A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 286, 2013

*Grzegorz Kończak**

ON THE USE OF THE EXTREME VALUE DISTRIBUTION IN MONITORING PRODUCTION PROCESSES

Abstract. In many statistical applications the main point of interest is estimating some population central characteristic such as the mean or the median. The Shewhart's control chart $\overline{X}$ is based on monitoring the average process level. However, in some areas the main interest is based on estimating the maximum or the minimum. The proposal of the monitoring processes based on the properties of the Gumbel distribution is presented in the paper. The properties of the proposed method have been analyzed in the Monte Carlo study.

Key words: process control, extreme value distribution, Gumbel distribution, Monte Carlo.

I. INTRODUCTION

The control chart is a useful tool in the statistical process control (SPC). It is a graphical display of a quality characteristic which has been measured or computed (for example a sample mean or a standard deviation) from a sample (D.C. Montgomery, 1996). The concept of control charts was developed in 1924 by Walter A. Shewhart. The main assumptions in a classical control chart for variables are the normality of the distribution and the independence of samples.

In many situations we may have reason to doubt the validity of the normality assumption. There are some methods for monitoring non-normal processes (P.K. Bhattacharya., D. Frierson, 1981, G. Kończak, 2010). This article presents the proposal of monitoring non-normal processes based on the properties of the extreme value distribution. The proposed control chart could be used even if the distribution is not normal.

II. EXTREME VALUE DISTRIBUTION

In many statistical applications the main problem is estimating some population central characteristic. The process average could be monitored by the Shewhart control chart. However, in some statistics areas the main interest

* Ph.D., Associate Professor, Department of Statistics, Katowice University of Economics, grzegorz.konczak@ue.katowice.pl. The work paper funded by resource on science period 2012-2015, National Science Centre Nr DEC-2011/03/B/HS4/05630.

is focused on estimating the extremes. Often is it important that the weight or the height of the produced elements should not be too high or too small. In these cases the properties of the extreme values distribution could be used in monitoring processes.

2.1. Exact distribution of extreme values

Let $(X_1, X_2, \ldots, X_n)$ be an *n* element independent random sample taken form the continuous distribution $F(x)$. Let X_M and X_m be the random variables given by

$$X_M = \max(X_1, X_2, \ldots, X_n) \tag{1}$$

and

$$X_m = \min(X_1, X_2, \ldots, X_n) \tag{2}$$

The forms of distributions $G_M^{(n)}(x)$ and $G_m^{(n)}(x)$ of these variables can be found in the following way

$$\begin{aligned} G_M^{(n)}(x) &= P(X_M < x) = P(X_1 < x \wedge X_2 < x \wedge \ldots \wedge X_n < x) = \\ &= P(X_1 < x) \cdot P(X_2 < x) \cdot \ldots \cdot P(X_n < x) = \left(F(x)\right)^n \end{aligned} \tag{3}$$

and

$$\begin{aligned} G_m^{(n)}(x) &= P(X_m < x) = P(X_1 < x \vee X_2 < x \vee \ldots \vee X_n < x) = \\ &= 1 - P(X_1 \geq x) \cdot P(X_2 \geq x) \cdot \ldots \cdot P(X_n \geq x) = 1 - \left[1 - F(x)\right]^n \end{aligned} \tag{4}$$

It could be noticed that the limit distributions of $G_M^{(n)}(x)$ and $G_m^{(n)}(x)$ have the following forms:

$$\lim_{n\to\infty} G_M^{(n)}(x) = \lim_{n\to\infty} F^n(x) = \begin{cases} 1 & \text{if } F(x) = 1 \\ 0 & \text{if } F(x) < 1 \end{cases} \tag{5}$$

$$\lim_{n\to\infty} G_m^{(n)}(x) = \lim_{n\to\infty}\left[1 - \left(1 - F^n(x)\right)\right] = \begin{cases} 1 & \text{if } F(x) = 1 \\ 0 & \text{if } F(x) < 1 \end{cases} \tag{6}$$

The limit distributions take value 0 and 1 only so it could be said they are degenerate. To avoid degeneracy, the linear transformations of the argument x are used

$$\lim_{n\to\infty} G_M^{(n)}(a_n + b_n x) = \lim_{n\to\infty}\left[F(a_n + b_n x)\right]^n = G_M(x) \tag{7}$$

$$\lim_{n\to\infty} G_m^{(n)}(c_n + d_n x) = \lim_{n\to\infty} 1 - \left[1 - F(c_n + d_n x)\right]^n = G_m(x) \tag{8}$$

where a_n, b_n, c_n and d_n are the constants depending on n.

A $F(x)$ distribution, is said to belong to the maximal domain of attraction of $G_M(x)$, if at least one pair of sequences $\{a_n\}$ and $\{b_n > 0\}$ that fulfill the formula (7) exist. Similarly, if $F(x)$ holds (8), we say that it belongs to the minimal domain of attraction of $G_m(x)$. The limit distribution of the maxima or the minima can only have one of the forms: the Gumbel distribution, the Frechet distribution or the Weibull distribution (Castillo E. et all., 2005). Table 1 presents the domains of attraction for the maximal values of some commonly used distributions.

Table 1. Domains of attraction for maximal values of some distributions

Distrbution	Domain of attraction
Normal Lognormal Gamma Exponentional	Gumbel
Cauchy Pareto	Frechet
Uniform	Weibull

Source: based on Castillo et all (2005).

Table 2 presents the domains of attraction for the minimal values of some distributions.

Table 2. Domains of attraction for minimal values of some distributions

Distrbution	Domain of attraction
Normal Lognormal	Gumbel
Cauchy	Frechet
Exponentional Gamma Pareto Uniform	Weibull

Source: based on Castillo et all (2005).

Further analysis will concentrate on the use of the Gumbel distribution as a description of the random variable $X_M = \max\limits_{i=1,2,\dots,n}(X_1, X_2, \dots X_n)$ where $(X_1, X_2, \dots X_n)$ is a random sample taken from the $F(x)$ distribution which belongs to the Gumbel distribution domain of attraction (see Table 1).

2.2. Gumbel distribution

The Gumbel distribution frequently appears in practical problems. It could take place when the maximal or the minimal values from some distributions are observed. The density function of the Gumbel random variable with parameters λ and δ is given by (M. Evans et all, 2000)

$$f(x) = \frac{1}{\delta} e^{\frac{-(x-\lambda)}{\delta} - e^{\frac{-(x-\lambda)}{\delta}}} \quad \text{for } x \in R \tag{9}$$

The cumulative distribution function is given by the formula

$$F(x) = 1 - e^{-e^{\frac{-(x-\lambda)}{\delta}}} \quad \text{for } x \in R \tag{10}$$

The mean and the variance of the random variable with parameters λ and δ have the form of

$$\mu = \lambda + 0{,}57772\delta \text{ and } \sigma^2 = \frac{\pi^2 \delta^2}{6} \tag{11}$$

The probability density function and the cumulative density function of the random variable with the Gumbel distribution with parameters $\lambda = 0$ and $\delta = 1$ is presented in Figure 1.

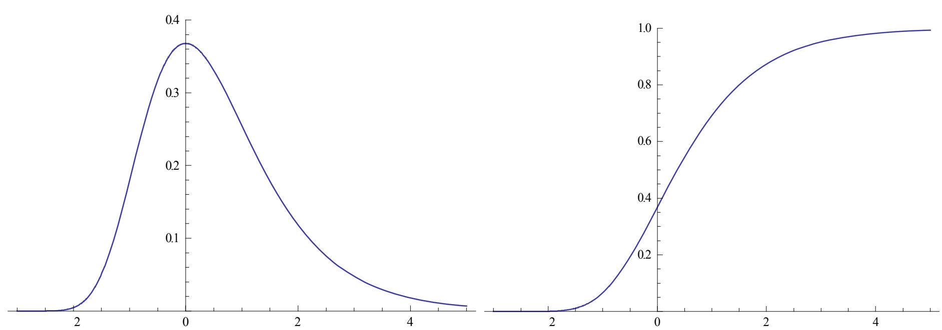

Fig. 1. Probability density function (left) and cumulative density function (right) of the random variable within the Gumbel distribution with parameters $\lambda = 0$ and $\delta = 1$

Source: own study in Mathematica.

III. CONTROL CHART $\overline{\mathbf{X}}$ AND EXTREME VALUES CONTROL CHART *ExV*

The control chart $\overline{X}$ is used for monitoring the average level of the process. This control chart could be used if the samples were taken from the normal population but D.C. Montgomery (1996) notes that this tool is robust to the normality assumption and can be employed unless the population is extremely non-normal.

Let $(X_1, X_2, \ldots X_n)$ be a random sample taken from the distribution $F(x)$ which belongs to the Gumbel distribution domain of attraction (see Table 1). Due to the robustness to the normality assumption the control chart $\overline{X}$ is often used to monitor this process. Let us consider a point over the upper control limit (UCL) as an indication that the process is out of control. The UCL is plotted on the level equal to the mean of the process plus 3 times the standard deviation.

Let us consider the maximum value for each n element sample. The distribution of these extremes could be approximated by the Gumbel distribution. The parameters of the Gumbel distribution could be estimated based on the N_0 data samples from the in-order process. The extreme values control chart *ExV* could be used to monitor such processes. On the chart *ExV,* the maximum values are marked for each sample. The central line (CL) of this control chart is plotted on the level equal to the median estimated the Gumbel distribution. The 10 points in a row over the CL is an indication of the incorrectness of the process.

The main characteristic for the control chart is the average run length (ARL). The ARL is the average number of points marked on the chart until an out-of-control condition is signaled. It is the expected value of the run length distribution (G. Kończak, 2007). The ARL is often used to compare the properties of control charts.

IV. MONITORING NORMAL AND NON-NORMAL PROCESSES – MONTE CARLO STUDY

The properties of the control chart $\overline{X}$ and the proposed control chart *ExV* based on the extreme values distribution have been analyzed in the Monte Carlo study. Computer simulations have been done for the data from normal, log-normal, gamma and exponential distributions. All the analyzed distributions belong to the domain of attraction of the Gumbel distribution (see Table 1). The estimated values of ARL_0 for the ordered processes and ARL_1 for the out-of-order processes were obtained in the Monte Carlo study.

The simulation procedure consisted of the following steps:

1. Control limits for the $\overline{X}$ and *ExV* control charts were established based on the samples data from distribution $F(x)$. These control limits were estimated based on $N_0 = 100$ samples of the sizes n ($n = 5, 10$).

2. For $k = 1000$ samples of the size n ($n = 5, 10$) from the in-order processes (see Table 3) the probabilities of an out-of-order signal were estimated and ARL_0 was obtained.

3. For the out-of-order processes (see Table 3) the ARL_1 values were estimated.

The control chart is better if ARL_0 is as high as possible and ARL_1 as low as possible. The main characteristics of the analyzed normal, exponential, log-normal and gamma processes are presented in Table 3. The densities of the analyzed distributions are presented in Figure 2.

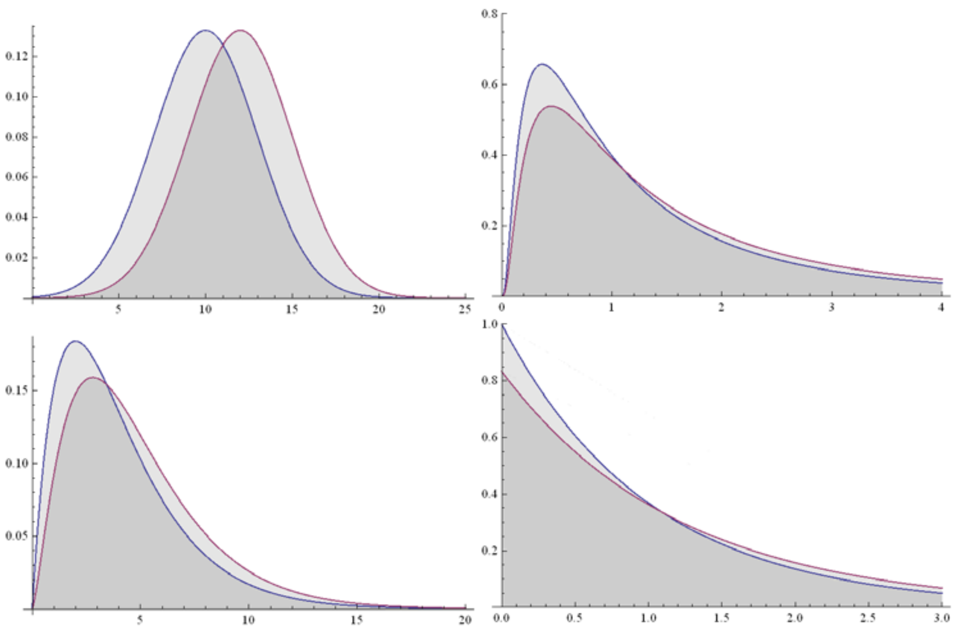

Fig. 2. Densities of distributions analyzed in Monte Carlo study (normal, log-normal, gamma and exponential)

Source: own study in Mathematica.

The first of the considered distributions is symmetric (normal distribution) and the others are characterized by various degrees of right skew.

Table 3. The main characteristic of the analyzed processes

Distribution	Expected value EX	In-order process	Out-of-order process
Normal N(μ, σ)	μ	N(10,3)	N(12,3)
Log-normal LN(μ, σ)	$e^{\mu+\frac{\sigma^2}{2}}$	LN(0; 1)	LN(0.2; 1)
Gamma G(α, β)	$\alpha\beta$	G(2; 2)	G(2.4; 2)
Exponential E(λ)	$\frac{1}{\lambda}$	E(1)	E(1/1.2)

Source: own research.

The estimated values of ARL_0 and ARL_1 for the analyzed processes are presented in Table 4 and in Figure 3.

Table 4. Estimated ARL values for the ordered and out-of-order processes

Distribution	$n = 5$				$n = 10$			
	ExV		$\overline{X}$		ExV		$\overline{X}$	
	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1	ARL_0	ARL_1
Normal	582.8	24.0	1078.7	70.1	546.7	11.6	902.5	5.8
Lognormal	4766.2	435.9	62.7	30.2	4209.2	297.5	73.1	26.4
Gamma	1166.4	72.0	169.5	41.5	1084.5	52.5	252.6	28.0
Exponential	1235.4	118.5	109.2	31.9	1069.2	71.2	167.8	29.9

Source: own research.

It could be noticed that the $\overline{X}$ control chart has a better performance only for the normal processes. For other processes the ARL_0 values for the control chart $\overline{X}$ are too low (due to the non-normality of these processes). The Monte Carlo study has shown that it is possible to use the presented proposal to monitoring non-normal processes.

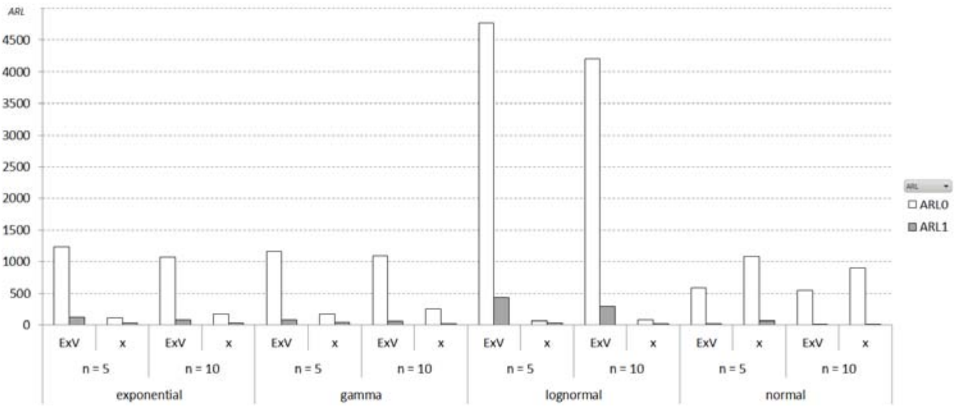

Fig. 3. Estimated ARL values for the ordered and out-of-order processes

Source: own study, data from Table 4.

V. CONCLUDING REMARKS

The construction of the Shewhart control charts assumes normality monitored characteristic. Due to the robustness to the normality assumption these control charts could be used even for the non-normal processes. Another way of monitoring such processes has been presented in the paper. The proposal of the extreme values control chart *ExV* was presented. This control chart could be evaluated for the samples taken from the distributions which belong to the Gumbel domain of attraction. The properties of the $\overline{X}$ and *ExV* control charts were analyzed in the Monte Carlo study for some typical distributions from the Gumbel domain of attraction. The Monte Carlo study has shown that the *ExV* extreme value control chart gives better performance than the $\overline{X}$ control chart for the non-normal processes. The $\overline{X}$ control chart should be used only for normal processes. The *ExV* control chart could be useful for monitoring the non-normal processes where the distribution of diagnostic variable belongs to the Gumbel domain of attraction.

REFERENCES

Bhattacharya P.K., Frierson D. (1981) *A nonparametric control chart for detecting small disorders*, Annals of Statistics, vol. 9, no. 3, p. 544-554.

Castillo E., Hadi A.S., Balakrishnan N., Sarabia J.M. *Extreme Value and Related Models with Applications in Engineering and Science*. A John Wiley & Sons, Inc. New Jersey 2005.

Evans M., Hastings N., Peacock B., *Statistical Distributions*, A John Wiley & Sons, Inc. New York 2000.

Kończak G. (2007) *Metody statystyczne w sterowaniu jakością produkcji*. AE Katowice.

Kończak G. (2010) *The Moving Average Control Chart Based on the Sequence of Permutation Tests*, [in:] Proceedings in Computational Statistics 2010, ed.: Y. Lechevallier, G. Saporta, s. 1199-1206. Physica-Verlag, Berlin Heilderberg.
Montgomery D.C. (1996) *Introduction to statistical quality control*, John Wiley & Sons, Inc.

Grzegorz Kończak

O WYKORZYSTANIU ROZKŁADU WARTOŚCI EKSTREMALNYCH W MONITOROWANIU PROCESÓW

Klasyczne metody pozwalające na monitorowanie poziomu przeciętnego procesów produkcyjnych odwołują się zwykle do założenia normalności rozkładu badanej zmiennej i niezależności kolejnych pomiarów. W wielu analizach statystycznych interesująca jest ocena poziomu przeciętnego badanej charakterystyki. Tak jest np. przy monitorowaniu procesów z wykorzystaniem karty kontrolnej $\overline{X}$. Jednak w wielu zastosowaniach może być interesująca ocena wielkości maksymalnych lub minimalnych. W artykule przedstawiono propozycję wykorzystania własności rozkładu wartości ekstremalnych w monitorowaniu procesów. Rozważania teoretyczne zostały uzupełnione analizami symulacyjnymi własności proponowanej metody.