Application problems of process capability evaluation methods in modern quality assurance systems

Sergey Egorov, Alexey Kapitanov and Aleksandra Kozlova*

MSUT «STANKIN», Department of automated information processing and control systems, 127994, Moscow, Vadkovsky Lane 3A, Russian Federation

*a.kozlova@stankin.ru

Abstract. The article discusses the most important issues for industrial quality assurance systems, such as calculation of capability and performance indices. Techniques being currently used at most enterprises, their advantages and disadvantages are assessed. Modern methods for calculating capability indices are proposed and a conclusion is made about the need to streamline the concepts of capability and performance. A conclusion is made about the need to use the proposed technique, which main difference is the ability to evaluate processes by the real distribution law, as well as about the risk of using EXCEL tables offered by many developers of quality assurance systems.

1. Introduction

Most software products for calculating capability and performance indices currently available on the market use outdated algorithms. These algorithms are no longer commonsensical and do not meet either the established world practice, or the accepted Russian (GOST) and international (ISO) norms.

As a rule, all these products are based on the methodology adopted in the 90s of the last century. By that time, the idea of calculating capability indices had already become quite widespread, but the index evaluation and application technique was quite primitive. Nevertheless, a methodology for calculating capability indices was developed and included, among other things, in the Quality System Standard for the Automotive Industry QS9000 that was in effect at that time. For better understanding, let us note the main points of this technique (captured in GOST R 50779.44-2001 and GOST R 51814.3-2001):

- A normal distribution pattern is adopted for all processes;
- On the basis of measurement results, two types of process variation are estimated: inherent variation $\sigma_I$ and total variation $\sigma_T$;
- Inherent variation $\sigma_I$ is calculated based on the Shewhart control charts using empirical coefficients;
- Total variation $\sigma_T$ is calculated as a standard deviation for all measurement results;
- A ratio of the specified tolerance to the inherent process variation (1) is the process capability index $C_P$;

$$\frac{U_{sl} - L_{sl}}{6\sigma_I} (1)$$
• A ratio of the specified tolerance to the total process variation (2) is the process performance index $P_p$;

$$\frac{Usl - Lsl}{6\sigma_T}$$

(2)

• To evaluate process centering, the indices $C_{pk}$ and $P_{pk}$ are used, calculated as minimum values from (3) and (4), respectively for the inherent and total process variation.

$$\frac{Usl - \bar{X}}{3\sigma_I (3\sigma_T)}$$

(3)

$$\frac{\bar{X} - Lsl}{3\sigma_I (3\sigma_T)}$$

(4)

Along with this, the end of the 90s of the last century was marked by the beginning of the widespread application of statistical methods of process management, primarily in the global automotive industry. The understanding was gained that these techniques do not always correspond to the actual state of affairs. In the mid-90s of the last century, Daimler and Ford carried out independent research and found that in a well-established automobile production, as a rule, from 2 to 5% of the processes are distributed according to the normal distribution law. All other processes are subject to other laws. Thus, the basic idea of the normal distribution law was destroyed, on which the existing method of process capabilities indices was based. As far back as in the beginning of 2000, large automobile companies began to introduce their own guidelines regulating the appropriate methods for calculating capability and performance indices, which differed from the method discussed above.

The unifying document was adopted by ISO in 2006 (with alterations in 2007) and became subsequently the Russian standard GOST R ISO 21747-2010 "Statistical Methods - Process Performance and Capability Statistics for Measured Quality Characteristics." With the release of this GOST, the above mentioned standard GOST R 50779.44-2001, which determines the technique discussed above, was canceled.

The purpose of this work is to justify the need for the introduction of modern methods of calculation of performance and reproducibility, considering a time-dependent distribution pattern.

The scientific novelty of the new methodology one are the ability to evaluate processes according to their real distribution law and to calculate the indicators of capabilities and productivity using one formula, together with the determination of their status (capability or productivity) depending on the stability of the process and considering a time-dependent distribution pattern.

The originality of this article is to try to practical justification of the introduction of a new method of evaluation of processes according to their real law of distribution. [1-3,5]

2. Formulation of the problem. Abstract theorems for the calculation of capability indices

Gaining a complete understanding of the changes that occurred in the methodology for determining capability and performance indices with the adoption of GOST R ISO 21747-2010, it is necessary to “recall” several concepts from mathematical statistics. The main of them at this stage is the concept of “quantile”. Imagine that we have a random variable and this variable is distributed according to some distribution law, which is determined by the distribution function showing the probability that the random variable will be less than or equal to the given value. That is, each point in the probability function graph shows the probability that the random variable will be less than or equal to this value. In Figure 1 the probability that the thread length is less than or equal to 14.069 mm is 80%. Thus, 14.069 mm is the 80% quantile of this distribution function [5].

2
If we turn to the normal distribution law, we will recall that there are 99.73% of all random variables distributed according to this law in the range of plus or minus six standard deviations (6σ). A probability density plot (the first derivative of the probability function) is shown in Figure 2.

There are 0.27% of the values outside the bounds of this zone, i.e. 0.135% on the right and on the left. As a consequence of this fact, for the normal distribution law, the corresponding quantiles (μ-3σ) and (μ + 3σ) restrain the range within which the sampling results (values) are expected to be found with a certain probability (confidence interval). The values of these quantiles in percent (this representation is usually called percentiles) are X0.135% and X99.865%. These two values are called lower (Q_{lower3}) and upper (Q_{upper3}) quantiles, respectively.

![Figure 1. Standard normal distribution and the concept of quantile.](image1)

![Figure 2. Probability density plot for the normal distribution law.](image2)

Now let us note an extremely important point for further consideration. The determination of the boundaries of the range of scatter through the number of standard deviations (for example, ± 3σ) refers only to the normal distribution law. But along with this the determination of a similar range through the upper and lower quantiles is possible for any distribution law, since it is possible to determine points with a significance value of 0.135% and 99.865% for any law. The interval constrained by the quantile boundaries of 0.135% and 99.865% is called a reference interval. In fact, the normal distribution law is only a particular case for which the reference interval is 6σ. Compare the representation in Figure 1 (for the normal distribution) and in Figure 3 (for the lognormal distribution law) [7-12].
3. Practical relevance. Modern Calculation Methodology for Calculation Indices

We can consider the first main change in the methodology that distinguishes GOST R ISO 21747-2010 from earlier existing standards. If the previously running standards operated only on the normal distribution law with a range of scatter of $6\sigma$, then the “new” GOST assumes the use of a reference interval instead of this, i.e. the range of scatter constrained by quantile boundaries of 0.135% and 99.865%. According to the new method, capability and performance indices are calculated from the formula (5). We will return to the matter of the same formula for the both indices later on, while we are interested in the calculation technique itself.

$$U_{sl} - L_{sl} \over X_{99.865\%} - X_{0.135\%}$$ (5)

What is the principal difference between the old (1) or (2) and the new (5)? A standard deviation $\sigma$ can be calculated for any set of random variables. But only for a set of random variables distributed according to the normal law, the reference interval is $6\sigma$. Therefore, when trying to calculate capability and performance indices from the above formulas (1) and (2), it was possible to get a sufficiently large calculation error.

A magnitude of this error can be estimated by examining several examples.

As has already been mentioned above, only a small part of processes are distributed according to the normal law in real production. Distribution laws of some other processes “are more similar” to the normal law, but there are a lot of processes for which distribution according to the normal law is impossible in theory. These are unilateral physical processes. This includes characteristics, such as defects of form and interrelation accuracy (roundness, flatness, beat, etc.), or surface texture indicators (for example, roughness). The values of these characteristics fundamentally cannot be negative. This is where the risk of the greatest mistake in calculation lies.

A big mistake lies in the essence of the normal law while trying to evaluate characteristics with a lateral physical boundary – in its symmetry. Having a certain set of data, all values wherein are in the positive region, the normal distribution law assumes that values with a certain probability may appear in the negative region. And since they are not there, the process capability is very good, as far as the process of scattering is currently very small. Let us consider this situation with an example.
Figure 4. Capability evaluation using the old technique.

There were 200 values of the flatness measurements obtained (the value graph is shown in Figure 4 at the top left). After the classification, a histogram was constructed (Figure 4, top right). Let us assess the capability using an old method, evaluating the process as normal. The result of approximating the curve of the normal law is shown in Figure 4, bottom left.

With this calculation, we obtain the value of the capability index $C_{pk} = 4.15$ (Figure 4, bottom right). Note that for characteristics with the unilateral physical boundary, modern software calculates the $C_p$ index conditionally, since the process centering will worsen this indicator (the closer the process to the upper boundary, the worse the indicators, in this case for the flatness error). Therefore, the $C_p$ index is displayed in gray. So, for the normal law, the index is 4.15. But what is in reality? Modern software products, designed for process analysis, automatically select the most appropriate distribution law. In this example, qs-STAT software of the German company Q-DAS was used, which is applied for process evaluation by nearly 200,000 users worldwide.

There are several distribution laws that describe the behavior of unilateral physical characteristics. These include the stowed normal distribution, Rayleigh distribution, lognormal distribution, and Weibull distribution. The verification of the conformity of the law of distribution is made by statistical criteria. In this case, the lognormal distribution was automatically selected (Figure 5 on the left). And for this distribution law, the calculation of the capability index $C_p$ showed a completely different value – 1.89 (Figure 5 on the right), which was more than two times lower than the one calculated according to the normal law.
For a final understanding of why this happened, let us draw your attention once again to Figs. 4 (bottom, left) and 5 (left). They clearly show the difference in the estimate of the interval $6\sigma$ and the reference intervals of 0.135% and 99.865%. The calculation shows that the interval $6\sigma$ is 2.164, while the reference interval is 2.798. With the same tolerance limits, this gives the difference in the values of the capability index. To finally dispel all doubts, let us align with the help of software two distribution laws on one histogram (Figure 6). Now it is clear that the normal distribution law is much less suitable for the given data set.

In this example, everything ended well – both distribution laws showed the process capability. However, this does not happen very often. The risk is that if a customer is guided and uses statistical managing methods in his practice, he will immediately see an incorrect choice of the distribution law. Along with this, the lack of process capability will inevitably lead to the appearance of scrapped parts.

Unilateral tolerance characteristics best demonstrate the difference when the capacity law is incorrectly chosen. But the error occurs also in the evaluation of “common” features. And this is not always towards the overestimation of the features.

In the case of automatic selection of the distribution law, the software selects a mixed distribution law (which in this case is quite natural). And the calculation of capability indices through the reference interval gives completely different results.

But, what is more important, with the correct process evaluation (from the selection of the most appropriate distribution law), you can avoid one of the most unpleasant situations in process
management – an attempt to achieve capability for the process that already has good capability indicators. Such an attempt can lead to complete process inoperability.

All mentioned above about the calculation procedure for indices $C_p$ from the reference interval equally applies to the calculation of indices $C_{pk}$, considering the process bias. In the old method, the upper and lower indices were calculated as (6) and (7), respectively.

$$U_{sl} - \bar{X}$$

$$\frac{3 \sigma_i (3 \sigma_T)}{3 \sigma_i (3 \sigma_T)}$$

$$\bar{X} - L_{sl}$$

$$\frac{X_{50\%} - L_{sl}}{X_{50\%} - X_{99.865\%}}$$

In the new method, everything is calculated through quantiles. The average value changes to a 50% quantile (this is the point of the probability function for which it is equally probable that the value of a random variable will be greater or less than the current value), and $3\sigma$ changes by half of the reference interval, i.e. the difference between the 50% quantile and the boundary quantiles. [6]

Then a smaller process index is calculated as the smallest value from the upper index (8) and the lower index (9).

$$U_{sl} - X_{50\%}$$

$$X_{99.865\%} - X_{50\%}$$

$$X_{50\%} - L_{sl}$$

$$X_{50\%} - X_{0.135\%}$$

It is extremely difficult to validate the correct selection of the distribution law by means of an Excel spreadsheet. In any case, this kind of product is unlikely to be free of charge.

4. Modern Methodology – Streamlining the Concepts of Capability and Performance

Along with the considered problem of the processes distributed according to laws different from the normal law, a methodology introduced in the middle of the 2000s solved another problem of the old methodology, which raised certain doubts. It is referred to the calculation of capability and performance indices from different formulas, based on different concepts of variability.

An example of a report on process capability indices (Table 1).

| capability index of the process   | $C_p = 1.68$ |
|----------------------------------|--------------|
| smaller capability index of the process | $C_{pk}=1.47$ |
| calculation method              | $M_{1,1}$    |
| number of values used for the calculation | 2000 |
| measurement uncertainty         | 0.002 mm     |
| model change of the distribution based on time | $A_{1}$ |

The old methodology determined one of the basic principles of the process analysis – it is necessary to clearly distinguish whether the process is in a state of statistical controllability or not (in other words, whether the process is stable in terms of statistics). As a solution, the old methodology suggested calculating two indices. One of them, the capability index, $C_p$, was calculated on the basis of the inherent process variation, assuming that the inherent variation depends only on common causes. The process, which is affected only by common causes, is statistically stable. Thus, the capability index $C_p$
showed the process capabilities in a stable state [4, 13]. The second index, the performance index, Pp, was calculated on the basis of the total process variation, which also takes into account the effect of random causes. If the process is affected by accidental causes, its stability cannot be confirmed. And the performance index Pp showed the process capabilities in a state when the stability could not be confirmed.

The new methodology has approached this problem from a slightly different perspective. It was determined in the “new” standards that the process cannot be repeatable being unstable. The process capability index in this case is calculated by the same formula — the ratio of the specified tolerance to the reference interval (10).

\[
\frac{Usl - Lsl}{X_{0.025} - X_{0.05}}
\]

But this index is denoted and called differently, depending on whether the process is in a stable or unstable state. For a process with confirmed stability, we obtain the process capability index Cp, and for a process which stability is not confirmed — the performance index Pp.

The process stability is confirmed by a control chart. If the chart does not contain the disturbance of the specified stability criteria, the process is stable. Thus, Pp and Cp are calculated from the same formula. If the plotted control chart has no stability disturbances, the index is denoted as Cp. If there were disturbances, the index is denoted as Pp.

All mentioned above about the indices Cp and Pp equally applies to the indices Cpk and Ppk, considering the process bias.

5. Conclusion
Free Excel spreadsheets being widely used in the industry for calculating capability and performance indices are based on the calculation techniques adopted in the mid-90s of the last century. At the end of the 2000s, a new calculation technique for capability and performance indices was normatively established. The main differences of the new methodology from the old one are the ability to evaluate processes according to their real distribution law (and not only to the normal law) and the calculation of capability and performance indices from one formula together with determination of their status (capability or performance) depending on the process stability.

In fact, in the new, existing methodology, there is one more component that always participates in the process analysis — a time-dependent distribution pattern.

Recently, the 22514 series standards have taken effect — “Statistical methods. Process management”. The standard GOST R ISO 22514-2-2015 “Statistical methods. Process management. Part 2. Evaluation of process capability and performance based on its time-dependent pattern”, which took effect in July 2016 and superseded the previously reviewed GOST R ISO 21747-2010. But all the techniques considered and proposed in this paper remain unchanged.

Acknowledgments
The research is performed within the framework of the grant Of the President of the Russian Federation for the state support of young scientists (MD-6629.2018.9).

References
[1] Egorov S., Kapitanov A., Loktev D., Implementation of statistical process control methods as a way to reduce production costs and improve product quality, MATEC Web of Conferences 129, 2017, ICMTMTE, Sevastopol, Russia.
[2] Egorov S.B., Kapitanov A.V., Loktev D.A., Egorova T.P., Modern methods of performance assessment processes (organizing concepts of reproducibility and suitability), International journal of experimental education, 2016. № 10-1. p. 127-129.
[3] Loktev D. A., Egorov S. B., Kapitanov A.V., Mitrofanov V. G. and Egorova T. P., Measuring Instruments for Shafts, Russian Engineering Research, 2016, Vol. 36, No. 4, pp. 312–317.

[4] Storto A., Variational quality control of hydrographic profile data with non-gaussian errors for global ocean variational data assimilation systems, Ocean Modelling. 2016. Т. 104 226-241.

[5] Egorov S.B., Kapitanov A.V., Egorova T.P., Mitrofanov V.G., Development of elements of system of quality management of engineering production on the basis of innovative Metrology equipment and multilevel software for statistical process control (in Russ.), International journal of experimental education. 2015. № 12-1. p. 128-129.

[6] Hawkins D. M., Deng Q., A Nonparametric Change-Point Control Chart. - Journal of Quality Technology, 2010, v.42, No.2, pp.165-173.

[7] Box G., Narasimhan S., Rethinking Statistics for Quality Control. - Quality Engineering, 2010, vol. 22, p.60-72.

[8] Kudrya A.V., Sokolovskaya E.A., Sukhova V.G., Skorodumov S.V., Limitations of classical statistics in quality control and certification of structural steel, Steel in Translation. 2010. T. 40. № 11 961-965.

[9] Balestracci D., Data Sanity: A Quantum Leap to Unprecedented Results. – Medical Group Management Association. – 2009, 304 P.

[10] Chakraborti S., Human S. W., Graham M. A. (2009) Phase I Statistical Process Control Charts: An Overview and Some Results. – Quality Engineering, 2009, v.21, pp.52-62.

[11] Hahn G., Doganaksoy N. The role of statistics in business and industry. – John Wiley & Sons, Inc. – 2008, 344 P.

[12] Jahnen A., Schilz C., Shannoun F., Schreiner A., Hermen J., Moll C., Optimage central organised image quality control including statistics and reporting, Radiation Protection Dosimetry. 2008. T. 129. № 1-3 p253.

[13] Laney D., Improved Control Charts for Attributes. – Quality Engineering, 2002, v.14(4), P.531-537.