Metrol. Meas. Syst., Vol. 31 (2024) No. 2, pp. 279–293
DOI: 10.24425/mms.2024.149701

METROLOGY AND MEASUREMENT SYSTEMS
Index 330930, ISSN 0860-8229
www.metrology.wat.edu.pl

CONSTRUCTION OF ASYMMETRIC CONFIDENCE/UNCERTAINTY INTERVALS USING SEMIVARIANCE ON THE EXAMPLE OF TRAFFIC NOISE INDICATORS

Bartosz Przysucha

Lublin University of Technology, 20-618 Lublin, Poland (✉ *b.przysucha@pollub.pl*)

Abstract

The occurrence of asymmetric probability distributions is quite common. Phenomena such as salary, number of failures, sound level values, *etc.* have skewed distributions. In such cases, estimating the mean using the interval method can be inaccurate as it ignores the distribution's asymmetry. Another method of constructing confidence intervals, which does not require symmetry of distributions, is the method based on Chebyshev's theorem. However, the intervals thus obtained are symmetrical. The approach proposed in this article uses the concept of Chebyshev's theorem and semivariances to construct new confidence and uncertainty intervals. The article examines the properties of semivariance-based confidence intervals for long-term noise indicators from acoustic monitoring of the city of Gdansk and compares them with classical confidence intervals. The new uncertainty assessment tool proposed in this article in the form of a semivariance-based uncertainty interval can therefore be the basis for new uncertainty assessment methodology and more effective uncertainty.

Keywords: metrology, asymmetry confidence intervals, uncertainty evaluation, semivariance, 3σ rule generalization.

1. Introduction

Measurement uncertainty assessment is one of the basic issues in metrology. In the measurement process, there are many measurement uncertainty components that constitute the uncertainty budget [1]. All uncertainties present in the uncertainty budget are aggregated based on the measurement function. The most common method of compounding measurement uncertainty is the uncertainty propagation method [1–4]. Two types of measurement uncertainty can be distinguished: type A uncertainty, which is related to measurement repeatability, and type B uncertainty, which is most often associated with the specification of measurement equipment or its accuracy. Type A uncertainty is expressed in terms of standard deviation or multiples thereof. This value is most often determined by a confidence interval. Confidence intervals, however, can be calculated using, for example, the classic Spława-Neyman definition or using Chebyshev's theorem. The first type of construction imposes a specific required percentage of coverage by the confidence

Article history: received September 27, 2023; revised February 15, 2024; accepted February 19, 2024; available online March 25, 2024.

interval of the true value. The second approach determines the minimum coverage value for a given distribution. The main disadvantage of classic confidence intervals is their symmetry. These intervals are constructed based on the assumption of normality of the characteristic distribution for which the confidence interval is determined. When a feature does not have a normal distribution, the Central Limit Theorem is used as well as the asymptotic behaviour of the mean value. However, in the case of measurements where the confidence interval is determined for small measurement samples, the distribution of $e.g.$ the mean value cannot be assumed to be normal. In this case, the determination of the classic confidence interval is not correct, which manifests itself primarily by not meeting the assumption of coverage by the confidence interval of the unknown parameter. Methods based on Chebyshev's theorem, which does not require an assumption regarding the normality of feature distribution, should then be used. Based on Chebyshev's inequality, the confidence interval created does not have an assumed percentage coverage level, only a minimum coverage level. However, the forms of confidence intervals constructed on the basis of this theorem either have the form of a symmetric confidence interval (1) or refer to forms that do not include sample variation characteristics ($e.g.$ variance). (1–8, 11).

This article focuses on the Chebyshev inequality approach and proposes its generalization using semivariance. The confidence intervals proposed in the paper, created using semivariance, on the one hand consider the asymmetry of characteristic distribution by including left and right semivariance. On the other hand, they are based on Chebyshev's theorem, which does not require assumptions about the normal distribution of the characteristic. This is a new approach not previously used.

Section 3 proposes a form of confidence interval constructed using semivariance. A minimum coverage level was also determined for sample probability distributions. Section 4 proposes the construction of an uncertainty interval using semivariance. Based on measurements from traffic monitoring in the city of Gdansk, a comparison was made between the percentage coverage obtained when applying the classic uncertainty interval and the uncertainty interval for semivariance depending on the measurement sample size.

2. Chebyshev's theorem and its generalizations

Chebyshev's inequality, well known in mathematics and statistics, is primarily used to estimate different types of probabilities when the mean and variance are estimated based on a sample. Let X denote the random variable for which $EX = \mu$ and $D^2X = \sigma^2 < \infty$. From the classical Chebyshev's inequality, also called the Bienaymé–Chebyshev inequality [5], it follows that for any $k > 0$ the following condition is satisfied

$$P(|X-\mu| > k\sigma) < \frac{1}{k^2}, \tag{1}$$

Chebyshev's inequality results directly from Markov's inequality [6]

$$P(|X| > \lambda) < \frac{E(|X|)}{\lambda}, \tag{2}$$

sometimes called Chebyshev's first inequality, after using the substitution $X{:} = (X-\mu)^2$ and $\lambda{:} = (k\sigma)^2$.

Under the additional assumption $\sigma^2 > 0$, Chebyshev's inequality takes the form

$$P(|X-\mu| \geq k\sigma) < \frac{1}{k^2}. \tag{3}$$

Only the case $k > 1$ is useful because for $k \leq 1$ the condition, $\frac{1}{k^2} \geq 1$ holds and the estimation given by the inequality becomes trivial.

Chebyshev's inequality guarantees that for a random variable with any probability distribution, only a certain percentage of the values are further away from the mean than a predetermined value. In particular, no more than $\frac{1}{k^2}$ values of the distribution can be k or more standard deviations from the mean. This rule is often referred to as Chebyshev's theorem, concerning the range of standard deviations around the mean in statistics. Chebyshev's inequality can be applied to any probability distribution where the mean and variance are defined. In this particular case, we know that $P(|X - \mu| > 2\sigma) < \frac{1}{4}$ and $P(|X - \mu| > 3\sigma) < \frac{1}{9}$, which means that at least 75% of the values of the random variable must be within two standard deviations of the mean and at least 88.89% within three standard deviations. The practical application of Chebyshev's inequality is similar to the well-known rule 68-95-99.7, which, however, applies only to normal distributions.

Many generalizations of Chebyshev's inequality have been developed. For example, lower limits are given for the probability of intervals that are not necessarily symmetric around the mean

$$P(l < X < u) \geq \frac{4[(\mu - l)(u - \mu) - \sigma^2]}{(l - u)^2}, \tag{4}$$

if $(\mu - l)(h - \mu) \geq \sigma^2$ and $(\mu - l)(h - \mu) - k^2 \leq 2\sigma^2$, where $k = \min(\mu - l, h - \mu)$ and $l < \mu < h$ [7,8], which for the case of symmetric intervals coincide with Chebyshev's inequality.

Inequalities have been proved for the case of two random variables that need not be independent [8]:

$$P\left(l_1 \leq \frac{X_1 - \mu_1}{\sigma_1} \leq u_1, l_2 \leq \frac{X_2 - \mu_2}{\sigma_2} \leq u_2\right) \geq 1 - \frac{4 + (u_1 + l_1)^2}{(u_1 - l_1)^2} - \frac{4 + (u_2 + l_2)^2}{(u_2 - l_2)^2}. \tag{5}$$

Inequalities are also given for the case of a bivariate distribution when the variables are correlated [9], and ρ denotes the correlation coefficient between them:

$$P\left(\bigcap_{i=1}^{2}\left[\frac{|X_i - \mu_i|}{\sigma_i} < k\right]\right) \geq 1 - \frac{1 + \sqrt{1 - \rho^2}}{k^2}. \tag{6}$$

More general estimates for the two correlated variables were also obtained by Lal [10]:

$$P\left(\bigcap_{i=1}^{2}\left[\frac{|X_i - \mu_i|}{\sigma_i} \leq k_i\right]\right) \geq 1 - \frac{k_1^2 + k_2^2 + \sqrt{(k_1^2 + k_2^2)^2 - 4k_1^2 k_2^2 \rho}}{2(k_1 k_2)^2}. \tag{7}$$

Isii [11] has shown that if

$$Z = ((-k_1 < X_1 < k_2) \cap (-k_1 < X_2 < k_2)), \tag{8}$$

$$\lambda = \frac{k_1(1 + \rho) + \sqrt{(1 - \rho^2)(k_1^2 + \rho)}}{2k_1 - 1 + \rho}. \tag{9}$$

where $0 < k_1 \leq k_2$, is:

1. if $2k_1^2 > 1-\rho$ and $k_2 - k_1 \geq 2\lambda$, then

$$Z \leq \frac{2\lambda^2}{2\lambda^2 + 1 + \rho}; \tag{10}$$

2. if the above conditions are not satisfied, *i.e.* $2k_1^2 \leq 1-\rho$ or $k_2 - k_1 < 2\lambda$ but $k_1k_2 \geq 1$ and $2(k_1k_2 - 1)^2 \geq 2(1-\rho^2) + (1-\rho)(k_2-k_1)^2$, then

$$Z \leq \frac{(k_2-k_1)^2 + 4 + \sqrt{16(1-\rho^2) + 8(1-\rho)(k_2-k_1)}}{(k_1+k_2)^2}; \tag{11}$$

3. if none of the above applies, there is no universal bound other than 1.

Inequalities were also derived for the case of multiple independent variables [12] or correlated variables with known correlation coefficients between each pair of variables [13–15]. Numerous modifications of Chebyshev's inequality using, for example, moments of higher orders or bounded random variables were given.

The universality of Chebyshev's inequality allows it to be used for any probability distribution, although the resulting estimates are not always optimal. In many cases, for certain types of distributions, these estimates can be improved [16]. Estimates are much better for variables with a standardized normal distribution, for example [17].

An alternative way to improve estimates is to use semivariance. The upper (σ_+^2) and lower (σ_-^2) semivariance are defined as follows:

$$\sigma_+^2 = E(X-\mu)_+^2 \text{ and } \sigma_-^2 = E(X-\mu)_-^2, \tag{12}$$

where $(X-\mu)_+ = \max\{0, (X-\mu)\}$, $(X-\mu)_- = \min\{0, (X-\mu)\}$. Of course, $\sigma^2 = \sigma_+^2 + \sigma_-^2$.

Chebyshev's inequality using lower semivariance [18]

$$P(X-\mu \leq -k\sigma_-) \leq \frac{1}{k^2} \tag{13}$$

after substitution $k := \frac{k\sigma}{\sigma_-}$ has the following form

$$P(X-\mu \leq -k\sigma) \leq \frac{1}{k^2}\frac{\sigma_-^2}{\sigma^2}. \tag{14}$$

The use of semivariance improves Chebyshev's inequality in this case $\frac{\sigma_-^2}{\sigma^2}$ times.

A similar estimate is also true for the upper semivariance, so by taking $\sigma_u^2 = \max\{\sigma_-^2, \sigma_+^2\}$ we obtain

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}\frac{\sigma_u^2}{\sigma^2}. \tag{15}$$

Semivariance has long been used in downside risk estimating in finance [19–22]. Markowitz argued that "semivariance is a more reliable measure of investment risk" than the mean-variance theory he developed [23]. Nowadays, semivariance is also successfully applied in agricultural sciences [18].

For symmetric distributions, the condition $\sigma_u^2 = \sigma_-^2 = \sigma_+^2 = \frac{1}{2}\sigma^2$ gives a known estimate for a variable with a normal distribution. Due to the inequality of $\sigma_u^2 \leq \sigma^2$ the use of semivariance for asymmetric distributions significantly improves the estimate.

Metrol. Meas. Syst.,Vol. 31 (2024), No. 2, pp. 279–293
DOI: 10.24425/mms.2024.149701

3. Construction of confidence intervals based on semivariance

For many distributions, especially asymmetric ones, there is a need to modify the classical 3σ rule. The following theorem, using semivariance, is a proposal for such a modification.

3.1. Theorem (Chebyshev's inequality for semivariance)

Let X be a continuous variable for which $EX = \mu$ and $D^2X = \sigma^2 < \infty$. Then

$$P(-k\sigma_- \le X - \mu \le k\sigma_+) \ge 1 - \frac{1}{k^2}\left(\frac{1}{P(X \ge \mu)} + \frac{1}{P(X < \mu)}\right). \tag{16}$$

Normally, for the 3σ interval we take $k = 3$, but in this case we have $\sigma^2 = \sigma_+^2 + \sigma_-^2$. For $k = 6$ finally we have

$$P(-6\sigma_- \le X - \mu \le 6\sigma_+) \ge 1 - \frac{1}{6^2}\left(\frac{1}{P(X \ge \mu)} + \frac{1}{P(X < \mu)}\right). \tag{17}$$

In the case of a symmetric distribution, we have $P(X \ge \mu) = P(X < \mu) = \dfrac{1}{2}$. Thus, the following condition holds

$$P(-6\sigma_- \le X - \mu \le 6\sigma_+) = P(-3\sigma \le X - \mu \le 3\sigma) = 1 - \frac{1}{6^2} \cdot 4 = \frac{8}{9}. \tag{18}$$

Based on the inequality between the arithmetic mean and the harmonic mean

$$\frac{2}{\dfrac{1}{P(X \ge \mu)} + \dfrac{1}{P(X < \mu)}} \le \frac{P(X \ge \mu) + P(X < \mu)}{2} = \frac{1}{2} \tag{19}$$

we have

$$\frac{1}{P(X \ge \mu)} + \frac{1}{P(X < \mu)} \ge 4. \tag{20}$$

Therefore, for any continuous random variable with parameters $EX = \mu$, $D^2X = \sigma^2$ the following condition holds

$$P(-6\sigma_- \le X - \mu \le 6\sigma_+) \ge P(-3\sigma \le X - \mu \le 3\sigma) = 1 - \frac{1}{9} \approx 0.89. \tag{21}$$

3.2. Examples of application of the 6σ *rule for selected distributions*

a) Let X be a random variable with normal distribution truncated on the interval $[a, b]$, whose density function is given by the formula

$$f(x) = \frac{1}{F(b) - F(a)} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}, \tag{22}$$

where $F(x)$ is the distribution function. Then the following equations are true

$$\sigma_-^2 = \sigma^2 \left[F(m) - F(a) - (m-a) f(a)\right] \quad \text{and} \quad \sigma_+^2 = \sigma^2 \left[F(b) - F(m) - (b-m) f(b)\right]. \tag{23}$$

b) Let X be a random variable with exponential distribution with parameter $\lambda > 0$, whose density function is given by the formula

$$f(x) = \lambda e^{-\lambda x}, \quad x \in (0, +\infty). \tag{24}$$

Then the following equations are true

$$\sigma_-^2 = \frac{1}{\lambda^2} - \frac{2}{\lambda^2} e^{-1}, \quad \sigma_+^2 = \frac{2}{\lambda^2} e^{-1}. \tag{25}$$

c) Let X be a random variable with gamma distribution with parameters $k > 0$ and $\theta > 0$, whose density function is given by the formula

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, x \in (0, +\infty). \tag{26}$$

Then the following equations are true

$$\sigma_-^2 = k\theta^2 F(k\theta) - \frac{\theta^2 k^k}{\Gamma(k)} e^{-k}, \tag{27}$$

$$\sigma_+^2 = k\theta^2 (1 - F(k\theta)) + \frac{\theta^2 k^k}{\Gamma(k)} e^{-k}. \tag{28}$$

The following inequalities:

a) for truncated normal distribution

$$P(X < m) = \int_a^m \frac{1}{F(b) - F(a)} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} = \frac{F(m) - F(a)}{F(b) - F(a)}, \tag{29}$$

$$P(X \geq m) = \frac{F(b) - F(m)}{F(b) - F(a)}, \tag{30}$$

b) for exponential distribution

$$P(X < 1/\lambda) = \int_0^{1/\lambda} \lambda e^{-\lambda x} \mathrm{d}x = -e^{-\lambda x}\Big|_0^{1/\lambda} = 1 - e^{-1}, \tag{31}$$

$$P(X \geq 1/\lambda) = e^{-1}, \tag{32}$$

b) for gamma distribution

$$P(X < k\theta) = \frac{\gamma(k, k\theta^2)}{\Gamma(k)}, \tag{33}$$

will be used in the estimation of

$$P(-6\sigma_- \leq X - \mu \leq 6\sigma_+) \geq 1 - \frac{1}{6^2}\left(\frac{1}{P(X \geq \mu)} + \frac{1}{P(X < \mu)}\right). \tag{34}$$

We obtain from here

a) for truncated normal distribution

$$P(-6\sigma_- \leq X - \mu \leq 6\sigma_+) \geq 1 - \frac{1}{6^2}\left(\frac{F(b) - F(a)}{F(b) - F(m)} - \frac{F(b) - F(a)}{F(m) - F(a)}\right), \tag{35}$$

Metrol. Meas. Syst., Vol. 31 (2024), No. 2, pp. 279–293
DOI: 10.24425/mms.2024.149701

b) for exponential distribution

$$P\left(-6\sigma_{-} \leq X-\mu \leq 6\sigma_{+}\right) \geq 1-\frac{1}{6^2}\left(e+\frac{e}{e-1}\right) \approx 0,88, \tag{36}$$

c) for gamma distribution

$$P\left(-6\sigma_{-} \leq X-\mu \leq 6\sigma_{+}\right) \geq 1-\frac{1}{6^2}\left(\frac{\Gamma(k)}{\gamma\left(k, k\theta^2\right)}+\frac{\Gamma(k)}{1-\gamma\left(k, k\theta^2\right)}\right). \tag{37}$$

4. Verification of confidence intervals based on semivariance for long-term sound level indicators

Long-term noise indicators are used to assess road noise: the day indicator L_D (38)

$$L_D = 10\log\left(\frac{1}{365}\sum_{i=1}^{365} 10^{0,1 L_{Aeq,D_i}}\right), \tag{38}$$

calculated as the logarithmic mean of day A sound levels from all days of the calendar year from 6 a.m. to 6 p.m.; the evening indicator L_E (39)

$$L_E = 10\log\left(\frac{1}{365}\sum_{i=1}^{365} 10^{0,1 L_{Aeq,E_i}}\right), \tag{39}$$

which is the logarithmic mean of evening A sound levels calculated from the whole calendar year from 6 p.m. to 10 p.m., the night indicator L_N (40)

$$L_N = 10\log\left(\frac{1}{365}\sum_{i=1}^{365} 10^{0,1 L_{Aeq,N_i}}\right), \tag{40}$$

calculated from the whole calendar year for the night time from 10 p.m. to 6 a.m., and the day-evening-night indicator calculated from the long-term indicators (41) as a weighted logarithmic mean of these indicators according to the formula [24]:

$$L_{DEN} 10\log\left(\frac{12}{24}10^{0,1 L_D}+\frac{4}{24}10^{0,1(L_E+5)}+\frac{8}{24}10^{0,1(L_N+10)}\right). \tag{41}$$

In measurement practice, for technical and economic reasons, noise indicators are rarely calculated from all days of the calendar year. In this approach [25], the number of measurement days is minimized, leading to noise indicator estimation based on a very small measurement sample. Consequently, such a process requires, in addition to providing the indicator's value, also determining the uncertainty with which that value was determined [1, 26]. This uncertainty is most often given in the form of a 95% confidence interval. The size and coverage of this interval is, obviously, influenced by many different factors, such as uncertainty in the measuring device, uncertainty related to weather conditions, *etc.*, but especially important is the uncertainty related to the measurement sample selection. It is this uncertainty that contributes the most to the uncertainty budget for noise indicators.

One of the assumptions that allow classical confidence intervals to be used to determine uncertainty is to make an assumption about the normality of the average energy of single-day indicators $e_{Aeq,DEN_i} = 10^{0,1L_{Aeq,DEN_i}}$ [27, 28]. This assumption is being made due to the additivity of acoustic energy, as opposed to decibel levels. For normal energy level distribution, one can assign to an energy mean $\overline{e}_{DEN} = 1/n \sum_{i=1}^{n} e_{Aeq,DEN_i}$ a Student's t-distribution and calculate the uncertainty interval for that mean then transform it by a logarithmic transformation and obtain the uncertainty interval for the given indicator

$$\left(10\log\left(\overline{e} - t_{1-\frac{\gamma}{2};n-1}\frac{s_{Aeq,DEN}}{\sqrt{n-1}}\right), 10\log\left(\overline{e} + t_{1-\frac{\gamma}{2};n-1}\frac{s_{Aeq,DEN}}{\sqrt{n-1}}\right)\right), \tag{42}$$

where $s_{Aeq,DEN}$ is the standard deviation of acoustic energy.

For small deviations from the normal distributions of energy distributions, the arithmetic mean of the energy levels can be assigned a normal distribution under the *Central Limit Theorem* (CLT) and the uncertainty propagation method can be applied to this mean [27, 28].

In the paper [28] in which the results of measurements from traffic acoustic monitoring in the city of Krakow were examined, and in the paper [29] where probability distributions from monitoring in the city of Gdansk were examined, it was shown that the parameters of noise indicators distributions (skewness and kurtosis) of noise are far from the normal distribution characteristics. These deviations may be so large that a large measurement sample would have to be taken to apply CLT, which is not economically justified when estimating noise indicators, for example, in large cities. Even larger deviations from normal distributions were observed for energy levels. This analogously affects the inapplicability of CLT for the energy average and, consequently, obtaining an uncertainty interval for the noise indicator with appropriate statistical properties, especially for small measurement samples [30–34]. The deviation of the distribution of the measurement sample mean energy from the normal distribution results in the lack of the required 95% confidence interval coverage. This can be seen in the work [31] where for 26 measurement points in Madrid it was shown that for classical confidence intervals, the coverage of the true value by the confidence interval is much less than the required 95%.

Due to the fact that the method of determining confidence intervals proposed in the article does not assume knowledge of probability distributions and considers the distribution's asymmetry, the proposed construction should better cover the true value of the mean.

To verify the applicability of confidence intervals based on semivariance to determine uncertainty intervals for noise indicators, the following interval form was proposed (43)

$$\left(10\log\left(\overline{e} - t_{1-\frac{\gamma}{2};n-1}\frac{2s_{Aeq,-}}{\sqrt{n-1}}\right), 10\log\left(\overline{e} + t_{1-\frac{\gamma}{2};n-1}\frac{2s_{Aeq,+}}{\sqrt{n-1}}\right)\right), \tag{43}$$

To maintain the proportion, an extension factor was chosen for the interval based on semivariance similar to the classical interval $k = \frac{t_{1-\frac{\gamma}{2};n-1}}{\sqrt{n-1}}$. In addition, from the fact that the variance of the distribution is the sum of the semivariances, which also translates into standard deviations as follows: $s = \sqrt{s_-^2 + s_+^2}$ in place of the standard deviation in formula (42), the two left and right deviations are taken, respectively. It is worth noting that the proposed method (43) does not assume a percentage of true value coverage, unlike the standard interval (42). This interval (43) only assumes a minimum value of coverage depending on the distribution (Section 3.2). The form of the interval (43), however, was chosen to assume $2s_{Aeq,+}$ and $2s_{Aeq,-}$ instead of s_{Aeq}, respectively, due to the fact that the sum of the left and right semivariances is the variance of $s^2 = s_-^2 + s_+^2$ so the assumed values ensure comparability of interval lengths while introducing asymmetry into the uncertainty interval.

Metrol. Meas. Syst., Vol. 31 (2024), No. 2, pp. 279–293
DOI: 10.24425/mms.2024.149701

The following experiment was then proposed to verify the properties of the interval thus constructed and to compare it with the classically used interval (42). The data for the experiment consisted of annual measurements of traffic sound levels in the city of Gdansk from 3 measurement stations. The average values of the noise indicators L_D, L_E, L_N and L_{DEN} were determined from the entire calendar year. These values were taken as population average sound level indicators. From the entire measurement year, 5-, 10-, 20-, 30- and 50-element samples were drawn separately for each station. This draw was repeated $n = 10^6$ times for each sample size. For each drawn sample, two uncertainty intervals were calculated as defined by equations (42) and (43). The coverage of the population mean value by each interval was then checked by counting the percentage coverage for each sample size separately.

4.1. Measurement data

The measurement data comes from continuous monitoring in the city of Gdansk from January 1, 2015 to December 31, 2015. Technical details of the monitoring station and monitoring can be found in the publication [29]. The measurement dataset chosen for the experiment is the nearly complete annual measurements of sound levels. The numbers of measurement days are presented in Table 1. Missing measurements for stations (130) 7 days, (134) 4 days, (141) 1 day were random and did not affect the quality of the measurements. Therefore, for each measuring station, a nearly complete set of indicator values was available to determine L_D, L_E, L_N, and L_{DEN} values for the entire year of 2015.

The coverage of confidence intervals, both classical (42) and based on semivariance (43), was checked for indicators L_D, L_E, L_N, L_{DEN} for measurement data for the following stations:

(130) 41 Pomorska Street – the station is located on a residential building near a two-lane single carriageway road.

(134) 26 Sienna Street – the station is located in a primary school building. Near the station is an intersection of two dual carriageway roads with a streetcar track.

(141) 1 Rybacka Street – the monitoring station is located on a service building near the intersection (a traffic circle) of two single-lane roads.

Basic statistics have been determined for the measurement data from the above stations and are included in the Table 1.

The measurement data were selected to ensure that the indicators represented different values of skewness and kurtosis.

Table 1. Summary statistics for measurements from individual monitoring stations by indicator L_D, L_E, L_N, L_{DEN}.

No. of the Measuring Station	Numbers of Measurement Days	Indicator Type	$\overline{L}$ (Mean) [dB]	$s_{\overline{L}}$ (Standard Deviation) [dB]	$\overline{L}_{\log}$ (Logarithmic Mean of the Sound Level) [dB]	ρ (Skewness) [–]	K (Kurtosis) [–]
130	358	L_D	69.65	1.53	69.91	0.18	-0.23
		L_E	68.49	1.13	68.61	0.88	1.60
		L_N	63.18	1.43	63.34	1.07	1.56
		L_{DEN}	71.88	1.17	72.02	0.88	1.18
134	361	L_D	66.15	2.82	67.08	-0.12	0.95
		L_E	64.69	2.23	65.32	0.30	1.48
		L_N	60.50	2.23	61.23	1.08	3.45
		L_{DEN}	68.88	2.09	69.43	0.37	1.53
141	364	L_D	66.13	2.86	67.47	1.16	3.10
		L_E	65.47	2.99	67.25	1.72	5.31
		L_N	59.87	3.73	62.91	1.76	4.83
		L_{DEN}	68.93	3.02	70.83	1.94	6.15

4.2. Results

5, 10, 20, 30 and 50-element measurement samples were drawn and two intervals were created based on them, a classical interval for the mean value of energy levels (42) and a confidence interval based on semivariance (43). Each sample, respectively, for each indicator L_D, L_E, L_N, L_{DEN} was determined separately so as not to burden the distributions with correlation. The calculations were conducted in the R-Studio software.

Each sampling was repeated one million times, then it was determined what percentage of the confidence intervals cover the true value of the index. Table 2 presents the results obtained.

Table 2. Percentage coverage of confidence intervals for Station no. 130.

Measurement sample size	Classic interval				Interval based on semivariance			
	L_D	L_E	L_N	L_{DEN}	L_D	L_E	L_N	L_{DEN}
5	93.3%	89.7%	87.1%	89.3%	96.6%	94.9%	92.2%	94.7%
10	93.5%	90.8%	88.4%	90.0%	97.0%	95.3%	92.4%	94.8%
20	94.6%	92.3%	90.2%	92.2%	97.8%	95.9%	93.2%	95.7%
30	95.1%	93.3%	92.0%	93.4%	98.2%	96.4%	94.3%	96.1%
50	95.9%	94.7%	93.9%	95.0%	98.6%	97.0%	95.2%	96.9%

Table 3. Percentage coverage of confidence intervals for Station no. 134.

Measurement sample size	Classic interval				Interval based on semivariance			
	L_D	L_E	L_N	L_{DEN}	L_D	L_E	L_N	L_{DEN}
5	87.3%	88.7%	72.2%	85.5%	95.8%	94.1%	82.7%	93.4%
10	89.8%	88.4%	76.4%	87.7%	95.3%	93.6%	86.3%	94.5%
20	88.7%	89.0%	79.6%	88.9%	94.7%	93.6%	84.7%	94.3%
30	89.0%	90.%	83.8%	90.7%	94.7%	94.2%	85.8%	94.7%
50	90.8%	92.2%	87.9%	92.9%	95.0%	94.8%	87.3%	95.4%

Table 4. Percentage coverage of confidence intervals for Station no. 141.

Measurement sample size	Classic interval				Interval based on semivariance			
	L_D	L_E	L_N	L_{DEN}	L_D	L_E	L_N	L_{DEN}
5	73.5%	63.4%	50.8%	55.9%	78.1%	66.4%	52.8%	60.2%
10	74.7%	65.6%	52.8%	59.7%	78.2%	71.9%	55.1%	62.7%
20	77.3%	65.3%	60.0%	64.1%	78.8%	72.7%	64.4%	71.1%
30	80.3%	68.4%	63.7%	68.7%	81.6%	74.7%	66.9%	73.2%
50	83.7%	71.9%	70.8%	75.4%	84.5%	75.8%	72.7%	77.9%

The coverage percentage for the classical interval differs from the coverage value for the semivariance-based confidence interval, as can be observed in Tables 2–4. Each time, an interval based on semivariance gives more coverage than a classically determined interval. For Station no. 130, 5-element measurement samples already give coverage close to 95%, while for the classic range, the coverage is a few percent less, depending on the indicator. It can also be observed that as the skewness and kurtosis values increase, the coverage percentages are much lower than the required 95%. However, it is worth noting that for a confidence interval based on semivariance, this is a few percent more than for a classical interval. In addition, the confidence interval based on semivariance does not have a defined required coverage of the true value, as is the case with the classical confidence interval, which should be 95%.

5. Conclusions

The article presents a generalization of Chebyshev's theorem for semivariance. Inequality variants are presented for specific distributions: truncated normal, exponential and gamma. A generalization of the 3-sigma rule using semivariance is also presented.

It also demonstrates how to create confidence intervals based on Chebyshev's theorem and shows how it can be used in determining uncertainty from long-term noise indicators used in environmental noise protection.

Confidence intervals based on semivariance were also compared with classical confidence intervals based on long-term noise indicators from acoustic monitoring data from the city of Gdansk. Classical confidence intervals require that the probability distribution of the random variable of sound level energy is normal or the measurement sample is large enough to use CLT. Measurement practice shows, however, that such a situation occurs very rarely (*e.g.* for industrial noise). It is shown that a confidence interval based on semivariance gives better coverage values than the classical confidence interval applied to energy means. This is due to the fact that a confidence interval based on semivariance considers the asymmetry of empirical distributions from which the measurement sample is drawn, as opposed to classical intervals. Confidence intervals based on semivariance can therefore be a better alternative for determining measurement uncertainty than classical confidence intervals and the classical Chebyshev's theorem not only for noise indicators but also for other indicators in the natural sciences or engineering, where measurement functions are nonlinear and indicator probability distributions differ significantly from normal distributions. The new uncertainty assessment tool proposed in this paper can form the basis of a new methodology for determining uncertainty when small measurement samples are available and when the distribution of the characteristic's mean value is not normal.

Acknowledgements

The author would like to express his gratitude to the Department of Environment of the city of Gdańsk for providing the measurement data.

Appendices

Theorem Proof (Chebyshev's inequality for semivariance)

From Markov's inequality we obtain

$$P(X-\mu \geq \varepsilon) = \frac{1}{P(X \geq \mu)} P((X-\mu)_+ \geq \varepsilon) \leq \frac{1}{P(X \geq \mu)} \frac{E(X-\mu)_+^2}{\varepsilon^2} = \frac{1}{P(X \geq \mu)} \frac{\sigma_+^2}{\varepsilon^2},$$

$$P(X-\mu \leq -\varepsilon_1) = P(-X+\mu \geq \varepsilon_1) = \frac{1}{P(X<\mu)} P((X-\mu)_- \geq \varepsilon_1) \leq$$

$$\frac{1}{P(X<\mu)} \frac{E(X-\mu)_-^2}{\varepsilon_1^2} = \frac{1}{P(X<\mu)} \frac{\sigma_-^2}{\varepsilon_1^2}.$$

By substituting $\varepsilon := k\sigma_+$, and $\varepsilon_1 := k\sigma_-$, we obtain

$$P(X-\mu \ge k\sigma_+) \le \frac{1}{P(X\ge\mu)}\frac{\sigma_+^2}{(k\sigma_+)^2} = \frac{1}{k^2 P(X\ge\mu)},$$

$$P(X-\mu \le -k\sigma_-) \le \frac{1}{P(X<\mu)}\frac{\sigma_-^2}{(k\sigma_-)^2} = \frac{1}{k^2 P(X<\mu)}.$$

As a result, we obtain

$$P(-k\sigma_- \le X-\mu \le k\sigma_+) \ge 1 - \frac{1}{k^2}\left(\frac{1}{P(X\ge\mu)} + \frac{1}{P(X<\mu)}\right).$$

Proofs of semivariance formulas for individual distributions:

a) normal truncated

$$\sigma_-^2 = \int_a^m (x-m)^2 \frac{1}{F(b)-F(a)}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}\mathrm{d}x =$$

$$= \left| \begin{matrix} f(x) = (x-m) & g'(x) = (x-m)\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \\ f'(x) = 1 & g(x) = \sigma^2\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \end{matrix} \right| =$$

$$= \frac{1}{F(b)-F(a)}\frac{1}{\sqrt{2\pi}\sigma}\left(-(x-m)\sigma^2\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}\right)\Bigg|_a^m +$$

$$+\sigma^2\int_a^m \frac{1}{F(b)-F(a)}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}\mathrm{d}x =$$

$$= (a-m)\,\sigma^2 f(a) + [F(m)-F(a)]\,\sigma^2.$$

Calculations for σ_+^2 are carried out in the same way.

b) exponential distribution

$$\sigma_-^2 = \int_0^{1/\lambda} (x-1/\lambda)^2\,\lambda e^{-\lambda x}\mathrm{d}x = \int_0^{1/\lambda}\left(x^2 - \frac{2x}{\lambda} + 1/\lambda^2\right)\lambda e^{-\lambda x}\mathrm{d}x =$$

$$= \int_0^{1/\lambda} x^2\lambda e^{-\lambda x}\mathrm{d}x - \int_0^{1/\lambda} 2xe^{-\lambda x}\mathrm{d}x + \int_0^{1/\lambda}\frac{1}{\lambda}e^{-\lambda x}\mathrm{d}x = \left| \begin{matrix} f(x) = x^2 & g'(x) = \lambda e^{-\lambda x} \\ f'(x) = 2x & g(x) = -e^{-\lambda x} \end{matrix} \right| =$$

$$= -x^2e^{-\lambda x}\Big|_0^{\frac{1}{\lambda}} + \int_0^{\frac{1}{\lambda}} 2xe^{-\lambda x}\mathrm{d}x - \int_0^{\frac{1}{\lambda}} 2xe^{-\lambda x}\mathrm{d}x + \int_0^{\frac{1}{\lambda}}\frac{1}{\lambda}e^{-\lambda x}\mathrm{d}x = -\frac{1}{\lambda^2 e} - 1/\lambda^2 e^{-\lambda x}\Big|_0^{\frac{1}{\lambda}} =$$

$$= \frac{1}{\lambda^2} - \frac{2}{\lambda^2}e^{-1}.$$

c) gamma distribution

$$\sigma_-^2 = \frac{1}{\Gamma(k)\,\theta^k}\int_0^{k\theta}(x-k\theta)^2\,x^{k-1}e^{-\frac{x}{\theta}}\mathrm{d}x = \frac{1}{\Gamma(k)\,\theta^k}\int_0^{k\theta}(x^2-2xk\theta+k^2\theta^2)x^{k-1}e^{-\frac{x}{\theta}}\mathrm{d}x =$$

$$= \frac{1}{\Gamma(k)\,\theta^k}\int_0^{k\theta}x^{k+1}e^{-\frac{x}{\theta}}\mathrm{d}x - \frac{2k}{\Gamma(k)\,\theta^{k-1}}\int_0^{k\theta}x^k e^{-\frac{x}{\theta}}\mathrm{d}x + \frac{k^2}{\Gamma(k)\,\theta^{k-2}}\int_0^{k\theta}x^{k-1}e^{-\frac{x}{\theta}}\mathrm{d}x =$$

$$= \left|\begin{array}{cc} f(x) = x^{k+1} & g'(x) = e^{-\frac{x}{\theta}} \\ f'(x) = (k+1)\,x^k & g(x) = -\theta e^{-\frac{x}{\theta}} \end{array}\right| =$$

$$= -\frac{1}{\Gamma(k)\,\theta^{k+1}}x^{k+1}e^{-\frac{x}{\theta}}\Big|_0^{k\theta} + \frac{(k+1)}{\Gamma(k)\,\theta^{k-1}}\int_0^{k\theta}x^k e^{-\frac{x}{\theta}}\mathrm{d}x - \frac{2k}{\Gamma(k)\,\theta^{k-1}}\int_0^{k\theta}x^k e^{-\frac{x}{\theta}}\mathrm{d}x +$$

$$+\frac{k^2\theta^2}{\Gamma(k)\,\theta^k}\int_0^{k\theta}x^{k-1}e^{-\frac{x}{\theta}}\mathrm{d}x = -\frac{k^{k+1}}{\Gamma(k)}e^{-k} + \frac{(1-k)}{\Gamma(k)\,\theta^{k-1}}\int_0^{k\theta}x^k e^{-\frac{x}{\theta}}\mathrm{d}x + k^2\theta^2 F(k\theta) =$$

$$= \left|\begin{array}{cc} f(x) = x^k & g'(x) = e^{-\frac{x}{\theta}} \\ f'(x) = kx^{k-1} & g(x) = -\theta e^{-\frac{x}{\theta}} \end{array}\right| =$$

$$= k^2\theta^2 F(k\theta) - \frac{k^{k+1}}{\Gamma(k)}e^{-k} - \frac{(1-k)}{\Gamma(k)\,\theta^{k-2}}x^k e^{-\frac{x}{\theta}}\Big|_0^{k\theta} + \frac{(1-k)}{\Gamma(k)\,\theta^{k-2}}\int_0^{k\theta}kx^{k-1}e^{-\frac{x}{\theta}}\mathrm{d}x =$$

$$= k^2\theta^2 F(k\theta) - \frac{k^{k+1}}{\Gamma(k)}e^{-k} - \frac{(1-k)}{\Gamma(k)\,\theta^{-2}}k^k e^{-k} + \frac{(1-k)\,k}{\theta^{-2}}F(k\theta) =$$

$$= k\theta^2\frac{\gamma(k,k\theta^2)}{\Gamma(k)} - \frac{\theta^2 k^k}{\Gamma(k)}e^{-k},$$

where $F(k\theta) = \frac{\gamma(k,k\theta^2)}{\Gamma(k)}$.

References

[1] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement* (JCGM 100:2008). https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf

[2] Liu, X., Xia, Y., & Rui, X. (2023). Uncertainty evaluation of multilateration-based geometric error measurement considering the repeatability of positioning of the machine tool. *Metrology and Measurement Systems*, 30(1), 49–63. https://doi.org/10.24425/mms.2023.144864

[3] Neska, M. C., & Opara, T. A. (2021). Uncertainty of pressure measurement in a single-bed adsorber. *Metrology and Measurement Systems*, *29*(1), 93–108. https://doi.org/10.24425/mms.2022.138544

[4] Gutiérrez, R., Ramírez, M., Olmeda, E., & Díaz, V. (2015). An Uncertainty Model of Approximating the Analytical Solution to the Real Case in the Field of Stress Prediction. *Metrology and Measurement Systems*, *22*(3), 429–442. https://doi.org/10.1515/mms-2015-0031

[5] Tchebichef, P. L. (1867). Des valeurs moyennes, *Journal de Mathématiques Pures et Appliquées 2ᵉ serie*. *12*, 177–184.

[6] Markov, A. (1884). *On Certain Applications of Algebraic Continued Fractions* [Unpublished doctoral dissertation]. St. Petersburg.

[7] Steliga, K., Szynal, D. (2010). On Markov-Type Inequalities. *International Journal of Pure and Applied Mathematics*, *58*(2), 137–152.

[8] Ferentios, K. (1982). On Tcebycheff's type inequalities. *Trabajos de Estadistica y de Investigacion Operativa*, *33*(1), 125–132. https://doi.org/10.1007/bf02888707

[9] Berge, P. O. (1938). A Note on a Form of Tchebycheff's Theorem for Two Variables. *Biometrika*, *29*(3/4), 405. https://doi.org/10.2307/2332015

[10] Lal, D. N., (1955). A note on a form of Tchebycheff's inequality for two or more variables. *Sankhya*, *15*(3), 317–320.

[11] Isii, K. (1959). On a method for generalizations of Tchebycheff's inequality. *Annals of the Institute of Statistical Mathematics*, *10*(2), 65–88.

[12] Kotz, S., Balakrishnan, N., Johnson, N. L. (2000). *Continuous Multivariate Distributions*, *Volume 1: Models and Applications* (2nd ed.). John Wiley & Sons, Inc.

[13] Birnbaum, Z. W., Raymond, J., & Zuckerman, H. S. (1947). A Generalization of Tshebyshev's Inequality to Two Dimensions. *The Annals of Mathematical Statistics*, *18*(1), 70–79. https://doi.org/10.1214/aoms/1177730493

[14] Olkin, I., & Pratt, J. W. (1958). A Multivariate Tchebycheff Inequality. *The Annals of Mathematical Statistics*, *29*(1), 226–234. https://doi.org/10.1214/aoms/1177706720

[15] Godwin H. J. (1964). *Inequalities on Distribution Functions*. Hafner Publishing Company.

[16] Savage, I. R. (1961). Probability inequalities of the Tchebycheff type. *Journal of Research of the National Bureau of Standards – B, Mathematics and Mathematical Physics B*, *65B*(3), 211–222.

[17] Ion, R. A. (2001). Nonparametric Statistical Process Control. [Doctoral dissertation, University of Amsterdam]. UvA-DARE.

[18] Berck, P., & Hihn, J. M. (1982). Using the Semivariance to Estimate Safety-First Rules. *American Journal of Agricultural Economics*, *64*(2), 298–300. https://doi.org/10.2307/1241139

[19] Boasson, V., Boasson E., Zhou Z., (2011). Portfolio optimization in a mean-semivariance framework, *Investment Management and Financial Innovations*, *8*(3), 58–68.

[20] Nawrocki, D. N. (1999). A Brief History of Downside Risk Measures. *The Journal of Investing*, *8*(3), 9–25. https://doi.org/10.3905/joi.1999.319365

[21] Nantell, T. J., & Price, B. (1979). An Analytical Comparison of Variance and Semivariance Capital Market Theories. *The Journal of Financial and Quantitative Analysis*, *14*(2), 221. https://doi.org/10.2307/2330500

[22] Neave, E. H., Ross, M. N., & Yang, J. (2008). Distinguishing upside potential from downside risk. *Management Research News*, *32*(1), 26–36. https://doi.org/10.1108/01409170910922005

[23] Markowitz, H. (1991). *Portfolio Selection: Efficient Diversification of Investment (2E)*, Malden, MA: Blackwell Publishers Inc.

[24] International Organization for Standardization. (2017). *Acoustics – Description, measurement and assessment of environmental noise. Part 2: Determination of sound pressure levels* (ISO Standard No. 1996-2:2017). https://www.iso.org/standard/59766.html

[25] Gaja, E., Gimenez, A., Sancho, S., & Reig, A. (2003). Sampling techniques for the estimation of the annual equivalent noise level under urban traffic conditions. *Applied Acoustics*, *64*(1), 43–53. https://doi.org/10.1016/s0003-682x(02)00050-6

[26] Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method* (JCGM 101:2008). https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf

[27] Przysucha, B., Pawlik, P., Stępień, B., Surowiec, A. (2021). Impact of the noise indicators components correlation L_d, L_e, L_n on the uncertainty of the long-term day–evening–night noise indicator L_{den}. *Measurement*, *179*, 109399. https://doi.org/10.1016/j.measurement.2021.109399

[28] Makarewicz, R., & Galuszka, M. (2011). Empirical revision of noise mapping. *Applied Acoustics*, *72*(8), 578–581. https://doi.org/10.1016/j.apacoust.2010.10.012

[29] Przysucha, B., Szeląg, A., & Pawlik, P. (2020). Probability distributions of one-day noise indicators in the process of the type A uncertainty evaluation of long-term noise indicators. *Applied Acoustics*, *161*, 107158. https://doi.org/10.1016/j.apacoust.2019.107158

[30] Batko W., Przysucha B., & Tekiel M. (2014, September). *Analysis of asymmetry of probability distributions of the sound level measurement results* (pp. 1–5). Proceedings of the 7th Forum Acusticum, the Polish Acoustical Society, Kraków, Poland.

[31] Barrigón Morillas, J. M., & Prieto Gajardo, C. (2014). Uncertainty evaluation of continuous noise sampling. *Applied Acoustics*, *75*, 27–36. https://doi.org/10.1016/j.apacoust.2013.07.005

[32] Ruggiero, A., Russo, D., & Sommella, P. (2016). Determining environmental noise measurement uncertainty in the context of the Italian legislative framework. *Measurement*, *93*, 74–79. https://doi.org/10.1016/j.measurement.2016.07.007

[33] Peters, R. (2021). *Uncertainty in Acoustics. Measurement, Prediction and Assessment* (1st ed.). CRC Press.

[34] Miao, Y., Yang, X., Zhu, G., Yang, W., Tian, J., & Dai, Y. (2023). Measurement uncertainty analysis of sound intensity using double-coupler calibration system. *Measurement*, *220*, 113315. https://doi.org/10.1016/j.measurement.2023.113315

Bartosz Przysucha received his PhD from the AGH University of Science and Technology in Krakow. He is the acting head of the Department of Quantitative Methods in Management at the Lublin University of Technology and associate dean for education at the Faculty of Management at the Lublin University of Technology. He specializes in measurement uncertainty issues in particular in acoustics and the application of machine learning and artificial intelligence in industrial tomography. Author of more than 30 scientific publications and one monography.