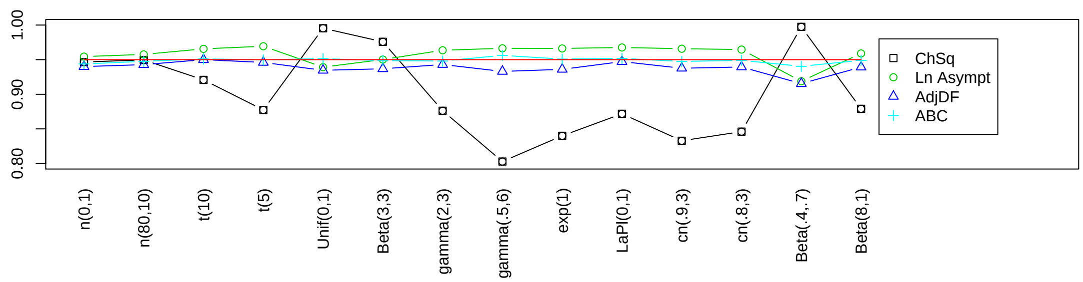
(3a) Sample size = 15



(3b) Sample size = 25



(3c) Sample size = 50



(3d) Sample size = 100

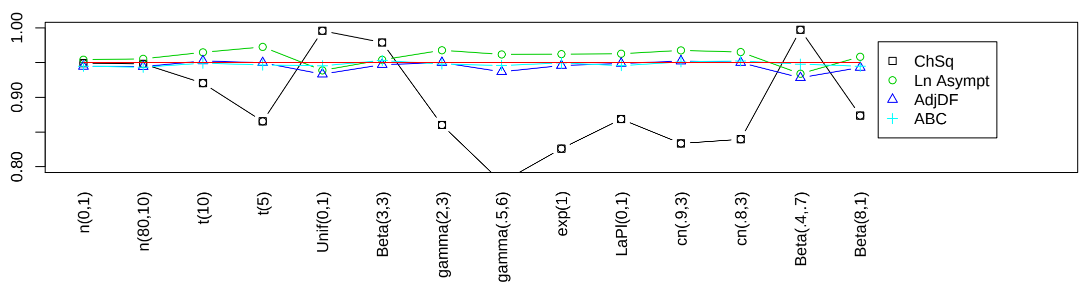
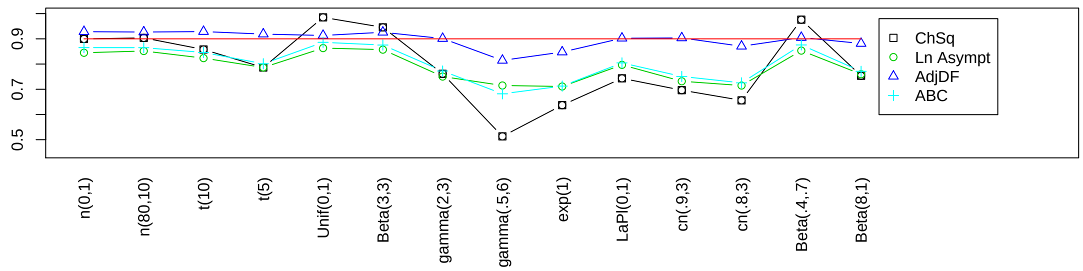
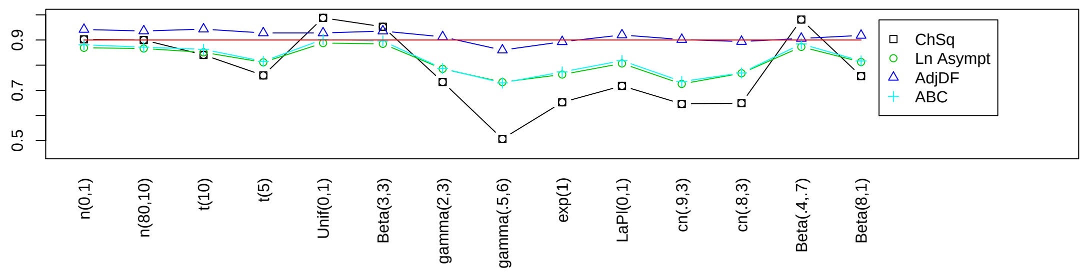


Figure 3: Achieved one-sided lower coverage probability for varying distributions; nominal confidence level = 0.95.

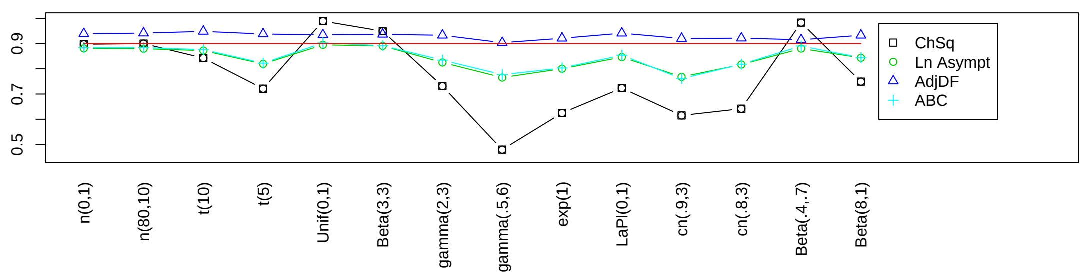
(4a) Sample size = 15



(4b) Sample size = 25



(4c) Sample size = 50



(4d) Sample size = 100

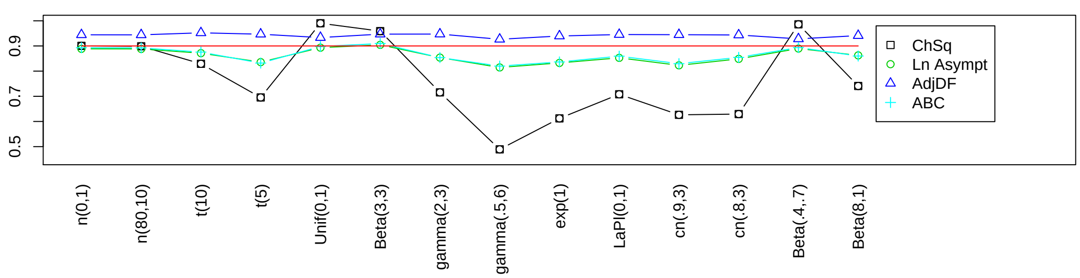


Figure 4: Achieved two-sided coverage probability for varying distributions; nominal confidence level = 0.90.

4 Anderson-Darling Selection Method

It is conceivable that, because we know the traditional Chi-square method is exact when we are sampling from a normal distribution, and (as is evident in the simulations), the adjusted degrees of freedom Chi-square method seems to be the most robust method, perhaps we should develop a method which assesses the normality of the sample and tells us to use the traditional Chi-square if the sample is "normal enough," and otherwise to use the adjusted degrees of freedom. In this spirit we use the following adaptive method:

We first test for normality based on an Anderson Darling test at a specified level. If the sample fails to reject, we use the traditional Chi-square method. Otherwise, we use the Chi-square method with adjusted degrees of freedom. As a test of the usefulness of this method versus always choosing the adjusted degrees of freedom method, we use α-levels of 0.2 and 0.6 as the acceptance/rejection level for the Anderson Darling normality test. (These are referred to as AD2 and AD6 in Figure 5.)

Through a simulation study we construct two-sided confidence intervals, at the nominal 90% level, to compare the confidence level achieved by using this two-step adaptive scheme to the achieved confidence level of each of the two methods (classical Chi-square and adjusted degrees of freedom) on their own. The results are shown in Figure 5, where we can see that the performance of this Anderson-Darling Selection method generally falls between the performance of the traditional Chi-square method and the adjusted degrees of freedom Chi-square method (and most frequently is *considerably* worse than the adjusted degrees of freedom Chi-square method). In the few cases where the Anderson-Darling Selection method performs better than both, it is only slightly better than the adjusted degrees of freedom Chi-square method. These results indicate that, since it is generally worse and only ever slightly better than the adjusted degrees of freedom method, the Anderson-Darling Selection method is an inferior method to the adjusted degrees of freedom method.

5 Correcting the Upper Confidence Limit

From Figure 2 and Table 2 we can see that the lower confidence limits usually yield coverage probabilities that are close to the nominal. The upper confidence limits, on the other hand, yield coverage probabilities that are systematically far from the nominal coverage level, and increasingly so as the distributions become more nonnormal. When the upper confidence limits are calculated using the true γ_e, however, as in Figure 6, then the simulated coverage probabilities are very close to the nominal coverage. These results suggest that the upper confidence limit is biased.

In an attempt to adjust for this bias we consider the adjusted degrees of freedom method, which seems to perform "best" based on our simulation study. As mentioned in Section 2.2, an approximate two-sided $100(1-\alpha)$ percent confidence interval may be given by

$$\frac{\hat{r}S^2}{\chi^2_{\hat{r},\alpha/2}} \leq \sigma^2 \leq \frac{\hat{r}S^2}{\chi^2_{\hat{r},1-\alpha/2}}$$

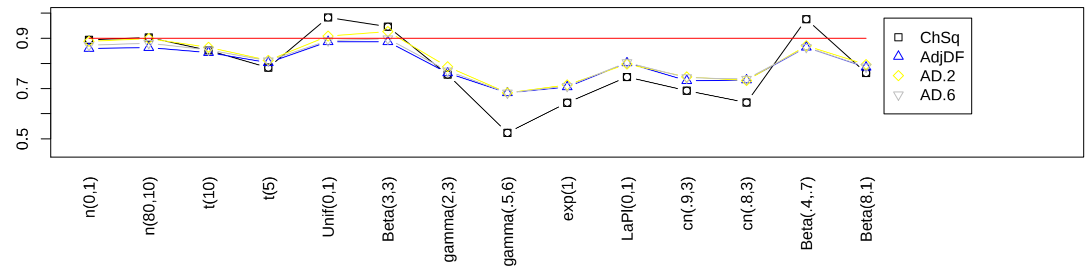
where

$$\hat{r} = \frac{2n}{\hat{\gamma}_e + 2n/(n-1)}$$

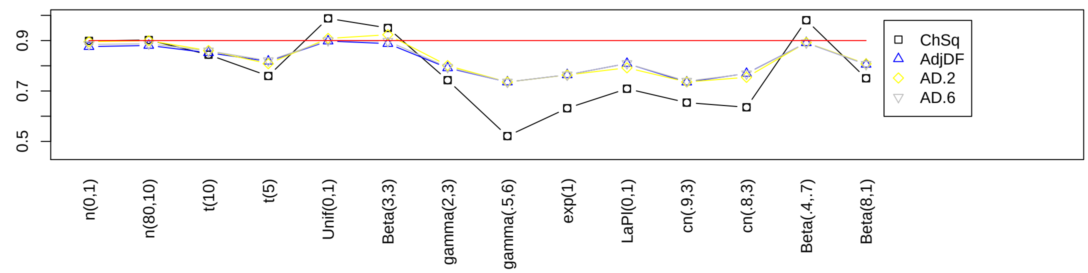
is an estimator of the true degrees of freedom

$$r = \frac{2n}{\gamma_e + 2n/(n-1)}.$$

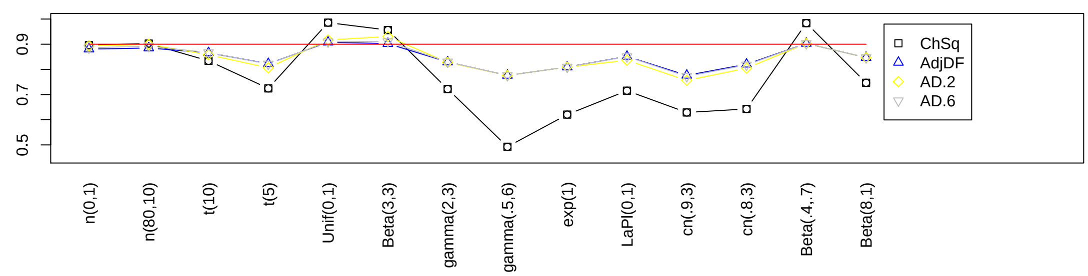
(5a) Sample size = 15



(5b) Sample size = 25



(5c) Sample size = 50



(5d) Sample size = 100

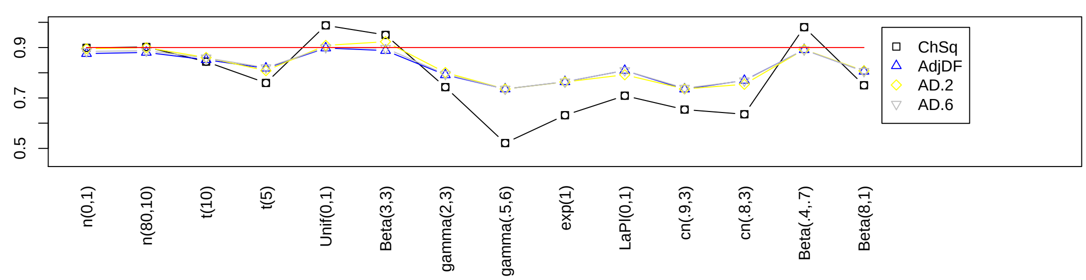
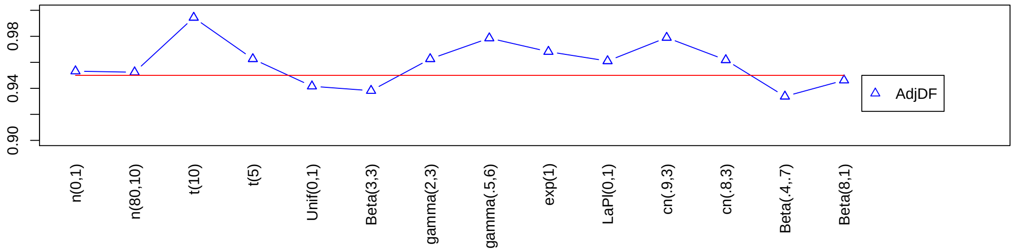
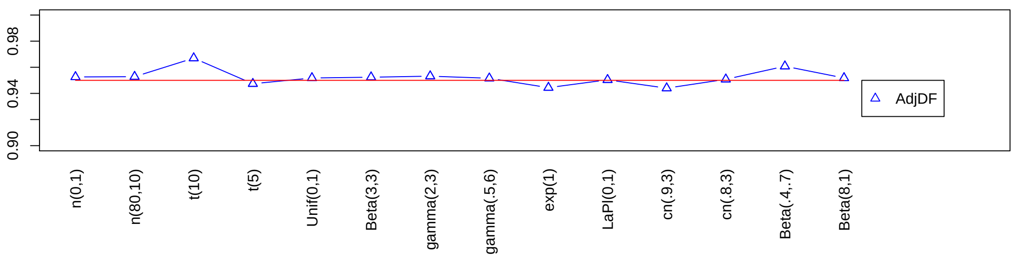


Figure 5: Comparison of achieved two-sided total coverage probability of the Anderson-Darling Selection Method to the coverage of the Classical Chi-square and Adjusted Chi-square for varying distributions; nominal confidence level = 0.90.

(5a) One-sided upper nominal confidence level = 0.95



(5b) One-sided lower nominal confidence level = 0.95



(5c) Two-sided nominal confidence level = 0.90

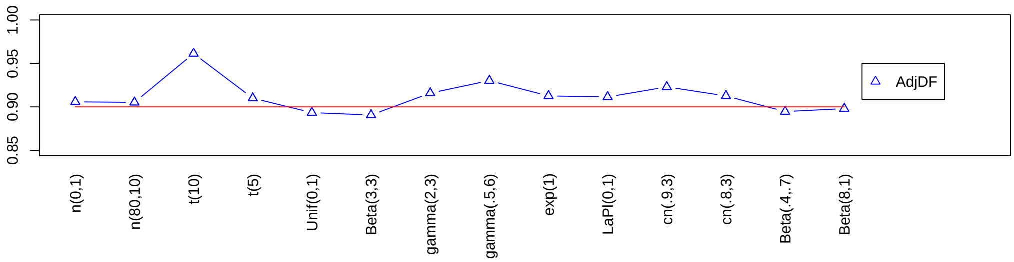


Figure 6: Achieved coverage probability when kurtosis is fixed at the true value; for varying distributions; sample size = 50.

Moreover, an approximate one-sided $100(1-\alpha)$ percent upper confidence limit on the variance is given as

$$\hat{u} = \frac{\hat{r}S^2}{\chi^2_{\hat{r},1-\alpha}},$$

which can be viewed as a biased estimator of u, which is a function of the true kurtosis.

We approximate the small-sample bias in the upper confidence limit which is induced through a bias in the kurtosis excess (see Appendix for details).

This gives us the following bias-correction for the upper confidence limit $\hat{u}$:

$$2\hat{u} - S^2(1 - C_{1-\alpha}(\frac{n+1}{(n-1)(2+\hat{r})})) = S^2\Big(\frac{2\hat{r}}{\chi^2_{\hat{r},1-\alpha}} + C_{1-\alpha}(\frac{n+1}{(n-1)(2+\hat{r})}) - 1\Big) \tag{4}$$

where

$$C_\alpha(r) = \sqrt{2}z_\alpha r^{1/2} + \frac{2}{3}(z_\alpha^2 - 1)r + \frac{1}{9\sqrt{2}}(z_\alpha^3 - 7z_\alpha)r^{3/2} - \frac{1}{405}(6z_\alpha^4 + 14z_\alpha^2 - 433)r^2 \tag{5}$$

$$+\frac{1}{4860\sqrt{2}}(9z_\alpha^5+256z_\alpha^3-433z_\alpha)r^{5/2}.$$

This method is referred to as "B-C AdjDF."

6 Second simulation: Investigating Bias Correction

The following plots illustrate the improvement in achieved coverage that the bias correction described in Section 5 can provide. The samples used for this comparison are the same samples used in the previous simulation (Section 3), so the only change in these plots is the addition of the performance of the bias-corrected method, "B-C AdjDF."

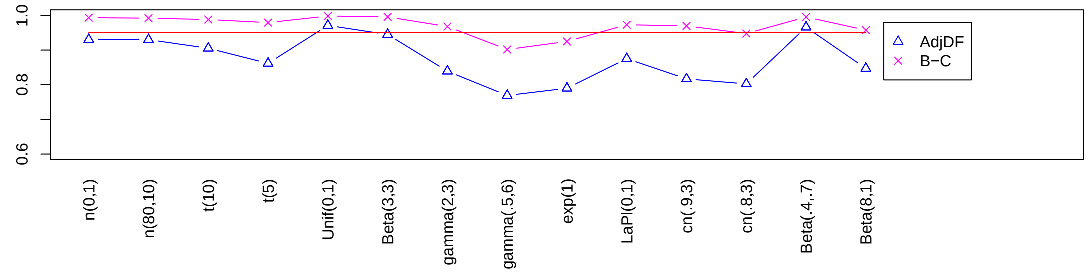
Here we see the achieved one-sided upper coverage probability for varying sample sizes and varying distributions, with B-C AdjDF compared to the adjusted degrees of freedom method at the nominal 0.95 level.

Note that the lower end point has not been corrected (for reasons discussed in Section 5), so we do not include plots of the lower coverage (because they have not changed). The Total Coverage plot results from adding the upper missed coverage to the lower missed coverage and plotting the resultant achieved total coverage. Consequently, the Total Coverage plots provide additional information (compared to the Upper Coverage plots) about the B-C AdjDF performance only in that the Total Coverage plots summarize the two-sided 90% confidence level performance, versus the upper one-sided 95% confidence level performance assessment provided in the Upper Coverage plots.

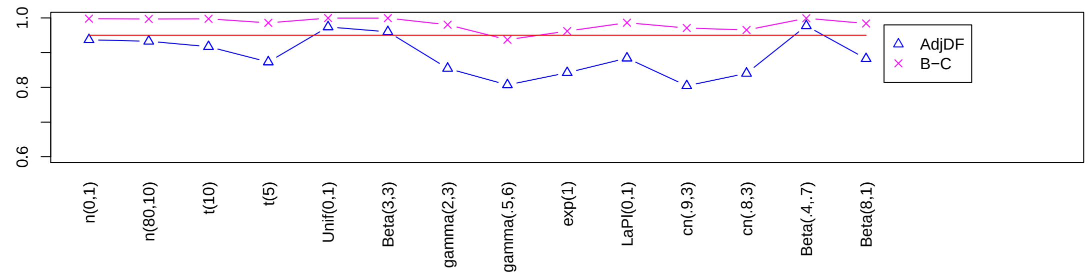
7 Summary and Conclusions

From the plots and tabled results, it seems that the "best" method varies a bit according to the characteristics of the distribution being sampled. For our heavy-tailed distributions (be they symmetric or skewed), the Ln Asymptotic Approximation, Chi-square Adjusted, and ABC methods yield achieved coverage probabilities that are, in general, lower than the nominal level. For the light-tailed and U-shaped distributions, the Chi-square Adjusted method is (debatably) the closest to the nominal level. Of course, if the distribution is precisely normal, the traditional Chi-square method is the best method. But even for slight deviations from normality, like Contaminated Normal distributions or the Student t with 10 degrees of freedom, the traditional Chi-square is already being outperformed by all of our alternative methods. For this reason, it seems most useful to consider a "normality not assumed" concept when developing a confidence interval for a variance. Given that there is no evidence to expect the sample to have come from a precisely normal distribution, some heuristic assumptions about whether the underlying distribution is most likely heavier or lighter than a normal distribution would lead to our "best" choice of method. With absolutely no distributional assumptions, the authors' recommendation at this time is to use the Adjusted Chi-square method, based on its reasonable performance (compared with the other methods proposed) and its theoretically logical extension from the traditional Chi-square method. In the near future it would be profitable to consider an adaptive scheme using distributional features of the sample to determine which method should be used. We anticipate that this adaptive structure would provide a single "best" method across all the distributions we have tested.

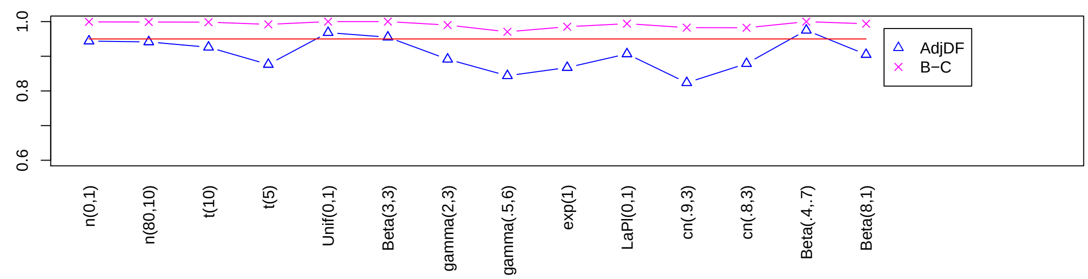
(7a) Sample size = 15



(7b) Sample size = 25



(7c) Sample size = 50



(7d) Sample size = 100

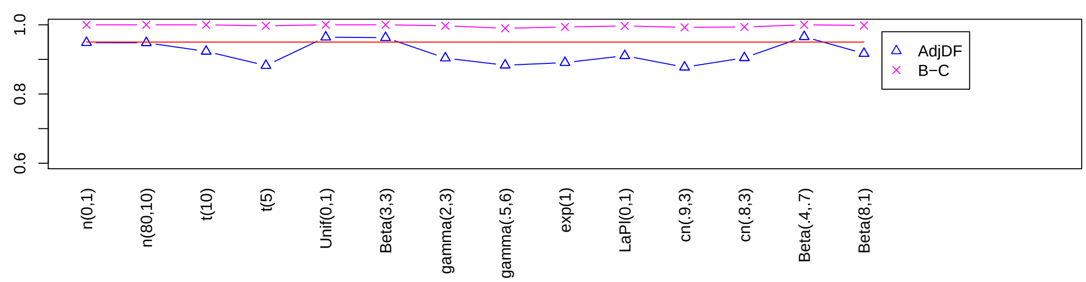
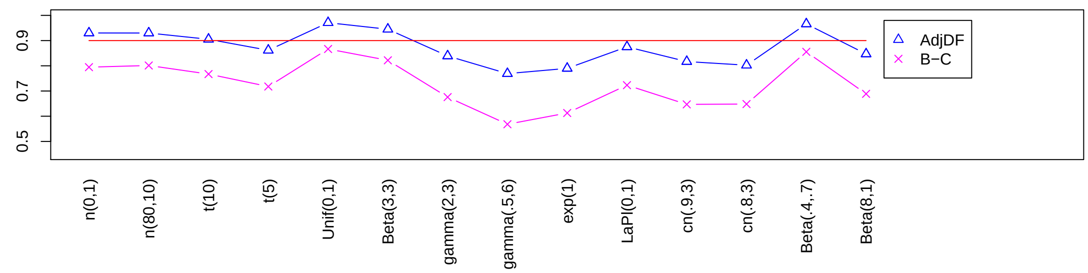
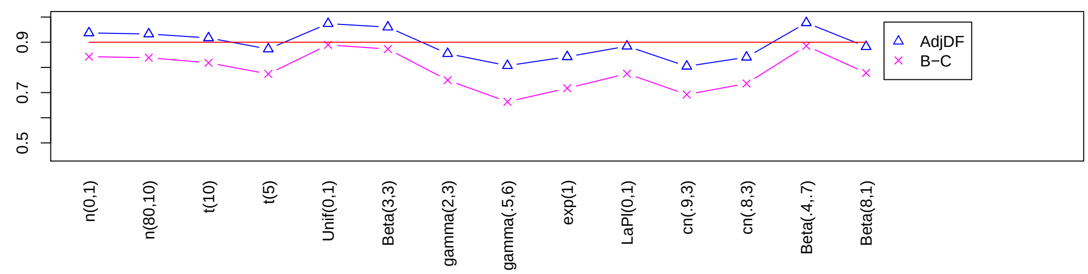


Figure 7: Achieved one-sided upper coverage probability for varying distributions; nominal confidence level = 0.95.

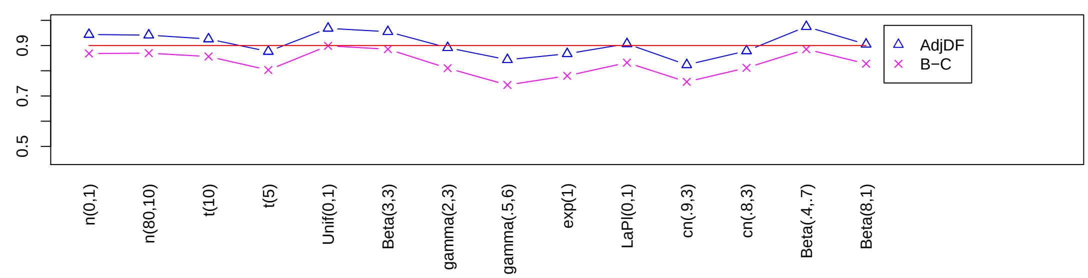
(8a) Sample size = 15



(8b) Sample size = 25



(8c) Sample size = 50



(8d) Sample size = 100

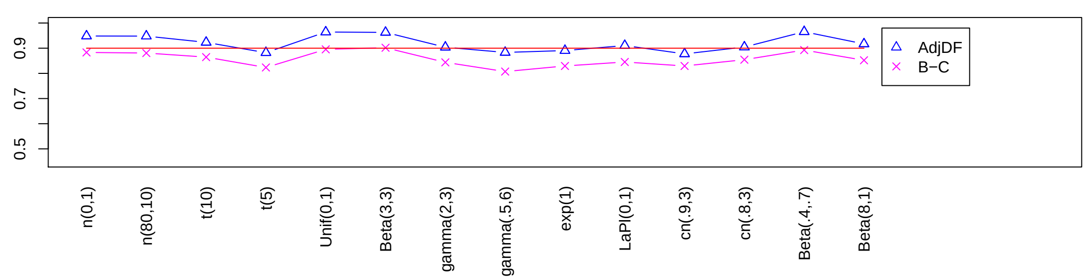


Figure 8: Achieved two-sided coverage probability for varying distributions; nominal confidence level = 0.90.

Table 2: Upper Coverage for Varying Sample Sizes, one-sided confidence level = 0.95

Upper Achieved Coverage															
Method	*n*	*n(0,1)*	*n(80,10)*	*t(10)*	*t(5)*	*Unif(0,1)*	*Beta(3,3)*	*gamma(2,3)*	*gamma(.5,6)*	*exp(1)*	*LaPl(0,1)*	*cn(.9,3)*	*cn(.8,3)*	*Beta(.4,.7)*	*Beta(8,1)*
ChSq	15	0.9514	0.9538	0.9286	0.88	0.988	0.967	0.8602	0.6668	0.7638	0.8544	0.831	0.7928	0.9772	0.856
LnAsympt	15	0.8896	0.899	0.8578	0.8102	0.9504	0.916	0.784	0.7416	0.7464	0.8284	0.7502	0.7468	0.9486	0.803
ABC	15	0.846	0.8528	0.8132	0.7578	0.9164	0.8748	0.7154	0.6048	0.6546	0.7686	0.6832	0.6942	0.9036	0.734
AdjDF	15	0.9302	0.93	0.9052	0.8616	0.971	0.945	0.839	0.769	0.7898	0.8752	0.8168	0.8024	0.966	0.847
B-C AdjDF	15	0.9934	0.992	0.988	0.9792	0.9982	0.9956	0.9678	0.9016	0.9248	0.973	0.9696	0.9476	0.9954	0.9574
ChSq	25	0.9496	0.9508	0.9182	0.8624	0.9912	0.9726	0.851	0.674	0.795	0.8446	0.7984	0.7974	0.9832	0.8668
LnAsympt	25	0.9086	0.908	0.882	0.8354	0.9566	0.9366	0.8242	0.7612	0.7994	0.846	0.758	0.8032	0.9646	0.8514
ABC	25	0.8932	0.8912	0.8642	0.8144	0.9376	0.9244	0.7984	0.7046	0.7622	0.8268	0.7404	0.787	0.941	0.8236
AdjDF	25	0.937	0.9328	0.9172	0.873	0.974	0.96	0.8546	0.807	0.8422	0.8842	0.8044	0.8404	0.9772	0.8824
B-C AdjDF	25	0.9978	0.997	0.9974	0.9858	0.9996	0.9994	0.9804	0.9372	0.962	0.9862	0.971	0.9652	0.999	0.984
ChSq	50	0.9508	0.9504	0.9218	0.8438	0.9938	0.9734	0.855	0.6766	0.7846	0.8518	0.7822	0.7958	0.9858	0.8702
LnAsympt	50	0.9268	0.922	0.9068	0.85	0.9558	0.9406	0.8616	0.799	0.8342	0.879	0.8022	0.8524	0.9616	0.8846
ABC	50	0.9246	0.9222	0.9054	0.8532	0.9474	0.937	0.8622	0.7874	0.8292	0.8804	0.8086	0.8626	0.9452	0.879
AdjDF	50	0.9438	0.9414	0.926	0.8764	0.9682	0.955	0.8914	0.8438	0.8674	0.9068	0.8236	0.8788	0.975	0.905
B-C AdjDF	50	0.9992	0.9988	0.9984	0.992	1	1	0.99	0.9706	0.9852	0.994	0.9826	0.9822	0.9998	0.9938
ChSq	100	0.9512	0.9506	0.9088	0.83	0.9944	0.98	0.8554	0.7116	0.786	0.8394	0.7928	0.7898	0.9888	0.8672
LnAsympt	100	0.9344	0.9328	0.9066	0.8632	0.9544	0.9504	0.886	0.8528	0.8704	0.89	0.8554	0.8834	0.9558	0.9042
ABC	100	0.9384	0.9374	0.9154	0.8768	0.95	0.9488	0.8958	0.8614	0.8798	0.8996	0.8784	0.902	0.9448	0.9072
AdjDF	100	0.9484	0.948	0.9232	0.8822	0.9642	0.9628	0.9038	0.8832	0.8906	0.9106	0.8774	0.9046	0.9654	0.917
B-C AdjDF	100	0.9998	1	0.9998	0.997	1	1	0.9972	0.9898	0.9938	0.9966	0.9926	0.9936	1	0.9982

Table 3: Lower Coverage for Varying Sample Sizes, one-sided confidence level = 0.95

Lower Achieved Coverage															
Method	*n*	*n(0,1)*	*n(80,10)*	*t(10)*	*t(5)*	*Unif(0,1)*	*Beta(3,3)*	*gamma(2,3)*	*gamma(.5,6)*	*exp(1)*	*LaPl(0,1)*	*cn(.9,3)*	*cn(.8,3)*	*Beta(.4,.7)*	*Beta(8,1)*
ChSq	15	0.9484	0.95	0.9292	0.9062	0.9972	0.9786	0.9012	0.8464	0.8732	0.889	0.8654	0.8632	0.9988	0.8976
LnAsympt	15	0.9552	0.9528	0.9658	0.9762	0.913	0.9412	0.9664	0.9734	0.9646	0.9684	0.9816	0.968	0.9048	0.9568
ABC	15	0.9486	0.9484	0.954	0.96	0.9502	0.947	0.9604	0.9634	0.9582	0.9546	0.9638	0.954	0.9518	0.9548
AdjDF = B-C	15	0.935	0.935	0.9408	0.9396	0.9152	0.9304	0.9334	0.913	0.9228	0.9294	0.934	0.923	0.9098	0.9242
ChSq	25	0.953	0.9492	0.9228	0.8968	0.9968	0.98	0.8824	0.8332	0.8574	0.8732	0.8478	0.8514	0.998	0.8898
LnAsympt	25	0.9604	0.9582	0.969	0.976	0.9312	0.9482	0.9612	0.9714	0.9634	0.9608	0.9674	0.9646	0.9082	0.9612
ABC	25	0.9494	0.9474	0.954	0.9596	0.952	0.9484	0.9508	0.9588	0.955	0.9484	0.9518	0.949	0.9446	0.9546
AdjDF = B-C	25	0.944	0.9388	0.9458	0.9422	0.9284	0.936	0.9322	0.9226	0.931	0.9338	0.9308	0.9286	0.9072	0.9342
ChSq	50	0.9466	0.9496	0.9208	0.8774	0.9954	0.9758	0.8762	0.8028	0.84	0.8718	0.8328	0.846	0.9974	0.879
LnAsympt	50	0.9544	0.9576	0.9656	0.9694	0.9392	0.9502	0.9636	0.9664	0.9662	0.9676	0.9658	0.9646	0.9188	0.959
ABC	50	0.9438	0.9476	0.951	0.9498	0.9514	0.9486	0.948	0.9562	0.9508	0.9518	0.9474	0.949	0.9404	0.949
AdjDF = B-C	50	0.94	0.9428	0.95	0.946	0.9346	0.9368	0.943	0.933	0.936	0.9472	0.9376	0.9394	0.9154	0.9392
ChSq	100	0.9492	0.9482	0.9204	0.8654	0.9958	0.979	0.8602	0.7776	0.8262	0.8686	0.8336	0.8396	0.9972	0.8738
LnAsympt	100	0.9542	0.9554	0.9648	0.9726	0.9388	0.954	0.9678	0.9618	0.9622	0.9628	0.9676	0.9652	0.934	0.9584
ABC	100	0.9448	0.9434	0.9492	0.9466	0.9454	0.953	0.948	0.9458	0.9496	0.9456	0.951	0.9524	0.9478	0.9448
AdjDF = B-C	100	0.9446	0.9442	0.9524	0.9498	0.9334	0.947	0.95	0.9368	0.9458	0.9488	0.9522	0.95	0.9282	0.943

Table 4: Total Coverage for Varying Sample Sizes, confidence level = 0.90

Total Achieved Coverage															
Method	*n*	*n(0,1)*	*n(80,10)*	*t(10)*	*t(5)*	*Unif(0,1)*	*Beta(3,3)*	*gamma(2,3)*	*gamma(.5,6)*	*exp(1)*	*LaPl(0,1)*	*cn(.9,3)*	*cn(.8,3)*	*Beta(.4,.7)*	*Beta(8,1)*
ChSq	15	0.8998	0.9038	0.8578	0.7862	0.9852	0.9456	0.7614	0.5132	0.637	0.7434	0.6964	0.656	0.976	0.7536
LnAsympt	15	0.8448	0.8518	0.8236	0.7864	0.8634	0.8572	0.7504	0.715	0.711	0.7968	0.7318	0.7148	0.8534	0.7598
ABC	15	0.8652	0.865	0.846	0.8012	0.8862	0.8754	0.7724	0.682	0.7126	0.8046	0.7508	0.7254	0.8758	0.7712
AdjDF	15	0.9284	0.927	0.9288	0.9188	0.9134	0.926	0.9012	0.8146	0.8476	0.9024	0.9036	0.8706	0.9052	0.8816
B-C AdjDF	15	0.7946	0.8012	0.7672	0.7178	0.8666	0.8218	0.9012	0.8146	0.8476	0.9024	0.9036	0.8706	0.9052	0.8816
ChSq	25	0.9026	0.9	0.841	0.7592	0.988	0.9526	0.7334	0.5072	0.6524	0.7178	0.6462	0.6488	0.9812	0.7566
LnAsympt	25	0.869	0.8662	0.851	0.8114	0.8878	0.8848	0.7854	0.7326	0.7628	0.8068	0.7254	0.7678	0.8728	0.8126
ABC	25	0.881	0.8716	0.863	0.8152	0.9024	0.896	0.7868	0.7296	0.7732	0.818	0.7352	0.769	0.8844	0.8166
AdjDF	25	0.9418	0.9358	0.9432	0.928	0.928	0.9354	0.9126	0.8598	0.893	0.92	0.9018	0.8938	0.9062	0.9182
B-C AdjDF	25	0.8426	0.8386	0.8182	0.774	0.8896	0.8728	0.7492	0.6634	0.7172	0.7752	0.6922	0.736	0.8856	0.7782
ChSq	50	0.8974	0.9	0.8426	0.7212	0.9892	0.9492	0.7312	0.4794	0.6246	0.7236	0.615	0.6418	0.9832	0.7492
LnAsympt	50	0.8812	0.8796	0.8724	0.8194	0.895	0.8908	0.8252	0.7654	0.8004	0.8466	0.768	0.817	0.8804	0.8436
ABC	50	0.8838	0.8842	0.876	0.8224	0.9028	0.8918	0.8344	0.7768	0.8034	0.854	0.7612	0.8182	0.8904	0.8442
AdjDF	50	0.9392	0.9416	0.9484	0.938	0.9346	0.9368	0.933	0.9036	0.9212	0.9412	0.9202	0.9216	0.9152	0.933
B-C AdjDF	50	0.8684	0.8698	0.8564	0.803	0.8988	0.8856	0.8102	0.7436	0.78	0.8322	0.756	0.8116	0.8856	0.828
ChSq	100	0.9004	0.8988	0.8292	0.6954	0.9902	0.959	0.7156	0.4892	0.6122	0.708	0.6264	0.6294	0.986	0.741
LnAsympt	100	0.8886	0.8882	0.8714	0.8358	0.8932	0.9044	0.8538	0.8146	0.8326	0.8528	0.823	0.8486	0.8898	0.8626
ABC	100	0.893	0.8922	0.8756	0.832	0.8976	0.9098	0.8538	0.82	0.8364	0.8594	0.8296	0.8546	0.8936	0.86
AdjDF	100	0.9444	0.9442	0.9522	0.9468	0.9334	0.947	0.9472	0.9266	0.9396	0.9454	0.9448	0.9436	0.9282	0.9412
B-C AdjDF	100	0.8832	0.8808	0.8646	0.8234	0.8954	0.9018	0.8438	0.8072	0.8294	0.8452	0.8294	0.8544	0.8926	0.852

Table 5: Length of Confidence Intervals, two-sided confidence level = 0.90

Median Length of Confidence Intervals, alpha = 0.1

Method	*n*	*n(0,1)*	*n(80,10)*	*t(10)*	*t(5)*	*Unif(0,1)*	*Beta(3,3)*	*gamma(2,3)*	*gamma(.5,6)*	*exp(1)*	*LaPl(0,1)*	*cn(.9,3)*	*cn(.8,3)*	*Beta(.4,.7)*	*Beta(8,1)*
ChSq	15	1.468296	147.476	1.762504	2.178503	0.1270609	0.05334799	23.69937	19.551065	1.2616442	2.648721	2.179712	3.266606	0.16968047	0.01342549
LnAsympt	15	1.147099	114.9945	1.388985	1.640884	0.07060413	0.03724704	16.02909	5.883919	0.6970143	2.090976	1.355221	1.875081	0.09099703	0.00936163
ABC	15	2.006014	205.2449	3.288793	5.877214	0.08152542	0.05156152	61.54063	100.347901	4.4430627	6.890916	8.711028	11.158703	0.1110148	0.03056472
AdjDF	15	1.356937	136.9424	1.778104	2.425511	0.08006976	0.04276338	25.962	36.569983	1.6633677	3.460288	2.47263	4.849622	0.10405444	0.01577262
B-C AdjDF	15	2.963329	299.1007	3.881223	5.289972	0.17396361	0.0932319	56.57217	79.084085	3.603307	7.549549	5.398707	10.566634	0.22594506	0.03436315
ChSq	25	1.041762	104.38926	1.274497	1.616764	0.08964801	0.03779865	17.74801	15.44653	0.9374493	1.994648	1.655638	2.52478	0.11864749	0.009996384
LnAsympt	25	0.9072527	90.61639	1.204006	1.633849	0.05253254	0.02856864	17.71778	14.14899	0.990368	2.238701	1.534321	2.972796	0.06974578	0.01063685
ABC	25	1.4791686	152.30753	2.55227	5.126229	0.05943897	0.03715555	52.3882	92.22364	3.6783806	5.671804	7.107984	8.830563	0.08429891	0.025843274
AdjDF	25	0.9899457	98.61265	1.331569	1.937434	0.05633061	0.03074583	21.43579	30.86561	1.4248063	2.677168	2.124724	4.085251	0.07478411	0.012794904
B-C AdjDF	25	2.1379807	213.40967	2.883744	4.199489	0.12048081	0.06622895	46.54795	66.71096	3.0963926	5.814783	4.61362	8.870497	0.16014693	0.027755159
ChSq	50	0.4773981	47.9825	0.5909668	0.7703005	0.0401074	0.01713251	8.557053	8.09196	0.465814	0.94018	0.832675	1.219981	0.05312872	0.004676524
LnAsympt	50	0.4608546	46.18741	0.6505068	0.9986756	0.0249846	0.01375658	11.558776	16.63266	0.7587628	1.326779	1.375	2.038087	0.03381167	0.00628841
ABC	50	0.6294603	62.61869	1.1498001	2.6149081	0.02689981	0.01588528	25.449677	47.95331	1.886389	2.586149	3.140198	3.795101	0.03840317	0.010920543
AdjDF	50	0.4701316	47.10212	0.666812	1.0373201	0.02538438	0.01398863	11.985735	18.38391	0.8017375	1.380899	1.450254	2.153448	0.03434316	0.0065126
B-C AdjDF	50	0.9864073	98.82031	1.4049367	2.1997728	0.05259948	0.02918877	25.427231	39.52176	1.7107783	2.936332	3.103154	4.606193	0.0712281	0.013808077
ChSq	100	0.6918176	69.60697	0.85883	1.094779	0.0585448	0.02493389	12.08264	11.19511	0.6496604	1.354436	1.157266	1.734709	0.07783186	0.00672824
LnAsympt	100	0.6463136	64.46749	0.895087	1.288445	0.03600082	0.01970605	14.81844	18.7847	0.9310192	1.775389	1.654347	2.720031	0.04835714	0.008408241
ABC	100	0.9888614	98.87039	1.788197	3.939558	0.03968631	0.02414012	38.02702	66.49814	2.7917597	4.060496	4.915876	6.026659	0.05671673	0.017163087
AdjDF	100	0.6733875	67.16733	0.940132	1.384093	0.03720436	0.0203925	15.86176	24.06965	1.0513464	1.912621	1.833842	3.009127	0.04997953	0.008980576
B-C AdjDF	100	1.4349415	143.00499	2.01034	2.967709	0.07809366	0.04319901	34.09044	52.06628	2.2684333	4.11458	3.952466	6.499578	0.10494628	0.019292476

8 Appendix

We begin by introducing the following Vanderbeck and Cooke approximation to the χ^2 percentile.

Let $\chi^2_{r,\alpha}$ be the $100(1-\alpha)$ percentile of the chi-square distribution with r degrees of freedom. Then, by [5],

$$\chi^2_{r,\alpha} \approx r + \sqrt{2} z_\alpha r^{1/2} + \frac{2}{3}(z_\alpha^2 - 1) + \frac{1}{9\sqrt{2}}(z_\alpha^3 - 7z_\alpha) r^{-1/2} - \frac{1}{405}(6z_\alpha^4 + 14z_\alpha^2 - 433) r^{-1}$$
$$+ \frac{1}{4860\sqrt{2}}(9z_\alpha^5 + 256 z_\alpha^3 - 433 z_\alpha) r^{-3/2}$$

where z_α is the $100(1-\alpha)$ percentile of the standard normal distribution.

For easier reference, we rewrite the Vanderbeck-Cooke approximation as

$$\chi^2_{r,\alpha} \approx r(1 + C_\alpha(r^{-1})) \tag{6}$$

where

$$C_\alpha(r) = \sqrt{2} z_\alpha r^{1/2} + \frac{2}{3}(z_\alpha^2 - 1) r + \frac{1}{9\sqrt{2}}(z_\alpha^3 - 7z_\alpha) r^{3/2} - \frac{1}{405}(6z_\alpha^4 + 14z_\alpha^2 - 433) r^2 \tag{7}$$
$$+ \frac{1}{4860\sqrt{2}}(9z_\alpha^5 + 256 z_\alpha^3 - 433 z_\alpha) r^{5/2}$$

This approximation works well even when α is near zero or one, which is essential for our purpose of constructing confidence intervals with coverage level α.

We now approximate the expectation of $\hat{u}$ as

$$E(\hat{u}) = E\Big(\frac{\hat{r} S^2}{\chi^2_{\hat{r},1-\alpha}}\Big) \approx E\Big(\frac{S^2}{1 + C_{1-\alpha}(\hat{r}^{-1})}\Big) \approx E(S^2) E\Big(\frac{1}{1 + C_{1-\alpha}(\hat{r}^{-1})}\Big) \approx \sigma^2 (1 + C_{1-\alpha}(E(\hat{r}^{-1})))$$

The above expressions show that the bias in $\hat{u}$ is in part induced by the small sample bias of the estimated degrees of freedom which is a function of the estimated kurtosis. The estimated kurtosis we have used so far in our calculations is the so-called "bias-corrected" kurtosis because it is unbiased for normal populations and is asymptotically unbiased for any other population. For nonnormal populations the small-sample bias may be considerable. Once this small-sample bias (of the kurtosis) is assessed, the bias in $\hat{u}$ can be derived via that of $1/\hat{r}$.

To find the bias of $1/\hat{r}$, we assume that the parent distribution has a finite fourth moment so that

$$E(\hat{\gamma_e}) = \frac{\gamma_e}{\gamma_e/n + (n+1)/(n-1)} + O(n^{-1}) = \frac{n(n-1)\gamma_e}{(n-1)\gamma_e + n(n+1)} + O(n^{-1}).$$

which gives a small-sample bias for the kurtosis excess of

$$\text{Bias}(\hat{\gamma_e}) = -\frac{n(2\gamma_e + \gamma_e^2) - \gamma_e^2}{n^2 + n(\gamma_e + 1) - \gamma_e} + O(n^{-1}).$$

The expectation of the kurtosis to the order of n^{-2} or lower is a function of higher order moments, which imposes further restrictions on the applicability of the results.

Based on the above approximation of the expectation for the sample kurtosis, we can now find the expectation of $\hat{r}^{-1}$ to be as follows:

$$E(\hat{r}^{-1}) \approx \frac{n+1}{(n-1)(2+r)}$$

Finally, by using that

$$E(\hat{u}) \approx \sigma^2(1 + C_{1-\alpha}(E(r^{-1})))$$

and the expectation of r^{-1}, we have that a bias correction for the upper confidence limit $\hat{u}$ is

$$2\hat{u} - S^2(1 - C_{1-\alpha}(\frac{n+1}{(n-1)(2+\hat{r})})) = S^2\Big(\frac{2\hat{r}}{\chi^2_{\hat{r},1-\alpha}} + C_{1-\alpha}(\frac{n+1}{(n-1)(2+\hat{r})}) - 1\Big). \tag{8}$$

References

[1] Minitab release 14 for windows, 2003.

[2] Steve Arnold. *Mathematical Statistics.* Prentice Hall College Div, 1990.

[3] Dennis D. Boos and Cavell Brownie. Comparing variances and other measures of dispersion. Unpublished Paper.

[4] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*, pages 328–331. Chapman and Hall/CRC, 1993.

[5] N. L. Johnson and S. Kotz. *Continuous Univariate Distributions - 1*, volume 8. John Wiley and Sons, Inc.

[6] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*, page 229. McGraw-Hill, 3rd edition, 1974.

[7] Egon S. Pearson. The analysis of variance in cases of non-normal variation. *Biometrika*, 23:114–133, 1931.

[8] Donald T. Searls and Pichai Intarapanich. A note on an estimator for the variance that utilizes the kurtosis. *The American Statistician*, 44:295–296, 1990.

[9] Lewis H. Shoemaker. Fixing the f test for equal variances. *The American Statistician*, 57:105–114, 2003.