GENERAL PAPER

On testing of the homogeneity of variances for two-side power distribution family

A. V. Stepanov[1] · A. G. Chunovkina[1]

Abstract

Some widely used classical tests for homogeneity of variance (Fisher, Link, Cochran, and Bliss–Cochran–Tukey) are considered for the case when the random variable has a symmetric two-side power distribution. The corresponding critical values for the test statistics are obtained. Some comparative analysis of power of the tests is carried out. The case when the null hypothesis assumes a minor random scatter of the variance is considered separately.

Keywords Hypothesis testing · Homogeneity of variance · Test power · Monte Carlo method

Introduction

The GUM (Guide to the expression of Uncertainty in Measurement) is widely applied for evaluating the quality of measurement in different areas. Recently, ISO published ISO/TS 20914:2019 [1] for estimating measurement uncertainty in medical laboratories. This guidance follows the general GUM framework but provides simple practical approach for calculation of measurement uncertainty using available information, so called "top-down" approach. ISO/TS 20914 considers two main uncertainty inputs. One relates to calibrators (which are used by the laboratory for "in-house" internal calibration of an analytical system) and corresponding information about assigned values and associated uncertainties should be provided by a manufacturer. The other relates to laboratory imprecision for a long-term period, which is evaluated using internal quality control data. Evaluation of the second input needs collecting, analyzing and processing a huge number of data to obtained pooled estimates of imprecision and then updating these estimates when new information becomes available. When combining these datasets, the question often arises about

their homogeneity (or almost homogeneity, acceptable from a practical point of view).

The work considers the criteria for variance homogeneity and consists of two chapters. The first provides some critical values for certain variance homogeneity tests, obtained under assumption that the values of the samples has a symmetric two-side power (TSP) distribution [2]. Some comparative analysis of the tests power is provided as well. The approach applied in this section is similar to one used in [3–5].

The second section contains the similar results for some variance proximity tests, this can be useful when exact equality of variances is not required. We consider the case when the null hypothesis assumes a minor random scatter of the variance around some constant value.

TSP distribution

A pdf (probability density function) for a symmetric TSP distribution, having a unity variance, is given by the following expression:

$$
f(x) = \begin{cases} \frac{p}{2x_0}\left(1 + \frac{x}{x_0}\right)^{p-1}, & -x_0 < x < 0, \\ \frac{p}{2x_0}\left(1 - \frac{x}{x_0}\right)^{p-1}, & 0 \leq x < x_0, \\ 0, & x = \pm x_0, \quad p > 1, \end{cases}
$$

$$
x_0 = \sqrt{\frac{(p+1)(p+2)}{2}}.
$$

✉ A. V. Stepanov
stepanov17@yandex.ru

A. G. Chunovkina
A.G.Chunovkina@vniim.ru

1 D.I. Mendeleev Institute for Metrology (VNIIM),
Saint Petersburg, Russia

Here p is the distribution power parameter ($p = 1$ corresponds to a uniform distribution case, and $p = 2$ corresponds to a triangular one). This family of distributions is quite wide and at the same time has an uncomplicated mathematical description that facilitates modeling.

It is expected that the value of the parameter p for each specific application case can be estimated based on the information about the experimental data accumulated in laboratory (see [2]). The plots for some pdfs are given in Fig. 1 (the pdf for a normal distribution is also given for comparison).

Variance homogeneity tests

Variance homogeneity tests for two samples

Let us have two data samples x_1, x_2 from the same TSP(p) distribution, of length n_1 and n_2, correspondingly. The test hypothesis of the variance homogeneity (uniformity) is assumed to be the equality

$$H_0: \quad \sigma_1 = \sigma_2$$

(and the competing hypothesis H_1 implies inequality of variances).

Consider two variance homogeneity tests. The first one is the Fisher test (of course, here we cannot say that it is exactly the classic Fisher test, since the distribution of sample values differs from normal); corresponding test statistic is given by

$$F = s_1^2 / s_2^2,$$

where s_i^2 are the sample variances for x_i. The other is the Link test [6],

$$\tilde{F} = \omega_1 / \omega_2,$$

where ω_i are the sample ranges, $\omega_i = x_{i,\max} - x_{i,\min}$.

Here we consider two-tailed test, and for given critical values $c(n_1, n_2)$ the checks look like

$$\left(c(n_2, n_1)\right)^{-1} \le F, \tilde{F} \le c(n_1, n_2), \quad \text{if} \quad F, \tilde{F} \ge 1,$$
$$\left(c(n_1, n_2)\right)^{-1} \le F, \tilde{F} \le c(n_2, n_1), \quad \text{otherwise.}$$

In case when $n_1 = n_2 = n$ one can simply test if

$$c^{-1} \le F, \tilde{F} \le c, \quad c = c(n).$$

Some critical values (obtained by the Monte Carlo method) corresponding to the probability level of $P_0 = 0.95$ could be found in Table 1; some other (for unequal n_1, n_2) could be found at [7, 8].

Comparing the test power

To estimate and compare the power of the tests let us consider the probability of rejecting the null hypothesis when the variance ratio is given (and differs from unity):
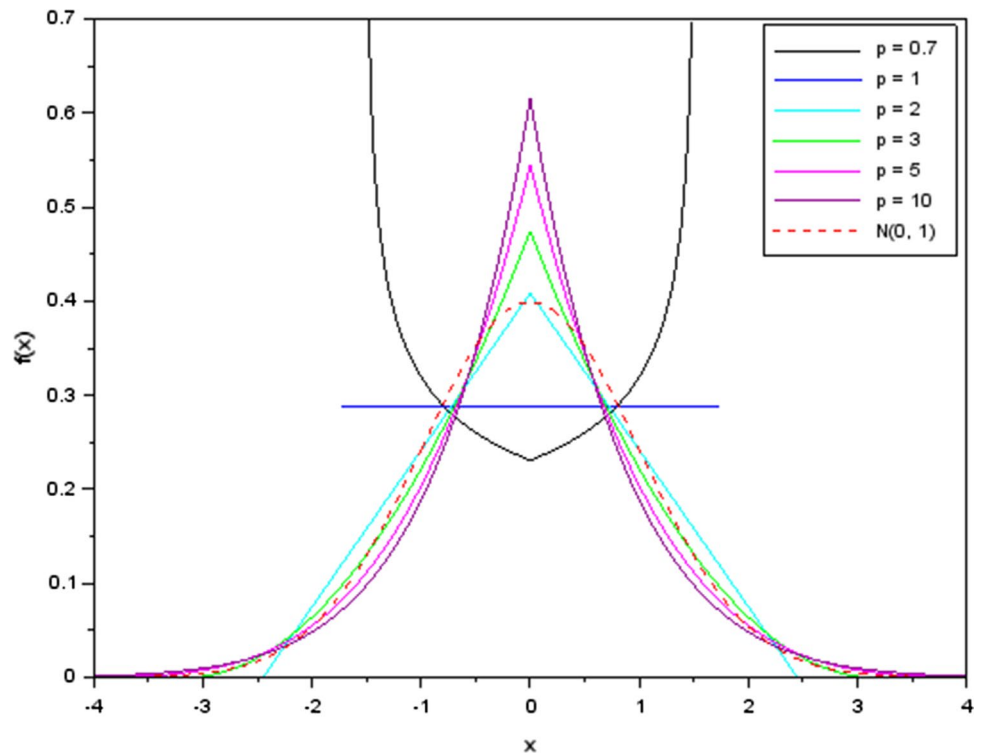
Fig. 1 TSP pdfs

Table 1 Critical values, $c = c(n)$

Test	n	p					
		0.7	1	2	3	5	10
	10	2.472	2.734	3.478	4.047	4.791	5.663
Fisher	15	1.933	2.114	2.600	2.956	3.423	3.976
	20	1.720	1.862	2.230	2.495	2.842	3.255
	10	1.415	1.545	1.871	2.081	2.322	2.566
Link	15	1.220	1.329	1.605	1.783	1.990	2.202
	20	1.143	1.235	1.478	1.639	1.827	2.021

$$P = P(k) = P\{H_0 \text{ rejected } \mid \sigma_2 = k\sigma_1, \ k > 1\}. \quad (1)$$

Hereinafter, it is assumed that the samples are taken from the TSP distribution having the same value of the parameter p (i.e., (1) assumes the "scaling" of the distribution rather than varying the distribution law due to the change of TSP parameter, and P represents the probability of detecting such "scaling").

The plots of $P(k)$ (also obtained by the Monte Carlo method) are given in the below Figs. 2, 3. The case of normal distribution is considered for comparison.

It could be noted that the power of the Link test is higher when the TSP power parameter is less than two; otherwise the Fisher test is preferable. The case $p = 3$ is closest to the case of normal distribution of all the considered above (but, of course, an even closer one can be found by varying the parameter p).

Variance homogeneity multi-sample tests

Let us have m data samples x_i of length n from the same TSP distribution. Here, the null hypothesis has the form

$$H_0: \quad \sigma_1 = \sigma_2 = \ldots = \sigma_m$$

(the competing hypothesis H_1 implies inequality of at least some pair of variances). Consider the following one-tailed variance homogeneity criteria:

Cochran test [9]:

$$C = \frac{\max_{1 \le i \le m} s_i^2}{\sum_{i=1}^{m} s_i^2}$$

Bliss-Cochran-Tukey (BCT) test [10]:
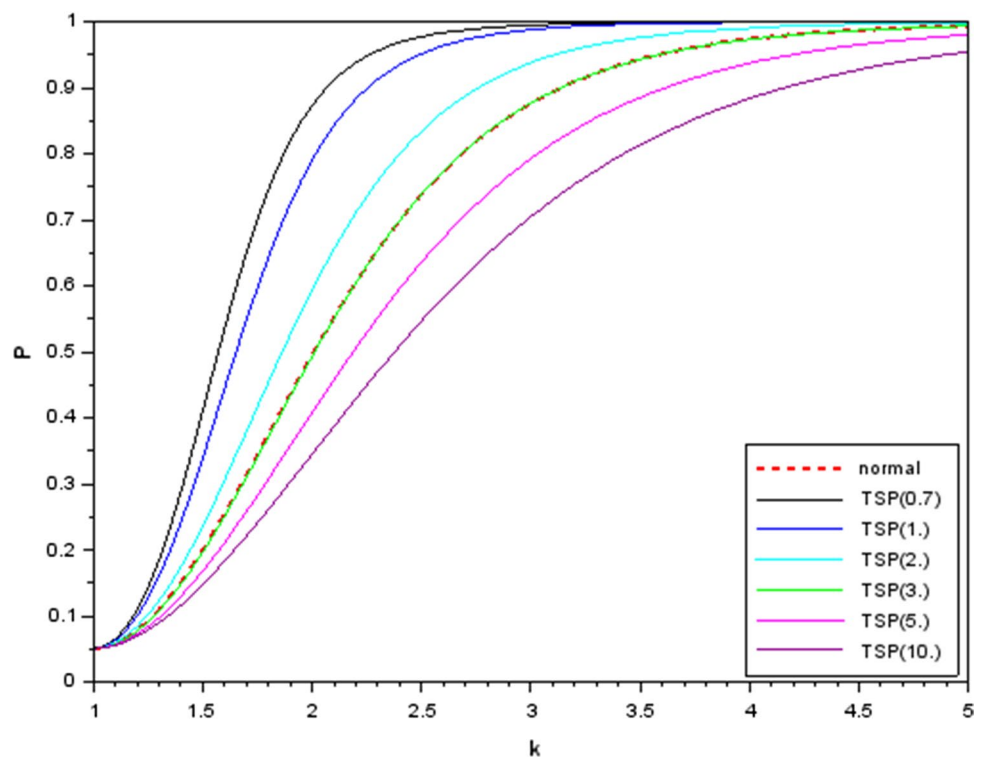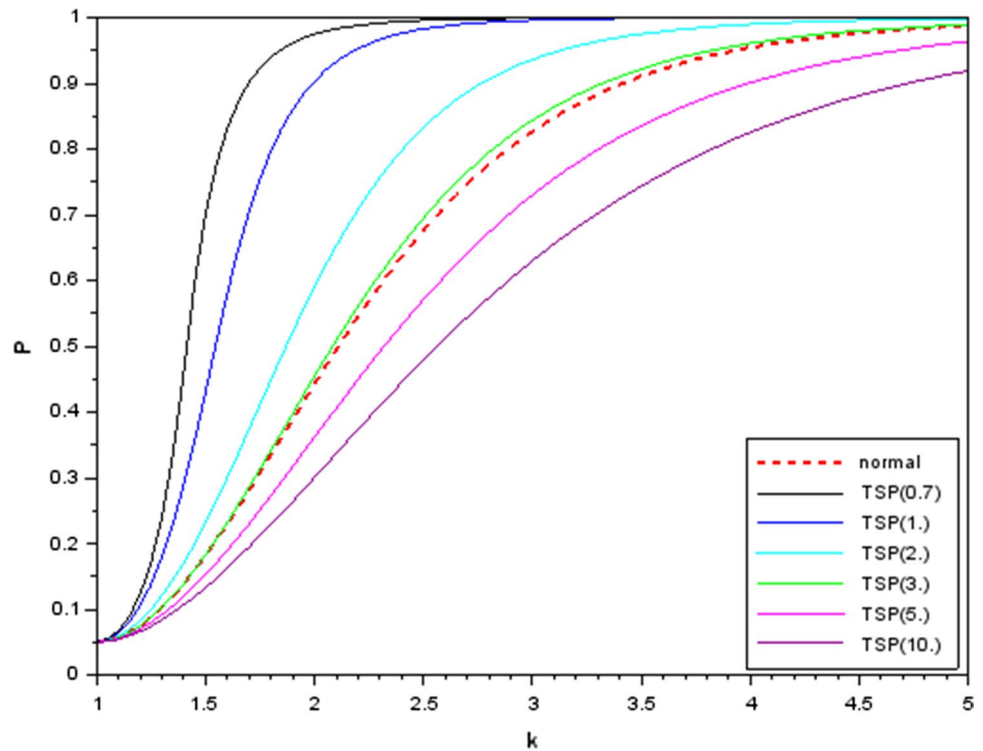
Fig. 2 $P(k)$, fisher test

Fig. 3 $P(k)$, link test

$$\tilde{C} = \frac{\max_{1 \le i \le m} \omega_i}{\sum_{i=1}^{m} \omega_i}$$

Some critical values (corresponding to a probability level of 95 %) for the tests are provided in Table 2.

In case of multiple samples formula (1) does not provide estimation of the test power, as we can have an outlier of variance for more than one sample. Nevertheless, consider a particular trivial case of single outlier (here, again, the parameter p is assumed to be immutable):

Table 2 Critical values, Cochran and Bliss–Cochran–Tukey tests

Test	m	n	p					
			0.7	1	2	3	5	10
		10	0.508	0.530	0.580	0.611	0.647	0.682
	3	15	0.323	0.342	0.386	0.414	0.449	0.489
		20	0.171	0.184	0.213	0.234	0.260	0.293
		10	0.465	0.485	0.528	0.555	0.587	0.620
Cochran	5	15	0.294	0.309	0.345	0.368	0.397	0.431
		20	0.279	0.292	0.323	0.343	0.367	0.397
		10	0.171	0.184	0.213	0.234	0.260	0.293
	10	15	0.155	0.164	0.188	0.204	0.224	0.250
		20	0.146	0.154	0.174	0.187	0.204	0.226
		10	0.394	0.413	0.456	0.481	0.507	0.531
	3	15	0.368	0.385	0.426	0.451	0.477	0.501
		20	0.357	0.372	0.401	0.434	0.459	0.484
		10	0.236	0.250	0.285	0.305	0.328	0.351
BCT	5	15	0.220	0.232	0.264	0.283	0.305	0.328
		20	0.214	0.224	0.253	0.271	0.293	0.314
		10	0.117	0.125	0.147	0.161	0.177	0.194
	10	15	0.110	0.116	0.135	0.148	0.163	0.179
		20	0.106	0.112	0.129	0.141	0.155	0.170

$$P = P(k) = P\{H_0 \text{ rejected } | \sigma_2 = \ldots = \sigma_m, \ \sigma_1 = k\sigma_2, \ k > 1\}.$$

Consider an example, $m = 3$, $n = 5$. A comparison of power for the Cochran and the Cochran-Bliss-Tukey tests is presented in the corresponding Fig. 4.

It can be noted again that the first or the second test may be preferred (as more powerful) depending on parameter p value (again, roughly, $p = 2$ is some "equilibrium" point).

Variance proximity tests: random scatter of the variances

If a homogeneity test passes, the pooled estimate of imprecision is calculated. In practice, quite often the pooled estimate is calculated when the critical values of the criteria are slightly exceeded, but the variances of the series differ insignificantly. So, let us test a null hypothesis which allows a small variability in the series variances.

Consider the following trivial case. Let the variance of the distribution varies in a random way between the subsequent samples, i.e.,

$$\sigma_1 = k_1\sigma_0, \quad \sigma_2 = k_2\sigma_0, \quad k_{1,2} \sim U(1 - d, 1 + d), \tag{2}$$

here U is for a uniform distribution, and d is a minor positive deviation (below we will consider d from a range

[0.05, 0.20], i.e., σ deviation is $\pm 5\%$...20% of σ_0). Again, we consider the TSP parameter p to be immutable, i.e., these changes of the deviation are not due to fluctuations in the form of the distribution law (this special case can be considered separately).

This model, on the one hand, is in good agreement with practical needs, and, on the other hand, is well suited for the Monte Carlo method (it allows simple randomization of samples, followed by obtaining the test statistics and, after some averaging, the corresponding critical values). Of course, this trivial model allows the further modification (another distributions for d, systematic effects, etc.).

The acceptance of the variance homogeneity hypothesis when the random effect is neglected

As a starting point, let us estimate the probability of acceptance of the "strict" null hypothesis neglecting the random effect

$$\tilde{H}_0 : \quad \sigma_1 = \sigma_2$$

when in fact it is present according to (2):

$$\tilde{P} = P\{\tilde{H}_0 \text{ accepted } | (2)\}$$

Some corresponding probability values when using the Fisher test are given in the below Table 3.

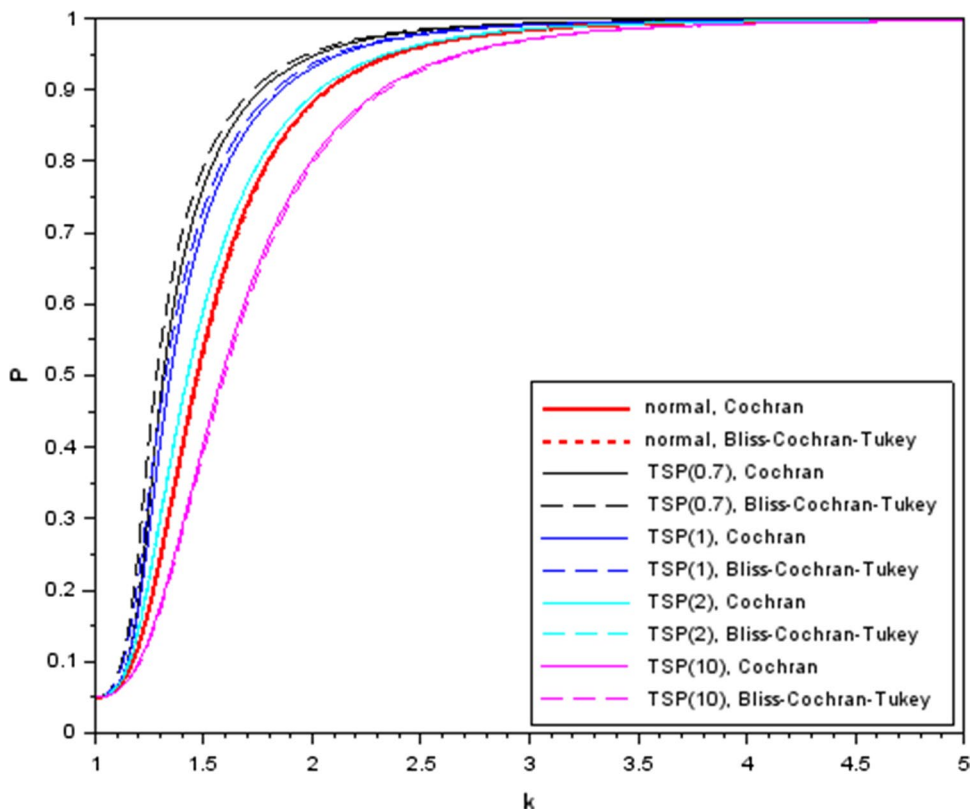**Fig. 4** $P(k)$, single outlier, $m = 3, n = 5$

Table 3 Some values for $\tilde{P}$, Fisher test

Distribution	d	n					
		5	10	20	50	100	200
	0.05	0.949	0.948	0.943	0.928	0.903	0.853
TSP(1)	0.10	0.948	0.941	0.920	0.859	0.768	0.639
	0.20	0.941	0.907	0.824	0.650	0.505	0.380
	0.05	0.950	0.948	0.946	0.937	0.923	0.895
TSP(2)	0.10	0.948	0.943	0.933	0.897	0.840	0.741
	0.20	0.941	0.921	0.875	0.749	0.612	0.475
	0.05	0.950	0.949	0.948	0.943	0.934	0.919
TSP(5)	0.10	0.948	0.945	0.940	0.921	0.888	0.824
	0.20	0.943	0.932	0.906	0.828	0.721	0.586
	0.05	0.950	0.949	0.947	0.941	0.931	0.911
$N(0, \sigma_0)$	0.10	0.948	0.944	0.937	0.912	0.872	0.796
	0.20	0.942	0.926	0.893	0.799	0.680	0.542

It could be noted that for relatively low values of n and d the probability do not differ significantly from $P_0 = 0.95$. i.e., someone using the Fisher test (being unaware of the random effect (2)) hardly notice the difference for relatively small values of d (0.05) and n (5..20). However, while increasing d and n, this effect becomes more and more significant, requiring the calculation of the correct critical values (which is done below).

For the other tests the situation is the same.

Two-sample tests: critical values and power of the tests

Some critical values $c(d, n)$ for $n_1 = n_2 = n$ for the Fisher and Link tests are given in the following Table 4. The case of normal distribution is also considered.

Note that in the range of relatively small values of d the critical values for the criteria also vary insignificantly. For example, the following Fig. 5 contains plots of the ratio $c(d, n) / c(0, n)$ for the Fisher and Link tests as a function of the parameter d, when $n = 10$. It can be seen that for almost all cases here the increment of c does not exceed 5 %, when $d \in (0, 0.1]$ (and even for larger d sometimes).

The plots of $P(k)$ for $n_1 = n_2 = 10$, $d = 0.1$ (see (1)) are given in the below Figs. 6 and 7. As it can be seen, the values of P do not differ sufficiently from the case when the random scatter of the standard deviation is absent (but the test power is expectedly lower as compared with the case of zero d).

Multi-sample tests

Let us restrict ourselves here to providing of some critical values for the Cochran and Cochran–Bliss–Tukey criteria. The values are given in Table 5.

Conclusion

Long-term imprecision variance is used by analytical, testing, medical and other laboratories in their internal quality control procedures. This information is also very valuable for uncertainty calculation in "top-down" approach. So, a laboratory should obtain and update the variance estimates of precision which are used in their internal quality control. The widely used ANOVA and corresponding

Table 4 Critical values, Fisher and Link tests

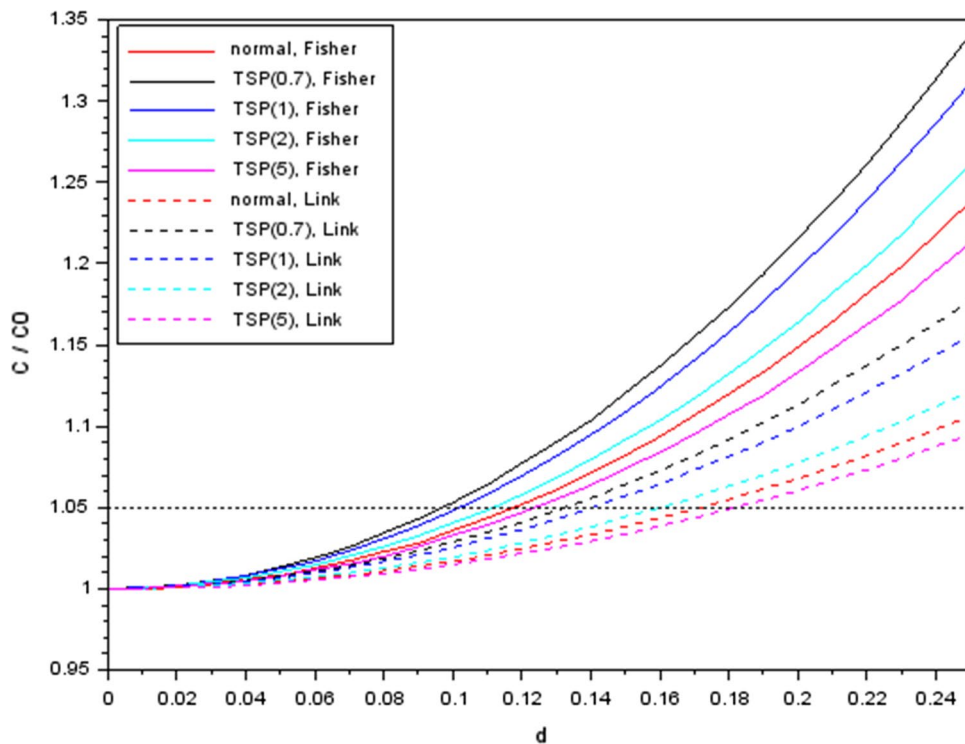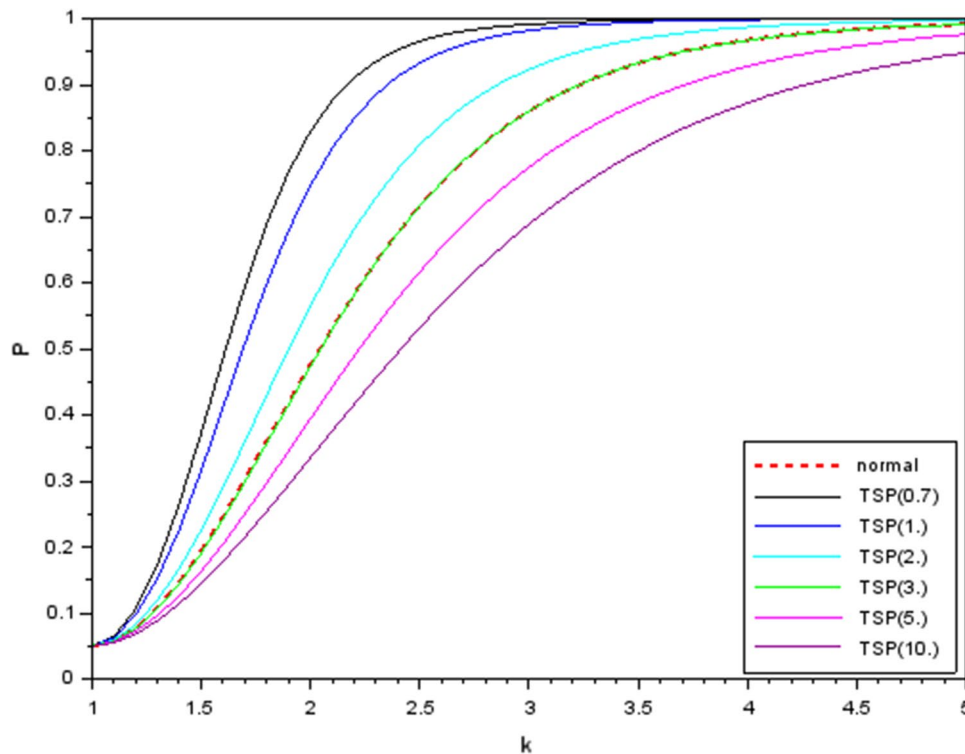| Test	d	n	p						$N(0, \sigma_0)$
			0.7	1	2	3	5	10	
Fisher	0.1	10	2.603	2.866	3.616	4.192	4.947	5.830	4.172
		20	1.873	2.008	2.369	2.632	2.980	3.396	2.665
	0.2	10	3.003	3.270	4.044	4.642	5.430	6.352	4.627
		20	2.298	2.426	2.781	3.046	3.401	3.831	3.081
Link	0.1	10	1.455	1.584	1.907	2.117	2.357	2.601	2.146
		20	1.220	1.296	1.524	1.680	1.865	2.058	1.767
	0.2	10	1.575	1.699	2.015	2.223	2.462	2.707	2.253
		20	1.400	1.452	1.652	1.800	1.979	2.168	1.884

Fig. 5 $c(d, n)/c(0, n)$, $n = 10$

Fig. 6 $P(k)$, fisher test, $n = 10$, $d = 0.1$

software should be accompanied by analysis of homogeneity of imprecision variance. The relating criteria are available but the critical intervals are based on the assumption of normality. In this paper we investigated the performance of the well-known criteria for TSP family distribution and discussed their properties and limitations of usage.

The work allows an expansion in the direction of considering other various criteria for the variance homogeneity (see, e.g., [3, 4]). Some further generalizations of
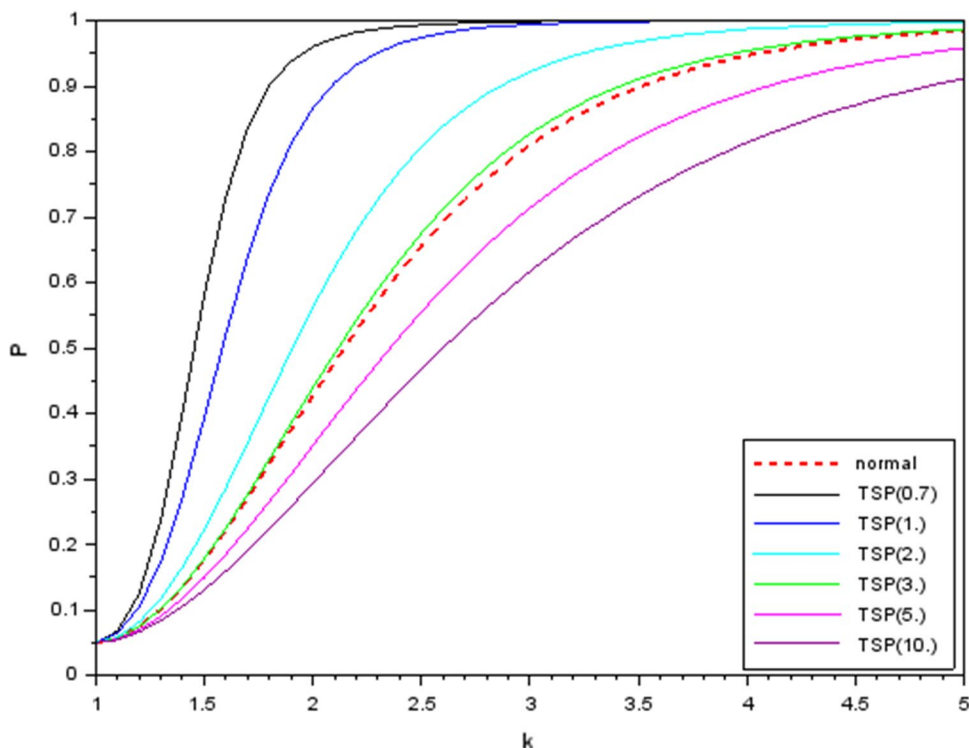
Fig. 7 $P(k)$, link test, $n = 10$, $d = 0.1$

Table 5 Critical values, Cochran and Bliss–Cochran–Tukey tests

Test	d	m	n	p						$N(0, \sigma_0)$
				0.7	1	2	3	5	10	
Cochran	0.1	3	10	0.522	0.543	0.591	0.620	0.654	0.688	0.625
			20	0.465	0.479	0.514	0.536	0.563	0.592	0.542
		5	10	0.337	0.355	0.397	0.424	0.458	0.496	0.433
			20	0.297	0.308	0.336	0.355	0.379	0.407	0.362
	0.2	3	10	0.558	0.576	0.617	0.644	0.674	0.705	0.648
			20	0.511	0.523	0.551	0.570	0.593	0.618	0.574
		5	10	0.370	0.386	0.424	0.449	0.481	0.516	0.457
			20	0.333	0.343	0.368	0.385	0.406	0.432	0.391
BCT	0.1	3	10	0.403	0.420	0.461	0.485	0.511	0.535	0.493
			20	0.374	0.385	0.418	0.440	0.465	0.488	0.455
		5	10	0.245	0.257	0.290	0.310	0.332	0.354	0.320
			20	0.227	0.235	0.260	0.277	0.297	0.318	0.292
	0.2	3	10	0.424	0.439	0.476	0.498	0.521	0.544	0.505
			20	0.403	0.411	0.438	0.457	0.478	0.500	0.470
		5	10	0.262	0.273	0.302	0.321	0.342	0.363	0.330
			20	0.248	0.254	0.275	0.290	0.309	0.328	0.303

the random model (2) could be considered as well. For example, [11] contains critical values for some variance proximity tests when the sample values are normally distributed, and the scatter of standard deviation has a TSP distribution.

References

1. ISO/TS 20914:2019, Medical laboratories – practical guidance for the estimation of measurement uncertainty
2. van Dorp JR, Kotz S (2002) The standard two-sided power distribution and its properties. Am Statistician 56(2):90–99
3. Lemeshko BY, Sataeva TS (2017) Application and power of parametric criteria for testing the homogeneity of variances Part III. Meas Tech 60(1):7–14

4. Sataeva TS, Lemeshko BY (2016) About properties and power of classical tests of homogeneity of variances, 11th International Forum on Strategic Technology (IFOST) 350–354

5. Lemeshko BYu, Mirkin EP (2004) Bartlett and Cochran tests in measurements with probability laws different from normal. Meas Tech 47:960–968

6. Link RF (1950) The sampling distribution of the ratio of two ranges from independent samples. Ann Math Statistics 21(1):112–116

7. https://github.com/stepanov17/varTests/blob/master/homogeneity/criticalValuesFisher.pdf

8. https://github.com/stepanov17/varTests/blob/master/homogeneity/criticalValuesLink.pdf

9. Cochran WG (1941) The distribution of the largest of a set of estimated variances as a fraction of their total. Ann Eugen 11:47–52

10. Bliss CI, Cochran WG, Tukey JW (1956) A rejection criterion based upon the range. Biometrika 43:418–422

11. https://github.com/stepanov17/varTests/blob/master/proximity/TSPVarScatter/TSPVarScatter.pdf