SELECTED NONPARAMETRIC METHODS OF STATISTICAL PROCESS CONTROL

SMAJDOROVÁ Tereza, NOSKIEVIČOVÁ Daria

VSB - Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Ostrava, Czech Republic, EU, tsmajdorova@seznam.cz , darja.noskievicova@vsb.cz

Abstract

This paper presents some possibilities of statistical process control that can be used when the basic requirements for the application of standard Shewhart control charts are not fulfilled. These basic assumptions that must be met include mainly a requirement on the normality of the data, the requirement for constant mean and variance, and last but not least the requirement for mutual independence of data. In practice, such as the metallurgical industry, those assumptions about the data are not necessarily always met. The stress in the article is put on the importance of the statistical process control as a part of the production management necessary for achieving a high product quality. Just the quality of the product decides on customer satisfaction and thus the success of the whole organization. The aim of this article is to describe some non-parametric control charts and concretely introduce one of the non-parametric control charts, namely Shewhart sign control chart, including a practical example from a metallurgical process. During preparation of this article accessible pieces of knowledge on the issue were compared. During this comparison of parametric and nonparametric methods it was found that nonparametric methods have many advantages and for cases where some of the basic assumptions about the data are not met they are appropriate. This article is the basis for further investigation of the problem, including a more detailed treatment of the particular non-parametric control charts.

Keywords: Statistical process control, nonparametric methods, production management, Shewhart sign control chart

1. INTRODUCTION

Statistical process control (SPC) is an immediate and continuous process control based on the mathematical-statistical evaluation of the product quality. If a company wants to achieve the high quality consistently, it has to collect, process and analyze systematically data available from the production and conclusions of the analysis must be used for continuous improvement. To use the classic Shewhart control charts, the certain basic assumptions about the data must be met. In the manufacturing practice, however, it is not always possible to meet these basic assumptions. This paper aims to answer the question how to control the production process if the obtained data do not meet these requirements.

2. NONPARAMETRIC METHODS IN SPC

Statistical process control allows interventions in the process based on the early detection of deviations from a predetermined process level. It is implemented by regular monitoring of the controlled process variable or output variable. It is found out whether it corresponds to the process level required by the customer. Achieving the desired level of the process requires a thorough analysis of the process variability. [13]

Most publications about the SPC deal with the processes that meet the basic requirements about the data needed for the use of the classic Shewhart charts. These assumptions include:

- compliant capability of the measurement system
- normal distribution of the quality characteristic,
- constant mean and variance,
- mutual independence of quality characteristic values.

- a sufficient number of data,
- sensitivity to greater changes in process,
- monitoring one quality characteristic per unit of product. [1, 14]

In non-compliance with some assumptions, it is possible to apply nonparametric methods. Nonparametric methods are based on a smaller number of observations. Compared to the model based methods most often it is only assumed that the probability distribution of the given data set is of the continuous type. [16]

Nonparametric methods have, compared to parametric methods, a number of advantages:

- conclusions obtained are independent of the distribution shape,
- they can be used even when the type of distribution is unknown,
- they are used in cases where sample size is too small,
- they can be used for ordinal (serial) variables, some also for nominal (verbal) variables,
- for small sample size the calculation is relatively simple,
- they have a greater robustness to the occurrence of outliers.

Disadvantages of non-parametric methods include the increased probability of missing signal, which means that it often leads to incorrect non-rejection of untrue null hypothesis. This probability can be reduced by increasing the sample size. [9, 15]

Nonparametric statistical process control (NSPC) is based on methods that are not dependent on a specific type of the probability distribution. The use of these control charts is suitable not only for processes that do not meet normality and independence of the data, but especially at the beginning of the SPC implementation, when there are not enough data available. [9]

2.1. Nonparametric Control Charts

Below there are some nonparametric methods that can be used if the basic assumptions, such as data normality, mutual independence or constant mean and variance are not met.

- Shewhart Sign Control Chart
  It is one of the simplest non-parametric control charts. It is based on simple statistics that tracks the difference between the number of observations above and below a predetermined target value. [2, 9]
- EWMA-DFCC (Exponentially-Weighted Moving Average - Distribution Free Control Chart)
  It is a nonparametric control chart of exponentially-weighted moving averages. It combines the properties of the classical EWMA chart with the robustness of nonparametric charts. Hackl and Ledolter [7] considered the use of nonparametric control chart for individual observations using a standardized series of the observations. The simulation studies showed that the method is resistant to the outliers and works well even with sudden changes in the process. [1, 5, 9]
- CUSUM-DFCC (Cumulative Sum - Distribution Free Control Chart)
  CUSUM charts are suitable in the case of sequential nature of the process control. One of the problems of the Shewhart sign control chart is the fact that the target value must be known. Application of the nonparametric CUSUM control chart is one of the ways to avoid this problem. This technique was originally developed by McGilchrist and Woodyer [12] for monitoring the amount of rainfall. [1, 5, 9]
- Pre-control
  Ledolter and Swersey [11] considered the possibility of using the pre-control as one of the alternatives to the conventional control charts. When comparing the standard control charts to the pre-control, they found out that the pre-control is of some importance, especially in machining. However, in general the pre-control is not an adequate substitute for the control charts. [9]
- Change-point chart

This is a problem, when the change of distribution type occurs in the number of independent random variables after the first observation. It is an issue dealt by Bhattacharya and Frierson [3]. The aim is to detect an unknown change-point without a large number of false signals and without having to know any assumptions about the type of probability distribution data. The nonparametric control chart was designed based on the weighted sum of values in a row and on the asymptotic behavior of the cumulative sums, assuming that there are small changes in distribution after a large number of observations. [9]

- Bootstrap Method
  It is a relatively young method. It was not practically possible to use this method before the advent of computers [8]. It is a resampling of the original data set. From the original data there are generated bootstrap samples from which there are calculated samples characteristics $\hat{\theta}$, it is repeated k-times. The values $\hat{\theta}$ obtained from bootstrap samples are used to calculate the characteristics θ of the original data set. [10]

3. SHEWHART SIGN CONTROL CHART

Let $X_{i1}, X_{i2}, \ldots, X_{in}$ (*I* = 1, 2,…) denote (*i* = 1,2,...) sample or subgroup of independent observations of size *n* > 1 from a process with an unknown continuous distribution function F. Let θ_0 denote the known or specified target value. Let's compare each x_{ij} (*j* = 1, 2,…,*n*) with θ_0. Let's record the difference between θ_0 and each x_{ij} (x_{ij} - θ_0). There will be *n* of such differences, for the i^{th} sample. Let n^+ denote the number of observations with values greater than θ_0 in the i^{th} sample. Let n^- denote the number of observations with values less than θ_0 in the i^{th} sample. Sum $n^+ + n^- = n$.

Shewhart sign statistic is defined as

$$SN_i = \sum_{j=1}^{n} sign(x_{ij} - \theta_0) \tag{1}$$

where $sign(x_{ij} - \theta_0)$ = -1, 0 or +1, if $(x_{ij} - \theta_0)$ < 0, = 0 or > 0.

Then SN_i is the difference between n^+ and n^- in the i^{th} sample, i.e. SN_i is the difference between the number of observations with values greater than θ_0 and the number of observations with values less than θ_0 in the i^{th} sample. The control limits and the center line of the two-sided nonparametric Shewhart-type sign chart (for the median) are given by UCL = *c*, CL = 0 a LCL = -*c*.

$$c = 2 \cdot t - n \tag{2}$$

where $c \in \{1,2,...,n\}$ and *n* is subgroup size. Value *t* can be obtained from table for binomial distribution for θ =0.5 from Table G see Gibbons and Chakraborti Nonparametric statistical inference [4].

When the control chart includes all points between the control limits, the process is in-control. If any point is located on one of the control limit, if it is below the lower control limit or above the upper control limit, it means that the process is out-of-control. In this case, we have to find the cause and implement corrective measures. [6]

3.1. EXAMPLE

The following example illustrates the application of Sign Shewhart Control Chart on the data obtained from the steelmaking process (**Table 1**). There was measured carbon content in the steel each day. The measured values are recorded in the table, in the columns x_1 to x_5. The target value θ_0 = 1.29%. With this value we will compare the data in the thirty-one subgroups each of five units. For each value we compute the difference between the measured and the target value $sign(x_{ij} - \theta_0)$ and we write it into the table -1, 0, or +1 depending on whether the released $sign(x_{ij} - \theta_0)$ < 0, = 0 or > 0.

Table 1 Data

Date	x_1	x_2	x_3	x_4	x_5	SN_i	Date	x_1	x_2	x_3	x_4	x_5	SN_i
1.1.2015	1.272	1.767	1.601	1.348	1.534	3	17.1.2015	1.281	1.214	1.391	1.327	1.051	-1
	-1	1	1	1	1			-1	-1	1	1	-1	
2.1.2015	1.454	1.324	1.435	1.193	1.237	1	18.1.2015	1.106	1.407	1.508	1.558	1.173	1
	1	1	1	-1	-1			-1	1	1	1	-1	
3.1.2015	1.509	1.886	1.546	0.882	1.787	3	19.1.2015	1.295	0.99	1.636	1.383	1.507	3
	1	1	1	-1	1			1	-1	1	1	1	
4.1.2015	1.038	1.349	0.898	1.393	1.671	1	20.1.2015	0.912	1.271	1.633	1.378	1.483	1
	-1	1	-1	1	1			-1	-1	1	1	1	
5.1.2015	1.094	1.232	1.612	1.8	1.376	-1	21.1.2015	1.695	1.044	1.65	1.536	1.46	3
	-1	-1	1	-1	1			1	-1	1	1	1	
6.1.2015	1.361	1.53	1.083	1.167	1.325	1	22.1.2015	1.317	1.185	1.218	1.778	1.613	1
	1	1	-1	-1	1			1	-1	-1	1	1	
7.1.2015	1.024	1.542	1.313	1.371	1.072	1	23.1.2015	1.507	1.244	1.158	0.941	0.855	-3
	-1	1	1	1	-1			1	-1	-1	-1	-1	
8.1.2015	0.933	1.394	1.13	1.335	1.376	1	24.1.2015	0.716	1.029	1.248	1.046	0.757	-5
	-1	1	-1	1	1			-1	-1	-1	-1	-1	
9.1.2015	1.168	1.121	1.264	0.95	0.903	-5	25.1.2015	1.206	1.145	0.772	1.109	1.137	-5
	-1	-1	-1	-1	-1			-1	-1	-1	-1	-1	
10.1.2015	1.182	1.062	1.016	1.345	1.064	-3	26.1.2015	1.213	1.549	1.463	1.468	1.713	3
	-1	-1	-1	1	-1			-1	1	1	1	1	
11.1.2015	1.528	1.571	1.501	1.242	1.217	1	27.1.2015	1.276	1.223	1.32	1.199	1.321	-1
	1	1	1	-1	-1			-1	-1	1	-1	1	
12.1.2015	1.077	0.826	1.062	1.283	1.396	-3	28.1.2015	1.419	0.968	1.032	1.107	1.066	-3
	-1	-1	-1	-1	1			1	-1	-1	-1	-1	
13.1.2015	1.169	1.289	1.476	1.745	1.479	1	29.1.2015	0.982	0.706	1.174	1.044	1.137	-4
	-1	-1	1	1	1			-1	-1	-1	-1	-1	
14.1.2015	1.373	1.194	1.001	1.082	0.909	-3	30.1.2015	1.067	1.309	1.046	1.031	0.989	-2
	1	-1	-1	-1	-1			-1	1	-1	-1	-1	
15.1.2015	1.214	1.198	1.413	1.49	1.137	-1	31.1.2015	1.169	0.909	1.637	1.038	1.4	-2
	-1	-1	1	1	-1			-1	-1	1	-1	-1	
16.1.2015	1.281	0.986	1.509	1.579	1.388	1							
	-1	-1	1	1	1								

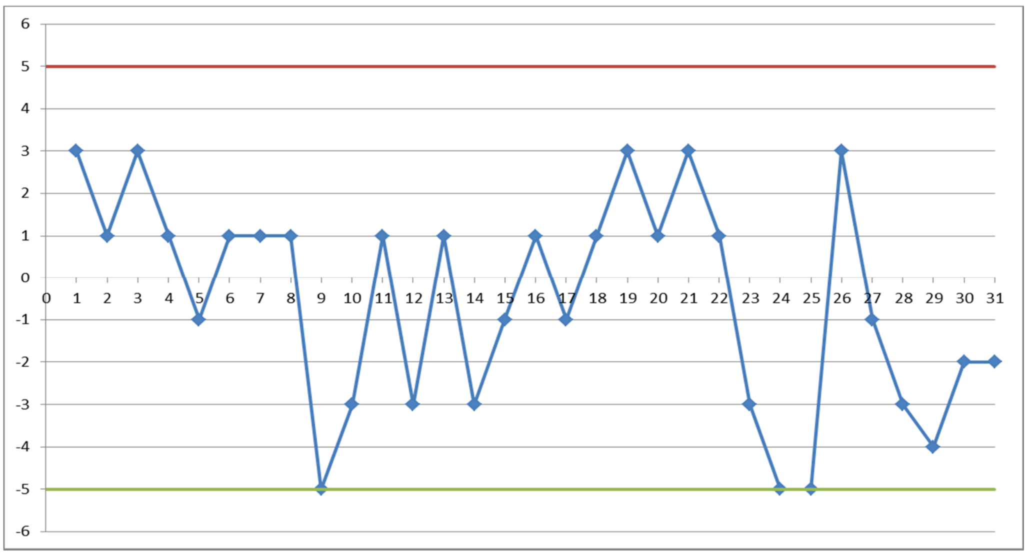

Figure 1 Shewhart sign control chart

Subsequently, for each subgroup we calculate the value SN_i according to the formula (1). Now we can proceed to the calculation of control limits, according to the formula (2) and construct the control chart (**Figure 1**).

Subgroup size is n = 5 and from table [4] t = 5, so the value $c = 2 \cdot 5 - 5 = 5$. It follows that UCL = 5, CL = 0 and LCL = -5.

The chart shows that in the ninth, twenty-fourth and twenty-fifth subgroup point lies on the lower control limit, which may mean exposure to assignable causes of variability that should be analyzed and subsequently eliminated. In the conventional Shewhart control chart for average there did not appear the assignable cause in the process.

CONCLUSION

The aim of the next work is a detailed look at how to control the production process, which violates one of the basic assumptions about the data and creating a methodology for production process control that do not meet these assumptions. The results will contribute to the development of statistical process control and process capability analysis. The proposed methodology could help in the decision-making processes in practice.

ACKNOWLEDGEMENTS

The work was supported by the specific university research of the Ministry of Education, Youth and Sports of the Czech Republic No. SP2016/107.

REFERENCES

[1] BAKIR, S. T. C*lassification of distribution-free quality control charts*. 2001. Dostupné z: http://www.amstat.org/sections/SRMS/Proceedings/y2001/Proceed/00088.pdf

[2] BAKIR, S. T., PRATER, T., KISER, S. *A Simple Nonparametric Quality Control Chart For Monitoring Students' GPAs*. 2015. available from: http://www.researchgate.net/publication/273984642_A_Simple_Nonparametric_Quality_Control_Chart_For_Monitoring_Students%27_GPAs

[3] BHATTACHARYA, P. K.; FRIERSON, D. *A nonparametric control chart for detecting small disorders*. The Annals of Statistics, 1981, 544-554.

[4] GIBSON, J. D., CHAKRABORTI, S. *Nonparametric statistical inference*. 4th ed., rev. and expanded. New York: Marcel Dekker, c2003. ISBN 0824740521.

[5] GRAHAM, M. S., HUMAN, S. W., CHAKRABORTI, S. *A nonparametric exponentially weighted moving average signed-rank chart for monitoring location*. available from: http://repository.up.ac.za/bitstream/handle/2263/19376/Graham_Nonparametric(2011).pdf?sequence=1

[6] GRAHAM, M. *Theory and applications of univariate distribution-free Shewhart, CUSUM and EWMA control charts*. Pretoria, 2008. Disertační práce. University of Pretoria, Faculty of Natural & Agricultural Sciences, Mathematical Statistics. Vedoucí práce C.F. Smit.

[7] HACKI, P., LEDOLTER, J. *A new nonparametric quality control technique. Communications in Statistics-Simulation and Computation*, 1992, 21.2: 423-443.

[8] HALL, P. *The bootstrap and Edgeworth expansion*. New York: Springer-Verlag, 1992, xiii, 352 s. Springer series in statistics. ISBN 0387945083.

[9] CHAKRABORTI, S., VAN DER LAAN, P., BAKIR, S. T. *Nonparametric Control Charts: An Overview And Some Results*. 2001. available from: http://207.67.83.164/pub/jqt/past/vol33_issue3/qtec_33_3_304.html

[10] CHERNICK, M. R. *Bootstrap methods: a guide for practitioners and researchers*. 2nd ed. Hoboken, N.J.: Wiley-Interscience, c2008. ISBN 0471756210.

[11] LEDOLTER, J., SWERSEY, A. *An evaluation of pre-control.Journal of quality technology*, 1997, 29.2: 163.

[12] MCGILCHRIST, C. A.; WOODYER, K. D. *Note on a distribution-free CUSUM technique*. Technometrics, 1975, 17.3: 321-325.

[13] NENADÁL, J. *Moderní management jakosti: principy, postupy, metody*. Vyd. 1. Praha: Management Press, 2008, 377 s. ISBN 978-80-7261-186-7.

[14] NOSKIEVIČOVÁ, D. *Pokročilejší metody statistické regulace procesu*. První vydání. Praha: Grada Publishing, 2015. Expert (Grada). ISBN 978-80-247-5355-3.

[15] ŠTIGLIC, M. *Neparametrické štatistické metódy a ich ekonomické aplikácie*. Košice, 2009. available from: http://web.ics.upjs.sk/svoc2009/prace/3/Stiglic.pdf. Študentská vedecká konferencia. Slovenská technická univerzita v Bratislave, Stavebná fakulta. Vedoucí práce Prof. RNDr. Magda Komorníková CSc.

[16] ZVÁROVÁ, J. *Základy statistiky pro biomedicínské obory*. 2., dopl. vyd. Praha: Karolinum, 2011, 219 s. Biomedicínská statistika. ISBN 978-80-246-1931-6.