



# The Use of the Extended Generalized Lambda Distribution for Controlling the Statistical Process in Individual Measurements

Sajad Noorian<sup>1,\*</sup>, Majid Nili Ahmadabadi<sup>2</sup>

<sup>1</sup> *Department of Statistics, University of Qom, Iran*

<sup>2</sup> *Department of Management, University of Qom, Iran*

**Abstract** In quality control (QC) of the individual pieces tested, the distribution of the intended variable is not normal in many cases. Therefore, the use of conventional statistical methods to design the control charts (X chart) can lead to misleading and inaccurate results, and hence, it can cause problems to users; because in such a situation it would not be expected that the failure in the production line will be detected in a timely manner. Therefore, in this study, the use of Extended Generalized Lambda distribution (XGLD) has been proposed for approximating the statistical distribution of data and finding a quality control chart based on it. For this purpose, the distribution function of data was firstly estimated by the XGLD, and then, based on this distribution, the control limits were calculated. The research data consists of the tensile strength of 149 aluminum plates categorized into 22 classes. For the goodness of fit test, Chi-square test was utilized and to evaluate the effectiveness of the proposed model, the average run length (ARL) method was used. The results showed that the ARL values in the proposed method are lower than that in the conventional method. Since this is true at all points, it can be concluded that the use of the new method can lead to the faster detection of the systems failure and consequently, the subsequent costs can be reduced. In addition, since the XGLD is a highly flexible distribution and can estimate many conventional (and even unconventional) distributions with high precision, the use of the proposed method instead of using other cases which are based on distributions other than normal, will increase the accuracy of the systems failure detection.

**Keywords** Average Run Length (ARL), Control X Chart, Extended Generalized Lambda Distribution (XGLD), Non-Normal Quality Control

**AMS 2010 subject classifications** 62G99, 62P30

**DOI:** 10.19139/soic.v6i4.420

## 1. Introduction

There are many situations in reality in which the size of a sample is  $n=1$  to monitor the process, or, in other words, samples can consist of individual or individual units. For example, in destructive tests, samples are tested individually to be nearly destroyed. [16] has introduced some of these conditions, which include the use of automated measurement and inspection technology, slow production rates, high interval between the products production, the difference in repeated measurements due to laboratory or analytical errors, very low standard deviation, costly repeated sampling, irrationality of the repeated sampling, impossibility of the repeated sampling, etc.

In such cases, the use of control methods that assume the normal distribution of variables is accompanied by a significant error. Therefore, in these situations, the use of individual value (IV) and moving range (MR) charts for continuous variables is proposed. In these charts, using the moving range of individual observations, the estimation of the standard deviation and the subsequent determination of the control limits of the graphs are discussed. These

\*Correspondence to: Sajad Noorian (Email: s.noorian@qom.ac.ir). Department of Statistics, University of Qom, Qom, Iran.

graphs are used in cases where only one sample can be used ([16]). One of these tools is the moving range graph, in which the moving range between two sequential observations is defined as the absolute value of the difference between these two observations  $MR_i = |y_i - y_{i-1}|$ , in which  $i = 2, 3, \dots, T$  and  $T$  is the number of sampling times. In the routine analysis of such data, first, the average of the observations and then the average of the moving ranges are calculated. The control limits are calculated as follows:

$$\begin{cases} UCL = \bar{X} + 3\frac{\overline{MR}}{d_2} \\ CL = \bar{X} \\ LCL = \bar{X} - 3\frac{\overline{MR}}{d_2} \end{cases} \quad (1)$$

And

$$\begin{cases} UCL = D_4\overline{MR} \\ CL = \overline{MR} \\ LCL = D_3\overline{MR} \end{cases} \quad (2)$$

where the constants  $D_3$  and  $D_4$  are defined using the mean and the standard deviation ( $d_2$  and  $d_3$ , respectively) of the relative range random variable  $W = R/\sigma$  as follows:  $D_3 = 1 - 3\frac{d_3}{d_2}$ ,  $D_4 = 1 + 3\frac{d_3}{d_2}$ , ([16]).

One of the drawbacks to this approach is that, in order to estimate the changes of the process, the user may want to estimate the standard deviation of the process, using the standard deviation of the sample (S). But it should be noted that typically, the sampling units are assumed to be independent of each other while sequential observations are not necessarily the same. In fact, using the average of the moving range (MR) for estimating  $\sigma$ , emphasizes the variability between sequential observations, while utilizing the standard deviation of the sample emphasizes the variability in the entire set of data. When there is an unstable trend in the data, different estimations may be obtained by MR and S for the standard deviation of the process ([3]).

[2] studied the performance of the above graphs for individual items in the absence of normal distribution of data. By examining several distributions for skewness and asymmetric representation of pseudo-normal data, they found out that the performance of control charts is significantly affected when the distribution of data is slightly distanced from the normal distribution. In this case, it is necessary to test the normality hypothesis, using statistical tests, before the data analysis is started, in order to avoid the consequences of misusing these graphs ([17].)

[11] obtained the X and R control charts in skewed data using Generalized Lambda Distribution (GLD), the percentile matching method and weighted variance method with left and right tail. [12] also obtained some control charts based on weighted bootstrap using some weights for each stratum in skewed populations. The novelty of this paper with respect to [11] and [12] consists in introducing the application of Extended Generalized Lambda distribution (XGLD) for individual sampling scheme when the population doesn't necessarily follow the normal distribution.

Therefore, it is necessary to improve the existing methods for drawing control charts in a situation where the distribution of data is not necessarily normal.

[19] have used the generalized 5-parameter lambda distribution ([13]) to control the process. In this method, they used the MM method (Moment Matching) to estimate the distribution parameters and did not refer to the statistical control of individual items in their proposal. Therefore, in addition to the circumstances in which the distribution of the items under study is unclear, it is still possible to consider the circumstances in which individual items are to be studied.

In another paper, [18] introduced a new distribution called the Extended General Lambda Distribution (XGLD), which is much more flexible than the generalized 5-parameter distribution of Lambda; they proposed it to estimate the uncertain distribution of data as well as the conventional functions.

In this paper, with the assumption that the statistical distribution of the random variable is unknown, we propose a new method for its statistical control. In the proposed method, the XGLD has been used. This distribution is very flexible and can cover a wide range of conventional and unconventional statistical distributions. The method of moments is used to estimate the parameters of this distribution. The difference between this article and the work of [19] is the development of methods of drawing quality control charts for individual samples, using an XGLD.

The rest of the paper is as follows: First, after an introduction to GLD, we introduce the XGLD and the method for estimating its parameters. In the following, after introducing the research method, using the selected data, the proposed method will be used. Finally, the proposed method will be validated using the Average Run Length (ARL).

## 2. Generalized Lambda Distribution

Many efforts have been made to provide a distribution of probability density that can provide the necessary flexibility to estimate other statistical distributions. [27] introduced the distribution of Lambda family as a Quantile function

$$Q(p) = \begin{cases} \frac{p^\lambda - (1-p)^\lambda}{\log(p)^\lambda}, & \lambda \neq 0 \\ \frac{\log(p)}{1-p}, & \lambda = 0 \end{cases} \quad (3)$$

where  $0 \leq p \leq 1$  and the form of the distribution is based on the value of the parameter  $\lambda$ .

[21] introduced the following four-parameter Quantile function for Lambda distribution:

$$F^{-1}(p, \lambda) = F^{-1}(p, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2} \quad (4)$$

Where  $p, p \in (0, 1)$ , is the cumulative probability,  $\lambda_1$  is the location parameter,  $\lambda_2$  is the scale parameter, and  $\lambda_3$  and  $\lambda_4$  are the shape parameters of the distribution. This distribution can be considered as the extended form of Tukey distribution because if  $\lambda_2 = \lambda_3 = \lambda_4 = \lambda$  and  $\lambda_1 = 0$ , the Tukey distribution is obtained.

Following this, [13] introduced the generalized five-parameter Lambda distribution as the Quantile function, in which  $\lambda_1$  is the location parameter,  $\lambda_2$  and  $\lambda_3$  are the scale parameters of variables, and  $\lambda_4$  and  $\lambda_5$  are exponential parameters:

$$F^{-1}(p, \lambda) = F^{-1}(p, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = \lambda_1 + \lambda_2 p^{\lambda_4} - \lambda_3 (1-p)^{\lambda_5}, \quad 0 \leq p \leq 1, \quad (5)$$

There are several methods for estimating the parameters of this distribution. For a review on these methods, see [23], [10], [21], [9], [14], [7], [22], [4], [8] and [19].

## 3. Extended Generalized Lambda Distribution (XGLD)

[18], by rewriting the generalized five-parameter Lambda distribution in equation (5) as:

$$F^{-1}(p, \lambda) = F^{-1}(p, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = \lambda_1 + \lambda_2 p^{\lambda_4} - \lambda_3 (1-p)^{\lambda_5} = \alpha_0 + \alpha_1 p^{\beta_1} + \alpha_2 (1-p)^{\beta_2}$$

introduced the XGLD as:

$$\begin{aligned} x &= XGLD(p, \alpha_0, \alpha_1, \dots, \alpha_k, \beta_1, \beta_2, \dots, \beta_k, c_1, c_2, \dots, c_k) \\ &:= \alpha_0 + \sum_{i=1}^k \alpha_i (p - c_i)^{\beta_i} \end{aligned}$$

where,  $(p - c_i)^{\beta_i}$  is the monotony power. Generalized five-parameter Lambda distribution is a special case of high distribution in which  $k = 2$ ,  $c_1 = 0$  and  $c_2 = 1$ . When  $k$  increases, the precision of estimators will increase, although it is natural that along with that, the volume of the calculations also increases.

In this paper, a simple form of XGLD is used to determine the control limits in single-sampling, considering  $k = 3$ ,  $c_1 = 0$ ,  $0 \leq c_2 = c \leq 1$ , and  $c_3 = 1$ .

$$\begin{aligned} x &= XGLD(p, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, c_1 = 0, c_2 = c, c_3 = 1) \\ &:= \alpha_0 + \alpha_1 (p)^{\beta_1} + \alpha_2 (p - c)^{\beta_2} + \alpha_3 (p - 1)^{\beta_3} \\ &= \alpha_0 + \alpha_1 (p)^{\beta_1} + \alpha_2 (p - c)^{\beta_2} - \alpha_3 (1 - p)^{\beta_3} \end{aligned}$$

In this case, the number of the parameters of the model is 8. Some relationships between the GLD and XGLD distributions are presented in [18]. The density function of XGLD is as follows:

$$f(x) = \frac{dp}{dx} = \frac{1}{dx/dp} = \frac{1}{d(\alpha_0 + \sum_{i=1}^k \alpha_i (p - c_i)^{\beta_i}) / dp} = \frac{1}{\sum_{i=1}^k \alpha_i \beta_i |p - c_i|^{(\beta_i - 1)}}$$

The  $m^{\text{th}}$  moment for this density function is obtained as follows:

$$E(X^m) = \int_{-\infty}^{+\infty} x^m f(x) dx = \int_0^1 (\alpha_0 + \sum_{i=1}^k \alpha_i (p - c_i)^{\beta_i})^m dp.$$

#### 4. XGLD Parameters Estimation

There are several methods for estimating GLD parameters, and subsequently XGLD parameters, some of which are Maximum Likelihood (ML), Method of Moments (MM), Percentile Matching (PM), Probability Weighted Moments (PWM), Minimum Cramer von-Mises estimators (MCM), Partial Least Squares estimators (PLS). [25] compared these methods in estimating the distribution parameters of GLD in the grouped data. [18] also suggested how to use some of the above methods to estimate XGLD parameters. In this research, the method of moments is used to estimate XGLD parameters. If the number of parameters is equal to  $k$ , then by equating the 1<sup>st</sup> to  $k^{\text{th}}$  distribution moments with the corresponding sample moments, a nonlinear  $k$ -equation and  $k$ -unknown system will be obtained and then used to estimate the model parameters, which can be answered by using numerical methods for solving nonlinear equations.

It should be noted that this system may have more than one acceptable answer. To validate the answers, the Chi-square goodness of fit test can be used. In this way, after classifying the data in  $h$  categories, the frequency observed in each category,  $O_i$ , is computed and compared with the expected frequency in each category,  $E_i$ . To obtain the  $E_i$ s, first, the probability of each one of the  $h$  categories is computed through the density function of XGLD. Then the expected frequency in each category is obtained by multiplying the total number of data in the probability of that category (the calculation methods are completely explained in [18]). The Chi-square test statistic is as follows:

$$\chi_0^2 = \sum_{i=1}^h \frac{(O_i - E_i)^2}{E_i} \quad (6)$$

and under the null hypothesis, it has the chi-square distribution with a degree of freedom  $df = h - k - 1$ , in which,  $h$  is the number of categories and  $k$  is the number of XGLD parameters (which in this study is equal to 8 parameters). Note that  $h$  should be greater than  $k + 1$  to have a positive  $df$  value. The null hypothesis in this test indicates that the sample is extracted from the obtained XGLD and the alternative hypothesis implies the opposite. The null hypothesis is rejected in favor of the alternative hypothesis, whenever  $\chi_0^2 > \chi_{h-k-1, 1-\alpha}^2$ , in which  $\chi_{h-k-1, 1-\alpha}^2$  is the higher percentile of the chi-square distribution with a degree of freedom  $h - k - 1$ , and the area of the right is equal to  $\alpha$ .

#### 5. Non-Normal Statistical Quality Control

To obtain the limits of the quality control charts, the data distribution density function should be available. In the absence of a density function, [28] used B-splines to estimate the density function and subsequently draw control charts. [5] used the Gram-Charlier density function to obtain a standard space model. [1] used the power normal family and [15] also used boron distribution to estimate the density function. [29] estimated the shape of the probability density function analytically using the Fokker-Planck-Kolmogorov equation in a random system and a form control procedure. [24] pointed out that the GLD could be used as a very good approximation for the density

function of the product data of a product line. [6] introduced how to use GLD to control the statistical process, and [20] also used lambda distribution to compute the efficiency indicators of the process.

In this paper, we present the control graphs of individual values using the XGLD function and compare them with the moving range graph for individual samples. These charts can monitor the production process very well and help to identify any deviations or changes that have occurred in the distribution, while these deviations or changes may not be detected when using the summarized statistics or when it is assumed that the sample is selected from a normal society.

In this research, the high flexibility and capabilities of XGLD (even compared to the GLD) enable the researcher to use the XGLD in many quality control schemes without worrying about not knowing the distribution of the community.

## 6. Control limits based on XGLD

Control limits are used to detect signals in process data that indicate that a process is not in control. In other words, if during the graphing of the data, a point or points of the sample are outside the scope of the control limits, then the user should be aware that the data generation process may be out of control. In individual samples, the IV and MR charts are commonly used. This paper suggests utilizing the XGLD to obtain the control limits. This method can also be used for almost all control charts.

Here, because the distribution of the population from which the samples are extracted is unknown, without diminishing the whole, it is assumed that the existing data is the same as the population under study and the empirical distribution function is used as an estimation for the unknown distribution function of the population ([26]) The algorithm used to obtain the control limits based on the XGLD function is as follows:

1. Start the production line and make sure that it is under control.
2. Determine the acceptable values for the probabilities of type I and type II errors, as well as the amount of deviation, according to the desired value for the average.
3. Take individual samples from the production process.
4. Draw a histogram of the dataset.
5. Estimate the XGLD parameters using the method of moments. It means that:
  - a. Using [13], estimate the parameters of the four-parameter GLD.
  - b. Estimate the XGLD parameters, using the relationship between the GLD and XGLD parameters and the given algorithm in [18].
6. Calculate the upper, middle and lower control limits based on the information obtained in step 5. If in step 4 the distribution of population was asymmetric, instead of using the approximation approach  $3\sigma$ , use the probabilistic approach. Otherwise, use the approximation approach  $3\sigma$ .
7. After obtaining the control limits, repeat the sampling step 3 again. It is not required to update the control limits for new individual samples. Using the calculated limits in step 6, check if the production process is in control. If yes, then go to step 5. Otherwise, by stopping the production process, you will have to correct the process by identifying and eliminating the causes of the deviations. Then go to step 3.

Using this algorithm, at each sampling time, a distribution will be obtained which represents the actual distribution of data. Therefore, at each sampling time, the changes that occur in the process, both in the average and in the standard deviation, can be recognized.

## 7. Application

### 7.1. Using XGLD to design a control chart (numerical example)

In this section, 149 data, which is related to the resistance of the aluminum plates against the pressure, is used to illustrate how the XGLD function is used in drawing the control charts. Since the number of the parameters

Table 1. Observed and expected frequencies to estimate the chi-square goodness of fit.

Number of category	Category limits	Observed frequency ( $O_i$ )	Expected frequency ( $E_i$ )	$\frac{(O_i - E_i)^2}{E_i}$
1	Below 1.10	0	0.53	0.53
2	1.10- 1.14	3	2.86	0.01
3	1.14-1.18	4	3.81	0.01
4	1.18-1.22	5	4.83	0.01
5	1.22-1.26	6	5.88	0.00
6	1.26-1.30	8	7.32	0.06
7	1.30-1.34	9	8.10	0.10
8	1.34-1.38	8	8.23	0.01
9	1.38-1.42	10	9.53	0.02
10	1.42-1.46	9	9.81	0.07
11	1.46-1.50	9	10.02	0.10
12	1.50-1.54	10	10.20	0.00
13	1.54-1.58	12	11.82	0.00
14	1.58-1.62	11	11.15	0.00
15	1.62-1.66	10	10.36	0.01
16	1.66-1.70	10	9.84	0.00
17	1.70-1.74	9	8.64	0.02
18	1.74-1.78	6	6.06	0.00
19	1.78-1.82	5	5.05	0.00
20	1.82-1.86	3	2.49	0.10
21	1.86-1.90	2	1.66	0.07
22	Above 1.90	0	0.80	0.80
	Total	149	149	1.94

Table 2. Estimating the model parameters of XGLD, using the sample data.

index ( $\theta_i$ )	$\alpha$	$\beta$	$c$
0	1.7337	—	0.5202
1	0.6425	0.6145	—
2	0.6145	0.5882	—
3	0.0018	0.0012	—

of XGLD is 8 and the degree of freedom of Chi-square test statistic is  $df = h - k - 1$ ,  $k = 8$ , so the number of categories should be greater than the number of parameters. Hence, the data is classified into 22 categories, which can be seen in Table 1. In the following, the method for estimating XGLD parameters and then the use of the parameters in drawing the control charts are discussed. For these data, eight sample moments of order 1 to 8,  $M_1$  to  $M_8$ , are obtained and by equating them to the distribution moments, and solving the following nonlinear system of eight- equation and eight-unknown, the XGLD parameters will be estimated:

$$\int_0^1 (\alpha_0 + \alpha_1(p)^{\beta_1} + \alpha_2(p-c)^{\beta_2} - \alpha_3(1-p)^{\beta_3})^m dp = M_m, \quad m = 1, 2, \dots, 8, \quad (7)$$

After solving the equation (7), the model parameters are estimated as shown in Table 2:

Using the above estimates, the 1<sup>st</sup> to 8<sup>th</sup> moments of the XGLD distribution, along with the sample moments calculated by the data, are given in Table 3. As you can see, the moments estimation is very close, which indicates the high accuracy of XGLD in estimating the distribution of data.

To test if the data really follows the XGLD distribution with the above parameters, the Chi-square goodness of fit test given in formula (6) was used. In each category from Table 1, the limits of categories, the observed and expected frequencies, as well as the amount of the test statistic in that category are given.

The critical value in the chi-square distribution with a degree of freedom of 13 ( $df = 22 - 8 - 1 = 13$ ), and the first type error rate of 0.05 ( $\alpha$ ) is equal to 5.89. Given the fact that the test statistic is less than the critical value

Table 3. Average, standard deviation and central moments, using the data and XGLD distribution.

XGLD Approximation	Main Data	Moments
1.5030	1.5010	Average
0.1889	0.1889	Standard deviation
0.0357	0.0357	2 <sup>nd</sup> Central moment
-0.0006	-0.0007	3 <sup>rd</sup> Central moment
0.0027	0.0028	4 <sup>th</sup> Central moment
-0.0001	-0.0001	5 <sup>th</sup> Central moment
0.0003	0.0003	6 <sup>th</sup> Central moment
0.0000	0.0000	7 <sup>th</sup> Central moment
0.0000	0.0000	8 <sup>th</sup> Central moment

in the chi-square distribution, the null hypothesis in the level of  $\alpha = 0.05$  is not rejected in favor of the alternative hypothesis. Therefore, it can be concluded that on this collection of data, the distributed XGLD is fitted; in other words, the data follows this distribution.

The control limits are determined with  $\alpha = 0.05$ . According to the calculated values for average and standard deviation of data in Table 3, the control limits for individual observations using the Montgomery equation are calculated as follows:

$$\begin{cases} UCL = 1.501 + 1.96(0.189) = 1.871 \\ CL = 1.501 \\ LCL = 1.501 - 1.96(0.189) = 1.131 \end{cases}$$

But the control limits computed by using the XGLD approach is obtained as (1.123, 1.840). It is clear that the control limits computed by XGLD approach is tighter than that of Montgomery approach. Figure 1 shows the control charts of the data using both Montgomery and XGLD approaches.

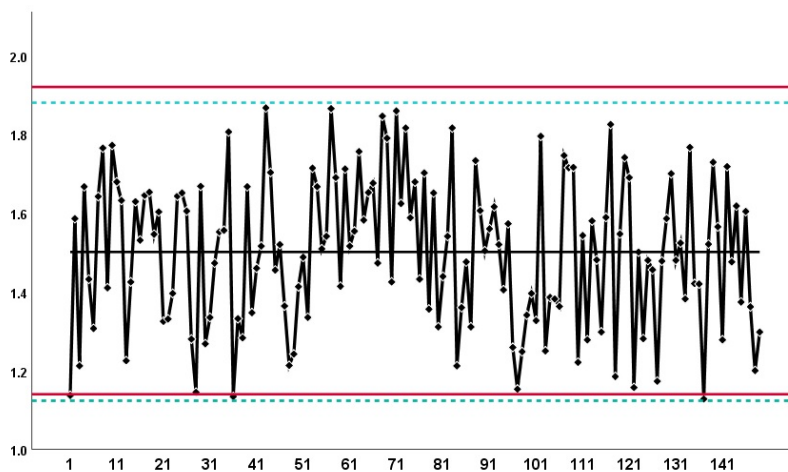


Figure 1. Individual control charts of the data using Montgomery (continuous red lines) and XGLD (dashed green lines) approaches

**7.2. Validating the proposed model using ARL**

The average run length (*ARL*) of the control chart is one of the ways to assess and evaluate the quality control method. *ARL* is the average of the number of points that are drawn on the control charts before the observation of an alert, and for a Shewhart control chart, it is defined as the opposite probability of drawing a point out of

the control limits (Montgomery, 2012). In other words, ARL is equal to the reverse of probability of plots out of control. So,  $ARL_0 = \alpha^{-1}$  for the in-control process and  $ARL_1 = (1 - \beta)^{-1}$  for the out-of-control process where  $\alpha$  and  $\beta$  are the probability of the first-type and second-type error, respectively. Determining the amount of  $\alpha$  usually depends on the features of the product line, as well as the type of the product. To obtain the control limits with a distance of three standard deviations ( $3\sigma$ ) from the average, the value of  $\alpha$  is considered 0.0027.

Because the values of  $\alpha$  and  $\beta$  are related to each other, the control charts are distinguished based on their ability to detect the second-type error. In other words, if two control charts are drawn with the same values of  $\alpha$ , the control chart which has a lower  $ARL_1$  and can stop the out of control system sooner is the more efficient control chart. In this research, the value of  $\alpha$  for each graph is 0.05.

If the distribution of the data is clear, the probability of the second-type error ( $\beta$ ) is equal to:

$$\beta = P(LCL < X < UCL | \mu = \mu_1)$$

To express it better, if the data has normal distribution the value of  $\beta$  is:

$$\begin{aligned} \beta(\delta) &= P(LCL < X < UCL | \mu = \mu_1) = P(\mu_0 - z_{(1-\alpha/2)}\sigma < X < \mu_0 + z_{(1-\alpha/2)}\sigma | \mu = \mu_1) \\ &= P\left(\frac{\mu_0 - z_{(1-\alpha/2)}\sigma - \mu_1}{\sigma} < \frac{X - \mu_1}{\sigma} < \frac{\mu_0 + z_{(1-\alpha/2)}\sigma - \mu_1}{\sigma} \mid \mu = \mu_1\right) \\ &= P(-z_{(1-\alpha/2)} - \delta < Z < z_{(1-\alpha/2)} - \delta) \\ &= \Phi(z_{(1-\alpha/2)} - \delta) - \Phi(-z_{(1-\alpha/2)} - \delta) \\ &= \int_{-z_{(1-\alpha/2)} - \delta}^{z_{(1-\alpha/2)} - \delta} \varphi(z) dz \end{aligned}$$

Where  $\delta = (\mu_1 - \mu_0)/\sigma$ ,  $\Phi$  and  $\varphi$  represent the distribution and density functions of a standard normal random variable.

The algorithm which is used to estimate the  $\beta(\delta)$ , when the variable has an XGLD distribution, is as follows:

1. Using the presented algorithm in the previous section, the XGLD parameters are estimated and then the control limits are calculated.
2. Consider the value of  $\delta$  that is equal to  $ks$ , where  $s$  is the standard deviation of the set which contains all individual samples, and  $k$  is usually the real number in the range  $-3$  to  $3$ . Add the fixed value to all the available data and then obtain the XGLD parameters for the new data set, again. So,

$$x(p) = \alpha_0 + \sum_{i=1}^k \alpha_i (p - c_i)^{\beta_i} \quad (8)$$

3. Find the values of  $y_{Lower}$  and  $y_{Upper}$ , in a way that the following equations are obtained:  $x(y_{Lower}) = LCL$  and  $x(y_{Upper}) = UCL$ . Then calculate  $\beta(\delta)$ , using the equation  $\beta(\delta) = y_{Upper} - y_{Lower}$ , and finally  $ARL_1$  is calculated using the relation  $ARL_1 = (1 - \beta)^{-1}$ . Note that  $y_{Lower}$  and  $y_{Upper}$  are two values that the equation (8) is established for them.
4. Repeat steps 2 to 3 for all  $k$  values, and get the  $ARL_1$  value.

In Figure 2, ARL values are drawn under normal and XGLD distributions. The value of ARL in all states is equal to  $(1 - \beta)^{-1}$  and can be computed at any point. For obtaining the ARL graphs, first  $\beta$  is computed. According to the definition of the second type error, when the real mean of the distribution is shifted to  $k\sigma$ , the surface under the new curve (displaced) at the  $LCL < X < UCL$  interval represents the value of  $\beta$ . Therefore, the value of  $\beta$  can be calculated for different values of  $k$ . The used mathematical relationships are explained in the previous sections. As it can be seen, when we assume that the data have normal distribution, all points of the chart are upper than the points of other two graphs. This means that in the case of system failure, then the true alarms is occurred with a bad delay and it leads to increasing in failure cost.



In the ARL graph of XGLD,  $\beta$  is calculated for each value of  $k$  (which indicates the system deviation from the mean) using the control limits calculated by the XGLD and the 4-step algorithm at each point. As can be seen, for every  $k$ , the value of  $\beta$  is lower, indicating that the system needs less defective products to have a true alarms.

In the Ideal ARL chart it is seen that immediately after the failure (the first positive or negative  $k$  values), the failure is detected by the control system and the production line stops. Such a control system does not exist in practice, and all the efforts of researchers in the field of quality control is to find a control system near to it. That's why it is called the ideal system. With such a control system, the probability of stopping the production line when it is not out of control (healthy) is equal to  $\alpha$ , also the probability of type II error ( $\beta$ ) (that an out of control system continues to generate products) is equal to zero.

As you can see, the ARL diagram is closer to the ideal ARL diagram, when the XGLD is used instead of the normal distribution as the fitted distribution on data. It should be noted that, unlike  $ARL_1$ , the  $ARL_0$  must be large in value, so that the possibility of a wrong alert is low. Given the  $\alpha = 0.05$ , the  $ARL_0$  value in both graphs is 20. Therefore, this index cannot be considered as an indicator of the effectiveness of the graphs in relation to each other.

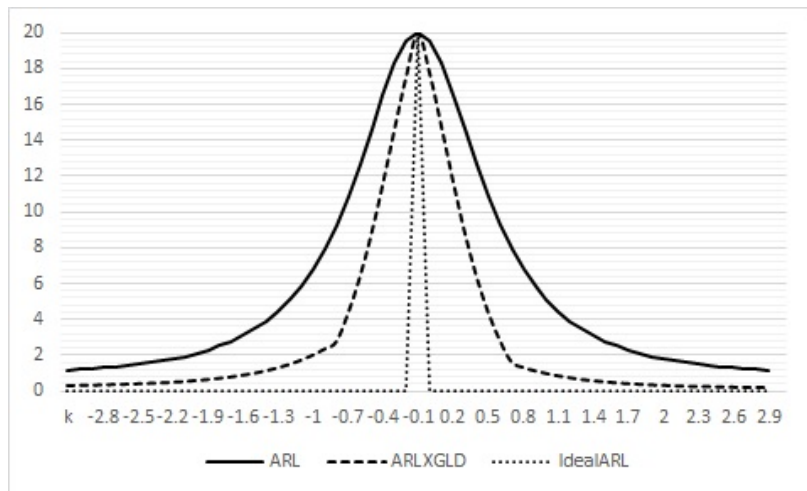


Figure 2. ARL graphs in three ideal modes (stipple), distribution of XGLD (line) and normal distribution (continuous line)

It should be noted that in this study, to compare the graphs, the value of the first-type error in both control graphs is  $\alpha = 0.05$ . As shown in Figure 2, at  $k = 0$  (when the system is completely healthy), the ARLs value is 20. That is, when using both charts, detecting a defective product among the 20 healthy products is considered normal and does not stop the production line. This value is equal to 20 in all the other control charts that are drawn with the same alpha (regardless of the distribution of data and the method for drawing the control chart). In addition, at this point in the graph, both ARLs (that is,  $ARL_0$  and  $ARL_1$ ) are equal. But in other places  $K$  (that is, when the system goes out of control), it is observed that the amount of  $ARL_1$  is lower in the proposed chart. This means that when the system is out of control and the XGLD is used in drawing a control chart, the system will stop earlier (with a smaller number of defective products) and less amount of costs.

### 7.3. Simulation

As discussed in the previous section, after extracting the values of  $\alpha$ ,  $\beta$  and  $c$ , the XGLD function for the initial data was formed as follows:

$$x = 1.7337 + 0.6425p^{0.6145} - 0.6145(1 - p)^{0.5882} + 0.0018(p - 0.5202)^{0.0012}$$

According to [18], in the distribution of XGLD,  $p$  is the cumulative relative probability of  $x$  ( $p = F(x)$ ). So, for generating different values of  $x$ , different values of  $p$ , ( $0 < p < 1$ ) can be generated randomly. By doing this, the

Table 4. The True and Estimated Value of XGLD parameters.

Parameter	True Value	Estimated Value	Difference
$\alpha_0$	1.7337	1.7574	0.02374
$\alpha_1$	0.6425	0.6369	0.00561
$\alpha_2$	0.6145	0.6122	0.00232
$\alpha_3$	0.0018	0.0018	0.00001
$\beta_1$	0.6145	0.6142	0.00028
$\beta_2$	0.5882	0.6033	0.01513
$\beta_3$	0.0012	0.0012	0.00003
$c$	0.5202	0.5328	0.01259
Differences Average			0.00746

distribution of generated  $x$  will be based on the XGLD. Accordingly, using the following process, 1000 random values were generated for  $x$ :

1. Using random numbers, generate 1000 values between 0 and 1 and name them  $p_i$ ,  $1 < i < 1000$ .
2. Put any  $p_i$  in the above formula to calculate the numerical value for each  $x_i$ .
3. Categorize  $x_i$ s. In this simulation study, we set the number of categories to 22.

After generation of 1000 data ( $x_i$ ), the first to eighth moments were calculated. Then, using a mechanism similar to that described in the previous section, the XGLD parameters for the simulated data are estimated. These values are shown in the Table 4 and compared to the true values of XGLD distribution:

It can be seen that in almost all cases, there is no significant difference between the estimated and true values of the parameters. In some cases this difference is zero or very close to it and the total average is 0.007, which is a small value. This suggests that XGLD works well in individual measurements paradigm and it can be used as a precision estimator for the population as well.

## 8. Conclusion:

In this paper, a new XGLD-based approach is developed for drawing the control charts in individual sampling cases. In Many cases, it is not logical to pick up more than one sample from the production line. In these situations, the X chart is usually used to control the process, which is not sensitive enough to detect the detectable deviations. In addition, when the distribution of the process is not normal or when this distribution is unclear, one can no longer rely on the X chart, which is often the case in single sampling. In this paper, the distribution of data is estimated, using the XGLD. With regard to the high flexibility of this distribution compared to the GLD (which in turn is more flexible than many distributions), the distribution of data is highlighted precisely. Using this distribution, the control limits were calculated in a manner similar to those used by [19] for distributing the GLD. Comparison of the ARL index in the conventional method and the proposed method showed that the reliability of the proposed method is higher at all points.

The difference between this research and the proposed method of ([19]) is the use of XGLD. Another innovation in this paper is the emphasis on using the XGLD to control the statistical quality of individual samples. Using the proposed method in this study will also be effective when the statistical distribution of the manufactured items is unclear. This research proposes the use of the Chi-square test for the goodness of fit. Another strong point of this chart is that it is not necessarily symmetrical, compared with the normal  $ARL_1$  curve (which is necessarily symmetrical). In individual sampling, in many cases, the distribution of data is not normal (in these cases, the use of the proposed method can lead to discovery of the data distribution). But even if the data distribution function is normal or close to normal, it has skewness or kurtosis, practically. In these cases, the XGLD distribution can also specify these features and draw a control chart with these considerations. Finally, some data has distributions that are not yet named (such as two-peak distributions). The XGLD's ability to transform can lead to determining these distributions and drawing control charts for them.

## REFERENCES

1. W. Albers, W. C. Kallenberg, and S. Nurdianti, *Parametric control charts*, Journal of Statistical Planning and Inference, vol. 124, no. 1, pp. 159–184, 2004.
2. C. Borrer, D. C. Montgomery, and G. C. Runger, *Robustness of the EWMA control chart to non-normality*, Journal of Quality Technology, vol. 31, no. 3, pp. 309–316, 1999.
3. E. N. Cruthis, and S. E. Rigdon, *Comparing Two Estimates of the Variance to determine the Stability of a Process*, Quality Engineering, vol. 5, no. 1, pp. 67–74, 1992.
4. L. Devroye, *Random variate generation in one line of code*, Simulation Conference, 1996. Proceedings. Winter, 265–272, 1996.
5. M. G. Forbes, M. Guay, and J. F. Forbes, *Control design for first-order processes: shaping the probability density of the process state*, Journal of process control, vol. 14, no. 4, pp. 399–410, 2004.
6. B. Fournier, L. Minguet, L. Niamien, and N. Rupin, *Les lambda distributions appliquées au SPC*, unpublished report, 2003.
7. M. Freimer, G. Kollia, G. S. Mudholkar, and C. T. Lin, *A study of the generalized tukey lambda family*, Communications in Statistics-Theory and Methods, vol. 17, no. 10, pp. 3547–3567, 1988.
8. W. Gilchrist, *Statistical modelling with quantile functions*, CRC Press, 2000.
9. D. C. Hoaglin, *The small-sample variance of the Pitman location estimators*, Journal of the American Statistical Association, vol. 70, no. 352, pp. 880–888, 1975.
10. B. L. Joiner, and J. R., Rosenblatt, *Some properties of the range in samples from Tukey's symmetric lambda distributions*, Journal of the American Statistical Association, vol. 66, no. 334, pp. 394–399, 1971.
11. S. C. Kao,  *$\bar{X}$  and R control charts based on weighted variance with left-right tail-weighted ratio for skewed distributions*, Communications in Statistics-Simulation and Computation, vol. 46, no. 4, pp. 2714–2732, 2017.
12. S. C. Kao, *A control chart based on weighted bootstrap with strata*, Communications in Statistics-Simulation and Computation, vol. 47, no. 2, pp. 556–581, 2018.
13. Z. A. Karian, and E. J. Dudewicz, *Fitting statistical distributions: the generalized lambda distribution and generalized bootstrap methods*, CRC press, 2000.
14. H. Lam, K. O. Bowman, and L. R. Shenton, *Remarks on the generalized Tukey's lambda family of distributions*, Chinese Univ. of Hong Kong, Shatin; Union Carbide Corp., Oak Ridge, TN (USA). Nuclear Div.; Georgia Univ., Athens (USA), 1980.
15. Y. C., Lin, and C. Y. Chou, *On the design of variable sample size and sampling intervals X charts under non-normality*, International Journal of Production Economics, vol. 96, no. 2, pp. 249–261, 2005.
16. D. C. Montgomery, *Introduction to statistical quality control*, John Wiley & Sons (New York), 2012.
17. D. C. Montgomery, C. L. Jennings, and M. E. Pfund, *Managing controlling and improving quality*, Philosophy, vol. 31, pp. 2–3, 2010.
18. M., Nili Ahmadabadi, Y. Farjami, and M. B. Moghadam, *Approximating distributions by extended generalized lambda distribution (XGLD)*, Communications in Statistics-Simulation and Computation, vol. 41, no. 1, pp. 1–23, 2012a.
19. M., Nili Ahmadabadi, Y. Farjami, and M. B. Moghadam, *A process control method based on five-parameter generalized lambda distribution*, Quality & Quantity, vol. 46, no. 4, pp. 1097–1111, 2012b.
20. S. Pal, *Evaluation of nonnormal process capability indices using generalized lambda distribution*, Quality Engineering, vol. 17, no. 1, pp. 77–85, 2004.
21. J. S. Ramberg, and B. W. Schmeiser, *An approximate method for generating asymmetric random variables*, ommunications of the ACM, vol. 17, no. 2, pp. 78–82, 1974.
22. J. M. Sarabia, *A hierarchy of lorenz curves based on the generalized tukey's lambda distribution*, Econometric Reviews, vol. 16, no. 3, pp. 305–320, 1997.
23. S. S. Shapiro, and M. B. Wilk, *An analysis of variance test for normality (complete samples)*, Biometrika, vol. 52, no. 3/4, pp. 591–611, 1965.
24. T. A. Spedding, and P. Rawlings, *Non-normality in statistical process control measurements*, International Journal of Quality & Reliability Management, vol. 11, no. 6, pp. 27–37, 1994.
25. A. Tarsitano, *Estimation of the generalized lambda distribution parameters for grouped data*, Communications in Statistics, Theory and Methods, vol. 34, no. 8, pp. 1689–1709, 2005.
26. S. Teyarachakul, S. Chand, and J. Tang, *Estimating the limits for statistical process control charts: A direct method improving upon the bootstrap*, European Journal of Operational Research, vol. 178, no. 2, pp. 472–481, 2007.
27. J. W. Tukey, *The future of data analysis*, The Annals of Mathematical Statistics, vol. 33, no. 1, pp. 1–67, 1962.
28. H. Wang, *Bounded dynamic stochastic systems: modelling and control*, Springer Science & Business Media, 2012.
29. L. Wang, and F. Qian *Shaping the PDF of the state variable based on piecewise linear control for non-linear stochastic systems*, Science China Information Sciences, vol. 59, no. 11, pp. 102–112, 2016.