



Pareto Distribution-Based Shewhart Control Chart for Early Detection of Process Mean Shifts

Aamir Saghir¹ · Gadde Srinivasa Rao² · Muhammad Aslam³  · Azhar Ali Janjua⁴

Received: 6 September 2023 / Accepted: 16 January 2024 / Published online: 19 February 2024
© The Author(s) 2024

Abstract

The Pareto distribution is of paramount importance in actuarial science, wealth distribution, finance, etc. This paper introduces a control chart inspired by Shewhart's methodology, designed for monitoring shifts in the Pareto distribution through a repetitive sampling approach. The chart employs a modified statistic that combines shape and threshold parameters as its plotting statistic. Coefficients for the Shewhart-type Pareto chart are computed for two-phase limits. The performance of the suggested chart is assessed in terms of run length characteristics, assuming a shift in the process mean. Additionally, we conduct an efficiency comparison with existing control charts. The findings suggest that, on average, the proposed Pareto chart demonstrates greater efficiency in promptly detecting changes compared to alternative methods. To illustrate the practical application of our approach, we present an example using revenue data.

Keywords Control chart · Pareto distribution · Repetitive sampling · Performance measure · Average run length

✉ Muhammad Aslam
aslam_ravian@hotmail.com

Aamir Saghir
aamirstat@yahoo.com

Gadde Srinivasa Rao
gaddesrao@gmail.com

Azhar Ali Janjua
azharjanjua12@gmail.com

- ¹ Department of Statistics, Mirpur University of Science and Technology (MUST), Mirpur 10250, AJK, Pakistan
- ² Department of Mathematics and Statistics, the University of Dodoma, PO Box: 338, Dodoma, Tanzania, The University of Dodoma, P.O. Box: 259, Dodoma, Tanzania
- ³ Department of Statistics, Faculty of Science, King Abdulaziz University, 21551 Jeddah, Saudi Arabia
- ⁴ Deputy Director Colleges Hafizabad Higher Education Department, Punjab, Pakistan, Higher Education Department, Hafizabad, Pakistan

1 Introduction

Control charts are commonly used in statistical process control (SPC) monitoring. Typically, the assumption is made that the quality of interest follows a normal distribution. However, in reality, this assumption is often not met, as discussed by various authors, including Saghir and Lin [12]. In such cases, relying on existing control charts may mislead industrial engineers and lead to an increase in non-conforming items, see Montgomery [9]. Many authors in the literature have worked on the development of variable control chart based on non-normal distributions including for example, Bai and Choi [3], Al-Oraini and Rahim [1], Lin and Chou [7], McCracken and Chakraborti [8], Saghir and Lin [12] and Saghir et al. [11].

Control charts, taking into consideration the assumption that the quality characteristic adheres to a Pareto distribution, characterized by its exceptionally heavy tails, have been formulated by researchers such as Guo and Wang [5], Nasiru [10], and Balamurali and Jeyadurga [4]. These studies collectively underscore the significance of Pareto distribution in the realm of statistical process monitoring. Aslam et al. [2] introduced a variance control chart utilizing repetitive sampling, specifically when the average sample size aligns with a fixed sample size. Saleh et al. [13] similarly advocated for the use of repetitive sampling under the condition of the average sample size being equal to the fixed sample size.

The Shewhart control chart's primary drawback is its inadequate capacity to identify even minute variations in the process mean. Further interpretation rules have been proposed to improve the Shewhart chart's ability to identify minute changes in the process mean. Repetitive sampling is used to create a Pareto control chart, which helps overcome this type of problem. Upon a thorough examination of the existing literature and to the best of our knowledge, there is a noticeable absence of research concerning the control chart for the Pareto distribution employing repetitive sampling. This paper addresses this gap by concentrating on the development of a Shewhart Pareto control chart within a repetitive sampling framework. Building upon the groundwork laid by Guo and Wang [5], we extend their work by incorporating repetitive sampling. The goal of this study is to design a control chart for the Pareto distribution using repetitive sampling, particularly when the average sample size aligns with a fixed sample size. Subgroups in sample plans are chosen to decide how much of the plan to accept or reject, whereas subgroups in control charts are used to monitor process variance when the process is intact or in flow. Based on the outcomes of the selected subgroup and the pattern of behavior of earlier subgroups, the quality supervisor is instructed to either leave the process as it is or delay the subgroup and make process adjustments. They address a change in the process configuration when many samples are delayed because they are not plotted. The suggested method is to cap the number of postponed subgroups in repeat sampling. These limits found in the literature fall between 2-sigma limits and 3-sigma, or, in many cases, they are inside 1-sigma limits in the charts under discussion because there are two additional limits, or repetitive limits, inside the typical control limits in repetitive

sampling-based control charts. The efficacy of the proposed control chart will be demonstrated in comparison to existing counterparts, and we will showcase its practical application.

2 The Proposed Pareto Control Chart Using Repetitive Sampling

Within this section, we introduce a novel framework for a control chart that employs repetitive sampling under the assumption of a Pareto-distributed quality characteristic. As only one type of Shewhart Pareto chart is available in the literature (called the [5] control chart), Therefore, we have made comparisons with it and discussed the results with reference to it. The Shewhart control chart's primary drawback is its inadequate capacity to identify even minute variations in the process mean. Further interpretation rules have been proposed to improve the Shewhart chart's ability to identify minute changes in the process mean. Repetitive sampling is used to create a Pareto control chart, which helps overcome this type of problem. Additionally, we derive the performance measure for this proposed control chart.

2.1 Design of the Proposed Control Chart

Assume that X is a random variable following the Pareto distribution with parameters a and b . Then, its probability density function and the cumulative distribution function are, respectively, given by

$$f(x) = \frac{ab^a}{x^{a+1}}, x > b, a < 0, \quad (1)$$

and

$$F(x) = 1 - \left(\frac{b}{x}\right)^a. \quad (2)$$

The mean and the variance of the underlying distribution are given by $\mu = ab/(a-1)$ for $a > 1$ and $\sigma^2 = ab^2/(a-1)^2(a-2)$ for $a > 2$, respectively.

Further, suppose a random variable $Y = \bar{x} \pm ks$ in the interval, where \bar{x} and s are the sample mean and standard deviation respectively and k is an acceptable constant for a given type-I error. It is well known in the literature that when the underlying distribution is normal, \bar{x} and s are independently distributed as normal and chi-square, respectively and Y is also normally distributed. When the underlying distribution is non-normal, the asymptotic distributions of Y is normal with the means, respectively, given by $\mu_y = \mu \pm k\sigma$ and with the common variance expressed by”.

$$\sigma_y^2 = \frac{\sigma^2}{n} \left[1 + \frac{k^2}{4} (\alpha_4 - 1) \pm k\alpha_3 \right].$$

where α_3 and α_4 represents the measure of the skewness and kurtosis of the underlying distribution and $k=3$ (as the asymptotic distribution of Y is normal).

In our case, the measures of skewness and kurtosis of Pareto distribution, which are the measures for identifying the asymmetric nature of the distribution, are expressed as:and

$$\alpha_3 = \left(\frac{2(1+a)}{a-3} \right) \sqrt{\frac{a-2}{a}} \quad \text{for } a > 3$$

$$\alpha_4 = \frac{3(a-2)(3a^3+a+2)}{a(a-3)(a-4)} \quad \text{for } a > 4. \quad (3)$$

This work develops a Pareto control chart for the random variable Y using the double group sampling scheme for decision making. Therefore, the proposed control chart has double control limits, outer and inner control limits. “The process is declared:

- (i) out-of-control if the plotted statistic Y falls outside the outer control limits
- (ii) in-control if the plotted statistic Y falls within the inner control limits.
- (iii) otherwise, resampling should be made with the same sample size”.

where the outer bound is

$$\begin{aligned} \text{UCL}_1 &= \mu_y + k_1 \sigma_Y \\ \text{LCL}_1 &= \mu_y - k_1 \sigma_Y \end{aligned}$$

and the inner bound is

$$\begin{aligned} \text{UCL}_2 &= \mu_y + k_2 \sigma_Y \\ \text{LCL}_2 &= \mu_y - k_2 \sigma_Y \end{aligned} \quad (4)$$

“where k_1 and k_2 are the control chart coefficients to be calculated for a fixed ARL_0 , sample size (n) and λ . μ_y and σ_Y are the mean and standard deviation of the random variable Y . In case of unknown parameters, the sample average \bar{Y} and the sample dispersion S_y of statistic Y , calculated from the in-control process initial samples are used in Eq. (4)”.

2.2 The Performance Measure

“The average run length (ARL), which is defined as the average number of samples until the process shows an out-of-control indication, is the most commonly used measure for the performance evaluation of the proposed control charts. When the process is in-control, the ARL should be sufficiently large to avoid many false alarms and it is denoted by ARL_0 while, ARL should be sufficiently small when the process is out of control and generally it is denoted by ARL_1 .

2.2.1 In-Control ARL

“The in-control ARL denoted by ARL_0 is defined as:

$$ARL_0 = \frac{1}{P_{out}}, \quad (5)$$

where P_{out} denotes the probability that the process is declared as out-of-control when it is actually in control and in a repetitive sampling scheme defined as”:

$$P_{out} = \frac{P_{out}^0}{1 - P_{Res}}$$

where P_{out}^0 is the probability that the process is declared to be out of control based on a single sample is:

$$P_{out}^0 = P(Y \geq UCL_1 | \mu_Y = \mu_0) + P(Y \leq LCL_1 | \mu_Y = \mu_0) \quad (6)$$

where $\mu_Y = \mu_0$ be in-controlled mean of the Pareto process. As mentioned earlier, the asymptotic distribution of Y is normal with mean $\mu_y = \mu \pm k\sigma$ and variance $\sigma_y^2 = \frac{\sigma^2}{n} [1 + \frac{k^2}{4}(\alpha_4 - 1) \pm k\alpha_3]$. Thus, when the process is in-control,

$$\begin{aligned} P_{out}^0 &= P(Y \geq UCL_1 | \mu_Y = \mu_0) + P(Y \leq LCL_1 | \mu_Y = \mu_0) \\ &= P\left(\frac{Y - \mu_y}{\sigma_y} \geq \frac{UCL_1 - \mu_y}{\sigma_y} | \mu_Y = \mu_0\right) + P\left(\frac{Y - \mu_y}{\sigma_y} \leq \frac{LCL_1 - \mu_y}{\sigma_y} | \mu_Y = \mu_0\right) \\ &= P(Z \geq k_1) + P(Z \leq -k_1) \\ &= 1 - F(., k_1) + F(., k_1) \end{aligned} \quad (7)$$

where Z is a standardized normal variable and $F(., t)$ is the cumulative probability of standard normal random variable Z up to point t .

The probability that the process is declared to be out of control based on a single sample is:

$$P_{in}^0 = P(Y \leq UCL_2 | \mu_Y = \mu_0) - P(Y \leq LCL_2 | \mu_Y = \mu_0) \quad (8)$$

and when the process is in-control

$$P_{in}^0 = P(Z \leq k_2) - P(Z \leq -k_2) = F(., k_2) - F(., -k_2) \quad (9)$$

The probability of resampling, when the process is actually in control, is defined as:

$$\begin{aligned} \text{Probability of resampling} &= P_{Res} = P(UCL_2 < Y < UCL_1 | \mu_Y = \mu_0) \\ &\quad + P(LCL_1 < Y < LCL_2 | \mu_Y = \mu_0). \\ &= P(k_2 < Z < k_1) + P(-k_1 < Z < -k_2) \\ &= F(., k_2) - F(., k_1) + F(., -k_1) - F(., -k_2) \end{aligned} \quad (10)$$

Thus,

$$P_{out} = \frac{P_{out}^0}{1 - P_{Res}} = \frac{1 - F(., k_1) + F(., k_1)}{1 - F(., k_2) + F(., k_1) - F(., -k_1) + F(., -k_2)} \tag{11}$$

The average sample size (ASS) for the proposed control chart is defined by

$$ASS_0 = \frac{n}{1 - P_{Res}} \tag{12}$$

2.2.2 Out-of-Control ARL

Suppose that the process mean does not remain at the in-control level, i. e μ_0 and further let $\mu_1 = \mu_0 + \delta\sigma_y$ be a shifted level of mean value under δ standard units shift depending upon the definition of Y . Then, the out-of-control ARL_1 is:

$$ARL_1 = \frac{1}{P_{out,shift}} \tag{13}$$

where

$$P_{out,shift} = \frac{P_{out,shift}^1}{1 - P_{Res1}} \tag{14}$$

The probability of an out-of-control process when a shift occurs is

$$\begin{aligned} P_{out,shift}^1 &= P(Y > UCL_1 | \mu_Y = \mu_1) + P(Y < LCL_1 | \mu_Y = \mu_1) \\ P_{out,shift}^1 &= P(Z > UCL_1) + P(Z < LCL_1) \\ &= P\left(\frac{Y - \mu_1}{\sigma_y} \geq \frac{UCL_1 - \mu_1}{\sigma_y}\right) + P\left(\frac{Y - \mu_1}{\sigma_y} \leq \frac{LCL_1 - \mu_1}{\sigma_y}\right) \\ &= P(Z \geq k_1 - \delta) + P(Z \leq \delta - k_1) \\ &= 1 - F(., k_1 - \delta) + F(., \delta - k_1) \end{aligned} \tag{15}$$

and the probability of resampling for the shifted mean is:

$$\begin{aligned} P_{Res1} &= P(k_2 - \delta < Z < k_1 - \delta) + P(-k_1 - \delta < Z < -k_2 - \delta) \\ &= F(., k_2 - \delta) - F(., k_1 - \delta) + F(., -k_1 - \delta) - F(., -k_2 - \delta) \end{aligned} \tag{16}$$

Thus,

$$P_{out,shift} = \frac{1 - F(., k_1 - \delta) + F(., \delta - k_1)}{1 - F(., k_2 - \delta) + F(., k_1 - \delta) - F(., -k_1 - \delta) + F(., -k_2 - \delta)} \tag{17}$$

The average sample size (ASS) for the proposed control chart is defined by

$$ASS_1 = \frac{n}{1 - P_{Res1}} \tag{18}$$

The following algorithm was applied to obtain the values of ASS and ARL.

- Step-1:** Fix the values of n and specified values of ARL, say r_0
- Step-2:** Find the values of k_1 and k_2 such that $ARL_0 \geq r_0$ and ASS_0 is close to r_0 .
- Step-3:** Determine ARL_1 and ASS_1 using k_1 and k_2 for various values of shift.

3 Results and Discussion

The proposed control chart depends on two constants k_1 and k_2 , which are called control charting constants. The numerical values of these constants can be determined using the Eq. (11) by fixing the parameter of Pareto distribution and the in-control average run length ARL_0 . Once, the control charting constants are determined and the Pareto control chart is constructed, the performance of the proposed control chart can further be investigated when a shift occurs in the parameter of Pareto distribution. Consider $\mu_1 = \mu_0 + \delta\sigma_y$ be shifted mean value of the process when the shift of amount δ occurs in the process mean in standard units depending upon the definition of $Y = \bar{x} \pm ks$ and $k=3$ is considered in this work. The statistical software *R* has been used in this study for all computations.

The performance of the Shewhart Pareto Chart using a repetitive sampling scheme is evaluated in terms of ARL_1 by fixing the ARL_0 , sample size and parameters of Pareto distributions. Some of the results are reported here in Tables 1, 2, 3 and 4 for various choices. The values of ARL_1 and ASS_1 for any other choice of parameter or sample size or of ARL_0 can be obtained easily using the Eqs. (13) and (18).

Table 1 The ARLs for the proposed chart when $a=4.5, b=2$ and $ARL_0=370$

δ	$n=5$		$n=7$		$n=10$		$n=15$	
	$k_1=2.9231$ $k_2=1.4559$		$k_1=2.8985$ $k_2=1.4806$		$k_1=2.7345$ $k_2=1.4810$		$k_1=3.2881$ $k_2=1.6807$	
	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1
1.00	370.42	5.08	370.53	7.09	370.56	10.11	370.66	15.21
0.95	311.17	5.09	337.79	7.12	319.67	10.14	283.71	15.26
0.90	295.92	5.12	280.55	7.15	265.38	10.17	225.05	15.31
0.85	224.97	5.15	218.51	7.18	196.52	10.22	195.29	15.39
0.80	188.53	5.18	159.60	7.23	151.11	10.28	210.49	15.51
0.75	144.98	5.23	120.99	7.29	114.65	10.34	186.49	15.69
0.70	95.99	5.30	102.34	7.37	97.60	10.46	103.06	15.94
0.65	80.00	5.39	102.77	7.52	95.07	10.66	91.17	16.26
0.60	75.21	5.55	81.35	7.74	57.65	10.93	56.06	16.81
0.55	61.91	5.83	51.06	8.10	47.10	11.32	42.52	17.73
0.50	37.25	6.24	34.09	8.62	29.04	11.99	23.03	19.07

Table 2 The ARLs for the proposed chart when $a = 5.0, b = 2$ and $ARL_0 = 370$

δ	$n = 5$ $k_1 = 3.3974$ $k_2 = 1.7061$		$n = 7$ $k_1 = 3.2498$ $k_2 = 1.6925$		$n = 10$ $k_1 = 3.4018$ $k_2 = 1.8121$		$n = 15$ $k_1 = 2.8741$ $k_2 = 1.6611$	
	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1
1.00	370.67	5.10	370.75	7.11	370.43	10.17	372.96	15.16
0.95	332.09	5.12	308.60	7.14	263.38	10.21	250.46	15.20
0.90	264.97	5.16	248.54	7.18	200.19	10.27	203.62	15.26
0.85	199.42	5.19	174.90	7.23	148.05	10.33	166.16	15.32
0.80	158.42	5.25	134.14	7.29	145.98	10.44	131.13	15.41
0.75	114.88	5.32	116.37	7.37	137.95	10.59	119.29	15.57
0.70	109.87	5.43	115.51	7.52	105.31	10.83	91.35	15.82
0.65	90.05	5.58	93.58	7.73	69.50	11.18	60.20	16.11
0.60	77.72	5.89	58.08	8.11	44.72	11.66	47.83	16.67
0.55	40.02	6.27	39.26	8.54	35.97	12.52	30.81	17.55
0.50	28.15	6.61	27.99	9.13	21.94	13.35	18.13	18.96

Table 3 The ARLs for the proposed chart when $a = 4.5, b = 3$ and $ARL_0 = 370$

δ	$n = 5$ $k_1 = 2.2365$ $k_2 = 1.2635$		$n = 7$ $k_1 = 2.2365$ $k_2 = 1.2635$		$n = 10$ $k_1 = 2.4335$ $k_2 = 1.3625$		$n = 15$ $k_1 = 3.1544$ $k_2 = 1.6510$	
	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1
1.00	370.21	5.05	370.82	7.10	370.87	10.09	370.55	15.20
0.95	310.52	5.06	296.19	7.12	291.70	10.10	283.93	15.24
0.90	234.88	5.07	256.91	7.15	279.10	10.14	221.36	15.30
0.85	186.38	5.09	194.04	7.19	220.92	10.17	195.06	15.37
0.80	150.44	5.11	151.26	7.24	163.10	10.22	177.73	15.48
0.75	122.52	5.14	119.26	7.31	127.01	10.28	165.09	15.64
0.70	102.67	5.19	91.17	7.40	97.30	10.36	105.64	15.87
0.65	69.79	5.25	95.81	7.54	81.64	10.48	86.38	16.20
0.60	52.12	5.34	78.42	7.80	72.22	10.73	57.65	16.71
0.55	42.35	5.48	48.74	8.17	45.12	11.05	36.86	17.48
0.50	38.08	5.77	30.62	8.67	31.82	11.57	24.70	18.95

Tables 1, 2, 3 and 4 reveal that

1. The control charting constant (k_1 and k_2) is significantly different than the standard value 3.
2. ARL_1 decreases as shift size decreases but ASS_1 increases slightly for the fixed value of a, b, n and ARL_0
3. ARL_1 decreases as sample size increases for the fixed value of a, b and ARL_0 .

Table 4 The ARLs for the proposed chart when $a=5, b=3$ and $ARL_0=370$

δ	$n=5$ $k_1=3.4275$ $k_2=1.7232$		$n=7$ $k_1=2.9576$ $k_2=1.5713$		$n=10$ $k_1=2.5155$ $k_2=1.4416$		$n=15$ $k_1=3.2365$ $k_2=1.7521$	
	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1	ARL1	ASS1
1.00	370.01	5.10	370.01	7.09	370.39	10.08	370.30	15.20
0.95	327.79	5.12	315.10	7.11	311.02	10.10	284.89	15.25
0.90	256.63	5.16	285.59	7.15	290.21	10.14	212.95	15.31
0.85	203.11	5.20	216.53	7.18	232.51	10.17	187.81	15.39
0.80	145.41	5.25	153.71	7.23	160.93	10.22	183.59	15.52
0.75	111.01	5.33	120.76	7.30	118.27	10.29	164.79	15.73
0.70	107.44	5.44	91.85	7.40	91.65	10.38	93.65	16.00
0.65	91.15	5.60	97.14	7.56	80.61	10.55	77.41	16.37
0.60	75.33	5.92	72.03	7.83	60.98	10.83	49.62	17.03
0.55	39.27	6.30	45.53	8.25	38.80	11.19	31.51	18.06
0.50	28.88	6.66	29.17	8.87	26.35	11.90	19.98	19.83

4. ARL_1 decreases as the location parameter value increase by fixing constant other factors This are expected results as our focus is to detect changes in the location parameter of Pareto distribution by assuming the shape parameter to be control-able.
5. There is no significant effect on ARL_1 when the shape parameter increases for the fixed value of n, a and ARL_0 .
6. There is no significant effect on the performance of the proposed chart in terms of ARL_1 when the simultaneous change in both parameters occurred for fixed value of n and ARL_0 .

Thus, the proposed Pareto control chart can effectively detect the change in location parameter and also efficiency of this chart increase with a moderate sample size. Note that it is desirable situation that the sample size must be close to ASS_1 . It is note from tables that when sample size is small, the value of ASS_1 are little bit away from the fixed sample size when δ is greater than 0.60. It is worth to noting that in case of the repetition, the fixed sample size is taken from the production process.

4 Performance Comparison with Existing Chart

In this section, the proposed control chart will be compared with the existing Shewhart-type Pareto control chart proposed by Guo and Wang [5], when $k_1 = k_2 = L$. “The control chart with smaller ARL is said to be more efficient”. Hence the performance of the proposed control chart will be studied based on ARL values. To defend a comparison, consider when the process is in-control at $ARL_0=370$, the out-of-control ARLs of the Guo and Wang [5] chart and proposed

Pareto control chart for different mean shifts, $n=5, a=5, b=2$ are presented in Table 5.

According to these results, the proposed control chart gives smaller ARL values for different shifts (δ); hence, the proposed control chart is more quick in detecting shifts as compared to the existing Shewhart type Pareto control [5] chart. For example, when $ARL_0=370, n=5, a=5, b=2$ and $\delta=0.7$ from Table 5, we noticed that the proposed control chart $ARL_1=109.87$ whereas $ARL_1=114.74$ for Guo and Wang [5] Pareto control chart. The ARL curves of the proposed Pareto control chart using repetitive sampling and Guo and Wang [5] control chart for the mean of specified shifts are displayed in Fig. 1. After exploring the ARL curves, it is clear that the proposed control charts using repetitive sampling show better performance than the Shewhart type control [5] chart. To accentuate this simulation study will be carried out in the next subsection. Similar results have been observed for other choices of parameters and sample sizes.

5 A Real Example Based on a Tax Revenue Data

The developed control chart is demonstrated by using a real data set in this section, the data set is taken from Klakattawi [6]. The data set is reported as monthly records of tax revenue of Egypt between January 2006 and November 2010, in 1000 million Egyptian pounds. The data along with statistics is reported in Table 6. It is established that the tax revenue data comes from the Pareto distribution with parameters $\hat{a}=1.707$ and $\hat{b}=6.641$ and the maximum distance between the real time data and the fitted Pareto distribution is found from the Kolmogorov–Smirnov test as 0.094 and also the p-value is 0.6978. The demonstration of the goodness of fit for the given model is shown in Fig. 2, the histogram and theoretical density and P–P plots for the Pareto distribution for the tax revenue data. The chart coefficients for the estimated parameters of a and

Table 5 The ARLs for the proposed chart and existing Shewhart type Pareto control chart when $n=5, a=5, b=2$ & $ARL_0=370$

δ	Guo and Wang [5] control chart $L=4.4198$	Proposed chart $k_1=3.3974, k_2=1.7061$
1.00	370.49	370.67
0.95	339.39	332.09
0.90	292.54	264.97
0.85	239.43	199.42
0.80	183.12	158.42
0.75	149.83	114.88
0.70	127.23	109.87
0.65	114.51	90.05
0.60	91.83	77.72
0.55	59.44	40.02
0.50	37.59	28.15

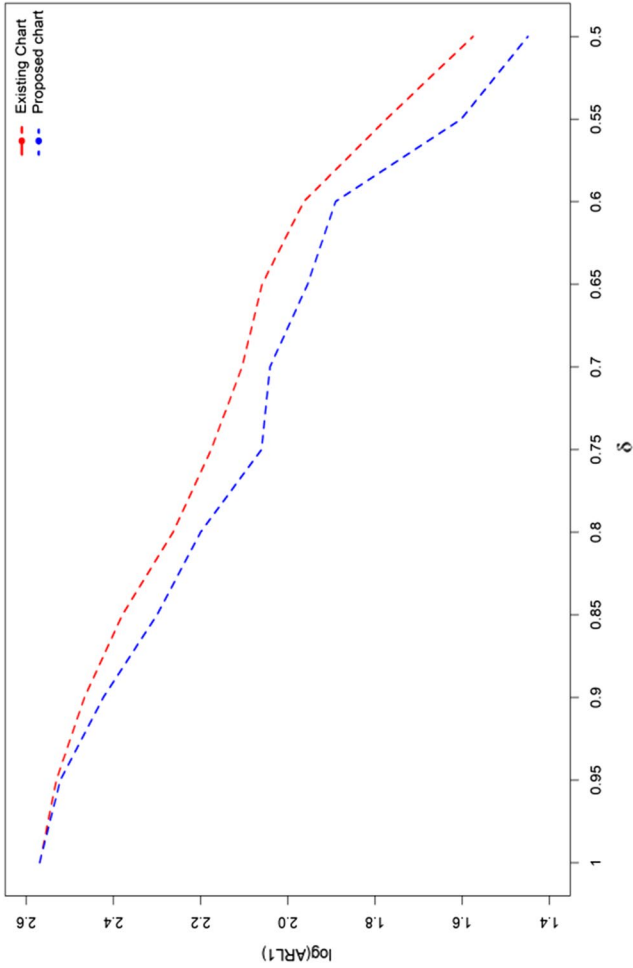


Fig. 1 Logarithmic ARL_1 curves of proposed control chart and Shewhart type control chart when $n = 5$, $a = 5$, $b = 2$ & $ARL_0 = 370$

Table 6 The monthly records of tax revenue of Egypt along with statistics

S. no.	Sample				\bar{x}	s	$Y = \bar{x} - ks$	$Y = \bar{x} + ks$	
1	5.9	8.5	8.5	19.1	11.2	10.64	5.0870	-4.6211	25.9011
2	20.4	21.6	8.0	20.5	6.1	15.32	7.5939	-7.4616	38.1016
3	14.9	18.5	9.2	7.1	8.4	11.62	4.8669	-2.9808	26.2208
4	16.2	5.1	26.2	7.7	11.0	13.24	8.3470	-11.8011	38.2811
5	17.2	6.7	21.9	18.1	11.6	15.1	5.9678	-2.8035	33.0035
6	7.8	17.0	16.7	16.5	11.9	13.98	4.0431	1.8506	26.1094
7	6.1	8.6	21.3	11.9	5.2	10.62	6.5113	-8.9139	30.1539
8	9.2	9.7	35.4	7.0	6.8	13.62	12.2434	-23.1103	50.3503
9	10.2	15.7	14.3	8.6	8.9	11.54	3.2532	1.7805	21.2995
10	9.6	9.7	8.5	12.5	7.1	9.48	1.9880	3.5161	15.4439
11	13.3	10.0	10.6	10.3	10.8	11	1.3210	7.0370	14.9630

b for $n=5$ and $ARL=370$ are $k_1=9.8274$ and $k_2=0.8366$. Using the chat coefficients the inner bound limits are $LCL_2 = -3.9007$ and $UCL_2 = 28.6571$; and the outer bound limits are $LCL_1 = -178.847$ and $UCL_1 = 203.6039$. In Fig. 3 the proposed Pareto control chart using repetitive sampling for tax revenue data is presented. The visual presentation shows that some points are in between inner and outer control limits; hence repetition of sampling should be made with the same sample size.

6 Application Using Simulated Tax Revenue Data

In this section, using simulated data we study the performance of the proposed control chart. To establish the performance of the proposed control chart as compared with existing control charts, 40 samples are generated from normal distribution. The first 20 subgroups of each size 5 are generated from the Pareto distribution with in-control parameters $a=5$, $b=2$ and the last 20 subgroups of each size 5 are generated from the Pareto distribution with in-control parameters $a=5$ and out-of-control $b_1 = \delta b = 0.8 \times 2 = 1.6$. That is, the process mean is shifted after 20 subgroups with a variance shift of $\delta=0.8$.

The chart coefficients when $n=5$ for $ARL_0=370$ are available in Table 5. The proposed control charts using repetitive sampling with $k_1=3.3974$, $k_2=1.7061$ is displayed in Fig. 4 and the Shewhart type control chart with $L=4.4198$ is presented in Fig. 5. From Fig. 4, the Shewhart type control chart fails to detect the shift. From Fig. 5, it can be observed that the proposed control charts under a repetitive sampling scheme show the out-of-control signals at subgroup numbers 28, 30, 36 and 39. This example shows that the proposed control chart under a repetitive sampling scheme is more effective in discovering process shift as compared to the Shewhart type control chart scheme.

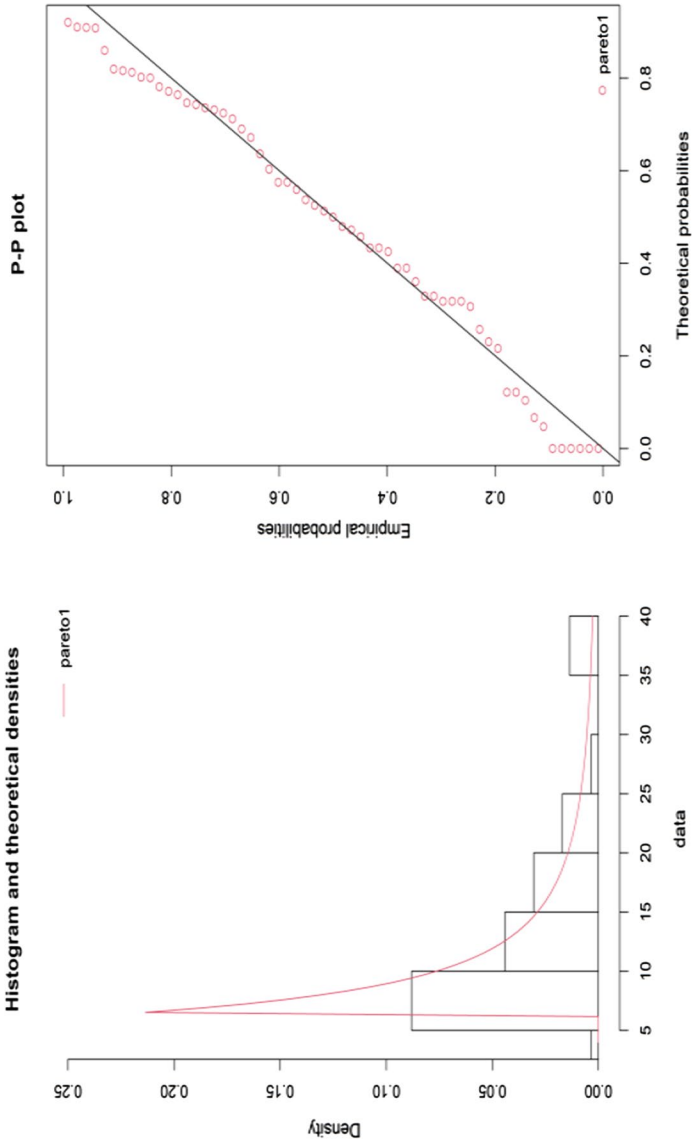


Fig. 2 Visual presentation of tax revenue data

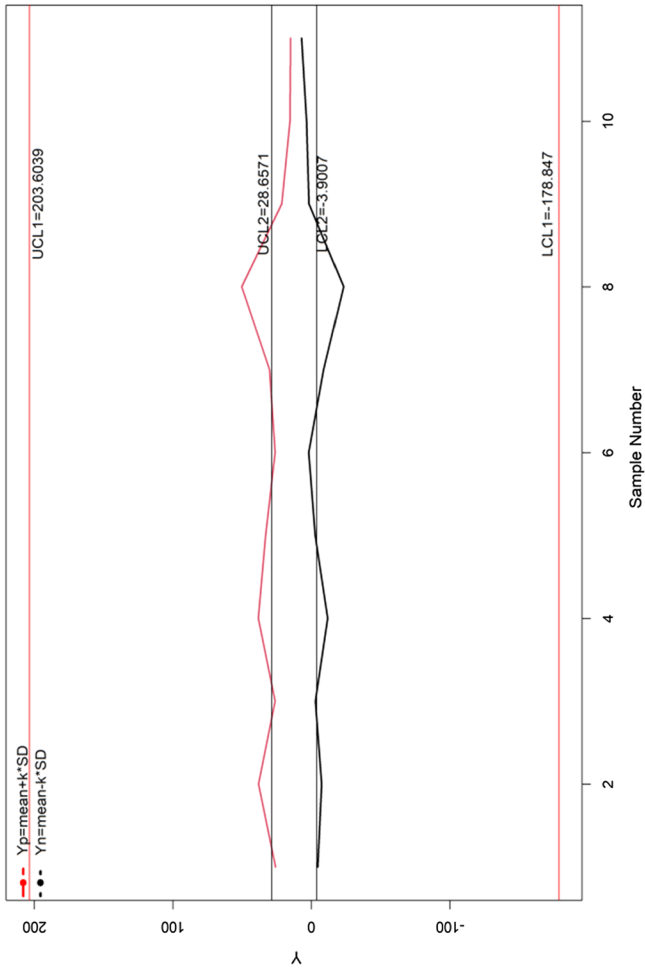


Fig. 3 Proposed Pareto control chart using repetitive sampling for tax revenue data

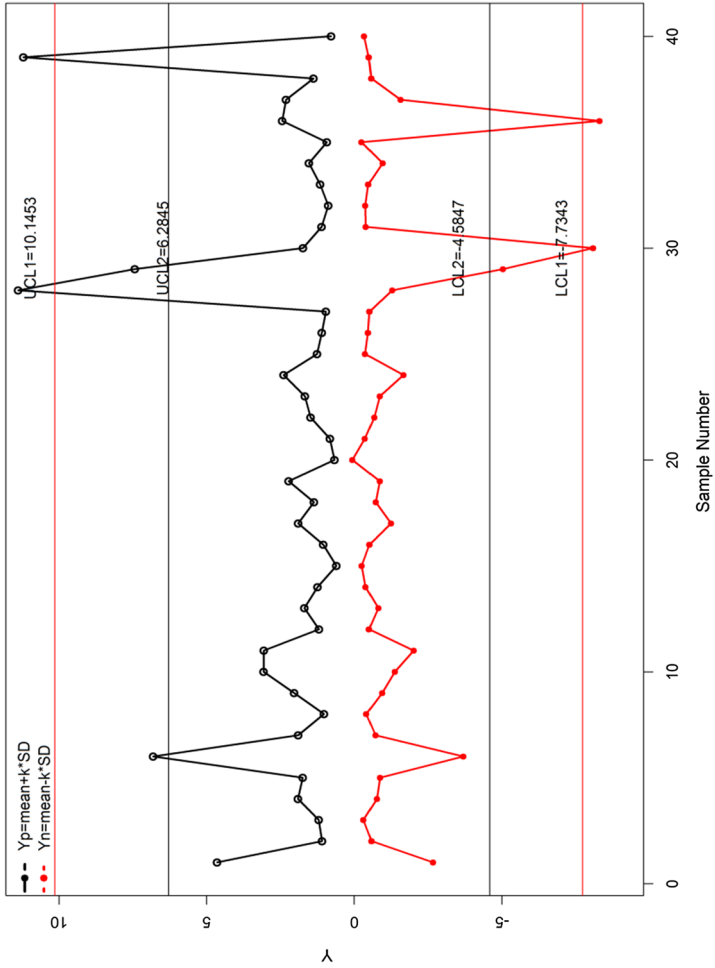


Fig. 4 Proposed Pareto control chart using repetitive sampling for simulated data

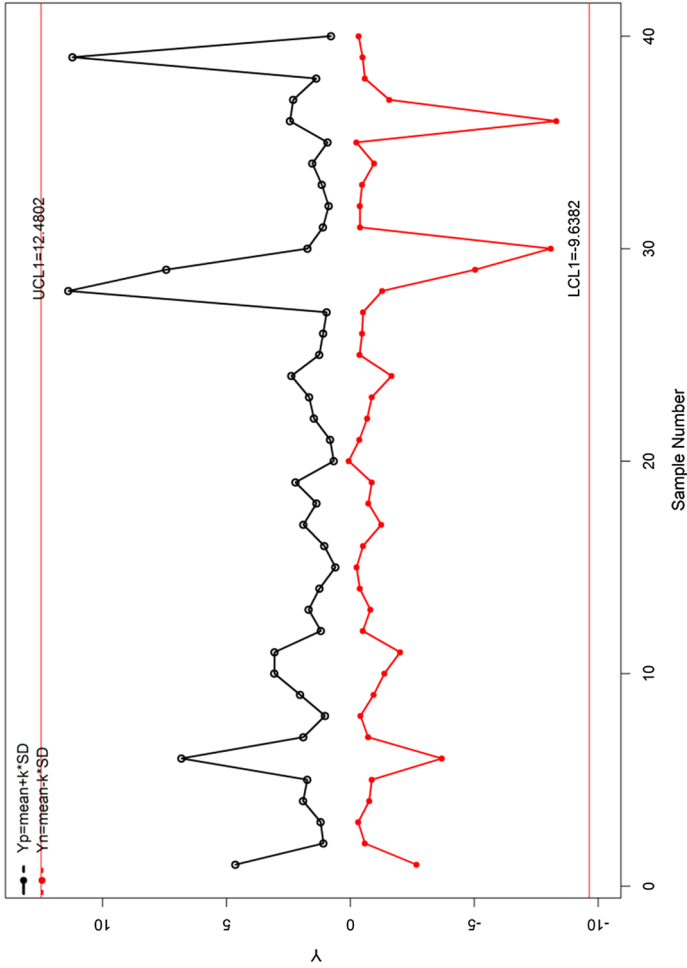


Fig. 5 Existing Shewhart type Pareto Guo and Wang [5] control chart for simulated data

7 Conclusions and Recommendations

In conclusion, this paper has introduced a novel Shewhart Pareto control chart designed for monitoring shifts in the Pareto distribution through a repetitive sampling approach. The chart, utilizing a modified statistic that combines shape and threshold parameters, has been evaluated for its performance in terms of run length characteristics, assuming a shift in the process mean. Through a comprehensive efficiency comparison with existing control charts, our findings indicate that the proposed Pareto chart demonstrates superior efficiency in promptly detecting changes, on average, compared to alternative methods. The practical application of our approach has been exemplified through an illustrative example employing revenue data. We have highlighted the significance of Pareto distribution in statistical process monitoring, particularly emphasizing the importance of considering heavy-tailed distributions in control chart design. The extension of the work by incorporating repetitive sampling adds depth to the methodology. The absence of prior research on control charts for the Pareto distribution employing repetitive sampling was addressed in this paper, contributing to the existing body of knowledge in statistical process control. By building upon the groundwork of [5], we have presented a comprehensive design for a Shewhart Pareto control chart under a repetitive sampling scheme. As we move forward, this work opens avenues for further exploration and refinement of control chart methodologies under non-normal distributions, ensuring robust and accurate statistical process control in diverse industrial settings. The proposed control chart faces a constraint as it is not suitable for data conforming to a normal distribution. Subsequent research endeavors could focus on developing and exploring an exponentially weighted moving average (EWMA) scheme tailored for the Pareto distribution through repetitive sampling. Additionally, investigating the efficacy of the proposed Pareto chart by employing a cost model represents a promising avenue for future research.

Acknowledgements The authors are deeply thankful to the reviewers and the editor for their valuable suggestions to improve the quality and presentation of the paper.

Author Contributions AS, GSR, MA and AAJ wrote the paper.

Funding Not applicable.

Availability of Data and Materials The data is given in the paper.

Declarations

Conflict of interest No conflict of interest regarding the paper.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article

are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Al-Oraini, H.A., Rahim, M.: Economic statistical design of X control charts for systems with Gamma ($\lambda, 2$) in-control times. *Comput. Ind. Eng.* **43**(3), 645–654 (2002)
2. Aslam, M., Khan, N., Jun, C.-H.: A new S 2 control chart using repetitive sampling. *J. Appl. Stat.* 1–12 (2015). (**ahead-of-print**)
3. Bai, D., Choi, I.: X and R control charts for skewed populations. *J. Qual. Technol.* **27**(2), 120–131 (1995)
4. Balamurali, S., Jeyadurga, P.: An attribute np control chart for monitoring mean life using multiple deferred state sampling based on truncated life tests. *Int. J. Reliab. Qual. Saf. Eng.* **26**(01), 1950004 (2019)
5. Guo, B.-C., Wang, B.-X.: Control charts for the Pareto distribution. *Appl. Math. J. Chin. Univ.* **30**(4), 379–396 (2015)
6. Klakattawi, H.S.: The Weibull-gamma distribution: properties and applications. *Entropy* **21**(5), 438 (2019)
7. Lin, Y.-C., Chou, C.-Y.: Non-normality and the variable parameters X control charts. *Eur. J. Oper. Res.* **176**(1), 361–373 (2007)
8. McCracken, A., Chakraborti, S.: Control charts for joint monitoring of mean and variance: an overview. *Quality Technol. Quant. Manag.* **10**(1), 17–36 (2013)
9. Montgomery, D.C.: *Introduction to Statistical Quality Control*. Wiley, New York (2020)
10. Nasiru, S.: One-sided cumulative sum control chart for monitoring shifts in the shape parameter of Pareto distribution. *Int. J. Product. Quality Manag.* **19**(2), 160–167 (2016)
11. Saghir, A., Akber Abbasi, S., Faraz, A.: The exact method for designing the Maxwell chart with estimated parameter. *Commun. Stat.-Simul. Comput.* **50**(1), 270–281 (2021)
12. Saghir, A., Lin, Z.: Designing of Gini-chart for exponential, t, logistic and Laplace distributions. *Commun. Stat.-Simul. Comput.* **44**(9), 2387–2409 (2015)
13. Saleh, N.A., Mahmoud, M.A., Woodall, W.H.: A re-evaluation of repetitive sampling techniques in statistical process monitoring. *Quality Technol. Quant. Manag.* 1–19 (2023). <https://doi.org/10.1080/16843703.2023.2246770>