Valuation of the normality of distribution in metrology using the fractal principle

To cite this article: T Stejskal et al 2018 J. Phys.: Conf. Ser. 1065 212008

View the article online for updates and enhancements.
Valuation of the normality of distribution in metrology using the fractal principle

T Stejskal¹, M Dovica¹, S Šuriš², R Palenčár² and J Palenčár²

¹The Technical University of Košice, Faculty of Mechanical Engineering, Slovakia, ²Slovak University of Technology in Bratislava, Faculty of Mechanical Engineering, Slovakia

E-mail: miroslav.dovica@tuke.sk,

Abstract. The questions of normality of distribution play a significant role in metrology. On the basis of the central limit theorem, large number of combinations of multiple distributions leads to an overall normal distribution. Deviation from the normal distribution is also significant in measurement because it points to hidden properties of the measurement process. If the measured distribution is not a normal one, there is at least one dominant influence in the measurement process that affects the measurement result. This effect may also be of diagnostic importance, as it might indicate a possibly wrong measurement procedure. It can also point to specific properties of a measured object that could have been otherwise overlooked. This paper presents a method based on the fractal principle, which supports evaluation the normality of the population. The normality evaluation method is based on the assumption that the population sequences under evaluation are self-similar.

1. Introduction

There is a whole range of tests and methods for evaluating normal distribution. The simplest is the graphical method. Less obvious but more accurate are statistical methods. The regression test is the Shapiro-Wilk (SW) test [1]. This type of tests is based on the distance of individual points from the regression line, it is also suitable for populations with fewer than 50 samples. The tested statistics is determined by the relationship derived from the regression curve parameters. Another method uses good compliance tests, which are based on empirical and hypothetical distribution functions. The main test used in this area is the Kolmogorov-Smirnov test [2]. It is based on a match comparison. A great disadvantage of this test is that we need to know the exact values of the standard deviation and the mean value. Chi-square test [3] was one of the first normality tests. It is a universal test for discrete and continuous distribution functions with a sufficiently large n range. It is applied to a lesser degree nowadays because its test strength is weaker.

The new approach to normality testing proposed by the authors is the assumption of an internal distribution structure that is self-similar [4]. This means that the distribution nature of the selected section of consecutively measured data is similar to total distribution. This principle is met by every fractal. In order to verify this principle, it is appropriate to compare derived distributions from the original distribution. This procedure is based on the assumption that, in distribution, the subsequent

¹ Miroslav Dovica E-mail: miroslav.dovica@tuke.sk
value of a sample does not depend on the previous value. So, changing the order of the samples, if it is completely random, should result in the original distribution. Therefore, it follows that even distribution created by subtracting adjacent samples should display the character of a normal distribution. The proposed test also takes into account the time factor, which gives a more stringent normality requirement than the one resulting from the central limit theorem.

2. Fractal principle testing procedure
The proposed fractal test of normality consists of two analyses
- The first one tests the independence of neighboring samples of the population.
- The second one checks the normality of sample sequence with SW test.

2.1. Test based on verifying the independence of neighboring samples
Let us suppose we have a sequence \( x_n, x_{n+1}, \ldots \) of measured values, for example, of a dimension.
\[
\{x_n\}_{n=1}^\infty = \{x_1, x_2, x_3, \ldots\}
\]
Let us create a derived sequence by subtracting adjacent samples.
\[
y_n = x_n - x_{n+1}
\]
(1)
If there is no dependency between the values in the order of the measurement, then we will get a new population of values with independent samples \( y_n, y_{n+1}, \ldots \). The new population can be compared to the original population. The simplest way is to compare the standard deviations and mean values. This method shall yield two independent populations. The theory of comparison of two independent populations asserts that if \( X \) and \( Y \) are independent variables, the dispersion will be
\[
D_X = D_Y = s^2
\]
\[
D_{(X,Y)} = D_X + D_Y = 2s^2
\]
(2)
or alternatively, the following holds for the ratio \( p \) of standard deviations
\[
p = \frac{s_Y}{s_X} = \sqrt{2} = 1.4142
\]
(3)
where \( s, s_X, s_Y \) are the corrected sample standard deviations. The difference between the measured value \( p \) and the theoretical value of 1.4142 is tested. The deviation from the independence of neighboring samples can be expressed as a relative error as:
\[
\delta = \frac{|p - 1.4142|}{1.4142}
\]
(4)
The parameter \( \delta \) has many independent random effects, so it will have a normal distribution. The boundary of assuming a zero-hypothesis depending on the number of population members can be mathematically derived, but the most appropriate way is to use the Monte Carlo simulation [5].

2.2. Test based on verification of the normality of the sample sequences
The modified selection procedure lies in comparing the same adjacent segments of the measurement sequence. In this case, too, the distributions should be similar, independent of the size of the selected segment. Here, it is assumed that even the subdivisions that created the overall distribution were also normal. This idea is based on the structure of binomial distribution, which approximates the shape of the normal distribution. The binomial distribution is fractal in nature because it is self-similar in every
part. The Galton plate is the obvious proof of this claim. The Galton plate is a mechanical model of binomial distribution. The principle lies in the likelihood of choosing the trajectory of a free-falling ball through the partitions of the board. The partition structure of the board can be divided into different self-similar objects that have the same function (Fig.1).

![Self-similar object](image)

**Figure 1.** Binomial distribution with Galton plate.

On the basis of the above, the selection of the center of the first step of the trajectory does not matter. The result will still be binomial distribution. Because the binomial distribution approximates the normal distribution, similar considerations can be used for normal distribution as well. It does not matter in what order the members will be evaluated or in what population size. If this is a normal distribution, the result of each confusion will also be normal distribution. Since the condition of self-similarity in random selection is met, the self-similarity in sequential selection must hold, too. Individual sequences of the selected sequences of the measured values should have the same distribution parameters as the entire distribution. This consideration is the basis for the method of checking the normality of the population data. To test data sequences, a comparison of standard deviations according to the formula (3) is used.

3. **Example of modeling of test values and verification of the method**

A Random Number Generator with Normal Distribution was used to test the method. 1000 samples of normal distribution with standard deviation $s = 1$ were generated. By using formulas (1) and (3) we have obtained a good result $p = 1.4434$. In the comparison sample $p_{nm}$, the first index ($n$) refers to the number of sequences in that population and the second index ($m$) expresses order of the sequences. The entire population is divided into the same sequences. Individual sequence sizes at the three levels are 1000, 500, and 250 samples (Fig.2a). Subsequently, the distribution was artificially modified so that the individual samples of sequences were aligned according to their size (Fig.2b). The independence of successive samples has thus been broken. An interesting observation in this modification is that the histogram will not change (Fig.3a). So SW test does not detect the change in the population. The population as a whole will have unchanged statistical parameters. As a result, the modified population is considered a normal distribution population.

![Histograms](image)

**Figure 2.** Original (a) and Modified (b) log of the normal distribution samples.
The problem of normality evaluation is that only the overall shape of the distribution of random values is taken into account by the existing analyses, which do not take into account the internal structure of the resultant population. So, there can be distributions with the same shape, but with a different structure of creation.

![Figure 3. (a) Random distribution histogram (b) Parameter $P_{am}$](image)

However, the fractal test method shows significant variations in the standard deviation ratio (tab.1), (Fig.3b) and at the same time small SW test deviations (all tested distributions were normal). The experimental population was tested on a beam with a strain gauge sensor. Deviations were elicited by a timekeeper. The object was subjected to several independent impacts, such as strike force, strike timing, magnitude of beam vibration at time of stroke, and the time of the maximum value recording. These impacts should create a normal distribution of maximum beam swings.

### Table 1. Changes in the $P_{am}$ parameter and SW test in the original, modified population and the experimental population.

|        | Original population | Modified population | Experimental population |
|--------|---------------------|----------------------|-------------------------|
| $P_{00}$ | 1.4434              | 0.6130               | 1.4002                  |
| $P_{10}$ | 1.4469              | 0.5646               | 1.3344                  |
| $P_{12}$ | 1.4403              | 0.6557               | 1.5008                  |
| $P_{20}$ | 1.4883              | 0.5580               | 1.3416                  |
| $P_{21}$ | 1.4036              | 0.5748               | 1.3677                  |
| $P_{22}$ | 1.4330              | 0.6643               | 1.4879                  |
| $P_{23}$ | 1.4469              | 0.6487               | 1.4953                  |

| SW test 50 samples [W] | 0.9741388 | 0.9741388 | 0.968269 |

### 4. Conclusion

The methodology of fractal evaluation of the normality distribution is based on the fundamentals of the normal distribution creation. The model is essentially a binomial distribution represented by a Galton plate. Unlike other methods, high test strength is expected for different random populations.

### Acknowledgments

This contribution has been supported by research grants: APVV-15-0149 and APVV-15-0295, VEGA 1/0224/18, VEGA 1/0098/18 VEGA 1/0556/18 and VEGA 1/0437/17, KEGA 006TUKE-2/2016 and KEGA039STU-4/2017.

### References

1. Shapiro S S and Wilk M B 1965 An analysis of variance test for normality *Biometrika.* 52 591.
2. Lilliefors H W 1967 On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association,* 62 399.
3. Pearson K 1900 Philosophical Magazine. 50 157.
4. Mandelbrot B B 1982 *The Fractal Geometry of Nature.* New York: WH freeman vol 173 p 468
5. Palenčár R, Ďuriš S, Dovica M and Palenčár J 2015 Application of Monte Carlo method for evaluation of uncertainties of temperature measurement by SPRT. XXI IMEKO World Congress “Measurement in Research and Industry p 1194